ANALYZING THE PERFORMANCE OF DIFFERENT PARAMETER SETTINGS WITH THE ENSEMBLE NOWCASTING OF TROPICAL CYCLONE PRECIPITATION

Steven Rearden[1], Steven Martinaitis[2], Jackson Anthony[2], and Dean Meyer[2]

[1]National Weather Center Research Experiences for Undergraduates Program
Norman, Oklahoma
[2]OU/CIWRO & NOAA NSSL
Norman, Oklahoma

ABSTRACT

Flash flooding is a dangerous weather hazard that can result in substantial property damage and loss of life if the public is not prepared. Because of this, continuous advancements dedicated to extending the lead times of flash flood warnings is of great importance. To allow for new progress in this, a new precipitation nowcasting scheme which uses data from the MRMS system was tested in the STEPS framework. This study focused on nowcasting precipitation with tropical cyclones. Hurricanes Ian (2022) and Henri (2021) were chosen as case studies due to their differing precipitation structures. Varying parameter value combinations were used to analyze nowcast performance. These parameters include the advection tracking threshold, number of ensemble members, and seed value for ensemble perturbations. Results indicate that differences in advection tracking threshold caused the most significant changes in performance, while differences in number of ensemble members and seed value cause generally insignificant performance differences. Performance when the seed value was randomized was also found to have a low variability, leading to the conclusion that the particular seed value is generally insignificant in nowcast performance. Analysis of both cases revealed that the best performance for each tropical cyclone utilized different advection tracking thresholds. This underscores the challenge of finding the right parameters to use within the MRMS system.

1. INTRODUCTION

Tropical cyclones (TCs) contain a wide range of hazards when impacting land. One of the most deadly hazards associated with them is flash flooding (Rappaport 2014), which often occurs when a band of intense rain remains over the same area for an extended period of time. Flash floods are typically characterized by quickly rising, fast moving, and debris filled rivers/streams which can pose a severe threat to infrastructure, property, and life (Creutin et al. 2013). When flash floods are imminent or occurring, flash flood warnings are issued by the National Weather Service for communication of the threat to the public. There have been significant advancements in lengthening the lead time for these alerts, especially between 1990 and 2010, when lead times increased from roughly 16 minutes to over 60 minutes; however, this progress has become stagnant over the past decade (Martinaitis et al. 2023). A new, innovative way of monitoring flash flood potential must be developed to further increase lead times to allow the public more time to protect themselves and their property.

A system of accurate nowcasts for impactful TC events would be beneficial for improving the accuracy of flash flood forecasting and allowing for longer lead times for flash flood warnings. Nowcasts refer to high resolution (< 5 km), short-term forecasts of generally two hours or less. They typically use existing radar data and extrapolate it by following an estimated vector field (Prudden et al. 2020). This method is known as Lagrangian persistence and is an effective nowcasting technique in TCs due to the frequent persistence of large scale features over short periods of time (Bowler et al. 2006).

The Short-Term Ensemble Prediction System (STEPS) is a versatile nowcasting framework that allows for the simulation of an ensemble of future potential precipitation rates

[1] *Corresponding author address:* Steven Rearden, University of North Carolina Asheville, srearden@unca.edu

TABLE. 1. Start and end times with coordinate boundaries for each TC case study.

	Ian	Henri
Date	28-29 September 2022	22 August 2021
Start Time (UTC)	1900	1700
End Time (UTC)	0100	2200
Northwest Corner	29.5°N, -84.5°W	44.0°N, -76.0°W
Southeast Corner	24.5°N, -79.5°W	39.0°N, -71.0°W

(Bowler et al. 2006). The main foundation of STEPS is the use of the Lucas Kanade (LK) optical flow method, which estimates an advection field through the use of image sequences (Sharmin and Brad 2012). This is done through calculating the movement of different colored pixels using a sequence of precipitation rate observations. A randomization of small scale features and advection uncertainty is then implemented, along with a second-order autoregressive (AR-2) model, which describes uncertainty in precipitation evolution through the calculation of AR-2 parameters (Bowler et al. 2006). The determined uncertainties in advection and precipitation evolution are then represented through the differences and overall spread of the individual ensemble members, while also adding a noise component to replicate the unpredictability of small-scale features.

The purpose of this study is to use STEPS to develop the most accurate precipitation nowcast for TCs through testing a wide variety of unique nowcasting iterations and evaluating their statistical performance. Two TC cases with differing precipitation structures and intensities were analyzed and compared to determine if these differences require unique nowcast settings for maximum predictability. These analyses will contribute to a broader flash flood forecasting framework that is in the process of development for the purpose of improving flash flood prediction and warning lead times.

2. DATA

The Multi-Radar Multi-Sensor (MRMS) system developed at the National Severe Storms Laboratory (NSSL) provides the radar-derived data used in this nowcast study. The MRMS system combines existing radar data with numerical weather prediction data, atmospheric environmental data, satellite data, lightning observations, and rain gauge observations (Zhang et al. 2016). All of this information allows for a wide range of products that are used for the observation and forecasting of severe weather and precipitation. This study utilizes the high spatiotemporal resolution (1-km, 2-min) MRMS instantaneous precipitation rates for the initiation of STEPS nowcasts and verification. These rates were derived from a synthetic dual-polarization scheme (Ryzhkov et al. 2022; Zhang et al. 2020) using mosaicked radar data (Qi and Zhang 2017) with an evaporation correction scheme (Martinaitis et al. 2018).

The cases chosen for this study are Hurricane Ian (2022) and Hurricane Henri (2021). Table 1 displays the chosen times and coordinate boundaries for analysis. Each case reflects significant differences in precipitation structure. Ian maintained a very organized structure as it approached land, characteristic of a strong tropical cyclone; however, Henri experienced extratropical transition (Jones et al. 2003), which caused it to have a more frontal-based precipitation structure as it approached land.

3. METHODOLOGY

The MRMS instantaneous precipitation rates were utilized in the STEPS nowcasting scheme through the use of pySTEPS, an open source Python package used for running STEPS nowcasts and statistically analyzing their performance (Pulkkinen et al. 2019). The study examines different nowcast outputs by evaluating different parameters used to run the STEPS scheme (Table 2). The product resolution refers to the resolution of the gridded input MRMS rain rates. Advection tracking threshold refers to the minimum rain rate used in the LK optical flow method. Ensemble members refers to the number of unique nowcast iterations used to create the ensemble mean. The seed refers to a variable that produces a unique and constant set of numbers which gives a pseudo-randomization of small-

TABLE. 2. Case parameters for testing different parameter combinations with the STEPS methodology.

Case Parameter	Parameter Values Tested			
Product Resolution	1-km			
Advection Tracking Threshold	0.5 mm h^{-1}	10 mm h^{-1}	20 mm h^{-1}	40 mm h^{-1}
Ensemble Members	10	20	30	
Seed	0	24	42	

scale features for each ensemble member. Each unique combination of these variable values were tested and analyzed for Ian and Henri.

Each iteration will produce a unique set of 60 minute rainfall rate simulations at two-minute resolution. An ensemble mean rain rate nowcast is used as the final output. The performance of each variable was examined across all initiation hours for both cases. An additional test was implemented to test the variability of performance statistics when the seed value is randomly selected. This was done through setting seed equal to "None", which causes the system to randomly choose a seed value between 0 and 100,000,000. This was done for the Ian case for the 2100 and 0000 UTC run times, using a 40 mm h^{-1} advection tracking threshold while changing the number of ensemble members. Each identical input combination was repeated 10 times for an analysis of the differing performance statistic values.

Simulated forecast rainfall rates were compared to observational data to produce a variety of statistical output.

3.1 *Fractions Skill Score*

The fractions skill score (FSS) describes the fraction of forecast and observed rainfall rate grid cells that surpass a defined threshold rate (Roberts and Lean 2008). The FSS can consider a neighborhood of grid cells; however, this study uses the neighborhood of one grid point, and the resulting fractions are binary ones and zeros for each grid cell analysis. The mean squared error (MSE) between observed and predicted binary grids is then calculated, and the FSS is calculated through

$$FSS = 1 - \frac{MSE}{MSE_{ref}}, \quad (1)$$

where MSE_{ref} refers to the MSE for a random forecast. Scores range from 0 (no skill) to 1 (perfect forecast), with scores of at least 0.5 considered to be skillful. A threshold of 1.0 mm h^{-1} was used for FSS calculations, and the FSS at the end of the 60-minute forecast period was evaluated.

3.2 *Area Under ROC Curve*

The relative operating characteristic (ROC) curve displays the probability of detection of a threshold rain rate compared to a corresponding false alarm rate at various classification thresholds. The area under the ROC curve (AURC) can then be used to determine forecast skill (Marzban 2004). Perfect forecasts have an area of 1, while random forecasts have an area of 0.5. These values were output at the end of each 60 minute nowcast, and the evaluated rain rate threshold was 0.1 mm h^{-1}.

3.3 *Pearson Correlation Coefficient*

The Pearson correlation coefficient (CC) displays the degree to which the forecast is related to observations. It is given by the equation,

$$r = \frac{\sum_{i=1}^{n} (x_i - \underline{x})(y_i - \underline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \underline{x})^2 \sum_{i=1}^{n} (y_i - \underline{y})^2}}, \quad (2)$$

where n refers to the number of grid points, x refers to observed rain rates, and y refers to forecasted rain rates. These values range from -1 (strong negative correlation) to 1 (strong positive correlation), while a CC of 0 indicates no correlation. CCs were collected for each two-minute interval across the entire 60-minute nowcast period.

3.4 *Reliability Diagram*

Reliability diagrams display the observed relative frequency of an event versus the forecast probability. This was done for every grid point, with forecast probabilities being the percentage that a

rain rate threshold (0.1 mm h^{-1}) was exceeded. This will indicate whether precipitation is over-forecasted (line is below 1:1 line), under-forecasted (line is above the 1:1 line), or is forecasted reliably. Each reliability diagram is valid for the end of each 60-minute nowcast.

4. RESULTS

Subtle differences in nowcast performance were observed with changes in number of ensemble members and seed value. Fig. 1 displays differences in averaged 60-minute performance statistics with these changes for Ian, and Fig. 2 displays the same for Henri. Figs. 3-4 display the changes in CC for every lead time in the simulations for Ian and Henri, respectively. The spread between these different parameter value choices was observed to be small throughout the entire nowcast for both cases, though patterns do exist. AURC and FSS values increased by slightly more than 0.01 when the number of ensemble members was changed from 10 to 30 in the Ian case, while these spreads were lower for Henri. CC ranges were observed to be 0.036 for Ian and 0.021 for Henri with this same ensemble member parameter. Changing the seed value revealed a less concrete pattern, though similar spreads were observed for both cases. For Ian, 24 seed performed best using AURC and FSS, and was virtually tied with seed=0 using CC. However, seed=0 performed best using all performance statistics in the Henri case.

Greater overall performance differences were observed in response to changes in the advection tracking rate for both Ian and Henri for all performance statistics (Fig. 5). Greater spreads in CC also developed, especially in Ian (Fig. 6). Differences between the best and worst CC values were 0.181 for Ian, and 0.064 for Henri. These results indicate a noteworthy change in the best performing advection tracking threshold for each case. Figs. 5 and 6 indicate that 40 mm h^{-1} had the

FIG. 1. Ian performance differences in response to changes in number of ensemble members (left) and seed value (right). Displayed are the AURC (blue), FSS (orange), and CC (green).

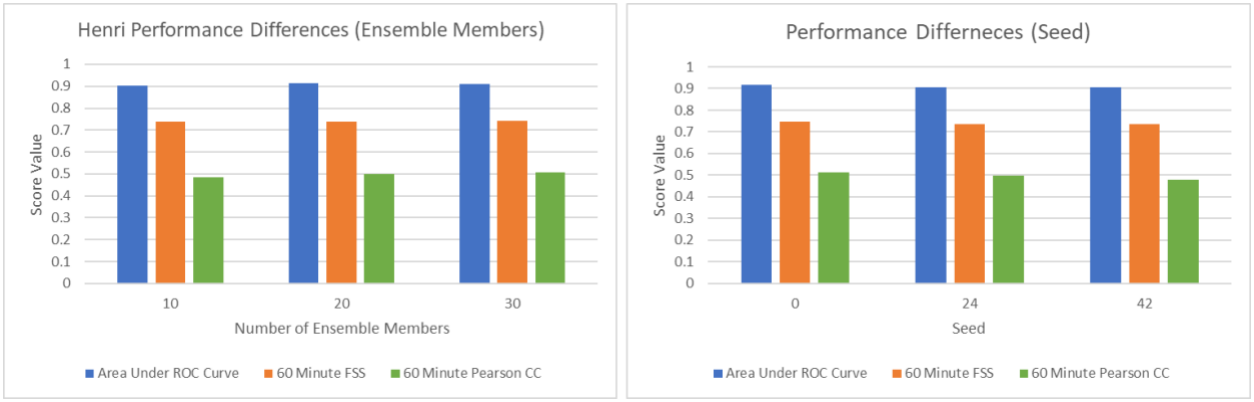

FIG. 2. As in Fig. 1, but for Henri.

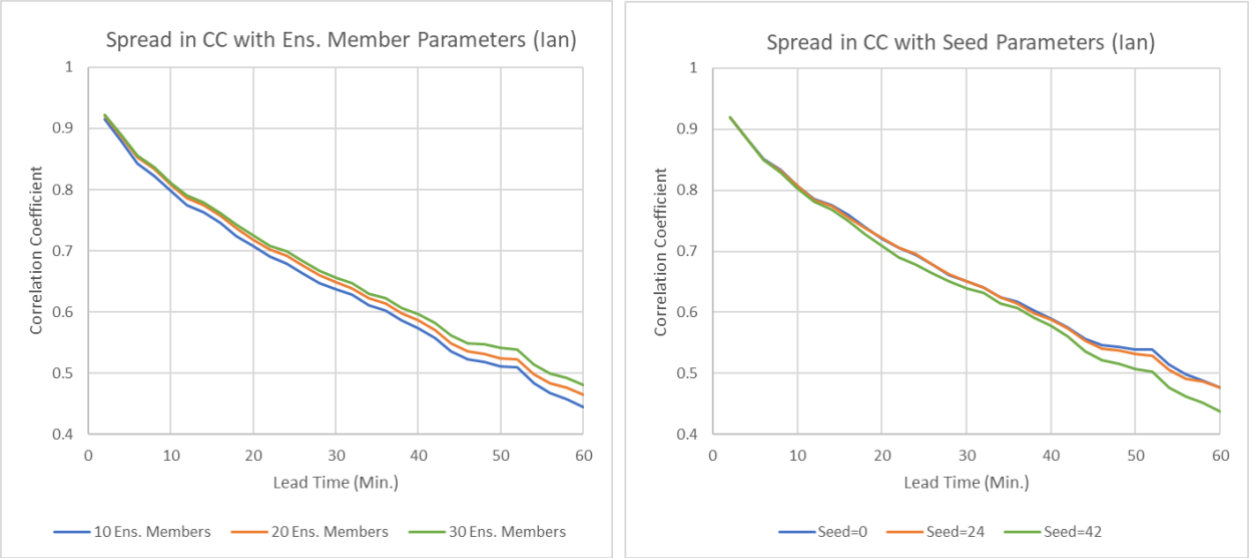

FIG. 3. Ian CC spread between the different number of ensemble members (left) and seed values (right).

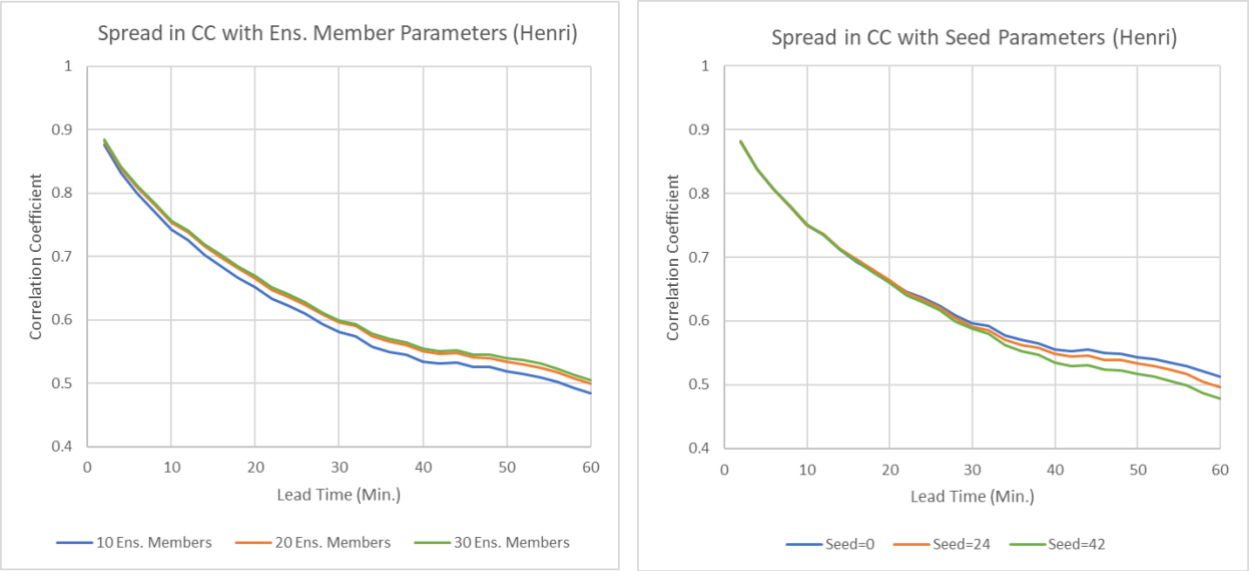

FIG. 4. As in Fig. 3, but for Henri.

best performance for Ian, yet it had the worst performance for Henri, in which 0.5 mm h^{-1} performed best.

The significance of some of these differences were determined through t-tests, assuming equal variances and using a significance level of α=0.05. Differences between CCs in the best (worst) performing advection tracking rate in Ian (Henri) and all others were found to be strongly significant (Tables 3-4). Differences in seed value CCs were mostly insignificant, and all differences in number of ensemble members were found to be insignificant (not shown).

Differences in ensemble mean output along with corresponding reliability diagrams demonstrate the substantial changes in output that occur when advection tracking threshold values are changed (Figs. 7-8). Changing the advection tracking threshold from 10 mm h^{-1} to 40 mm h^{-1} increased the AURC by 0.129, the FSS by 0.258, and the CC by 0.419. The reliability diagrams also indicate a reduction of the over-forecasting of precipitation through this parameter value change

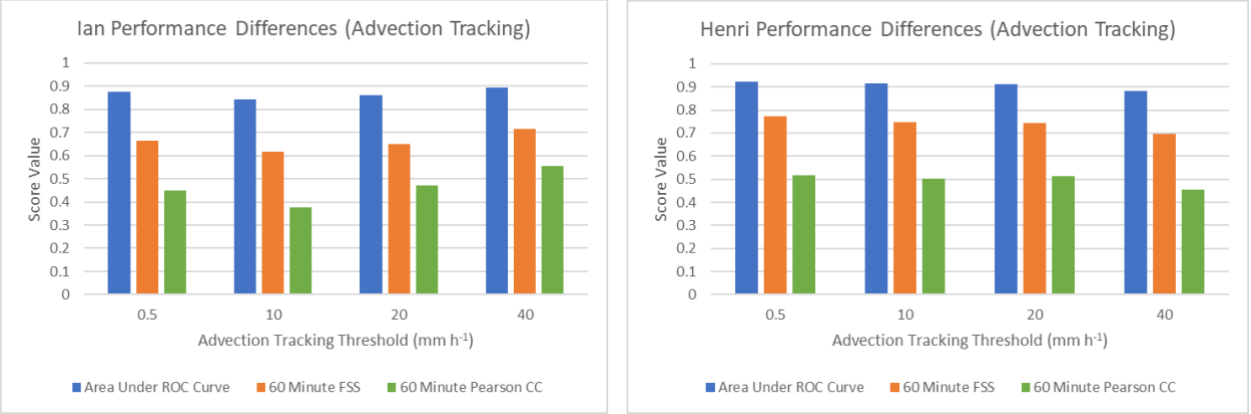

FIG. 5. Performance differences in response to changes in advection tracking thresholds for Ian (left) and Henri (right).

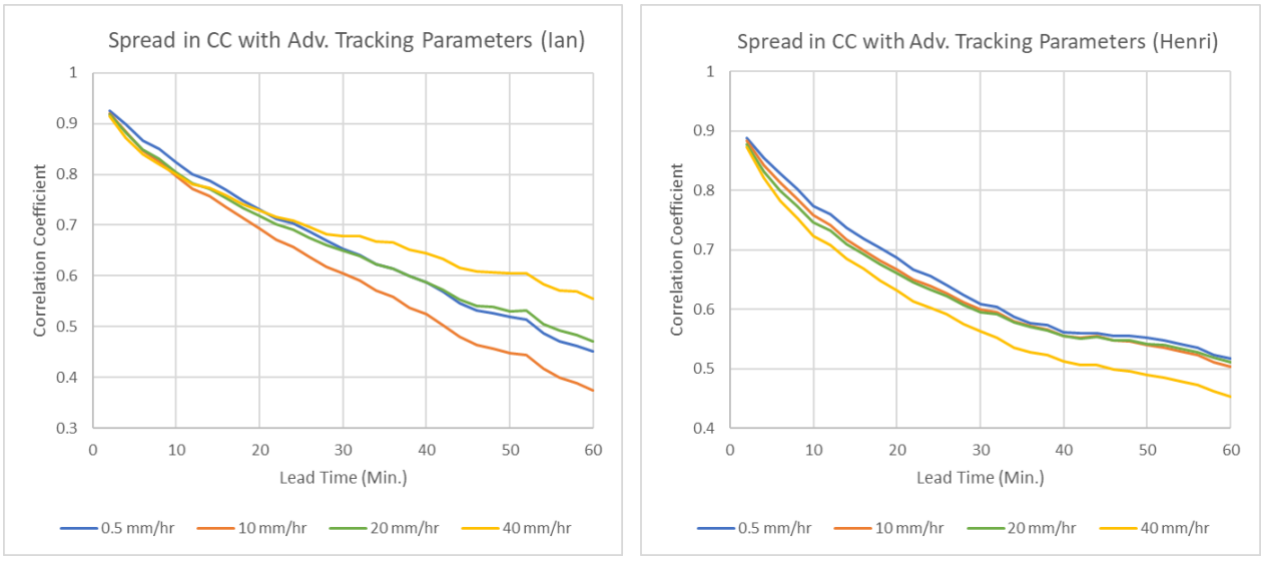

FIG. 6. CC spread between different advection tracking threshold values in Ian (left) and Henri (right).

(Fig. 7). Changing the advection tracking threshold from 40 mm h^{-1} to 0.5 mm h^{-1} increases the AURC by 0.121, the FSS by 0.187, and the CC by 0.187, along with some improvement in the forecast probabilities (Fig. 8).

Observed performance when setting seed to “None” revealed a generally small amount of variability (Figs. 9-11), especially in areas under ROC curves and FSSs. The average range of the AURC, FSS, and CC values were 0.015, 0.041, and 0.074, respectively. Ranges were also observed to decrease as the number of ensemble members increased, with the notable exception of 2100 UTC FSS ranges which has the lowest range with 10 ensemble members.

5. DISCUSSION

Results indicate that values for advection tracking thresholds played a significant role for each unique tropical cyclone being evaluated. Based on this study, Ian would need a far higher value than Henri. One possible explanation for this is that Henri lacked areas with rain rates of at least 40 mm h^{-1}, so the movement of only a few very small portions of the storm would be analyzed to generate the complete advection field estimation. In contrast, Ian had a much larger area of rain rates that exceeded 40 mm h^{-1}, especially around the eyewall. The southern portion of Ian was largely sheared off after making landfall, which would mean the LK method would only catch the

TABLE. 3. Ian t-test results for differences in avg CC per advection tracking threshold.

x_1	x_1(CC avg)	x_1(CC var)	x_2	x_2(CC avg)	x_2(CC var)	p
40 mm h^{-1}	0.556	0.019	0.5 mm h^{-1}	0.451	0.025	0.00014
40 mm h^{-1}	…	…	10 mm h^{-1}	0.375	0.027	<0.00001
40 mm h^{-1}	…	…	20 mm h^{-1}	0.471	0.033	0.00388

TABLE. 4. As in Table 3, but for Henri.

x_1	x_1(CC avg)	x_1(CC var)	x_2	x_2(CC avg)	x_2(CC var)	p
40 mm h^{-1}	0.453	0.008	0.5 mm h^{-1}	0.517	0.002	0.00001
40 mm h^{-1}	…	…	10 mm h^{-1}	0.503	0.002	0.00053
40 mm h^{-1}	…	…	20 mm h^{-1}	0.511	0.005	0.00027

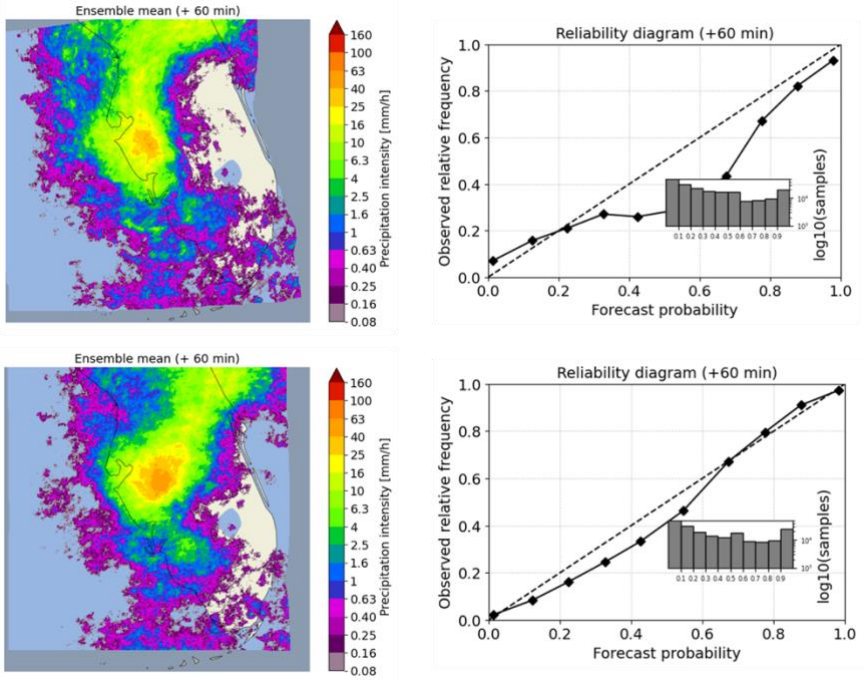

FIG. 7. Ian 0000 UTC run time ensemble mean output (left) and corresponding reliability diagrams (right) using 20 ensemble members, seed=42, and 10 mm h^{-1} advection tracking rate (top) / 40 mm h^{-1} advection tracking rate (bottom).

cyclonic movement of that section of the system when generating the advection field. This is the likely reason that some nowcasts for Ian had the forecast precipitation too far to the west (Fig. 7). Only taking into account movement of higher rain rates that were concentrated mostly around the eyewall likely allowed for a better representation of the forward movement of the storm, thus leading to results more closely aligned with observations.

The number of ensemble members was, surprisingly, found to be insignificant to performance, though raising this value did increase performance slightly in both cases. This has major implications for future work with this nowcasting scheme, since running 10 ensemble members takes far less run time than 30 members. More case studies need to be analyzed for this conclusion to be solidified. The seed value

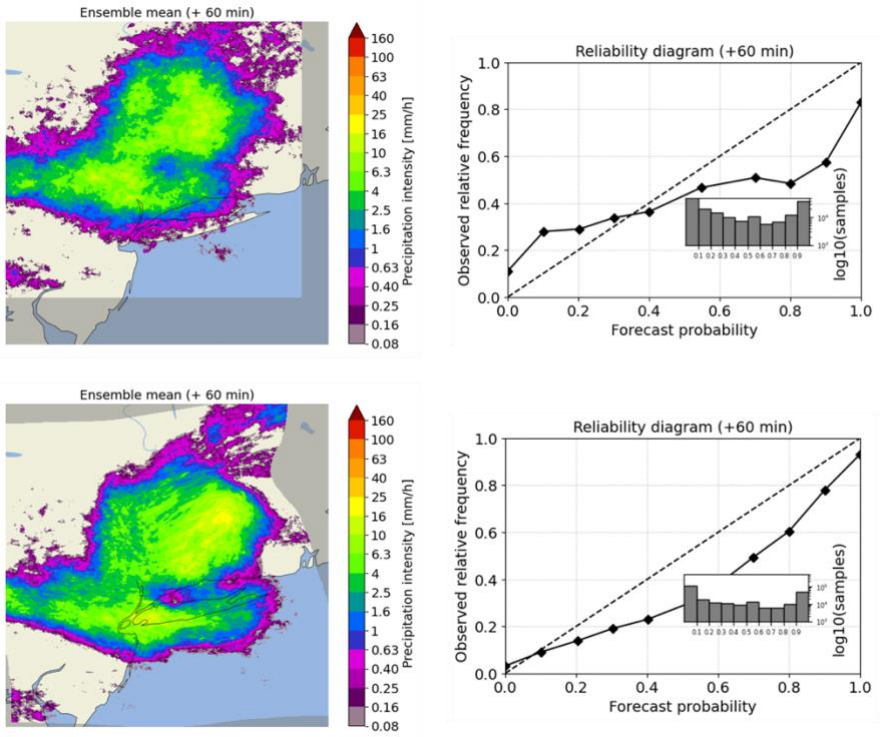

FIG. 8. Henri 1800 UTC run time ensemble mean output (left) and corresponding reliability diagrams (right) using 10 ensemble members, seed=0, and 40 mm h^{-1} advection tracking rate (top) / 0.5 mm h^{-1} advection tracking rate (bottom).

was also found to be insignificant to performance. All but one t-test using differences in CC revealed no significance between both cases. The additional analysis in which seed was set equal to "None" strengthened this conclusion that the particular seed value chosen is not of critical importance. This is valuable information, because there is a massive amount of seed value options that can be chosen. If more testing continues to support this conclusion, much time can be saved by focusing efforts elsewhere.

More case studies using a variety of disturbance types should be analyzed and compared to existing results in order to ensure a high level of versatility of these nowcasts. The testing of more parameter value combinations would also be beneficial for a more complete analysis of existing and future case studies. For example, maybe 30 mm h^{-1} would have performed better than a 40 mm h^{-1} advection rate for Ian. The testing of higher rain rate thresholds for FSS and AURC would benefit in determining accuracy in predicting the location of potential flash flooding. Higher resolution input data would likely be beneficial in improving accuracy. The initial plan in this study was to run all simulations using 500-m resolution input data as well; however, programming issues and time constraints prevented this.

6. SUMMARY

The performance of a parameter-based STEPS nowcasting method has been evaluated using MRMS instantaneous precipitation rates. Two TC case studies were analyzed, with 36 unique nowcasts containing different parameter combinations for each initiation run time. Performance statistics were averaged between all run times for analysis, which led to a number of conclusions:

- The advection tracking threshold was the most critical parameter. Changes in this value led to the greatest spread in performance, and the most significance in CC differences.
- Different advection tracking thresholds were necessary for TCs with significantly different precipitation field characteristics in order to have the best performance. For Henri, the 0.5 mm h^{-1} threshold was best, while in Ian, the 40 mm h^{-1} threshold performed the best.

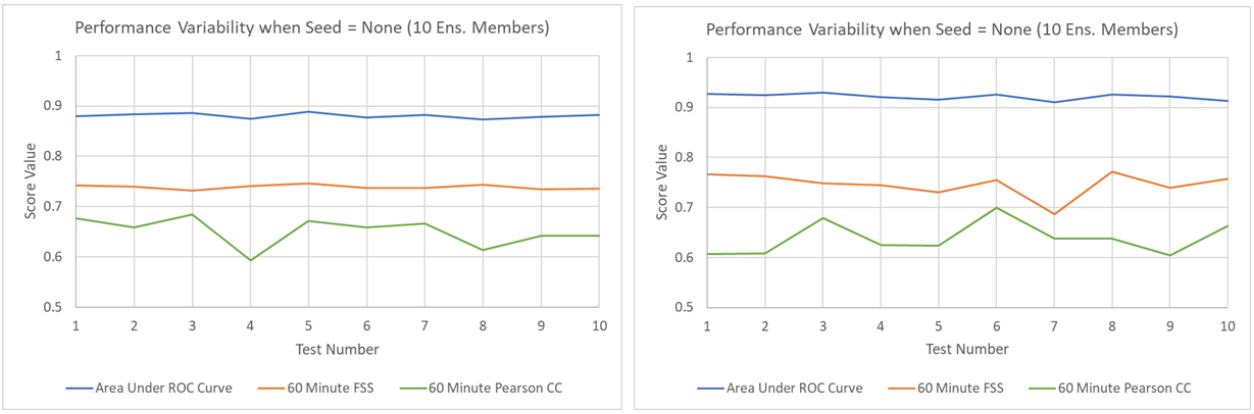

FIG. 9. Ian performance variability when seed=None for 10 ensemble members using 2100 UTC (left) and 0000 UTC (right) run times.

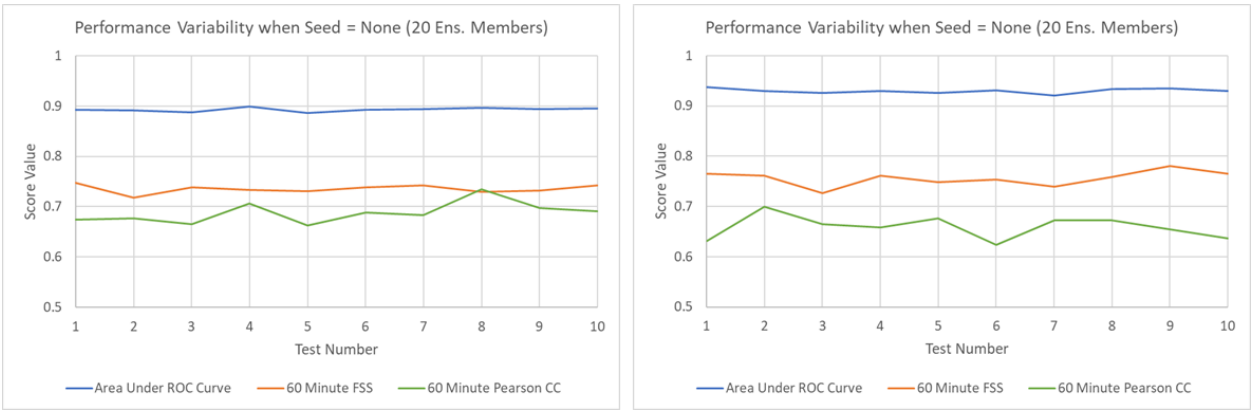

FIG. 10. As in Fig. 9, but for 20 ensemble members.

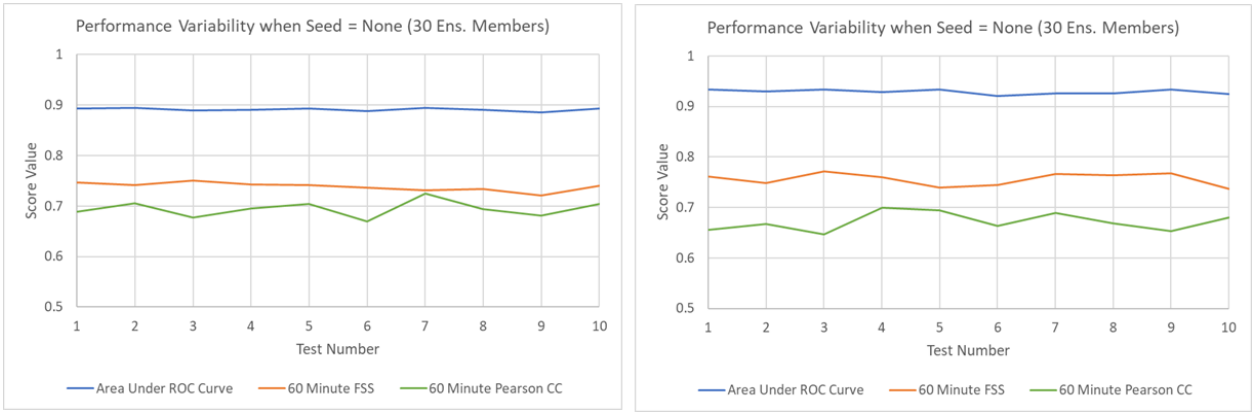

FIG. 11. As in Fig. 9, but for 30 ensemble members.

seed value were less critical parameters. Changes in these values led to lower spreads in performance, and little to no significance in CC differences.

This study has only examined a very small portion of the capabilities and tendencies of this nowcasting scheme. Much more work must be done before it can be implemented operationally (see discussion section). If successful, these nowcasts will allow for the resumption of progress in increasing flash flood warning lead times, which will help keep the public safer when these hazards occur.

7. ACKNOWLEDGEMENTS

The authors would like to thank Alex Marmo and Dr. Daphne LaDue for leading the National Weather Center REU and for their encouragement and support. This work was prepared by the authors with funding provided by National Science Foundation Grant No. AGS-2050267, and NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA11OAR4320072, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, NOAA, or the U.S. Department of Commerce.

7. REFERENCES

Bowler, N. E., C. E. Pierce, and A. W. Seed, 2006: STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Quart. J. Roy. Meteorol.*, **132**, 2127–2155, https://doi.org/10.1256/qj.04.100.

Creutin, J. D., M. Borga, E. Gruntfest, C. Lutoff, D. Zoccatelli, and I. Ruin, 2013: A space and time framework for analyzing human anticipation of flash floods. *J. Hydrol.*, **482**, 14–24, https://doi.org/10.1016/j.jhydrol.2012.11.009.

Jones, S. C., and Coauthors, 2003: The extratropical transition of tropical cyclones: forecast challenges, current understanding, and future directions. *Wea. Forecasting*, **18**, 1052–1092, https://doi.org/10.1175/1520-0434(2003)018<1052:TETOTC>2.0.CO;2.

Martinaitis, S. M., H. M. Grams, C. Langston, J. Zhang, and K. Howard, 2018: A real-time evaporation correction scheme for radar-derived mosaicked precipitation estimations. *J. Hydrometeor.*, **19**, 87–111, https://doi.org/10.1175/JHM-D-17-0093.1.

Martinaitis, S. M., and Coauthors, 2023: A path toward short-term probabilistic flash flood prediction. *Bull. Amer. Meteor. Soc.*, **104**, E585–E605, https://doi.org/10.1175/BAMS-D-22-0026.1.

Marzban, C., 2004: The ROC curve and the area under it as performance measures. *Wea. Forecasting*, **19**, 1106–1114, https://doi.org/10.1175/825.1.

Prudden, R., S. V. Adams, D. Kangin, N. H. Robinson, S. V. Ravuri, S. Mohamed, and A. Arribas, 2020: A review of radar-based nowcasting of precipitation and applicable machine learning techniques. *ArXiv*, https://doi.org/10.48550/arXiv.2005.04988.

Pulkkinen, S., D. Nerini, A. A. Pérez Hortal, C. Velasco-Forero, A. Seed, U. Germann, and L. Foresti, 2019: Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geosci. Model Dev.*, **12**, 4185–4219, https://doi.org/10.5194/gmd-12-4185-2019.

Qi, Y., and J. Zhang, 2017: A physically based two-dimensional seamless reflectivity mosaic for radar QPE in the MRMS system. *J. Hydrometeor.*, **18**, 1327–1340, https://doi.org/10.1175/JHM-D-16-0197.1.

Rappaport, E. N., 2014: Fatalities in the United States from Atlantic tropical cyclones: new data and interpretation. *Bull. Amer. Meteor. Soc.*, **95**, 341–346, https://doi.org/10.1175/BAMS-D-12-00074.1.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Ryzhkov, A., P. Zhang, P. Bukovcic, J. Zhang, and S. Cocks, 2022: Polarimetric radar quantitative precipitation estimation. *Remote Sens.*, **14**, 1695, doi: https://doi.org/10.3390/rs14071695.

Sharmin, N., and R. Brad, 2012: Optimal filter estimation for Lucas-Kanade optical flow. *Sens.*, **12**, 12694–12709, https://doi.org/10.3390/s120912694.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

Zhang, J., L. Tang, S. Cocks, P. Zhang, A. Ryzhkov, K. Howard, C. Langston, and B. Kaney, 2020: A dual-polarization radar synthetic QPE for operations. *J. Hydrometeor.*, **21**, 2507–2521, https://doi.org/10.1175/JHM-D-19-0194.1.