# Machine Learning Prediction of Precipitation in Metro Manila, Philippines

*Akira Noda(1), Mitsuteru Sato(2), Yukihiro Takahashi(2), Hisayuki Kubota(2), Kenichi Fukui(3)
1. Department of Cosmoscience, Graduate School of Science, Hokkaido University.
2. Faculty of Science, Hokkaido University.
3. Graduate School of Information Science and Technology, Osaka University.

## Abstract

It is difficult to accurately predict the occurrence and rain volume of torrential rains such as guerrilla rain, rain band with typhoon and linear precipitation zones, even in Japan where meteorological observations from the ground and space and weather forecasts using numerical models are well established. One of the reasons for this is that the spatial narrowness of the rainfall area and spatial resolution of conventional weather observation networks. Furthermore , in Southeast Asia, where the meteorological observation infrastructure is weak, disasters such as heavy rains and floods caused by monsoons and typhoons occur every year. There is an urgent need to establish low-cost and highly viable weather forecasting technology. Our research group has developed an automatic weather and lightning observation device called P-POTEKA since 2017, and has been deploying it in Metro Manila, Philippines, which is frequently affected by heavy rains and associated flooding. Using the P-POTEKA rainfall data, RGB image data corresponding to the spatial distribution of rainfall in Metro Manila was created. By training this time-series image data of rainfall on a machine learning model called ConvLSTM (Convolutional Long-Short Term Memory), we predicted the distribution and amount of rainfall at 10-minutes intervals from the present to one hour from now, based on the observation data of the past hour. As a result, it was found that the prediction using ConvLSTM is relatively accurate up to 30 minutes from the present time, but the prediction accuracy of the spatio-temporal change of prediction becomes significantly worse after 40-60 minutes.

## Introduction

According to the statistical research of rainfalls in Japan by Japan Meteorological Agency [1], the number of the occurrences per year of the heavy rainfalls are increasing year after year. (Figure 1.1)
In recent years, the development of machine learning technology has been remarkable and in the field of meteorology, research on the development of forecasting methods using machine leaning has been actively conducted. (Aifang et al. [2020], Chul-Min et al. [2020]) Both studies compared the precipitation prediction ability of traditional methods with machine learning methods. They shows as a result that machine learning methods predicted more accurate than traditional methods did but the prediction ability are still limited and need more improvement.
Main purpose of this study is to process P-POTEKA data and build a base machine learning model.

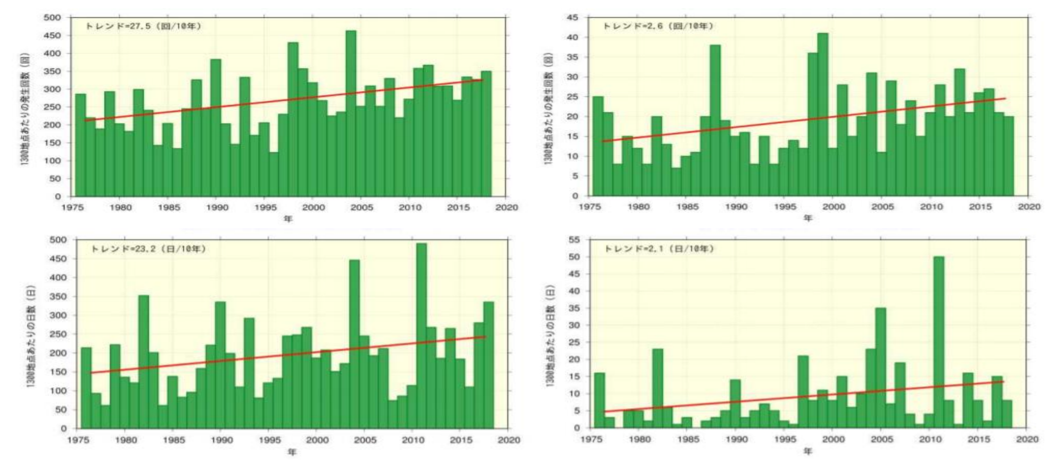


Figure1.1: The number of rainfalls (upper left:50mm/h, upper right:80mm/h, lower left:200mm/h, lower right: 400mm/h)1976-2018. The red line shows long-term change trend.

## Data

Our research group has developed an automatic weather and lightning observation system called as P-POTEKA(Figure 2.1) since 2017. To datem 35 P-POTEKA units have been installed in Metro Manila and continue the acquisition of meteorological data. While AMeDAS (the Automated Meteorological Data Acquisition System) in Japan is deployed with the average interval of 17km, P-POTEKA is deployed with the average interval of 2-3km, making the observation network with the highest spatial and temporal resolution suitable for caputuring extreme weather events(Figure 2.2).

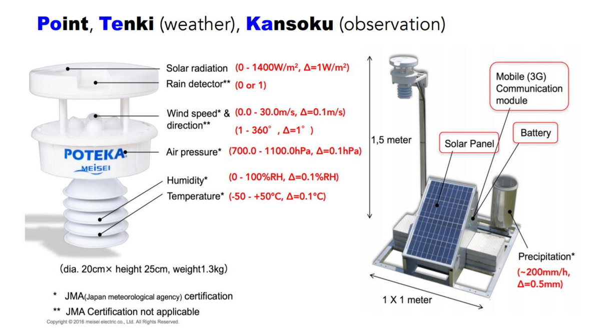



Figure2.1: P-POTEKA AWS (left), P-POTEKA observation device (right)

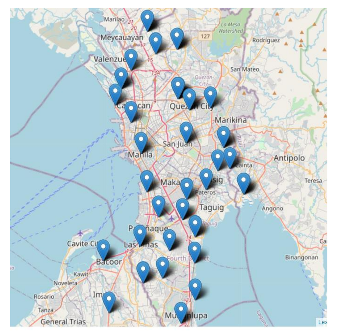
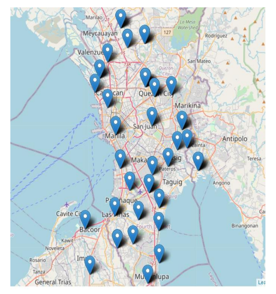


Figure2.2: P-POTEKA distribution map.

## Methodology

To create rainfall images corresponding to the spatial distribution of actual rainfall in MetroMania using P-POTEKA, we interpolate the amount of the hour rainfall of unobserved point by ordinary kriging method. Ordinary kriging method estimate unknown data using spatial autocorrelation of obtained data.



| mm/h | (lon, lat) |
|---|---|
| A | (x1, y1) |
| B | (x2, y2) |
| ... | ... |
| Y | (x34, y34) |
| Z | (x35, y35) |

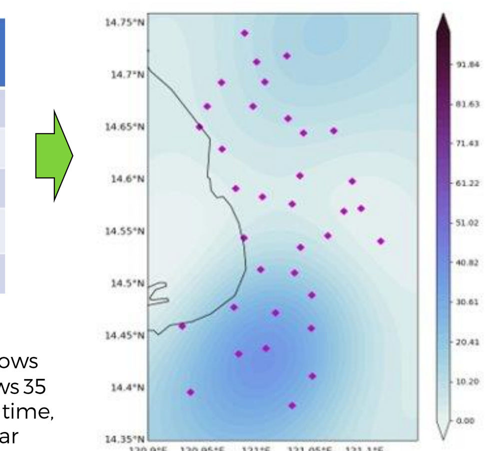


Figure3.1: The schematic image of making rainfall distribution image data using ordinary kriging method. The left image shows the distribution of P-POTEKA devices, the middle image shows 35 observed value of the amount of the hour rainfall at a certain time, the right image is created using interpolated data with ordinar ykriging method.

ConvLSTM (Convolutional Long-Short Term Memory) [Xingjian et al. 2015]combines tow machinel learning models CNN (Convolutional Neural Network) [Hubel et al. 1962] and LSTM (Long-Short Term Memory) [Schmidhuber et al. 1997]. CNN is suitable model to learn image and LSTM is suitable to learn time series data. Therefore, ConvLSTM is suitable model to learn time series image data.

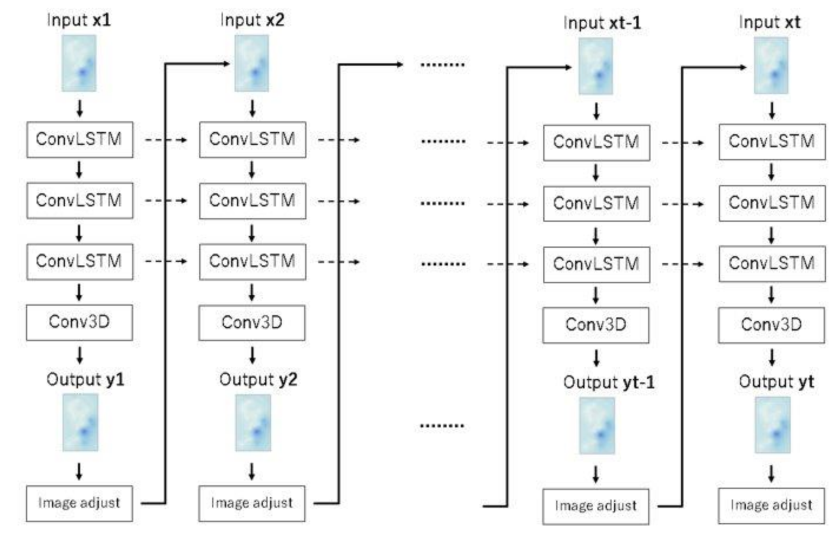



Figure3.2: The schematic image of ConvLSTM model in this study

## Results

About 20000 images (2020/04 ~ 0202/10) are used for training and evaluation. To evaluate model, the average of RMSE (Root Mean Squared Error) between the the RGB values of the predcted images and the label images is calculated for the 40 unlearned rainfall sample images.

The way to train the model and prediction would affect the prediction ability. We compared the two ways (Figure 4.1, Way1 and Way2). We trained two models using the same data but in the different way. As a result (Table4.1), the RMSE average of Way2 is smaller than that of Way2. This may be because Way2 can use closer and more data to make a prediction.

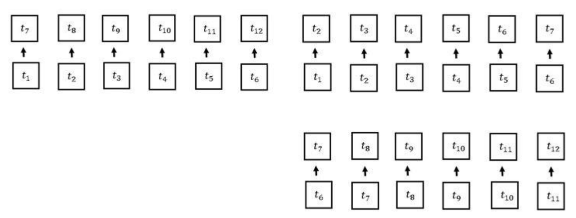


Figure4.1: The schematic image of Way1 (Left) and Way2 (Right). In Way1, predction image's time is 1 hour apart from input image's itme. In Way2, 10 minutes apart from input and using predction image to predict next image.

| Way | RMSE average |
|---|---|
| Way1 | 21.0082 |
| Way2 | 16.0390 |

Table 4.1: The result of evaluation calculating RMSE average of 40 unlearned samples. The Way2's valuse is smaller, which means the prediction ability is higher than Way1.

For more detailed evaluation, we analysed case studies. One of them shows that our model cannot predict correctly because the model is highly affected by the rainfall trend of rainfall image. If the rainfall is getting stronger in the input images, the model predict the rainfall will be getting stronger and vice versa. It is still difficult to predict starting or stopping rain in a short time.

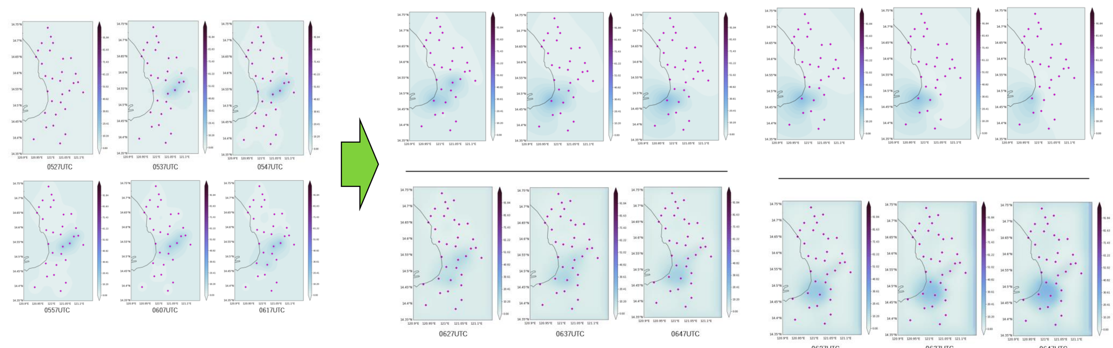


Figure 4.2: The time series images of the rainfall at 2019/11/01 0527UTC~0717UTC. he images on the left side of the arrow are the input images. The images on the right side are the label images (upper row) and the predict images (lower row).

In Figure4.2, relatively small rain moved from the center of the area to southwest. In the prediction image, the southwest of the area getting ticken, but faild to predict getting weak at the center of the are.

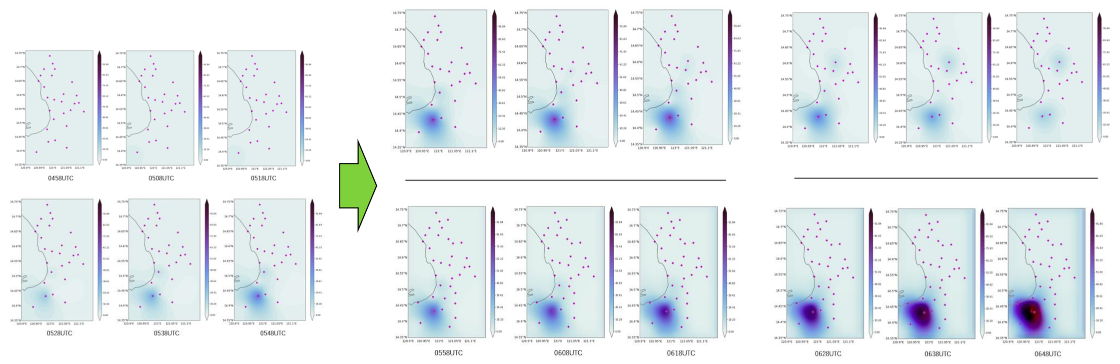


Figure 4.3: The time series images of the rainfall at 2019/10/04 0458UTC~0648UTC. The images on the left side of the arrow are the input images. The images on the left side are the label images (upper row) and the predict images (lower images).

In Figure4.3, it started to rain and was getting heavier at the relatively narrow place. In the label images, the rain was getting weak but the rain is getting heavier too much in the prediction images.

## Results

| time | RMSE |
|---|---|
| 0627UTC | 5.1276 |
| 0637UTC | 7.8675 |
| 0647UTC | 9.3733 |
| 0657UTC | 10.7375 |
| 0707UTC | 13.7348 |
| 0717UTC | 18.0579 |

| time | RMSE |
|---|---|
| 0558UTC | 3.7078 |
| 0608UTC | 7.1577 |
| 0618UTC | 13.4037 |
| 0628UTC | 26.8758 |
| 0638UTC | 43.0286 |
| 0648UTC | 66.0148 |

Table4.2: RMSE between the label image and the predict image of each time of these case studies. The left table is the first case and the right table is the last case study.

From Table4.2, the later RMSE is getting bigger in both cases. One of the reasons for this reuslt is the model uses the predicted images to predict the next time images. The number of predict images used for the next prediction increase if the time later.

## Conclution

Our final goal is to create a machine lerning model that can predict the torrential rainfalls. And the purpose of this study is to process the P-POTEKA data, to create the base machine learning model and evaluate its basic prediction ability.
In this study, we made the time series images from the P-POTEKA rain data interpolated by ordinary kriging method and train ConvLSTM model with these images. Then, we made the predict images by the trained model and evaluate model performance by RMSE of RGB value of between the predict image and the label image.
If the input image upload every 10 minutes and only first 3 predict images (30 minutes future), the prediction accuracy would be somewhat imported (Figure 5.1).

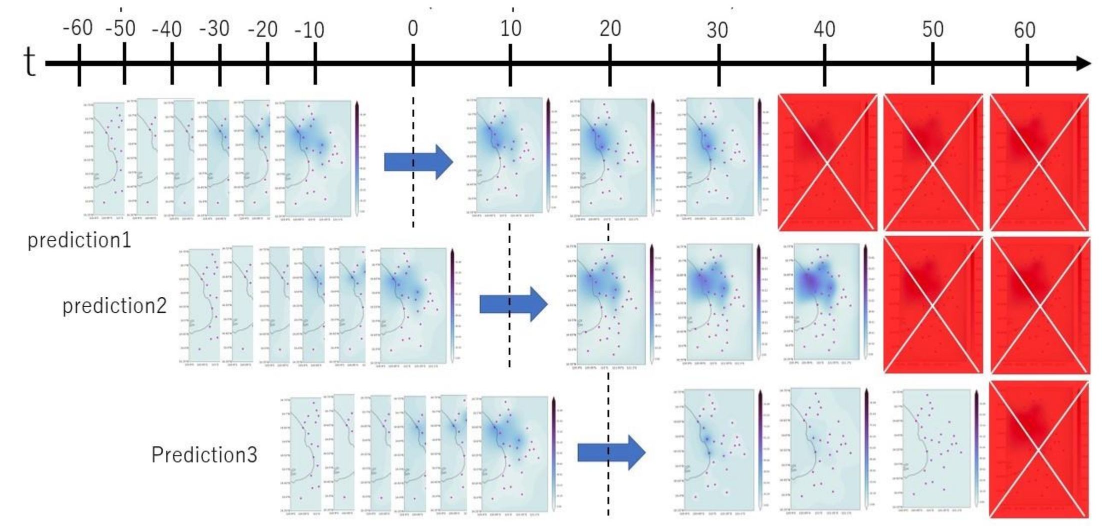



Figure 5.1: The schematic image updating the input images and making prediction every 10 minutes.

However the prediction accuracy is too bad to use in practice. We can get more P-POTEKA data in the future. And model's hyper parameters can be more suitable. We are also planning to use additional data like temperature.

## References

[1] 気象庁, "気象庁業務評価レポート（2020年度版）", 気象庁, 2020
[2] Aifang Su, Han Li, Liman Cui and Yungang Chen, "A Convection Nowcasting Method Based on Machine Learning", Hindawi, China, 2020
[3] Chul-Min Ko, Yeong Yun Jeong, Young-Mi Lee and Byung-Sik Kim, "The Development of a Quantitative Precipitation Forecast Correction Technique Based on Machine Learning for Hydrological Applications", MDPI, Korea, 2020
[4] Z. C. H. W. D.-Y Y. W.-k W. W.-c W. Xingjian Shi, "Convolutional LSTM Network: A Machine Learning Approach For Precipitation Nowcasting", Advances in Neural Information Processing Systems, 2015.
[5] D. H. H. a. T. N. Wiese, "Receptive fields, binocular", The Journal of Physiology, 1962
[6] S.H.a.j. Schmidhuber, "Long short-term memory", Neural Computation, vol. 9, no 8, 1997