# STORM CELL DISPLACEMENT ERRORS IN THE WARN-ON-FORECAST SYSTEM

Brina M. Lemke[1, 2], Derek R. Stratman[3, 4], and Corey K. Potvin[4, 5]

[1]National Weather Center Research Experiences for Undergraduates Program
Norman, Oklahoma
[2]University of California, Los Angeles
Los Angeles, California
[3]Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma
Norman, Oklahoma
[4]NOAA/OAR/National Severe Storms Laboratory
Norman, Oklahoma
[5]School of Meteorology, University of Oklahoma
Norman, Oklahoma

ABSTRACT

The Warn-on-Forecast System (WoFS) is a rapidly updating ensemble data assimilation and prediction system that provides probabilistic forecasts of severe weather and stands to increase lead times for localized severe weather forecasts. We use matched object pairs between WoFS-forecast storms and Multi-Radar Multi-Sensor (MRMS)-observed storms, then filter for storms with cell-like characteristics in order to study patterns in storm cell centroid displacements. Identifying patterns and correlations in these storm cell displacement errors could provide insight into their causes and be used to increase forecast accuracy or provide compensation for inaccuracies. This study found that there is little variation in displacement errors across years, ensemble members, or planetary boundary layer schemes. However, error spread does increase as lead time increases and minimum error spread for shorter lead times occurs within 1-3 hours of a storm's assimilation. Storms forming after the initialization time in a forecast also have the smallest bias. Additionally, higher mean intensity storms were found to have larger northeast biases, a correlation that has been previously observed and documented and which could be indicative that the WoFS is underpredicting deviant motion in rotating storms.

## 1. INTRODUCTION

At present, there are two types of public severe weather alerts: watches and warnings. While watches are based on forecasts and cover large areas wherein severe weather may be expected over a period of a few hours, warnings are currently only based on observed conditions and are localized to a presently occurring individual storm or cluster of storms. Between 2012 to 2016, the mean tornado lead time was 15.6 minutes (Brooks and Correia 2018). Unfortunately for many, this may only be the minimum amount of lead time necessary for an adequate response, especially for tornado warnings. People in need of longer warning times include mobile home residents who need to seek alternative shelter, hospital staff who need to move vulnerable patients and prepare for a potential influx of patients, and emergency services, who need to prepare to make a timely response (Hoekstra et al. 2011; Stensrud et al. 2009). Likewise, advance notice of severe thunderstorms gives residents time to prepare for potential hail, flooding, lightning, and damaging wind threats. Although, while watches are meant to encourage people to take precautions, it is not feasible for such large populations—many of whom may be impacted minimally or not at all—to brace themselves for the worst possible impacts every time a severe weather watch is issued. As a result, many residents of severe weather-prone areas are desensitized to watches and may take little notice, if any, until a warning is issued, by which time it may be too late to react adequately before impact.

This is where the Warn-on-Forecast System (WoFS; Jones et al. 2020) may be able to provide a solution. The WoFS is a convective-scale ensemble forecast system capable of producing output that allows for more localized severe weather alerts with longer lead times, closing the gap between watches and warnings. This would open the possibility of warnings being issued based on numerical model forecasts rather than waiting for observed conditions (Stensrud et al. 2009; Stensrud et al. 2013).

However, there are still improvements to be made in its accuracy; there are often notable errors in forecast storm locations when compared to observed

[1] *Corresponding author address:* University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095
Email: blemke@g.ucla.edu

outcomes, known as storm displacement errors (SDEs; Britt et al. 2023; Flora et al. 2019; Skinner et al. 2018). SDEs can also self-perpetuate, directly impacting the WoFS by resulting in suboptimal data assimilation and thus leading to larger storm location errors in successive forecasts (Stratman et al. 2018; Stratman and Potvin 2020). Errors in any system used for official forecast output could risk improperly informed populations. Since the WoFS is especially intended for storm-scale forecasting, accuracy—even on the scale of only a few kilometers—is essential. Additionally, the more accurate and precise a forecast system proves itself to be, the more trust the forecasters will have in its projections and the more likely they are to use it in operational settings.

Finding and describing patterns (or a lack thereof) and identifying potential (as well as unlikely) sources of SDEs in WoFS output are important steps toward improving WoFS accuracy. Likewise, giving forecasters a better understanding of these SDEs will provide a basis for manual correction or compensation before the dissemination of information to the public.

## 2. DATA AND METHODS

### 2.1 *WoFS Output*

The WoFS is a rapidly updating ensemble data assimilation and prediction system designed to provide probabilistic forecasts of severe and hazardous weather with 3-km horizontal grid spacing over a 900 x 900-km square domain. It uses 36 ensemble members for data assimilation and 18 members for forecasts. Different initial and boundary conditions along with various physics scheme combinations are used to create the ensemble diversity, including two shortwave and longwave radiation schemes and three planetary boundary layer (PBL) schemes: the Yonsei University (YSU) scheme, the Mellor-Yamada-Jankic (MYJ) scheme, and the Mellor-Yamada-Nakanishi-Niino (MYNN) scheme. It has been shown that the forecast differences arising from the PBL schemes dominate those arising from the radiation schemes (e.g. Potvin et al. 2020), so only PBL scheme differences are examined further in this study. More information about the WoFS configuration can be found in Jones et al. (2020).

Designed for short-term predictions, the WoFS produces ensemble forecasts every 30 minutes with lead times up to 6 hours and output in 5-minute increments. The forecasts are initialized from ensemble analyses—which are created from the assimilation of the latest observations, such as radar, satellite, and conventional observations—every 15 minutes using an ensemble data assimilation method. For this study, forecast composite reflectivity, which is the maximum reflectivity in a model column, will be used to assess the SDEs.

### 2.2 *MRMS Observations*

WoFS output is verified using observational data from the Multi-Radar Multi-Sensor (MRMS) System (Smith et al. 2016), an operational product on a 1-km CONUS-sized grid. The MRMS system combines radar data from operational weather radars, such as WSR-88D radars, to generate multiple products, including a CONUS-wide composite reflectivity field. The 1-km composite reflectivity is interpolated to the 3-km WoFS grid using the Cressman method before being used to create observed storm objects.

### 2.3 *Object-Based Analysis*

Skinner et al. (2018) demonstrated the utility of using object-based verification to assess WoFS efficacy. Thus, in order to quantify SDEs, we use object-based analysis to identify storm objects in WoFS composite reflectivity forecasts and MRMS composite reflectivity observations. Several criteria, such as intensity and area size, are used to define these storm objects. An MRMS (WoFS) storm object must have a minimum intensity threshold of 38 dBZ (43 dBZ) and its maximum intensity must exceed 43 dBZ (48 dBZ; Skinner et al. 2018). A percentile threshold is used to account for the WoFS's high bias in comparison to the MRMS, so as a result, different intensity threshold values are used in storm object identification for WoFS and MRMS (Guerra et al. 2022).

These objects can then be matched at each forecast output time using a Total Interest (TI) score, defined as

$$\mathrm{TI}_{\mathrm{match}} = 0.5\left[\frac{(D_m - D_{\mathrm{min}})}{D_m} + \frac{(D_m - D_{\mathrm{cnt}})}{D_m}\right], \qquad (1)$$

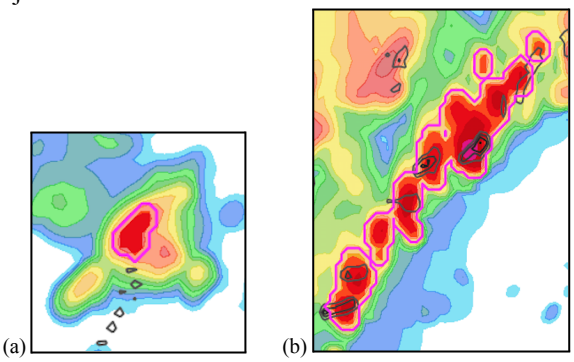
in which $D_m$ is the constant distance threshold for matching, $D_{min}$ is the minimum boundary distance between a pair of storms, and $D_{cnt}$ is the distance between object centroids. This score lies on a scale of 0 to 1, with smaller $D_{min}$ and $D_{cnt}$ values equally responsible for generating higher scores, which in turn indicate higher favorability of a match. MRMS-WoFS storm pairs are defined by the highest TI scores between objects and with a minimum TI score threshold of 0.2 (Guerra et al. 2022). We define SDEs as the difference in storm object centroid location between

WoFS and MRMS objects in a matched storm pair (Skinner et al. 2018).

### 2.4 *Storm Object Filtering*

There is less predictability with increased lead time, so we will consider only storm objects from forecasts with a lead time of less than 3 hours. This decreases the risk of any results being muddled or even obscured by uncertainty at longer lead times.

Likewise, we will examine discrete storm cell behavior exclusively, as SDEs in discrete cells tend to be more distinct and consistent. Discrete cells are generally more isolated and have smaller, more circular shapes with an order of magnitude of 100 km$^2$ (Fig. 1a), whereas mesoscale convective systems (MCSs) are much larger and less circular. As a result, small variations in MCS object shapes can make their geographic centroids jump around erratically between forecast times. In particular, quasi-linear convective systems (QLCSs) also have high eccentricities (are much thinner and longer), so small changes in the major axis can make this especially true (Fig. 1b). These inconsistencies make centroid-based SDEs less indicative of overall forecast location accuracy. Thus, to avoid the influence of possibly different dynamics and SDE patterns associated with QLCSs or other non-cellular behavior, it is necessary to attempt to eliminate these discrepancies by isolating these cell-like objects from the data.



(a) (b)

**Figure 1.** Example of (a) a standard storm cell reflectivity signature and (b) a typical QLCS reflectivity signature. Adapted from Potvin et al. (2022).

In order to do this, we first apply a maximum area threshold of 5000 km$^2$. We then apply the preliminary cell classification technique presented by Potvin et al. (2022), with the exception of the solidity criteria, as solidity was not a variable included in or calculable with this study's data. The remainder of the classification technique is unchanged and applied both to WoFS objects and MRMS objects, in order to avoid potentially including, for example, a QLCS that has been matched to a cell-like object. Only matched pairs between WoFS and MRMS objects that are both identified as cell-like are included in this study.

| | | WoFS | | |
|---|---|---|---|---|
| | | Cell-like | Other | |
| MRMS | Cell-like | **2,451,246** | 725,168 | 3,176,414 |
| | Other | 1,596,205 | 907,639 | 2,503,844 |
| | | 4,047,451 | 1,632,807 | 5,680,258 |

**Table 1.** Incidence of matched object pair classifications for both WoFS and MRMS using a 5000-km$^2$ area maximum and the Potvin et al. (2022) preliminary classification process.

Out of 5,680,258 object pairs, 2,503,844 object pairs are disqualified by applying all criteria to MRMS objects alone, while 725,168 additional object pairs are disqualified when applying the criteria to WoFS objects as well (Table 1). 907,639 objects are disqualified by both MRMS and WoFS, indicating these are likely accurately matched QLCS-like/other objects and likewise will not be considered. This leaves 2,451,246 matches between cell-like objects, which we will examine further.

### 2.5 *Evaluation Methods*

This study uses two measurements of SDE: bias and root-mean-square error (RMSE). Bias measures the mean displacement between the forecast and observed centroids, including both direction and magnitude, while

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(f_k - o_k)^2} \qquad (2)$$

provides a measure of the spread of forecast object displacements around their observed objects. Here, $n$ is the sample size and $f_k$ and $o_k$ are the forecast and observed locations ($f_k - o_k$ represents the SDEs).

Subjects of interest in this study include yearly variations, differences across forecast ensemble members and their PBL schemes, lead time, relative storm age, and storm mean intensity. While the absolute storm age at a given valid forecast time is the time between an observed storm's formation and the valid forecast time in question, relative storm age is the time between initialization and an observed storm's formation (Eq. 3). When considering relative storm age, the time at which a forecast will be valid is irrelevant; we are instead interested in correlations between the

age a storm has already reached by "now" (initialization) as opposed to the age a storm "will be" (valid time). As a result, relative storm age can be negative for storms projected to form after the initialization time. Note that storms with unknown formation times (i.e. forming before a WoFS run begins) instead use the earliest initiation time available.

$$relative\ storm\ age\ =\ absolute\ storm\ age\ -\ lead\ time \quad (3)$$

The two main visualizations used in this study include density plots and heat maps with vector map overlays. The density plots illustrate SDEs with i- and j-coordinate displacements from the origin while using a logarithmic color scale to indicate the proportional density of data points within 5-km pixels. These also indicate bias with a black x and RMSE as the radius of a black circle.

The heat maps show lead time on the x-axis in 15-minute intervals, with the other variable having intervals on the y-axis. The color scale indicates the RMSE values and the vectors indicate the direction and magnitude of bias, with these values determined by the mean RMSEs and biases over the x- and y-intervals. Each pixel has a minimum threshold of 200 data points.

## 3. RESULTS

### 3.1 *By Year*

There is little variation across the density plots for 2017, 2018, 2019, and 2020 (Fig. 2). All density plots show a radially symmetric distribution of SDEs around the origin, with the logarithmically-scaled proportion increasing steadily with decreasing SDE radius. This is despite the variation in sample size across years, with 300,593 instances in 2017, 692,868 in 2018, 830,428 in 2019, and 627,357 in 2020. RMSE values are similar across years and while there is some variation in bias, namely between 2017 and the other years, this can likely be attributed in part to the smaller 2017 sample size (less than half of the others).

Other potential affecting factors include that WoFS was only run during May and June, two of the most active months for severe weather, in 2017, while the other three years encompass events from May-July (2018), April-August (2019), and April-September (2020). There is generally also substantial variation in weather patterns between years, including predominant directions of propagation. If the direction of storm propagation were to have a substantial impact on bias and there was a contrasting common direction of propagation in May and June of 2017 compared to other years, this could account in part for the difference.

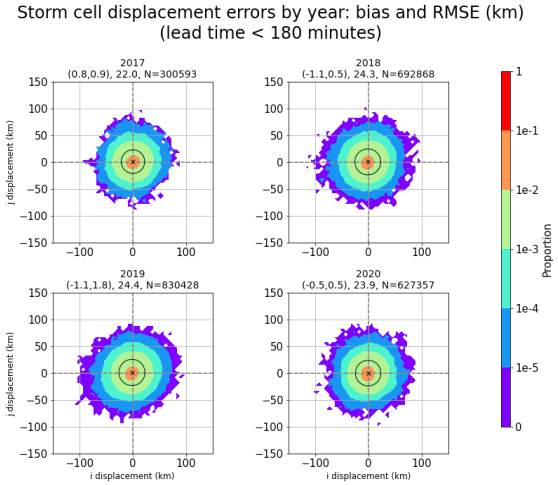



**Figure 2.** Storm cell displacement errors by year, including bias and RMSE, for the years 2017-2020. A logarithmic color scale indicates the proportional data distribution while bias is marked with a black x and the radius of the black circle indicates the RMSE.

There was also considerable variation in WoFS operations practices throughout this four-year period, which could likewise contribute to the differences. Notably, while the data assimilation software used in the WoFS was changed between 2018 and 2019, there is no clear difference in error patterns between the former two years and the latter.

### 3.2 *By Ensemble Member and PBL Scheme*

The density plots for each of the 18 forecast ensemble members show the proportional distribution is radially symmetrical with the logarithmically-scaled proportion increasing steadily with decreasing SDE radius. The variation in bias and RMSE across ensemble members is extremely minimal (Fig. 3). The overall SDEs for the MYNN scheme are the smallest, with a bias of (-0.6, 0.8) and an RMSE of 23.8. The overall SDEs for the YSU scheme are the largest, with a bias of (-0.7, 1.5) and an RMSE of 24.2 (Table 2). However, since no significance testing was performed, these small variations may not be significant.

| | YSU | MYJ | MYNN |
|---|---|---|---|
| Sample Size | 805,694 | 828,317 | 817,235 |
| Bias (km) | (-0.7, 1.5) | (-0.8, 0.7) | (-0.6, 0.8) |
| RMSE (km) | 24.1 | 23.9 | 23.8 |

**Table 2.** Storm cell displacement sample size, bias, and RMSE by PBL scheme.

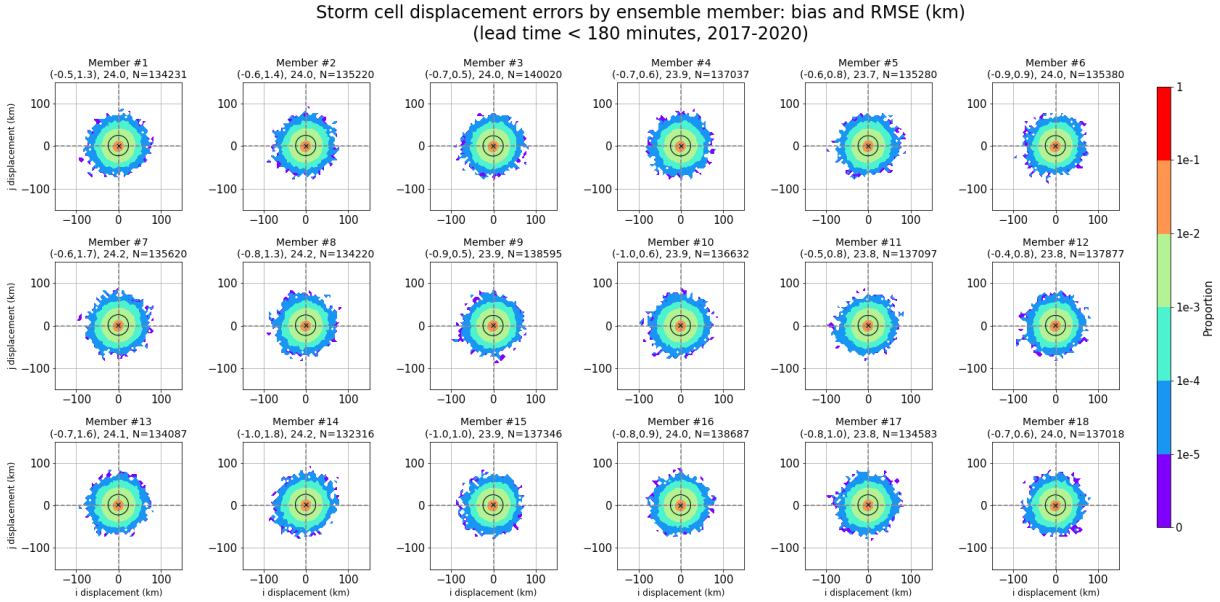



**Figure 3.** Storm cell displacement errors (#1-#18) by ensemble member and by PBL scheme: (#1-2, #7-8, #13-14) YSU, (#3-4, #8-9, #16-16) MYJ, and (#5-6, #11-12, #17-18) MYNN. A logarithmic color scale indicates the proportional data distribution while bias is marked with a black x and the radius of the black circle indicates the RMSE.

### 3.3 *By Relative Storm Age*

We can examine RMSEs with respect to relative storm age by taking the 10-minute rolling mean RMSE over multiple lead time intervals (note that these intervals are not of equal size). This shows that there is a clear increase in RMSE with increasing lead time (Fig. 4). For lead time intervals of less than 90 minutes, there is also an immediate, even dramatic, decrease in RMSE with relative storm age once a storm is assimilated (i.e. as relative storm age becomes greater than 0 minutes), at which point a few data-assimilating cycles have passed and allowed WoFS to integrate the now-observed storm. This minimum RMSE range occurs approximately between relative storm ages of 60 and 180 minutes before RMSE values begin to increase more dramatically again. This minimum RMSE range does not clearly exist for the lead time interval of [90,120) minutes. Still, there is a more delayed minimum RMSE range for lead time intervals of [120,180) minutes, beginning to decrease around relative storm ages of 60 minutes, with the minimum RMSE range occurring between relative storm ages of 120 and 240 minutes.

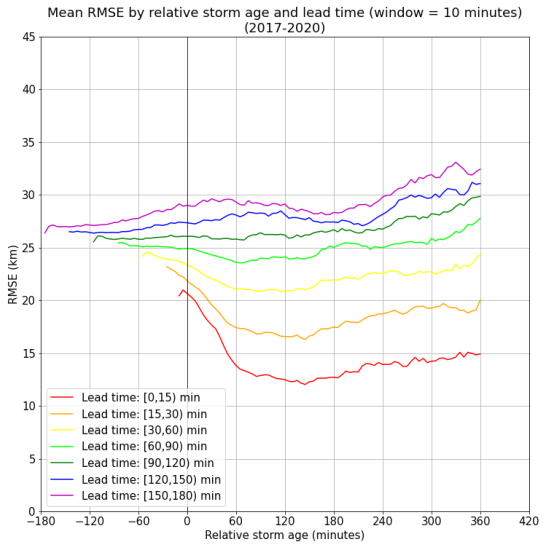



**Figure 4.** Mean RMSE by relative storm age and lead time intervals with a rolling mean window of 10 minutes.

Surprisingly, RMSE increases near-monotonically with relative storm age beyond these RMSE minimums; the oldest storms have higher RMSE values despite having the most time to be well-assimilated. This may result from storms increasing in size with time, causing proportionately larger SDEs: enough to offset the benefit of their age. However, for lead time intervals of less than 60 minutes, universal RMSE maximums occur for storms with the smallest relative storm ages, likely due to the lack of certainty resulting from not yet being well-assimilated.

Patterns in RMSE values are also obvious in the heat map visualization. Absolute RMSE minimums occur at less than 15 minutes of lead time for storms

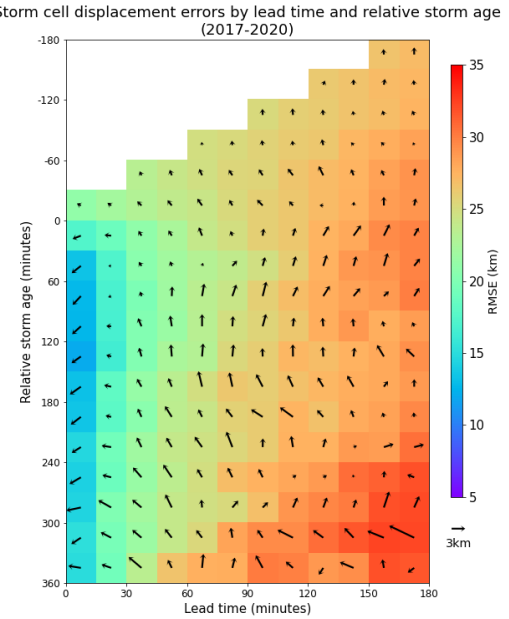



**Figure 5.** Storm cell displacement errors by relative age and lead time. A color scale represents the RMSE values while vector direction and magnitude correspond to bias.

with relative ages between 60 and 180 minutes and RMSE increases overall with lead time (Fig. 5). Yet, this is not where the smallest biases occur. Storms forming after initialization (having a negative relative storm age) have the smallest biases, which is to be expected; the storms have not yet been assimilated so there is no obvious reason SDEs would be favored to occur in any particular direction. As a result, biases become larger primarily after the storms have been assimilated. Still, even unassimilated storms have increasing RMSE with lead time. The largest RMSEs occur for the oldest storms at the longest lead times just as in the line plot (Fig. 4).

We note that for lead times greater than or equal to 15 minutes, there is a northerly bias for well-assimilated storms (i.e. storms with a relative age greater than 0 minutes). Though it could be argued that this bias somewhat exists among unassimilated storms, their biases are small enough for this to be negligible in comparison to well-assimilated storms. Finally, there is a generally increasing easterly bias with increasing lead time for well-assimilated storms. For the [0,15)-minute lead time interval, however, there is a clear southwesterly bias.

### 3.4 *By Mean Intensity*

It is clear from Figure 6 that objects with shorter lead times and higher mean intensities have the smallest RMSE values. This fits with previous observations, as storms are likely to have been well-assimilated by the time they would reach (or be forecast to reach) higher intensities and the previous section shows that well-assimilated storms (especially within 60-180 minutes of assimilation) have the lowest RMSEs. Storms of higher mean intensity have a distinctly larger northeasterly bias. Northeasterly biases have been observed for supercellular storms in previous storm-scale modeling studies (e.g., Stratman et al. 2018) but also observed in WoFS forecasts (Skinner et al. 2018). This pattern suggests that WoFS may be under-predicting deviant motion in rotating storms. As cyclonic supercellular storms intensify, they tend to veer to the right (in the Northern Hemisphere) from the predominant direction of propagation while their forward motion decreases. Since storms most commonly track in a northeastward direction, if the WoFS is not fully taking this strengthening motion into account, it would explain why it favors a larger northeastward displacement than is observed in storms of these intensities.

Storms with lower mean intensities have larger northwesterly biases. However, the magnitudes of biases and RMSEs at the smallest mean intensities (39-41dBZ) were notably larger before the implementation of storm cell filtering criteria. Additionally, the biases and RMSEs of the lowest intensity interval (39-40dBZ) remain disproportionately large after filtering. This may be an indication that some MCSs are still making it through the filtering process and are at least in part responsible for this bias pattern. A northwesterly bias was likewise observed in QLCSs in Britt et al. (2023), further supporting this

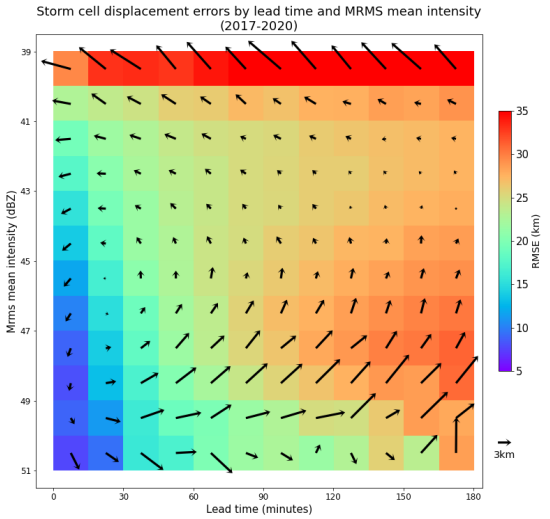



**Figure 6.** Storm cell displacement errors by MRMS mean intensity and lead time. A color scale represents the RMSE values while vector direction and magnitude correspond to bias.

theory. Also, just as in the relative storm age figure, the lead time interval of [0,15) minutes features the most notable southerly biases. We observe that as mean intensity increases in this lead time interval, the westerly bias apparent at the lowest mean intensities becomes more southerly and less westerly, resulting in southeasterly biases at the highest mean intensities (i.e. the j-component decreases and the i-component increases). Finally, overall near-identical patterns in bias and RMSE were found when examining a heat and vector map for maximum intensity rather than mean intensity.

## 4. DISCUSSION AND CONCLUSIONS

The goal of this study was to identify and describe qualities and patterns in the SDEs between WoFS forecasts and MRMS observations of cell-like storms. We looked at SDEs across 2017-2020, forecast ensemble members, PBL schemes, lead times, relative storm ages, and mean intensity. There was little variation across the years, ensemble members, and PBL schemes. However, we found that RMSE increases as lead time increases and that minimum RMSEs occur within 1-3 hours of assimilation. Storms forming after initialization have smaller biases, but their RMSEs follow the overall pattern of increasing with lead time. Higher mean intensity storms also have larger northeasterly biases, which have been previously documented and observed by forecasters. This could be an indication that WoFS is underpredicting deviant motion in rotating storms. Interestingly, lower-intensity storms had a pattern of northwesterly biases, but of smaller magnitude than the higher-intensity northeasterly correlation. This could be accounted for in part by MCSs not being fully filtered out, but the strength of the overall pattern seems to point to the existence of a trend inclusive of storm cells.

Though this study has identified multiple patterns, there is still much room for improvement regarding the methodology. Most notably, the outlying high RMSE values and bias magnitudes in the 39-40 dBZ interval from Figure 6 indicate that there is likely still a considerable number of non-cell-like objects being evaluated. Though the pattern shown in Figure 6 is distinct overall, This is an important reminder that this cell filter is imperfect and that this must be considered when evaluating SDEs in varied contexts. Additionally, this study added a maximum area filter of 5000-km$^2$ to the original Potvin et al. (2022) preliminary cell classification technique. This value was chosen somewhat arbitrarily based on observations of the data spread and a general understanding of what could be considered a reasonable cell area maximum. Further studies may want to consider refining the cell classification approach, including this area maximum, based on a more objective approach.

There is also a potential benefit in conducting another, similar study, but wherein storms are matched only after cell filtering (i.e. matching storm objects within classification groups) rather than calculating TI scores with all objects considered together. This too will not be a perfectly sound approach, as cells can and do merge into MCSs and thus some pairing of cell-like and non-cell-like objects must inevitably occur in true-to-life object matching. Still, this approach may be more successful in avoiding high-likelihood erroneous pairing. For example: a discrete cell object may pair with a QLCS despite a large centroid displacement if they overlap in boundary, causing an inflated TI score.

It may be of interest to study SDEs in relation to aspects such as geographical region, topography, and terrain. However, WoFS data only includes coordinates relative to the domain of observation, not geographic coordinates, making this more difficult to investigate directly. Additionally, this study only examined four years of WoFS outputs, including earlier years during which the system was run far fewer times. Though there is no distinct indication of a significant difference between any of these four years, weather patterns in any one year can vary dramatically from those in another year. Thus having a larger pool of data would contribute to a higher confidence level in the existence of any large-scale patterns, rather than leaving open the possibility of results being somewhat unique to one or more of only a few years studied.

Finally, the findings regarding mean intensity that are suggestive of correlations between deviant motion and SDE characteristics may benefit from being further examined in relation to predominant directions and speeds of storm propagation. The observed directional bias to the northeast in higher mean intensity storms and to the northwest in lower mean intensity storms may be related in part to the most common northeastward direction of severe weather propagation. It would be of interest to examine if these directional bias patterns vary for storms with more irregular directions of propagation, as well as if the bias magnitude and RMSE values likewise vary with the forward speed of propagation.

The patterns (or lack thereof) identified in this work may lead to the identification of potential root causes of SDEs, especially the northeasterly bias with higher mean intensity. A stronger understanding of SDEs not only contributes to efforts to increase WoFS accuracy (e.g. by helping to improve model physics schemes) but also lends context to WoFS output. This will allow end users, such as NWS meteorologists, to

understand and compensate for WoFS errors (e.g. accounting for model biases in short-term forecasts). and disseminate the most accurate, relevant, and potentially life-saving information to the public.

## 5. ACKNOWLEDGMENTS

The corresponding author would like to thank Alex Marmo and Daphne LaDue, co-directors of the Research Experience for Undergraduates at the National Weather Center, for the outstanding work they have put into the program and their support throughout for the students involved, the author included. The corresponding author would also like to thank their mentors, Derek Stratman and Corey Potvin for their guidance and support, especially through trying circumstances, as well as Pat Skinner and Monte Flora for their help and insight.

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

Britt, K. C., P. S. Skinner, P. L. Heinselman, C. K. Potvin, M. L. Flora, B. Matilla, K.H. Knopfmeier, and A. E. Reinhart, 2023: Verification of Quasi-Linear Convective Systems Predicted by the Warn-on-Forecast System. In review, Wea. Forecasting.

Brooks, H. E., and J. Correia, 2018: Long-Term Performance Metrics for National Weather Service Tornado Warnings. *Wea. Forecasting*, **33**, 1501–1511, https://doi.org/10.1175/WAF-D-18-0120.1.

Guerra, J. E., P. S. Skinner, A. Clark, M. Flora, B. Matilla, K. Knopfmeier, and A. E. Reinhart, 2022: Quantification of NSSL Warn-on-Forecast System Accuracy by Storm Age Using Object-Based Verification. *Wea. Forecasting*, **37**, 1973–1983, https://doi.org/10.1175/WAF-D-22-0043.1.

Hoekstra, S., K. Klockow, R. Riley, J. Brotzge, H. Brooks, and S. Erickson, 2011: A Preliminary Look at the Social Perspective of Warn-on-Forecast: Preferred Tornado Warning Lead Time and the General Public's Perceptions of Weather Risks. *Wea. Climate Soc.*, **3**, 128–140, https://doi.org/10.1175/2011WCAS1076.1.

Jones, T. A., and Coauthors, 2020: Assimilation of GOES-16 Radiances and Retrievals into the Warn-on-Forecast System. *Mon. Wea. Rev.*, **148**, 1829–1859, https://doi.org/10.1175/MWR-D-19-0379.1.

Potvin, C. K., and Coauthors, 2020: Assessing Systematic Impacts of PBL Schemes on Storm Evolution in the NOAA Warn-on-Forecast System. *Mon. Wea. Rev.*, **148**, 2567–2590, https://doi.org/10.1175/MWR-D-19-0389.1.

Potvin, C. K., and Coauthors, 2022: An Iterative Storm Segmentation and Classification Algorithm for Convection-Allowing Models and Gridded Radar Analyses. *J. Atmos. Oceanic Technol.*, **39**, 999–1013, https://doi.org/10.1175/JTECH-D-21-0141.1.

Skinner, P. S., and Coauthors, 2018: Object-Based Verification of a Prototype Warn-on-Forecast System. *Wea. Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Stensrud, D. J., and Coauthors, 2013: Progress and challenges with warn-on-forecast. *Atmos. Research*, **123**, 2–16, https://doi.org/10.1016/j.atmosres.2012.04.004.

Stratman, D. R., C. K. Potvin, and L. J. Wicker, 2018: Correcting Storm Displacement Errors in Ensembles Using the Feature Alignment Technique (FAT). *Mon. Wea. Rev.*, **146**, 2125–2145, https://doi.org/10.1175/MWR-D-17-0357.1.