SSAS: Spatiotemporal Scale Adaptive Selection for Improving Bias Correction on Precipitation

Yiqun Liu, Junping Zhang, Lei Chen, Hai Chu, James Z. Wang, Leiming Ma

***Abstract*—By utilizing physical models of the atmosphere collected from current weather conditions, the Numerical Weather Prediction (NWP) model developed by the European Centre for Medium-range Weather Forecasts (ECMWF) can provide the indicators of severe weather such as heavy precipitation for an early-warning system. However, the performance of precipitation forecasts from ECMWF often suffers from considerable prediction biases owing to the high complexity and uncertainty for the formation of precipitation. The Bias Correcting on Precipitation (BCoP) was thus utilized for correcting these biases via forecasting variables including the historical observations and variables of precipitation, and these variables as predictors from ECMWF are highly relevant to precipitation. Existing BCoP methods, such as Model Output Statistics (MOS) and Ordinal Boosting Autoencoder (OBA), do not take advantage of both spatiotemporal dependencies of precipitation and scales of related predictors that can change with different precipitation. We propose an end-to-end deep-learning BCoP model, named the Spatiotemporal Scale Adaptive Selection (SSAS) model, to automatically select the spatiotemporal scales of the predictors via Spatiotemporal Scale-Selection Modules (S3M/TS2M) for acquiring the optimal high-level spatiotemporal representations. Qualitative and quantitative experiments carried out on two benchmark datasets indicate that SSAS can achieve state-of-the-art performance, compared with eleven published BCoP methods, especially on heavy precipitation.**

I. INTRODUCTION

SEVERE weather events such as flooding caused by heavy precipitation leave a trail of damages and destruction in their wake. For example, severe floods in China in June and July of 2019 killed over 200 people and caused more than 7 billion dollars lost [1]. For mitigating the damage and avoiding loss of life, a series of advanced Weather Prediction Systems (WPS) has been developed in the past decades. Unlike classic WPS focusing on the theory of meteorology and experience, Numerical Weather Prediction (NWP) based on the set of kinetic and thermodynamic equations [2] can obtain spatiotemporal gridded quantitative analysis and thus more objective results. Among various NWP systems, the Integrated Forecasting System (IFS) [3] is a representative model developed by the European Centre for Medium-range Weather Forecasts (ECMWF).[1] Especially, IFS has been widely used by meteorological observatories in China [4].

Quantitative precipitation forecasts of IFS can provide early warnings of heavy rainfalls for preventing floods [2]. Nevertheless, the dynamic mechanism of the formation of precipitation is uncertain and sophisticated [5]. Besides, the value range of precipitation takes on more maldistribution than other statistical postprocessing,[1] as shown in Fig. 1. Moreover, the forecast of NWP in the long-range (100+ hours) will suffer considerable biases because of accumulated errors from all earlier forecasts [6]. It is a challenging task to design a new Bias Correction on Precipitation (BCoP) for improving the precipitation accuracy from NWP.

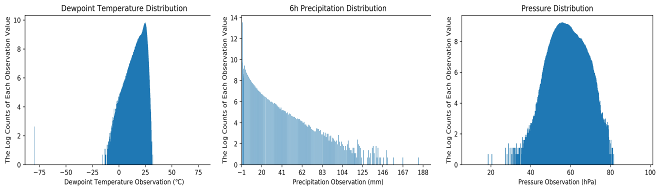


Fig. 1. Ranging from January to December of 2016 and 2017, the distributions of dewpoint temperature, pressure, and 6-hour precipitation observations consist of a total of 886,872 samples. The horizontal axis is the value range of observation and the vertical axis represents the logarithm of the number of observations. On the one hand, as for dewpoint and pressure, except for few outliers in the extremely low dewpoint temperature and pressure cases, their main distributions are close to Gaussian distribution and the value range is in a short interval (*e.g.*, $[-15, 30]^{\circ}C$). On the other hand, the distribution of precipitation displays the maldistribution and long-span value range ($[0, 188]mm$). For example, there are numerous rainless ($[0.1, 1)mm$) samples [4] but few rainstorm (above $150mm$) samples.

As a branch of statistical post-processing method[1], the BCoP is often utilized for correcting the predictive precipitation bias to the observation. The traditional BCoP represented by Model Output Statistics (MOS) [7] can be roughly categorized into Frequency Matching Method (FMM) and Multiple Linear Regression (MLR), according to the number of predictors. FMM is an approximation approach for fitting

---

Manuscript received xx x, 2021; revised xx x, 2021 and xx x, 2021; accepted xx xx, 2021. This work was supported by Science and Technology Commission of Shanghai Municipality (18DZ1200404), the National Key Research and Development Program of China (2018YFB1305104), and the Shanghai Municipal Science and Technology Major Project (2018SHZDZX01), ZJLab, and The Pennsylvania State University (J. Z. Wang). (Corresponding author: Junping Zhang.)

Yiqun Liu, Junping Zhang are with Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai, 200433, China. Email: {yqliu17, jpzhang}@fudan.edu.cn

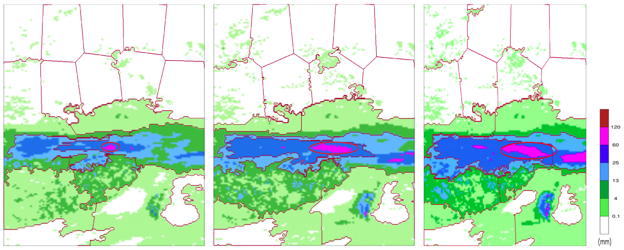
L. Chen, H. Chu, and L. Ma are with the Shanghai Central Meteorological Observation, China. Emails: qqydss@163.com, labyrin@qq.com, malm@typhoon.org.cn

J. Z. Wang is with the College of Information Sciences and Technology, The Pennsylvania State University, USA. He contributed to discussions and revising the manuscript. Email: jwang@ist.psu.edu

This paper has supplementary downloadable source codes provided by the authors. They are available at https://github.com/EchoItLiu/xxx .

[1] **Medium range** is the meteorological forecasting range beyond 72-hour ensemble forecasting and up to 240-hour description of weather parameters; **Predictors** are the main physical variables of weather forecasts from ECMWF; **Statistical postprocessing** is a method of improving numerical weather models' ability to forecast by relating model outputs to observational or additional model data using statistical methods.

the frequency distribution between correcting precipitation and observation. Concretely, FMM could learn a transfer (correction) function for simulating the mapping from the given distribution of precipitation to the distribution of observation, and then new precipitation would be corrected by coefficients of the transfer function [8]. In FMM, there are three kinds of transfer functions including linear function [9], non-linear function [10], and quantile mapping function [11]. Although they each have their respective advantages, these functions have several common limitations. For example, precipitation itself is regarded as the only predictor for simulating the transfer function, which is not easy to fit precipitation distribution because the formation of rainfalls is subject to multiple meteorological predictors [6]. Besides, the MLR approach could utilize the multiple precipitation-related predictors as inputs for obtaining the corrected prediction. Unlike FMM with a single predictor, MLR can make use of the correlation between multiple predictors and precipitation [12]. However, MLR has low generalizability to deal with the predictors from different NWP systems and variational spatial resolution [13].



(a) $t-2$ (b) $t-1$ (c) t

Fig. 2. **The visualization of precipitation delivered from IFS in the 3 continuous timestamps with region segmentation**. The Simple Linear Iterative Clustering [14] method is employed for segmenting different precipitation subregions by clustering the pixels possessing similar semantic information. The precipitation with similar color is enclosed to the same subregion by the purple irregular grid, which can reflect the spatial dependency contained in precipitation. Besides, the scope of the same modena precipitation marked by a red oval gradually becomes larger over time, and this reflects temporal dynamical relevance in precipitation. The colormap on the side indicates different levels of rainfall intensity.

Artificial Neural Network (ANN), known as a universal approximator, is capable of theoretically approximating any continuous input and output mapping. Compared with FMM and MLR, ANN [6] simultaneously inherits the advantages of linear and non-linear (sigmoid function) transformation [15]. In addition to that, ANN has the ability of automatic feature extraction from predictors.

However, conventional ANN cannot capture spatiotemporal (ST) weather patterns contained in predictors. These ST weather patterns including the precipitation have a spatial dependency and vary over time [16]. Specifically, we visualize the ST variation of precipitation illustrated in Fig. 2. Deep learning models based on station-wise BCoP[2], such as Convolutional Neural Network (CNN) [17], Long Short-Term Memory (LSTM) [18], and Ordinal Boosting Autoencoder (OBA) [19] can extract spatial or temporal patterns from predictors [4], [5], [20]. However, they heuristically select the spatial or temporal scale of predictors around the given station to be a fixed size, which is unreasonable. Due to the intricate mechanism of precipitation, the formation of different precipitation is determined by predictors that have different ST scales [21]–[23].

To address these issues, we take into account the selection strategy of the adaptive ST scales of predictors correlated with corresponding precipitation. We propose a deep ST Scale Adaptive Selection (SSAS) model to capture the optimal ST features from station-wise predictors through selecting the adaptive ST scales of them. Concretely, two pre-trained modules termed Spatial Scale-Selection Module (S3M) and Temporal Scale-Selection Module (TS2M) are added for selecting ST scales and fine-tuned by end-to-end optimization. The predominance of SSAS is summarized as follows:

- **Adaptive ST feature selection strategy:** SSAS applies the proposed S3M and TS2M to select the different ST scales for station-wise predictors, which can learn the high-level ST discriminative representations for BCoP.
- **Effectiveness on correcting the heavy precipitation:** Qualitative and quantitative experiments indicate that our model achieves better correction performance than the eleven published machine learning methods on two benchmark datasets and precipitation forecasts before correcting, especially for correcting heavy precipitation.

The remainder of the paper is organized into five sections. Section II surveys the relevant techniques involved in our method. Section III describes two benchmark datasets in our papers and preprocessing procedure. Section IV details our SSAS framework. Section V provides comprehensive evaluations by comparing the SSAS with twelve published machine learning methods on two benchmark datasets and precipitation forecasts before correcting. Finally, Section VI summarizes the main conclusions and discusses future work.

II. RELATED WORK

A. Traditional Bias Correcting on Precipitation

Compared with forecasts for other variables such as temperature, sophisticated and chaotic mechanisms of formation of rainfalls will lead to a larger bias for precipitation forecast. Therefore, the Bias Correcting on Precipitation (BCoP) has a recent surge in demand. The traditional BCoP known as Model Output Statistics (MOS) [7], [8], [12], [13] utilized the precipitation-related predictors and historical observations of precipitation for creating the correction function for fitting the distribution of given precipitation. The MOS provides a good

---

[2] **Station-wise BCoP** is a fine-grained method that corrects the precipitation in units of every station from a specific region; **Our datasets** are available at https://www.ecmwf.int/en/forecasts/documentation-and-support (HR-ECMWF) and https://cds.climate.copernicus.eu/ (ERA5); **The time of TS-Predictors** is Coordinated Universal Time(UTC) while MICAPS time is Beijing Time (BJT). $BJT = UTC + 8$.

basis for the research of deep-learning BCoP and is introduced in detail in paragraph 3 of Sec I.

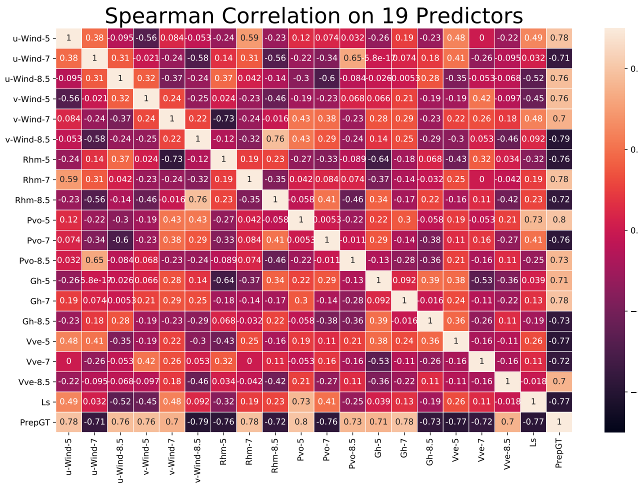
B. Deep-Learning-Based Bias Correcting on Precipitation

The deep-learning approaches represented by CNN [17] and LSTM [18] are capable of extracting deep features from big data in an end-to-end manner. Until now, by learning the ST features, the deep-learning models have been widely applied to many downstream tasks in computer vision such as action recognition [24], sentiment analysis [25], electroencephalogram classification, and prediction of traffic flow [26]. Further, Dual-stage Attention-based Recurrent Neural Network (DA-RNN) [27] is the deep ST model based on an advanced attention mechanism to adaptively extract the relevant-driving temporal features locally and globally in the time dimension and is applied to indoor temperature forecasting. Meanwhile, the T-GCN [28] can combine with the Graph Convolutional Network (GCN) and the Gated Recurrent Unit (GRU) to apply to traffic prediction by learning the complex topological structures and dynamic changes.

In the weather forecasting domain, the Convolutional LSTM (ConvLSTM) [29]was proposed to construct an end-to-end trainable model for the precipitation nowcasting through extracting ST features from radar echo data on precipitation [30]. Because the physical fields of predictors contain the rich time-space dynamical information naturally, Therefore, LSTM and CNN could obtain better prediction performance than ANN [10], [25] on BCoP by capturing the temporal and spatial dependency. Further, Miao et al. combined CNN and LSTM to correct precipitation based on atmospheric dynamical fields from ERA-Interim [31]. Benefitting from ST dependency of the physical fields, these deep learning-based BCoP methods obtain higher accuracy in medium-range precipitation forecasts [5]. Therefore, it is interesting to exploit the potential of ST patterns as far as possible for a better weather forecast.

C. Adaptive Strategy

Generally speaking, most of the deep network models utilize some heuristic operations such as empirically building network architecture based on the off-the-shelf modules, which possibly affects the performance of the network. To address this issue, the automatic selection of network architecture is implemented by employing a novel Neural Architecture Search (NAS) strategy [32]. Further, some adaptive strategies can be integrated into the end-to-end network. For example, Zhang *et al.* proposed a tracking framework with a scale-adaptive strategy [33] to handle the scale estimation problem and reduce the effect of the background of images by adaptively adjusting the size of filters according to the different tracking targets. Meanwhile, the multiagent system with an adaptive event-triggered control [34] is utilized for reducing the controllers' update frequency to shorten the computational time. Besides, Feature Selective Anchor-Free (FSAF) module [35] was proposed for addressing the limitation of heuristic-guided feature selection from object detection via an anchor-free adaptive strategy. As discussed above, a deep model with an adaptive strategy is a new trend because of less human intervention in virtue of end-to-end superiority.



	u-Wind-5	u-Wind-7	u-Wind-8.5	v-Wind-5	v-Wind-7	v-Wind-8.5	Rhm-5	Rhm-7	Rhm-8.5	Pvo-5	Pvo-7	Pvo-8.5	Gh-5	Gh-7	Gh-8.5	Vve-5	Vve-7	Vve-8.5	Ls	PrepGT
u-Wind-5	1	0.38	-0.095	-0.56	0.084	-0.053	-0.24	0.59	-0.23	0.12	0.074	0.032	-0.26	0.19	-0.23	0.48	0	-0.22	0.49	0.78
u-Wind-7	0.38	1	0.31	-0.021	-0.24	-0.58	0.14	0.31	-0.56	-0.22	-0.34	0.65	5.8e-17	0.074	0.18	0.41	-0.26	-0.095	0.032	-0.71
u-Wind-8.5	-0.095	0.31	1	0.32	-0.37	-0.24	0.37	0.042	-0.14	-0.3	-0.6	-0.084	-0.026	0.0053	0.28	-0.35	-0.053	-0.068	-0.52	0.76
v-Wind-5	-0.56	-0.021	0.32	1	0.24	-0.25	0.024	-0.23	-0.46	-0.19	-0.23	0.068	0.066	0.21	-0.19	-0.19	0.42	-0.097	-0.45	0.76
v-Wind-7	0.084	-0.24	-0.37	0.24	1	0.22	-0.73	-0.24	-0.016	0.43	0.38	-0.23	0.28	0.29	-0.23	0.22	0.26	0.18	0.48	0.7
v-Wind-8.5	-0.053	-0.58	-0.24	-0.25	0.22	1	-0.12	-0.32	0.76	0.43	0.29	-0.24	0.14	0.25	-0.29	-0.3	0.053	-0.46	0.092	-0.79
Rhm-5	-0.24	0.14	0.37	0.024	-0.73	-0.12	1	0.19	0.23	-0.27	-0.33	-0.089	-0.64	-0.18	0.068	-0.43	0.32	0.034	-0.32	-0.76
Rhm-7	0.59	0.31	0.042	-0.23	-0.24	-0.32	0.19	1	-0.35	0.042	0.084	0.074	-0.37	-0.14	-0.032	0.25	0	-0.042	0.19	0.78
Rhm-8.5	-0.23	-0.56	-0.14	-0.46	-0.016	0.76	0.23	-0.35	1	-0.058	0.41	-0.46	0.34	-0.17	0.22	-0.16	0.11	-0.42	0.23	-0.72
Pvo-5	0.12	-0.22	-0.3	-0.19	0.43	0.43	-0.27	0.042	-0.058	1	0.0053	-0.22	0.22	0.3	-0.058	0.19	-0.053	0.21	0.73	0.8
Pvo-7	0.074	-0.34	-0.6	-0.23	0.38	0.29	-0.33	0.084	0.41	0.0053	1	-0.011	0.29	-0.14	-0.38	0.11	0.16	-0.27	0.41	-0.76
Pvo-8.5	0.032	0.65	-0.084	0.068	-0.23	-0.24	-0.089	0.074	-0.46	-0.22	-0.011	1	-0.13	-0.28	-0.36	0.21	-0.16	0.11	-0.25	0.73
Gh-5	-0.26	5.8e-17	-0.026	0.066	0.28	0.14	-0.64	-0.37	0.34	0.22	0.29	-0.13	1	0.092	0.39	0.38	-0.53	-0.36	0.039	0.71
Gh-7	0.19	0.074	0.0053	0.21	0.29	0.25	-0.18	-0.14	-0.17	0.3	-0.14	-0.28	0.092	1	-0.016	0.24	-0.11	-0.22	0.13	0.78
Gh-8.5	-0.23	0.18	0.28	-0.19	-0.23	-0.29	0.068	-0.032	0.22	-0.058	-0.38	-0.36	0.39	-0.016	1	0.36	-0.26	0.11	-0.19	-0.73
Vve-5	0.48	0.41	-0.35	-0.19	0.22	-0.3	-0.43	0.25	-0.16	0.19	0.11	0.21	0.38	0.24	0.36	1	-0.16	-0.11	0.26	-0.77
Vve-7	0	-0.26	-0.053	0.42	0.26	0.053	0.32	0	0.11	-0.053	0.16	-0.16	-0.53	-0.11	-0.26	-0.16	1	-0.16	0.11	-0.72
Vve-8.5	-0.22	-0.095	-0.068	-0.097	0.18	-0.46	0.034	-0.042	-0.42	0.21	-0.27	0.11	-0.36	-0.22	0.11	-0.11	-0.16	1	-0.018	0.7
Ls	0.49	0.032	-0.52	-0.45	0.48	0.092	-0.32	0.19	0.23	0.73	0.41	-0.25	0.039	0.13	-0.19	0.26	0.11	-0.018	1	-0.77
PrepGT	0.78	-0.71	0.76	0.76	0.7	-0.79	-0.76	0.78	-0.72	0.8	-0.76	0.73	0.71	0.78	-0.73	-0.77	-0.72	0.7	-0.77	1

Fig. 3. **The visualization of Spearman correlation analysis** for 19 predictors. In the colormap, the change from stronger positive correlation to stronger negative correlation corresponds to the light to dark colors. The $\{1\}-\{2\}$ represents the abbreviated name shown in Table I of given predictor $\{1\}$ in the $\{2\}$-th level including 5 (500**hPa**), 7 (700**hPa**), or 8.5 (850**hPa**).

III. THE DATASET

A. ECMWF IFS and Observations

In our work, datasets[2] are collected by the Integrated Forecasting System (IFS) [36]. IFS can generate 613 meteorological variables. IFS can mainly implement the data assimilation and numerical calculation by jointly utilizing multiple functional components. Among ECMWF IFS products, the High-Resolution ECMWF (HR-ECMWF) [36] provides the finer grid, *i.e.*, the horizontal resolution is $0.125^\circ \times 0.125^\circ$ (*Ground*) and the temporal resolution is 3 hours. Therefore, it is selected as the main tool on NWP systems served for the Shanghai Central Meteorological Observation [4].

ERA5 based on the fifth-generation atmospheric reanalysis is a global dataset by combining numerous data and historical observations across the world. By adopting a more recent model and advanced assimilation system, ERA5 can replace its former version named the ERA-Interim [31] for more advanced atmospheric reanalysis [37]. The ST resolution of ERA5 is $0.25^\circ \times 0.25^\circ$ (*Single altitude*) for one hour.

TABLE I
PREDICTORS ACQUIRED BY CORRELATION ANALYSIS AND THE NUMBER OF PREDICTOR CHANNEL FOR HR-ECMWF AND ERA5. THE PREDICTORS OF HR-ECMWF ARE SPLIT INTO 3 HIGH ALTITUDES (8) AND GROUND (5), AND PREDICTORS OF ERA5 ONLY HAVE A SINGLE ALTITUDES (15). FOR CONVENIENCE, THE NAMES OF MOST PREDICTORS ARE ABBREVIATED AND FULL NAMES ARE LISTED IN SUPPLEMENTARY. FIVE BASIC METEOROLOGICAL FACTORS AS DESCRIPTORS ARE LISTED, AND (.) IS THE ABBREVIATED NAME OF THE DESCRIPTOR.

Predictors		Basic Meteorological Factors (Descriptors)
HR-ECMWF	ERA5	
Altitudes(hPa) (500,700,850): **1∼2:** Wind-U/V **3∼4:** Temperature, Divergence **5∼8:** Rhm, Pvo, Gh, Vve **Ground:** **1∼3:** Tcv, Tc, Tcc **4∼5:** precipitationFactor- Ls, Cp **Channel Num:** 57	**1∼3:** Wind-10mU/10V/Nw **4∼6:** Temperature-2mD/2m/ss **7∼9:** Fcv, Sp, Vid **10∼12:** Tcv, Tc, Tcc **13∼15:** precipitationFactor- Ls, Cp, Tp **Channel Num:** 27	**1:** Temperature (**T**) **2:** Pressure (**P**) **3:** Dewpoint Temperature (**DT**) **4:** Wind (**W**) **5:** Rainfall (**R**)

From the perspective of our BCoP task, HR-ECMWF contains more fine-grained temporal features because of higher horizontal resolution than other NWP products such as $0.5^\circ \times 0.5^\circ$ (*Ground*) for American Global Forecast System (GFS) [37]. Therefore, the predictors of HR-ECMWF are suitable as station-wise input data. Moreover, HR-ECMWF forecasts more than 100 hours whereas ERA5 predicts only a single time step. Hence, HR-ECMWF is more suitable for training the BCoP model for the early warning system.

Nevertheless, it is worth pointing out that the predictors of ERA5 have lower noise than those of HR-ECMWF because reanalyzing is implemented under more recently collected data. Consequently, ERA5-based models can obtain more accurate and credible results. As a balance, HR-ECMWF and ERA5 are selected as two benchmark datasets to evaluate the performance. Furthermore, the real-time observations of surface automated stations from the China Meteorological Agency are used as labels to be predicted [38].

B. Predictor Selection and Preprocessing

It is unnecessary to use all variables from ECMWF IFS for training the BCoP model because of the high computational cost and redundancy in variables. Further, considering existed the linear and non-linear correlations simultaneously between precipitation and other variables, we perform the Spearman correlation analysis [39] respectively on 613 meteorological variables and eliminate the effects of collinearity [40] simultaneously. The related formula of Spearman is defined as:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} , \tag{1}$$

where n denotes the number of samples ranging from January to December of 2017 including a total of 550,361 samples. d_i is the positional deviation between i-th pair-wise samples acquired by sorting (descend/ascend). As a result, we select the 29 predictors highly correlated with corresponding precipitation ground truth. To illustrate the correlation result obtained via Spearman more clearly, we illustrate the correlated heatmap between 19 predictors and precipitation shown in Fig. 3. Specifically, we regard the two successive predictors in a one-time interval as the selected predictor channel of input data for augmenting. The details of selected predictors are shown in Table I.

In this paper, our goal is to achieve the station-wise BCoP for improving the accuracy of precipitation forecast in Eastern China (**Range:** $20^\circ - 40^\circ N, 112^\circ - 124^\circ E$). The task-specific preprocessing shown in Fig. 4 is mainly divided into three steps described below:

Spatiotemporal cropping: For the selected predictors shown in Table I, we crop the Regional Predictors (R-Predictors) in the range of Eastern China mentioned above. Next, we employ the time-slider to slice the Time-Series Predictors (TS-Predictors) in the unit of time.

Aligning stations in time span: By switching the timestamps of TS-Predictors to the corresponding timezone of MICAPS stations[2], we crop the Regional Stations (R-Stations) the same as the previous step. Generally speaking, there exist some differences in the distribution numbers of stations in different timestamps of the time span [4]. To ensure the stations consistent in the same TS-Predictor, we align the locations of all R-Stations from the same TS-Predictor for obtaining the Aligned Stations (A-Stations). Meanwhile, we can acquire the observations of basic meteorological factors listed in Table I from A-Stations as **Station-wise Labels** and the corresponding longitude and latitude.

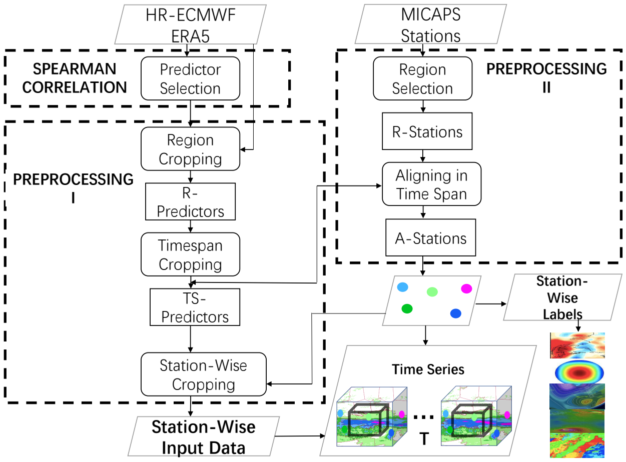


Fig. 4. **The flow diagram of preprocessing the two benchmark datasets and the corresponding stations**. The main operations can be divided into three parts including the Spearman Correlation Analysis [39], Preprocessing I, and II (Pre I, II) enclosed by the dashed boxes. The details of the Spearman Correlation described in Sec. III-B and the candidates of predictors are showed in Table I. Pre I is the preprocesses of cropping the **Station-Wise Input Data** and interacts with Pre II that is the preprocesses of acquiring the **Station-Wise Labels** in the end. Among them, the preprocessing is mainly divided into three key steps detailed in the following subparagraphs and rectangles are the corresponding results. The colorful circles are aligned stations providing the location information and observations as labels for Bias Correcting on Precipitation (BCoP).

Station-wise cropping: For achieving the station-wise BCoP, we crop a specific subregion of the TS-Predictor in the center of longitude and latitude of one A-Station as **Station-Wise Input Data**. We fix the size of the subregion Θ to 29×29 (HR-ECMWF) or 16×16 (ERA5).

In summary, a **Station-Wise Sample** of BCoP consists of station-wise input data and a corresponding label.

IV. SSAS: A SPATIOTEMPORAL SCALE ADAPTIVE SELECTION MODEL

The left plot of Fig. 5 shows our proposed SSAS framework. We define basic notations for our method and then elaborate on an existing problem on BCoP in Sec. IV-A. Whereafter, we propose a Spatial Scale-Selection Module (S3M) to adaptively select the station-wise spatial scale for helping the Spatial-Encoder (S-Encoder) to extract the high-level spatial features in Sec. IV-B. These spatial features are fed into the Temporal Scale-Selection Module (TS2M) to select the optimal sequence for acquiring the adaptive Spatiotemporal (ST) patterns via Temporal-Encoder (T-Encoder) in Sec. IV-C. In Sec. IV-D, the obtained ST patterns are then utilized for outputting the regressing results of precipitation by the Ordinal Regression (OR) method. Besides, the RRC outputs the binary results of precipitation. Finally, we describe the training and testing processes of our model in Sec. IV-E.

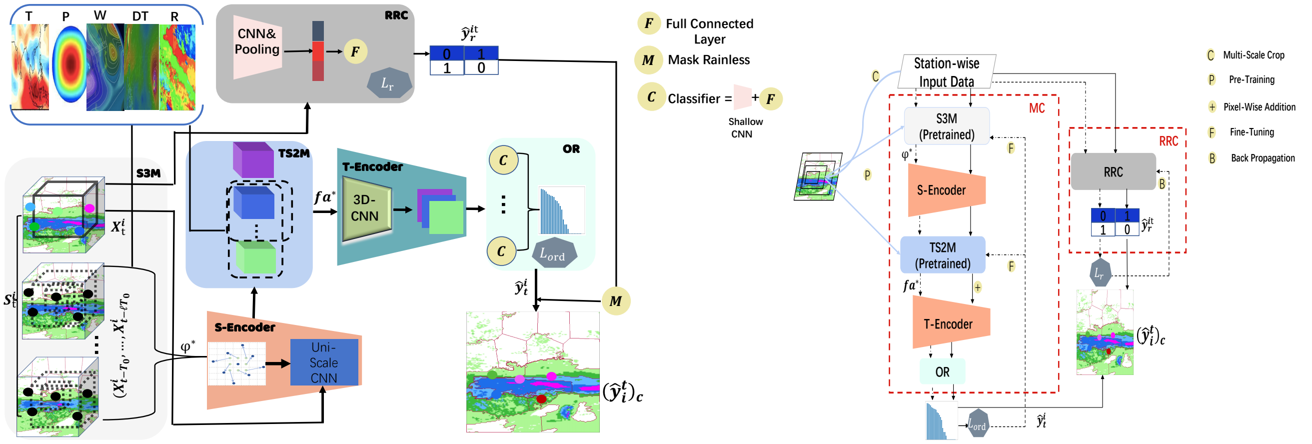


Fig. 5. Left: **The framework of the SSAS.** Every module is filled with different colors and jointly conducts the BCoP. Concretely, the station-wise input data $\boldsymbol{S}_t^i$ in given i-th station at timestamp t firstly are split into **corrector** $\boldsymbol{X}_t^i$ and **auxiliary slice** $[\boldsymbol{X}_{t-T_0}^i, \ldots, \boldsymbol{X}_{t-\ell T_0}^i]$ components. The former is BCoP subject and the latter is utilized for selecting the optimal ST scales by S3M and TS2M, to capture the high-level ST features. The black dots in the auxiliary slice represent stations that are not BCoP targets (**mask**). The color change for colorful dots at the end indicates that SSAS has corrected the station-wise precipitation $(\hat{y}_i^t)_c$ which is closer to labels. Right: **The pipeline of training and testing processes of SSAS.** The blue curve and arrows respectively represent the multi-scale cropping and pre-training process for S3M and TS2M. The training pipeline is performed along the dotted paths and the testing is the solid paths detailed in Sec. IV-E.

A. Notations and Problem Description

First, we define each station-wise input data and label mentioned in Sec. III-B for Bias Correcting on Precipitation (BCoP). A station-wise input data at timestamp t is defined as $\boldsymbol{S}_t^i = [\boldsymbol{X}_t^i, \boldsymbol{X}_{t-T_0}^i, \ldots, \boldsymbol{X}_{t-\ell T_0}^i]$, where i is the ID of the aligned station shown in Fig 4. T_0 is the time interval between two successive frames of $\boldsymbol{S}_t^i$ and ℓ is the length of time sequence set to be 6. Therefore, $[t - \ell T_0, t]$ is the time span of $\boldsymbol{S}_t^i$, and $\boldsymbol{S}_t^i \in \mathbb{R}^{C \times \ell \times \Theta}$. C is the number of predictor channel shown in Table I and Θ given in Sec. III-B is the fixed spatial size of $\boldsymbol{S}_t^i$. Specifically, for the sake of describing the subsequent module more clearly, $\boldsymbol{S}_t^i$ is divided into **corrector** $\boldsymbol{X}_t^i$, which is represented as BCoP subject at timestamp t and ID i, and **auxiliary slice** $[\boldsymbol{X}_{t-T_0}^i, \ldots, \boldsymbol{X}_{t-\ell T_0}^i]$. Note that **auxiliary slice** is not recognized as BCoP subject, but is used for extracting the Spatiotemporal (ST) features detailed in Sec. IV-B-C. Besides, the corresponding station-wise precipitation label of $\boldsymbol{S}_t^i$ is y_t^i.

According to the results of Spearman correlation analysis, the precipitation is mainly subject to selected predictors listed in Table I. Note that precipitation under different locations and time is affected by these predictors in different spatiotemporal scales, especially for heavy rainfall [22]. Hence, selecting the appropriate ST scales of predictors of $\boldsymbol{S}_t^i$ is important for correcting the station-wise precipitation. However, until now, the model based on station-wise BCoP is empirically set to a fixed spatial scale for prediction [19], and this scale is not usually suitable for predicting all station-wise precipitation. The rules-of-thumb may affect the generalization ability of the model and impair predictive accuracy for BCoP. To solve this problem, we come up with the S3M detailed in Sec. IV-B and TS2M in Sec. IV-C to adaptively search the optimal spatiotemporal scales for boosting the performance of BCoP.

B. Spatial Scale-Selection Module and Spatial-Encoder

S3M can adaptively search the optimal spatial scale of the predictor channel of the auxiliary slice. Concretely, we first crop the multi-scale candidates from every element $\boldsymbol{E}$ in the auxiliary slice of $\boldsymbol{S}_t^i$ by providing a step size of cropping $\kappa = 2$, and obtain the multi-scale candidate set $\boldsymbol{\Phi} = \{\boldsymbol{E}^{\Theta}, \boldsymbol{E}^{\Theta-\kappa}, \boldsymbol{E}^{\Theta-2\kappa}, ..., \boldsymbol{E}^{3\times3}\}$. Every candidate in $\boldsymbol{\Phi}$ sequentially forwards the S3M shown in Fig. 6. Concretely, candidates first pass a shallow CNN block. Then, a Deformable CNN (D-CNN) [41] block is used as the backbone of S3M for enhancing the transformation modeling capability detailed in supplementary and dealing with the chaotic property of predictors [42]. Subsequently, the extracted features are separately fed into five S3M branches named after basic meteorological factors shown in Table I including Temperature (T), Pressure (P), Dewpoint Temperature (DT), Rainfall (R), and Wind (W). The first four branches output the probability distributions of T/P/DT/R in virtue of the modified Compact yet efficient Cascade Context-based Age Estimation model (C3AE) [43] detailed in supplementary. These four C3AEs can be optimized by the Kullback-Leibler divergence loss $KL(\cdot,\cdot)$ [44]. Meanwhile, the wind branch is a classifier to recognize the wind speed and wind direction and optimized by the Cross-Entropy Loss $CE(\cdot,\cdot)$ [45]. The $KL(\cdot,\cdot)$ and $CE(\cdot,\cdot)$ for the candidate in one scale ϕ are formulated as:

$$\mathcal{L}_{KL}^{\phi}(\boldsymbol{D}) = KL(S3M_{\boldsymbol{D}}(\boldsymbol{E}^{\phi}), \boldsymbol{p_D}) \,, \qquad (2)$$
$$\mathcal{L}_{CE}^{\phi}(W) = CE(S3M_W(\boldsymbol{E}^{\phi}), \boldsymbol{o}_W) \,,$$

$$\mathcal{L}_{S3M}^{\phi} = \omega_1 \sum_{\boldsymbol{D}} \mathcal{L}_{KL}^{\phi}(\boldsymbol{D}) + \omega_2 \mathcal{L}_{CE}^{\phi}(W) \,, \qquad (3)$$

where $S3M_{\boldsymbol{D}}$ is the S3M branch identified by Descriptors $\boldsymbol{D} = \{T, P, DT, R\}$. $\boldsymbol{E}^{\phi}$ is the candidate having the scale ϕ from Φ. $\boldsymbol{p_D}$ is labeled probability distribution generated from the corresponding observation of $\boldsymbol{D}$. Similarly, the second row of Eq 2 is the cross-entropy loss of wind in the scale ϕ for one element in auxiliary slice and $\boldsymbol{o}_W$ is the one-hot label of wind. The $\boldsymbol{p_D}$ and $\boldsymbol{o}_W$ are detailed in supplementary. Further, the loss of S3M in the scale ϕ for one element is a weighted summation of 4 $\mathcal{L}_{KL}^{\phi}$s and $\mathcal{L}_{CE}^{\phi}(W)$. ω_1 and ω_2 are weights set to be 0.2. Finally, we select the candidate in spatial scale ϕ^* corresponding to minimum $\mathcal{L}_{S3M}$ among $\boldsymbol{\Phi}$ as optimal spatial scale formulated as:

$$\phi^* \quad = \quad \arg\min \mathcal{L}_{S3M}^{\Phi} \ . \tag{4}$$

Toward all elements of the auxiliary slice in respective ϕ^*s, we feed them into Spatial-Encoder (S-Encoder) to conduct 3 steps one by one shown on the left of Fig. 5. Concretely, 1) $\boldsymbol{E}^{\phi^*}$ and the corresponding corrector pass the Deformable CNN (D-CNN) for learning the dense spatial transformation, and then 2) employ the uni-scale CNN block for upsampling or downsampling to unify the spatial scale $U = 10 \times 10$. 3) Finally, all elements in the same auxiliary slice and corresponding corrector in the time dimension is concatenated to generate the station-wise spatial features $\boldsymbol{F}_t^i = [\boldsymbol{f}_t^i, \boldsymbol{f}_{t-T_0}^i, \cdots, \boldsymbol{f}_{t-\ell T_0}^i]$, and $\boldsymbol{F}_t^i \in \mathbb{R}^{C_1 \times \ell \times U}$. C_1 is the number of channels of $\boldsymbol{F}_t^i$. Same as $\boldsymbol{S}_t^i$, $\boldsymbol{F}_t^i$ can also be separated into **corrector** $\boldsymbol{f}_t^i$ and **auxiliary slice** $\boldsymbol{FA}$ =$[\boldsymbol{f}_{t-T_0}^i, \cdots, \boldsymbol{f}_{t-\ell T_0}^i]$.

C. Temporal Scale-Selection Module and Temporal Encoder

Because precipitation is affected by the predictors in different time scales, TS2M is employed for choosing the optimal temporal scale for $\boldsymbol{F}_t^i$. Concretely, we define the $\boldsymbol{\Psi} = \{[\boldsymbol{f}_{t-\ell T_0}^i], [\boldsymbol{f}_{t-\ell T_0}^i, \boldsymbol{f}_{t-2\ell T_0}^i], ..., [\boldsymbol{f}_{t-T_0}^i, \cdots, \boldsymbol{f}_{t-\ell T_0}^i]\}$ as the subsequence set of $\boldsymbol{FA}$ in $\boldsymbol{F}_t^i$. In Fig. 6, we feed every subsequence $\boldsymbol{fa}$ in $\boldsymbol{\Psi}$ into the TS2M, entering the two-layers 3D-CNN [46] as the backbone of TS2M for capturing the Spatiotemporal (ST) features. Like the S3M, the fa having the minimum loss of TS2M can be regarded as the sequence $\boldsymbol{fa}^*$ having the optimal temporal scale for $\boldsymbol{F}_t^i$. The ST features learned from 3D-CNN have richer semantic information than the spatial features captured by D-CNN alone from S3M. Therefore, the outputs from five TS2M branches including regression results, in addition to probability distributions and softmax results shown on the right part of Fig. 6. Mean Absolute Error $MAE(\cdot,\cdot)$ [45], $KL(\cdot,\cdot)$ and $CE(\cdot,\cdot)$ are employed for the process of TS2M, and the related formulas of TS2M are defined as:

$$\begin{aligned}
\mathcal{L}^{\boldsymbol{fa}}(\boldsymbol{D}) &= KL(TS2M_{\boldsymbol{D}}^{12}(\boldsymbol{fa}), \boldsymbol{p_D}) \\
&\quad + MAE(TS2M_{\boldsymbol{D}}^{13}(\boldsymbol{fa}), y_{\boldsymbol{D}}) \ , \\
\mathcal{L}^{\boldsymbol{fa}}(W) &= CE(TS2M_{W}^{12}(\boldsymbol{fa}), \boldsymbol{o}_W) \\
&\quad + MAE(TS2M_{\boldsymbol{D}}^{34}(\boldsymbol{fa}), y_W) \ , \\
\mathcal{L}_{TS2M}^{\boldsymbol{fa}} &= \omega_3 \sum_{\boldsymbol{D}} \mathcal{L}^{\boldsymbol{fa}}(\boldsymbol{D}) + \omega_4 \mathcal{L}^{\boldsymbol{fa}}(W) \ , \\
\boldsymbol{fa}^* &= \arg\min \mathcal{L}_{TS2M}^{\boldsymbol{\Psi}} \ ,
\end{aligned} \tag{5}$$

where $TS2M_{*}^{12}$ is the specific TS2M branch including 3D-CNN and two fully connected layers $F1(\cdot)$ and $F2(\cdot)$, as similar with $TS2M_{\boldsymbol{D}}^{13}$ and $TS2M_{\boldsymbol{D}}^{34}$. Parameters $y_{\boldsymbol{D}}$ and y_W are the observations of $\boldsymbol{D}$ and W. Besides, ω_3 and ω_4 set to be 0.2 are regarded as the weights of $\mathcal{L}_{TS2M}^{\ell}$. Subsequently, the $\boldsymbol{F}_t^i$ including $\boldsymbol{fa}^*$ are fed into the Temporal Encoder (T-Encoder) shown on the left of Fig. 5. As a result, the optimal station-wise spatiotemporal representation $(\boldsymbol{F}_t^i)^* \in \mathbb{R}^{C_2 \times U}$ can be obtained for prediction, where C_2 denotes the number of channels of $(\boldsymbol{F}_t^i)^*$.

D. Ordinal Regression and Rainfall Classifier

The traditional regression methods may suffer from maldistribution because our datasets only include a small proportion of moderate and heavy rain samples reflected in the middle of Fig. 1. Besides, precipitation fluctuation caused by the long-span value range will make the model hard to converge. Therefore, a rank learning method named Ordinal Regression (OR) [47] is introduced for solving these issues. OR can divide the value of precipitation into multiple ranges to conduct the binary classification for every range on BCoP [19]. In this way, a complicated regression task was converted into multiple binary classification tasks that output the ordinal distribution of precipitation shown on the left of Fig. 5. To refine the performance of our model in the rainless prediction, we introduce a binary classifier [19] to directly classify the rainy or rainless samples, as shown on the left of Fig. 5. The function of this classifier is equivalent to regress the precipitation if and only if the classification result is rainy, and is detailed in Sec. IV-E. $\mathcal{L}_r$ is a binary cross-entropy loss for this classifier.

E. Training and Testing

Here we separately discuss the processes of training and testing for better understanding our proposed SSAS model. For convenience, the model will be split into Master Components (MC) and Rainy or Rainless Classifier (RRC).

Training: We firstly crop station-wise input data $\boldsymbol{S}_t^i$ along the time and space dimensions and acquire multi-scale ST data for pre-training the S3M and TS2M alone shown on the right of Fig. 5. The purpose of pre-training is to learn the ability to predict five descriptors shown in Table I. We integrate the pre-trained S3M and TS2M into the MC. Second, the RRC and MC are separately trained. Concretely, the binary results from RRC can be used for calculating binary cross-entropy loss $\mathcal{L}_r$ and formulated as:

$$\mathcal{L}_r = -\sum_i y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i) \ , \tag{6}$$

where $\hat{p}_i$ is the predicted binary possibility for i-th station output from RRC and y_i is the corresponding binary label. Subsequently, the MC are trained and utilized to select the optimal ST scales for every $\boldsymbol{S}_t^i$, capture the ST representation in line with the selected scales, and output the ordinal distribution of precipitation shown on the right of Fig. 5. The $\mathcal{L}_{ord}$ with Focal loss is selected for optimizing the entire MC and formulated as:

$$\begin{aligned}
\mathcal{L}_{ord} = &-\sum_{d=0}^{v-1} \alpha(1 - \hat{p}_d)^{\gamma} log(\hat{p}_d) * y_d \\
&+ (1 - \alpha)\hat{p}_d^{\gamma} log(1 - \hat{p}_d) * (1 - y_d) \ ,
\end{aligned} \tag{7}$$

where v is the number of binary classifiers of ordinal regression (OR). y_d is the discrete one-hot ordinal label that is generated from the station-wise label of precipitation, and $\hat{p}_d$ is the predicted ordinal possibility from d-th classifier.

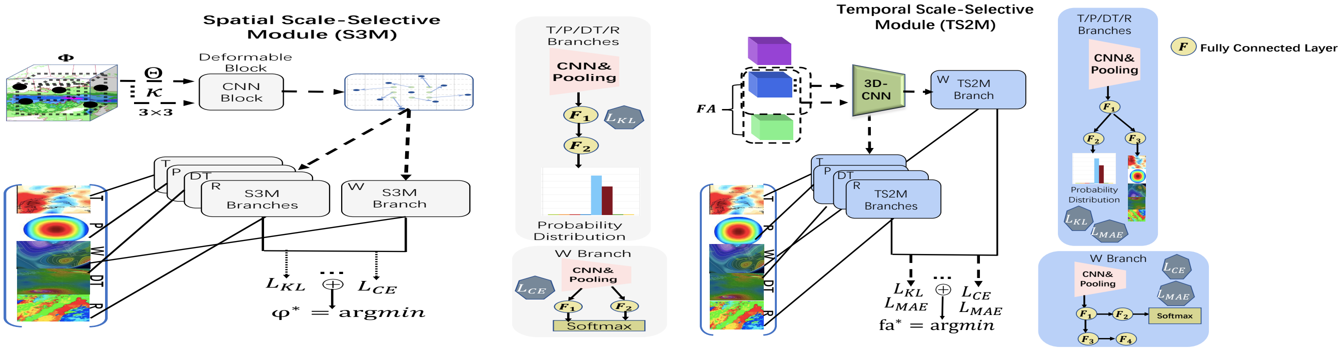


Fig. 6. Left: **The structure of the Spatial Scale-Selection Module (S3M)** can select the adaptive spatial scale ϕ^* corresponding to minimum S3M loss. The labeled probability distributions utilized for calculating the S3M loss, are generated from the observations of station-wise descriptors visualized by color code. Concretely, we crop the multi-scale candidate set $\boldsymbol{\Phi} = \{\boldsymbol{E}^{\Theta}, \boldsymbol{E}^{\Theta-\kappa}, \boldsymbol{E}^{\Theta-2\kappa}, ..., \boldsymbol{E}^{3\times3}\}$ from the element E in Auxiliary Slice, where κ is the step size of cropping and Θ is the fixed spatial scale. Right: **The structure of the Temporal Scale-Selection Module (TS2M)** can adaptively select the time sequence $\boldsymbol{fa}^*$ in optimal time-scale from subsequence set of auxiliary slice $\boldsymbol{\Psi}$ for station-wise spatial features $\boldsymbol{F}_t^i$. Different with S3M, the TS2M also performs the regressions in five branches.

$\alpha = 0.25$ and $\gamma = 2$ are the hyperparameters. Particularly, the abilities to select optimal ST scales and fit the distribution of corresponding station-wise precipitation labels for S3M and TS2M can be learned (fine-tuned) via an end-to-end optimization. **Testing:** According to the description of S3M and TS2M in Sec. IV-B-C, the outputs of every S3M/TS2M branch include the probability distribution and softmax results. During the testing phase, the max confidence score $CS_{max,*}$ for one element $\boldsymbol{E}$ under one scale ϕ or subsequence $\boldsymbol{fa}$ is defined as the largest value of the corresponding predicted possibility distribution or softmax results. The $*$ represents ϕ or $\boldsymbol{fa}$, and $CS_{max,*}$ can be calculated below:

$$CS_{\max,*} = \theta_1 \sum_{\bar{\boldsymbol{D}}} \max(\hat{\boldsymbol{p}}^*_{\bar{\boldsymbol{D}}}) + \theta_2 \max(\hat{\boldsymbol{p}}^*_R) \ , \tag{8}$$

where $\hat{\boldsymbol{p}}^*_{\bar{\boldsymbol{D}}}$ and $\hat{\boldsymbol{p}}^*_R$ are respectively the $*$-specific predicted possibility distributions and softmax results, and $\bar{\boldsymbol{D}} = \{T, P, W, DT\}$. θ_1 and θ_2 are weights set to be 0.1 and 0.6. Further, for S3M, according to the calculated $\{CS^1_{\max,\phi^1}, CS^2_{\max,\phi^2}, ..., CS^k_{\max,\phi^k}\}$ as the top k max confidence scores and corresponding scale set $\boldsymbol{\Omega} = \{\phi^1, ..., \phi^k\}$, we crop the element E to multi-scale elements by $\boldsymbol{\Omega}$ and feeding them into S-Encoder for extracting the k spatial features. Enlightened by multi-scale feature pyramid network [48], we conduct the pixel-wise addition operation on k spatial features shown on the left of Fig. 5 and regard the results as the spatial features of given E. Subsequently, for TS2M, we directly select the subsequence $\boldsymbol{fa}$ having the largest $CS_{\max,\boldsymbol{fa}}$ as the input of T-Encoder. Finally, the ordinal distribution of precipitation is utilized for regressing the precipitation $\hat{y}_t^i$. we expand $\hat{y}_t^i$ to regression vector $\hat{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{i}} = [0, \hat{y}_t^i]$ by padding zero for the subsequent operation. Besides, the two one-hot predictions $\hat{\boldsymbol{y}}_{\boldsymbol{r}}^{\boldsymbol{it}}$ ($[0,1]$ is rainy and $[1,0]$ is rainless) are output from RRC. Above all, the final formulation on BCoP is shown as:

$$(\hat{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{i}})_c = \hat{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{i}} \times (\hat{\boldsymbol{y}}_{\boldsymbol{r}}^{\boldsymbol{it}})^T \ , \tag{9}$$

where the last element in $(\hat{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{i}})_c$ is station-wise corrected precipitation shown on the left of Fig. 5. Specifically, Eq (9) can mask the $(\hat{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{i}})_c$ into a zero vector when $\hat{\boldsymbol{y}}_{\boldsymbol{r}}^{\boldsymbol{it}} = [1, 0]$, which is equivalent to only output the rainless classification result from RRC.

V. RESULTS AND ANALYSIS

In this section, we evaluate the performance of our method. First, we split the datasets for all experiments, and elaborate on the experimental setups and the network layers of our proposed model. Second, we formulate the evaluation criteria of BCoP. Finally, we show our method has a preferable performance of BCoP and also effectively correct the heavy precipitation through six experiments: 1) **Contrastive Experiments**, 2) **Pure Rainfall Experiments**, 3) **Ablation Study**, 4) **Qualitative Analysis**, and 5) **Parameter Sensitivity Analysis**, and 6) **Statistical Tests**.

A. Datasets and Experiment Setups

We split the station-wise samples preprocessed in Sec. III-C into two datasets, i.e., High-Resolution ECMWF (HR-ECMWF) and ERA5. In consideration of the climate character of precipitation in Eastern China, these two datasets cover a total of 9 months from July to September, 2016 to 2018 [4]. HR-ECMWF benchmark (**HR-ECb**) includes 300,205 samples and ERA5 benchmark (**ERA5b**) contains 335,946 samples. Further, we adopted different partitioning strategies for the Contrastive Experiment and Pure Rainfall Experiment described below:

Contrastive Experiment: we partition both HR-ECb and ERA5b into three parts: training set, cross-validation set, and testing set. Concretely, we use the samples of the given 6 months of 2016-2017 as the training set, the August-September of 2018 as the testing set, and samples from July 2018 as the cross-validation set. The sample distribution ratio in the training, cross-validation, and testing sets is roughly 70%:10%:20%. For alleviating the unbalanced issue

TABLE II
THE LAYER STRUCTURE OF SSAS FRAMEWORK. $[\cdot,\cdot]$ MEANS A 2D OR 3D CONVOLUTIONAL LAYER WITH BATCH NORMALIZATION [49] AND LEAKYRELU, WHERE THE FIRST ELEMENT IS KERNEL SIZE AND THE SECOND ONE IS THE NUMBER OF OUTPUT CHANNELS. BESIDES, $\{\cdot,\cdot\}$ IS A BOTTLENECK BLOCK EMPLOYED IN LIGHTWEIGHT MOBILENETV2 [50]. THE TWO ELEMENTS OF $\{\cdot,\cdot\}$ DENOTES THE NUMBER OF INPUT AND OUTPUT CHANNELS RESPECTIVELY. THE ABBREVIATIONS OF ADAPTIVE AVERAGE POOLING AND GLOBAL AVERAGE POOLING ARE AAP AND GAP, AND THE INP IS THE INTERPOLATION. THE AAP AND INP ARE UTILIZED FOR UNIFYING THE SPATIAL SCALE TO 10×10. THE FC MEANS FULL CONNECTED LAYER AND THE $(\cdot) - d$ IS THE NUMBER OF DIMENSION OF FC. ALL FEATURE EXTRACTION COMPONENTS SUCH AS S3M ARE CONNECTED WITH THE RESIDUAL LAYER [51] IDENTIFIED BY RES.

Inputs: $57 \times 6 \times 29 \times 29$ / $27 \times 6 \times 16 \times 16$		
S3M	Res I	$[1 \times 1, 128] \times 1$
	Bottleneck I	$\{128, 64\} \times 1$
	Conv1	$[3 \times 3, 32] \times 1$
	D-CNN Block I	$[1 \times 1, 64] \times 1$ $[3 \times 3, 128] \times 1$
	S3M-Bottleneck on (T/P/DT/W/R)	$\{128, 64\} \times 1$
	S3M-Output on (T/P/DT/R)	$GAP,\ prob - d\ FC$
	S3M-Output on W	$GAP,\ cls - d\ FC,\ softmax$
S-Encoder	D-CNN Block II	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 128 \\ 1 \times 1, 64 \\ 3 \times 3, 64 \end{bmatrix}$
	Uni-Scale CNN Block	ResII$[1 \times 1, 256] \times 1$, Bottleneck II$\{256, 64\} \times 1$, Conv2$[3 \times 3, 256] \times 1$ $AAP/INP, \quad Downsampling \| Upsampling 10 \times 10$
TS2M	3DCNN Block I	$[2 \times 3 \times 3, 64] \times 2$
	Res III	$[3 \times 3, 64] \times 1$
	TS2M-Bottleneck on (T/P/DT/W/R)	$\{64, 128\} \times 1$
	TS2M-Output on (T/P/DT/R)	$GAP,\ prob - d\ FC,\ 1 - d\ FC$
	TS2M-Output on W	$GAP,\ cls - d\ FC,\ softmax,\ 1 - d\ FC$
T-Encoder	3DCNN Block II	$[2 \times 3 \times 3, 64] \times 2$
Ordinal Regression	Single Precipitation	$[3 \times 3, 128] \times 2,\ [1 \times 1, 32] \times 2$
	Binary Classifier	$GAP,\ 1 - d\ FC,\ sigmoid$
Rainfall Classifiers	Conv4	$[3 \times 3, 64] \times 2$ $[3 \times 3, 128] \times 2$
	Binary Layer	$GAP,\ 2 - d\ FC,\ softmax$

of data distribution in three divided sets, we adjust ratios of 4 rainfalls[3] in these three sets to roughly approximate No:Light:Moderate:Heavy = $12 : 9 : 3 : 1$.

Pure Rainfall Experiments: we only adopted the HR-ECb dataset in this experiment because HR-ECb supports longer-range forecasting compared with ERA5b discussed in Sec. III-A. The pure rainfall samples from HR-ECb are partitioned into three different datasets based on rainfall intensity: 1) **HR-ECbL** only including Light rainfall about 62,000 samples, 2) **HR-ECbM** only including Moderate rainfall about 41,500 samples and 3) **HR-ECbH** only including Heavy rainfall about 20,700 samples. The sample distribution ratios between training and testing in these three datasets are all roughly 80%:20%.

In our experiments, each network model and pre-training module are trained 60 epochs with a batch size of 256, and tested once every 3 training epochs with a batch size of 64 (20 tests in total). Meanwhile, we save all test results and corresponding weights. We employ Adam as optimizers, and then the initial learning rate, dropout rate, and weight decay are set to 0.001, 0.3, and 0.0001, respectively. The learning rate is progressively decreased by a decay strategy with step size 20 on the multiplicative change factor of 0.1. Besides, for labeling the probability distribution of descriptors for S3M and TS2M, the value range of descriptors and corresponding intervals need to be estimated experimentally and empirically. The ranges of dividing bins[4] of descriptors are fixed to be a set $B\_R = \{(-4, 43, 2)_T, (18, 82, 3)_P, (-16, 34, 2)_{DT}, (0, 90, 3)_R\}$ by parameter sensitivity analyses in Sec. V-F, where the three elements of $(\cdot,\cdot,\cdot)$ in order as follows: 1) minimum value, 2) maximum value, and 3) intervals of T/P/DT/R. The details of generating the labeled possibility distribution of T/P/DT/R and the class labels of wind are discussed in supplementary. Similarly, as for the discrete one-hot label of ordinal regression, the rank range of precipitation and interval are set to 0-90mm and 0.5 detailed in supplementary. Finally, all experiments are implemented in PyTorch 1.5.0 for deep models with 4 NVIDIA GPUs and scikit-learn 0.19.1 for machine learning models with multi-core Intel CPUs.

Table II sequentially lists the main layers of the SSAS framework. Specifically, the S3M-Bottleneck on (T/P/DT/W/R) and S3M-Output on T/P/DT/R represent the five bottleneck blocks and four output layers with independent parameters respectively, and the same as notations of TS2M. Meanwhile, the value of d in $prob - d\ FC$ depends on the bin interval of T/P/DT/R, and the d in $cls - d\ FC$ is subject to the rank of wind speed and angle segmentation of wind direction.

TABLE III
THE SIX METRICS DEFINED BETWEEN 12 METHODS AND SSAS ON HR-ECB AND ERA5. IFS-LP IS LARGE PRECIPITATION (LP) PREDICTOR AS PRECIPITATION FORECASTS BEFORE CORRECTING AND DELIVERED FROM THE INTEGRATED FORECASTING SYSTEM (IFS). FOR ACQUIRING THE STATION-WISE IFS-LP RESULTS, WE AVERAGE THE SUMMATION OF THE NEAREST FOUR GRIDS OF LP PREDICTOR AROUND THE SPECIFIC STATION. IN THIS TABLE, MLR: MULTIPLE LINEAR REGRESSION; SVR: SUPPORT VECTOR REGRESSION; RF: RANDOM FOREST; ANN: ARTIFICIAL NEURAL NETWORK; CNN: CONVOLUTIONAL NEURAL NETWORK; FPN: FEATURE PYRAMID NETWORK; TPN: TEMPORAL PYRAMID NETWORK; DA-RNN: DUAL-STAGE ATTENTION-BASED RECURRENT NEURAL NETWORK. T-GCN: TEMPORAL GRAPH CONVOLUTIONAL NETWORK. CONVLSTM: CONVOLUTIONAL LONG SHORT-TERM MEMORY; OBA: ORDINAL BOOSTING AUTO-ENCODER. THE COMPLEXITY OF ALL DEEP-LEARNING-BASED METHODS ARE ANALYZED BY FLOATING POINT OPERATIONS (**FLOPs**) AND THE NUMBER OF PARAMETERS (**#Params**).

		Criteria						**Complexities**	
Datasets	**Methods**	**MAE**	**RMAE**	**MAPE**	**$TS_{0.1}$**	**TS_1**	**TS_{10}**	**FLOPs**	**#Params**
HR-ECb	IFS-Lp [36]	1.98±0.00	20.48±0.00	5.58±0.00	0.30±0.00	0.28±0.00	0.19±0.00	–	–
	MLR [12]	1.83±0.17	14.57±0.95	5.15±0.22	0.37±0.03	0.36±0.04	0.20±0.05	–	–
	SVR [52]	1.69±0.15	13.80±0.94	4.76±0.20	0.40±0.03	0.38±0.04	0.20±0.05	–	–
	RF [53]	1.61±0.15	13.47±0.94	4.54±0.20	0.41±0.03	0.39±0.03	0.21±0.05	–	–
	ANN [15]	1.58±0.11	13.23±0.92	4.45±0.19	0.43±0.03	0.40±0.03	0.21±0.04	0.00**G**	0.19**M**
	CNN [5]	1.36±0.08	11.85±0.81	3.83±0.15	0.44±0.03	0.44±0.03	0.22±0.04	0.54**G**	0.63**M**
	DA-RNN [27]	1.33±0.08	11.17±0.82	3.75±0.15	0.45±0.03	0.44±0.03	0.21±0.04	0.50**G**	1.82**M**
	FPN [54]	1.30±0.08	10.58±0.79	3.66±0.13	0.47±0.03	0.45±0.03	0.22±0.04	0.38**G**	3.18**M**
	T-GCN [28]	1.32±0.09	10.91±0.81	3.71±0.15	0.45±0.03	0.44±0.03	0.21±0.04	30.09**G**	7.20**M**
	ConvLSTM [5]	1.27±0.08	8.94±0.76	3.58±0.13	0.47±0.03	0.48±0.03	0.23±0.04	44.36**G**	11.8**M**
	TPN [55]	1.23±0.07	7.80±0.64	3.46±0.11	0.55±0.03	0.50±0.03	0.27±0.03	22.56**G**	9.08**M**
	OBA [19]	1.02±0.05	6.89±0.52	2.87±0.10	0.59±0.02	0.53±0.03	0.29±0.03	0.29**G**	12.67**M**
	SSAS(ours)	**0.92**±0.03	**4.90**±0.33	**2.61**±0.08	**0.69**±0.02	**0.57**±0.02	**0.39**±0.03	33.17**G**	26.78**M**
ERA5b	IFS-Lp [36]	1.89±0.00	18.99±0.00	5.32±0.00	0.34±0.00	0.32±0.00	0.20±0.00	–	–
	MLR [12]	1.77±0.15	14.02±0.90	5.01±0.20	0.38±0.03	0.37±0.04	0.20±0.04	–	–
	SVR [52]	1.63±0.13	13.69±0.88	4.59±0.19	0.41±0.03	0.39±0.04	0.20±0.04	–	–
	RF [53]	1.57±0.12	13.39±0.87	4.42±0.19	0.43±0.03	0.40±0.03	0.21±0.04	–	–
	ANN [15]	1.49±0.10	12.53±0.85	4.20±0.18	0.46±0.03	0.41±0.03	0.21±0.03	0.00**G**	0.18**M**
	CNN [5]	1.20±0.07	9.39±0.72	3.38±0.13	0.49±0.02	0.48±0.02	0.23±0.03	0.16**G**	0.62**M**
	DA-RNN [27]	1.18±0.07	8.83±0.72	3.32±0.13	0.51±0.02	0.47±0.02	0.22±0.03	0.21**G**	1.71**M**
	FPN [54]	1.14±0.06	8.39±0.67	3.21±0.11	0.53±0.02	0.50±0.02	0.24±0.03	0.11**G**	3.16**M**
	T-GCN [28]	1.15±0.06	8.74±0.71	3.24±0.12	0.52±0.02	0.49±0.02	0.23±0.03	8.05**G**	6.98**M**
	ConvLSTM [5]	1.13±0.07	7.01±0.64	3.18±0.10	0.53±0.02	0.50±0.02	0.24±0.03	13.08**G**	9.43**M**
	TPN [55]	1.06±0.06	5.66±0.49	2.99±0.09	0.55±0.02	0.55±0.02	0.28±0.02	6.65**G**	7.88**M**
	OBA [19]	0.98±0.04	5.03±0.46	2.76±0.06	0.62±0.01	0.54±0.02	0.30±0.02	0.10**G**	12.65**M**
	SSAS(ours)	**0.88**±0.02	**4.17**±0.22	**2.48**±0.05	**0.73**±0.01	**0.59**±0.02	**0.39**±0.02	19.07**G**	24.12**M**

B. Evaluation Criteria

In BCoP, MAE (Mean Absolute Error) [45] and RMAE (Rainy Mean Absolute Error) can be regarded as two eval-

---

[3] 4 rainfalls in Eastern China can be roughly grouped, based on intensity, into no rainfall $[0, 0.1)mm$, light rainfall $[0.1, 1)mm$, moderate rainfall $[1, 10)mm$, and heavy rainfall that is above $10mm$ [4], where a unit of rainfall is millimeter (mm).

[4] Bin detailed in supplementary is defined as the specific dividing point in the value range of T/P/DT/R descriptors listed in Table I or classes of wind.

TABLE IV

The four metrics between three methods and SSAS on HR-ECbL/HR-ECbM/HR-ECbH divided by precipitation intensity. N/A(Not Applicable) indicate that the TS_ρ for current specific dataset not have to research. For example, for the evaluation results of TS_{10} on HR-ECbL, it is not the purpose of the pure rainfall experiments.

	Methods	Criteria			
Datasets		RMAE	$TS_{0.1}$	TS_1	TS_{10}
HR-ECbL	LSTM [4]	6.15±0.18	0.43±0.02	N/A	N/A
	CNN [5]	4.92±0.25	0.45±0.02	N/A	N/A
	OBA [19]	2.18±0.10	0.61±0.01	N/A	N/A
	SSAS	**1.97**±0.08	0.72±0.01	N/A	N/A
HR-ECbM	LSTM	12.77±0.77	0.39±0.02	0.40±0.03	N/A
	CNN	9.91±0.70	0.41±0.02	0.42±0.02	N/A
	OBA	6.16±0.48	0.59±0.02	0.50±0.02	N/A
	SSAS	**4.47**±0.32	0.70±0.01	**0.54**±0.02	N/A
HR-ECbH	LSTM	18.27±0.99	0.30±0.04	0.33±0.06	0.14±0.05
	CNN	15.83±0.91	0.34±0.03	0.39±0.05	0.16±0.05
	OBA	10.89±0.82	0.55±0.03	0.45±0.03	0.20±0.05
	SSAS	**7.15**±0.56	0.64±0.02	0.51±0.03	**0.28**±0.05

uation metrics to calculate the difference between station-wise corrected precipitation and corresponding observation. RMAE is a variant of MAE not including rainless samples. Meanwhile, Threat Score (TS) is an important evaluation index for guiding forecasting products [4] as follows:

$$TS = H/(H + M + FA) \ , \tag{10}$$

where H (Hit) is the number of samples in the case of correction=1, observation=1, and M (Miss) is the number of samples in the case of correction=0, observation=1. Similarly, FA (False Alarm) is regarded as the case of correction=1, observation=0, in which 1 is rainy (above $0.1mm$) and 0 is rainless, *i.e.*, $[0, 0.1)mm$ in line with 4 rainfalls[3]. Further, for the sake of obtaining the fine-grained and targeted evaluation results, a new criterion TS_ρ, as a variant of TS, is introduced for evaluating the performance of models in a specific value range of precipitation and not only limited the forecasting on 1 and 0. The TS_ρ formulates as:

$$TS_\rho = H_\rho/(H_\rho + M_\rho + FA_\rho) \ , \tag{11}$$

where H_ρ is the number of samples in the case of correction$>=\rho$, observation$>=\rho$ and M_ρ is the number of samples in the case of correction$<\rho$, observation$>=\rho$. Similarly, FA_ρ (False Alarm) is regarded as the case of correction$>=\rho$, observation$<\rho$, where ρ is set to be 0.1, 1, and 10 as range partitions according to rainfall intensity. The larger for ρ, the higher the evaluating difficulty for heavy rainfall.

C. Contrastive Experiments on HR-ECb and ERA5b

We list the assessment results of 12 methods and our proposed Spatiotemporal Scale Adaptive Selection (SSAS) on the HR-ECb and ERA5b shown in Table III. The reported results are the average of 20 tests, each of which is the mean of evaluation results of all batches. First, the majority of methods obtain slightly preferable performance on the ERA5b, compared with the same methods on HR-ECb. One possible reason is that the predictors of ERA5b are closer to the corresponding observation [31] than those of HR-ECb. Therefore, models trained by ERA5b may be easier to converge with the lower losses. Moreover, the absolute differences of $RMAE$ in the adjacent rows of Table III are much larger than those of MAE. The main reason for this phenomenon is that the number of rainless samples is not used as the denominator to calculate the $RMAE$. Hence, compared with MAEs, the $RMAE$s are more sensitive to correction errors for the precipitation itself, especially for heavy precipitation. Second, CNN and ConvLSTM obtain better correction performance than MLR and ANN because of the stronger abilities of feature extraction. Furthermore, ConvLSTM achieves better performance than CNN. The reason is that both temporal and spatial features can be leveraged by ConvLSTM, whereas spatial features are given alone for CNN. In addition to that, SVR has a preferable performance than MLR because SVR is more suitable for simulating non-linear task such as BCoP by searching the divided hyperplane with kernel function, compared with MLR. Meanwhile, RF has more influence on BCoP than SVR. We speculate that the main cause is that RF consists of multiple classified Decision Trees (DTs) in the different ranges of precipitation. This ensemble idea of RF is similar to ordinal regression. However, RF is sensitive to many key hyperparameters such as the number of DTs and depth of the forest, which will disturb the performance of BCoP to some extent. As for DA-RNN and T-GCN, the performance of DA-RNN has slight increment because LSTM with dual attention can indeed adaptively extract more relevant temporal features by referring to the previous encoder hidden state. Nevertheless, the advantage of long-term time-series prediction for DA-RNN will exploit at a discount, because the assimilation operation [3] as the process of generating the EC data is equivalent to time-series prediction. The T-GCN has inferior results to FPN, because the edge weights of the graph that are essential to learning the strong spatial correlation are not easy to calculate. We only heuristically link the edge by metrics such as Euclidean distance instead of spatial association of atmospheric dynamics. Therefore, T-GCN cannot make full use of extracting the spatiotemporal (ST) representations. In contrast, FPN is designed on the structure of a multi-scale spatial pyramid to capture the highest-resolution, strongly semantic spatial features by bottom-up residual fusion. The key is, of course, the motivation of FPN basically meet our assumption, so it makes sense that FPN has superior performance than T-GCN and DA-RNN. Further, TPN as an advanced FPN adds the temporal module to capture the multi-scale ST features and achieve more competitive results than FPN and ConvLSTM in the same cases. Specifically, information flows aggregation in TPN with IO stream strategy is utilized for aligning the extracted features, so the more enhanced and richer level-wise representations are obtained for precipitation prediction of TPN. As a whole, it is worth mentioning that the higher performance enhancement from ANN to ConvLSTM is obtained on ERA5b and the evaluating results on ERA5b have smaller standard deviations. There are mainly two reasons for these aforementioned phenomena: 1) ERA5b has fewer noisy samples than those of HR-ECb, and 2) compared with HR-ECb, the ST representations learned from ERA5b can be more robust since the noise in ERA5b is too small to severely disturb the performance in the process of feature extraction. The OBA has superior performance than TPN and ConvLSTM, which mainly benefits from the component of ordinal regression that can mitigate the long-span and maldistribution problems. Finally, SSAS outperforms all the other methods on five metrics and its TS_{10}s are respectively 34.48% and 30.00% higher than the second-highest results from OBA in these two datasets. The outperformance of our model may well be benefited from extracted more high-level and informative representations from the predictors under the adaptive ST scales, and the detailed demonstration of this analysis is described in the ablation study.

D. Pure Rainfall Experiments on HR-ECbL/M/H

For investigating the performance of models under the different rainfall intensities, especially in heavy rainfall, we compare the performance of the three targeted methods and SSAS on HR-ECbL, HR-ECbM, and HR-ECbH as shown in Table IV. For the pure rainfall experiments, we select $RMAE$ and three TSs as evaluation metrics. Overall, for HR-ECbM and HR-ECbH, the performance of all methods on TS_1 and TS_{10} decreases, compared with the same TSs of these methods in Table III. The possible reason is that the HR-ECbM and HR-ECbH inclining to the sampling distributions of moderate and heavy rainfall are hard to correct because long-span and maldistribution problems will become more serious. Meanwhile, $RMAE$s of all methods sharply rise from HR-ECbL to HR-ECbH because heavy rainfall samples can make more contribution to the absolute differences when calculating the $RMAE$s. Furthermore, both CNN and LSTM have worse performance in TS_{10} than OBA because OBA not only captures the temporal features but also introduces the Ordinal Regression (OR) with the focal loss to relieve the long-span and maldistribution issues. Moreover, the performance of CNN is better than LSTM on the three datasets. One possible explanation is that the single-step prediction from LSTM is easy to induce larger cumulative errors if no history observations were utilized for adjusting these errors [56]. Hence, LSTM possibly provides inferior dynamic information than that of CNN. Specifically, SSAS obtains a SOTA result of TS_{10} on HR-ECbH and it indicates that our method has superior correction performance for heavy rain, compared with the other three methods.

E. Ablation Study

Here, we perform ablation experiments on subsets of HR-ECb to verify the effectiveness of each new-introduced component in SSAS, especially researching the influence of the adaptive ST scale relative to the fixed scale in BCoP. We investigate the TS_1 and $RMAE$ for each ablation. The sample distribution ratio between training and testing sets is roughly 70%:30% and the number of tests is 10. Specifically, in Table V, the results in the last row are calculated by the whole SSAS and utilized for contrasting with the results from all ablative operations on SSAS.

Impact of S3M and TS2M: For the ablative operation of the second row is to remove the whole S3M and replace its function with three fixed spatial scales $fs_{(\cdot)}$ set to be 10×10, 15×15, and 23×23. Further, we replace the Deformable CNN (D-CNN) block in Spatial-Encoder (S-Encoder) with the naive CNN layers. Meanwhile, we remove the TS2M from SSAS and set two fixed temporal scales $ft_{(\cdot)}$ to be 3 and 6 as a substitute for the adaptive temporal scale in the third row. For the second row and the third row, the TS_1 sharply decreases after removing S3M or TS2M, which showed the effectiveness of the adaptive ST scale engined by S3M/TS2M. Furthermore, the average TS_1 under the three $fs_{(\cdot)}$s of without (w/o) S3M is approximately 0.02 lower than the average TS2M under two $ft_{(\cdot)}$s. One possible reason is that the role of spatial scale is more important in BCoP than that of temporal scale.

S3M vs TS2M for branches: We study the impacts on correction performance when removing the Rainfall (R) branch or the other four branches (T/P/DT/W) of S3M and TS2M shown in the fourth row to the seventh row. We noticed that S3M and TS2M with the T/P/DT/W branches boost a certain performance. Furthermore, the R branch is more helpful for guiding the ST scales in BCoP than the T/P/DT/W branch in this case, which benefits from autocorrelation between rainfall observation and corresponding precipitation predictor illustrated in Fig. 3.

Impact of deformable CNN and C3AE: We replace all D-CNN blocks or modified C3AEs in SSAS with naive CNN blocks that have the same number of layers as both, aiming at researching the influences of these two components in BCoP shown in the 8-9th rows. $RMAE$ of w/o C3AE increases because C3AE is a ranking learning method, quite like ordinal regression for solving the long-span range of precipitation distribution. Besides, the decrement of TS_1 on w/o D-CNN indicates that D-CNN does work in selecting adaptive pixels [41] of predictor channel for different station-wise precipitation in the process of transforming spatial features.

Impact of MAE and KL divergence loss: We perform the ablation studies on removing the KL loss of Spatial Scale-Selection Module (S3M) and the KL/MAE loss of Temporal Scale-Selection Module (TS2M) to verify the effectiveness of these two losses in our SSAS. Specifically, we adopt the same experiment settings of Table V and the results are shown in Table VI. First, the performance of the 3 methods with ablated loss decreases, compared with the same case of the baseline. It indicates that both MAE loss and KL divergence loss have certain effectiveness, so we may combine KL divergence and MAE loss function in the total loss. Besides, we observe that the performance of S3M w/o KL drops more sharply than that of TS2M w/o KL. One possible reason is that S3M will lose 4 spatial scale-selection submodules shown on the left of Fig. 6 except for the wind submodule when removing the KL loss, but TS2M w/o KL still preserves MAE loss calculated among 4 submodules shown on the right of Fig. 6. All in all, mixing the KL divergence and MAE loss function in the total loss is reasonable and effective.

F. Qualitative Analysis and Parameter Sensitivity Analysis

In Fig. 7, We visualize corrected results from the two rainy cases on HR-ECb. Among these two rainy cases, except IFS-Lp and corresponding observations, we also select the three trained models and SSAS from **Contrastive Experiments** as

TABLE V

ABLATION EXPERIMENTS CONDUCTED ON 80% SAMPLES OF HR-ECB USING TS_1 AND $RMAE$. $\surd$ (**Reserve**) IS DEFINED AS AN EXISTING COMPONENT IN CURRENT ABLATED SSAS FOR EVERY ROW OF TABLE. ON THE OTHER HAND, × (**Remove**) IS NO THIS COMPONENT IN THE ABLATION. S3M-BR AND TS2M-BR ARE THE BRANCHES OF S3M AND TS2M.

S3M	TS2M	S3M-Br					TS2M-Br					D-CNN	C3AE	TS_1	$RMAE$
		R	T	P	W	DT	R	T	P	W	DT				
×	√	N/A					√	√	√	√	√	×	√	fs_{10}:0.54 fs_{15}:0.53 fs_{23}:0.54	fs_{10}:8.69 fs_{15}:8.90 fs_{23}:8.40
√	×	√	√	√	√	√	N/A					√	√	ft_3:0.56 ft_6:0.55	ft_3:7.92 ft_6:8.01
√	√	√	×	×	×	×	√	√	√	√	√	√	√	0.58	6.59
√	√	×	√	√	√	√	√	√	√	√	√	√	√	0.57	7.71
√	√	√	√	√	√	√	√	×	×	×	×	√	√	0.59	5.26
√	√	√	√	√	√	√	×	√	√	√	√	√	√	0.59	5.81
√	√	√	√	√	√	√	√	√	√	√	√	×	√	0.57	6.99
√	√	√	√	√	√	√	√	√	√	√	√	√	×	0.58	6.52
√	√	√	√	√	√	√	√	√	√	√	√	√	√	0.60	4.96

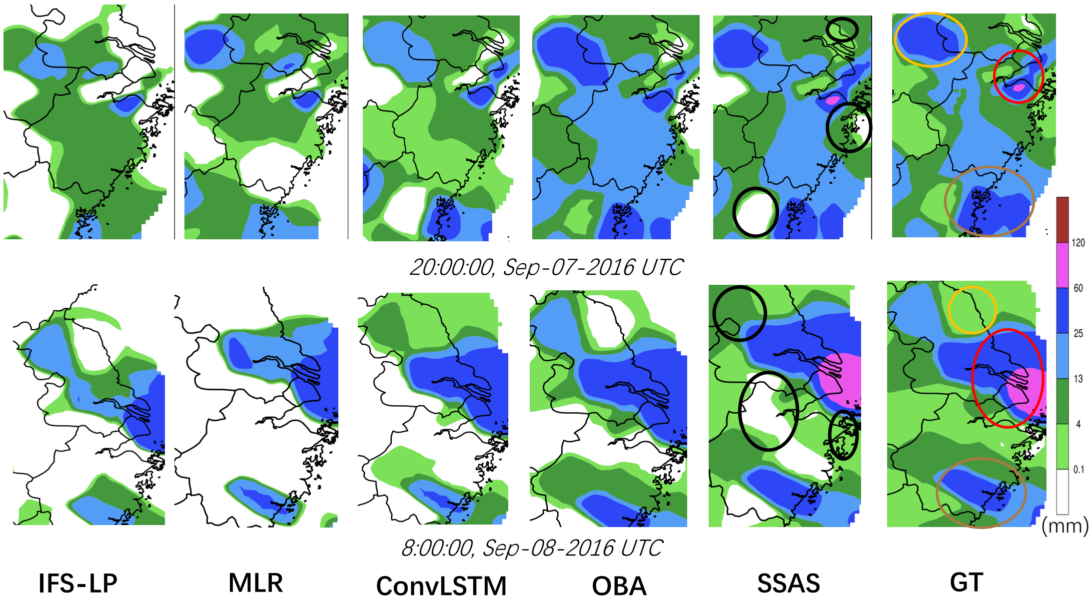


Fig. 7. Visualized comparisons between corrected results from three published methods and SSAS in the specific East China region from the two cases. For the convenience of observation and analysis, we also add the corresponding Ground-Truth (GT) and precipitation forecasts before correcting (IFS-Lp). Among the pictures in the two rows, the areas surrounded by colored ovals represent three subregions of interest for qualitative analysis discussed in Sec. V-F. Besides, the black ovals are the subregions that have the missed correction for SSAS. The timestamp of the case is displayed at the bottom of every row. The color map on the right of the figure is used for distinguishing rainfall intensity by changing light color to dark color.

TABLE VI

The ablated experiments of loss function of S3M and TS2M on 80% samples of HR-ECb. The S3M w/o KL is the S3M without KL divergence when fixing the loss of TS2M. Similarly, The TS2M w/o KL/MAE is TS2M without KL divergence and MAE when fixing the loss of S3M. The results of SSAS (Baseline) with entire loss functions are regarded as benchmarks

		Criteria					
Datasets	Methods	MAE	RMAE	MAPE	$TS_{0.1}$	TS_1	TS_{10}
HR-ECb	S3M w/o KL	0.99±0.04	8.58±0.71	2.79±0.09	0.61±0.03	0.54±0.02	0.31±0.03
	TS2M w/o KL	0.97±0.04	7.95±0.67	2.73±0.08	0.63±0.03	0.56±0.02	0.32±0.03
	TS2M w/o MAE	0.95±0.05	6.61±0.54	2.68±0.09	0.66±0.03	0.58±0.02	0.35±0.03
	SSAS (Baseline)	*0.91±0.03*	*4.96±0.34*	*2.56±0.07*	*0.69±0.02*	*0.60±0.02*	*0.39±0.03*

correction methods. By noticing the change of pictures from left to the right in these two rows, the performance of correction for the different methods has obvious discrimination. Further, from IFS-Lp to SSAS, we observe that the given subregions marked by the colored ovals are gradually filled with the different colors of color-patch and adjusted their shapes, to approach the same subregions from precipitation GT. It indicates that the four methods on BCoP work to some extent. Concretely, in the red subregions in the two rows, SSAS can almost correct heavy rain (purple) accurately while the other three methods cannot correct exactly. Besides, toward the brown regions in the two rows, SSAS has the preferable correction results in both shapes and locations. This suggests that our method based on adaptive ST scales can not only boost the accuracy of correction but also help to refine the ST granularity for station-wise precipitation. However, the black regions report the missed correction from our methods. Therefore, the robustness of SSAS needs to be enhanced for

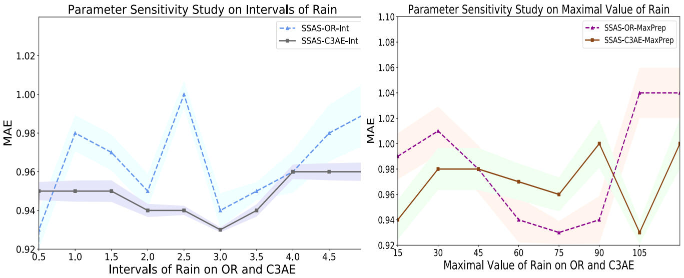


Fig. 8. **Comparisons** of BCoP performance (**MAE**) under different intervals and max value of rain among Ordinal Regression (OR) and Cascade Context-based Age Estimation (C3AE). Shallowed area represented ± standard deviation.

adapting to different cases.

To select the adaptive hyperparameters of SSAS to boost the generalization ability, we conduct the Parametric Sensitivity Analysis (PSA) to 4 key hyperparameters of modules and a combination of weights of our loss functions on the validation set of **HR-ECb**. We investigate all criteria for each study with the 5-fold cross-validation.

PSA on intervals of rain: For studying the sensitivity of SSAS to its four key hyperparameters, i.e., 1) the rank interval of rain in Ordinal Regression (OR) ξ_o, 2) the bin interval of rain ξ_c in Cascade Context-based Age Estimation (C3AE), 3) maximal value of rain in OR ρ_o 4) and maximal value of rain in C3AE ρ_c. Concretely, when we vary ξ_o (ξ_c) or ρ_o (ρ_c), we keep the others fixed. By setting ρ_o = ρ_c = 80, we plot the **MAEs versus** different intervals on OR and C3AE shown on the upper left of Fig. 8. It is easily observed that the best **MAE** results are obtained when $\xi_o = 0.5$ and $\xi_c = 3$. We also notice that the minimum intervals cannot obtain the best performance for both OR and C3AE. One possible reason is that these fine-grained divided labels contain too many zero/one elements, which is easy to lead to over-fitting problems and therefore influences the performance. Similarly, by fixing the ξs are equivalent to 0.5 and 3, we also plot the **MAEs versus** the different maximal values of rain on OR and C3AE shown on the upper right of Fig. 8. We notice that SSAS usually achieves the best performance when $\rho_o = 75$ and $\rho_c = 105$ respectively. For unifying the value range of rain among OR and C3AE, we average the sum of ρ_o and ρ_c, then obtain the $\rho_m = 90$ as the maximal value of rain.

PSA on the combination of weights: To select the adaptive combination of weights for loss functions of S3M (Eq (3)) and TS2M (Eq (5)). Heuristically, as shown in Table VII, the 3 groups of combinations of ω_1, ω_2 of Eq (3)) and same groups for ω_3, ω_4 of Eq (5) are regarded as candidates in this study. First, 1) we observe that it has little effect on the performance of our model for adopting the different combinations. In contrast, the standard deviations of the combination change more sharply. The reason is that the improper combination has negative effects on balancing the order of magnitudes between different losses, which may increase the standard deviation.

2) Besides, different combinations of S3M loss with a more obvious change in performance, compared with the same case of TS2M. One possible reason is that the S3M module itself has greater performance gain than the contribution of TS2M. This explanation has been proved in **Impact of S3M and TS2M** and shown in the second row to the third row of Table V. 3) Finally, we select the combination $\{0.2, 0.2\}$ for losses of S3M and TS2M because our model with this combination has better performance and smaller standard deviation.

TABLE VII

The parametric sensitivity for the weights of loss function of S3M and TS2M on the validation set of HR-ECb. The $\{\cdot, \cdot\}$ represents the combination of weights ω_1, ω_2 in loss of S3M when the loss of TS2M is obtained with the same weights, and the same as loss of TS2M when fixing the losses of S3M.

		Criteria					
Loss	Combinations	MAE	RMAE	MAPE	$TS_{0.1}$	TS_1	TS_{10}
S3M	{0.24,0.04}	0.95±0.03	5.04±0.36	2.68±0.07	0.66±0.02	0.56±0.02	0.35±0.03
	{0.1,0.6}	0.96±0.06	5.11±0.42	2.70±0.10	0.67±0.03	0.56±0.04	0.34±0.04
	{0.2,0.2}	**0.94±0.03**	**4.97±0.30**	**2.65±0.07**	0.67±0.02	0.55±0.02	**0.36±0.03**
TS2M	{0.24,0.04}	0.95±0.03	4.91±0.36	2.65±0.08	0.66±0.02	0.56±0.02	0.35±0.03
	{0.1,0.6}	0.95±0.03	4.95±0.34	2.68±0.06	0.65±0.02	0.54±0.03	0.34±0.03
	{0.2,0.2}	**0.94±0.03**	4.97±0.30	2.65±0.07	**0.67±0.02**	**0.55±0.02**	**0.36±0.03**

G. Statistical Tests

To verify the significance of testing results of our model, we perform a two-tailed T-test like air quality prediction [57] to verify the significance of **MAE** metric for Spatiotemporal Scale Adaptive Selection (SSAS) on HR-ECb. In the statistical tests, the value of the two-tailed T-test ($p < 0.01$) is greater than the critical value of the table lookup when the significance degree $\alpha = 0.01$. We set the number of tests m to be 20 (tested once every 3 training epochs). Here the critical value of this test τ_t is calculated with 3 formulas in the sequence below:

$$\begin{aligned} \mu &= \frac{1}{m}\sum_{i=1}^{m} \hat{\varepsilon}_i = 0.9246 \ , \\ \sigma^2 &= \frac{1}{m-1}\sum_{i=1}^{m} (\hat{\varepsilon}_i - \mu)^2 = 3 \times 10^{-4} \ , \\ \tau_t &= \sqrt{m}|\mu - \varepsilon_0|/\sigma = 21.8529 \ , \end{aligned} \tag{12}$$

where $\hat{\varepsilon}_i$ represents the i-th test **MAE** value and assumed maximum **MAE** value ε_0 is 1.0092 calculated by given critical value $\tau_g = 2.845$ by looking up the two-tailed **T-test Table**. The critical value τ_t is larger than τ_g. The results show that the performance of SSAS is less than the assumed test **MAE** ε_0 with a confidence degree ($1 - \alpha = 0.99$). Therefore, we prove that the significant difference between **MAEs** under 20 tests can be restrained to a given range.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed the SSAS approach that can automatically select the optimal ST scales of the predictors for capturing the high-level representations on BCoP.

Because the ST scales of predictors related to precipitation always continuously change over time and location, our method effectively extracts both optimal temporal and spatial features based on adaptive ST scales for station-wise precipitation. Our experiments on two benchmark datasets over Eastern China indicate that SSAS achieves the highest threat score (TS) on BCoP than eleven other published algorithms, and has a better correction ability in dealing with different intensities of precipitation, especially for heavy precipitation.

In the future, we will explore a more flexible strategy inspired by self-supervision learning [58] to remove ST scales around the precipitation (*scale-free*). Concretely, we can directly learn the chaotic and sophisticated regime of the predictors related to different precipitation by pretraining under the various predictors augmenting tasks such as rotation, deformation, and permutation without a label or with a few labels. Designing such a system is beyond the scope of this work. Besides, from the perspective of the early-warning system, we will investigate how to employ the adaptive strategy on more complex heavy precipitation scenarios such as squall lines, severe convection, and severe thunderstorms.

REFERENCES

[1] Insurance Information Institute, "Natural catastrophes 2019." https://www.iii.org/fact-statistic/facts-statistics-global-catastrophes, 2019.

[2] Z. Liu, A. Sha, L. Hu, Y. Lu, W. Jiao, Z. Tong, and J. Gao, "Kinetic and thermodynamic modeling of Portland cement hydration at low temperatures," *Chemical Papers*, vol. 71, no. 4, pp. 741–751, 2017.

[3] I. M. Navon, "Data assimilation for numerical weather prediction: a review," in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, pp. 21–65, Springer, 2009.

[4] C. Zhang, J. Zeng, H. Wang, L. Ma, and H. Chu, "Correction model for rainfall forecasts using the LSTM with multiple meteorological factors," *Meteorological Applications*, vol. 27, no. 1, p. e1852, 2020.

[5] Q. Miao, B. Pan, H. Wang, K. Hsu, and S. Sorooshian, "Improving monsoon precipitation prediction using combined convolutional and long short term memory neural network," *Water*, vol. 11, no. 5, p. 977, 2019.

[6] F. J. Tapiador, R. Roca, A. Del Genio, B. Dewitte, W. Petersen, and F. Zhang, "Is precipitation a good metric for model performance?," *Bulletin of the American Meteorological Society*, vol. 100, no. 2, pp. 223–233, 2019.

[7] D. J. Stensrud and N. Yussouf, "Short-range ensemble predictions of 2-m temperature and dewpoint temperature over New England," *Monthly Weather Review*, vol. 131, no. 10, pp. 2510–2524, 2003.

[8] C. Piani, J. Haerter, and E. Coppola, "Statistical bias correction for daily precipitation in regional climate models over Europe," *Theoretical and Applied Climatology*, vol. 99, no. 1-2, pp. 187–192, 2010.

[9] G. Lenderink, A. Buishand, and W. Van Deursen, "Estimates of future discharges of the river rhine using two scenario methodologies: direct versus Delta approach," *Hydrology and Earth System Sciences Discussions, European Geosciences Union*, vol. 11, no. 3, pp. 1145–1159, 2007.

[10] R. Leander and T. A. Buishand, "Resampling of regional climate model output for the simulation of extreme river flows," *Journal of Hydrology*, vol. 332, no. 3-4, pp. 487–496, 2007.

[11] J. P. Vidal and S. Wade, "A framework for developing high-resolution multi-model climate projections: 21st century scenarios for the UK," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 28, no. 7, pp. 843–858, 2008.

[12] L. J. Wilson and M. Vallée, "The Canadian updateable model output statistics (UMOS) system: design and development tests," *Weather and Forecasting*, vol. 17, no. 2, pp. 206–222, 2002.

[13] Q. Mao, R. T. McNider, S. F. Mueller, and H.-M. H. Juang, "An optimal model output calibration algorithm suitable for objective temperature forecasting," *Weather and Forecasting*, vol. 14, no. 2, pp. 190–202, 1999.

[14] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[15] S. Moghim and R. L. Bras, "Regression-based regionalization for bias correction of temperature and precipitation," *International Journal of Climatology*, vol. 39, no. 7, pp. 3298–3312, 2019.

[16] M. Mu, W. Duan, and B. Wang, "Conditional nonlinear optimal perturbation and its applications," *Nonlinear Processes in Geophysics*, vol. 10, no. 6, pp. 493–501, 2003.

[17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[19] X. Xu, Y. Liu, H. Chao, Y. Luo, H. Chu, L. Chen, J. Zhang, and L. Ma, "Towards a precipitation bias corrector against noise and maldistribution," *ArXiv Preprint ArXiv:1910.07633*, 2019.

[20] H. U. Kim and T. S. Bae, "Preliminary study of deep learning-based precipitation," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 35, no. 5, pp. 423–430, 2017.

[21] D. Li, J. Sun, S. Fu, J. Wei, S. Wang, and F. Tian, "Spatiotemporal characteristics of hourly precipitation over central Eastern China during the warm season of 1982–2012," *International Journal of Climatology*, vol. 36, no. 8, pp. 3148–3160, 2016.

[22] S. Fu, R. Liu, and J. Sun, "On the scale interactions that dominate the maintenance of a persistent heavy rainfall event: a piecewise energy analysis," *Journal of the Atmospheric Sciences*, vol. 75, no. 3, pp. 907–925, 2018.

[23] Y. Zhang, J. Sun, and S. Fu, "Main energy paths and energy cascade processes of the two types of persistent heavy rainfall events over the Yangtze River–Huaihe River Basin," *Advances in Atmospheric Sciences*, vol. 34, no. 2, pp. 129–143, 2017.

[24] M. Fayyaz and J. Gall, "SCT: Set constrained temporal transformer for set supervised action segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 501–510, 2020.

[25] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 839–847, 2018.

[26] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3322–3332, 2018.

[27] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. W. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 2627–2633, 2017.

[28] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[29] S. Xingjian, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing System*, pp. 802–810, 2015.

[30] L. Chen, Y. Cao, L. Ma, and J. Zhang, "A deep learning-based methodology for precipitation nowcasting with radar," *Earth and Space Science*, vol. 7, no. 2, p. e2019EA000812, 2020.

[31] D. P. Dee, S. M. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, d. P. Bauer, *et al.*, "The ERA-Interim reanalysis: Configuration and performance of the data assimilation system," *Quarterly Journal of The Royal Meteorological Society*, vol. 137, no. 656, pp. 553–597, 2011.

[32] S. You, T. Huang, M. Yang, F. Wang, C. Qian, and C. Zhang, "GreedyNAS: Towards fast one-shot NAS with greedy supernet," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 1999–2008, 2020.

[33] S. Zhang, W. Lu, W. Xing, and L. Zhang, "Learning scale-adaptive tight correlation filter for object tracking," *IEEE Transactions on Cybernetics*, vol. 50, no. 1, pp. 270–283, 2018.

[34] Y. Zhang, J. Sun, H. Liang, and H. Li, "Event-triggered adaptive tracking control for multiagent systems with unknown disturbances," *IEEE Transactions on Cybernetics*, 2018.

[35] C. Zhu, Y. He, and M. Savvides, "Feature selective anchor-free module for single-shot object detection," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 840–849, 2019.

[36] T. Bergeron, "ECMWF forecast user guide." https://www.ecmwf.int/en/about/media-centre/focus/user-guide-ecmwf-forecast-products, 2019.

[37] G. Balsamo, E. Dutra, C. Albergel, S. Munier, J.-C. Calvet, J. Munoz-Sabater, and P. de Rosnay, "ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better?," *Hydrology and Earth System Sciences*, vol. 22, no. 6, pp. 3515–3532, 2018.

[38] C. M. Agency, "2019 CMA observation." http://www.nmc.cn/publish/satellite/FY4A-true-color.htm, 2019.

[39] C. Spearman, "The proof and measurement of association between two things. by c. Spearman, 1904.," *American Journal of Psychology*, vol. 100, no. 3-4, pp. 441–471, 1987.

[40] R. P. Freckleton, "Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error," *Behavioral Ecology and Sociobiology*, vol. 65, no. 1, pp. 91–101, 2011.

[41] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 764–773, 2017.

[42] H. Millán, A. Kalauzi, M. Cukic, and R. Biondi, "Nonlinear dynamics of meteorological variables: multifractality and chaotic invariants in daily records from Pastaza, Ecuador," *Theoretical and Applied Climatology*, vol. 102, no. 1-2, pp. 75–85, 2010.

[43] C. Zhang, S. Liu, X. Xu, and C. Zhu, "C3AE: Exploring the limits of compact model for age estimation," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 12587–12596, 2019.

[44] T. Van Erven and P. Harremos, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.

[45] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Advances in Neural Information Processing Systems*, pp. 8778–8788, 2018.

[46] L. Zhang, G. Zhu, P. Shen, J. Song, S. Afaq Shah, and M. Bennamoun, "Learning spatiotemporal features using 3DCNN and convolutional LSTM for gesture recognition," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 3120–3128, 2017.

[47] H. Zhu, Q. Zhou, J. Zhang, and J. Z. Wang, "Facial aging and rejuvenation by conditional multi-adversarial autoencoder with ordinal regression," *ArXiv Preprint ArXiv:1804.02740*, 2018.

[48] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 2980–2988, 2017.

[49] Y. Wu and K. He, "Group normalization," in *Proceedings of the European Conference on Computer Vision*, pp. 3–19, 2018.

[50] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[52] D. Cho, C. Yoo, J. Im, and D. Cha, "Comparative assessment of various machine learningbased bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas," *Earth and Space Science*, vol. 7, no. 4, 2020.

[53] D. Wolfensberger, M. Gabella, M. Boscacci, U. Germann, and A. Berne, "Rainforest: A random forest algorithm for quantitative precipitation estimation over Swizerland," *Atmospheric Measurement Techniques Discussions*, pp. 1–35, 2020.

[54] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, 2017.

[55] C. Yang, Y. Xu, J. Shi, B. Dai, and B. Zhou, "Temporal pyramid network for action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[56] Y. Liu, Y. He, S. Li, Z. Dong, J. Zhang, and U. Kruger, "An auto-adjustable and time-consistent model for determining coagulant dosage based on operators' experience," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

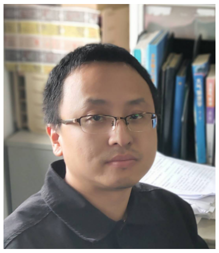
[57] H. Jun and W. Zheng, "Multistage attention network for multivariate time series prediction," *Neurocomputing*, vol. 383, pp. 122–137, 2020.

[58] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," *arXiv preprint arXiv:2002.05709*, 2020.

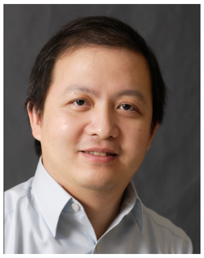
Yiqun Liu received an M.S. degree in Computer Technology from the School of Information Science and Engineering, Lanzhou University, Lanzhou, Gansu Province, China, in 2017. He is currently pursuing a Ph.D. degree at the School of Computer Science, Fudan University, Shanghai, China. His research interests include machine learning, smart city, and spatiotemporal prediction.

Junping Zhang (M'05) received a B.S. degree from Xiangtan University, P.R. China, in 1992, and obtained an M.S. degree from Hunan University, Changsha, P.R. China, in 2000. In 2003, Dr. Zhang received a Ph.D. degree from the Institution of Automation, Chinese Academy of Sciences. In 2006, he joined Fudan University as an Associate Professor and became a Professor in the School of Computer Science in 2011. His research interests include machine learning, intelligent transportation systems, image processing, and biometric authentication. His research results have been published in highly-ranked international journals including the IEEE Journals IEEE TPAMI, TNN, TSMCB, and TITS, and the leading conferences ICML, CVPR, AAAI, and ECCV. Dr. Zhang is an associate editor of IEEE Intelligent Systems.

Lei Chen received a Ph.D. degree from the University of Chinese Academy of Sciences in 2015. In the same year, he joined Shanghai Central Meteorology Observatory (SCMO) as a weather forecaster. He is now an engineer in the technology development department of SCMO. His research interest is the application of deep learning technology in nowcasting.



Hai Chu received an M.S. degree from the School of Atmosphere of Nanjing University in 2011. In the same year, he joined Shanghai Central Meteorology Observatory (SCMO) as a weather forecaster. He is now deputy senior engineer and section chief of the technology development department of SCMO. His research interests include nowcasting, mesoscale weather forecasting, and numerical weather prediction calibration.



James Z. Wang is a Professor of Information Sciences and Technology at The Pennsylvania State University. He received the bachelor's degree in mathematics and computer science summa cum laude from the University of Minnesota, and the M.S. degree in mathematics, the M.S. degree in computer science, and the Ph.D. degree in medical information sciences, all from Stanford University. His research interests include image analysis, image modeling, image retrieval, and their applications. He was a visiting professor at the Robotics Institute at Carnegie Mellon University (2007-2008), a lead special section guest editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence (2008), and a program manager at the Office of the Director of the National Science Foundation (2011-2012). He was a recipient of a National Science Foundation Career award (2004) and the Amazon Research Awards (2018, 2019, and 2020).

Leiming Ma received a Ph.D. degree from the School of Atmosphere of Nanjing University in 2009. In 2017, he joined Shanghai Central Meteorological Observatory (SCMO) as a deputy director, before which he had been a senior researcher (2005)/researcher (2000) in Shanghai Typhoon Institute (STI). He is now a professor/senior researcher and director of SCMO. His research interests include numerical weather prediction, tropical cyclones, atmospheric convection, and nowcasting.