Robust Precipitation Bias Correction Through an Ordinal Distribution Autoencoder

Youcheng Luo, Xiaoyang Xu, Yiqun Liu, Hanqing Chao, Hai Chu, Lei Chen, Junping Zhang*, *Member, IEEE,* Leiming Ma, and James Z. Wang

***Abstract*—Numerical precipitation prediction plays a crucial role in weather forecasting and has broad applications in public services including aviation management and urban disaster early warning systems. However, Numerical Weather Prediction (NWP) models are often constrained by a systematic bias due to coarse spatial resolution, lack of parametrizations, and limitations of observation and conventional meteorological models, including constrained sample size and long-tail distribution. To address these issues, we present a data-driven deep learning model, named the Ordinal Distribution Autoencoder (ODA), which principally includes a Precipitation Confidence Network and a combinatorial network that contains two blocks, *i.e.*, a Denoising Autoencoder block and an Ordinal Distribution Regression block. As an expert-free model for bias correction of precipitation, it can effectively correct numerical precipitation prediction based on meteorological data from the European Centre for Medium-Range Weather Forecasts (ECMWF) and SMS-WARMS, an NWP model used in East China. Experiments in the two NWP models demonstrate that, compared with several classical machine learning algorithms and deep learning models, our proposed ODA generally performs better in bias correction.**

***Index Terms*—Precipitation bias correction, ordinal distribution autoencoder, weather prediction, deep learning.**

I. Introduction

As an indispensable public service, reliable forecasting of precipitation is significant in many aspects of society such as emergency management and disaster early warning. Providing accurate meteorological predictions for the coming six hours can greatly help to warn local people and to make better decisions.

Numerical weather prediction (NWP) estimates the incoming state of the various weather components, including precipitation, from several minutes to several days quantitatively with aerodynamic and thermodynamic models. However, NWP models always contain two kinds of biases: 1) systematic bias due to coarse spatial resolution and imperfectness of parameterizations, and 2) observation bias due to systematic representativeness errors in surface observations [1][1].

To reduce the bias, postprocessing of the model output is important for further usage of models. Several statistical methods have been used for the bias correction of prediction produced from NWP models [2], [3]. These techniques have been applied to NWP models in several countries to improve performance.

Corresponding author: Junping Zhang

[1]According to the prediction time range, weather prediction can be categorized into nowcasting (0-2 hours), and very short-range (0-12 hours), short-range (12-72 hours), medium-range (72-240 hours), extended-range (10-30 days), and long-range (>30 days) forecasting.

ANNs, as opposed to traditional methods in meteorology, are based on self-adaptive mechanisms that learn from the examples and capture functional relationships between data, even if the relationships are unknown or difficult to describe [4].

In this experimental study, we use a set of real-world data from hundreds of meteorological stations located in Southeast China as labeled precipitation and try to enhance the 6-hour precipitation prediction from ECMWF and SMS-WARMS with 37 physical predicted parameters as input which also suffer from noise. Besides, treating this task as a regression problem would be less effective due to the highly imbalanced data distribution and a large span between the maximum and minimum precipitations.

To address the aforementioned issues, we transform the original regression problem into a classification one. We propose a novel deep convolutional framework named Ordinal Distribution Autoencoder (ODA) for bias correction of precipitation prediction through effective designing deep model architecture and corresponding loss function, the focal loss [5]. Compared with the conventional regression strategy, the proposed network is superior in tasks with long-span data and can take full advantage of the majority of factors that may influence precipitation.

The main **contributions** of this paper are summarized as follows:

1) We propose a novel deep learning model to solve the bias-correction problem instead of conventional machine learning methods or ensemble models, and achieves better correcting performance and meteorological index.
2) The Ordinal Distribution Regression transfers the regression problem into several binary classification subproblems, greatly alleviating the impact of the large span. To the best of our knowledge, it is the first time to introduce an ordinal regression model to correct the bias of EC precipitation models. Meanwhile, a focal loss in this block is introduced to address the issue of the long-tail distribution of rainfall.
3) The evaluation in different areas and different NWP models shows the robustness and effectiveness of our proposed network in correcting the 6-hour precipitation prediction.

II. Related Work

A. Bias Correction of Numerical Weather Prediction

The most commonly used bias-correction methods in the precipitation prediction fields are the Model Output Statistics

(MOS) and Kalman Filter (KF) techniques. As a statistical post-processing method, MOS is frequently used to improve the forecast of NWP models through applying a statistical model (mostly multi-linear regression), to correct the NWP model output according to the past model results and observation data [2]. However, linear regression restricts the performance of MOS methods. As an alternative recursive nonlinear algorithm, KF can estimate a signal from noisy measurements, which has been widely used in bias-correcting the prediction of NWP models [3]. Although these conventional methods are feasible based on few features, they cannot construct models for a variety of stations. Without spatial information, further, the performance of these models is limited.

To address these issues, machine learning techniques have been applied to improve the prediction accuracy under various types of high-impact weather phenomena. For example, Support Vector Regression (SVR) and Random Forest (RF) are two typical methods for bias correction [6]. Different from MOS and KF, these machine learning methods are less sensitive to the correlation among input features and can deal with multiple input variables as well as utilize the spatial distributions of parameters from several stations for bias correction.

It is worth noting that the bias-correcting ability of a single machine learning method is limited due to complex atmosphere surface interactions. Although ensemble models can take advantage of different machine learning approaches and make the bias correction more accurate [7], this kind of combination greatly relies on expert experience and thus works only in some specific scenarios.

In recent years, Deep Neural Network (DNN) has been used as an effective technique for the prediction or correction of precipitation [8]. Because of its end-to-end manner, deep learning can automatically extract representative features to assist estimation, less depending on expert knowledge than those in (ensemble) machine learning methods. For example, Snderby et al. used sequences of maps of precipitation to predict future precipitation. Besides, axial self-attention is utilized to aggregate the global context [9]. Wang et al. [10] focused on prediction through using precipitation from different stations in the past 9 days to predict the precipitation of the next 1/2/.../9 days, NLE loss is proposed in this paper to improve the generalization of point estimation. However, these prediction methods are to forecast the precipitation instead of bias-correcting the precipitation, greatly restricting the performance of these methods to output corrected precipitation. Moreover, few parameters taking as input also limits the performance.

For bias correction, Tao et al. proposed a DNN framework with a denoising Autoencoder and FCN with four layers to improve the accuracy of satellite precipitation products by reducing the bias and false alarms [11]. Zjavka found that a polynomial neural network could successfully bias-correct the National Oceanic and Atmospheric Administration (NOAA) mesoscale model [12]. However, these methods can hardly correct the precipitation of different levels at the same time, as they both treat the correction problem as a regression one and suffer from the large span of precipitation.

B. Ordinal Regression

Ordinal Regression aims at finding a function to predict the labels in ordinal variables. Most ordinal regression algorithms are derived from conventional classification algorithms. For instance, Herbrich et al. [13] proposed a support vector method for ordinal regression. And Shashua and Levin [14] developed a novel support vector machine to handle multiple thresholds.

An alternative way is to regard ordinal regression as a series of binary classification problems. Ordinal regression is proved useful in different areas, Niu et al. [15] combined ordinal regression with Convolutional Neural Networks (CNNs) to estimate facial age. Fu et al. [16] adopted a similar strategy for monocular depth estimation. Considering the uneven distribution of precipitations, we thus utilize ordinal regression for better precipitation prediction correction.

III. The Framework

In this section, we present the ODA for bias correction of numerical precipitation prediction. After defining this task, we introduce our datasets, the preprocessing method, and the framework of the proposed DNN. Then we describe the design for handling the noisy features. To ease the impact of the large span and uneven distribution of precipitation, further, an Ordinal Distribution Regression block is implemented, which will be described in detail in Section III-E2.

A. Problem Definition

Generally, bias correction of numerical precipitation prediction can be regarded as a regression problem. Specifically, let $\mathcal{X}$ and $\mathcal{Y}$ denote the input and output spaces, respectively. Each input sample $x \in \mathcal{X}$ consists of N features $\{\boldsymbol{M}_1, \boldsymbol{M}_2, \cdots, \boldsymbol{M}_N\}$ which are predicted by NWP for the coming 6-th hour. Here, each $\boldsymbol{M}_k$ is a 2-D matrix with size $l \times l$, with each element recording a value of the k-th feature on the corresponding latitude and longitude. Each output $y \in \mathcal{Y}$ is a scalar representing the corrected accumulative total precipitation for 6-hour.

All features in x are the multiple predicted meteorological parameters generated by the NWP model (*i.e.*, ECMWF or SMS-WARMS) including predicted precipitation. The task is to correct the predicted prediction by learning a mapping function $g : x \rightarrow y$.

B. Data Preprocessing

We evaluate ODA on a 6-hour Integrated Forecasting System (IFS) data provided by the ECMWF and SMS-WARMS V2 [17], which is collected from Shanghai Central Meteorological Observatory (SCMO). The data from ECMWF are filled in a chart of the world with latitude-longitude grid cells of 0.125° of latitude by 0.125° of longitude, and the frequency is one sample every three hours. And the precise precipitation y is provided by SCMO, precisely, collected from the ground stations every six hours. Although the geographical coordinates of the ground stations mismatch those of the grid data, given the high density of the grid, we make a reasonable

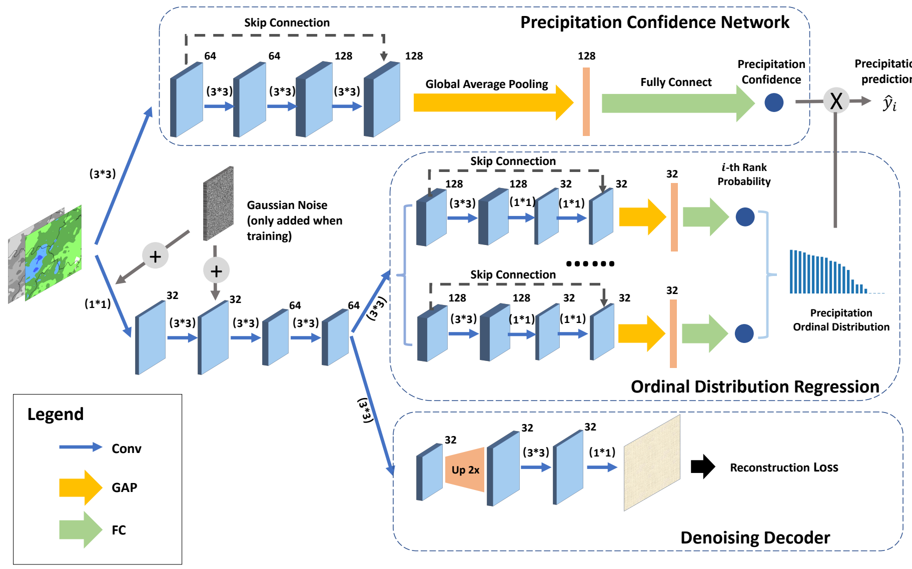

Fig. 1. An overview of the proposed approach for bias correction of numerical precipitation prediction. 'AP' represents Average Pooling and 'FC' is the abbreviation of the Fully Connection Layers. Squared blocks represent input data or feature maps of convolutional blocks in our model, while circles are the output scalars of different sub-nets. The output vectors of the Ordinal Distribution Regression blocks form the ordinal distribution histogram.

assumption that the precise precipitation inside a grid cell is static.

For bias correction, we first filter out some meteorological parameters according to the ranks of correlation with precipitation, which are calculated with the Pearson correlation coefficient as the input features. The picking threshold is set as 0.2, and for parameters with the same name but different heights, we retain the one with the highest coefficient. Then, considering the spatial impact of rainfalls, we slice the entire grid matrix to several smaller ones centering at the ground observatory with the slicing window of 2.0° latitude by 2.0° longitude. After pre-processing the entire grid, we choose 37 different features, and every input sample for the model is resized to $37 * 17 * 17$.

Besides, when preprocessing the data from SMS-WARMS, some of the parameters used in ECMWF are replaced with other similar parameters because of the different data provided.

In our experiment, the whole dataset consists of 237,498 samples of Eastern China collecting from July to September in both 2016 and 2017. We got 4 samples one day in the training process as the output of ground stations is utilized as the label, and when being used online or for testing, this model would output the corrected predicted results 8 times per day. The ratio of the training set to the testing set is 4:1, and samples are randomly selected without overlapping.

C. Framework Design

The framework shown in Fig. 1 includes two parts: a Precipitation Confidence Network as a binary classifier to judge whether precipitation exists, and a combinatorial network consisting of two blocks: a Denoising Autoencoder for robust features extraction, and multiple Ordinal Distribution Regression blocks for ordinal distribution.

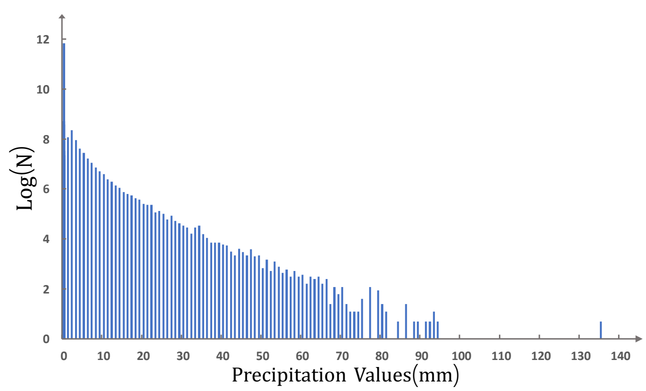

Fig. 2. The distribution of ground-truth precipitation in our dataset. The y-axis represents the logarithm of the number of each precipitation, while the x-axis represents the range of the ground-truth precipitation.

The Precipitation Confidence Network is a 4-layer Convolution Network with an average pooling and a full connection with an output of 0/1 as $\boldsymbol{O}_1$.

The combinatorial network consists of a Denoising Autoencoder block for feature generation and an Ordinal Distribution Regression block for precipitation correction, which is the main body of our framework. This part is trained with the multiplication of construction loss from Autoencoder and focal loss from distribution classification with the output of accurate classification of precipitation as $\boldsymbol{O}_2$, which can be seen as corrected precipitation value. The final output should be

$$\boldsymbol{O} = \boldsymbol{O}_1 \times \boldsymbol{O}_2 \,. \quad (1)$$

D. Precipitation Confidence Network

As shown in Fig. 2, nearly 78% of samples in our dataset are non-precipitation, and in only 1% of samples, the precipitation is larger than 10 mm. Because of the imbalanced distribution of precipitation, we create the Precipitation Confidence Network to identify non-precipitation and precipitation. Then, the Ordinal Distribution Regression module can focus on the finer classification of precipitation for achieving better bias-correction performance.

The Precipitation Confidence Network is a 4-layer Convolution Network, while its details are described in Table I. This part is trained individually using binary classification error (BCE) and aims at distinguishing whether precipitation exists. Then, we can only send the rainfall samples to the Ordinal Regression module when training.

E. Combinatorial Network

The combinatorial network aims at correcting the predicted precipitation. With input x, the Denoising Autoencoder first extracts noisy robust features as bottleneck output. Then, Ordinal Distribution Regression takes this output as input to generate corrected precipitation.

1) Denoising Autoencoder: If input features contain observational and systematic errors, it will mislead the regression model to generate precipitation with bias. To address this problem, we refine a Denoising Autoencoder with a fully convolutional neural network by introducing a noisy level, and formulate it as follows:

$$z_i = g_{E_1}\left(x_i + \varepsilon \parallel \theta\right) \ , \tag{2}$$

$$h_i = g_{E_2}\left(z_i \parallel \theta'\right) \ , \tag{3}$$

$$\hat{x}_i = g_D\left(h_i \parallel \theta''\right) \ , \tag{4}$$

where $\varepsilon \sim N\left(\mu, \sigma\right)$ is random normal noise. $g_{E_i}\left(\cdot\right)$ denotes each part of the enhanced denoising encoder, $g_D\left(\cdot\right)$ denotes the enhanced denoising decoder, and θ, θ', and θ'' denote their respective parameters. z_i is a feature map of the perturbation layer with random normal noise added which only works when training. h_i is the bottleneck output of Denoising Autoencoder. Besides, to ensure the bottleneck output h_i precisely represents the input x, the gap between x and the output of Decoder $\hat{x}_i$ should be as small as possible. With such a setting, the objective function of the Denoising Encoder-Decoder is formulated as

$$\theta^*, \theta'^*, \theta''^* = \underset{\theta, \theta', \theta''}{\arg\min} \sum_{i}^{N} E_x \left\| x_i - \hat{x} \right\|_2^2 \ . \tag{5}$$

During the test, we take original samples x_i as inputs and remove the perturbation layer. h_i can be regarded as noise-free and squeezed features utilized in the Ordinal Distribution Regression block to correct the precipitation.

2) Ordinal Distribution Regression: As a regression problem, one common objective function is the least square loss:

$$\frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2 \ . \tag{6}$$

However, due to the large span and long-tail distribution of rainfall, learning such a distribution with Eq. (6) forces the regression model to favor small precipitation, while the large precipitation still produces a large loss, making a model hard to converge to a global optimum.

Furthermore, precipitation is a typical ordinal variable having an obvious ordering relation on an arbitrary scale. Therefore, we adopt multiple Ordinal Distribution Regression blocks to correct precipitation. Similar to [15], we transform a regression problem into a series of binary classification problems. In our work, we partition the 6-hour cumulative precipitation ranging from $y_{\min}$ to $y_{\max}$ into K ranks $r_i \in \{r_1, r_2, \cdots, r_{K-1}\}$ followed by training $K-1$ binary classifiers with an interval 0.5 to judge what the rank of the predicted values of sample x_i belongs to.

More concretely, we first transform the continuous ground-truth precipitation y_i to discrete binary vectors $\mathbf{y}_i = \{d_1, d_2, \cdots, d_{K-1}\}$, where d_i is calculated as follows:

$$d_i = \begin{cases} 1, & \text{if} \quad y_i > r_k \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

Then, we directly utilize $K-1$ four-layers CNNs as the distribution subnets, each of which generates a scalar p_i representing the probability of $d_i = 1$. The details of each distribution subnet will be shown in Section IV-B. Finally, the Ordinal Distribution Regression blocks calculate the bias-corrected predicted precipitation of sample x_i as follows:

$$\tilde{y}_i = \eta * \sum_{i}^{K-1} \left(p_i \geq \xi\right) \ , \tag{8}$$

where η is the partitioning interval 0.5 and ξ is a threshold that is set to 0.5 in our experiment. In such a setting, the samples with larger precipitation contribute equally to the ordinal regression model as those models with smaller values. As a result, the impact of the large span issue is alleviated. It is worth noting that the partitioning interval is larger than the minimal ranging interval in some cases, resulting in inevitable gaps between the predicted and ground-truth precipitation. Therefore, the partitioning interval is an important hyperparameter in our approach and will be discussed further in Section V-B.

3) Loss Function: The distribution of precipitation of samples that are larger than 10 mm is long-tail which further causes the ordinal ranked-classes r_i to be imbalanced. Focal loss, stemming from the cross-entropy loss function, is thus adopted in our approach for addressing such an imbalance issue. Similar to [5], we introduce two hyperparameters in cross-entropy loss function and rewrite it as:

$$\begin{aligned} \mathcal{L}_{\text{FOR}} = \frac{1}{N} \sum_{i=1}^{N} & (1-\boldsymbol{\alpha})\mathbf{p}_i^{\gamma}\mathbf{y}_i \log(1-\mathbf{p}_i) + \\ & \boldsymbol{\alpha}(1-\mathbf{p}_i)^{\gamma}(1-\mathbf{y}_i)\log(\mathbf{p}_i) \ , \end{aligned} \tag{9}$$

where both $\mathbf{y}_i$ and $\mathbf{p}_i$ are vectors representing the labels and probabilities of each ordinal rank class r_i, respectively. γ is a scalar for differentiating the samples and $\boldsymbol{\alpha}$ is a vector to balances each r_i rank class.

Therefore, the final loss function of the combinatorial network would be:

$$\mathcal{L} = \mathcal{L}_{\text{Cons}} + \mathcal{L}_{\text{FOR}} = \sum_{i}^{N} E_x \left\| x_i - \hat{x} \right\|_2^2 + \mathcal{L}_{\text{FOR}} \ , \tag{10}$$

where $\mathcal{L}_{\text{Cons}}$ represents the construction loss of denoising autoencoder.

IV. Experimental Results

In this section, we describe the evaluation metrics followed by model configurations and training details. After that, We report the experiments to validate the effectiveness of ODA in bias correction tasks.

A. Evaluation Metrics

We report the mean average error (MAE, Eq. (11)) and the mean positive average error (MPAE, Eq. (12)) as accuracy assessments:

$$MAE = \sum_{1}^{n} \frac{|\hat{y} - y_i|}{n} \ , \tag{11}$$

$$MPAE = \sum_{1}^{n} \frac{|\hat{y}_{positive} - y_i|}{n} \ , \tag{12}$$

where $\boldsymbol{y}_i$ is the corrected precipitation, $\hat{\boldsymbol{y}}_i$ is the ground truth precipitation, and n is the number of samples.

As an important meteorological index for precipitation prediction, the Threat Score (Ts) is calculated as:

$$Ts_\delta = \frac{TP_\delta}{TP_\delta + FP_\delta + FN_\delta} \ , \tag{13}$$

where the parameter δ defines the boundary between positive and negative samples. With different δ we can evaluate the model in a different range of precipitation as drizzle or torrential rain have different characteristics and models can hardly predict or correct both of them well, so we determined δ as 0.1, 1, and 10.

B. Model Configurations and Training Process

Details of model configurations are shown in Table I and each Conv block is equipped with a convolutional layer, a batch normalization layer, and a leaky ReLU with a negative slope of 0.01.

The Denoising Encoder contains four Conv blocks and one noise perturbation layer. We generate random noise from a normal distribution with $\mu = 0$ and $\sigma = 0.001$ during the training phase and remove them during the testing phase. The decoder consists of three up-sampling Conv blocks, whose sampling operation is bilinear interpolation with stride 2.

Although all of the Ordinal Distribution Regression subnets share one common configuration, their parameters are completely independent. Each ordinal subnet contains one skip connection to accelerate the training procedure and dropout with a ratio of 0.2 is used for relieving the overfitting of the fully connected layers (FC).

We adopt feature normalization before features are fed into the models and use the Adaptive moment estimation

TABLE I
The Model Details of the Framework

Network			**Input**: 37 × 17 × 17
Bonded Network	**Encoder**	Conv Block	$\begin{bmatrix} 32, & 1\times 1 \\ 32, & 3\times 3 \end{bmatrix} \times 1$
		Noise	$N(0, 1e-3)$
		Conv Block	$\begin{bmatrix} 64, & 3\times 3 \end{bmatrix} \times 2$
	Ordinal Subnets	Conv Block	$\begin{bmatrix} 128, & 3\times 3 \\ 128, & 3\times 3 \\ 32, & 1\times 1 \\ 32, & 1\times 1 \end{bmatrix} \times 1$
		Down Sampler	$\begin{bmatrix} 32, & 1\times 1 \end{bmatrix} \times 1$
		Binary Classifier	Average pooling, 1-D FC
	Decoder	Conv Block	$\begin{bmatrix} 64, & 3\times 3 \end{bmatrix} \times 1$
		Upsampling	Bilinear (17 × 17)
		Conv Block	$\begin{bmatrix} 32, & 3\times 3 \\ 37, & 1\times 1 \end{bmatrix} \times 1$
Precipitation Confidence Network		Conv Block	$\begin{bmatrix} 64, & 3\times 3 \\ 64, & 3\times 3 \\ 128, & 3\times 3 \\ 128, & 3\times 3 \end{bmatrix} \times 1$
		Down Sampler	$\begin{bmatrix} 128, & 1\times 1 \end{bmatrix} \times 1$
		Rainfall Classifier	Average pooling, 1-D FC

(Adam) [18] with a mini-batch of 256. The learning rate of the Denoising Autoencoder starts from $1e-3$ with a weight decay of $1e-2$. Both the learning rate and weight decay of the Ordinal Distribution Regression blocks are set to $1e-4$.

In testing, for comparison with other methods, we adopt standard 5-fold cross-validation in which the dataset is divided into 5 folds. Each time we choose one fold to test, and the other four folds to train, repeating this until 5 folds are all tested. And we repeat the experiment with different dataset divisions. Noted that all experiments in Section V are implemented with a fixed dataset of 80% for training and 20% for testing without overlapping.

C. Validation on $6-$hour IFS

We validate ODA on a $6-$hour IFS dataset by comparing it with several baseline methods. Original input provided by ECMWF and SMS-WARMS is the basic comparison. Bilinear interpolation (BI) means that we directly interpolate each sliced predicted precipitation grid to the corresponding ground observatory. Linear regression (LR) and Support Vector Regression (SVR) are commonly used machine-learning methods.

FPN [19] is one baseline deep learning model, in which the precipitation and other features are taken as different channels. As another baseline, the Convolutional Long Short-Term Memory (ConvLSTM) model [20] predicts meteorological features with spatiotemporal sequence forecasting. Furthermore, the networks of [8] and [11] are employed as two baseline deep learning models, which both use Denoising Autoencoder as a feature generator. Note that [8] used a Multilayer perceptron (MLP-A) with one hidden layer as the core module and [11] used a fully convolutional network with four Conv blocks as the core module (FC-A). Besides, we

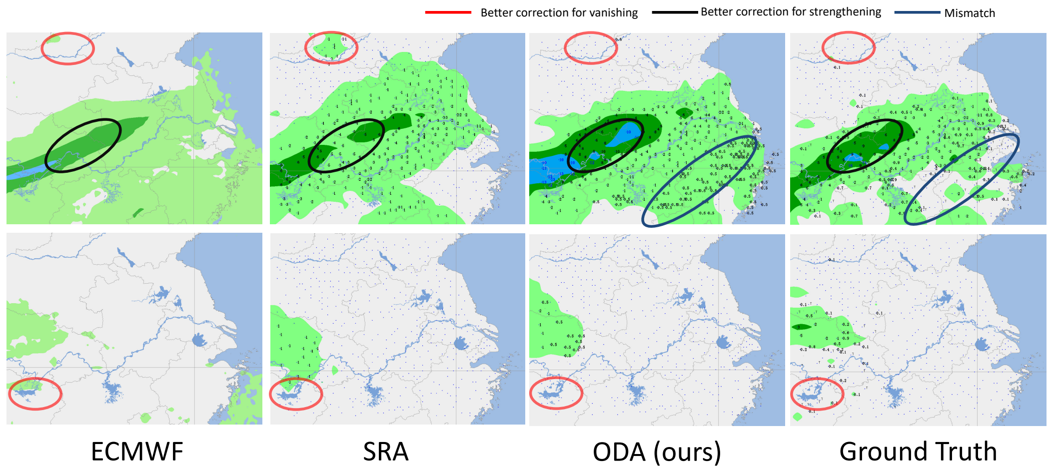

Fig. 3. Visualization of two examples of our correction results. Each row represents one example. While different circles mean different performance parts. First column: the raw ECMWF precipitation grid data. Second column: corrected precipitation of Single Regression Autoencoder. Third column: corrected precipitation of Ordinal Distribution Autoencoder. Last column: the ground-truth(GT) precipitation.

TABLE II
COMPARISON OF OUR PROPOSED MODEL AGAINST BASELINE METHOD ON 6−HOUR PREDICTION OF ECMWF

Methods	MAE	MPAE	$Ts_{0.1}$	Ts_1	Ts_{10}
BI	1.31 ± 0.01	4.53 ± 0.1	$0.44 \pm 3e-4$	$0.44 \pm 4e-4$	$0.24 \pm 1e-4$
LR	$1.6 \pm 8e-3$	4.43 ± 0.08	$0.31 \pm 7e-4$	$0.36 \pm 4e-4$	$0.21 \pm 1e-4$
SVR	$1.35 \pm 7e-3$	4.98 ± 0.13	$0.23 \pm 5e-4$	$0.34 \pm 4e-4$	0.00 ± 0.00
FPN [19]	$1.21 \pm 8e-4$	4.31 ± 0.03	$0.41 \pm 3e-4$	$0.47 \pm 4e-4$	$0.25 \pm 2e-4$
ConvLSTM [20]	3.09 ± 0.02	5.46 ± 0.2	$0.23 \pm 3e-3$	$0.21 \pm 5e-3$	$0.27 \pm 3e-3$
MLP-A [8]	$1.18 \pm 1e-3$	4.00 ± 0.1	$0.40 \pm 4e-4$	$0.46 \pm 5e-4$	$0.26 \pm 2e-4$
FCN-A [11]	1.28 ± 0.01	4.40 ± 0.09	$0.38 \pm 4e-4$	$0.44 \pm 4e-4$	$0.26 \pm 7e-3$
SR-A	$1.27 \pm 5e-4$	4.47 ± 0.13	$0.39 \pm 9e-4$	$0.45 \pm 4e-4$	$0.26 \pm 1e-4$
ODA (ours)	$\mathbf{1.04 \pm 1e-3}$	$\mathbf{3.88 \pm 0.1}$	$\mathbf{0.59 \pm 2e-4}$	$\mathbf{0.52 \pm 3e-4}$	$\mathbf{0.28 \pm 6e-4}$

TABLE III
COMPARISON OF OUR PROPOSED MODEL AGAINST BASELINE METHOD ON 6−HOUR PREDICTION OF SMS-WARMS

Methods	MAE	MPAE	$Ts_{0.1}$	Ts_1	Ts_{10}
BI	2.97 ± 0.15	7.45 ± 0.12	$0.16 \pm 2e-4$	$0.11 \pm 1e-4$	$0.03 \pm 2e-5$
LR	2.64 ± 0.03	6.1 ± 0.05	$0.27 \pm 4e-4$	$0.21 \pm 3e-4$	$0.01 \pm 4e-5$
SVR	1.81 ± 0.03	6.4 ± 0.05	$0.24 \pm 6e-4$	$0.12 \pm 1e-4$	0.0 ± 0.0
FPN [19]	1.93 ± 0.02	6.1 ± 0.04	$0.32 \pm 4e-4$	$0.27 \pm 3e-4$	$0.06 \pm 4e-5$
ConvLSTM [20]	5.82 ± 0.07	9.25 ± 0.3	0.12 ± 0.01	$0.10 \pm 1e-3$	$0.06 \pm 3e-4$
MLP-A [8]	1.9 ± 0.02	6.0 ± 0.04	$0.31 \pm 3e-4$	$0.27 \pm 3e-4$	$0.058 \pm 5e-5$
FCN-A [11]	1.95 ± 0.03	6.1 ± 0.03	$0.286 \pm 3e-4$	$0.28 \pm 4e-4$	$0.06 \pm 5e-5$
SRA	2.0 ± 0.02	6.0 ± 0.03	$0.28 \pm 3e-4$	$0.26 \pm 3e-4$	$0.06 \pm 5e-5$
ODA (ours)	$\mathbf{1.81 \pm 0.02}$	$\mathbf{6.0 \pm 0.04}$	$\mathbf{0.38 \pm 5e-4}$	$\mathbf{0.32 \pm 5-e5}$	$\mathbf{0.065 \pm 5e-5}$

deploy a Single Regression Autoencoder (SR-A) in which a single regression block is used instead of an ordinal regression block for comparison.

We summarize the results of ODA and other algorithms in Table II with the ECMWF model and in Table III with the SMS-WARMS model where the result is repeated more than 20 times and presented by 'mean ± standard deviation'. From the tables, it can be seen that ODA is stable and outperforms others on all metrics in either ECMWF or SMS-WARMS. The result of ConvLSTM shows that focusing on prediction rather than bias correction could limit its correction performance. The results of MLP-A, FC-A, and SR-A indicate that by transforming the regression problem to several binary classification problems, the proposed ODA could achieve better performance in correcting the biased precipitation.

We visualize examples of precipitation bias correction results in Fig. 3. The ODA model correctly revises prediction in the non-precipitation region, as illustrated in the red circles. Meanwhile, ODA successfully corrects small rainfall into large rainfall, as indicated by the black circles. Besides, the blue circles show a more interesting circumstance, where multiple areas are small rainfall in the ground truth, while the output of ODA is mostly 0.5mm.

Note that in the real world, when a region has a small amount of rainfall, the minimal precipitation would be 0.1mm. While in the ODA algorithm, we set the most fine-grained precipitation prediction value to 0.5 mm. When the Precipitation Confidence Network indicates that the region has a high probability of rainfall, as a result, the minimal prediction generated by ODA would be 0.5 mm. Although this limitation can result in a higher MAE or MSE, such a prediction is acceptable because, in a real-world application, the region indeed has a high probability to rain and such a small difference has a negligible influence on the rainfall level.

As a comparison, these examples show that even if some imperfectness of ODA, it still effectively corrects the biased precipitation and rearranges the non-precipitation and precipitation.

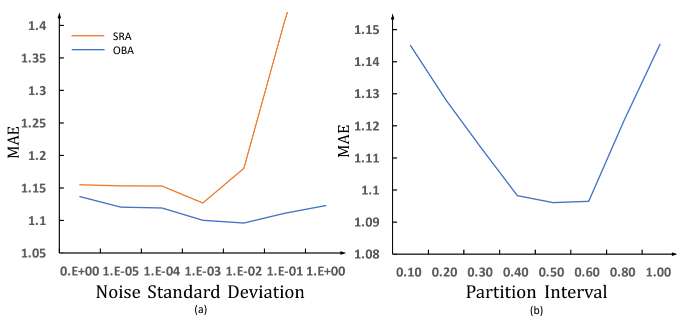

Fig. 4. Performance changes by varying (a) two types of standard deviation and (b) partition interval.

V. Discussion

A. Ablation Study

We analyze the effectiveness of the two novel parts of our approach by ablation experiments and show the results as in Fig. 4(a).

1) The impact of Ordinal Distribution Regression: The orange line in Fig. 4 (a) represents the results of a single regression model and the blue one represents the results of ODA. The comparison of the two lines shows that the MAE of using the Ordinal Distribution Regression model is always less than that of the single regression model. The main reason for this improvement is that converting a regression problem to multiple classification problems alleviates the negative impact of the large span of the training labels.

2) Influence of the Noise Perturbation: For analyzing the influence of the noise perturbation, we vary the standard deviation σ of the normal distribution. As shown in Fig. 4(a), the MAE of our ODA goes down first and then up with the increase of σ. When the σ is set as 0, the Denoising Autoencoder is degraded to a simple autoencoder and the features of its bottleneck layer are squeezed without noise handling. Therefore, the features with noise lead to the worst performance among all of the σ settings. With the increase of σ, the Denoising Autoencoder starts to learn how to deal with the noise and the performance of our model begins to improve. However, owing to features having been normalized, the noise contained in the features is normalized as well. With the σ going larger, as a result, Enhanced Denoising Autoencoder cannot remove noise, even may regard useful information as noise.

B. Comparison of Different Partition Intervals

Unlike regarding the partition interval of ages on facial images as only one region, we split precipitation into several finer ranging intervals in this task. And the partition interval decides not only how many binary classifiers are necessary, but also the number of parameters. Therefore, it is crucial to choose a suitable partition interval for ordinal regression in the bias correction task. It can be seen from Fig. 4(b) that when η is set to 0.5, the best partition interval can be obtained. We also notice that the minimum ranging interval cannot obtain the best partition interval. A possible reason is that some neighbor rank classes contain the same training samples when choosing the minimum ranging interval as the partition interval. This situation causes some samples to be incorrectly classified as positive in these neighbor rank classes, leading to inaccurate ordinal regression results.

VI. Conclusions and Future Work

We proposed a novel Ordinal Distribution Autoencoder to correct the bias of the NWP model in precipitation prediction. The ODA can extract more robust features from the highly noisy predictive data, and correct the biased precipitation against the uneven distribution and a large span of the labels.

Experiments on the ECMWF and SMS-WARMS datasets indicate that compared with several NWP correction baseline methods and three deep learning methods, ODA achieves the best correcting performance as well as robustness. Different from the past works, the proposed ODA change the regression problem into several binary classification problems with ordinal distribution regression. Meanwhile, denoising autoencoder and focal loss are utilized for solving highly noisy and imbalanced data distribution. In summary, ODA is shown to be effective through both quantitative indices and qualitative visualization with little help from experts.

However, when correcting heavy rainfall, the advantages of ODA compared with other models are not significant, which still needs to be improved. The reasons are that: 1) in Shanghai, the amount of heavy rainfall cases in our used dataset is sporadic, resulting in an under-fitting problem to our model, and 2) the design of our ordinal distribution regression limits the performance when correcting heavy rainfall.

In the future, we will explore how to utilize finer partition intervals to obtain more precise bias correction performance, especially for heavy rainfall. It is also interesting to study how to combine the timing information of the predictive weather data with ordinal regression for bias correction of numerical weather prediction.

Acknowledgment

We acknowledge support from Shanghai Municipal Science and Technology Project (18DZ1200404), the Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01), and ZJLab. J.Z. Wang was supported by Penn State University. We also acknowledge the invaluable comments and constructive suggestions by the associate editor and anonymous reviewers.

REFERENCES

[1] A. Aghakouchak, A. Mehran, H. Norouzi, and A. Behrangi, "Systematic and random error components in satellite precipitation data sets," *Geophysical Research Letters*, vol. 39, no. 9, p. 9406, 2012.

[2] Z. Sokol, "MOS-based precipitation forecasts for river basins," *Weather and Forecasting*, vol. 18, no. 5, pp. 769–781, 2003.

[3] S. Shin, J.-S. Kang, and Y. Jo, "The local ensemble transform kalman filter (letkf) with a global nwp model on the cubed sphere," *Pure and Applied Geophysics*, vol. 173, no. 7, pp. 2555–2570, 2016.

[4] D. E. O'Leary, "Artificial intelligence and big data," *IEEE Intelligent Systems*, vol. 28, no. 02, pp. 96–99, 2013.

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

[6] Y. Shao and R. S. Lunetta, "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 70, pp. 78–87, 2012.

[7] D. Cho, C. Yoo, J. Im, and D.-H. Cha, "Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas," *Earth and Space Science*, vol. 7, no. 4, p. e2019EA000740, 2020.

[8] E. Hernández, V. Sanchez-Anguix, V. Julian, J. Palanca, and N. Duque, "Rainfall prediction: A deep learning approach," in *Proceedings of the International Conference on Hybrid Artificial Intelligence Systems*. Springer, 2016, pp. 151–162.

[9] C. K. Sønderby, L. Espeholt, J. Heek, M. Dehghani, A. Oliver, T. Salimans, S. Agrawal, J. Hickey, and N. Kalchbrenner, "Metnet: A neural weather model for precipitation forecasting," *arXiv preprint arXiv preprint arXiv:2003.12140*, 2020.

[10] B. Wang, J. Lu, Z. Yan, H. Luo, T. Li, Y. Zheng, and G. Zhang, "Deep uncertainty quantification: A machine learning approach for weather forecasting," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2087–2095.

[11] Y. Tao, X. Gao, K. Hsu, S. Sorooshian, and A. Ihler, "A deep neural network modeling framework to reduce bias in satellite precipitation products," *Journal of Hydrometeorology*, vol. 17, pp. 931–945, 2016.

[12] L. Zjavka, "Numerical weather prediction revisions using the locally trained differential polynomial network," *Expert Systems with Applications*, vol. 44, pp. 265–274, 2016.

[13] R. Herbrich, T. Graepel, and K. Obermayer, "Support vector learning for ordinal regression," in *Proceedings of International Conference on Artificial Neural Networks*, 1999, pp. 97–102.

[14] A. Shashua and A. Levin, "Ranking with large margin principle: Two approaches," in *Advances in Neural Information Processing Systems*, 2003, pp. 961–968.

[15] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output CNN for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4920–4928.

[16] H. Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, "Deep ordinal regression network for monocular depth estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2002–2011.

[17] X. Tong, L. Jia, Y. Yuhua, W. Xiaofeng, and C. Baode, "Verification of SMS-WARMS V2.0 model forecast results," *Meteorological Monthly*, vol. 42, no. 10, pp. 1176–1183, 2016.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[19] T. Lin, P. Dollr, 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*, July 2017, pp. 936–944.

[20] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *arXiv preprint arXiv:1506.04214*, 2015.

Youcheng Luo is pursuing an M.S. degree from the School of Computer Science, Fudan University. His research interests include machine learning, intelligent systems, and computer vision. Contact him at luoyc18@fudan.edu.cn.

Xiaoyang Xu received the M.S. degree from the School of Computer Science, Fudan University, China, in 2019. His research interests include machine learning, intelligent systems, and computer vision. Contact him at xuxy17@fudan.edu.cn.

Yiqun Liu is pursuing a Ph.D. degree from the School of Computer Science, Fudan University, China. His research interests include machine learning, intelligent systems, and computer vision. Contact him at yqliu17@fudan.edu.cn.

Hanqing Chao received an M.S. degree from the School of Computer Science, Fudan University, China, in 2019. His research interests include machine learning, intelligent systems, and computer vision. Contact him at hqchao16@fudan.edu.cn.

Hai Chu is a forecaster of the Shanghai Central Meteorological Observatory (SCMO). He is also working on the bias correction of numerical weather model outputs. Contact him at chuhai@163.com.

Lei Chen is a forecaster of the Shanghai Central Meteorological Observatory (SCMO). He is also working on the bias correction of numerical weather model outputs. Contact him at qqydss@163.com.

Junping Zhang is a professor in the Shanghai Key Laboratory of Intelligent Information Processing and School of Computer Science at Fudan University. His research interests include machine learning, pattern recognition, image processing, biometric authentication, and intelligent transportation systems. Contact him at jpzhang@fudan.edu.cn.

Leiming Ma is a senior researcher and the deputy director of the Shanghai Central Meteorological Observatory (SCMO). His research interests include cyclone prediction, numerical weather prediction, and convection detection. Contact him at malm@typhoon.org.cn.

James Z. Wang is a Professor of information sciences and technology at The Pennsylvania State University. His research interests include image analysis, image modeling, image retrieval, and their applications. He contributed to discussions and composing the manuscript. Contact him at jwang@ist.psu.edu.