Assessing the Performance of Deterministic Precipitation Nowcasting Algorithms with Weather Radar Data and Multimetric Verification

Abdullah Ali[1], Alif Adiyasa[1], Andri Ramdhani[1]
[1] Public Weather Services Directorate, Indonesia Agency for Meteorology Climatology and Geophysics (BMKG)

e-mail: abdullah.ali@bmkg.go.id

Received: 15-04-2025; Revised: 28-04-2025; Approved: 15-07-2025

Abstract. Precipitation nowcasting plays a crucial role in disaster mitigation, especially in tropical regions where convective rainfall develops rapidly. Evaluating deterministic radar-based nowcasting methods is essential for improving real-time forecasts. This study assesses the performance of four deterministic radar-based nowcasting algorithms—Spatial Prognosys (SPROG), Extrapolation, Autoregressive Nowcasting using Vertically Integrated Liquid (ANVIL), and Lagrangian Integro-Difference equation model with Autoregression (LINDA)—using a single case from the Tangerang C-band weather radar. Forecasts were verified up to +96 minutes through spatial inspection, ROC analysis, and Taylor/Target diagrams. At short lead times (+8 to +24 minutes), all algorithms achieved high discrimination skill (AUC > 0.90), with SPROG reaching the highest AUC of 0.96, followed closely by LINDA at 0.95. However, while SPROG excelled in categorical discrimination, statistical evaluations showed that LINDA consistently preserved rainfall structure, achieving correlation coefficients ≥ 0.80 at short range and approximately 0.65 at +80 minutes. Target diagrams indicated low bias (< ±0.1) and relatively stable uRMSD values across lead times for LINDA, whereas SPROG exhibited increasing overdispersion and structural error at longer forecasts. These findings highlight the complementary strengths of different nowcasting approaches and emphasize LINDA's robustness in maintaining structural accuracy, supporting its potential for operational nowcasting applications in tropical convective environments.

Keywords: *weather radar, deterministic nowcasting, ROC, Taylor Diagram, Target Diagram*

1 INTRODUCTION

Precipitation nowcasting refers to very short-range forecasting, typically covering 0 to 6 hours ahead. This capability plays a crucial role in mitigating hydrometeorological disasters by enabling timely warnings of severe weather events. Previous studies have extensively evaluated nowcasting techniques in temperate regions using benchmark datasets such as the Sydney, Zürich, or Finnish radar domains (e.g., Germann and Zawadzki, 2002; Seed, 2003; Pulkkinen et al., 2019; 2022). However, there is limited evaluation of these methods in tropical environments, particularly in Southeast Asia, where convective storms develop faster and on smaller spatial scales. This creates challenges for methods originally designed for stratiform or mixed precipitation regimes. Moreover, while several studies have compared deterministic nowcasting methods using either categorical scores (e.g., CSI, POD, FAR) or correlation-based metrics, relatively few incorporate multimetric frameworks such as ROC analysis alongside Taylor and Target diagrams to evaluate spatial pattern similarity and error characteristics comprehensively. This gap is particularly relevant for assessing performance consistency across lead times.

Therefore, this study aims to evaluate and compare the performance of four deterministic radar-based nowcasting algorithms—S-PROG, Extrapolation, ANVIL, and LINDA—under tropical convective conditions. While BMKG has implemented the probabilistic STEPS algorithm operationally (Ali et al., 2022), deterministic methods remain relevant due to their simplicity and suitability for real-time applications. The evaluation is conducted using a single case from the Tangerang C-band weather radar, with forecasts verified through multiple

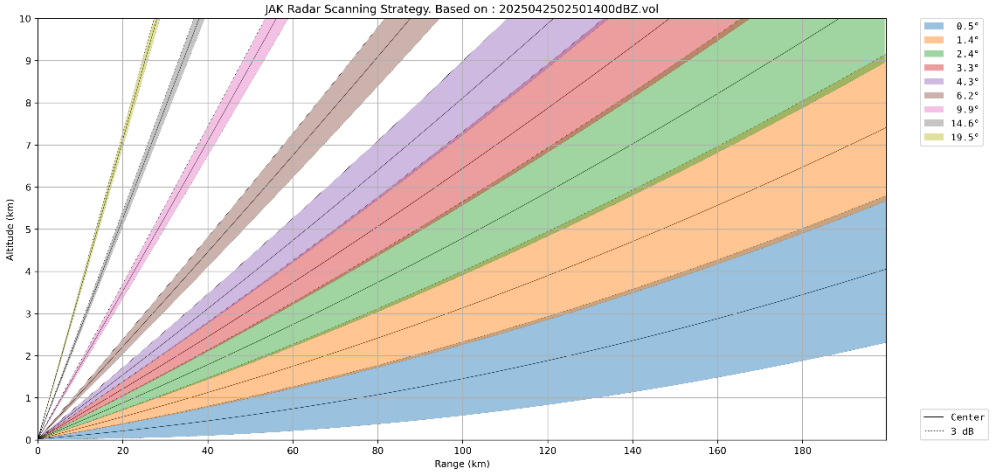


Figure 2-1: Tangerang radar beam spreading.

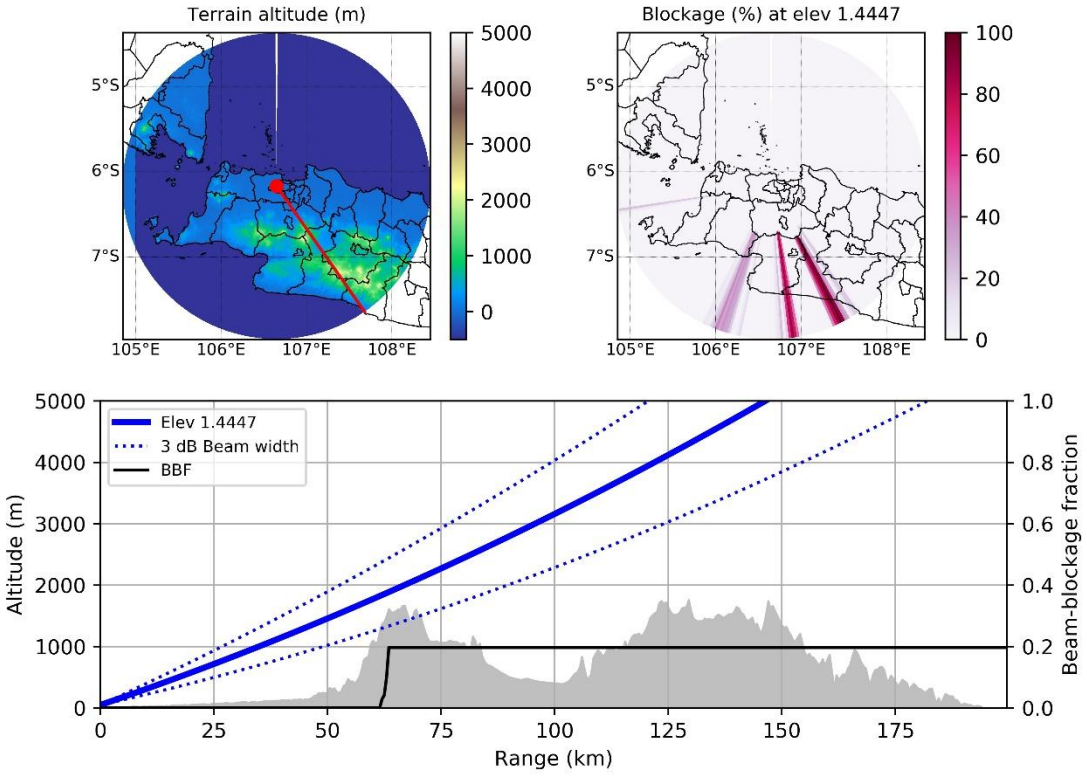


Figure 2-2: Tangerang radar beam blockage analysis.

statistical metrics, including Receiver Operating Characteristic (ROC) curves at a 1 mm/h threshold, Taylor diagrams, and Target diagrams. This multimetric approach is intended to identify the most consistent and accurate algorithm, with implications for both operational use and future research in the Indonesian context.

2 MATERIALS AND METHODOLOGY

This study is based on a single convective rainfall event observed on 5 May 2020 by the Tangerang C-band radar. Radio interference filter already applied to the radar data (Ali et al., 2021) Radar reflectivity data from a two-hour window prior to the forecast time (07:00–09:00 UTC) were used as input for all nowcasting models. Although the single-case approach enables a detailed comparison of algorithm performance under specific tropical conditions, it limits the generalizability of the results. Therefore, further validation across multiple events and diverse meteorological conditions is required to confirm the broader applicability of these findings. The technical specifications of the Tangerang radar used in this study are detailed in Table 2-1, and the associated radar beam spreading characteristics are visualized in Figure 2-1. Additionally, a beam blockage analysis (Figure 2-2) indicates that only a very small fraction of the radar coverage area is obstructed by terrain features, confirming the radar data's suitability and representativeness for precipitation nowcasting purposes.

Table 2-1: Tangerang weather radar specification and scanning strategy

Parameter	Value
Polarisation	Single
Installation year	2009
Latitude	-6.1669° S
Longitude	106.6502° E
Altitude	10 m
Tower height	20 m
Frequency	5.6 GHz
Beam width	1°
Pulse width range	0.5 - 2.0 µs
Pulse width used	0.8 µs
PRF Used	600 Hz
Staggering	¾
Range step	250 m
Range maximum	220 km
Number of sweeps	9
Nyquist Velocity	24.1 m/s

The Spectral Prognosis (S-PROG) algorithm, introduced by Seed (2003), is a deterministic radar-based nowcasting method that utilizes the scale-dependent predictability of rainfall fields. The algorithm decomposes the log-transformed rainfall field into multiple spatial scales using a Fourier transform. Each spectral component is then evolved forward in time using a second-order autoregressive model [AR(2)]:

$$X_l(t) = \phi_{i,l}\, X_l(t-1)\, \phi_{2,l}\, X_l(t-2) + \epsilon_l(t)$$

where $X_l(t)$ is the spectral coefficient at time t, $\phi_{1,l}, \phi_{2,l}$ are AR coefficients, and $\epsilon_l(t)$ is a white noise term. After temporal evolution, an inverse Fourier transform is applied to reconstruct the forecast in the spatial domain. To correct bias in rainfall intensity, a cumulative distribution function (CDF) matching is performed between the forecast and the most recent observation. S-PROG assumes that large-scale features are more predictable than small-scale ones. As such, small-scale variability is filtered out during evolution, resulting in smoother fields at longer lead times. While this approach improves skill for lead times around 30–60 minutes, it tends to blur convective peaks, underestimating intense rainfall and overestimating the spatial coverage of rainfall. Despite these limitations, S-PROG remains widely used due to its simplicity and overall better performance compared to pure extrapolation (Seed, 2003; Ritvanen et al., 2025).

The Autoregressive Nowcasting using VIL (ANVIL) algorithm enhances deterministic radar nowcasting by accounting for the local growth and decay of precipitation. ANVIL applies a multiscale decomposition to the rain field and models the temporal evolution of each scale using an Autoregressive Integrated (ARI) model. Unlike S-PROG, which evolves the signal directly, ANVIL evolves the temporal derivative of each scale component, allowing it to better capture local intensity changes. The ARI

model used is typically of order (p=2, d=1), given by:

$$\nabla X_l(t) = \phi_{i,l}\,\nabla X_l(t-1)\,\phi_{2,l}\,\nabla X_l(t-2) + \epsilon_l(t)$$

where $\nabla X_l(t)$ is the temporal derivative of the spectral coefficient at scale l, ϕ are AR coefficients, and ϵ is a stochastic error. ANVIL does not require post-processing corrections because it preserves both small- and large-scale features more faithfully. This makes ANVIL well-suited for convective nowcasting over short to medium lead times (Pulkkinen et al., 2019; Pulkkinen et al., 2021).

LINDA, or the Lagrangian Integrated Nowcasting using a Difference-based Autoregressive model, was introduced by Pulkkinen et al. (2021) to improve the handling of convective rain systems. The method operates in a Lagrangian frame, where the rain fields are advected, and the difference between observed and advected rainfall fields is computed and forecasted. A convolution-based smoothing is applied to filter out unpredictable scales before and after the update. Its core formulation is:

$$R(t+\Delta t) = R(t) + \sum_{i=1}^{p} \phi_i[R(t-i\Delta t) - R(t-(i+1)\Delta t] + \epsilon(t) \quad [2]$$

where R is the rain rate field, ϕ_i are autoregressive coefficients, and ϵ is white noise. LINDA improves nowcasting performance by capturing both smooth and convective features while reducing over-smoothing, making it effective even for small, fast-growing convective cells (Pulkkinen et al., 2021).

The Extrapolation method assumes that the observed rainfall field moves at a constant velocity, without change in intensity or structure. Motion vectors were estimated using the Lucas–Kanade optical flow algorithm based on a pair of recent radar images. The forecasted field at lead time t+Δt is generated by translating the latest rain field backward by the estimated displacement vector::

$$R(t+\Delta t, x) = R(t, x - v\Delta t)$$

where R is the rain rate at location x, v is the motion vector derived from radar image pairs (e.g., using optical flow), and Δt is the lead time. While computationally efficient and accurate for very short lead times (<15 minutes), this method lacks any capacity to predict growth or decay of rainfall, limiting its usefulness in rapidly evolving convective conditions (Germann and Zawadzki, 2002).

To evaluate the performance of all four nowcasting algorithms, the forecasts were verified using a combination of categorical and continuous statistical metrics. Three primary verification techniques were applied: the Receiver Operating Characteristic (ROC) curve, the Taylor diagram, and the Target diagram. The ROC curve was used to assess the binary detection capability of the forecasts for rainfall exceeding a threshold of 1 mm/h. Each forecast field was binarized and compared with the observed field at matching lead times, allowing the calculation of the probability of detection (POD), false alarm rate (FAR), and area under the curve (AUC). ROC analysis is widely adopted in meteorological forecast evaluation for its ability to capture discrimination skill (Mason, 1982, Imhoff et al., 2020).

In addition to categorical metrics, the Taylor diagram was employed to quantify the spatial correspondence between the predicted and observed rainfall fields (Shi et al, 2015; Wang et al., 2023). This diagram simultaneously depicts the correlation coefficient, normalized standard deviation, and centered root mean square error (RMSE), providing a compact visualization of forecast skill in terms of pattern similarity and amplitude accuracy (Taylor, 2001).

Lastly, the Target diagram was used to assess the normalized bias and RMSE components of the forecasts. By plotting the bias and the centered RMSE as Cartesian coordinates, this diagram enables the identification of systematic over- or under-prediction, as well as the spread of forecast errors relative to the observations. Target diagrams have been increasingly used for hydrometeorological verification to visualize forecast quality in a unified

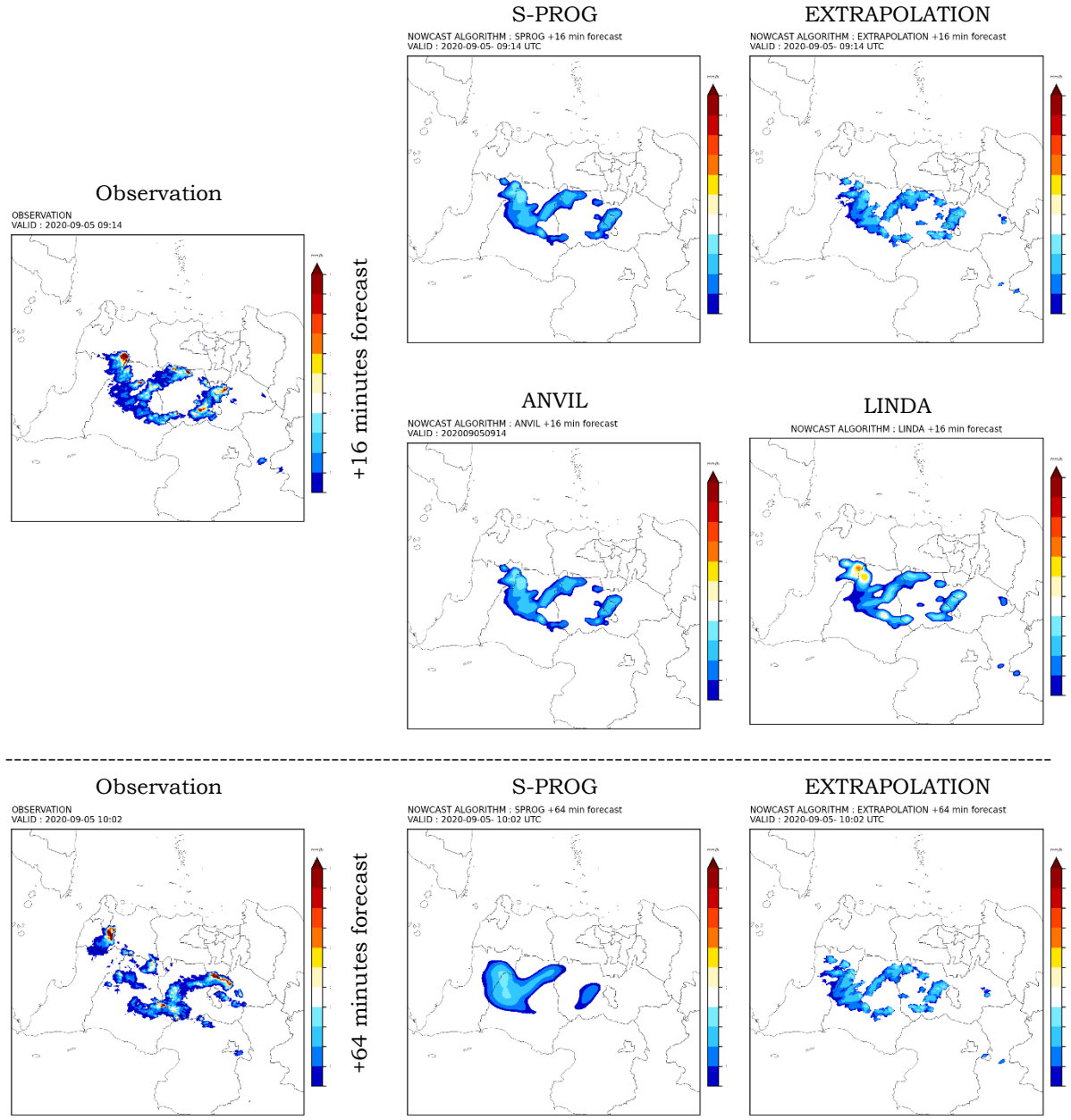
framework (Jolliffe and Stephenson, 2012; Ebert, 2008). The combination of these three verification approaches allows for a comprehensive evaluation of both the classification skill and structural accuracy of the nowcasts across multiple lead times.

All nowcasting algorithms were implemented using the open-source pysteps library (Pulkkinen et al., 2019). The verification metrics, including ROC curves, Taylor diagrams, and target diagrams, were computed using python with additional support from libraries such as NumPy, SciPy, scikit-learn, and Matplotlib.

3 RESULTS AND DISCUSSION

3.1. Spatial Evaluation of Rainfall Forecasts

The spatial performance of each nowcasting algorithm is qualitatively evaluated by comparing radar-derived rainfall forecasts with observed fields across a range of lead times. The evaluation emphasizes several key aspects, including rainfall coverage, intensity distribution, preservation of convective structures, and positional accuracy. For clarity, the analysis is divided into two temporal categories: forecasts up to 64 minutes (short-range) and those between 72-96 minutes.



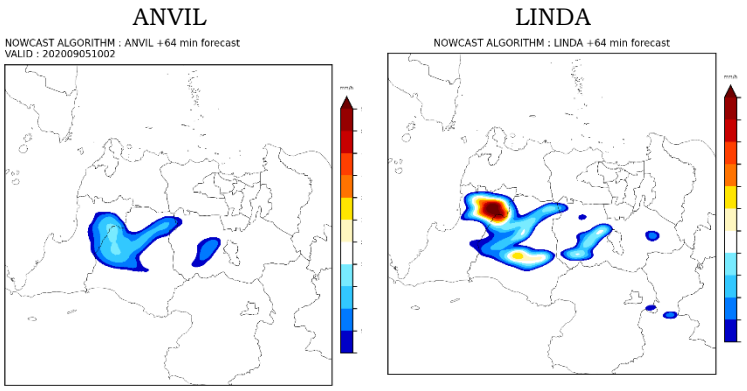
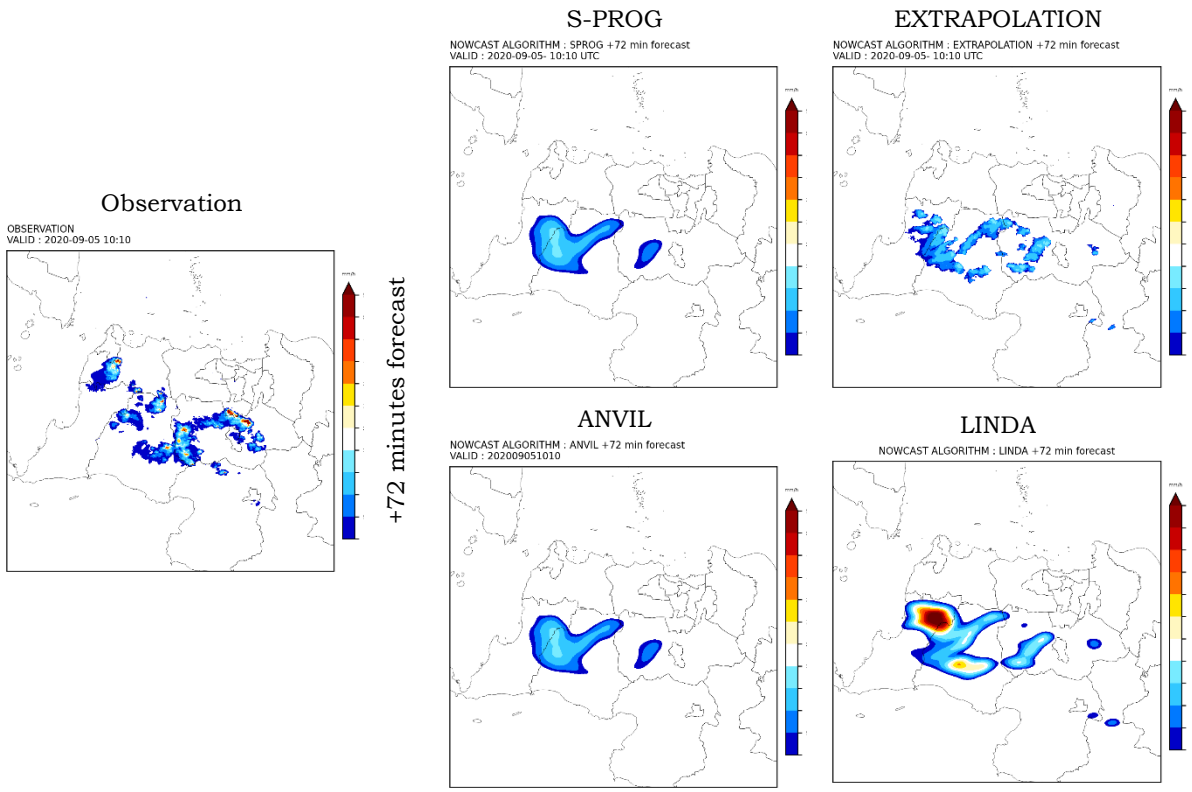


Figure 3-1: Spatial comparison of rainfall fields between radar observations and nowcast outputs from four algorithms at 16-64 minutes lead times. The first column shows the radar-observed rainfall at each time step, while the second and third columns show the corresponding forecasts generated by SPROG, Extrapolation, ANVIL, and LINDA. The panels highlight the ability of each method to capture rainfall intensity, areal extent, and storm displacement through time.



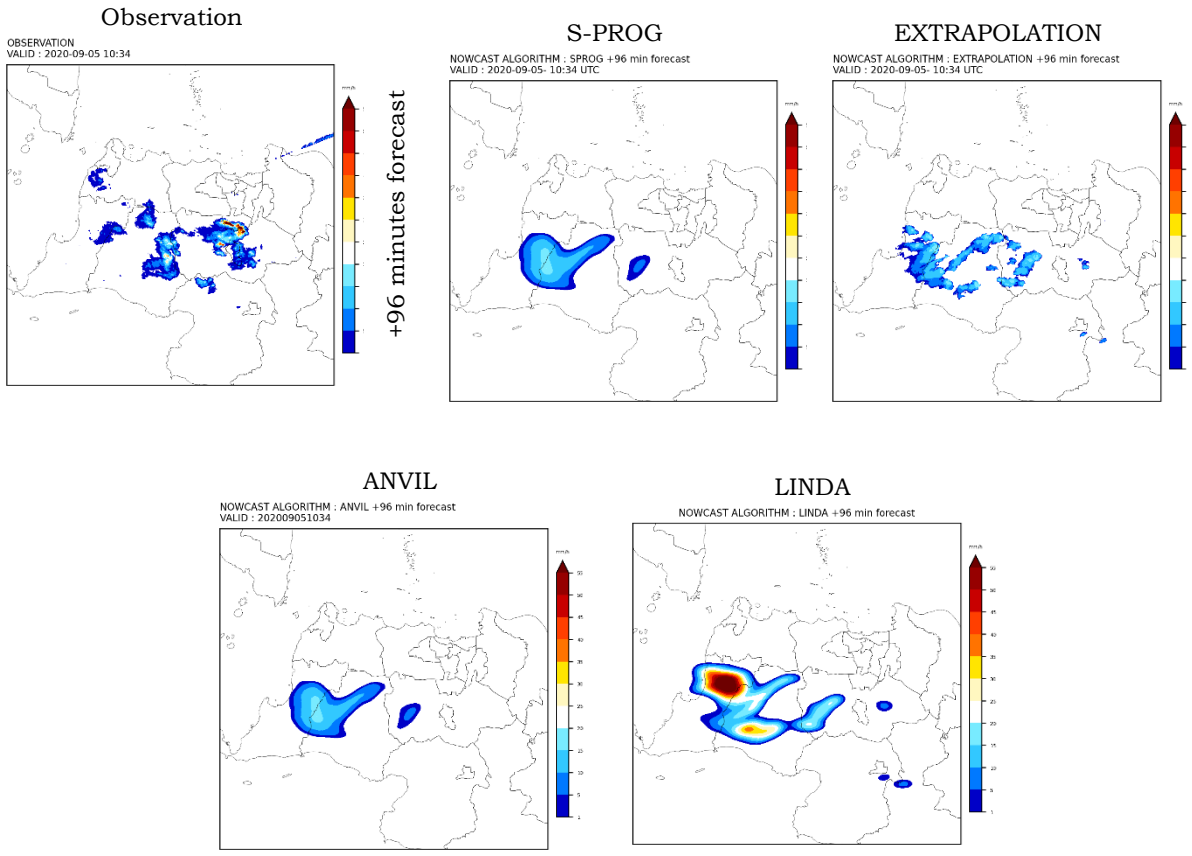


Figure 3-2: Forecast-observation comparison at 72-96 minutes lead times, following the same panel arrangement as Figure 2-1. Radar observations are shown in the first column, with subsequent columns presenting outputs from SPROG, Extrapolation, ANVIL, and LINDA

Within the first 60 minutes of lead time (Figure 3-1), all nowcasting methods reasonably capture the spatial extent and movement of the main convective systems. The Extrapolation method shows a high degree of spatial continuity with the observed field, preserving detailed rainfall features due to its conservative, non-evolving approach. However, it lacks any representation of convective growth or dissipation, which may lead to overestimation of system longevity or underestimation of newly developing cells.

The SPROG algorithm, which incorporates a scale decomposition strategy, demonstrates modest ability to simulate the decay of small-scale structures while maintaining the dominant mesoscale pattern. However, it slightly underestimates the rainfall intensity, as seen by the reduced presence of higher rainfall classes in comparison to LINDA and ANVIL.

ANVIL, which accounts for convective lifecycle stages through temporal filtering and statistical growth-decay modeling, produces smoother rainfall patterns with slightly dampened intensity peaks. It moderately follows the storm motion but tends to oversmooth smaller cells, potentially compromising short-range hazard prediction.

In contrast, LINDA consistently captured higher intensity cores (>35 mm/h) that closely matched observations, while other methods generally underestimated rainfall intensity. This suggests that LINDA captures convective growth more aggressively, though with slight overprediction in some sectors. Its

spatial distribution aligns well with observations, especially in the western domain, indicating better skill in preserving both intensity and morphology.

At longer lead times (72–96 minutes) as represented in Figure 3-2, performance discrepancies among the methods become more pronounced. Extrapolation shows visible displacement errors and fails to reflect the natural decay and evolution of convective systems. Rainfall areas appear overly static, resulting in increasing divergence from the true storm field over time.

SPROG exhibits improved realism compared to Extrapolation, as it progressively suppresses smaller-scale noise and attenuates the rainfall field. However, by 80 minutes, the rain areas become excessively smoothed and underestimated, with limited preservation of high-intensity cores.

ANVIL maintains structural consistency up to +64 minutes but begins to over-dissipate the system at longer lead times. Its performance appears better than SPROG in representing mesoscale convective structures, though still underrepresenting localized extremes.

On the other hand, LINDA retains significant convective features even at +80 minutes, including strong rainfall cores exceeding 40 mm/h. While this reflects LINDA's emphasis on storm growth, it may also result in false alarms if convective initiation is not verified by observation. Nevertheless, among all models, LINDA sustains the most coherent depiction of convective intensity and areal coverage over 2 hours.

3.2. Categorical Verification Using ROC Analysis

While qualitative spatial comparisons offer valuable insights into the structural realism and storm evolution captured by each algorithm, they remain inherently subjective and limited in statistical rigor. To complement the visual assessment, a quantitative evaluation using Receiver Operating Characteristic (ROC) analysis was conducted. This method enables objective comparison of detection capabilities by evaluating the trade-off between hit rates and false alarms for binary rainfall thresholds. The resulting Area Under the Curve (AUC) values provide a concise metric to summarize the forecast skill at different lead times across all models.

At very short lead times (+8 and +16 minutes), all algorithms demonstrate excellent discrimination capability, with AUC values above 0.83. SPROG shows the highest performance in these initial forecasts (AUC = 0.96 and 0.95), followed closely by LINDA and ANVIL, while Extrapolation also performs strongly (AUC > 0.9 at +8 min).

As lead time increases, a gradual decline in AUC values is observed across all methods. At +24 and +32 minutes, SPROG continues to outperform others with AUCs of 0.92 and 0.90 respectively, indicating its robustness in maintaining probabilistic accuracy. LINDA follows with AUCs of 0.83 and 0.80, showing consistent early skill. In contrast, ANVIL and Extrapolation begin to diverge, with both dropping below 0.80 beyond +32 minutes.

Forecast skill diminishes further at medium-range lead times (+40 to +64 minutes). SPROG remains the most resilient, with AUCs gradually decreasing from 0.87 to 0.70. LINDA and ANVIL show a steeper performance drop, with LINDA declining to 0.63 and ANVIL to 0.60 by +64 minutes. Extrapolation consistently yields the lowest AUC values beyond 32 minutes, reflecting its limited ability to account for storm evolution.

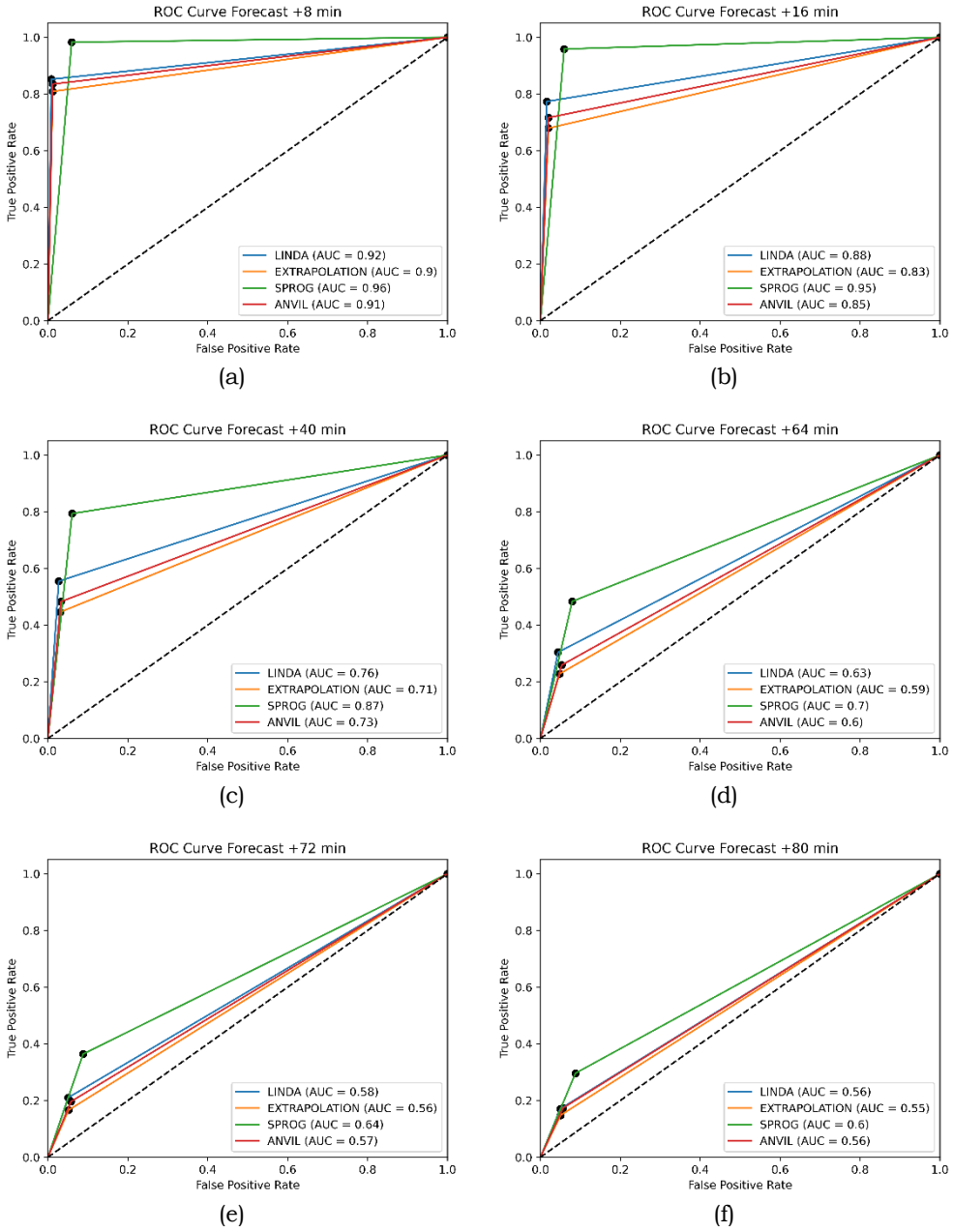


Figure 3-3: Selected ROC curves at forecast lead times of +8, +16, +40, +64, +72 and +80 minutes (a-f) for four nowcasting algorithms.

By +72 and +80 minutes, all models exhibit notable degradation in discriminative skill. SPROG still leads marginally (AUC = 0.64 and 0.60), whereas other algorithms—including LINDA and ANVIL—approach values near the no-skill threshold (AUC ≈ 0.55–0.58), indicating limited forecast value at this extended range.

These results confirm that while SPROG maintains the highest ROC-based performance throughout the forecast window, LINDA offers competitive skill in the short range and gradually degrades with time. Extrapolation and ANVIL underperform particularly at medium to long lead times, highlighting the challenges in preserving event structure and growth behaviour without explicit storm evolution modelling.

3.3. Statistical Performance Evaluation using Target and Taylor Diagrams

To complement the spatial and categorical assessments, statistical verification was conducted using Target Diagrams and Taylor Diagrams across varying lead times (Figure 3-4). While

SPROG demonstrated the highest ROC skill at short lead times, the Target and Taylor diagrams reveal a different aspect of forecast performance related to structural preservation and error characteristics.

During the initial short lead times (up to +24 minutes), all nowcasting methods shows generally strong statistical agreement with observations. In the Target Diagrams, LINDA, ANVIL, and Extrapolation cluster closely around the origin, indicating minimal bias and uRMSD errors. This suggests that these methods are capable of generating rainfall estimates with both low systematic error and acceptable random deviation at early stages.

Taylor Diagrams further support these findings. LINDA and ANVIL achieve relatively high correlation coefficients (≥ 0.8) and preserve the standard deviation structure of the observed rainfall fields. Extrapolation also demonstrates solid performance with low RMSD and reasonable spatial correlation, albeit slightly lower than LINDA. SPROG, however, begins to show signs of overdispersion. Its location in both diagrams is noticeably farther from the reference point, revealing an early tendency to overestimate variability and a lower pattern match with observations.

As the lead time increases between 32-80 minutes forecast, distinctions among the methods become more pronounced. The Target Diagrams indicate increasing bias and uRMSD, particularly for SPROG and ANVIL. SPROG's points move outward significantly, reflecting inflated variability and diminished accuracy, with correlation coefficients dropping below 0.5 by +56 to +80 minutes in the Taylor Diagrams.

Extrapolation retains low uRMSD and bias throughout the forecast window, suggesting stable but non-evolving output. However, its correlation coefficient declines steadily, indicating growing mismatch with the actual rainfall distribution as storms evolve spatially and temporally.

LINDA stands out for its consistent performance. In both diagrams, it remains relatively close to the reference across all lead times, maintaining favorable bias, uRMSD, and correlation. Its ability to preserve rainfall intensity structure and track convective features even beyond one hour confirms its robustness for medium-range nowcasting.

The statistical verification using Taylor and Target diagrams thus reinforces that although SPROG initially excels in categorical skill (ROC), LINDA offers superior consistency in structural accuracy and error control across lead times. This finding aligns with Shi et al. (2015), who emphasized the importance of maintaining spatiotemporal coherence in nowcasting systems like ConvLSTM, and with Ravuri et al. (2021), who demonstrated the operational advantages of models achieving high pattern fidelity. While LINDA remains deterministic rather than learning-based, its physical formulation ensures performance comparable to more complex machine learning frameworks under tropical convective conditions.

Similarly, the strong statistical behavior of LINDA resonates with the conclusions of Ravuri et al. (2021), whose deep generative radar model achieved high accuracy and expert-rated usability in short-term precipitation forecasts. While LINDA is a deterministic model and not learning-based, its robustness in both bias and correlation metrics suggests that properly calibrated physical algorithms can still provide competitive performance. These comparisons affirm that LINDA’s forecast consistency, as reflected in multimetric diagrams, is on par with trends observed in more advanced modeling frameworks.

+24 minutes forecast

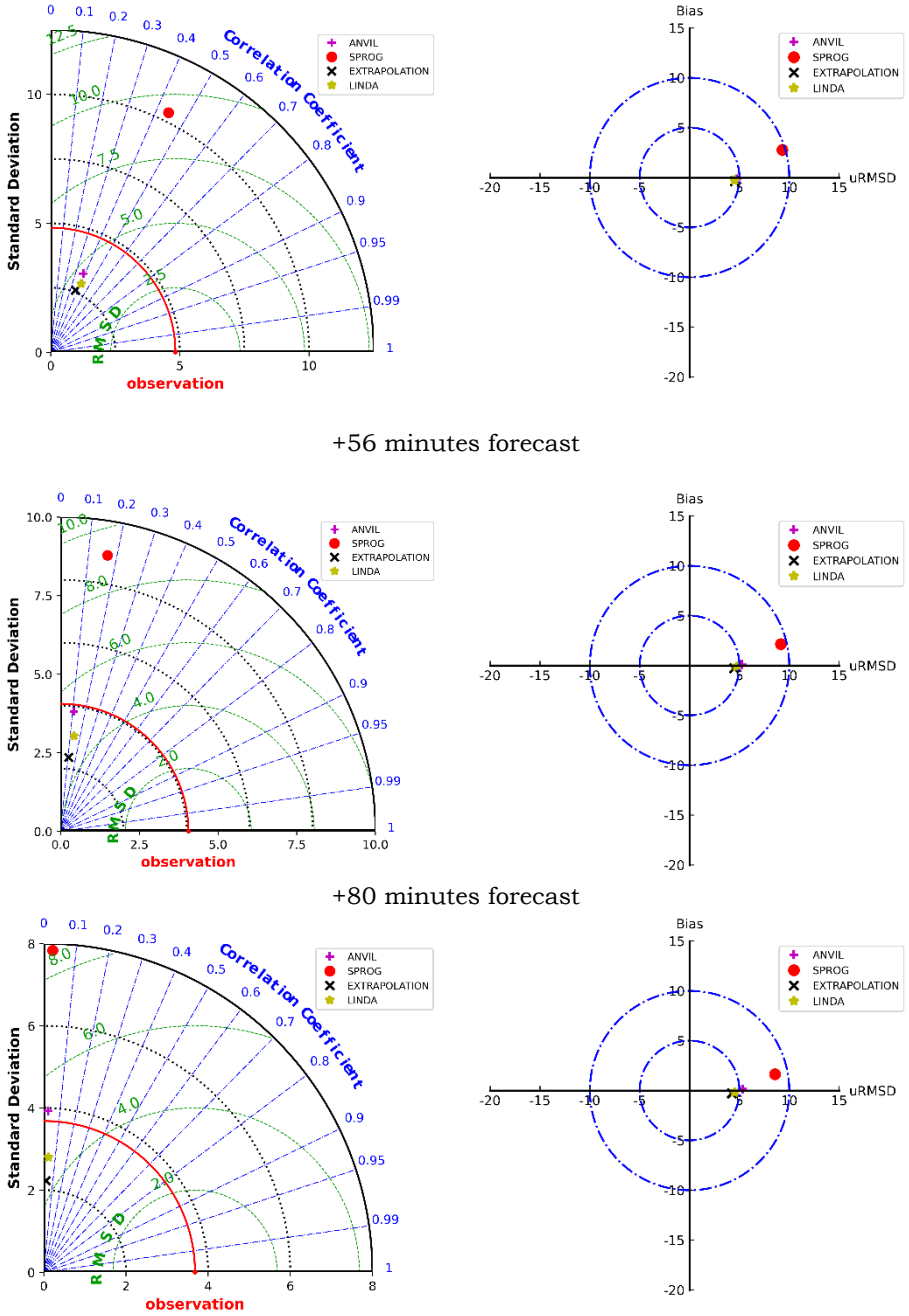


Figure 3-4: Selected Taylor and Target Diagram at forecast lead times of +24, +56 and +80 minutes for four nowcasting algorithms.

3.3. Overall Summary of Nowcasting Algorithm Performance

The comparative evaluation of four deterministic nowcasting algorithms—SPROG, Extrapolation, ANVIL, and LINDA—revealed distinct strengths and limitations across different verification approaches and forecast lead times.

Spatial analysis showed that all methods reasonably captured storm motion at short lead times, but LINDA more effectively preserved rainfall intensity structures and convective patterns, especially beyond 60 minutes. Extrapolation maintained spatial continuity but failed to account for storm evolution, while SPROG and ANVIL exhibited progressive smoothing and intensity underestimation at longer lead times.

ROC analysis confirmed that SPROG achieved the highest categorical discrimination skill at short lead times, with AUC values up to 0.96. However, its performance declined more rapidly over time compared to LINDA, which retained higher skill at extended lead times.

Taylor and Target diagram evaluations emphasized LINDA's superiority in maintaining statistical consistency across the forecast window. LINDA achieved the highest correlations, lowest bias, and most stable uRMSD values up to +80 minutes, reinforcing its robustness for medium-range nowcasting applications.

Overall, the results suggest that while SPROG excels in initial rain detection, LINDA offers more reliable performance in maintaining structural accuracy and minimizing forecast errors over time, making it particularly well-suited for operational nowcasting in tropical convective environments.

4 CONCLUSIONS

This study assessed the performance of four deterministic radar-based nowcasting algorithms—LINDA, SPROG, ANVIL, and Extrapolation—using radar observations from a single convective rainfall event over Tangerang. A comprehensive evaluation was conducted through spatial inspection, categorical verification (ROC analysis), and multimetric statistical diagnostics using Taylor and Target diagrams across forecast lead times of up to 80 minutes.

Qualitative spatial analysis revealed that while all algorithms were able to reproduce the general motion of convective systems at short lead times, their ability to maintain rainfall intensity and structural realism varied significantly. LINDA emerged as the most robust model in preserving convective features and spatial coherence beyond 60 minutes, whereas SPROG and ANVIL exhibited progressive smoothing and underestimation of peak rainfall. Extrapolation maintained spatial continuity but failed to account for storm evolution.

The ROC analysis further demonstrated that all methods performed well during the first 30 minutes, with SPROG and LINDA achieving the highest discrimination skill (AUC > 0.90). However, as lead time increased, SPROG and Extrapolation experienced notable skill degradation, while LINDA retained higher AUC values through to 80 minutes, reflecting a better balance between detection and false alarms.

Taylor and Target diagram evaluations reinforced LINDA’s strength in maintaining structural accuracy, with higher correlations, lower bias, and stable uRMSD across lead times. These findings are consistent with previous studies (e.g., Shi et al., 2015; Ravuri et al., 2021) that emphasize the importance of capturing spatiotemporal dynamics for skillful precipitation nowcasting.

Overall, within the context of this case study, LINDA demonstrated the most balanced and robust performance across multiple verification dimensions. However, these findings are based on a single convective event, and the generalizability to broader tropical conditions remains uncertain. Future work should include a wider range of storm types, multiple radar sites, and a broader verification framework—including multiple rainfall thresholds—to more comprehensively assess algorithm robustness and operational applicability.

ACKNOWLEDGEMENTS

The data of this research is fully supported by BMKG Weather Radar Data Management Subdivision. The paper was improved by the helpful suggestion of Dr. Hidde Leijnse from KNMI Netherland. The author wishes to thank all persons who help this research.

REFERENCES

Ali, A., Umam, I. H., Leijnse, H., & Sa'adah, U. (2021). Preliminary Study of A Radio Frequency Interference Filter for Non-Polarimetric C-Band Weather Radar In Indonesia (Case Study: Tangerang Weather Radar). International Journal of Remote Sensing and Earth Sciences (IJReSES), 18(2), 189-202.

Ali, A., Umam, I. H., Heningtyas, H., Charolidya, R., Sanditya, B., Cempaka, A. P., ... & Kiki, D. (2022).

Pengembangan Sistem Peringatan Dini Cuaca Ekstrem Terintegrasi Berbasis Y-Model Webgis Development Methodology (Y-WDM). Jurnal Geografi, Edukasi Dan Lingkungan (Jgel), 6(2), 87-100.

Ebert, E. E. (2008). Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. Meteorological Applications, 15(1), 51–64.

Germann, U., & Zawadzki, I. (2002). Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. Monthly Weather Review, 130(12), 2859–2873.

Imhoff, R. O., Brauer, C. C., Overeem, A., Weerts, A. H., & Uijlenhoet, R. (2020). Spatial and temporal evaluation of radar rainfall nowcasting techniques on 1,533 events. Water Resources Research, 56(8), e2019WR026723.

Jolliffe, I. T., & Stephenson, D. B. (2012). Forecast Verification: A Practitioner's Guide in Atmospheric Science (2nd ed.). Wiley.

Mason, I. (1982). A model for assessment of weather forecasts. Australian Meteorological Magazine, 30, 291–303.

Pulkkinen, S., Nerini, D., Zahraei, A., & Seed, A. (2019). *Pysteps: An open-source Python library for probabilistic precipitation nowcasting*. Geoscientific Model Development, 12, 4185–4219.

Pulkkinen, S., Nerini, D., & Seed, A. (2021). *A difference-based autoregressive model for radar-based nowcasting of convective rainfall*. Journal of Atmospheric and Oceanic Technology, 38(12), 2067–2084.

Ravuri, S., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878), 672–677.

Ritvanen, J., Pulkkinen, S., Moisseev, D., & Nerini, D. (2025). Cell-tracking-based framework for assessing nowcasting model skill in reproducing growth and decay of convective rainfall. Geoscientific Model Development, 18(3), 1851–1870.

Seed, A. W. (2003). A dynamic and spatial scaling approach to advection forecasting. Journal of Applied Meteorology, 42(3), 381-388.

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems* (pp. 802–810).

Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., ... & Pinto, J. (2014). Use of NWP for nowcasting convective precipitation: Recent progress and challenges. Bulletin of the American Meteorological Society, 95(3), 409-426.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres, 106(D7), 7183–7192.

Wang, H., Yin, J., Deng, H., & Xue, M. (2023). Comparison of Simulated Warm-Rain Microphysical Processes in a Record-Breaking Nocturnal Rainfall Event in South China Using Polarimetric Radar Observations. *Journal of Geophysical Research: Atmospheres*, 128(23)