ARL-TR-8854 ● Nov 2019

DEVCOM ARMY RESEARCH LABORATORY

Application of the Scale Decomposition Technique for Assessing Radar Reflectivity Forecasts of the Weather Running Estimate–Nowcast Modeling System

by John W Raby, Brian P Reen, Huaqing Cai, and Jeffrey A Smith

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-8854 ● Nov 2019

Application of the Scale Decomposition Technique for Assessing Radar Reflectivity Forecasts of the Weather Running Estimate–Nowcast Modeling System

John W Raby, Brian P Reen, Huaqing Cai, and Jeffrey A Smith
Computational and Information Sciences Directorate, CCDC Army Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
November 2019	Technical Report	1 October 2018–30 September 2019

4. TITLE AND SUBTITLE
Application of the Scale Decomposition Technique for Assessing Radar Reflectivity Forecasts of the Weather Running Estimate–Nowcast Modeling System

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
John W Raby, Brian P Reen, Huaqing Cai, and Jeffrey A Smith

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CCDC Army Research Laboratory
ATTN: FCDD-RLC-EM
White Sands Missile Range, NM 88002-5501

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-8854

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ORCID IDs: Reen, 0000-0002-2031-4731; Smith, 0000-0002-2920-749X

14. ABSTRACT

The Army Weather Running Estimate–Nowcast (WRE-N) Modeling System, which uses the advanced research version of the Weather Research and Forecasting model to generate nowcasts, was modified to improve hydrometeor forecasts in support of flight system testing at the Ronald Reagan Ballistic Missile Defense Test Site on Kwajalein Atoll. One of the modifications made was to enable assimilation of Kwajalein radar reflectivity data. An assessment of the WRE-N was needed to determine the value-added of this modification. Due to the difficulty of verifying forecasts of spatial fields of 1-km above-ground-level reflectivity using traditional methods, a spatial verification technique called scale decomposition was evaluated for possible use because it has the capability to focus on reflectivity objects and assess the quality of the forecast structure in terms of the spatial scale of the error and intensity. It has been used for assessment of precipitation fields that are discontinuous in nature, not unlike those of radar reflectivity. To apply this method of verification, the Wavelet-Stat tool of the National Center for Atmospheric Research's Model Evaluation Tools was used. The results showed that this technique provided an assessment of model skill as a function of reflectivity threshold value and spatial scale complementary with other verification methods.

15. SUBJECT TERMS
scale decomposition, scale separation, thresholds, observations, model evaluation tools, wavelet transform

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			John W Raby
Unclassified	Unclassified	Unclassified	UU	37	19b. TELEPHONE NUMBER (Include area code) (575) 678-2004

Contents

List of Figures

List of Tables

Acknowledgments

We offer our thanks to Mr Robert Dumais of the US Army Combat Capabilities Development Command Army Research Laboratory, who contributed guidance and suggestions without which the study could not have been completed. Also, special thanks go to the Technical Publishing Branch for its technical editing excellence, specifically to Mark Gatlin.

Summary

The US Army Weather Running Estimate–Nowcast (WRE-N) Modeling System, which uses the advanced research version of the Weather Research and Forecasting model to generate short-range nowcasts, was modified to improve hydrometeor forecasts in support of flight system testing at the Ronald Reagan Ballistic Missile Defense Test Site on Kwajalein Atoll. One of the modifications made was to enable the assimilation of Kwajalein radar reflectivity data. An assessment of the WRE-N with this modification was needed to determine the value-added as a result of the modification. Because of the difficulty of verifying high-resolution forecasts of spatial fields of 1-km above-ground-level reflectivity using traditional methods, a spatial verification technique called scale decomposition was used to focus on the placement of reflectivity objects and the quality of the forecast structure in terms of the spatial scale of the error. This method has been used for assessment of precipitation fields that are discontinuous in nature, not unlike those of radar reflectivity. To apply this method of verification to forecasts of reflectivity, the Wavelet-Stat (W-S) tool of the National Center for Atmospheric Research's Model Evaluation Tools was used. W-S uses the Intensity-Scale approach to perform verification on different spatial scales as a function of reflectivity intensity values. The results showed that this technique provided an assessment of model skill as a function of reflectivity threshold value and spatial scale complimentary with other verification methods.

1. Introduction

The US Army Combat Capabilities Development Command Army Research Laboratory (ARL), in collaboration with the Aviation and Missile Center (formerly the Army Aviation and Missile Research, Development, and Engineering Center), CFD Research Corporation, and Integration Innovation, Inc. (commonly, i3), is developing and testing weather models for operational decision support. The role of ARL is to provide expertise and technologies in weather prediction for Go/No Go launch-commit criteria in support of flight testing and mission planning. The focus of this effort involves targeting flight test mission planning operation–specific requirements for weather prediction and applying our expertise in Numerical Weather Prediction (NWP) modeling and uncertainty quantification to mitigate the risks associated with missile testing. To accomplish this goal, we investigate modeling technologies that can improve hydrometeorology forecasts over the Ronald Reagan Ballistic Missile Defense Test Site on Kwajalein Atoll by assimilating radar data into the Weather Research and Forecasting (WRF) model.

During development of the assimilation technique, an ongoing process of model assessment is used to quantify the improvement in forecast accuracy achieved as a result of changes made in the model. Various metrics can quantify the skill of forecasts of 1-km above-ground-level (AGL) radar reflectivity. Traditional grid-to-point methods can verify the skill of NWP in predicting continuous meteorological variables through the computation of such statistics as mean error and root-mean-square error, which characterize model accuracy over the entire domain. When these techniques are applied to high-resolution models such as the Weather Running Estimate–Nowcast (WRE-N), the results can give misleading error estimates compared with lower-resolution models, which often score better when using these techniques. The problem lies in requiring the exact match between the point observations and the forecast grid values. This leads to the so-called "double penalty", where the feature in the forecast being spatially displaced creates an offset in position that produces two types of errors. The first type results from the forecast placing the feature where it was not observed, and the second type results from the forecast not placing the feature where it was observed. Furthermore, the error statistics provide no information to distinguish between occurrences of "near-miss" or "complete miss" forecasts. A "near-miss" forecast places the forecast feature at a location very close to the observed feature, while a "complete miss" forecast places the feature further away. Using traditional verification, both forecasts are penalized equally due to the "miss". No credit is given for the near-miss forecast despite the close proximity of the forecast and observed features compared with the feature of the "complete-miss" forecast, which was further away. Another issue is

the inability of the verification technique to evaluate the true skill of higher-resolution forecasts of spatially heterogeneous fields, which replicate mesoscale atmospheric features in a way that is more representative of the actual phenomenon owing to their use of a reduced grid spacing over smaller domains, higher-resolution land-surface models, and better parameterization of subgrid physical processes.

To obtain more-quantitative evidence of the skill of the WRE-N in predicting radar reflectivity, spatial methods for verification were considered. Spatial verification approaches, in contrast to grid-to-point verification approaches, focus on features and coherent spatial structure that characterize meteorological fields. Since these approaches account for the intrinsic spatial correlation existing between nearby grid points, they do not suffer from double penalty errors that impact grid-to-point verification. Spatial verification approaches take into account the observation and forecast time–space uncertainties and seek to provide feedback on the forecast error in physical terms. In particular, a spatial method called "fuzzy" verification has been used for model assessment. This method, also known as "neighborhood" verification, uses an approach that does not require exact matching and instead focuses on how well the atmospheric feature or object is replicated by the model, even if there is a spatial displacement of the feature. The goal is to determine the amount of displacement by using a range of sizes of neighborhoods of surrounding forecast and observed grid points in the verification process. In this way, model performance as a function of spatial scale can be determined to allow selection of the scale required to have the desired accuracy. Many methods for fuzzy verification have been developed, mostly for evaluating model precipitation forecasts. Ebert (2008) reviews a number of such methods.

A different spatial method for verification called scale-decomposition (or scale-separation) was used in this study to evaluate its effectiveness in quantifying the uncertainty at a range of different scale sizes of the error field between the forecast and observed reflectivity fields. Weather features on different scales (e.g., frontal systems versus convective showers) are often driven by different physical processes. Verification on different spatial scales can therefore provide deeper insights into model performance. The spatial scale components are obtained usually by applying a single-band spatial filter to the forecast and observation fields (e.g., Fourier decomposition wavelets). The scale-decomposition approaches measure error, bias, and skill of the forecast on each different scale component. The scale-decomposition approach provides feedback on the scale dependency of the error and skill and on the capability of the forecast for reproducing the observed scale structure (NCAR 2016).

2. Model Configuration and Experimental Design

The NWP model used is a triple-nest implementation of the Advanced Research version of the Weather Research and Forecast model (WRF-ARW). This model was configured similar to that tested as a prototype ARL Weather Running Estimate–Nowcast (WRE-N) Modeling System (Dumais et al. 2013). The WRF is discussed in detail by Skamarock et al. (2008). The WRE-N used Obsgrid to apply a multiscan Cressman analysis to incorporate surface and upper air weather observations. WRF and WRE-N are used interchangeably throughout this report. For this project, the WRFV3.9.1.1 was the core of the WRE-N that was modified by Reen et al. (in preparation) to assimilate radar reflectivity observations from the Ronald Reagan Kwajalein Atoll Ballistic Missile Defense Test Site. The domain configuration, shown in Fig. 1, was a triple nest with the outer domain on a 9-km grid spacing, the middle domain on a 3-km grid, and the inner domain on a 1-km grid spacing. The modifications made were to allow application of the radar-derived latent heating term to the WRE-N preforecast without application of digital filter initialization and used the following parameterizations, which closely match the tropical parameterization suite:

- Planetary Boundary Layer: YSU (Yonsei State University)
- Microphysics: WSM6 (WRF Single Moment 6 Class)
- Cumulus parameterization: New Tiedke (coarsest domain only)
- Surface layer scheme: Revised MM5 (5th-generation Penn State/National Center for Atmospheric Research [NCAR] mesoscale model)
- Land surface model: Noah

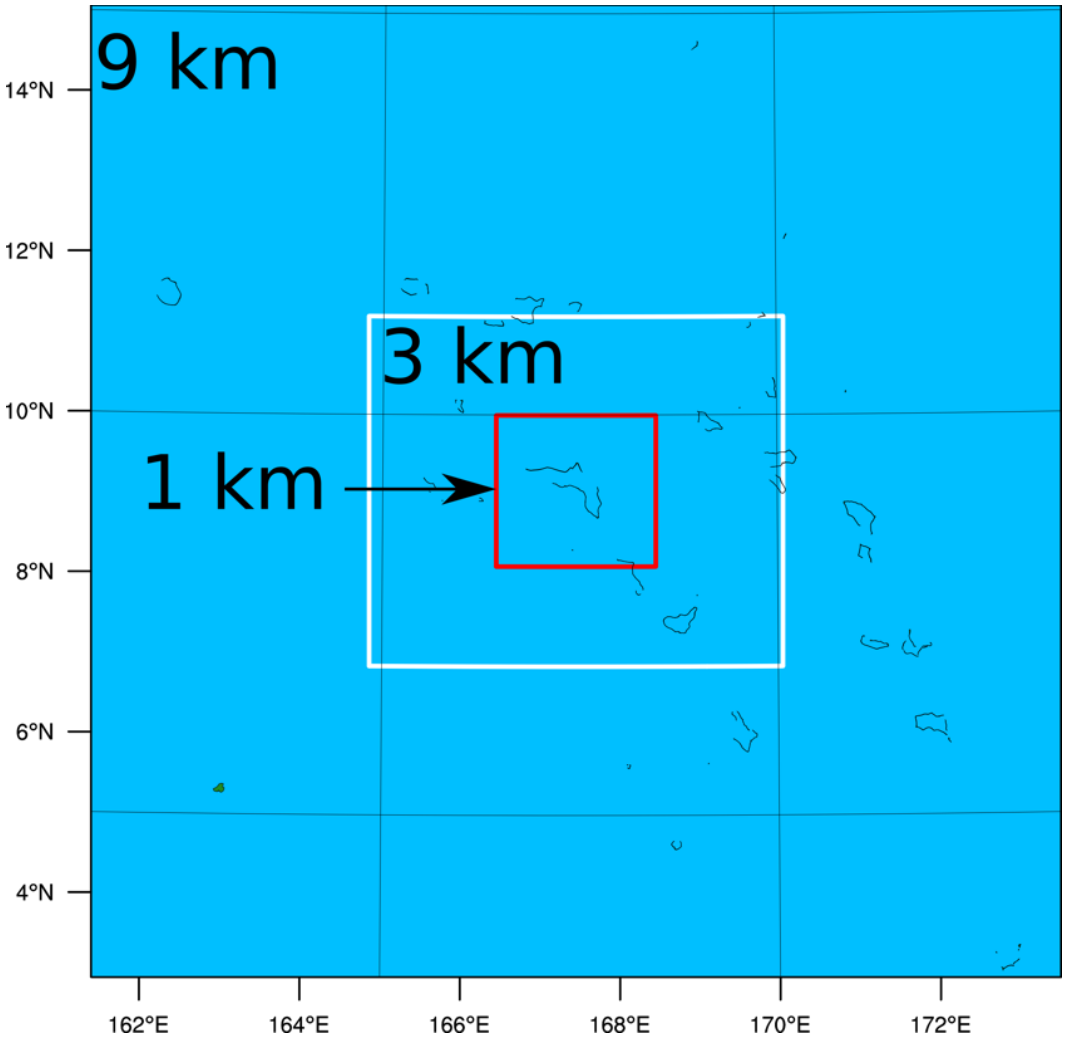

Fig. 1 WRE-N modeling domains

The experiment consisted of running the WRE-N to produce output plus collecting the radar observations for analysis to determine the benefit of assimilating the radar reflectivity data. The radar-derived latent heating term was applied to the WRE-N running on the 1-km grid. Two cases studies involving two different start times on the same day were examined. The length of the preforecast (assimilation phase) varied from 1, 2, and 6 h with the 0-hour forecast time set to 0100 and 0600 Coordinated Universal Time (UTC) on 10 September 2016. The evaluation metric was 1-km AGL reflectivity. For this report, a subset of the WRE-N output was selected to demonstrate the applicability of the scale-decomposition method of spatial verification. The forecast output was the 0-hour lead time valid at 0100 UTC from the 6-h preforecast of the 181015B experiment and the radar observation taken at 0059 UTC. The forecast and the observation comprise what is called the 181015B experiment later in this report.

3. Application of Scale Decomposition to Case Study Data

The Model Evaluation Tools (MET), which was developed by NCAR, is a suite of tools used for model verification (NCAR 2016). Previous assessments for this project, described by Reen et al. (in preparation), used a tool from MET called MET Grid-Stat that performs neighborhood verification using the radar reflectivity forecast and observations. The output generated consisted of the Fractions Skill Score (FSS) computed at a range of threshold values and neighborhood sizes and the frequency bias computed as a function of threshold values. The FSS compares the fraction of grid cells exceeding a reflectivity threshold within a given neighborhood size for the forecast with the same fraction from the observation. A score of 1.0 is considered a perfect forecast. The frequency bias is the ratio of the grid cells exceeding a reflectivity threshold in the model compared with the observations. A perfect forecast has a bias of 1.0.

The particular tool used to perform scale-decomposition verification in this report is the MET Wavelet-Stat (W-S) tool, which uses the Intensity-Scale technique developed by Casati et al. (2004). The Intensity-Scale technique evaluates the forecast skill as a function of the intensity or magnitude of the field values and of the spatial scale of the error. The scale components are decomposed by applying a 2-D Haar wavelet filter. Note that wavelets, because of their locality, are suitable for representing discontinuous fields such as precipitation. The technique uses a long-established and reliable categorical approach suitable for non-normally distributed variables, such as precipitation. The Intensity-Scale technique was specifically designed to handle the difficult characteristics of precipitation fields and for the verification of spatial precipitation forecasts (NCAR 2016). Since radar reflectivity fields are analogous to precipitation fields, this technique is equally appropriate for use in verification.

To evaluate the use of the Intensity-Scale method as a spatial technique for quantifying the skill of the WRE-N radar reflectivity forecast, the W-S tool was used to apply that method to the case study data from the 181015B experiment. The W-S tool allows the user to adjust the position of the tile to focus the verification on a specific feature of interest. In this case, the feature was elevated reflectivity associated with convection. Figure 2 shows the model 0-h forecast reflectivity (left image) and the observed reflectivity (right image) present in the inner domain and the smaller, 32 × 32 grid square tile outlined in red where a 2-D Haar wavelet filter was applied for decomposition into orthogonal component images for the different scales. The tile was positioned to encompass both reflectivity features. Note that the WRE-N feature matches the observed feature quite well despite some differences due to displacement and in the structure discussed later in this report.

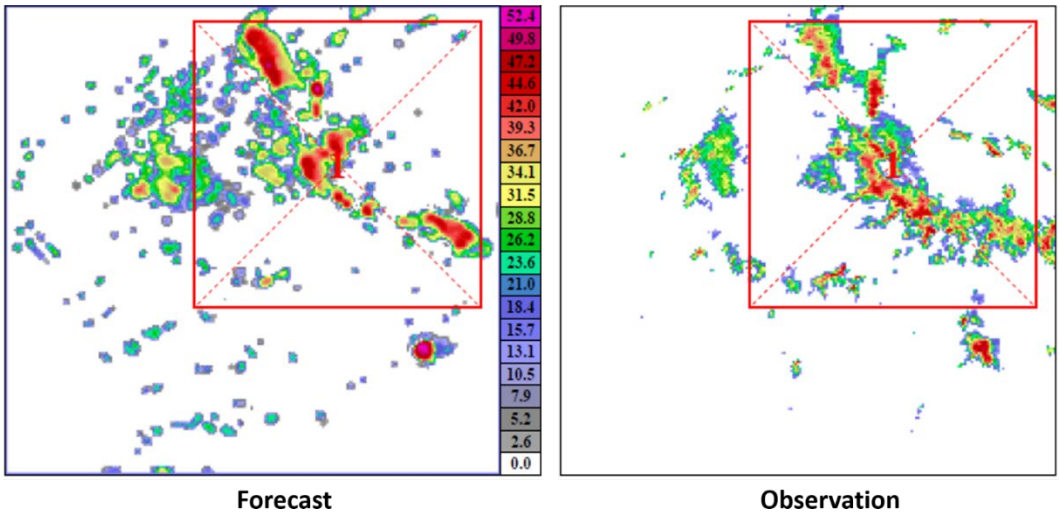

Fig. 2 Graphic produced by W-S showing the reflectivity feature from the forecast and from observations with 32 × 32 grid square tile

The scale-decomposition process of W-S begins with the application of the thresholds to the forecast and observations reflectivity fields that transforms them into binary fields by assigning a value of 1 where there is an event and a value of 0 where there is no event. The resulting binary difference or error field can then be used to generate 1) metrics that quantify the error as a function of scale for a given threshold, 2) forecast and observed energy metrics that relate to the numbers of events present in the forecast and observation fields for each scale at a given threshold, and 3) the Intensity Skill Score (ISS), which quantifies forecast skill for each scale at a given threshold. The following describes how these metrics are computed:

Given a forecast field (F) and an observed field (O), the associated binary fields (B_f and B_O respectively) are calculated as follows:

$$B_f(i,s,t) = \begin{cases} 0 \ if \ F(i,s) < t \\ 1 \ if \ F(i,s) \geq t \end{cases} \quad (1)$$

and

$$B_o(i,s,t) = \begin{cases} 0 \ if \ O(i,s) < t \\ 1 \ if \ O(i,s) \geq t \end{cases}, \quad (2)$$

where:

B_f = binary forecast field

B_o = binary observation field

F = forecast field

i = a specific horizontal grid point (i.e., this covers variations in x and y)

O = observed field

s = a specific scale

t = a specific threshold

Now we can define the binary difference field Z:

$$Z(i,s,t) = B_f(i,s,t) - B_o(i,s,t). \tag{3}$$

The non-scale-specific binary difference field for a specific threshold is

$$Z(i,t) = \sum_{s=1}^{S} Z(i,s,t)\,, \tag{4}$$

where S is the total number of scales.

The mean squared error (MSE) for a specific scale and threshold can be calculated as

$$MSE(s,t) = \frac{1}{N}\sum_{i=1}^{N} Z^2(i,s,t)\,, \tag{5}$$

where N is the total number of grid points.

The overall MSE for a given threshold can be calculated as

$$MSE(t) = \frac{1}{S}\sum_{s=1}^{S} MSE(s,t)\,. \tag{6}$$

The MSE% is the scale- and threshold-specific MSE normalized by the overall MSE for a given threshold and provides the fraction (not percent) of the MSE at a given threshold that is at a specific scale:

$$\mathrm{MSE\%(S,T)} = \frac{MSE(s,t)}{MSE(t)}\,. \tag{7}$$

The forecast energy squared ($En2_f$) and observation energy squared ($En2_o$) measure the amount of energy at a specific threshold and possibly a specific scale. The scale-specific version is

$$En2_p(s,t) = \sum_{i=1}^{N} B_p^2(i,s,t)\,, \tag{8}$$

where P is F for forecast or O for observed. We can also measure the overall energy at a specific threshold (i.e., over all spatial scales):

$$En2_p(t) = \frac{1}{S}\sum_{s=1}^{S} En2_p(s,t)\,. \tag{9}$$

The percentage version of En2 for forecast ($En2\%_f$) and observation ($En2\%_o$) measures how energy is distributed as a function of scale for a given threshold and provides the fraction (not percent) of the energy for a threshold that is at a specific scale:

$$En2\%_p(s,t) = \frac{En2_p(s,t)}{En2_p(t)}\,. \tag{10}$$

The energy relative bias (En2 bias) is

$$En2\ bias = \frac{En2_f - En2_o}{En2_f + En2_o}\,. \tag{11}$$

The ISS is

$$ISS(s,t) = 1 - \frac{(S+1) * MSE(s,t)}{MSE_{random}(t)}\,, \tag{12}$$

where MSE_{random} is the MSE for the random binary reference forecast.

Figure 3 shows the W-S output graphic that contains the binary difference or error field with various scores and error statistics for the reflectivity threshold value of ≥10.0 decibels of radar reflectivity (dBZ). The graphic shows the difference field (F-O, forecast minus observation) between the binary forecast field and the binary observation field at 1-km resolution.

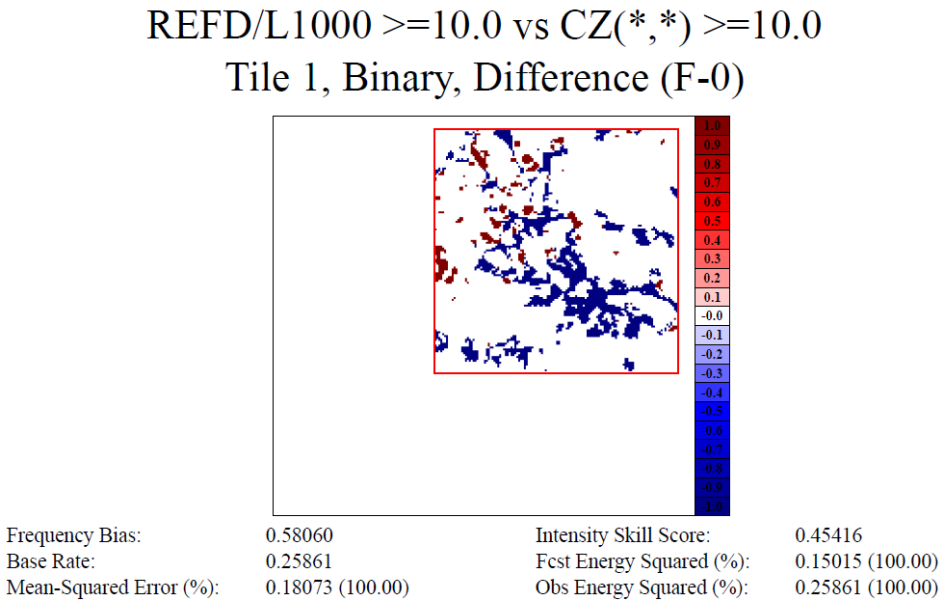

Fig. 3 W-S output graphic of the binary difference field for threshold value ≥10.0 dBZ

The value of the differences for the events is +1 (shaded red), 0, or –1 (shaded blue), which accounts for the lack of differences over the full range of color shades within the domain except for the darkest red or blue. A value of +1 (red) indicates the model forecasted a value ≥10 dBZ, but the radar did not observe this (i.e., overforecast), while a value of –1 (blue) indicates that the radar observed a value of ≥10 dBZ, but the model did not forecast this (i.e., underforecast). The value of the frequency bias (<1) indicates the prevalence of event underforecasting. The pattern of the under- and overforecast areas appears to be an indication of some spatial displacement errors as well as errors in the structure of the reflectivity feature.

W-S computes the same statistics for each decomposed scale. Figure 4 shows the first spatial scale component (Scale 1) of the binary field difference, which was obtained by taking the difference (F-O) between the spatial scale components of the binary forecast and observations fields per Fig. 3. Because the wavelet transform is a linear operator, decomposing the binary difference field into spatial scale components is equal to taking the difference between spatial scale components of the forecast and observations fields (Casati 2010). Scale 1 is the first mother wavelet component with a corresponding resolution of 1 km. The mother wavelet captures the variation about the mean component of the reflectivity binary difference field as opposed to the coarser first father wavelet, which captures the mean component of the reflectivity field after spatial averaging over a 2- × 2-km subarea of the binary difference field. The magnitude of the mother wavelet is thus ≤ |1|, which accounts for the presence of the full range of blue and red color shades

present in the difference field in the domain. The sum of the mother and father wavelets produces the original binary difference field. The value of the frequency bias (<1), which is not dependent on scale, indicates the prevalence of event underforecasting at the 10-dBZ threshold. The negative value of the ISS indicates no forecast skill for threshold value of 10 dBZ and errors of 1-km spatial scale.

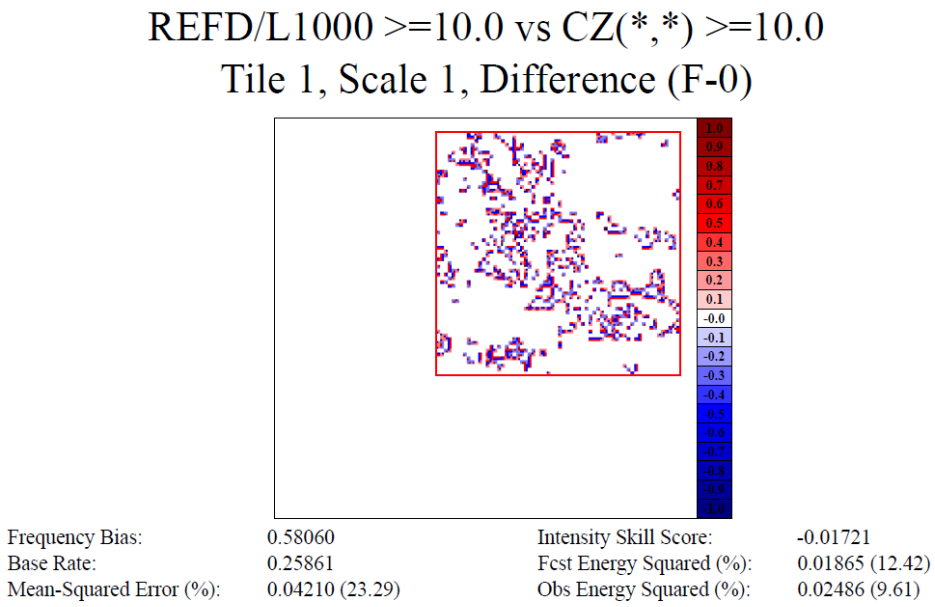

Fig. 4 W-S output graphic of the first spatial scale (1 km) for threshold value ≥10.0 dBZ

Figures 5–11 are analogous figures to Fig. 4, but for increasingly large scales. The scales represented by Figs. 4–11 are 2, 4, 8, 16, 32, 64, and 128 km with the values averaged over squares with sides of these lengths compared with values averaged over squares with sides twice these lengths. The exception is that in Fig. 11, since 128 km is the size of the entire comparison area, the value plotted is simply the difference between the mean forecast and mean observed values. After being negative and thus showing no skill at the finest scale (1 km; Fig. 4), the ISS becomes positive at 2 km (Fig. 5) and increases with increasing scale. The exception is that ISS is smaller at 128 km than 64 km, but the 128-km scale is simply an average over the entire area of focus.

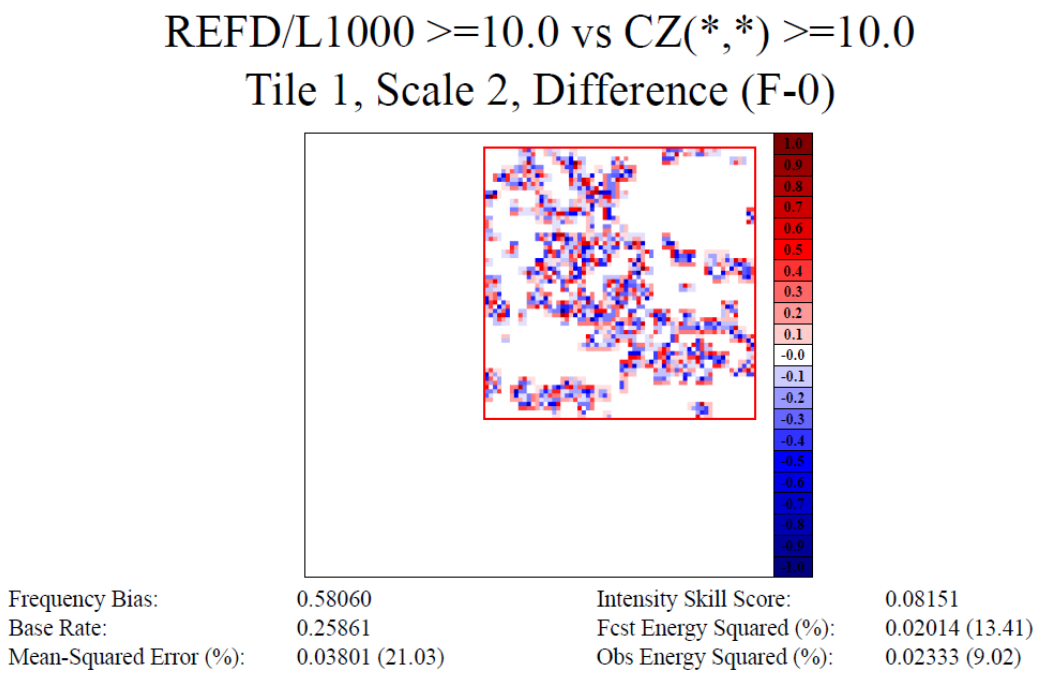

Fig. 5 W-S output graphic of the second spatial scale (2 km) for threshold value ≥10.0 dBZ

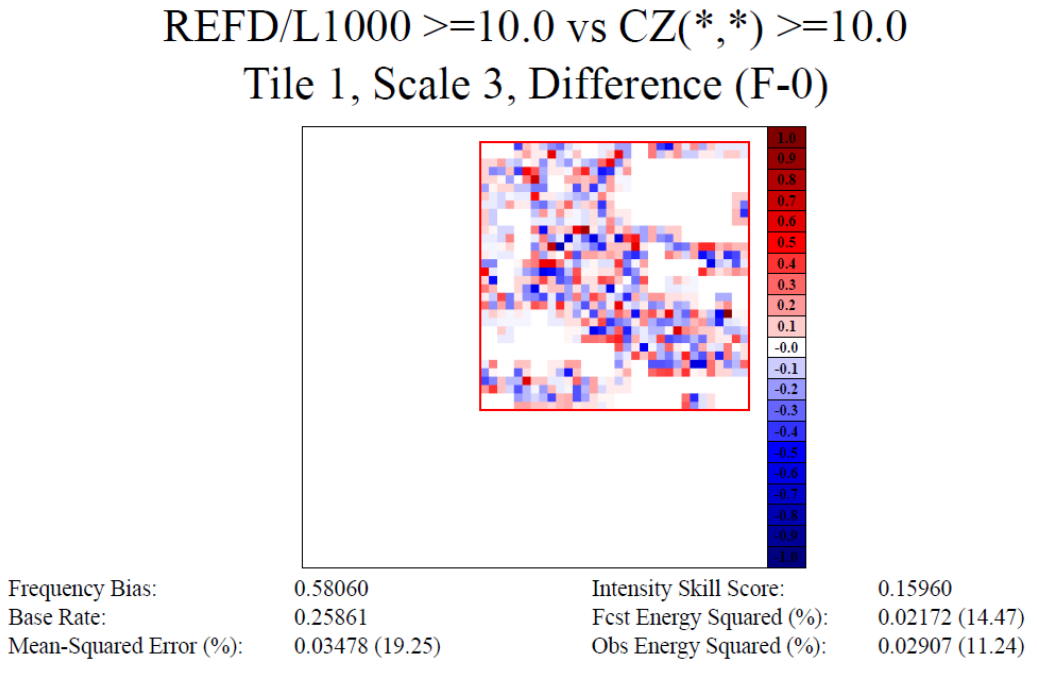

Fig. 6 W-S output graphic of the third spatial scale (4 km) for threshold value ≥10.0 dBZ

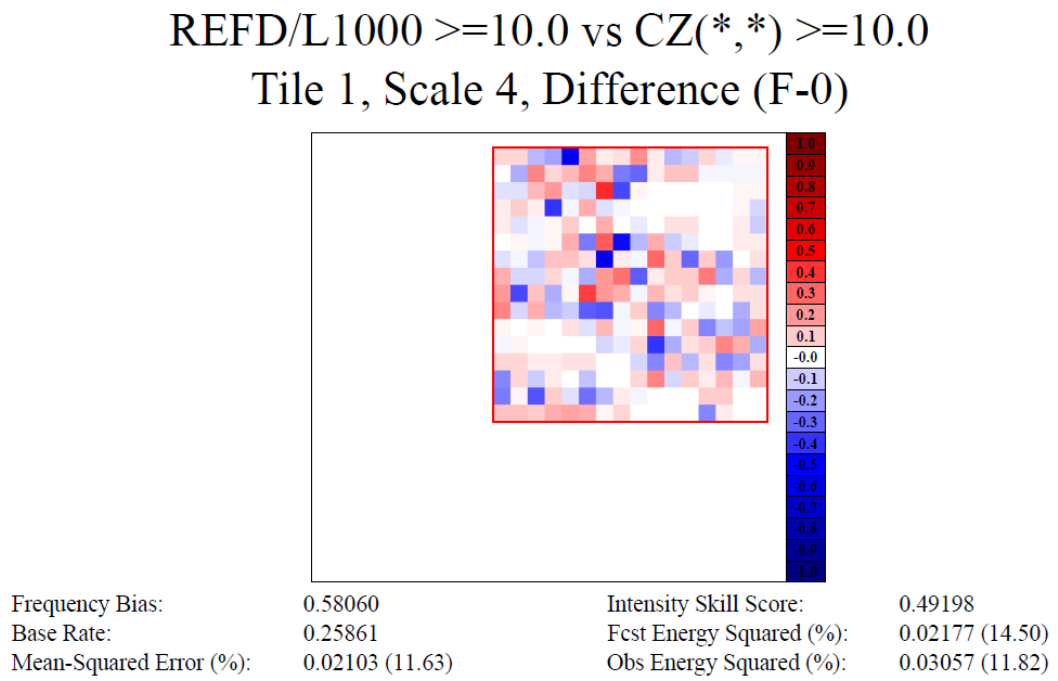

Fig. 7 W-S output graphic of the fourth spatial scale (8 km) for threshold value ≥10.0 dBZ

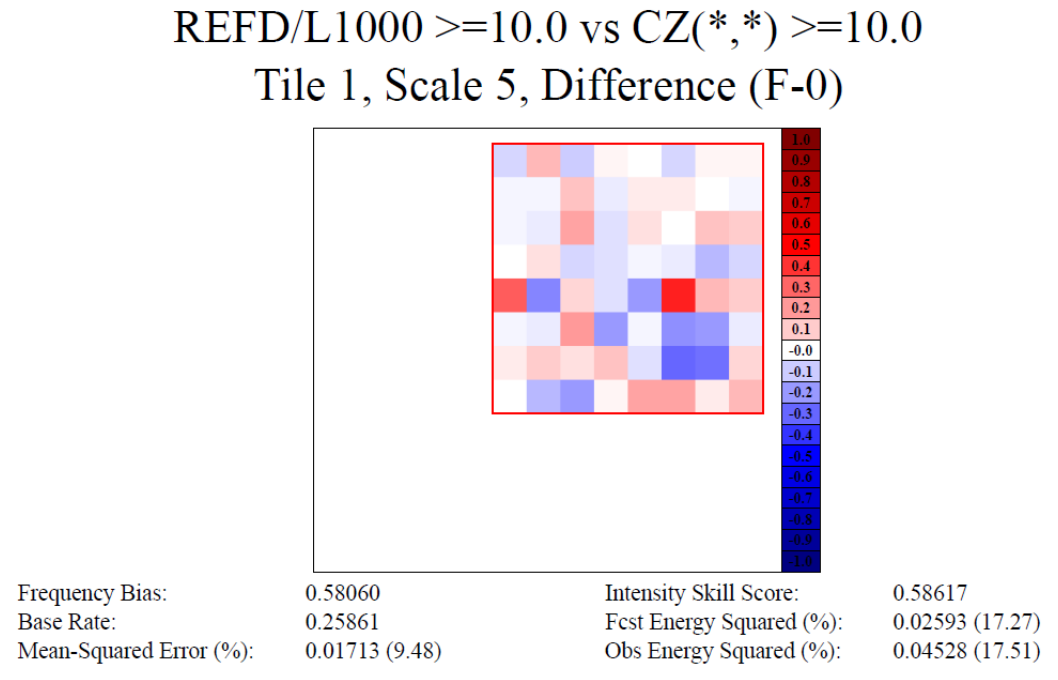

Fig. 8 W-S output graphic of the fifth spatial scale (16 km) for threshold value ≥10.0 dBZ

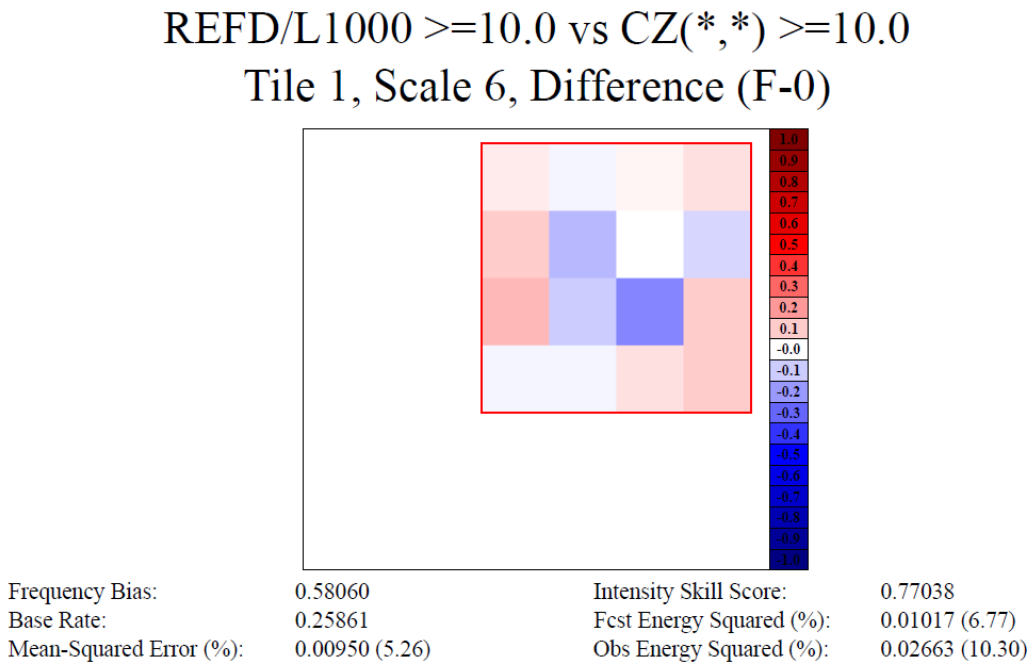

Fig. 9 W-S output graphic of the sixth spatial scale (32 km) for threshold value ≥10.0 dBZ

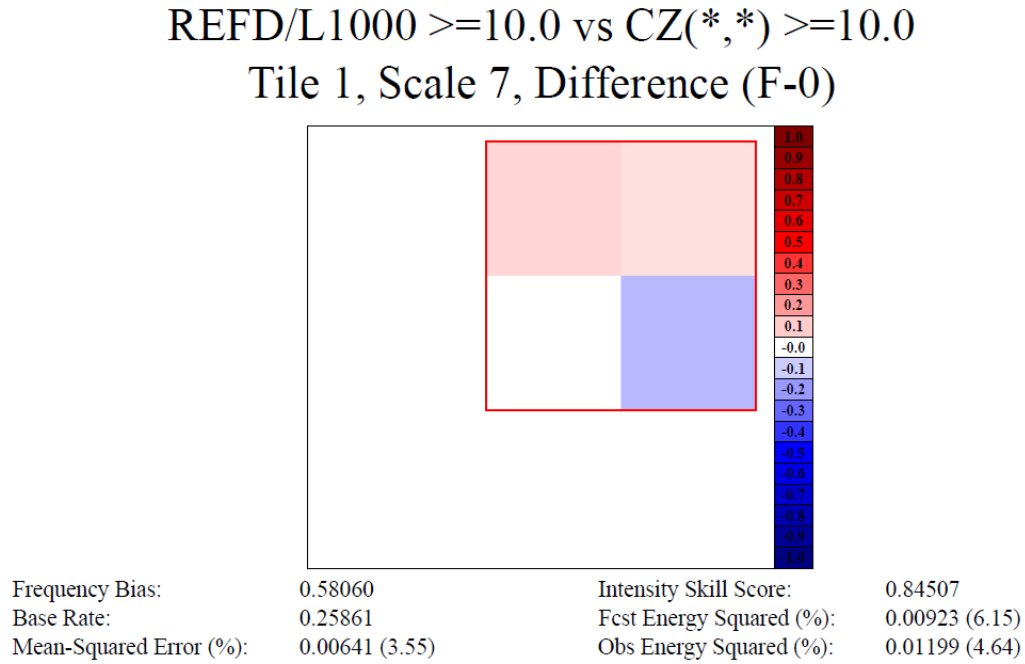

Fig. 10 W-S output graphic of the seventh spatial scale (64 km) for threshold value ≥10.0 dBZ

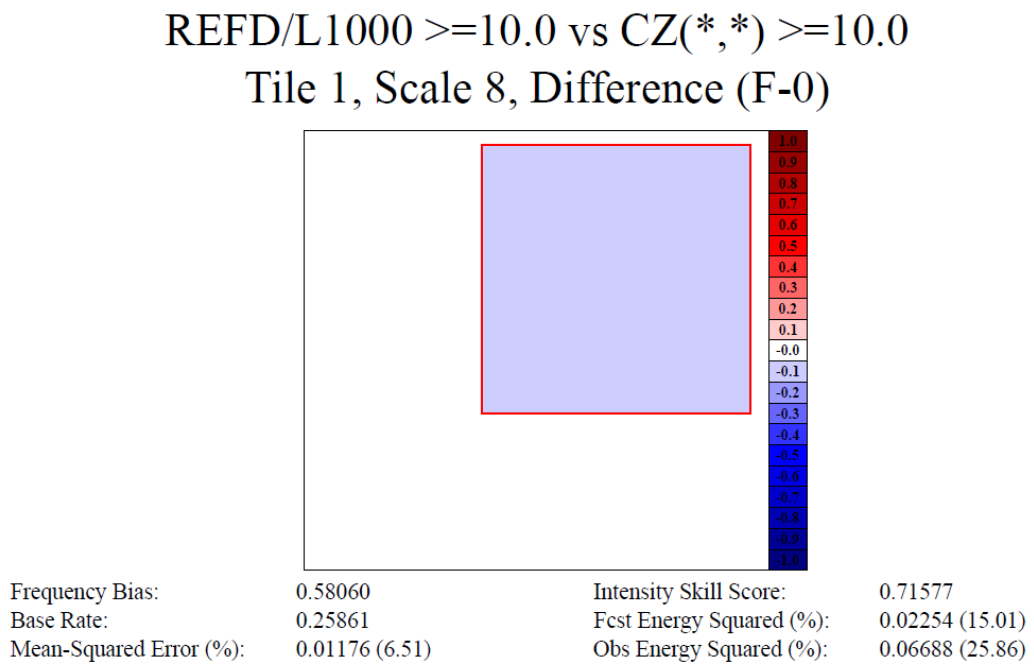

Fig. 11 W-S output graphic of the eighth spatial scale (128 km) for threshold value ≥10.0 dBZ

W-S generates a text file that contains all the mentioned scores and statistics for all thresholds and scale sizes. These data were plotted and put into tables for analysis.

4. Analysis of MET Wavelet-Stat Output

Table 1 is an ISS table as a function of threshold value and error scale. Figure 12 is a plot of the ISS at all thresholds and scales.

Table 1 ISSs for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

Threshold (dBZ)	1km	2km	4km	8km	16km	32km	64km	128km
>=5	-0.05	0.04	0.1	0.49	0.57	0.77	0.81	0.87
>=10	-0.02	0.08	0.16	0.5	0.59	0.77	0.85	0.72
>=15	-0.04	0.13	0.18	0.48	0.65	0.79	0.89	0.63
>=20	-0.13	0.09	0.23	0.53	0.7	0.79	0.92	0.65
>=25	-0.27	-0.02	0.26	0.57	0.72	0.77	0.93	0.73
>=30	-0.35	-0.1	0.21	0.53	0.72	0.77	0.94	0.8
>=35	-0.79	-0.15	0.08	0.53	0.69	0.77	0.95	0.86
>=40	-1.66	-0.56	-0.24	0.5	0.58	0.77	0.96	0.88
>=45	-3.13	-1.07	0.15	0.71	0.87	0.92	0.99	0.97

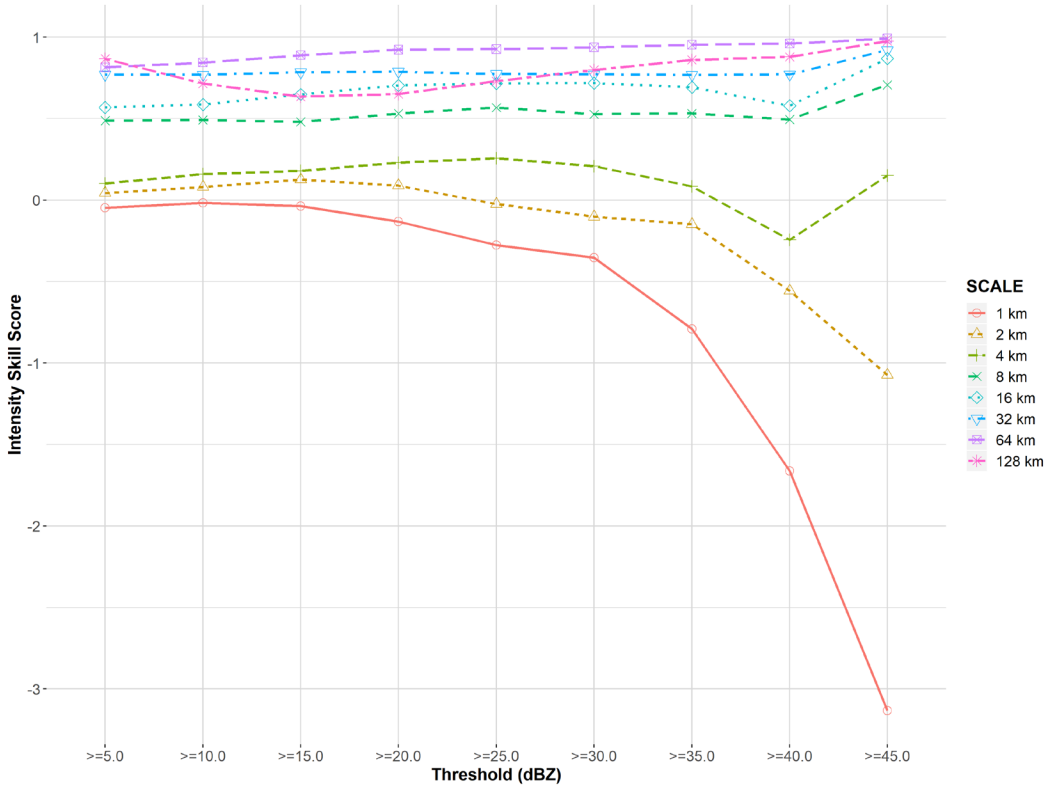

Fig. 12 Plot of ISSs for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

The ISS from this case study shows that the reflectivity forecasts exhibit the expected tendencies of improving scores with increasing scales. The smallest scales associated with convective showers are less predictable, while larger scales associated with larger, more organized convective activity are more predictable. For all threshold values and scales ≥8 km, the ISS is all positive, showing good skill. There were no occurrences of negative ISS value surrounded by positive ISSs, which suggests the lack of significant displacement errors with a possible exception being the 4-km scale at the 40-dBZ threshold, where there is an outlying negative ISS nearly disconnected from surrounding negative ISS values at lower thresholds and at all scales. For the 4-km scale at all thresholds, no other ISS value is negative except for that at the 40-dBZ threshold, which suggests that perhaps there is a slight displacement error for reflectivity features of 4-km spatial scale whose reflectivity is ≥40 dBZ but <45 dBZ. Forecast skill is degraded due to small spatial scale (<2 km) errors for all thresholds. Errors at this scale may be attributed to reduced skill in handling small-scale reflectivity structure.

Table 2 is a table of the frequency bias as a function of threshold value. Figure 13 is a plot of frequency bias as a function of threshold value.

Table 2 Frequency bias for threshold values from ≥5 to ≥45 dBZ

Threshold (dBZ)	Frequency bias
>5	0.71
>10	0.58
>15	0.50
>20	0.46
>25	0.44
>30	0.41
>35	0.38
>40	0.18
>45	0.10

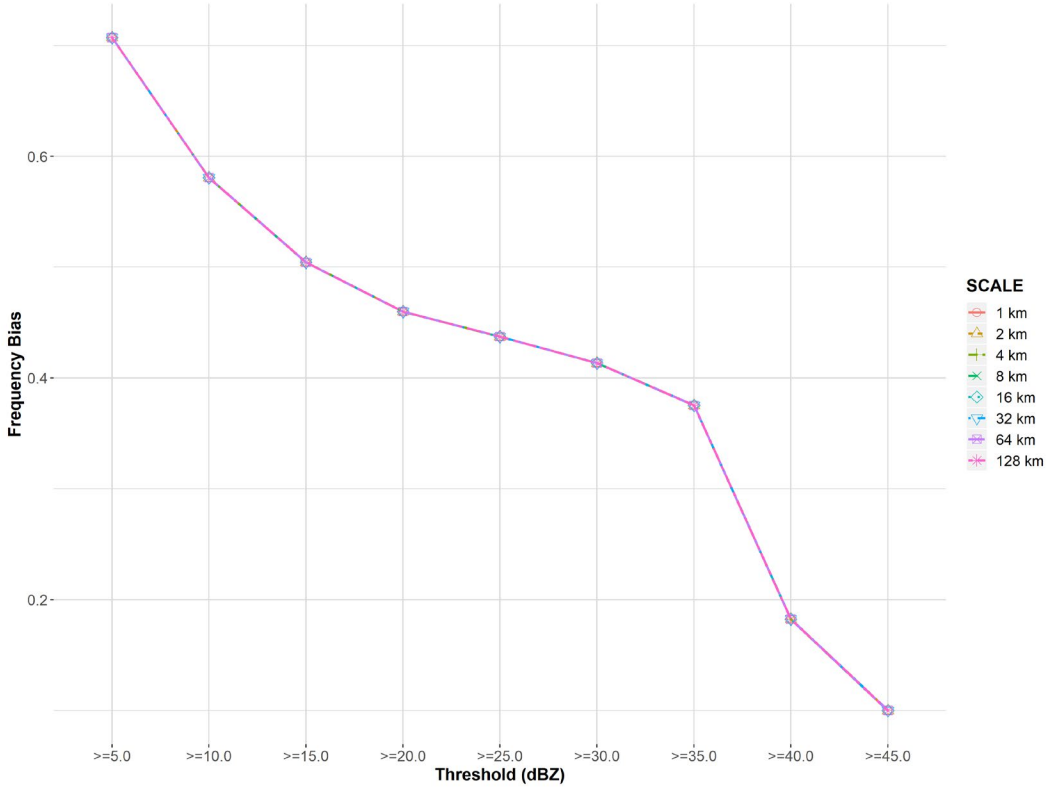

Fig. 13 Plot of frequency bias for threshold values from ≥5 to ≥45 dBZ

The values of frequency bias shows a general underforecast tendency at all thresholds with increasing bias occurring with increasing threshold values.

Table 3 is a table of the MSE% as a function of threshold value and error scale. Figure 14 is a plot of the MSE% at all thresholds and scales.

Table 3 MSE% for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

Threshold (dBZ)	1km	2km	4km	8km	16km	32km	64km	128km
>=5	23.8	21.8	20.4	11.7	9.8	5.2	4.2	3.0
>=10	23.3	21.0	19.3	11.6	9.5	5.3	3.6	6.5
>=15	24.1	20.4	19.1	12.1	8.2	5.0	2.6	8.5
>=20	26.9	21.6	18.3	11.1	7.0	5.0	1.8	8.3
>=25	29.5	23.7	17.2	10.0	6.6	5.2	1.7	6.2
>=30	30.1	24.5	17.6	10.5	6.3	5.1	1.4	4.5
>=35	35.5	22.7	18.2	9.3	6.1	4.6	0.9	2.8
>=40	39.3	23.0	18.4	7.5	6.2	3.4	0.6	1.8
>=45	54.5	27.3	11.2	3.9	1.7	1.0	0.1	0.3

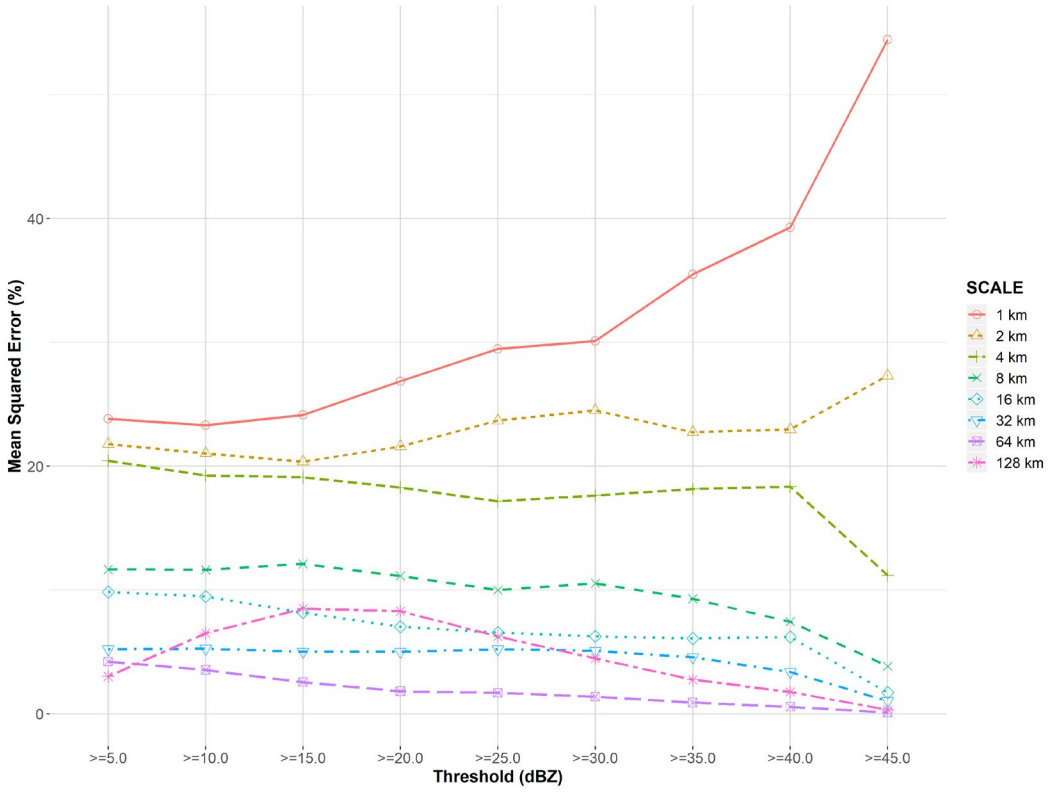

Fig. 14 Plot of MSE% for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

The MSE% values as a function of threshold and error scale show the expected tendency of smaller errors occurring for larger scales and larger errors occurring for the smallest scales, especially at the higher thresholds. The plot enables easier comparison of the variability of MSE% with scale and threshold. The MSE% for

forecast reflectivity decreases with increasing scale, which is expected, but the exception is for scale = 128 km for which the MSE% was higher than that of the second largest scale = 64 km. The highest MSE% was associated with the smallest scale and highest threshold, which is the expectation (NCAR 2016).

Table 4 is a table of the En2 bias as a function of threshold value and error scale. Figure 15 is a plot of the En2 bias at all thresholds and scales.

Table 4 En2 bias for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

Threshold (dBZ)	1km	2km	4km	8km	16km	32km	64km	128km
>=5	-0.05	0.01	-0.07	-0.14	-0.21	-0.38	0.05	-0.33
>=10	-0.14	-0.07	-0.14	-0.17	-0.27	-0.45	-0.13	-0.50
>=15	-0.29	-0.14	-0.18	-0.20	-0.32	-0.50	-0.23	-0.59
>=20	-0.45	-0.27	-0.22	-0.17	-0.35	-0.57	-0.32	-0.65
>=25	-0.53	-0.37	-0.29	-0.12	-0.34	-0.62	-0.41	-0.68
>=30	-0.59	-0.49	-0.25	-0.15	-0.40	-0.68	-0.45	-0.71
>=35	-0.60	-0.41	-0.28	-0.31	-0.52	-0.74	-0.45	-0.75
>=40	-0.71	-0.69	-0.58	-0.64	-0.82	-0.84	-0.64	-0.94
>=45	-0.91	-0.70	-0.72	-0.78	-0.91	-0.91	-0.80	-0.98

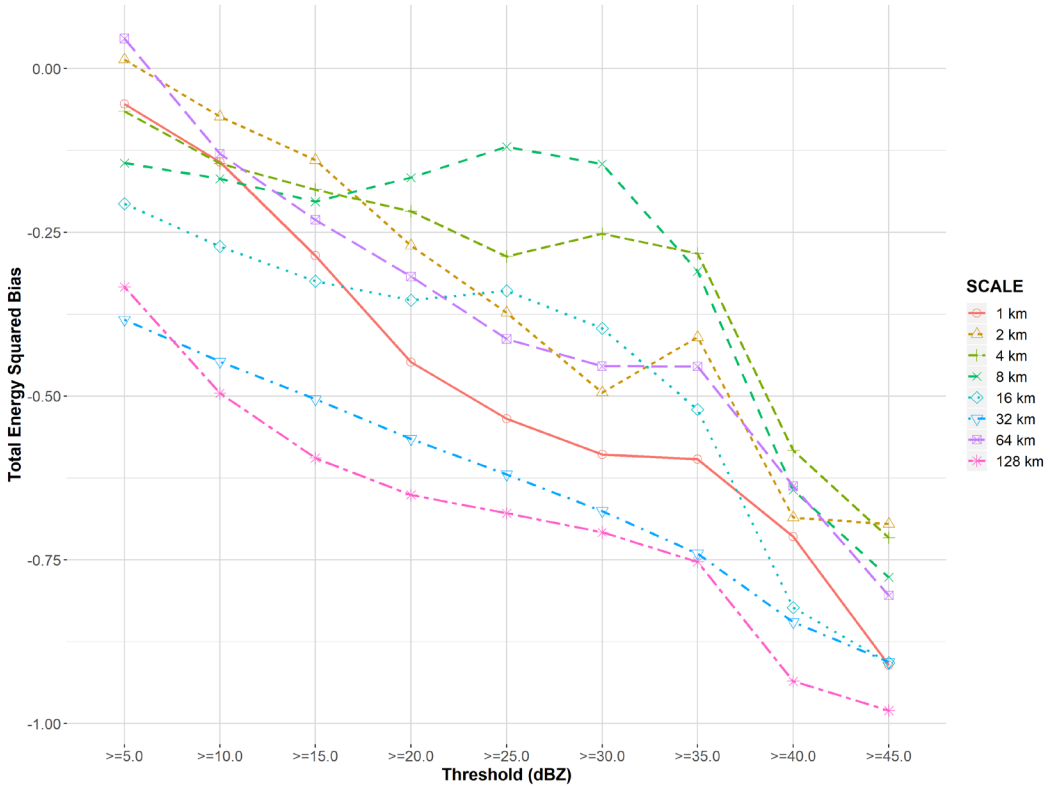

Fig. 15 Plot of En2 bias for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

En2 provides information on the number of events present in the forecast and observed reflectivity fields as a function of threshold and scale. Low thresholds typically have greater energy and high thresholds typically have less energy because there are many events for low thresholds and fewer events for high thresholds (Casati 2010). For this reason, En2 is not used for information on the distribution of energy across scales, but it can be used to obtain bias information across the scales and thresholds. The En2 bias, often referred to as energy relative difference, is computed by taking the difference between the forecast and observed energy divided by their sum and computed at each grid cell. En2 bias ranges in value from –1 to +1. Values greater than zero indicate overforecast, and values less than zero indicate underforecast (NCAR 2016). In Table 4 the values are mostly negative, indicating that the model underforecasted the reflectivity for all threshold values ≥10 dbZ. For threshold ≥5 dbZ there were two scales that had a very small positive bias. The plot shows that the magnitude of the underforecast bias at all scales increases with increasing threshold value.

Table 5 is a table of the forecast En2% as a function of threshold value and error scale. Figure 16 is a plot of the forecast En2% at all thresholds and scales.

Table 5 Forecast En2% for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

Threshold (dBZ)	1km	2km	4km	8km	16km	32km	64km	128km
>=5	12.21	13.08	13.96	12.44	16.31	6.49	7.18	18.33
>=10	12.42	13.41	14.47	14.50	17.27	6.77	6.15	15.01
>=15	12.12	13.51	16.48	16.58	17.39	6.66	5.39	11.87
>=20	11.27	13.31	18.63	20.79	17.02	5.75	4.40	8.83
>=25	11.60	14.68	19.85	23.15	16.85	4.44	3.30	6.13
>=30	11.72	14.22	24.51	25.94	13.82	3.11	2.66	4.02
>=35	17.31	19.95	28.49	18.84	9.31	2.01	1.87	2.22
>=40	31.33	23.02	26.88	11.47	4.10	1.71	1.01	0.48
>=45	25.00	49.11	19.42	4.85	0.93	0.51	0.13	0.04

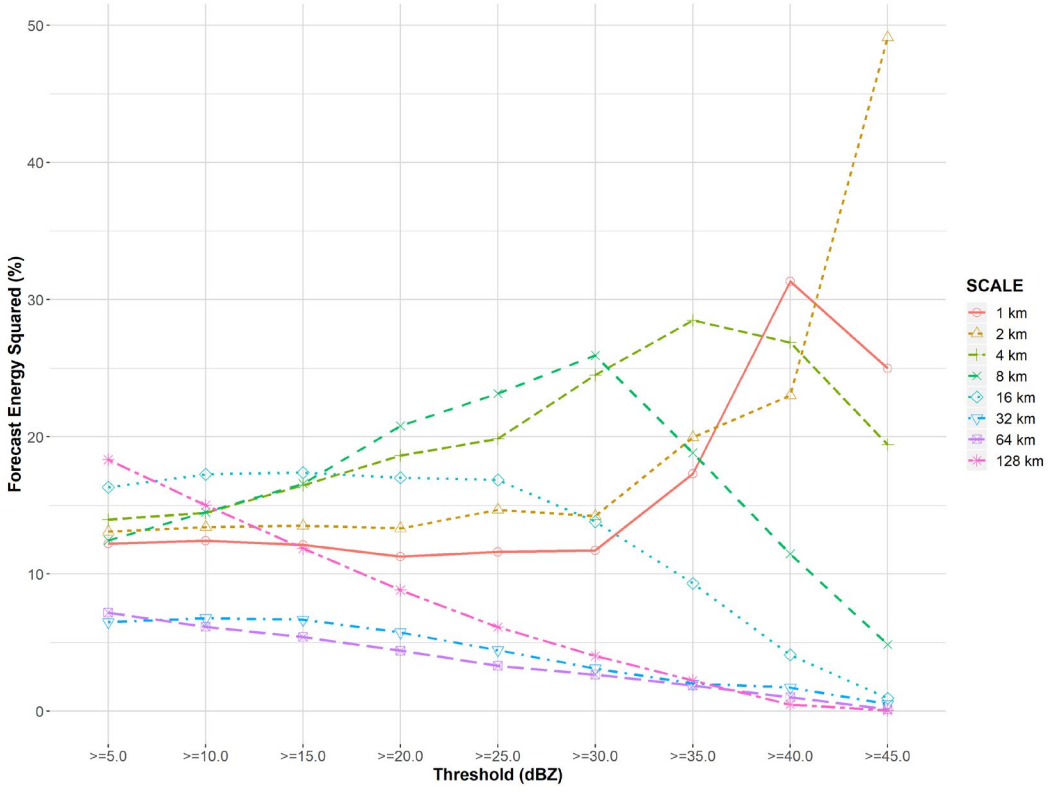

Fig. 16 Plot of forecast squared energy percentage (En2%) for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

En2% is the percentage that the En2 for each scale contributes to the total En2 computed for each threshold. Energy percentages provide information on how the events are distributed across the scales (Casati 2010). This avoids the threshold

dependency issues encountered with using En2. This is caused by low thresholds having many events and high thresholds having few events. The tabular values show that for higher thresholds the energy percentage is usually larger on small scales, since intense events are associated with small-scale features such as convective cells or showers. Low thresholds exhibit most of the energy percentage on large scales (and less percentage on the small scales), since low thresholds are associated with large-scale features, such as fronts (NCAR 2016) or larger, more organized convective activity more likely in this case study. The plot enables easier comparison of the variability of forecast En2% with scale and threshold. Most energy percentage at small scales (≤4 km) is associated with the higher thresholds and the peak of the energy percentage for larger scales occurs at lower thresholds, so that for the largest scales the peak in energy percentage occurs at the lowest threshold values. This seems to indicate that the model predicts high-threshold events such as convective cells at the small scales, which is the expected behavior. For scales ≥16 km the energy percentage is relatively larger for the lower-threshold values compared with the higher-threshold values because generally low-threshold features are low-intensity events associated with relatively larger areas of light precipitation. For scales ≤16 km, there is more energy percentage at higher thresholds than at lower thresholds because generally high-threshold features identify mainly small-scale, intense events associated with convective cells or showers. The significant observation here is that the model seems to be capable of forecasting small-scale convective precipitation events, which is a desirable skill for this project.

Table 6 is a table of the observed En2% as a function of threshold value and error scale. Figure 17 is a plot of the observed En2% at all thresholds and scales.

Table 6 Observed En2% for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

Threshold (dBZ)	1km	2km	4km	8km	16km	32km	64km	128km
>=5	9.62	9.00	11.25	11.76	17.53	10.29	4.63	25.92
>=10	9.61	9.02	11.24	11.82	17.51	10.30	4.64	25.86
>=15	11.00	9.02	12.08	12.63	17.19	10.20	4.35	23.54
>=20	13.60	10.66	13.34	13.39	16.39	9.52	3.91	19.19
>=25	16.72	14.05	15.67	12.87	14.94	8.27	3.47	14.01
>=30	18.76	17.38	16.95	14.38	13.22	6.65	2.93	9.73
>=35	25.67	17.90	19.11	13.40	11.07	5.07	1.87	5.92
>=40	34.35	22.50	18.60	9.64	7.71	3.72	0.83	2.64
>=45	53.57	27.32	11.74	3.86	1.92	1.04	0.12	0.43

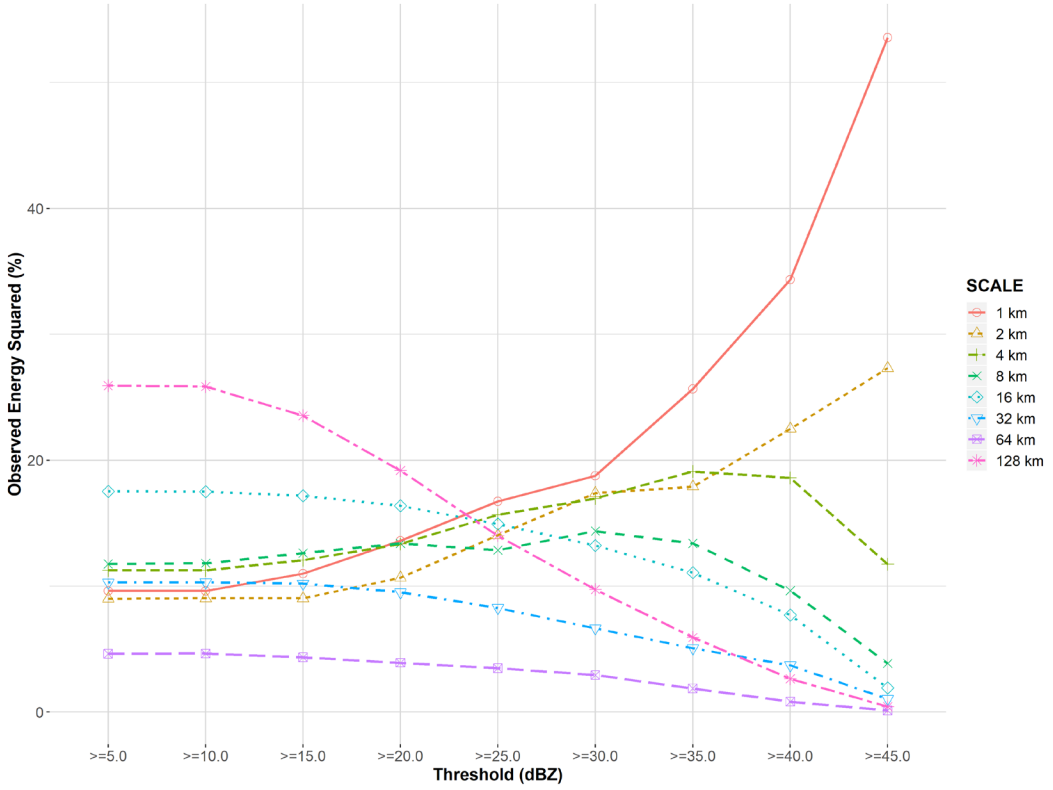

Fig. 17 Plot of observed En2% for spatial scales 1–128 km and threshold values from ≥5 to ≥45 dBZ

The same expectations apply for observed reflectivity. For higher thresholds, the energy percentage is usually larger on small scales since intense events are associated with small-scale features such as convective cells or showers. Low

thresholds exhibit most of the energy percentage on large scales (and less percentage on the small scales) because low thresholds are associated with large-scale features such as fronts (NCAR 2016) or larger, more-organized convective activity more likely in this case study. The plot enables easier comparison of the variability of observed En2% with scale and threshold. The distribution of energy percentage among the various scales for each threshold is very similar to that of the forecast energy, with most energy percentage at small scales (≤4 km) associated with the higher thresholds, and the peak of the energy percentage for larger scales occurring at lower thresholds. Thus, for the largest scales, the peak in energy percentage occurs at the lowest-threshold values. In general, these results indicate the model is doing a good job of forecasting small-scale convection since the bulk of the energy is at the small scales for the higher thresholds resembling the energy distribution in the observations.

A comparison of the forecast and observed energy percentage at specific scales was conducted to contrast the performance of the WRE-N for small-scale and large-scale events. Figure 18 is a plot of the forecast and observed En2% at all thresholds for the error scales of 4 and 128 km. Recall that En2% provides the distribution of energy as a function of scale for a given threshold.

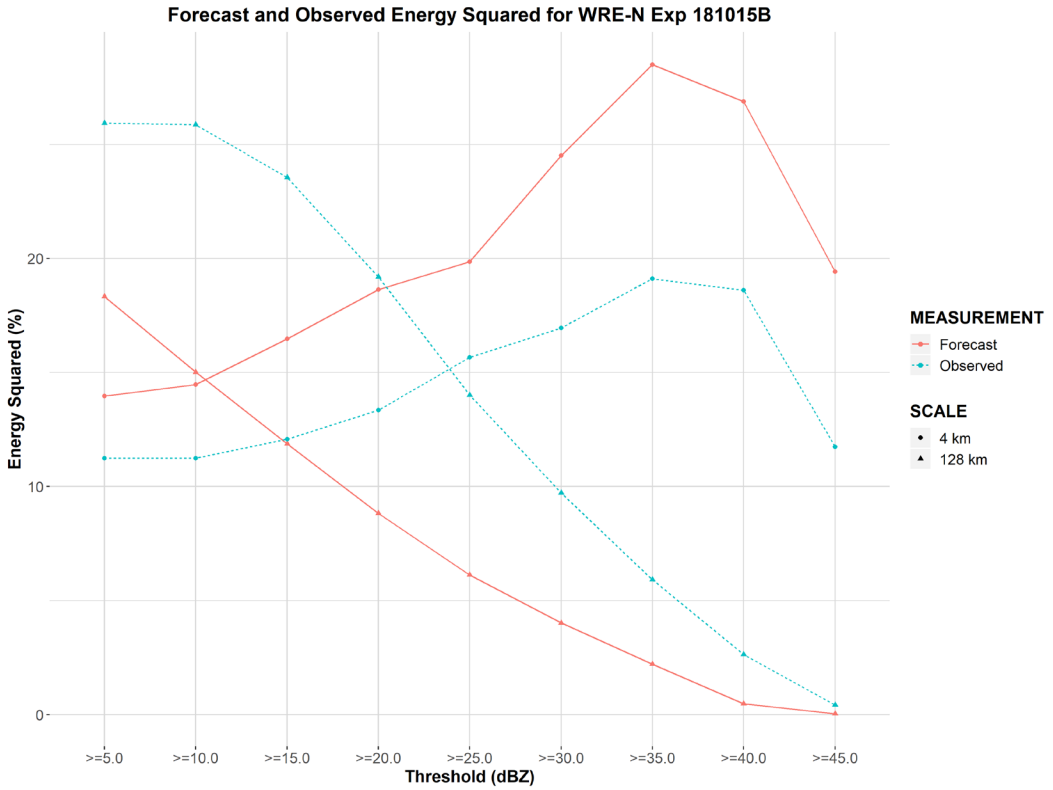

Fig. 18 Plot of forecast and observed En2% for spatial scales 4 and 128 km and threshold values from ≥5 to ≥45 dBZ

For purposes of this discussion, 4 km is considered representative of the small-scale events and 128 km is considered representative of the large-scale events. The distribution of energy (En2%) shows that for small-scale events, most of the energy percentage is at the higher thresholds associated with convective cells. The distribution of energy (En2%) shows that for the large-scale events most of the energy percentage is at the lower thresholds. Comparing the magnitude of the energy percentage in the large-scale (128 km) and small-scale (4 km) peaks, a significant portion of the energy at the thresholds where the respective peak occurs is associated with large-scale (128 km) and small-scale (4 km) events, respectively.

5. Summary and Conclusion

The ISS provides an assessment of the performance of the WRE-N as a function of threshold value and error scale size. It showed good skill at all thresholds for error scales of 8 km or greater. For all thresholds at or below the 2-km scale, the FSS reveals degraded skill, indicating some difficulty in handling small-scale reflectivity structure. Analysis of the ISS variability over ranges of threshold value and scale size reveals a possible small displacement error for errors with scale size of 4 km for reflectivity exceeding 40 dBZ. The frequency bias shows a general

underforecast tendency for reflectivity at all scales and thresholds that increases in magnitude with increasing threshold value; however, it does not have dependency on scale. En2 bias shows the bias at most scales and thresholds is negative, consistent with the frequency bias. MSE% shows decreasing magnitude of error with increasing scale and decreasing threshold consistent with expectations. En2% forecast shows that the WRE-N predicts more energy percentage present for small-scale convective precipitation events (≤4 km) than for larger-scale events based on comparison of magnitude of En2% over the range of scales and thresholds. This is consistent with the case study weather situation and is consistent with the obvious skill in predicting the convective feature shown in Fig. 2. The distribution of the En2% for the observations shows a strong resemblance to that from the forecast, which is also evidence of good skill.

In summary, the use of the Intensity-Scale verification technique in MET W-S provided some valuable insight into model performance not achievable using other MET tools. It provides assessment of model skill and errors as a function of threshold value and spatial scale by providing a skill score (ISS) and MSE for each combination of threshold and scale. This technique can enable the separation of the larger-displacement errors from the smaller-scale errors attributable to smaller-scale processes by analysis of the distribution and sign of ISS over all scales and thresholds. Use of this technique also enabled the determination of bias at specific scales and thresholds using metrics such as frequency bias, En2 bias, and En2. It provides a means to evaluate the distribution of En2 of forecast and observations leading to increased understanding of model handling of weather situation.

6. References

Casati B, Ross G, Stephenson D. A new Intensity-scale approach for the verification of spatial precipitation forecasts. Meteor Appl. 2004;11:141–154.

Casati B. New developments of the intensity-scale technique within the spatial verification methods intercomparison project. Wea Forecast. 2010;25:113–143.

Dumais RE Jr, Kirby SF, Passner JE, Knapp DI. Development of a nowcast modeling system to support Army aviation forecasts. Presented at the 16th Conference on Aviation, Range, and Aerospace Meteorology, 93rd Annual American Meteorological Society Conference; 2013 Jan 6–10; Austin, TX.

Ebert E. Fuzzy verification of high resolution gridded forecasts: a review and proposed framework. Meteorol Appl. 2008;15:51–64.

[NCAR] National Center for Atmospheric Research. Model evaluation tools (MET). Version. 5.2. Boulder (CO): NCAR; 2016 [accessed 2019 Aug 15]. https://dtcenter.org/community-code/model-evaluation-tools-met.

Reen BP, Cai H, Raby J. Radar assimilation over the Kwajalein Atoll. Adelphi (MD): CCDC Army Research Laboratory (US). In preparation.

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, Huang XY, Wang W, Powers JG. A description of the advanced research WRF version 3. Boulder (CO): National Center for Atmospheric Research; 2008. Report No.: NCAR/TN-4751STR.

List of Symbols, Abbreviations, and Acronyms

2-D	two-dimensional
AGL	above ground level
ARL	CCDC Army Research Laboratory
ARW	Advanced Research Version of Weather Research and Forecasting
dBZ	decibels of radar reflectivity
En2	energy squared
En2%	energy squared percent
F-O	forecast minus observation
FSS	Fractions Skill Score
ISS	Intensity Skill Score
MET	Model Evaluation Tools
MM5	Fifth generation Penn State/NCAR mesoscale model
MSE	mean squared error
MSE%	mean squared error percent
NCAR	National Center for Atmospheric Research
NWP	Numerical Weather Prediction
UTC	Coordinated Universal Time
W-S	Wavelet-Stat
WRE-N	Weather Running Estimate–Nowcast
WRF	Weather Research and Forecasting
WRF-ARW	Weather Research and Forecasting, Advanced Research WRF
WSM6	WRF Single Moment 6 Class
YSU	Yonsei State University

1 (PDF) DEFENSE TECHNICAL INFORMATION CTR DTIC OCA

1 (PDF) CCDC ARL FCDD RLD CL TECH LIB

1 (PDF) US NAVY RSRCH LAB DR J MCLAY

1 (PDF) US AIR FORCE 557TH WEATHER WING R CRAIG

1 (PDF) DCGS-A WEATHER EET LEAD J CARROLL

3 (PDF) UCAR T JENSEN J H GOTWAY B BROWN

2 (PDF) EARTH SYS RSRCH LAB M WANDISHIN K FENTON

10 (PDF) CCDC ARL FCDD RLC E B MACCALL T JAMESON FCDD RLC ED G VAUCHER FCDD RLC EM H CAI E CHIN R DUMAIS J PASSNER J RABY B REEN J SMITH