Raindrop Size Distribution Prediction by an Improved Long Short-Term Memory Network

Yongjie Zhu 1,2,3,4, Zhiqun Hu 1,2,3,4,*, Shujie Yuan 1, Jiafeng Zheng 1, Dejin Lu 5 and Fujiang Huang 1

1 School of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu 610225, China
2 State Key Lab of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China
3 Key Laboratory of Atmospheric Sounding, China Meteorological Administration, Chengdu 610225, China
4 Research Centre on Meteorological Observation Engineering Technology, China Meteorological Administration, Beijing 100081, China
5 Weather Modification Office of Anhui, Hefei 230031, China
* Correspondence: huzq@cma.gov.cn

Abstract: The observation of and research on raindrop size distribution (DSD) is important for mastering and understanding the mutual restriction relationship between cloud dynamics and cloud microphysics in a process of precipitation; it also plays an irreplaceable role in many fields, such as radar meteorology, weather modification, boundary layer land surface processes, aerosols, etc. Using more than 1.7 million minutes of raindrop data observed with 17 laser disdrometers at 17 stations in Anhui Province, China, from 7 August 2009 to 30 April 2020, a DSD training dataset was constructed. Furthermore, the data are fitted to a normalized Gamma function and used to obtain its three parameters, i.e., the normalized intercept $N_w$, the mass weighted average diameter $D_m$, and the shape factor $\mu$. Based on the long short-term memory network (LSTM), a DSD Gamma distribution prediction network (DSDnet) was designed. In the process of modeling based on DSDnet, a self-defined loss function (SLF) was proposed in order to improve the DSD prediction by increasing the weight values in the poor fitting regions according to the common mean square error loss function (MLF). By means of the training dataset, a DSDnet-based model was trained to realize the prediction of $N_w$, $D_m$, and $\mu$ minute-to-minute over the course of 30 min, and then was evaluated by the test dataset according to three indicators, namely, mean relative error (MRE), mean absolute error (MAE), and correlation coefficient (CC). The CC of $\lg N_w$, $D_m$, and $\mu$ can reach 0.93403, 0.90934, and 0.89741 for 12-min predictions, and 0.87559, 0.85261, and 0.84564 for 30-min predictions, respectively, which means that the DSD prediction accuracy within 30 min can basically reach the application level. Furthermore, the 12- and 30-min predictions of 3 precipitation processes were taken as examples to fully demonstrate the application effect of model. The prediction effects of $N_w$ and $D_m$ are better than that of $\mu$, and the stratiform precipitation is better than the convective and convective-stratiform mixed cloud precipitation.

Keywords: raindrop size distribution; normalized Gamma distribution; long short-term memory network; self-defined loss function; DSD prediction network

1. Introduction

The raindrop size distribution (DSD) reflects the variation in the particle number concentrations of raindrops with sizes in different unit volumes. From DSD, various parameters of microphysical characteristics, such as raindrop number concentration, average diameter, precipitation intensity, water content, etc., can be calculated in a precipitation process. In addition, the rainfall development can be more clearly understood through the variation of DSD, which is significant when optimizing the parameterization scheme in the weather and climate models [1,2], evaluating the effects of weather modification, and improving the radar quantitative precipitation estimation, etc.
Earlier in the filter paper measurement, Marshall and Palmer addressed that the DSD can be represented by a negative exponential function, namely, M-P distribution [3]. However, many studies have shown that the M-P distribution is applicable to stable stratiform cloud precipitation [4–6], and there is a relatively large deviation in small and large droplets for convective clouds. Comparatively, a Gamma function proposed by Ulbrich is universal [7], which can not only reflect the DSD of stratiform clouds, but is also an ideal expression of DSD for convective–stratiform mixed and convective clouds. Meanwhile, the Gamma distribution can greatly improve the fitting accuracy in the small and large droplet regions. Nevertheless, the parameter \( N_0 \) is affected by \( \mu \) in the Gamma function. In order to give an explicit physical meaning for describing DSD, a normalized Gamma function was proposed by Willis [8].

DSD is an important basis for studying the microphysical characteristics of clouds and precipitation. By analyzing the DSD in a precipitation process, the development and evolution of precipitation can be better understood [9–11]. Therefore, the prediction of DSD can provide a certain reference for the forecasting of precipitation. At present, a regional historical regression is usually used to evaluate the effect of artificial precipitation enhancement [12], but the results lack a quantitative basis. By comparing the predicted and measured DSD, the precipitation enhancement effect after weather modification operations can be quantitatively evaluated. In addition, the Z-R relationship fitted with DSD is a common method for radar quantitative precipitation estimation. However, because of the continuous changes of DSD during a rainfall process, the Z-R relationship is unstable, which usually results in a light rain being overestimated and a heavy rain being underestimated [13–15]—so it’s appropriate coefficients need to be fitted by means of a significant amount of DSD data. Therefore, accurate DSD measurements and predictions can better improve the radar quantitative precipitation estimation and forecast (QPE and QPF). To date, most studies have only focused on the DSD observation and analysis after a precipitation process has occurred but lacked the DSD predictions during the process. This paper attempts to perform this work by utilizing the current popular deep learning algorithm.

In recent years, with technological development and maturity, deep learning has become the preferred algorithm for data mining. As one of the important deep learning algorithms, long short-term memory (LSTM) can mine evolutionary information, which shows strong adaptability and decision-making ability on time-series data processing and plays an important role in dealing with non-linear issues and discovering the relationship between the elements [16]. Due to these characteristics of LSTM, it is widely used in meteorological prediction [17–19]. A refined model for 24-h temperature prediction was established by LSTM, which could attain 68.75% accuracy [20]. Aiming at the low accuracy of wind speed prediction, which has been caused by many factors, a multidimensional LSTM model was proposed for improving short-term wind speed prediction, in which the main influencing factors collected by a fuzzy rough dataset were used as the inputs of LSTM [21]. However, there has been little research on DSD prediction by means of deep learning. Using the data acquired from the laser raindrop disdrometers, a DSD dataset was constructed, and a DSD prediction network (DSDnet) based on LSTM was designed to predict the normalized Gamma distribution parameters, \( N_w, D_m \), and \( \mu \), minute-by-minute during the next 30 min. The study in this paper hopes to provide a reference for cloud microphysical, dynamic, and precipitation forecasting.

The remainder of the paper is organized as follows. Section 2 describes the data source and preprocesses, including the main performance parameters, locations of the instruments, and the steps for data quality control. Section 3 introduces the building process of the DSDnet-based model in detail, which is the essence of the full text, including a simple introduction of the LSTM algorithm, the design of the self-defined loss function (SLF), construction of the training set, and adjustment of the hyperparameters, the model training process, and the model performance in the test dataset. In Section 4, using three typical cases, i.e., stratiform, mixed convective-stratiform, and convective cloud precipitation, the parameters fitted by the observed data and predicted by the model are contrasted to
demonstrate the practical application effect of the model. The shortcomings of DSDnet, improvement suggestions, and prospects are discussed in the Section 5.

2. Data Source and Preprocessing

2.1. Data Source

The DSD data were observed with 17 raindrop disdrometers in Anhui Province, including 7 OTT Parsivel\(^2\) and 10 HSC-OTT Parsivel EF. Their main performance parameters are listed in Table 1. The disdrometer samples record per minute and give the number of particles within 1024 channels, i.e., the 32-level diameter size in a 32-level falling velocity. The diameter range was from 0.2 to 25 mm, and the terminal velocity was from 0.2 to 20 m s\(^{-1}\). A total of 1,788,915-min samples were collected during 6725 rainfall events from 7 August 2009 to 30 April 2020. The corresponding CINRAD/SA weather radar is in Hefei, with a beam width of 1°, radial resolution of reflectivity of 1 km, and maximum detection distance of 460 km. The radar adopts the VCP 21 (volume coverage pattern, scan strategy #2, version 1) scan mode, which comprises 9 elevation angles (0.5°, 1.5°, 2.4°, 3.4°, 4.3°, 6.0°, 9.9°, 14.6°, and 19.5°), and the time for a volume scanning is six minutes. The locations of the radar and disdrometers are shown in Figure 1, in which all disdrometers are within the radar detection range of 230 km.

Table 1. The performance parameters of the raindrop disdrometer.

| Items | Parameters |
|-------|------------|
| Average falling speed of channel 1–32 (m/s) | Range 0.2–20 |
| Classification | 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.10, 1.30, 1.50, 1.70, 1.90, 2.20, 2.60, 3.00, 3.40, 3.80, 4.40, 5.20, 6.00, 6.80, 7.60, 8.80, 10.40, 12.00, 13.60, 15.20, 17.60, 20.80 |
| Average particle diameter of 1–32 channels (mm) | Range 0.2–25 |
| Classification | 0.062, 0.187, 0.312, 0.437, 0.562, 0.687, 0.812, 0.937, 1.062, 1.187, 1.375, 1.625, 1.875, 2.125, 2.375, 2.750, 3.250, 3.750, 4.250, 4.750, 5.500, 6.500, 7.500, 8.500, 9.500, 11.000, 13.000, 15.000, 17.000, 19.000, 21.500, 24.500 |
| Accuracy | Liquid ±5% |
| Solid ±20% |
| Particle level | 32 (size) × 32 (velocity) |
| Differentiation of precipitation types | >97% |
| Measurement interval | 60 s |
| Measuring area | 54 cm\(^2\) (18 cm × 3 cm) |
| Wavelength | 780 nm (OTT Parsivel\(^2\)) |
| 650 nm (HSC-OTT Parsivel EF) |
| Output rating | 0.5 mW (OTT Parsivel\(^2\)) |
| 3 mW (HSC-OTT Parsivel EF) |
construction of the training set, and adjustment of the hyperparameters, the model training process, and the model performance in the test dataset. In Section 4, using three typical cases, i.e., stratiform, mixed convective-stratiform, and convective cloud precipitation, the parameters fitted by the observed data and predicted by the model are contrasted to demonstrate the practical application effect of the model. The shortcomings of DSDnet, improvement suggestions, and prospects are discussed in the last section.

2. Data Source and Preprocessing

2.1. Data Source

The DSD data were observed with 17 raindrop disdrometers in Anhui Province, including 7 OTT Parsivel 2 and 10 HSC-OTT Parsivel EF. Their main performance parameters are listed in Table 1. The disdrometer samples record per minute and give the number of particles within 1024 channels, i.e., the 32-level diameter size in a 32-level falling velocity. The diameter range was from 0.2 to 25 mm, and the terminal velocity was from 0.2 to 20 m s\(^{-1}\). A total of 1,788,915-min samples were collected during 6725 rainfall events from 7 August 2009 to 30 April 2020. The corresponding CINRAD/SA weather radar is in Hefei, with a beam width of 1°, radial resolution of reflectivity of 1 km, and maximum detection distance of 460 km. The radar adopts the VCP 21 (volume coverage pattern, scan strategy #2, version 1) scan mode, which comprises 9 elevation angles (0.5°, 1.5°, 2.4°, 3.4°, 4.3°, 6.0°, 9.9°, 14.6°, and 19.5°), and the time for a volume scanning is six minutes. The locations of the radar and disdrometers are shown in Figure 1, in which all disdrometers are within the radar detection range of 230 km.

Figure 1. The station distribution map, in which the dots are the disdrometers, the triangle in the center is the radar, and the circle is the 230 km detection range of the radar.

2.2. Data Preprocessing

To ensure the effectiveness of deep learning training, data cleaning was crucial. Therefore, before being added to the dataset, these unqualified DSD data were given up, including (1) in the first two diameter intervals, i.e., diameters less than 0.187 mm; (2) when the number of particles in a diameter interval is less than 2; (3) when the number of particles in a minute is less than 10; (4) when diameters are greater than 8 mm; and (5) when the difference in particle falling velocity between the measured and classical is greater than 5 m s\(^{-1}\) [22,23].

2.3. Normalized Gamma Distribution

The Gamma DSD expression proposed by Ulbrich is:

\[
N(D) = N_0 D^\mu \times \exp(-\lambda D)
\]

where \(N\) represents the corresponding number concentration of raindrops at the diameter \(D\), \(N_0\) is the intercept of the Gamma function, \(\mu\) is the shape parameter, and \(\lambda\) is the slope.

The unit of Gamma parameter \(N_0\) is related to another parameter \(\mu\), so it has less physical significance in Equation (1), and \(N_0\) can be discussed only under the same \(\mu\). To avoid this problem, Willis proposed a DSD form for the normalized Gamma function (Equation (2)):

\[
N(D) = N_w f(\mu) \left(\frac{D}{D_m}\right)^\mu \exp\left[-(\mu + 4)\frac{D}{D_m}\right]
\]
The three parameters of the normalized Gamma function, namely, normalized intercept $N_w$, mass weighted average $D_m$, and shape factor $\mu$, which have clear physical significance, can be calculated by the order-moment method, in which $N_w$, $D_m$, and $\mu$ can be calculated from the third-, fourth-, and sixth-order-moments of DSD [24], respectively (Equations (5)–(8)). The expression of the ith-order-moment is Equation (4):

$$M_i = \int_0^\infty D^i N(D) dD$$  \hspace{1cm} (4)$$

$$N_w = \frac{4^4}{\pi \rho_w} \cdot \frac{W \left( \frac{W}{D_m^4} \right)}{D_m^4} = \frac{256}{6} \times \frac{M_3^3}{M_4^4}$$  \hspace{1cm} (5)$$

$$D_m = \frac{M_4}{M_3}$$  \hspace{1cm} (6)$$

$$G = \frac{M_4^3}{M_3^3 M_6}$$  \hspace{1cm} (7)$$

$$\mu = \frac{11G - 8 + \sqrt{G(8 + G)}}{2(1 - G)}$$  \hspace{1cm} (8)$$

3. DSD Prediction Model

3.1. Introduction of the LSTM Algorithm

Traditional neural networks (NN) include an input layer, hidden layers, and output layer. These layers are fully connected, but the nodes between each layer are independent. Therefore, it is difficult to deal with time sequence issues by using NN. As shown in Figure 2, a recurrent neural network (RNN) is composed of repeated NN modules in a chain form, in which A represents a unit, $X_t$ is the input at time $t$, and $Y_t$ is the corresponding output through the unit and is used as another input factor for the next unit at time $t+1$ for information transmission. Therefore, the nodes between the hidden layers are connected and the information is transmitted.

![Figure 2. The schematic diagram of RNN network.](image-url)

However, RNN has the problem of gradient disappearance and gradient explosion in long sequence training. As a special RNN, LSTM not only inherits most of the characteristics of RNN, but also effectively avoids these defects. Compared with the RNN, LSTM has a more complex memory cell, which has three gates, namely, the forgetting gate, input gate, and output gate (Figure 3) [25,26].
3.2. DSDnet Design

On the basis of LSTM, a DSD prediction network was designed and named DSDnet, which includes two LSTM layers, a linear layer, and an output layer (Figure 4).

3.3. Training Dataset Construction

After preprocessing, the measured raindrop data were re-sorted into chronological order according to the stations. While \( N_w \) was not equal to 0 in more than 12 consecutive minutes, and the duration of precipitation was larger than 60 min, the sequence was marked as one record of the dataset.

Finally, the dataset was built by a total of 6725 sequences from 1,788,915-min samples. Additionally, these data were normalized with the min–max standardization method to map into the \([0, 1]\) interval, that is:

\[
y = \frac{y - \min}{\max - \min}
\]  

(9)

where the maximum \( N_w, D_m, \) and \( \mu \) were set to \( 10^{5.5} \text{ m}^{-3} \text{ mm}^{-1}, 4 \text{ mm}, 30, \) and the minimum of \( N_w \) and \( D_m \) were set to 0, while \( \mu \) was set to \(-5\), respectively. If the values of
the fitting parameters of the normalized Gamma were larger than the maximum or less than the minimum, their values were set to the corresponding maximum or minimum.

We randomly selected 20% from the dataset, i.e., 1345 sequences, as the test set, and the remaining 80%, i.e., 5380 sequences, were used as the training set. The samples in the training set were rearranged into the input and label form required by the DSDnet, namely, how many minutes (T_{step}) of data in the past were used to fit the parameter values at the prediction minute (M_{pred}). Therefore, the input of DSDnet is a matrix consisting of T_{step} row minutes and three column parameters, N_w, D_m and µ, and the corresponding labels are N_w, D_m, and µ at the T_{step} + M_{pred} minute. Without losing representativeness, T_{step} was set to 18 min and M_{pred} was set to 12 and 30 min, respectively. Namely, the input of the network was an 18 × 3 matrix, and the labels were the actual values of N_w, D_m, and µ at 12 or 30 min after 18 min, respectively. The inputs started from the 1st to the 18th row of the samples, and the matrix slid down one row to form the second input, and so on, until the labels were at the last row to complete the training dataset's construction. During the modeling process, the samples were shuffled every iteration, and 15% of the training set was randomly extracted as the validation set. Finally, the model was evaluated by the test set. Three precipitation systems, i.e., stratiform, mixed convective–stratiform, and convective clouds, were selected to demonstrate the prediction effects of the DSDnet-based model.

3.4. Self-Defined Loss Function

In the process of model training, the fitting value was compared with the real value by using a loss function to calculate the loss value E after each iteration. Then, through the back propagation of the loss value, the weight coefficients w_{j,k} at the kth node of the jth hidden layer were adjusted by an optimizer (Equation (10)):

$$
\left( w_{j,k} \right)_{\text{new}} = \left( w_{j,k} \right)_{\text{old}} - \alpha \frac{\partial E}{\partial w_{j,k}}
$$

where the subscript “new” and “old” represent the weight after and before adjustment, respectively, and α is the learning rate. In essence, the process of training a deep learning model is to minimize the loss function whose value reflects how far the fitting result is from perfection on a given dataset. For regression problems, mean squared error (MSE) is often used as a loss function (MLF), and its expression is:

$$
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left( y_{\text{pred}} - y_{\text{true}} \right)^2
$$

where y_{pred} and y_{true} are the normalized predicted and measured values, respectively.

Since the three parameters of the normalized Gamma function are usually a normal distribution, the frequencies are lower at large value and small value regions. If MLF is used as the loss function, the fitting results will gradually tend to the middle value regions, which have higher frequencies. Therefore, a SLF (Equation (12)) is suggested on the basis of MLF. In the model training process, according to the distribution of the three parameters in the large and small value regions, different weight coefficients are set, and higher weights are given to the large and small value regions (Table 2, in which W is the weight value vectors, and L is the normalized value vectors of the parameters).
3.5. Hyperparameter Setting

The hyperparameters have a great impact on model training efficiency and prediction accuracy, which include the number of hidden layers, number of stacking layers, batch size, epoch number, learning rate, and optimizer, etc. Since the number of DSD samples reaches millions, automatic parameter adjustment methods, such as GridSearchCV, which are suitable for small datasets are difficult to apply here. Therefore, the hyperparameters are continuously adjusted during modeling. It was found that the model output result and convergence speed were influenced by the number of hidden layer nodes, stacking layers, and batch size. When the number of stacking layers is greater than 4, or the batch size is less than 1000, the convergence speed of the model decreases significantly. After taking into account the principles of fast convergence and small error, the number of hidden layer nodes, stacking layers, and batch size are set to 128, 2, and 4000, respectively. Other hyperparameters are set as follows: Adam gradient descent optimizer with a learning rate of 0.001 and 500 epochs. Meanwhile, to improve the efficiency, an early stop mechanism was adopted. If the loss value of the validation set does not decrease after five consecutive iterations, the training process was stopped, and the model was saved.

3.6. Evaluation Indicator

The mean relative error (MRE), mean absolute error (MAE), and correlation coefficient (CC) were adopted as the evaluation indicators for the DSDnet-based model (Equations (13)–(15)).

\[
\text{MRE}(y_{\text{true}}, y_{\text{pred}}) = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_{\text{true}} - y_{\text{pred}}|}{|y_{\text{true}}|} \tag{13}
\]

\[
\text{MAE}(y_{\text{true}}, y_{\text{pred}}) = \frac{1}{n} \sum_{i=1}^{n} |y_{\text{true}} - y_{\text{pred}}| \tag{14}
\]

\[
\text{CC}(y_{\text{true}}, y_{\text{pred}}) = \frac{\sum_{i=1}^{n} (y_{\text{true}} - \bar{y}_{\text{true}})(y_{\text{pred}} - \bar{y}_{\text{pred}})}{\sqrt{\sum_{i=1}^{n} (y_{\text{true}} - \bar{y}_{\text{true}})^2} \sqrt{\sum_{i=1}^{n} (y_{\text{pred}} - \bar{y}_{\text{pred}})^2}} \tag{15}
\]

3.7. Modeling Flow Chart

The modeling flow chart is shown in Figure 5.

![Figure 5. The modeling flow chart.](image-url)
3.8. Model Evaluation by Test Set

After the model was finished, the model was evaluated by the test dataset. Figure 6 shows the scatter plot of $\text{lgN}_w$, $D_m$, and $\mu$, which were fitted by the observed data and predicted by the model (Figure 6a–c modeled with MLF, and Figure 6d–f with SLF). From the indicators, $N_w$ is better than $D_m$ and $\mu$. The model performs slightly worse in the large and small value regions, which is caused by the relatively less data and more difficult rule mining in these parts. In addition, after using SLF, the model accuracy was greatly improved. By comparing the model using SLF and MLF, it can be seen from Table 3 that the MRE of $\text{lgN}_w$, $D_m$, and $\mu$ decreased by 3.98%, 3.05%, and 6.48%; MAE decreased by 2.56%, 3.06%, and 5.78%; and CC increased by 0.25%, 1.44%, and 1.94%, respectively.

### Table 3. The 12-min prediction evaluation results of the test set.

| Model   | Evaluation Index | $\text{lgN}_w$ | $D_m$ | $\mu$ |
|---------|------------------|----------------|-------|-------|
| With MLF| MRE              | 0.05452        | 0.11925| 0.16559|
|         | MAE              | 0.23862        | 0.13497| 1.43105|
|         | CC               | 0.93162        | 0.89621| 0.87996|
|         | MRE              | 0.05235        | 0.11561| 0.15486|
| With SLF| MAE              | 0.23251        | 0.13084| 1.34834|
|         | CC               | 0.93403        | 0.90934| 0.89741|

Figure 7 is the same with Figure 6 but 30-min prediction scatter plot. Compared with the 12-min set, when the correlation decreases, the error of each parameter increases, and the model using SLF is more obviously improved. It can be seen from Table 4 that the MRE of $\text{lgN}_w$, $D_m$, and $\mu$ decreased by 12.87%, 7.18%, and 11.3%; MAE decreased by 13.7%, 7.29%, and 14.75%; and CC increased by 2.32%, 1.51%, and 2.82%, respectively.
4. Model Application

To further demonstrate the application effect of the model, the 12- and 30-min predictions were used as examples by means of three rainfall cases. Representative 0.5° PPI of the radar are shown in Figures 8, 12, and 15. To facilitate the comparison, the predicted (red line) and fitted values (blue line) are shown in Figures 9, 10, 13, and 16, in which the vertical dashed lines are at \( T_{\text{step}} + M_{\text{pred}} \) minute, that is, the start time of prediction, and \( N_w \) is recalculated with the logarithm of base 10. If \( N_w \) is less than 1, \( \lg N_w \) is set as 0. Overall, these figures illustrate that the model has a satisfactory prediction of the parameters.

4.1. Stratiform Cloud Precipitation

This case is a long-term winter stratiform cloud precipitation process with a total of 1184 min observed at the Lujiang station (117.17° E, 31.16° N) from 0333 to 2316 LST (local standard time, the same below) on 26 January 2020. Three representative 0.5° PPI (plan position indicators) of the radar at 0726, 1226, and 1332 LST are shown in Figure 8, corresponding to the 233rd, 533rd, and 599th minutes in Figures 9 and 10, respectively. The red triangle is the location of the disdrometer. The curves of the 12- and 30-min predictions and actual values of \( \lg N_w, D_m, \) and \( \mu \) are shown in Figures 9 and 10, respectively, in which the left ones (a, b, c) are modeled with MLF and the right ones (d, e, f) with SLF. The

Table 4. The 30-min prediction evaluation results of test set.

| Model      | Evaluation Index | \( \lg N_w \) | \( D_m \) | \( \mu \) |
|------------|-----------------|--------------|-----------|---------|
| With MLF   | MRE             | 0.06867      | 0.17442   | 0.27311 |
|            | MAE             | 0.29384      | 0.15411   | 2.36024 |
|            | CC              | 0.85564      | 0.83968   | 0.82761 |
|            | MRE             | 0.05983      | 0.16188   | 0.24224 |
| With SLF   | MAE             | 0.25354      | 0.14287   | 2.01193 |
|            | CC              | 0.87599      | 0.85261   | 0.84564 |

Figure 7. The 30-min scatter plot of \( \lg N_w, D_m, \) and \( \mu \) fitted by the observed data and predicted by the model, in which (a–c) are modeled with MLF and (d–f) with SLF. The red line denotes a linear trend line of the scattered points, and the color mark represents the Gaussian kernel density estimation.
curves show that, except for the slightly larger individual error, the model has satisfactory prediction results based on the two loss functions, and the fitting accuracies of $N_w$ and $D_m$ are better than that of $\mu$.

**Figure 8.** The 0.5° PPI of reflectivity of the Hefei radar at 0726 (a), 0127 (b), and 1332 LST (c) on 26 January 2020, in which the red triangle is the location of the disdrometer at Lujiang, and the scale on the coordinate axes represents the distance (km).

**Figure 9.** The curve of the 12-min prediction for the stratiform cloud precipitation case, in which the horizontal axes are rainfall duration time (min); vertical axes are (a, d) $\log_{10}N_w$, (b, e) $D_m$ (mm), and (c, f) $\mu$; the red and blue curves represent the predicted and actual values; and the vertical dotted line is at the minute of $T_{\text{step}} + M_{\text{pred}}$. Modeling with MLF is shown on the left (a–c), and with SLF on the right (d–f).
horizontal axes are rainfall duration time (min); vertical axes are (a,d) $\text{lg}N_w$, (b,e) $D_m$ (mm), and (c,f) $\mu$; the red and blue curves represent the predicted and actual values; and the vertical dotted line is at the minute of $T_{\text{step}} + M_{\text{pred}}$. Modeling with MLF is shown on the left (a–c), and with SLF on the right (d–f).

**Figure 10.** The curve of the 30-min prediction for the stratiform cloud precipitation case, in which the horizontal axes are rainfall duration time (min); vertical axes are (a,d) $\text{lg}N_w$, (b,e) $D_m$ (mm), and (c,f) $\mu$; the red and blue curves represent the predicted and actual values; and the vertical dotted line is at the minute of $T_{\text{step}} + M_{\text{pred}}$. Modeling with MLF is shown on the left (a–c), and with SLF on the right (d–f).

4.1.1. 12-min Prediction Results

Figure 9 shows the variation curves of the actual and predicted values of $\text{lg}N_w$, $D_m$, and $\mu$ from the 12-min prediction. The values of $D_m$ are between 0.8 to 1.2 mm, $\text{lg}N_w$ are between 4 to 4.8, and $\mu$ are between 4 to 12. By observing the change in the curves, when $D_m$ is small, $\text{lg}N_w$ is large; on the contrary, when $D_m$ is large, $\text{lg}N_w$ is small. The maximum value of $\mu$ is about 24, and most of the values of $\mu$ are larger than those of the convective–stratiform mixed clouds and convective precipitation (the cases are demonstrated below).

(d)

(e)

(a)

(f)

(b)

(c)

0

0.8

1.6

2.4

3.2

4

4.8

5.6

$\text{lg}N_w$ (m$^{-3}$ mm$^{-1}$)

0

0.2

0.4

0.6

0.8

1

1.2

1.4

$D_m$ (mm)

0

0.2

0.4

0.6

0.8

1

1.2

1.4

$\mu$

0

200

400

600

800

1000

1200

Time (min)

0

4

8

12

16

20

24

28

Time (min)
between 4 to 4.8, and μ are between 4 to 12. By observing the change in the curves, when $D_m$ is small, $\lg N_w$ is large; on the contrary, when $D_m$ is large, $\lg N_w$ is small. The maximum value of μ is about 24, and most of the values of μ are larger than those of the convective–stratiform mixed clouds and convective precipitation (the cases are demonstrated below). In addition, the parameter μ has large fluctuations and poor prediction for the prediction of $N_w$ and $D_m$. Due to the inherent smoothness of LSTM, it is difficult to fit such data, which have strong short-term fluctuations.

The evaluation results are listed in Table 5. The overall prediction effect was improved after the SLF is adopted, in which the MRE of $\lg N_w$, $D_m$, and $\mu$ decreased by 2.24%, 2.39%, and 13.71%; MAE decreased by 2.21%, 23.9%, and 13.69%; and CC increased by 2.31%, 6.26%, and 1.42%, respectively.

### Table 5. The 12-min prediction evaluation results of stratiform clouds.

| Model       | Evaluation Index | $\lg N_w$ | $D_m$  | $\mu$  |
|-------------|-----------------|-----------|--------|--------|
| With MLF    | MRE             | 0.02231   | 0.03545| 0.17484|
|             | MAE             | 0.09959   | 0.03395| 1.41840|
|             | CC              | 0.88919   | 0.86223| 0.84995|
| With SLF    | MRE             | 0.02181   | 0.02697| 0.15086|
|             | MAE             | 0.09739   | 0.02583| 1.22393|
|             | CC              | 0.90978   | 0.91626| 0.86207|

#### 4.1.2. 30-min Prediction Results

The actual and predicted values of $\lg N_w$, $D_m$, and $\mu$ of the 30-min prediction are shown in Figure 10. It shows that the performance of the 30-min prediction is relatively poor compared to that of the 12-min prediction, especially in the regions of large values of $\lg N_w$ and $D_m$, where the predicted values are generally lower than the measured ones. From the comparison of Tables 5 and 6, the prediction error for 30-min was larger than that of 12-min, and the CC prediction for 30-min was lower than that of 12-min. This is because the time-series correlation decreases when the prediction duration increases. Even so, the accuracy of the 30-min prediction was still satisfactory. After the model adopted the SLF, the prediction effect was greatly improved. The MRE of $\lg N_w$, $D_m$, and $\mu$ decreased by 17.68%, 15.05%, and 21.08%; MAE decreased by 16.69%, 19.72%, and 20.03%; and CC increased by 3.45%, 0.69%, and 5.81%, respectively (Table 6).

### Table 6. The 30-min prediction evaluation results for stratiform clouds.

| Model       | Evaluation Index | $\lg N_w$ | $D_m$  | $\mu$  |
|-------------|-----------------|-----------|--------|--------|
| With MLF    | MRE             | 0.04846   | 0.04277| 0.22196|
|             | MAE             | 0.21637   | 0.04096| 1.80071|
|             | CC              | 0.84809   | 0.82457| 0.79582|
| With SLF    | MRE             | 0.03989   | 0.03633| 0.17516|
|             | MAE             | 0.18024   | 0.03288| 1.43994|
|             | CC              | 0.87736   | 0.83032| 0.84207|

As an example, the DSDs of measurement, Gamma fitting, and model prediction at the 1063rd min are shown in Figure 11. Figure 11a,b are the 12- and 30-min prediction, respectively. Overall, the predicted DSD curve was very close to the curve fitted with the measured values, which indicated that the prediction effect of the model was satisfactory, and that the 12-min prediction was better than 30-min one.
Figure 11. The DSDs of measurement (black solid line), Gamma fitting (blue dotted line), and model prediction (red dotted line, (a) 12-min and (b) 30-min prediction) at the 1063rd min of this stratiform cloud precipitation case.

4.2. Mixed Convective-Stratiform Clouds

This case was a mixed convective–stratiform cloud precipitation process observed at ChuZhou station (118.17° E, 32.21° N) from 0733 to 2337 LST on 24 November 2015, for a total of 965 min in early winter. Figure 12 is the 0.5° PPI of the radar reflectivity corresponding to the three typical times, 0828, 0914, and 0949 LST in the process, which correspond to the 55th, 101st, and 136th min in Figure 13. The red triangle is the position of the disdrometer at ChuZhou station. To save the length of the article and not lose representativeness, this and the next case only demonstrate the 30-min prediction results. In Figure 13, although the prediction error of cumulus mixed cloud precipitation increases compared with stratiform precipitation, the performance of the DSDnet-based model remains satisfactory, except for individual points.

Figure 12. The 0.5° PPI of reflectivity of the Hefei radar at 0818 (a), 0914 (b), and 0949 LST (c) on 24 November 2015, in which the red triangle is the location of the disdrometer at ChuZhou.
Results of 30-min Prediction

Figure 13 shows the change curves of the actual and predicted values of lgN_w, D_m, and μ for the 30-min prediction. By contrasting Figures 10 and 13, the parameter fluctuations of the mixed convective–stratiform clouds were more extreme and wider than those of stratiform clouds, and the values of lgN_w and μ were smaller, but D_m was larger. lgN_w is distributed between 2.4 to 4.8, and μ appeared at values less than 0. The maximum of D_m was around 2.1 mm. Near the 580th min, lgN_w and μ decreased rapidly, but D_m increased, which may be related to the change of precipitation type, and the latter half of the case
was dominated by convective precipitation. In general, the fitting accuracy of $\lg N_w$ and $D_m$ was better than $\mu$. After modeling with the SLF, the prediction effect was improved. The MRE of $\lg N_w$, $D_m$, and $\mu$ decreased by 11.96%, 23.49%, and 19.36%; MAE decreased by 18.43%, 14.84%, and 7.71%; and CC increased by 2.07%, 2.50%, and 4.25%, respectively (Table 7).

Table 7. The 30-min prediction evaluation results of mixed convective-stratiform clouds.

| Model  | Evaluation Index | $\lg N_w$ | $D_m$  | $\mu$  |
|--------|------------------|-----------|--------|--------|
| With MLF | MRE             | 0.14235   | 0.08697| 0.41552|
|         | MAE             | 0.11261   | 0.33545| 2.63504|
|         | CC              | 0.90904   | 0.89730| 0.81276|
| With SLF | MRE             | 0.12532   | 0.06654| 0.33504|
|         | MAE             | 0.09185   | 0.28565| 2.43201|
|         | CC              | 0.92786   | 0.91974| 0.84736|

Similarly, as an example, the DSDs of measurement, Gamma fitting, and model prediction at the 594th min are shown in Figure 14, respectively. The Gamma function of the prediction basically coincides with the fitting one, and the 12-min prediction is better than the 30-min one.

![Figure 14](image-url)

Figure 14. The DSDs of measurement (black solid line), Gamma fitting (blue dotted line), and model prediction (red dotted line, (a) 12-min and (b) 30-min prediction) at the 594th min of this mixed convective-stratiform cloud precipitation case.

### 4.3. The Case of Convective Clouds

This case is a convective cloud precipitation process observed by the Dingyuan station (117.40° E, 32.32° N) from 0355 to 0845 LST on 29 June 2015, for a total of 291 min in summer. Figure 15 shows the PPI corresponding to the three typical times, 0436, 0522, and 0528 LST, which are the 41st, 87th, and 93rd min in Figure 16. The red triangle is the position of the disdrometer. The actual and 30-min prediction values are shown in Figure 16.
Figure 15. The 0.5° PPI of reflectivity of the Hefei radar at 0436 (a), 0522 (b), and 0528 LST (c) on 29 June 2015, in which the red triangle is the location of the disdrometer at Dingyuan.

Figure 16. The curve of the 30-min prediction for the convective cloud precipitation case, in which the horizontal axes are rainfall duration time (min); vertical axes are (a,d) $\lg N_w$, (b,e) $D_m$ (mm), and (c,f) $\mu$;
the red and blue curves represent the predicted and actual values; and the vertical dotted line is at
the minute of \( T_{\text{step}} + M_{\text{pred}} \). Modeling with MLF is shown on the left (a–c), and with SLF on the right
(d–f). Results of 30-min Prediction.

Figure 16 shows the evolution of the observation and 30-min prediction values in this
convective rainfall event. Compared with the two types of precipitation mentioned above,
the \( D_m \) values of cumulus were larger, and the maximum of \( D_m \) exceeded 2.4 mm. \( \lg N_w \)
were distributed between 3.5 to 4.2, and \( \mu \) appeared as negative values too. When \( D_m \) was
large, \( \mu \) was smaller, and, on the contrary, when \( D_m \) was small, \( \mu \) was larger. Obviously,
the prediction effect of the three parameters was lower than the above two cases, which
may be related to the lower number of samples of cumulus precipitation. According to
the evaluation results in Table 8, the prediction with SLF was greatly improved. The MRE
of \( \lg N_w \), \( D_m \), and \( \mu \) decreased by 18.71%, 17.92%, and 20.52%; MAE decreased by 12.07%,
11.35, and 13.51%; and CC increased by 6.99%, 1.94%, and 7.31%, respectively.

Table 8. The 30-min prediction evaluation results of convective clouds.

| Model   | Evaluation Index | \( \lg N_w \) | \( D_m \) | \( \mu \) |
|---------|------------------|--------------|-----------|--------|
| With MLF| MRE              | 0.08276      | 0.09283   | 0.47653|
|         | MAE              | 0.32051      | 0.15222   | 2.85232|
|         | CC               | 0.74374      | 0.86702   | 0.71285|
|         | MRE              | 0.06727      | 0.07619   | 0.37871|
| With SLF| MAE              | 0.28180      | 0.13494   | 2.46677|
|         | CC               | 0.79573      | 0.88387   | 0.76503|

Similarly, as an example, the DSDs of measurement, Gamma fitting, and model
prediction at the 86th min are shown in Figure 17. Although the error increased, the
Gamma function of the prediction was close to the fitting one, and the 12-min prediction
was better than the 30-min one.

![Figure 17](image)

Figure 17. The DSDs of measurement (black solid line), Gamma fitting (blue dotted line), and model
prediction (red dotted line, (a) 12-min and (b) 30-min prediction) at the 86th min of this convective
cloud precipitation case.

5. Conclusions and Discussion

In recent years, artificial intelligence technology has been rapidly applied in all walks
of life. It is one of the promising research directions of meteorological and hydrological
prediction, in which deep learning algorithms are used to analyze a large amount of data
that has been observed with various meteorologic detection instruments over many years
and to extract the rules and information from this big data.
On the basis of LSTM, a DSD network (DSDnet) was designed to predict the distribution of raindrops during a precipitation process. However, due to their intrinsic structure, deep learning algorithms are more inclined to fit the data with high frequency during the modeling process. Furthermore, with the depth of the network increasing, the information in the previous layers gradually attenuates and is difficult to transmit to the final output layer.

By aiming at the inherent problem of the LSTM method, a self-defined loss function was proposed to improve the smoothness by increasing the weight of the small and large diameter particles.

By means of a large amount of data observed with the laser raindrop disdrometers, the parameters of the normalized Gamma functions, i.e., \( N_w \), \( D_m \), and \( \mu \), were initially fitted, and the, as the input and output factors, a DSDnet-based model was trained to realize high-accuracy DSD predictions minute-by-minute. Compared to only using the common MSE as a loss function, the accuracy modeled with the SLF was significantly enhanced according to the multiple quantitative evaluation indicators. The prediction results were beneficial to cloud physics and dynamic research, radar quantitative precipitation estimation, and weather modification operations, which can provide new ideas and methods for DSD research.

During the process of modeling, the quality control and standardization of DSD data were indispensable. From the perspective of time, the sequence autocorrelation decreases when the time increases, and the prediction accuracy worsens. From the perspective of cloud types, the prediction of stratiform clouds was better than that of mixed convective—stratiform clouds, while the mixed convective—stratiform clouds were better than that of convective clouds.

For machine learning, the most important matter was the amount of data. Although the samples in this paper reach more than 1.7 million, more observational data are still needed to achieve better fitting.

Herein, the DSDnet-based model was built with all samples together. However, the DSD in convective and stratiform clouds are significantly different. Although the model prediction is relatively satisfactory, the data from stratiform clouds account for the vast majority of the training dataset, which results in a large error in convective cloud precipitation. Therefore, it would be better to model by distinguishing different clouds if there are sufficient samples. In addition, with the polarization upgrading of weather radar in most countries and regions, the DSD can be retrieved by dual polarimetric radar data first, and then the model can be used to obtain the prediction of DSD for the whole radar detection range.

At present, artificial intelligence technology is booming and a variety of new deep learning algorithms have endlessly emerged, and using more new algorithms to construct a DSD prediction network may achieve more accurate predictions.

Author Contributions: Y.Z. led manuscript writing and contributed to data analysis and research design. Z.H. supervised this study, contributed to the research design, manuscript writing and discussion of the results, and served as the corresponding author. S.Y., J.Z., D.L. and F.H. contributed to data analysis and model design. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Science and Technology Plan Projects of Sichuan Province (2021YJ0280), the Key Project of Monitoring, Early Warning and Prevention of Major Natural Disasters of China (2019YFC1510304), the Key-Area Research and Development Program of Guangdong Province (2020B1111200001), the Joint Fund of Key Laboratory of Atmosphere Sounding, CMA and Research Centre on Meteorological Observation Engineering Technology, CMA (U2021Z05), the National Natural Science Foundation of China (No. 42105141), the Basic Research Fund of CAMS (2020Y017), the Science & Technology Plan Project of Fujian Province (2021L3019).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to sincerely thank Fen Xu of the Nanjing Joint Institute for Atmospheric Sciences for providing technical guidance.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References
1. Gilmore, M.S.; Straka, J.M.; Rasmussen, E.N. Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme. Mon. Weather Rev. 2004, 132, 2610–2627. [CrossRef]
2. Krishna, U.V.M.; Reddy, K.K.; Seela, B.K.; Shirooka, R.; Lin, P.L.; Pan, C.J. Raindrop size distribution of easterly and westerly monsoon precipitation observed over Palau islands in the Western Pacific Ocean. Atmos. Res. 2016, 174, 41–51. [CrossRef]
3. Marshall, J.S.; Palmer, W.M. The distribution of raindrops with size. J. Meteor. 1948, 5, 165–166. [CrossRef]
4. Chen, B.J.; Li, Z.H.; Liu, J.C.; Gong, F.J. Model of raindrop size distribution in three types of precipitation. Acta Meteorol. Sin. 1998, 4, 506–512.
5. Zheng, H.J.; Chen, B.J. Comparative study of exponention and Gamma functional fits to observed raindrop size distribution. Sci. Meteorol. Sin. 2007, 27, 17–23.
6. Gong, F.J.; He, Y.J.; Wang, J.H. Characteristics of raindrop size distributions of Northeast cold vortex precipitation in China. Sci. Meteorol. Sin. 2007, 4, 365–373.
7. Ulbrich, C.W. Natural variations in the analytical form of the raindrop size distribution. J. Clim. Appl. Meteorol. 1983, 22, 1764–1775. [CrossRef]
8. Willis, P.T. Functional fits to some observed drop size distributions and parameterization of rain. J. Atmos. Sci. 1984, 41, 1648–1661. [CrossRef]
9. Kirankumar, N.V.P.; Rao, T.N.; Radhakrishna, B.; Rao, D.N. Statistical characteristics of raindrop size distribution in southwest monsoon season. J. Appl. Meteorol. Climatol. 2008, 47, 576–590. [CrossRef]
10. Wu, Y.H.; Liu, L.P. Statistical Characteristics of Raindrop Size Distribution in the Tibetan Plateau and Southern China. Adv. Atmos. Sci. 2017, 34, 727–736. [CrossRef]
11. Wang, G.L.; Li, R.; Sun, J.S.; Xu, X.D.; Zhou, R.R.; Liu, L.P. Comparative Analysis of the Characteristics of Rainy Season Raindrop Size Distributions in Two Typical Regions of the Tibetan Plateau. Adv. Atmos. Sci. 2022, 39, 1062–1078. [CrossRef]
12. Zeng, G.P.; Zheng, X.Z.; Fang, S.Z. Research on the Method of Evaluating the Efficiency of the Non-Randomized Artificial Pre-cipitation Experiments. Chin. J. Atmos. Sci. 1994, 18, 233–242.
13. Liu, H.Y.; Lei, H.C. Characteristics of Rain from Stratiform Versus Convective Cloud Based on the Surface Raindrop Data. Chin. J. Atmos. Sci. 2006, 30, 693–702.
14. Yang, J.M.; Chen, B.J.; Han, Y.X.; Li, P.R. Statistical characteristics of raindrop size distribution in different regions of Shanxi. J. Meteorol. Sci. 2016, 36, 88–95.
15. Chi, Z.P.; Liu, X.; Chen, J.M. Calculation and Analysis of Z-I Relation among Precipitation Processes Caused by Sheet Cloud in Spring and Autumn. Meteor. Mon. 2000, 26, 35–37.
16. Graves, A. Long short-term memory. In Supervised Sequence Labelling with Recurrent Neural Networks; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45.
17. Sekertekin, A.; Bilgili, M.; Arslan, N.; Yildirim, A.; Celebi, K.; Ozbek, A. Short-term air temperature prediction by adaptive neu-ro-fuzzy inference system (ANFIS) and long short-term memory (LSTM) network. Meteorol. Atmos. Phys. 2021, 133, 943–959. [CrossRef]
18. Liu, H.; He, B.; Qin, P.; Zhang, X.; Guo, S.; Mu, X. 2021: Sea level anomaly intelligent inversion model based on LSTM-RBF network. Meteorol. Atmos. Phys. 2021, 133, 245–259. [CrossRef]
19. Zhang, C.J.; Zeng, J.H.; Wang, Y.; Ma, L.M.; Chu, H. Correction model for rainfall forecasts using the LSTM with multiple meteorological factors. Meteorol. Appl. 2020, 27, e1852. [CrossRef]
20. Ni, Z.; Liang, P. Fine temperature forecast based on LSTM deep neural network. Comput. Appl. Softw. 2018, 35, 233–236+271.
21. Yao, W.; Huang, P.; Jia, Z. Multidimensional LSTM networks to predict wind speed. In Proceedings of the 2018 37th Chinese Control Conference (CCC), IEEE 2018, Wuhan, China, 25–27 July 2018; pp. 7493–7497.
22. Atlas, D.; Srivastava, R.C.; Sekhon, R.S. Doppler radar characteristics of precipitation at vertical incidence. Rev. Geophys. 1973, 11, 1–35. [CrossRef]
23. Li, H.; Yin, Y.; Shan, Y.P.; Jin, Q. Statistical Characteristics of Raindrop Size Distribution for Stratiform and Convective Precipitation at Different Altitudes in Mt. Huangshan. Chin. J. Atmos. Sci. 2018, 42, 268–280.
24. Testud, J.; Oury, S.; Black, R.A. The concept of “normalized” distribution to describe raindrop spectra: A tool for cloud physics and cloud remote sensing. J. Appl. Meteorol. 2001, 40, 1118–1140. [CrossRef]
25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780. [CrossRef]
26. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. Neural Comput. 2000, 12, 2451–2471. [CrossRef]