Evaluating Past Horizons for U-Net-Based Precipitation Nowcasting with Radar Data in Southeastern Pará, Brazil

Rafael Rocha[1,2], Eduardo Carvalho[1], Ewerton Oliveira[1],
Sergio Viademonte[1], Douglas Ferreira[1], Ronnie Alves[1]

[1] Vale Institute of Technology, Brazil
rafael.lima.rocha@pq.itv.org, eduardo.costa.carvalho@itv.org,
ewerton.oliveira@pq.itv.org, sergio.viademonte@itv.org,
douglas.silva.ferreira@itv.org, ronnie.alves@itv.org
[2] Federal University of Pará, Brazil
rafael.lima@icen.ufpa.br

Abstract.

Severe weather events significantly impact daily life, especially during emergencies, affecting human lives and the economy. Decision-making in such events, like heavy rainfall, strong winds, and flash floods, is challenging due to the rapid changes and strong interconnections between the variables involved. Nowcasting models use real-time data, commonly from weather radars, to forecast short-term rain up to 6 hours ahead, supporting decision-making in severe weather situations. These models provide early warnings and precise information about the location, intensity, and duration of these events. Recently, machine learning models for precipitation forecasting have gained prominence due to their ability to learn from data and offer reliable and fast predictions. This study explores precipitation nowcasting using weather radar data in the southeast of Pará, Brazil, focusing on a one-hour forecast horizon utilizing the U-Net architecture. Four models based on U-Net algorithm, investigating past horizons of 30, 60, 90, and 120 minutes, are evaluated using categorical and continuous metrics, and a visual comparison of the 60-minute forecast horizon. The results demonstrate that the model with a past horizon of 120 minutes outperforms the other models in all evaluated metrics, achieving 37.96 and 0.5476 scores in continuous and categorical metrics, respectively, improving the forecast of severe events and decision making up to a 60-minute forecast horizon.

CCS Concepts: • **Computing methodologies** → **Neural networks**.

Keywords: machine learning, past horizons, precipitation nowcasting, weather radar

1. INTRODUCTION

During emergencies, severe weather events profoundly affect various aspects of daily life impacting human life and the economy, with sectors that heavily rely on accurate meteorological information being particularly harmed. Decision-making in severe events like heavy rainfall, strong winds, and flash floods is extremely challenging, as the variables involved are strongly interconnected and change rapidly during the event.

Nowcasting models use real-time data, commonly from weather radar, to forecast short-term rain, up to 6 hours, to support decision-making in severe events. These models provide early warnings and precise information about the location, intensity, and duration of these events. Machine learning models for precipitation forecasting have gained prominence in recent years, standing out for learning from data and providing reliable and fast predictions.

Machine learning for precipitation nowcasting using weather radar data has been studied with a focus on various regions of Brazil to investigate the amount of input data (past horizon) required for models to achieve better performance, given a determined number of outputs, i.e., the forecast horizon.

The work by [Jorge et al. 2022] investigates precipitation nowcasting using weather radar data with a temporal resolution of 10 minutes, through the ConvLSTM architecture initially proposed by [Shi et al. 2015]. The study explores past horizons of 40, 50, and 60 minutes to forecast the next 40 minutes in the metropolitan area of São Paulo. The results presented by [Jorge et al. 2022] show lower prediction errors compared to persistence models, although the results do not indicate significant differences between models with different past horizons.

Similarly, [Caseri et al. 2022] utilizes ConvLSTM with weather radar data to forecast heavy rainfall around Campinas, in southeastern Brazil, due to the high number of floods in this region. The study investigates rain prediction for the next 10, 20, and 30 minutes, given the previous 50 minutes, with radar data generated every 10 minutes. The results indicate that the longer the forecast horizon, the lower the prediction accuracy, although these forecasts outperform the persistence model for different forecast horizons.

On the other hand, the work by [Bonnet et al. 2020] performs precipitation nowcasting using radar data from Ponta Nova (São Paulo) through the PredRNN++ algorithm. This architecture was previously proposed by [Wang et al. 2018] for solving space-time predictive learning problems. The study by [Bonnet et al. 2020] frames precipitation nowcasting as a video prediction problem, evaluating the forecast of a sequence of images over one hour, based on the sequence of images from the previous hour. This study demonstrates the potential of the model as a nowcasting tool to assist in decision-making and efficient alerts.

This work uses the U-Net architecture for precipitation nowcasting. U-Net was initially designed by [Ronneberger et al. 2015] for biomedical image segmentation. In the context of precipitation forecasting, [Ayzel et al. 2020] was one of the first to treat it as an image-to-image translation problem using the encoder and decoder components.

Therefore, this study aims to explore precipitation nowcasting using weather radar data in the southeast of Pará, Brazil, considering a one-hour forecast horizon utilizing the U-Net architecture. Four models are created to investigate the number of input images, determining which past horizon best captures and improves the complex patterns in precipitation nowcasting. The past horizons investigated are 30, 60, 90, and 120 minutes. Each model is evaluated using categorical and continuous metrics, as well as a visual comparison of the 60-minute forecast horizon. Additionally, a full day of forecasts is analyzed for municipalities within the radar coverage area.

2. MATERIALS AND METHODS

2.1 Coverage area and data

The data used in this work come from the weather radar located in southeastern Pará, in the northern region of Brazil. The radar operates with a sweep covering 150 km and has a temporal resolution of 5 minutes, generating data volumes (with a resolution of 300 x 300) that include information such as maximum reflectivity, which is the variable used for precipitation nowcasting in this work. Figure 1 shows the map of Pará with the radar's coverage area (red circle) in the southeast of the region.

The database consists of 5,353 data volumes generated by the weather radar, covering 28 days of 2021 within the rainy season in the region, which occurs between November and April. These data are used to generate the sequences for training the precipitation nowcasting model.

To evaluate the performance of the precipitation forecasts at specific points within the coverage area, three municipalities in the northern region of Brazil are investigated: Canaã dos Carajás, Curionópolis,

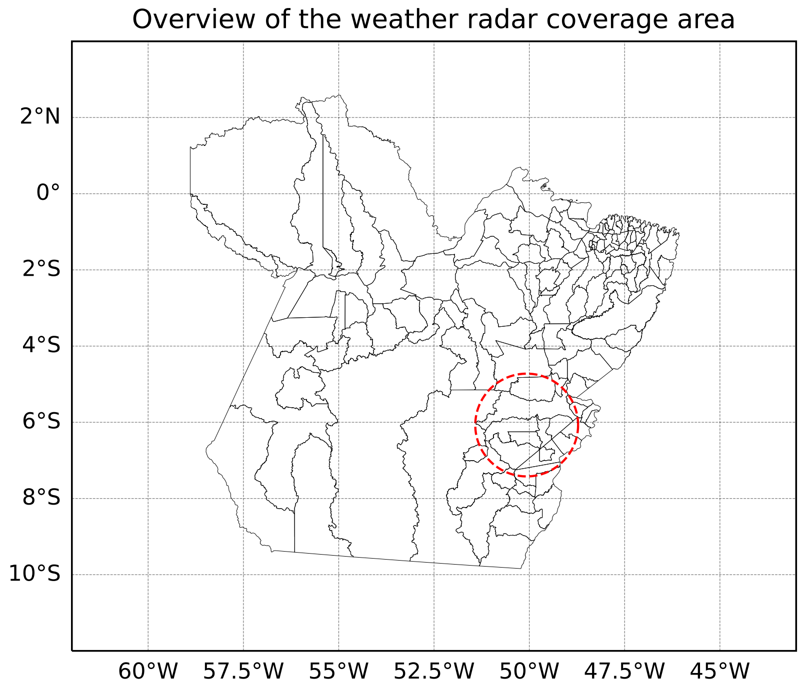


Fig. 1: Map of the state of Pará in the north of Brazil showing an overview of the weather radar coverage area located in the southwest region of the state.

and Parauapebas. Figure 2 shows the three municipalities and their locations within the radar's operational area.

2.2 Reflectivity and precipitation

The variable used in this work is maximum reflectivity (dBZ), which corresponds to the maximum value of each data volume at ten elevations of the weather radar, representing the data collection between 1-10 km in height. The relationship between reflectivity and precipitation (rainfall) is given by the power relationship presented in Equation 1:

$$Z = aR^b \tag{1}$$

Where a and b are the parameters that influence the transformation, R is the precipitation rate (mm/h), and Z is the reflectivity factor (mm^6/m^3). The reflectivity used here represents the logarithmic version of Z.

Table I presents the relationship between reflectivity and precipitation values. To generate the relationship, the parameters a and b were set to 200 and 1.6, respectively, which are widely used and were originally defined by [Marshall and Palmer 1948].

It is possible to observe that the relationship between Z and R is a power relationship (Table I), where any change in reflectivity values significantly alters the precipitation values. For example, a variation of 10 dBZ between 20 and 30 dBZ results in a 321% increase in precipitation. The value of 20 dBZ (in bold) indicates light rain.

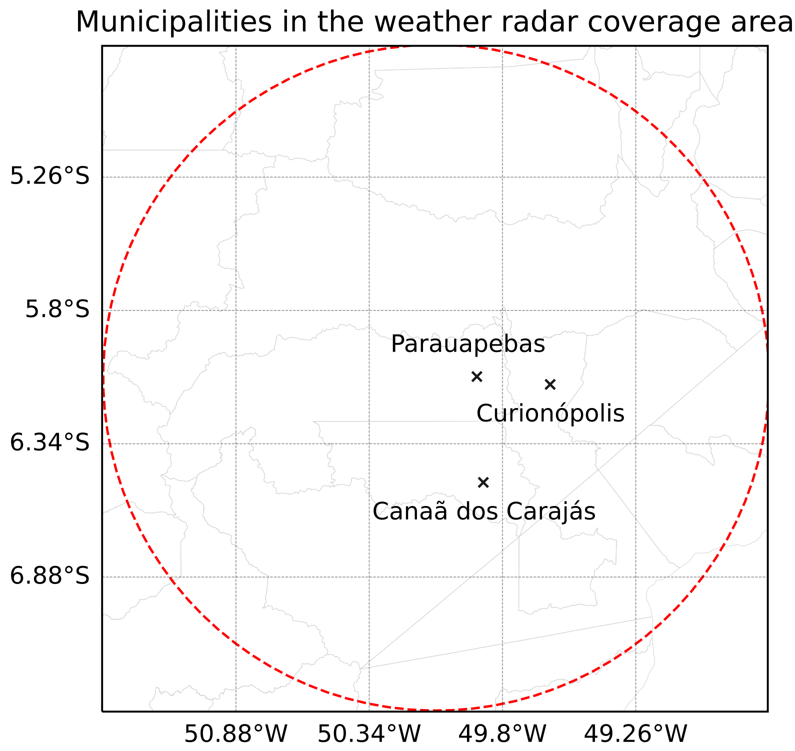


Fig. 2: Municipalities in southeastern Pará within the meteorological radar coverage area.

Table I: Relationship between reflectivity and precipitation values. The first line represents reflectivity (dBZ), and the line below shows precipitation values in mm/h.

Reflectivity (dBZ)						
5	10	**20**	30	40	50	60
0.0748	0.1537	**0.6484**	2.7343	11.5307	48.6246	205.0483

2.3 Next hour prediction approach

The approach used in this work aims to investigate the necessary past horizon to achieve better forecasts for a one-hour prediction horizon (60 minutes). In other words, it evaluates the number of radar data volumes/images needed (or not) to precipitation nowcasting for the next 60 minutes of a meteorological event.

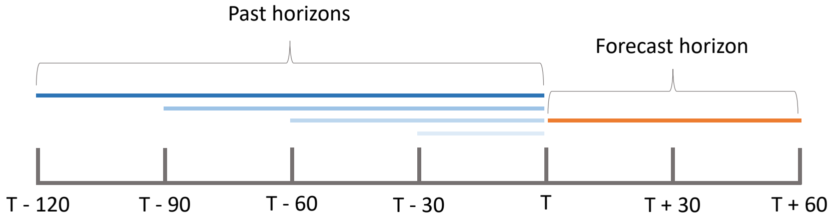


Fig. 3: Overview of the evaluated approach for next-hour precipitation nowcasting.

An overview of the approach evaluated for forecasting the next hour's precipitation is presented in Figure 3. Here, the focus is on the 60-minute forecast horizon, represented by the orange timeline from T to $T + 60$. The past horizons are represented by blue-toned lines. Four past horizons are investigated: 30 minutes, 60 minutes, 90 minutes, and 120 minutes, represented by the times $T - 30$, $T - 60$, $T - 90$, and $T - 120$ to T, respectively.

To ensure consistency in constructing precipitation nowcasting model sequences and to maintain a fixed forecast horizon for evaluation, all past horizons are based on a 120-minute past horizon, filtering data to represent specific past horizons for each model. For example, sequences for 30 and 60-minute past horizons, starting at 08:00, have fixed inputs from 08:00 to 10:00 and outputs from 10:05 to 11:00, with inputs for the 30-minute model from 09:30 to 10:00 and for the 60-minute model from 09:00 to 10:00.

2.4 Training setup

In this work, we use the U-Net architecture for precipitation nowcasting. The encoder portion consists of 4 blocks with 16, 32, 64, and 128 filters, respectively, each block having a kernel size of 3 x 3 and ReLU activation function. The central block contains 512 filters, with the same kernel size and activation function. The decoder portion has the same blocks, filters, kernel size, and activation function as the encoder, with the difference being the presence of a transposed convolutional layer at the beginning of each block, aimed at expanding the dimensionality of the filters with a kernel size of 2 x 2. Finally, the final convolutional layer uses a linear activation, with the number of outputs equal to 12, corresponding to the forecast horizon analyzed in this work.

The work investigates previous horizons, so the input data size of the U-Net architecture varies between 7, 13, 19, and 25 images, representing previous horizons of 30, 60, 90, and 120 minutes, respectively. To maintain a fair forecast horizon of 60 minutes during sequences building, the past horizon was set to 120 minutes, regardless of the past horizon, by selecting the appropriate number of input images for the specified past horizon during training.

The loss function used in this work is the mean squared error (MSE), which evaluates the errors between observed and predicted data sequences. Additionally, the Adam optimizer is used to minimize the loss function over 100 training epochs with a batch size of 4.

The complete dataset consists of 3,981 input and output sequences generated from 28 days of weather radar data. For model training, 80% of the data was used for training and 20% for testing, resulting in 3,184 and 797 sequences, respectively.

2.5 Evaluation metrics

To evaluate the quality of precipitation nowcasting, both continuous and categorical metrics are considered. The continuous metric is MSE, which measures the mean squared difference between observed and predicted values; the lower the MSE, the better the model's performance. MSE is defined by Equation 2, where y is the observed data, $\hat{y}$ is the forecast and M is total number of observations.

$$MSE = \frac{1}{M} \sum_{i=1}^{M} (y_i - \hat{y}_i)^2 \tag{2}$$

For categorical metrics, it is necessary to first transform the observed and predicted values through binarization, so the values become 0s or 1s (true or false). In this work, a threshold of 20 dBZ, representing light rain (bold in Table I), is used to binarize both forecasts, where values greater than 20 dBZ represent category 1 and values less than this threshold are designated as category 0.

Thus, these metrics are similar to those derived from the confusion matrix commonly used in machine learning, considering true positive (TP), false positive (FP), false negative (FN), and true negative (TN).

The categorical metric used in this work is the critical success index (CSI), a technique frequently used by the meteorological community for precipitation nowcasting, that the higher the value obtained by the CSI metric, the better the performance of the forecast. It can be defined as the fraction of predicted values correctly categorized as 1 relative to all values categorized as 1, whether observed or predicted. The categorical metric CSI is defined by Equation Y,

$$CSI = \frac{TP}{TP + FP + FN} \tag{3}$$

3. RESULTS AND DISCUSSIONS

The results presented in Table II show that the past horizon of 120 minutes outperforms for precipitation nowcasting in the next hour using weather radar reflectivity data. For both metrics, this past horizon achieves the best results, with an MSE of 37.9635 and a CSI of 0.5476.

Based on MSE, the 30-minute model has the poorest performance, presenting the highest error among the models, with an MSE of 43.1882. Despite this, it performs well in terms of CSI, ranking second only to the model with a past horizon of 120 minutes.

Table II: General result of the next hour's precipitation nowcasting given the different past horizons.

Metrics	30 min	60 min	90 min	120 min
MSE ↓	43.1882	40.2114	42.9521	**37.9635**
CSI ↑	0.5460	0.5401	0.5234	**0.5476**

The results for the 90-minute past horizon (Table II) indicate low performance for both metrics. Regarding CSI, it has the lowest value among all models. For MSE, it has the second worst performance, with an error of 42.9521.

Figure 4 presents a comparison between the observed data and the forecasts made by the models with past horizons of 90 and 120 minutes, specifically the data volumes at times $T + 5$, $T + 30$, and $T + 60$. The images in Figure 4 represent the forecast horizon from 20:05 to 21:00 on March 13, 2021, with model inputs from 18:30 to 20:00 and 18:00 to 20:00 for the 90-minute and 120-minute past horizons, respectively.

It is observed in Figure 4 that as time progresses from T, the forecasts degrade in terms of reflectivity values, although the movement of the meteorological event is captured efficiently. It is also noted that the results presented in Figure 4 align with the general results presented in Table II, where the 120-minute past horizon shows better performance compared to the 90-minute horizon. This is especially evident in the forecasts at $T + 30$ and $T + 60$, where reflectivity values above 20 dBZ are better distributed throughout the meteorological event, slightly closer to what is seen in the observed data.

A comparison between reflectivity and precipitation variables, in terms of observed data and forecasts for a full day in the municipality of Curionópolis, from the 60-minute and 120-minute past horizon models, is presented in Figure 5. The chosen day for analysis is January 8, 2022, meaning the data presented here are outside the training and testing sets.

Figure 5a presents the comparison between observed data and forecasts for the 60-minute and 120-minute horizons. Additionally, the red dashed line indicates the 20 dBZ threshold, which represents

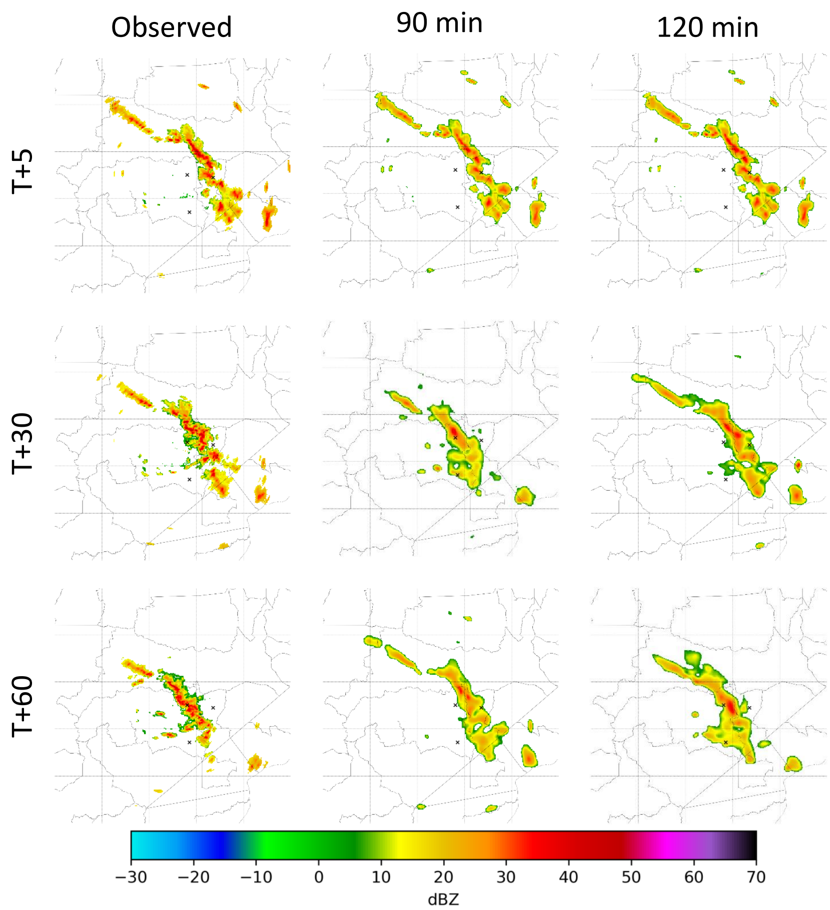
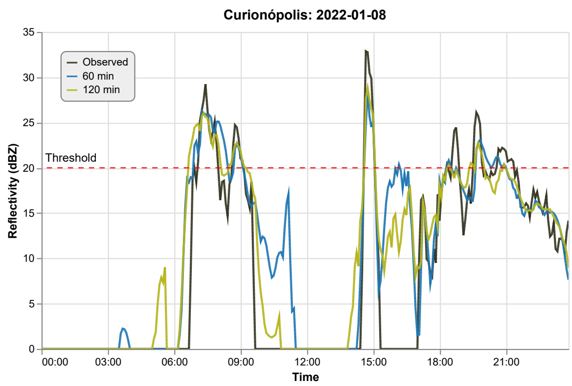
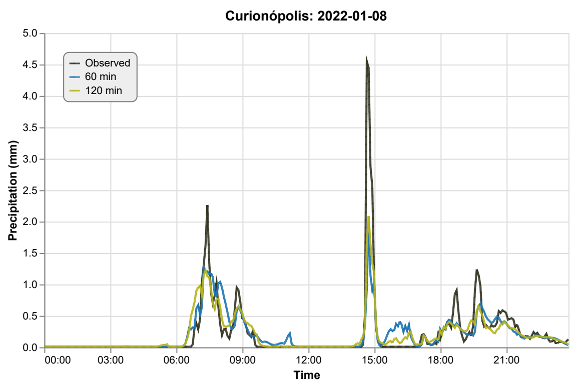


Fig. 4: Comparison between observed data and forecasts from past horizons of 90 and 120 minutes to forecast horizons of 5, 30 and 60 minutes for March 13, 2021.



(a) Reflectivity



(b) Precipitation

Fig. 5: Comparison between observed data and forecasts from past horizons of 60 and 120 minutes for the reflectivity and precipitation variables.

light rain. On the other hand, Figure 5b presents the same values as Figure 5a, but transformed into precipitation using Equation 1, as demonstrated in Table I.

When comparing the results of Figures 5a and 5b, it is generally observed that the model with a 120-minute past horizon performs better. Both models have significant errors, especially below the 20 dBZ threshold, although the forecast with the 60-minute past horizon is considerably worse during times with no reflectivity (Figure 5a), such as between 09:00 and 12:00 and 15:00 and 18:00.

Despite the large errors (reflectivity, Figure 5a) below the threshold, it is noted that when transformed into precipitation (mm/h), Figure 5b, the errors become negligible. Analyzing the forecasts above the threshold, it is possible to observe a distinct behavior compared to the results below the threshold. Here, forecasts with a difference of less than 5 dBZ, such as the results between 07:00 and 09:00 and around 15:00 (Figure 5a), generate much larger errors when converting from dBZ to mm/h, as shown in the same time intervals in Figure 5b.

4. CONCLUSIONS AND FUTURE WORKS

This study investigates precipitation nowcasting using weather radar data located in southeastern Pará, Brazil. Past horizons of 30, 60, 90, and 120 minutes are evaluated to estimate a 60-minute forecast horizon using the U-Net architecture. Results indicate that the 120-minute past horizon model outperforms others, achieving an MSE of 37.9635 and CSI of 0.5476. Additionally, visual analysis of forecasts and a comprehensive day of predictions in Curionópolis municipality contribute to these findings.

Future work will explore other forecast horizons for the various past horizons and include additional continuous and categorical metrics, as well as similarity metrics, to better evaluate precipitation nowcasting performance. Furthermore, persistence models widely used in meteorological community will be evaluated to compare with models for each past horizon, alongside other machine learning approaches.

REFERENCES

Ayzel, G., Scheffer, T., and Heistermann, M. Rainnet v1. 0: a convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development* 13 (6): 2631–2644, 2020.

Bonnet, S. M., Evsukoff, A., and Morales Rodriguez, C. A. Precipitation nowcasting with weather radar images and deep learning in são paulo, brasil. *Atmosphere* 11 (11): 1157, 2020.

Caseri, A. N., Santos, L. B. L., and Stephany, S. A convolutional recurrent neural network for strong convective rainfall nowcasting using weather radar data in southeastern brazil. *Artificial Intelligence in Geosciences* vol. 3, pp. 8–13, 2022.

Jorge, A. A. S., Quiles, M. G., Costa, I. C., and Santos, L. B. L. A convolutional LSTM neural network for precipitation nowcasting based on weather radar data. In *XXIII Brazilian Symposium on Geoinformatics - GEOINFO*, L. B. L. Santos and M. de Arruda Pereira (Eds.). MCTIC/INPE, São José dos Campos, SP, Brazil, pp. 384–388, 2022.

Marshall, J. S. and Palmer, W. M. K. The distribution of raindrops with size. *Journal of Atmospheric Sciences* 5 (4): 165–166, 1948.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, October 5-9, 2015, proceedings, part III 18*. Springer, Munich, Germany, pp. 234–241, 2015.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* vol. 28, pp. 802–810, 2015.

Wang, Y., Gao, Z., Long, M., Wang, J., and Philip, S. Y. Predrnn++ towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In *International conference on machine learning*. PMLR, Stockholmsmässan, Stockholm, Sweden, pp. 5123–5132, 2018.