Semi-Annual Report

Better Use of Ensembles in the Forecast Process: Scenario-Based Tools for Predictability Studies and Hazardous Weather Communication

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Collaborative Science, Technology, and Applied Research (CSTAR) Program

(Award Number: NA17NWS4680004)

FINAL REPORT

1 July 2017 – 30 June 2022

PI: Dr. Brian A. Colle and Co-PI: Dr. Edmund Kar-Man Chang

School of Marine and Atmospheric Science

Stony Brook University

State University of New York

Telephone (631) 632-3174

FAX (631) 632-6251

Email: brian.colle@stonybrook.edu

1. Background

Our project addressed CSTAR objectives to: "Improving the lead-time and accuracy of forecasts and warnings for high impact weather -- Improving the use of ensemble predictions systems in order to enable more effective forecaster assessment of uncertainty"; "Improving Impact-Based Decision Support Services"; and "Improving water resource information (precipitation) for decision support and situational awareness" Our focus area is the Eastern U.S. for high impact weather during the cool season; however, our approach can be expanded to other parts of the country and phenomena. The primary goals are: (1) To extend our newly developed fuzzy clustering approach to high impact weather events including precipitation, freezing level (2-m temperature), and 10-m wind for days 1-7 using the short-range and global ensembles; (2) Expand our new spread-anomaly ensemble tool; (3) Use these tools to verify these phenomena in the ensembles and understand the large-scale flows attached to the less predictable events; and (4) Integrate the Alan Alda Center for Communicating Science (www.aldacenter.org) into our CSTAR to help forecasters better communicate probabilistic information through a series of three workshops, some of which involving stakeholders.

2. Scientific Objectives and Accomplishments

a. Fuzzy-Clustering Tool Development

The CSTAR student has been expanding the fuzzy clustering tool to include other variables and approaches. The existing version online (http://breezy.somas.stonybrook.edu/CSTAR/Ensemble_Sensitivity/FC_Main.html) focuses the clusters around sea level pressure. A goal of this project was to develop the clustering around other variables using the global ensembles (GEFS, CMC, and EC – 90 members).

During winter storms one serious concern of forecasters is whether the precipitation will fall as snow, rain, or mixed phase. One useful guidance is the location of the zero degree Celsius contour within the planetary boundary layer, such as at 925 hPa level. Proximity to this line may indicate the transition region between snow and rain. We have experimented with a new clustering tool to highlight the uncertainty of this location. Clustering based on the 925 hPa temperature (T925) directly is not expected to highlight this region, since temperature variability can be large over regions with temperature far from freezing, and using temperature directly mainly highlights those regions with the largest temperature variability that are not necessarily close to the freezing contour. Instead, the zero degree contour is highlighted by a novel transformation, as follows: For each ensemble member, grid boxes where T925 is above freezing is assigned a value of 1, while grid boxes where T925 is below freezing is assigned a value of 0. For any forecast ensemble member (or the analysis), the entire field will be either 0 or 1. However, when averaging the entire forecast ensemble, regions where all ensemble members have temperature above 0 will show an ensemble mean of 1, while regions where all members have below freezing temperature will have a mean of 0. Over regions where some of the members have above freezing temperature and some below, the ensemble average is between 0 and 1, and corresponds to the fraction of ensemble members that have temperature above freezing.

To illustrate this new approach, we applied it to a winter storm that impacted the US Northeast coastal area on 15 November 2018. This storm brought several inches of snow to the tri-state area, and generated evening rush-hour traffic gridlock over New Jersey and New York due to more widespread snow accumulation than expected during the evening commute. Here, we examine the 1.5-day ensemble forecast valid at 0000 UTC 16 November 2018. The ensemble consists of 50-member ECMWF ensemble, 20-member NCEP GEFS ensemble, and 20-member CMC ensemble, with a total of 90 ensemble members. Figure 1a shows the ensemble mean and spread of T925. The ensemble-mean zero degree contour extends from just south of Long Island west-southwestward towards central New Jersey into southern Pennsylvania, and then southwestward into Maryland and Virginia. The shades in the figure shows the ensemble spread. Clearly, the largest spread occurs south of the zero degree ensemble mean contour. Thus by examining the ensemble mean and spread of T925, it is not clear how much uncertainty there is in the location of the zero degree contour, with the largest uncertainties highlighted over regions where T925 is well above freezing.

The situation is very different after our transformation is applied. The ensemble mean and spread of the transformed T925 field is shown in Fig. 1b. The ensemble spread (shades) now highlights the region around the ensemble mean zero degree contour, which is very close to the location where the ensemble mean of the transformed field has a value of 0.5. As expected, the ensemble spread is basically bounded by the ensemble mean contours of 0.01 and 0.99. North of the 0.01 contour, all members have below freezing temperature, while south of the 0.99 contour, all members have above freezing temperature. In between, some members forecast above freezing while others forecast below freezing, thus highlighting a region of potential transition between rain and snow. Figure 1b shows that even 1.5-day prior to the event, there is still significant uncertainty in the location of the zero degree contour, with the full ensemble indicating an uncertainty of around 150 km in its location. Such large uncertainties are not clear from the inspection of T925 directly (Fig. 1a).

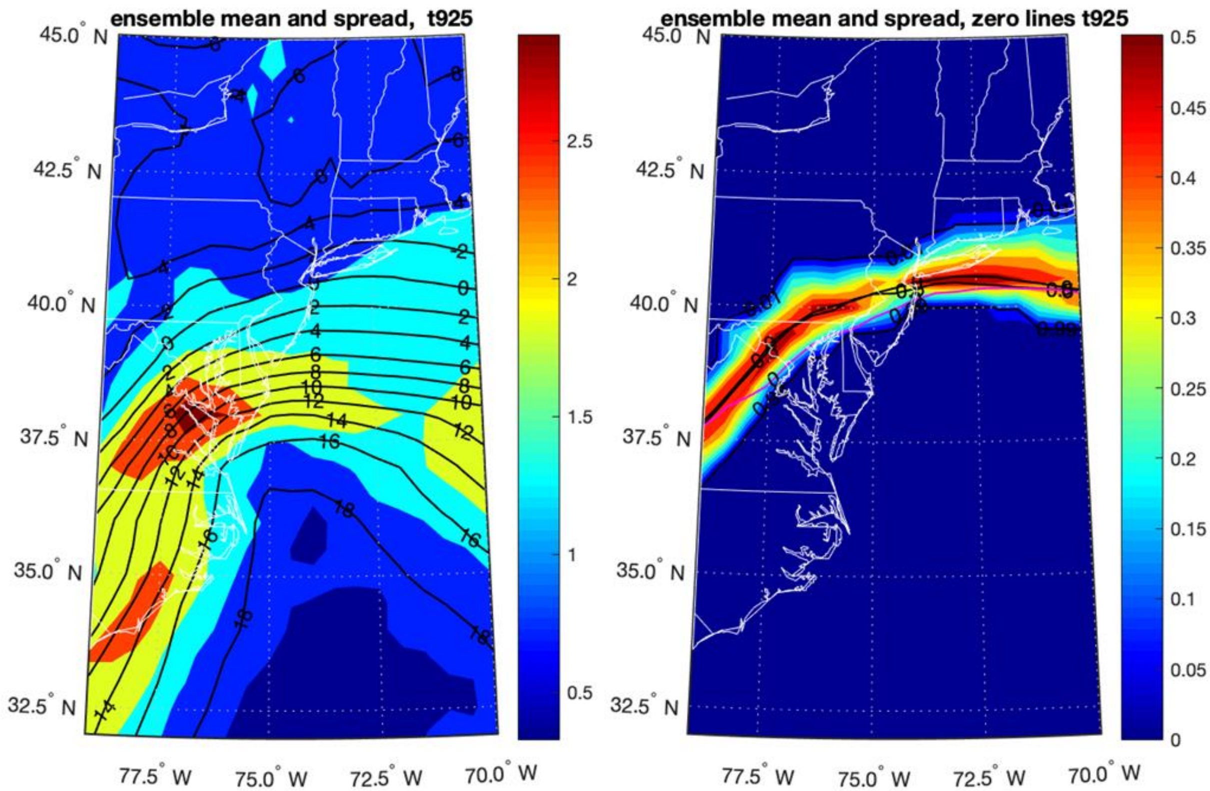


Figure 1. Left (a) Contours: Ensemble mean T925 (in °C) valid at 00Z 16 November 2018; Shades: Ensemble spread. Right (b) Shades Ensemble spread of transformed T925 (see discussions in text). Black contours: Ensemble mean of transformed T925 at values of 0.01, 0.50, and 0.99. The ensemble mean zero degree contour is also shown. The zero degree contour from the analysis is shown as the magenta line.

The zero degree contour from the analysis is plotted as the magenta line on Fig. 1b. Over most regions it lies south of the ensemble mean zero degree contour, indicating colder temperature than predicted by the ensemble mean. Nevertheless, Fig. 1b shows that the analysis lies within the ensemble and is not out of ensemble. Over southern New Jersey, northern Delaware and northeastern Maryland, the analysis lies close to the edge of the ensemble, but close to or below freezing temperature over these regions should not have been unexpected if the full ensemble is considered, since the analysis does lie within the full ensemble.

Clustering is performed using the two leading EOFs of the transformed T925 field. The leading EOFs are shown in Fig. 2. Positive EOF1 (Fig. 2a) shows negative values on both sides of the ensemble mean zero degree contour, indicating colder temperature over those regions, translating to a more southward location of the zero degree contour. Positive EOF2 (Fig. 2b) shows negative values to the southwest and positive values to the northeast, suggesting colder T925 over Virginia and warmer to the east of Long Island, or a slight counter-clockwise rotation of the zero degree contour. Clustering is performed using the two PCs, and the results of the clustering is shown in Fig. 3a.

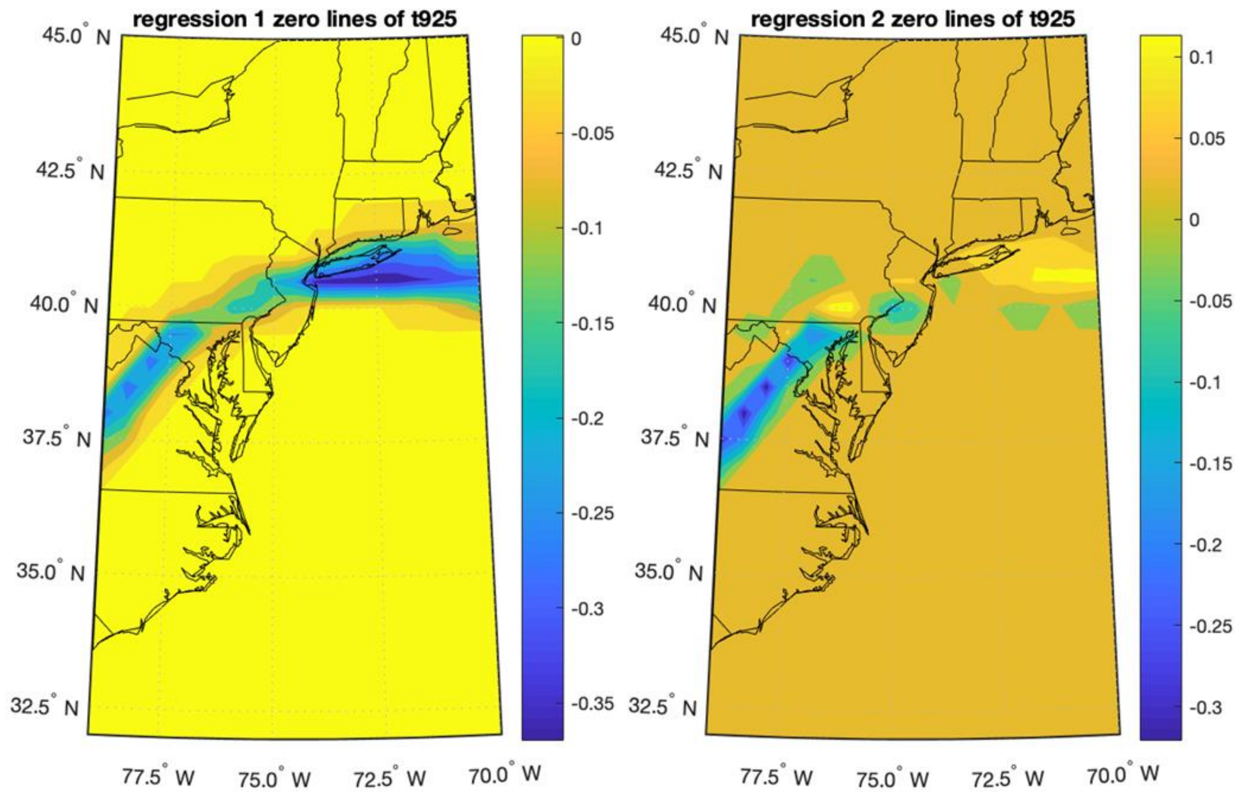


Figure 2. Left (a) Leading EOF of transformed T925. Right (b) Second leading EOF.

Figure 3a shows that all members with large negative PC2 (corresponding to more northward location of the zero degree contour) are from CMC, and this ensemble clearly has a warm bias for this case. Note that the analysis projected onto PC1 and PC2 is shown as the magenta cross on Fig. 3a. The analysis has positive PC1, indicating a more southward location of the zero degree contour than the ensemble mean as discussed above. Nevertheless, Fig. 3a shows that the analysis is located inside the full ensemble in PC1-PC2 space, consistent with the discussions above. The group mean zero degree contours for the five groups are shown in Fig. 3b. This also highlights the large spread in the predicted location of the zero degree contour. While none of the groups shows as far south a location as the analysis (yellow) over southern New Jersey and northern Delaware, the large spread displayed by the five groups as well as the spread shown in Figure 1b would serve as indicators that there is significant uncertainty in the location of the T925 zero degree contour.

One complication that we encountered relates to the fact that the main interest in the location of the zero degree contour lies over regions with significant precipitation. That is the reason why we restricted the analysis to the region shown in the figures. We have experimented with larger domains, but since the zero degree contour extends into regions without precipitation, the leading

EOFs could be dominated by variability over those regions and thus become less useful. This is different from clustering based on MSLP or precipitation, in which the variance is always dominated by regions around cyclones and thus areas with large variance are always of interest. More testing and fine-tuning will be needed to perfect this tool. We are currently coding up a new package to be run daily at EMC by our NCEP/EMC CSTAR collaborators, with the results to be posted on a new web page (password protected due to inclusion of ECMWF data) on the Stony Brook CSTAR site, to further test and evaluate this tool. We will also examine historical cases from the TIGGE archive and use this tool as a verification tool to evaluate the performance of the 90-member ensemble using the methodology described in Zheng et al. (2017).

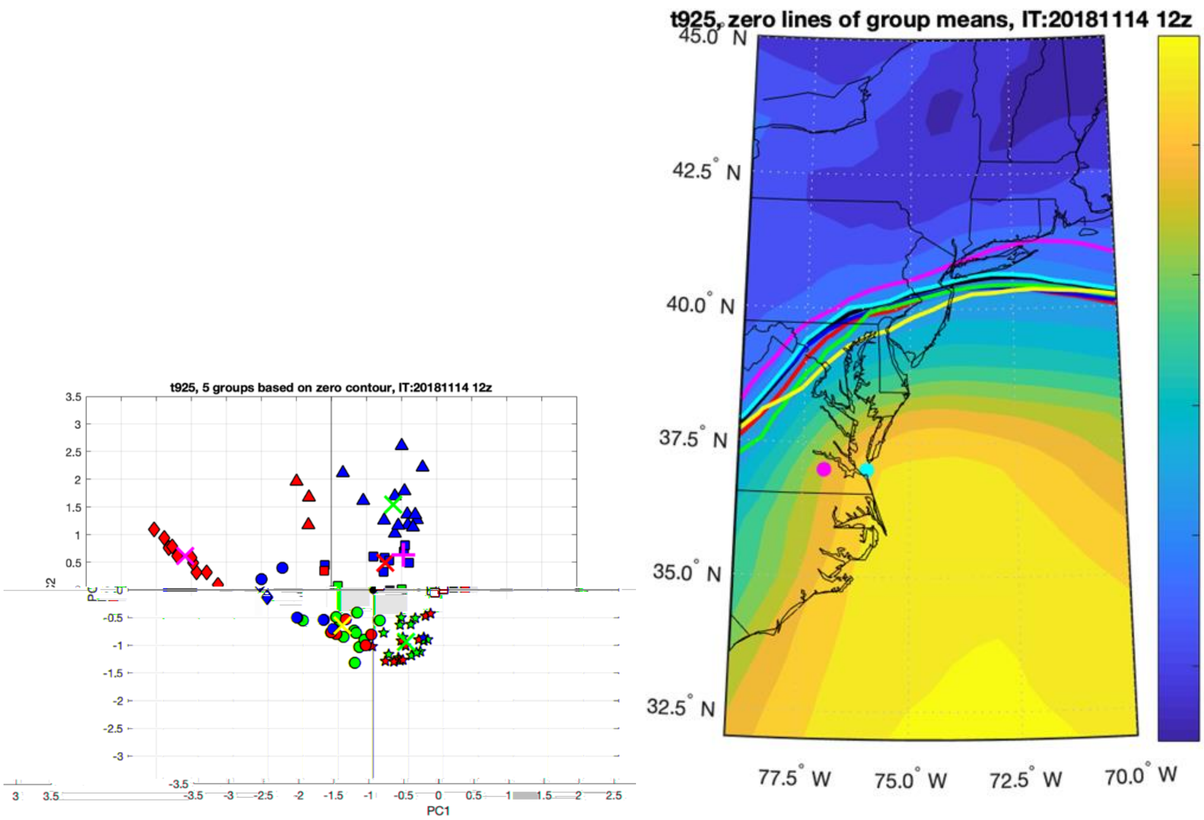


Figure 3. a) Projection of each ensemble member (filled colored symbols), group means from clustering analysis (X), and the analysis (magenta +) onto PC1-PC2 space. Red symbols indicate CMC members, green NCEP GEFS, and blue ECMWF. b) Zero degree contours from the 5 groups, together with that from the analysis (yellow).

b. Evaluation of Clustering Approaches

The current operational cluster approach uses a fuzzy clustering method (Zheng et al. 2017; 2019; hereafter ZH1 and ZH2 respectively), in which a Principal Component Analysis (PCA) assesses the modes of variability of each member to cluster. PCA describes creation of an orthogonal basis, or a rotation of axis, where the first vector (PC1) is oriented along the direction of greatest variance, and vectors thereafter along the next greatest variances (PC2, PC3… etc.).

For extratropical cyclones along the U.S. East Coast with MSLP, PC1 generally represents a spatial displacement, while PC2 represents variability in intensity (ZH2).

This project explored different clustering approaches using the 90-member ensemble of GEFS, CMC, and ECMWF for U.S. East coast and western Atlantic regions. The details are in a paper in revision (Kiel and Colle 2022), while some highlights are shown below. The dataset included 180 winter cyclones from 2007 to 2015, defined by minimum pressure less than 1005 hPa in the GFS Analysis. The 12-h MSLP and accumulated precipitation for the ensembles was obtained from the Observing System Research and Predictability Experiment's Interactive Grand Global Ensemble (TIGGE) archives. For the precipitation, the ERA5 data were obtained from the Copernicus Climate Data Store (Hersbach et. al. 2019). ERA5 is chosen due its availability of precipitation in 1-h accumulation, and it has shown consistency with historic observations in the Northeastern U.S. (Crosett et. al. 2020) All data are bilinearly interpolated to a 1° by 1° grid.

Several clustering spaces and methods have been used in atmospheric sciences, including hierarchical based approaches (ie, Agglomerative Hierarchical Clustering; Ferstl 2001), density based approaches (ie, Density Based Spatial Clustering With Noise; Li et. al. 2008), and centroid based approaches (ie, K-Means Clustering; Yesibudak 2016, Fuzzy Clustering; ZH1). They are summarized in Tables 1 and 2 and also in some detail here:

- Agglomerative Hierarchical Clustering (AHC): The closest pair of points are clustered together, and the mean of those points are taken to be a new point which represents a cluster. This is repeated iteratively until the desired number of clusters is reached. Ferstl (2001) used AHC on height contours to consolidate spread and forecast dynamics onto a single visual.
- K-Means Clustering (KMC): Clustering method where each point is assigned to the nearest mean cluster center where within cluster distances are minimized. Yesibudak (2016) used (KMC), to cluster wind vectors to find wind pattern similarities of efficient and inefficient wind farms in Turkey.
- Density-based Spatial Clustering with Noise (DBSCAN): Density-Based Spatial Clustering With Noise: A clustering method which identifies regions of high density by examining weather points fall within a certain fixed radius of each other. Li et al. (2008) used DBSCAN in a novel data-mining algorithm used to detect thunderstorm mesocyclones in real-time.
- Fuzzy Clustering (FC): Ensemble members are assigned weights based on strength of membership to a cluster depending on PC space distance from cluster mean. The cluster with the largest weight represents that ensemble member. Zheng et al. (2017) (Hereafter ZH1) and ZH2 used fuzzy clustering to generate scenarios from model ensembles.

Each clustering approach comes with its own pros and cons. DBSCAN has the ability to exclude outliers, preventing them from significant modification of a discrete group of points. However, careful choosing of radius is important, as too small of a radius may result in over exclusion or even no points clustered, and too large a radius leads to a single cluster encompassing all points. FC's computational complexity allows for more robust clustering, but it is possible that such complexity isn't needed. Further, methodologies using FC usually assign a point to the cluster that is most probable to belong to, but could disregard important information about significant probability that the point belongs to another cluster. AHC and KMC provide simpler alternatives, however, both, along with FC, attempt to include all points in clusters - possibly when it is not warranted.

Table 1: A description of all clusterable spaces tested.

Two Principal Components (2PC)	Clustered by the unweighted first two principal components of the EOF analysis of the dataset only.
Weighted Principal Components (WPC)	Clustered by the n principle components of the EOF analysis of the dataset representing 90% of and weighted by the variance explained.
Magnitude - Displacement (MDISP)	Clustered by a 3D space, one dimension representing the magnitude of the minimum pressure, the other two dimensions the displacement of the minimum pressure from the analysis scenario minimum pressure.
Euclidean	Clustered by applying ensemble members into the clustering algorithms directly.
SOM – Principal Component (SOM)	Ensembles members are not clustered directly. Rather, the clusters are fitted to a 6x6 node map, which acts a linearization step. The nodes are then clustred by their Principal Components.

Table 2: A description of clustering methods used for each clusterable space tested.

Euclidean distance was used for all clustering algorithms.

Agglomerative Heirarichal Clustering – Ward's Linkage (AHC)	All aggloreratve clustering is done by grouping points together one by one iteravely, based on a linkage metric. Ward's Linkage minimizes the intracluster variance between points.
Agglomerative Heirarichal Clustering – Full Linkage (AHF)	Clusters linked by the distance of the point furthest away from each other within each group.
Agglomerative Heirarichal Clustering – Average Linakge (AHA)	Clusters linked by the average distance between points between each group.
K-Means Clustering (KMC)	Initial groups are assigned, then iterations are preformed using an algorithm to minimize variance between points.
Fuzzy Clustering (FZC)	Like k-means, but instead assigns probability of each point to belong to each cluster on each iteration.

1. APCP Scenario Nearest Analysis Results

We compared how close each 12-h APCP cluster is to the precipitation analysis. This is done for (1) the magnitude of displacement error and its north/south and east/west components, (2) intensity and magnitude of intensity error, and (3) brier skill score (BSS) by threshold.

APCP displacement error (Fig. 4) is calculated through finding the weighted mean latitude and longitude of all grid points, using precipitation as the weight. We break this down into north/south (Fig. 5) and east/west components (Fig. 6). Component results were subtracted in the same way as MSLP, so a positive error indicates northward or westward bias from NCEP analysis in the ensemble or scenario mean, and a negative error a southward or eastward bias from NCEP analysis in the ensemble mean. We found that using weighted mean for the full multi-model ensemble mean becomes problematic at later lead times. As precipitation forecasts become increasingly dispersed, ensemble mean APCP becomes uniform, especially at later lead times (Fig. 7). Therefore, the center of mass location of the ensemble mean becomes less representative of precipitation errors and more representative of the center of the region. Figure 7f shows that 70% of cyclones are located to the south and 70% to the west of region's center. This translates into a northward bias of 20 km and 60 km as well as the eastward bias of 14 km and 114 km for the medium (days 4-6) and long lead (days 7-9) times, respectively. But this ensemble mean result is likely artifacts of region boundary choice. However, the scenario nearest analysis is resilient to this issue, most notably that the westward bias of around 27 km and 71 km for medium and long lead times, which contrasts to the eastward bias of the ensemble mean. This affirms that clustering 12 h APCP into scenarios is sufficiently effective to preserve sets of similar APCP fields, if they exist.

We also repeated using 12 h APCP the results for the scenario nearest analysis percentage error minimums of the magnitude of displacement error for each clustering algorithm (Table 3a), using the average between clustering spaces, and for clustering spaces (Table 3b), using the average between clustering algorithms. There exists sensitivity of the clustering algorithm results to lead time (e.g., k-means clustering has minimum displacement error for 10%, 19%, and 16% of the time at short (days 1-3), medium, and long lead times, respectively, indicating that the error are likely small, but there may be some evidence that SOM performs better than other clustering algorithms, having minimum value of displacement error 23%, 20% and 24% of the time for short, medium, and long lead times, respectively. And while IDISP is determined differently for 12 h APCP than MSLP, its scenario nearest analysis still performs most strongly, for example, for short lead time, the scenario nearest analysis has the minimum value of displacement error 42% of the time, well above the next highest, Euclidean, which has minimum value of displacement error 25% of time for short lead time.

Intensity error (Fig. 8) and its magnitude (Fig. 9) are calculated by finding the 10 most intense precipitation points in each ensemble member, averaging them, and comparing the average to the average of the 10 most intense precipitation points in the analysis. APCP intensity errors are calculated as scenario minus observed, so a positive value of intensity error indicates a systematic overestimation of APCP intensity, and a negative value a systematic underestimation of APCP intensity. The ensemble mean shows a systematic overestimation at 95% confidence for medium and long lead times of 1.6 mm and 4.7 mm respectively. Scenarios nearest analysis for 2PC and WPC give a 95% significant overestimation for long lead time at around 2.3 mm. IDISP has a lower overestimation at a long lead time of 1.2 mm. Euclidean space has a slightly higher estimation at a long lead time of 3.1 mm. For magnitude of intensity error (Fig 9), we find that 2PC and IDISP have a marginal improvement over ensemble mean at long lead times from about 7.7 mm to 6.3 mm, but in general results between the ensemble mean and the scenarios do not change with significance for magnitude of intensity error.

We may also find the percentage of times a scenario nearest analysis produces an intensity error minimum for 12 h APCP, for clustering algorithms (Table 4a), using the average between clustering spaces, and for clustering spaces for analysis (Table 4b), using the average between cluster algorithms. Sensitivity from lead time to lead time for clustering algorithms indicates that error differences between them are once again small on a case-by-case basis, and it isn't clear whether any algorithm or space gives a more frequent occurrence of an intensity error minimum.

For BSS by threshold using GEFS control as reference, we compared the results of using 0, 5, 10, and 15 mm thresholds for scenario nearest analysis in 2PC space (Fig. 10), and only the 10 mm threshold results are shown for clustering approach closest to the analysis (Fig.11). BSS decreases when increasing the threshold value for medium and long lead times, for example, at long lead time, the average 2PC BSS value for a scenario nearest analysis was 0.58 for 0 mm threshold, 0.57 for 5 mm threshold, 0.51 for 10 mm threshold, and 0.44 for 15 mm threshold (not shown). BSS by threshold also decreased with lead time, for example, for a 10 mm threshold the BSS threshold on average for any scenario nearest analysis in 2PC was 0.76, 0.62, and 0.51 for short, medium, and long lead times respectively. While this is a slight improvement on ensemble mean (0.67, 0.52, 0.44, for short, medium, and long lead time, respectively), there are no significant variations in behaviors of BSS by threshold for any different clustering technique (Fig. 10). For example, the 10 mm threshold values as reported above for 2PC did not change from clustering space to clustering space. Likewise, use of a different control member did not change these behaviors either (not shown).

2. Brier Skill Score and Weighted Indexing

We repeat the predictive skill evaluation of the scenario nearest analysis done in the last section, using APCP, by testing (1) BSS by cluster number and (2) a weighted indexing of the scenario nearest analysis.

We now evaluate results of BSS by cluster number for 12 h APCP as was done with MSLP, first unweighted (Fig. 12), weighting by the size of the largest cluster (Fig 13), and then weighting by the square of the largest cluster (Fig 14). After weighting by the square of the largest cluster (Fig 11b), we note that fuzzy clustering and SOM have lower BSS by cluster values than other clustering algorithms at short and medium lead times, for example, in WPC space, SOM has values of 0.01 and -0.01 and fuzzy clustering has values of -0.01 and -0.01 at short and medium lead times, respectively, lower than k-means or any AHC clustering algorithm (the average of them is 0.17 and 0.08 at short and medium lead times). Euclidean space shows greater spread between lead time values than any other clustering space. For example, Euclidean k-means has values of 0.29, 0.12, and 0.01 at short, medium and long lead times, respectively, whereas 2PC space has values of 0.11, 0.02, and -0.07, at short, medium, and long lead times respectively.

The index distribution was calculated for 12 h APCP (not shown), and the index distribution weighted by the average number of members in each index (not shown)). From comparing the weighted index distribution to the weighted index distribution, it is implied that the size of the largest cluster (index 1), is very large on average for Euclidean space, excluding SOM. Verifying this, we find that, for short lead times, index 1 contains on average 35% of members for Euclidan AHC Ward's linkage, 48% of members for Euclidan AHC full linkage, 44% of members for Euclidean fuzzy clusters, and 31% of members for K-means clustering. AHC average linkage also gives high cluster percentages in the first index, including in 2PC space (48% at short lead time), IDISP space (66% at short lead time) and WPC (56% at short lead time). This is evidence that not only does AHC average linkage still have outlier sensitivity, but Euclidean space also has similar difficulties in generating clusters. In fact, attempting to cluster in the Euclidean space using AHC average linkage results in on average 76% of members in cluster 1, which is evidence that this method shows extreme outlier sensitivity and almost always fails to generate rigorous clusters.

The outlier removal tendency of Euclidean space and average linkage related clustering techniques helps explain their weighted index values usually being above the random selection threshold of 20% for short lead times (e.g., 35% for Euclidean AHC Ward's linkage) at short lead times. Outlier removal means that the index cluster is the scenario nearest analysis more than the average number of clusters would indicate for Euclidean space and average linkage related clustering techniques at short lead times.

12 h APCP at long lead times, like with MSLP, gives the inverted distribution result where index 1 is less likely and index 5 more likely to represent the scenario nearest analysis than the number of members within them would suggest. [e.g.: weighted index results of 2PC SOM 1: 9% 2: 14% 3: 20% 4: 27% 5: 29%]. The behavior is more common with 12 h APCP than with MSLP, with cluster index 1 being less likely to represent the scenario nearest analysis then suggested by its number of members for every technique tested. Caution would need to be applied to interpreting clusters at long lead times, especially when clustering via 12 h APCP

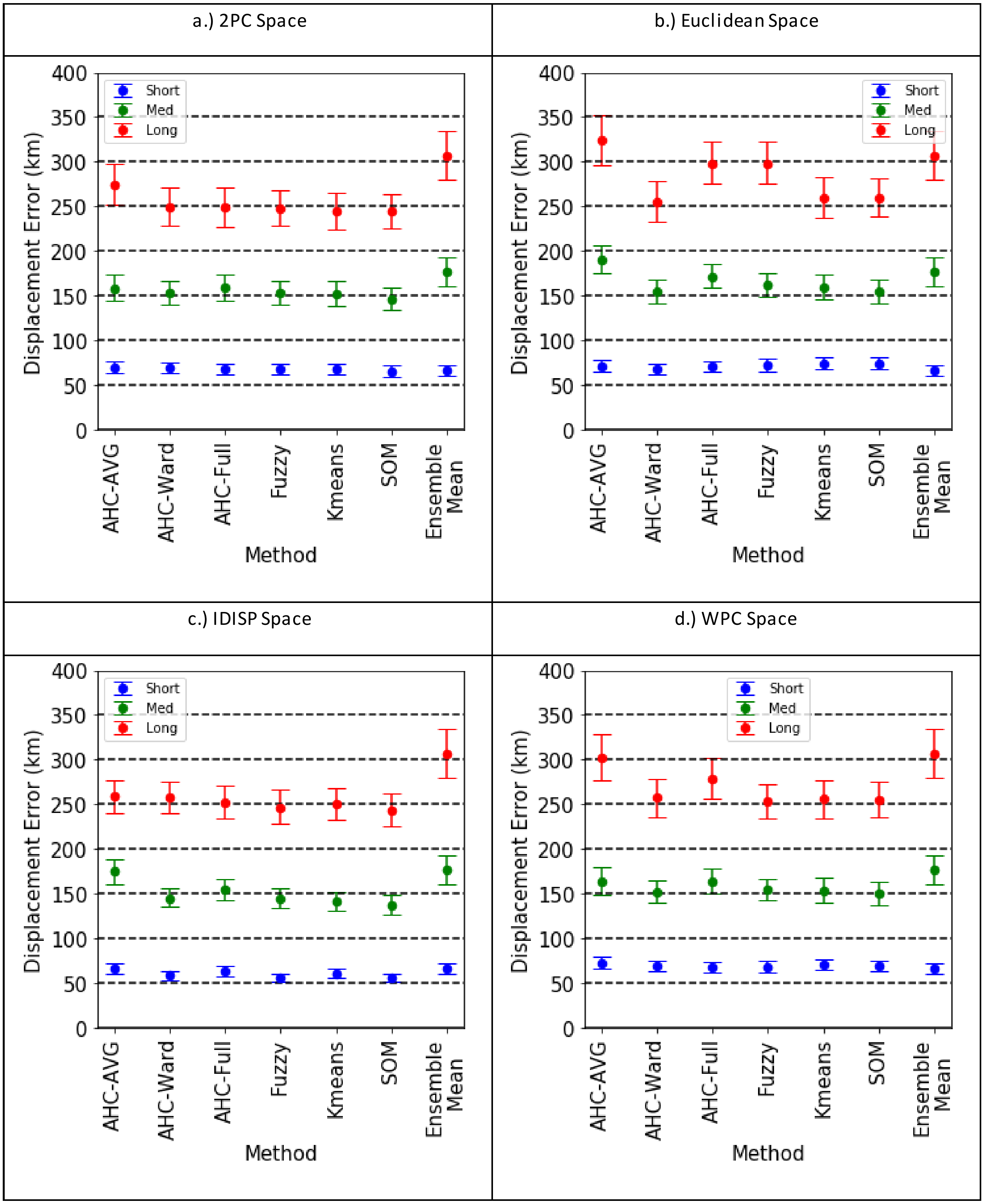


Figure 4: 12 h APCP Mean magnitude of displacement error for scenario closest to analysis and divided by short (blue) medium (green) and long (red) lead times. Plots are separated by space to cluster by, with a) 2PC Space, b) Euclidean Space, c) IDISP space, and d) WPC space. Also plotted are the 95% bootstrapped confidence intervals. Each is compared to the full ensemble mean.

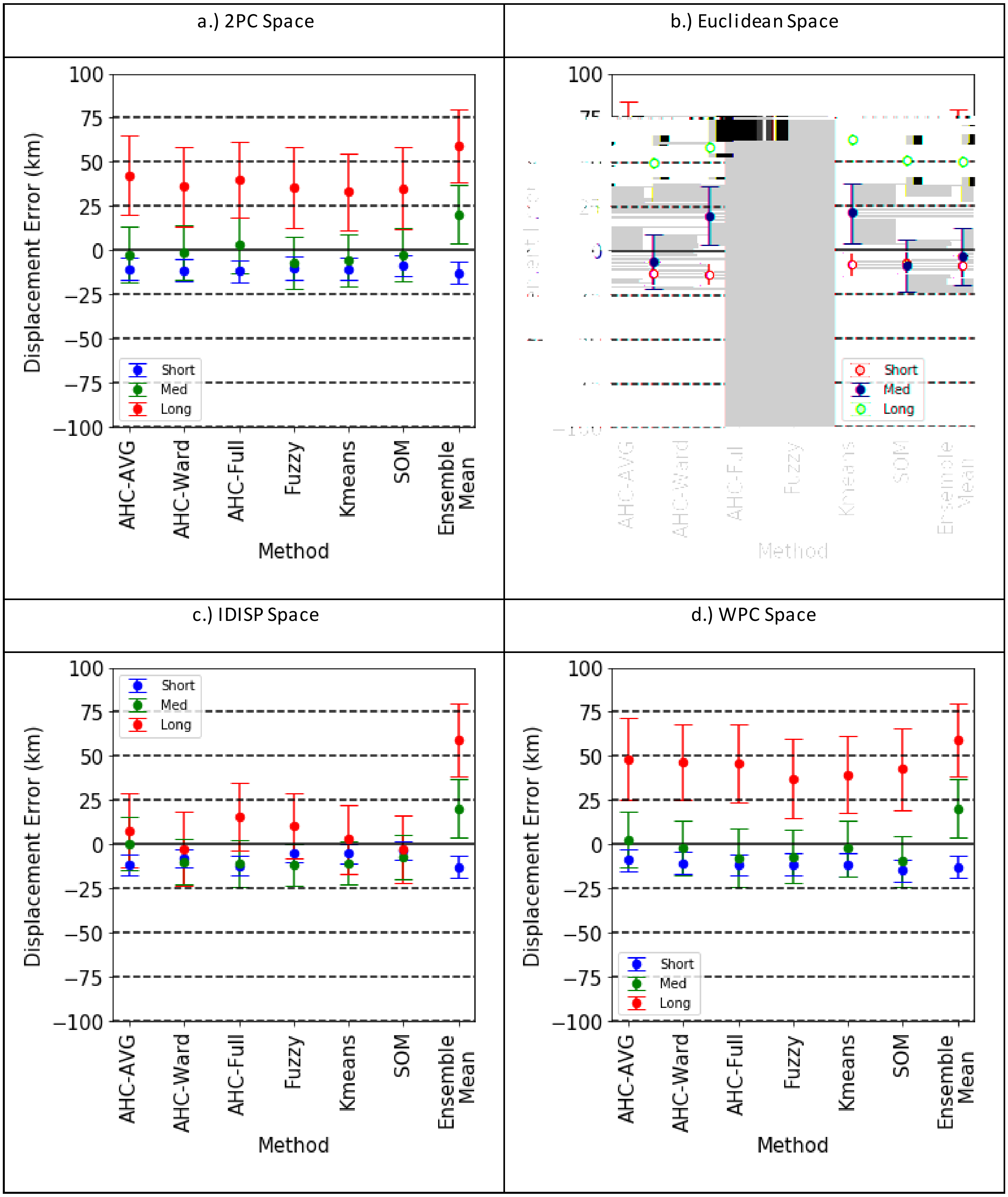


Fig 5: As in figure 4, but for 12 h APCP N/S component of displacement error

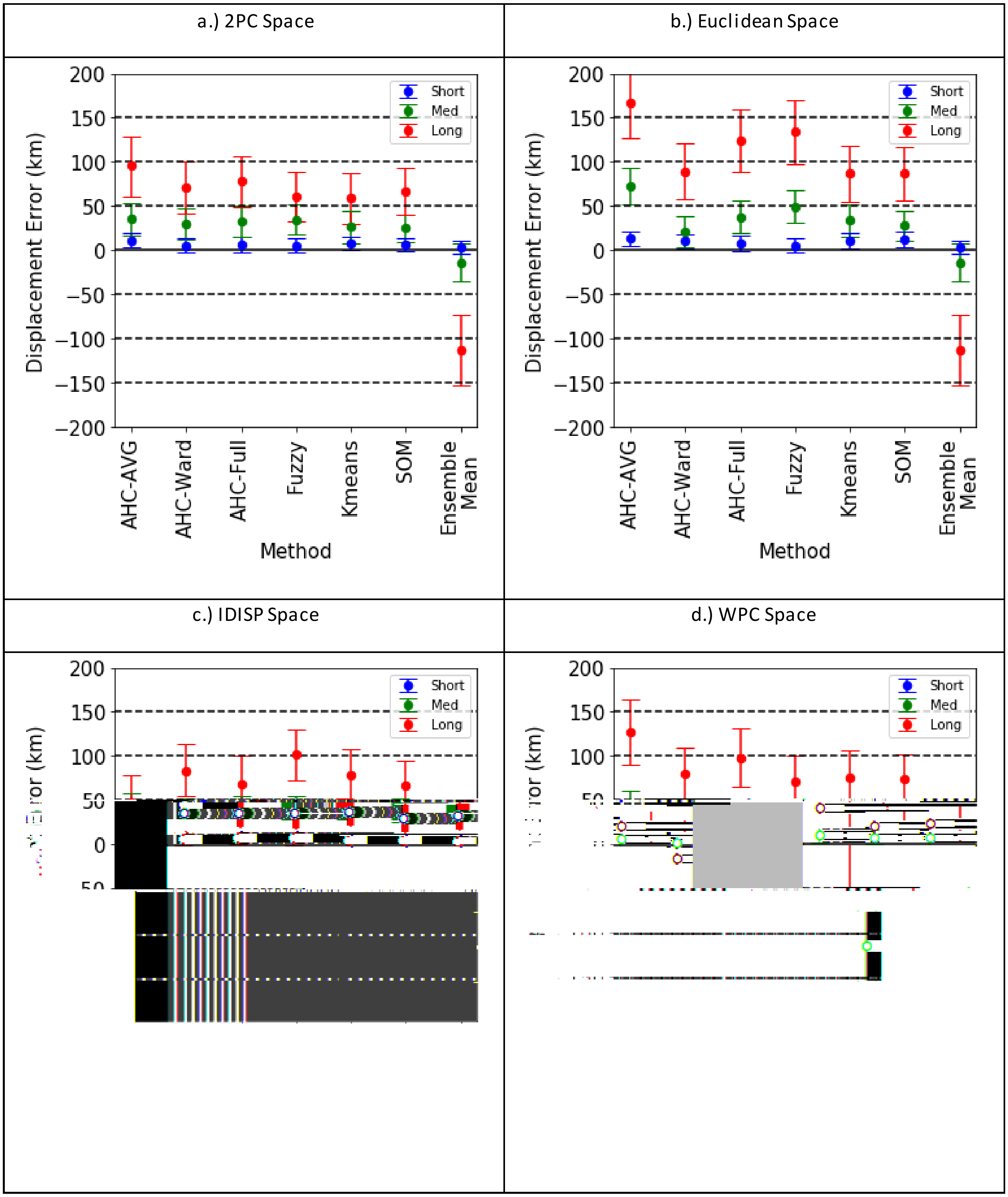


Fig 6: As in figure 4, but for 12 h APCP E/W component of displacement error.

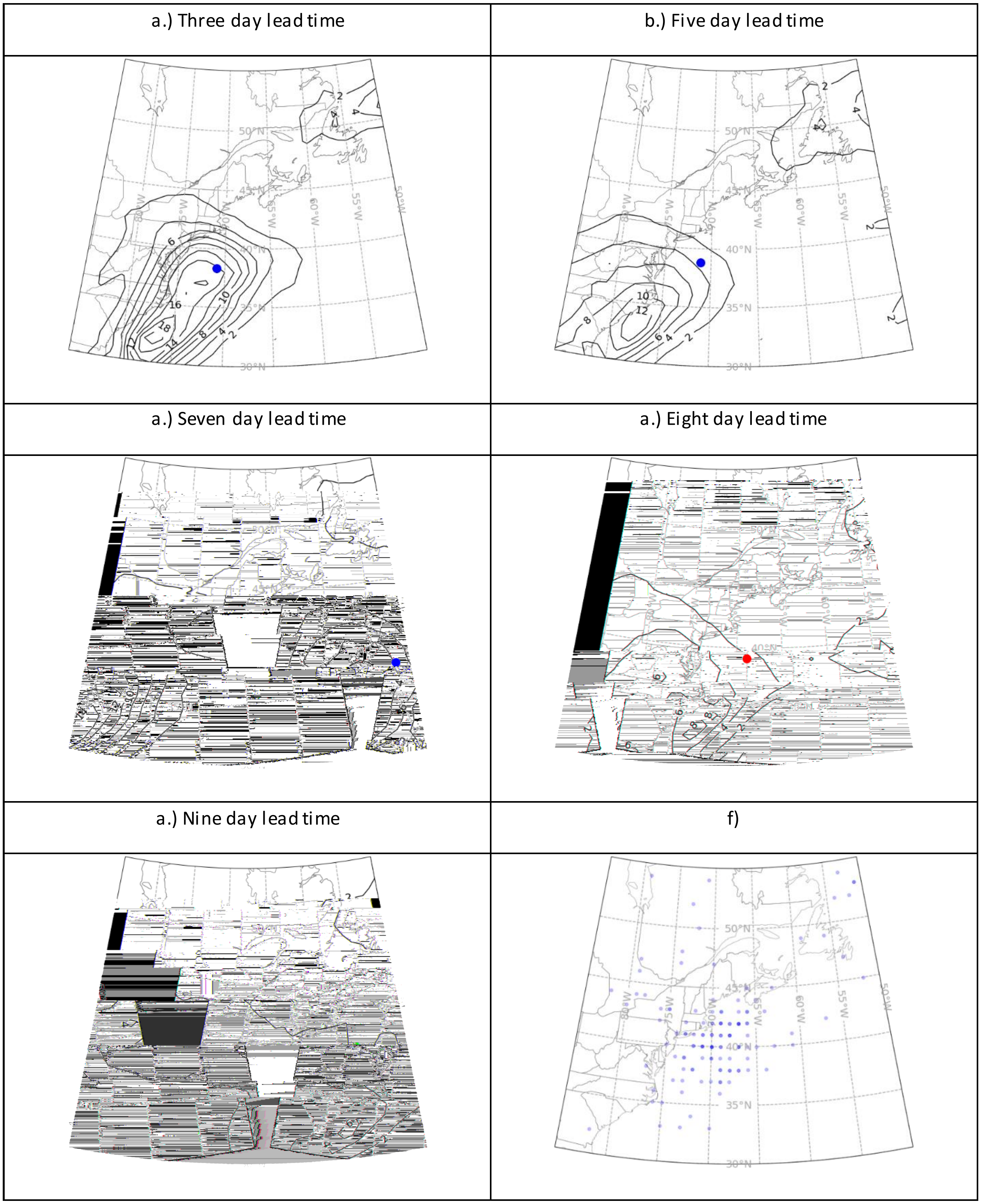


Fig 7. Plots showing 12 h APCP (solid every 2mm) ensemble means and their center of mass at a.) IT: 0000 UTC 27December 2012, b.) IT: 0000 UTC 25 December 2012, c.) IT: 0000 UTC 23 December 2012, d.) IT: 0000 UTC 22 December 2012, e.) IT: 0000 UTC 21 December 2012. VT: 0000 UTC 30 December 201.

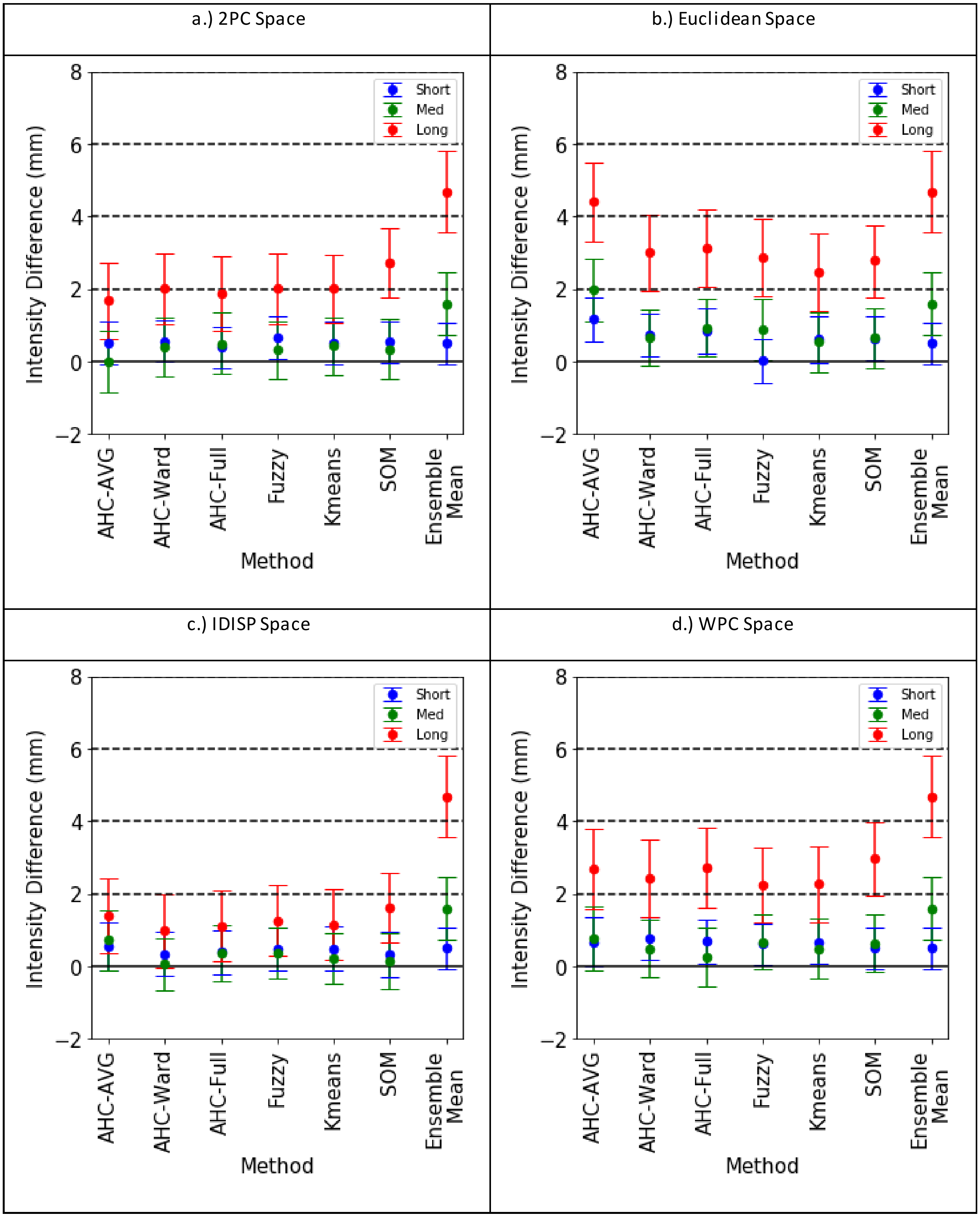


Fig 8: As in figure 4, but for 12 h APCP Intensity Error.

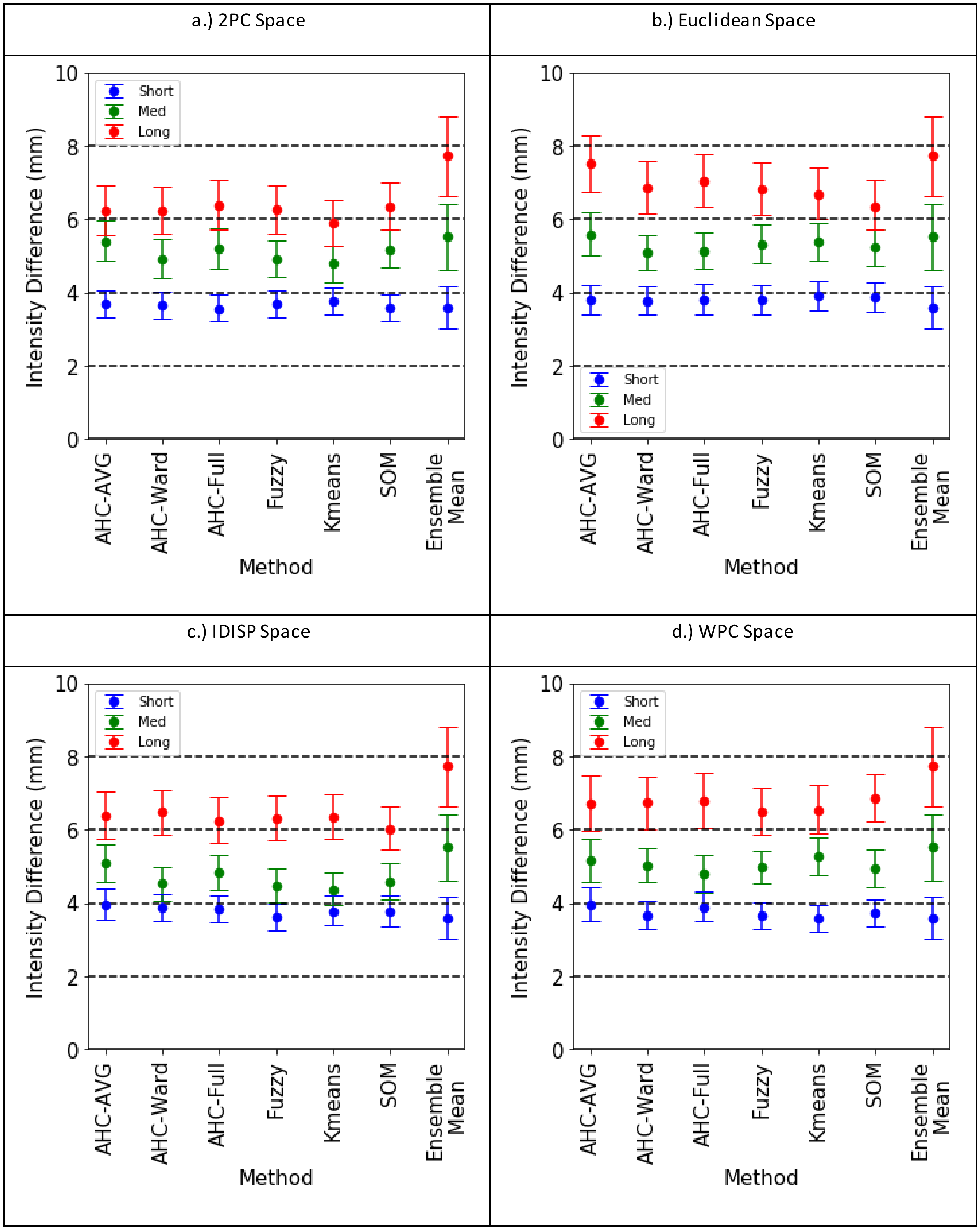


Fig 9: As in figure 4, but for 12 h APCP Magnitude of Intensity Error.

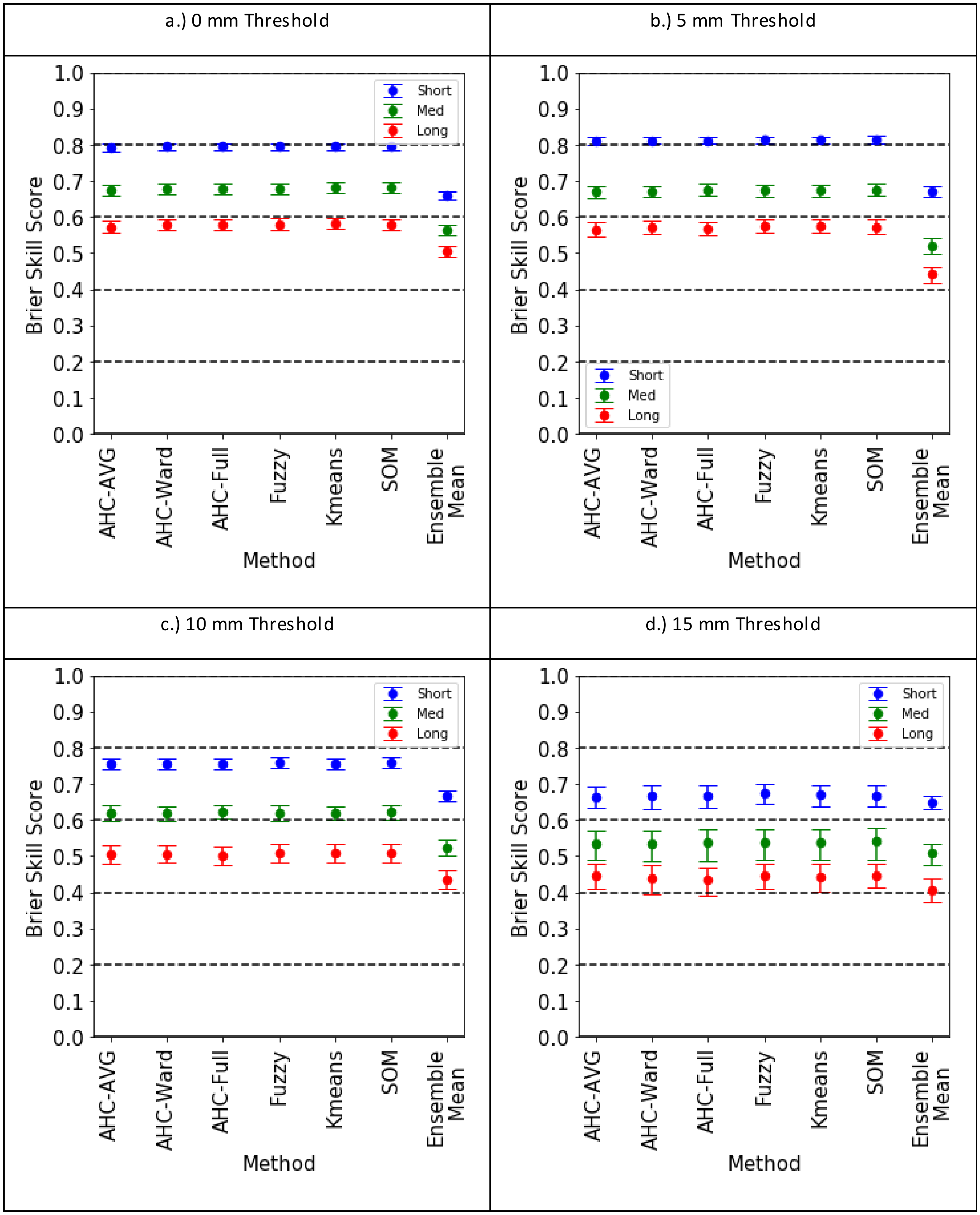


Fig 10: As in figure 4, but 2PC space comparisons of mean BSS by threshold of scenario nearest analysis using GEFS control as reference at thresholds of a.) 0 mm, b.) 5 mm, c.) 10 mm, d.) 15 mm.

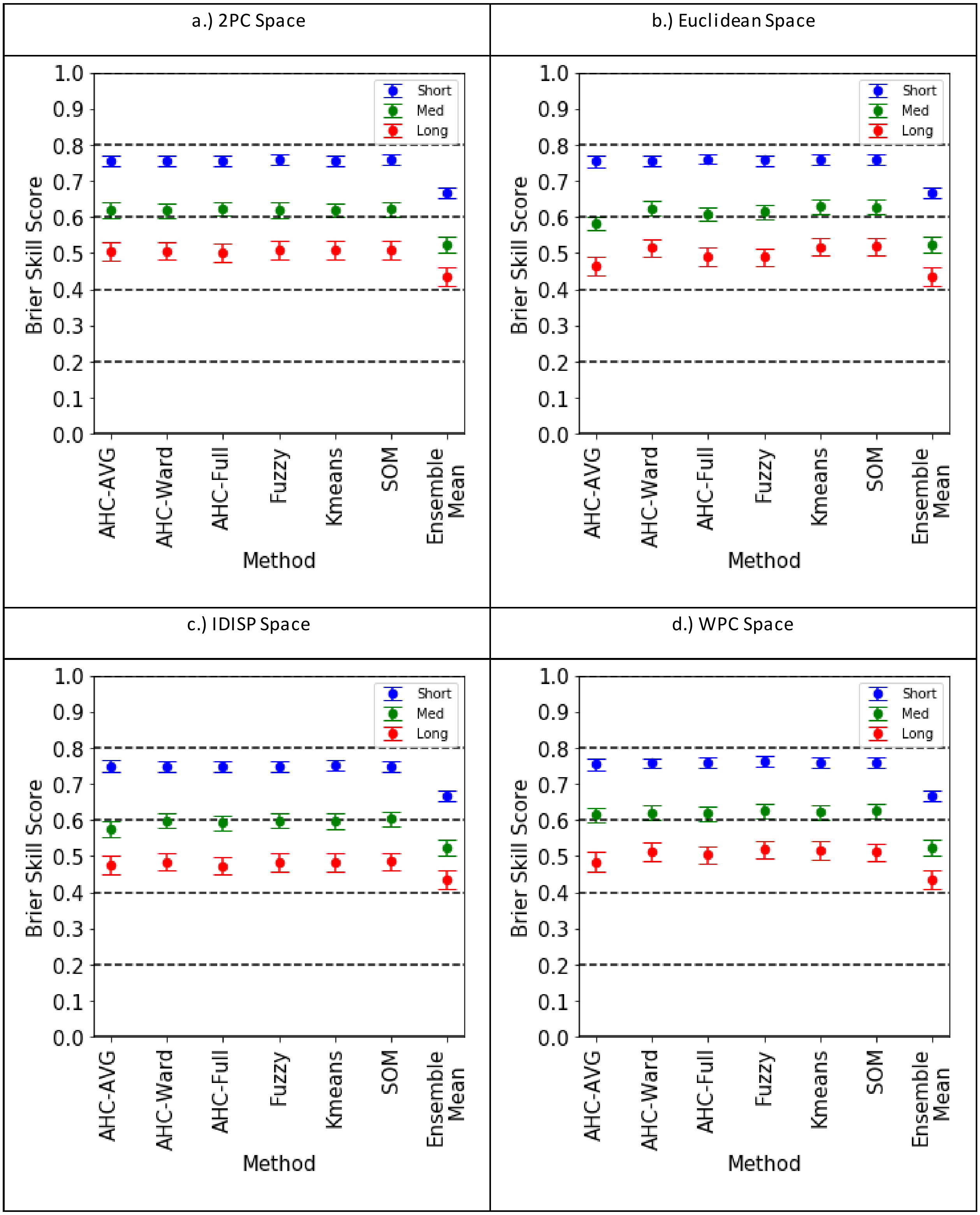


Fig 11: As in figure 4, but for mean BSS by threshold of scenario nearest analysis using 10 mm threshold and GEFS control as reference.

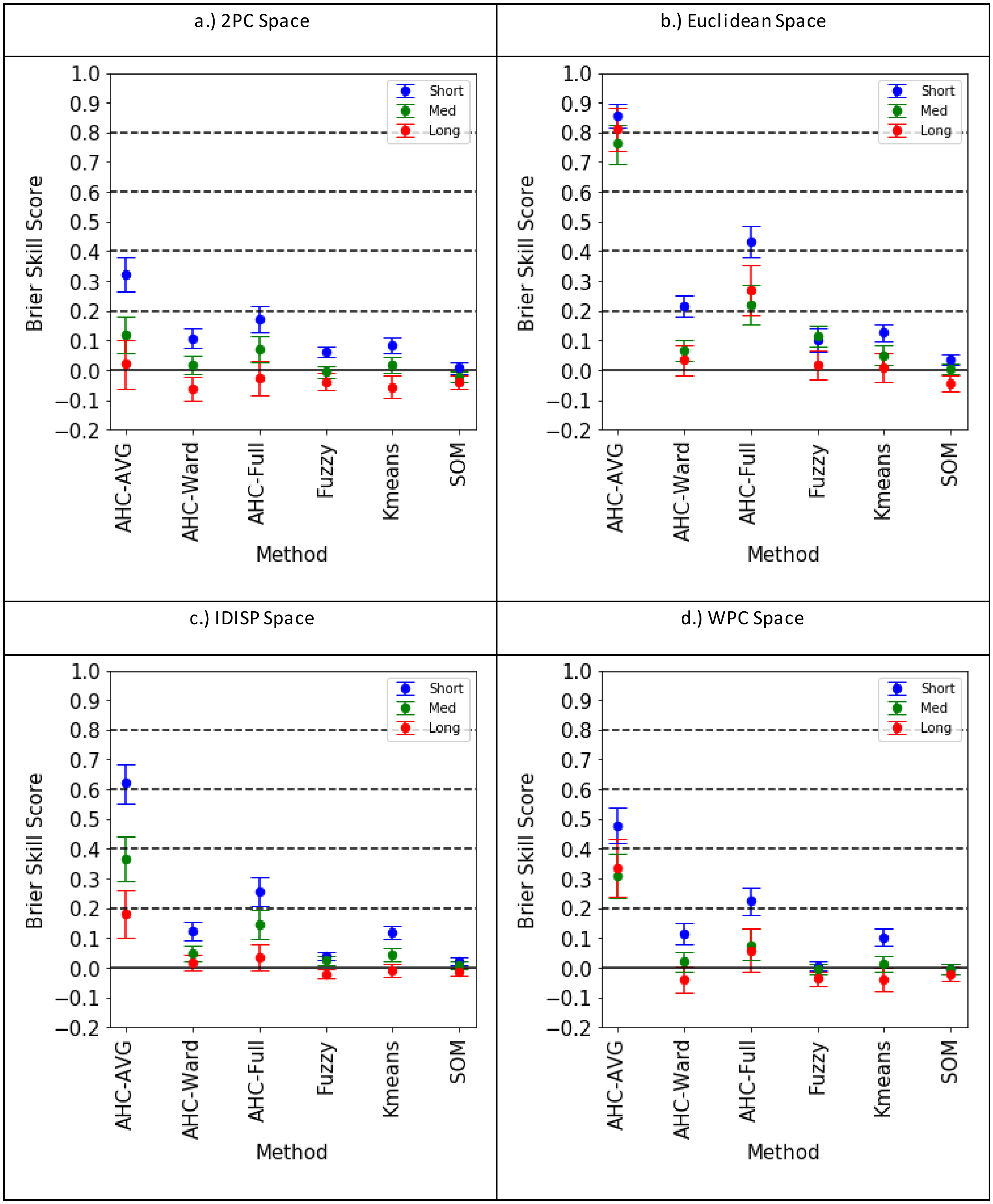


Fig 12: As in figure 4, but for mean BSS by cluster number.

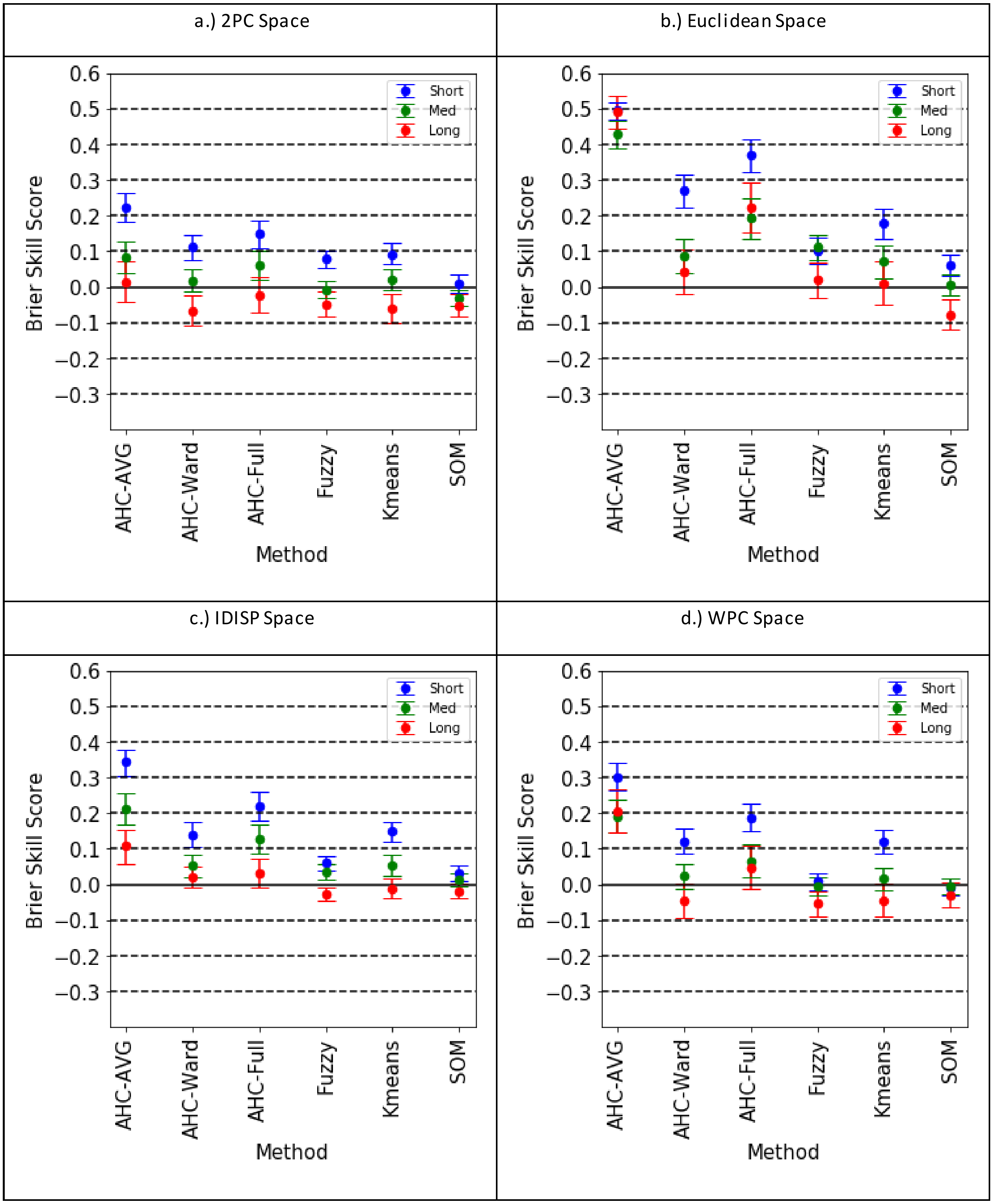


Fig. 13: As in figure 12 but weighting each BSS by the size of the largest cluster. Note that the y-axis range has been reduced to better represent the dataset.

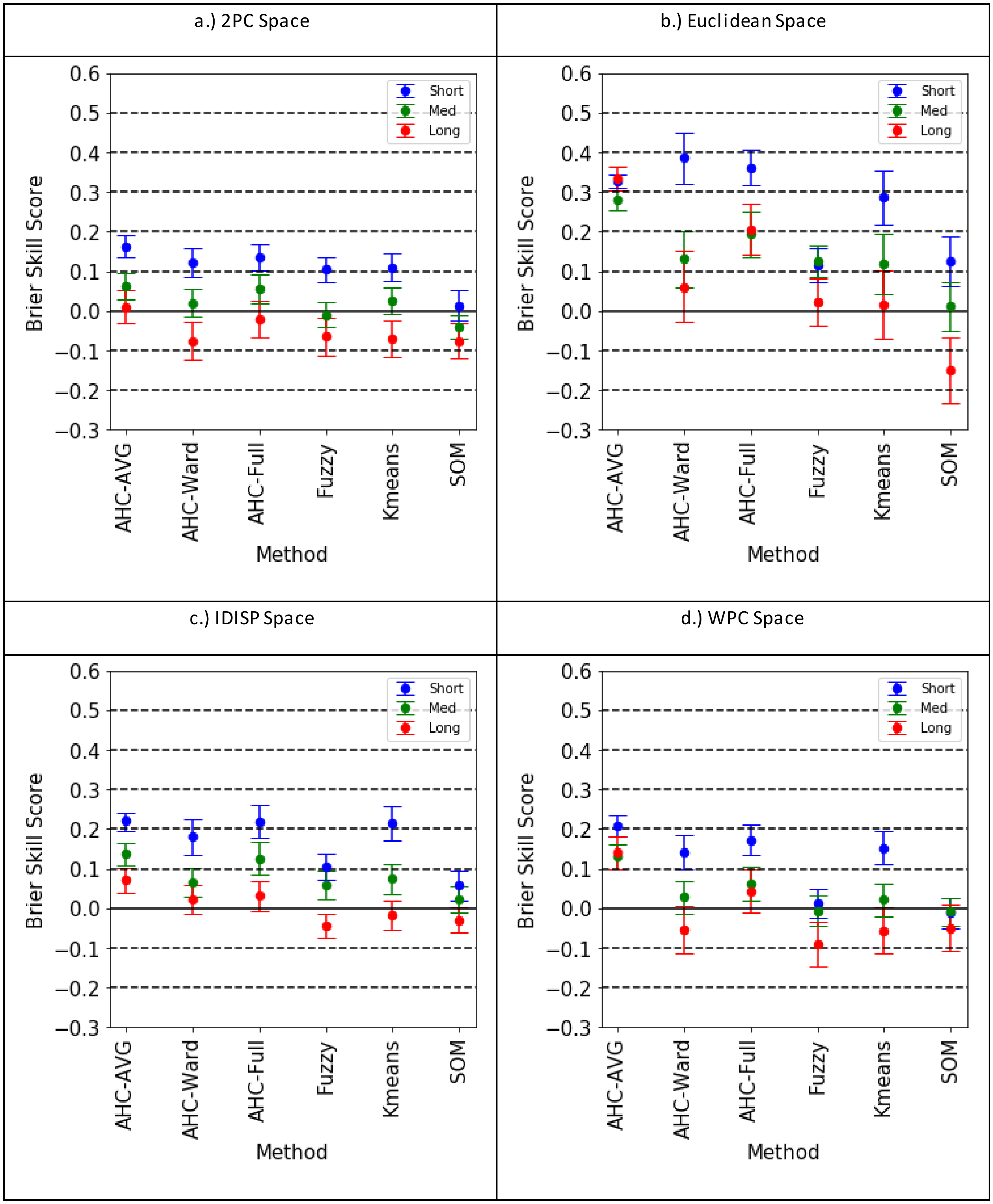


Fig 14: As in figure 12 but weighting each BSS by the square of the largest cluster prior to averaging.

Table 3: Percent of times a method produces a cluster with the minimum displacement error for 12-h APCP for a) clustering algorithm and b) clustering space, divided by lead time. Also included is the normal approximation to the binomial confidence interval for these results.

a. clustering algorithm

Algorithm	*Short Lead Time*	*Medium Lead Time*	*Long Lead Time*
K-means	10%±4	19%±5	16%±5
AHC: Avg	10%±4	14%±5	15%±5
AHC: Full	16%±5	14%±5	18%±5
AHC: Ward	19%±5	14%±5	13%±5
Fuzzy	22%±5	18%±5	13%±5
SOM	23%±5	20%±5	24%±6

b) clustering space

Space	*Short Lead Time*	*Medium Lead Time*	*Long Lead Time*
2PC	18%±5	19%±5	19%±6
Euclidean	25%±5	25%±6	23%±6
IDISP	42%±6	33%±6	44%±7
WPC	15%±4	22%±6	14%±5

Table 4: Percent of times a method produces a cluster with the minimum error in 12-h APCP intensity.

a. clustering algorithm

Algorithm	*Short Lead Time*	*Medium Lead Time*	*Long Lead Time*
K-means	15%±5	13%±5	16%±5
AHC: Avg	14%±4	15%±5	15%±5
AHC: Full	12%±4	19%±5	21%±6
AHC: Ward	18%±5	18%±5	13%±5
Fuzzy	21%±5	16%±5	21%±6
SOM	20%±5	19%±5	14%±5

b) clustering space

Space	*Short Lead Time*	*Medium Lead Time*	*Long Lead Time*
2PC	21%±5	22%±6	21%±6
Euclidean	23%±5	21%±5	25%±6
IDISP	29%±6	36%±6	33%±7
WPC	27%±6	21%±5	21%±6

b. Communication Uncertainty Workshop #1

The summary of both communication workshops are in Colle et al. (2021). The first CSTAR Communication Workshop was held from 4-5 March 2019, involving 14 forecasters from NWS Eastern Region, Northeast River Forecast Center, and Weather Prediction Center. Table 5 highlights the agenda of the meeting. The workshop began with welcomes and a discussion of the communication challenges for forecasting. The forecasters discussed some recent challenging forecast events as part of the homework assignment. The activities the rest of the day focused on how to better distill a message and communicate things in a more concise way. For example, one

exercise focused on communicating anything to do with weather in 1-minute, then the time allowed was reduced to 30 seconds, 15 seconds, and then 7 seconds. This forced participates to focus on the most important parts of their message.
During day2 participates gave 3-minute weather briefings using just 1-2 slides and the information they were taught in day1. The Alan Alda staff provided constructive feedback on their presentations and slides. Common issues were putting too much information or too many weather hazards on a slide, such that the message was lost. The oral presentation should also have a sense of urgency and clear statements of why it is important to the user. Figure 15 shows a few photos from the event. The "Recent Interactions" section below has the evaluations from the workshop and comments from the forecasters.

Table 5. 4-5 March 2019 workshop agenda.

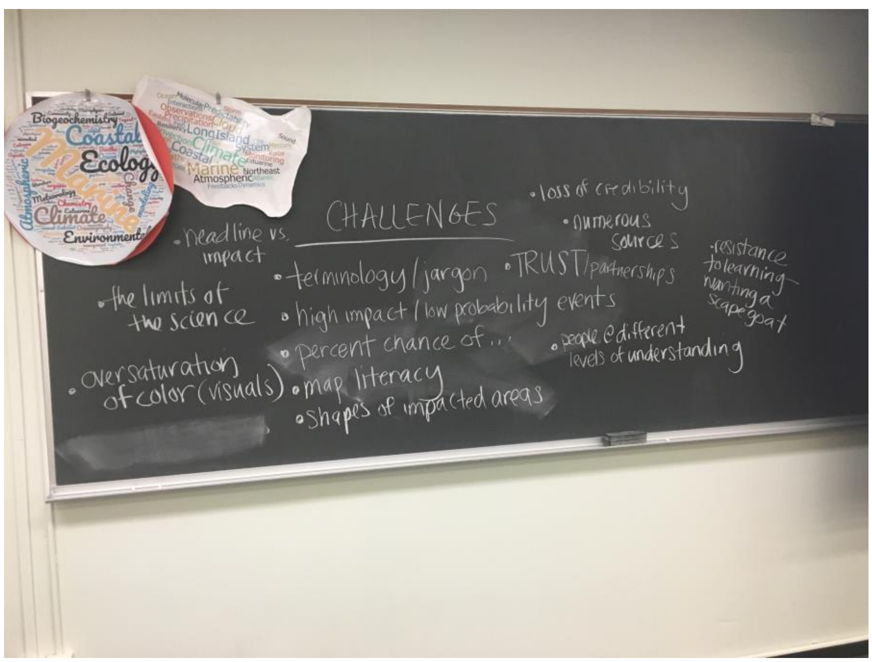
DAY 1

8:30-9:00am	Check-in & Registration All participants: Endeavor 120
9:00-10:00am Workshop	Welcome to the Alda Center's Science Communication - Communication challenges for forecasting, messaging probabilistic information, and communication with various audiences - The science of science communication - Examples of recent events All participants: Endeavor 120
10:00am-12:30pm	See and Be Seen Improvisation-based activities to help you focus on and connect with your audience. All participants: Endeavor 120
12:30-1:30pm	Lunch
1:30-2:00pm	Designing a Vivid Message Part I All participants: Endeavor
2:00-3:00	Designing a Vivid Message Part II Group A: Okubo Group B: EN 113

3:00-3:15pm	Break
3:15-5:15pm	Just a Minute (JAM) Session (Groups of 8) Practice talking about your work in clear, vivid and concise ways. Group A: Okubo Group B: EN 113
5:15-5:30pm	Reflection Routine & Wrap Up All participants: Room

DAY 2

9:00-10:00am	Talking to Challenging Audience and Listening All participants: Endeavor 120
10:00-12:45pm	Practice 3-minute Briefings Participants will practice and receive feedback presenting on a hazardous weather event
12:45-1:00pm	Celebrating the Journey


CHALLENGES
• headline vs. impact
• terminology / jargon
• TRUST / partnerships
• loss of credibility
• numerous sources
• resistance to learning — wanting a scapegoat
• the limits of the science
• high impact / low probability events
• percent chance of...
• people @ different levels of understanding
• oversaturation of color (visuals)
• map literacy
• shapes of impacted areas
Biogeochemistry
Coastal
Ecology
Marine
Climate
Environmental
Atmospheric
LongIsland
Climate
Coastal
Marine
Northeast
Atmospheric
System
Observations
Precipitation

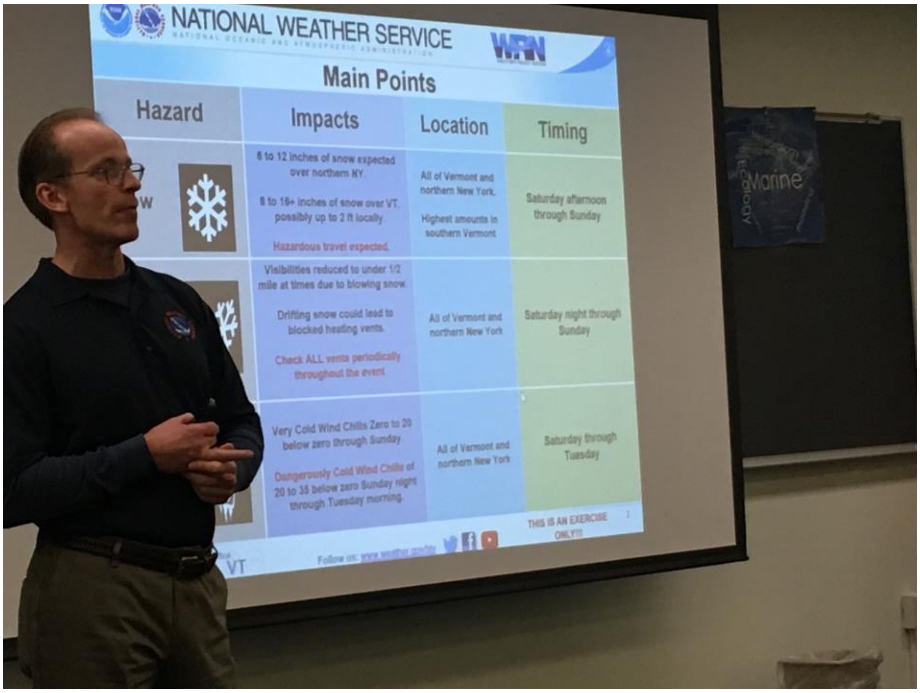


Figure 15. Photos from the 4-5 March Communication Workshop.

c. Communication Uncertainty Workshop #2

The second CSTAR Communication Workshop was held from 14-15 November 2019, involving 15 forecasters from NWS Eastern Region, Northeast River Forecast Center, and Weather Prediction Center. This workshop built on the ideas learned from workshop #1, so nearly all forecasters from the first workshop attended this second one. The first workshop emphasized distilling the message and communication skills through a series of exercises by the Alan Alda Center. The second workshop (also led by Alan Alda Center) focused on distilling the message for a major weather event to various stakeholders (e.g., emergency managers, media, etc.). As homework forecasters were required to bring a 1-page powerpoint slide communicating the forecast and hazard for an extreme weather event to an assigned stakeholder (emergency manager, department of transportation, and media). Table 6 gives the homework assignment. They arrived with version 1 of this slide (e.g, Fig. 16). After day1 of the workshop the participants were asked to make version2 given what they learned in class. Then after interacting with the stakeholders on the 2nd day of the workshop the participants were asked to make final version3. This allows us to track the modifications as they experienced the workshop from start to finish.

Table 7 highlights the agenda of the meeting and Figure 17 shows some photos. During the morning of day1 we reviewed what the participants learned from the first workshop (e.g., what stuck), and how they have applied some of the techniques back in the office and communicated this information to other staff. Then there were some Improvisation-based activities to help focus on and connect with the audience. The afternoon focused on how to match the message to the audience and how to do it in a shorter and shorter time period. We also rehearsed the weather briefing presentations to the stakeholders for day2, both the verbal (non-powerpoint) and powerpoint slide presentation. During day 2 the stakeholders were present from Nassau County Office of Emergency Management, New York City Emergency Management, Suffolk County Fire, Rescue and Emergency Services, NYS DOT LI Region, News 12 Long Island. There were some warm-up exercises with the stakeholders, and then we did the "circuit exercise," in which the forecasters in groups went to various rooms for 45 min to give the presentations to the stakeholders. The stakeholders asked questions and provided feedback.

Table 6: Homework assignment to the forecasters before the workshop

The focus of workshop no. 2 is Matching the Message to the Audience. Assignment 1 will provide the information for our exercises on day 1 (Thursday) and the interactions with stakeholders/partners on the morning of day 2 (Friday).

Assignment 1

Consider a previous high-impact winter weather event that your forecast office worked. This event will be the basis of your preparation on Thursday for providing a briefing Friday morning to a representative of an emergency management agency, Dept. of transportation, and media. Think about the key points your *partner* would need to know, including the inherent uncertainties and possible outcomes of this threatening event. What information has been most important to the different types of partners you have briefed in the past? Again, *Matching the Message to the Audience* is the theme for this workshop. No slides or other aids will be used in this briefing.

Assignment: Part 2

In this assignment, make a PowerPoint briefing slide to convey your message in about a minute and a half or less. We will send you an email shortly about which type of stakeholder to focus on.

Table 7. 14-15 November 2019 workshop #2 agenda at Stony Brook University.

DAY 1

8:30-9:00am	Check-in & Registration All participants: ***Endeavour 120***
9:00-10:00am Workshop	Welcome to the Alda Center's Science Communication - What stuck? - Brief review/overview for new participants

- How have you used what you learned?
- What challenges have you faced?
- Examples of recent events
- Stakeholder needs

All participants: ***Endeavour 120***

10:00am-12:30pm	Improv Refresh Improvisation-based activities to help you focus on and connect with your audience. Revisit the classics and introduce new exercises. All participants: ***Endeavour 120***
12:30-1:30pm	Lunch
1:30-3:00pm	Matching the Message to the Audience Revisit JAM Tool Tailoring to stakeholder audiences Half life talks Rehearsing and refining – Participants practice talks for

stakeholder audiences

Roleplaying – Participants listen and interact as stakeholders

Feedback from cohort and instructors

All participants: ***Endeavour 120***

3:30-3:45pm	Break
3:45-4:45pm	Visuals Wildcard Talks ***Endeavour 120*** ***NEED ANOTHER room***
4:45-5:00pm	Homework – revise slide Reflection Routine & Wrap Up All participants: ***Endeavour 120***

DAY2

9:00-9:45am

Talking to Stakeholders

Welcome, Introductions
Overview of the Alda Method for stakeholders
Warm up with improv exercises
Description of circuit training exercise
All participants: ***Endeavour 120***

9:45-10:00pm Break

10:00-12:15pm Weather briefings for Stakeholders
Endeavour 113
Endeavour 139
Endeavour 158
Dana 100

12:15-12:30pm Celebrating the Journey

All participants: *Endeavour 113*

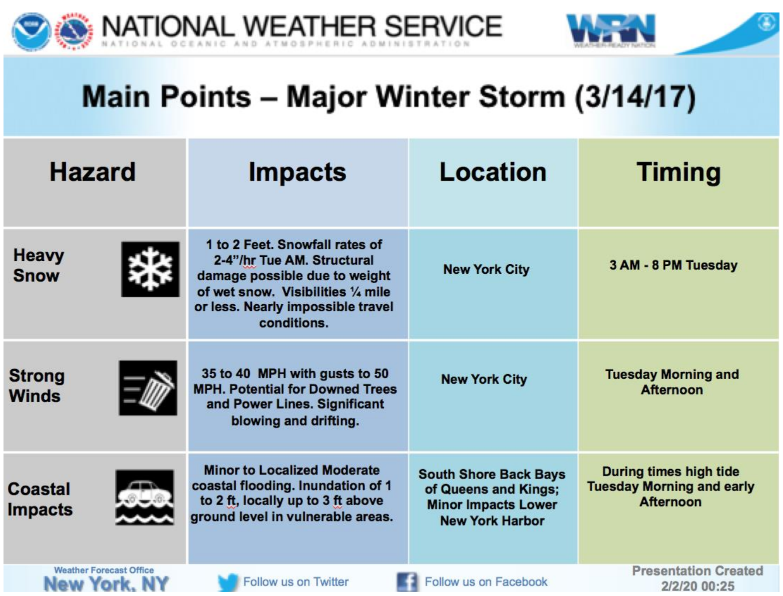


Figure 16. Sample one-pager used for briefing with stakeholders.

Figure 17. Select photos from the 14-15 November Workshop.

3. Interaction with Operational CSTAR Partners

a. Ensemble Tools

We have improved our fuzzy cluster tools for operations:

1. Updated the CSTAR web page to include the 6- and 12-h precipitation clusters as well as the 925 hPa freezing line clusters.
2. We implemented the latest code on the EMC machines in order to utilize the NCEP, CMC, and EC ensembles (90 total members). After the images are generated at EMC they are ftp'd back to Stony Brook to be displayed on a password-protected webpage.
3. We iterated with EMC on debugging the cluster code to run on their computer. This took a few months.
4. We added a floater domain to each of the cluster variables, which allows the user to select the region of interest using a Google map interface.

WPC implemented our clustering approaches for their 2019-2020 Winter Weather Experiment (WWE). An earlier report (January 2019) highlighted some examples from their internal cluster page. We have shared the precipitation and 0C cluster codes to hopefully be included in their 2019-2020 Winter Experiment.

We have maintained the ensemble tools webpage for our partners, which includes the Ensemble Sensitivity and Clustering.

http://breezy.somas.stonybrook.edu/CSTAR/Ensemble_Sensitivity/EnSense_Main.html
http://breezy.somas.stonybrook.edu/CSTAR/Ensemble_Sensitivity/FC_Main.html

b. Communication Uncertainty Workshops

In 2019 Stony Brook University (SBU), as part of the recent NOAA CSTAR project, and in collaboration with the Alda Center, held two 1.5-day workshops on effective forecast communication for 14 forecasters from several NWS forecast offices and operational centers. The goal of the Alda Center is to help scientists and other professionals learn how to communicate clearly with people outside their field using improvisational theater-based techniques and message design strategies. Improvisation techniques have been recognized as a tool for effective communication as they can help build empathy and connection with your audience. The Alda Center also focuses on message design strategies that communicate complex topics in "clear, vivid, and engaging ways." The Alda Center provides training on message design (both oral and visual) through games, exercises, and role-playing scenarios. The first workshop focused on the fundamentals of human communication using this unique improvisational approach, thus forcing forecasters to get out of their comfort zone. The second workshop focused more on adapting your message to the audience and included five different stakeholders in the NYC-Long Island area.

The two workshops were highly successful, and this work has resulted in a 2021 peer-reviewed Colle et al. (2021) paper in the *Bulletin of the American Meteorological Society*. Said one participant, "...When working forecast shifts, particularly during flooding events, I've caught myself making assumptions about how users or partners might interpret our forecasts. Due in part to the lessons taught at our workshops, I've consciously tried to avoid making those assumptions and to actually take the time to communicate with the partners to ensure clarity. This includes collaboration with our Weather Forecast Offices, but also core partners, such as the US Army Corps of Engineers, which as an example, has enabled us to better incorporate reservoir releases into our river forecasts." Overall, as summarized in the Colle et al. (2021) paper: participants embraced the unique, highly interactive exercises, games, peer and stakeholder presentation sessions. Several forecasters also viewed the improvisational teaching methods as an adaptive staff training opportunity. The participants have reported that clear, concise messaging and personal engagement has been key to enhancing their connection with NWS core partners.

These two in-person workshops only involved a limited number of forecasters and stakeholders. Unfortunately, because of COVID-19 we were not able to have our final (3^{rd}) workshop for this project. As noted below in the "problems and difficulties" part of the report we are exploring an online version. We are considering an online workshop in earl 2022, which would (1) provide workshops in an online format that can train many more NWS forecasters; (2) allow for more stakeholder interaction and help improve IDSS support; and (3) complete a more social science assessment to understand how effectively the forecasters utilize operationally what they learned.

Covid-19 travel restrictions and travel budget limitations within the NWS necessitate the potential use of online versions of our previous workshops and the development of a new capstone workshop. Even after Covid-19, a new online approach can positively impact more NWS forecasters than the less frequent and logistically more challenging in-person meetings, and the larger sample size for online meetings can provide more data for assessment. Over the course of 2020-2021, the Alda Center has successfully transitioned several in-person workshops to online offerings. Our experience and participant feedback indicates that much of the experiential nature of the Alda Method® can be retained, and, in some cases, enhanced in online forms. Since many stakeholder interactions often occur online (online group chats, social media, and phone calls), there is a need to evaluate these online communication approaches.

4. Products and Presentations

a. Fuzzy Clustering and other Ensemble Tool

The fuzzy clustering and other ensembles tools (ensemble sensitivity, wave packets, cyclone tracks, etc) are currently maintained and accessible from our CSTAR page:

http://breezy.somas.stonybrook.edu/CSTAR/Models.html

b. Clustering tools

This CSTAR has motivated the clustering approach to be made operational at WPC on a webpage (Fig. 18). The links are https://origin.wpc.ncep.noaa.gov/wpc_ensemble_clusters/day_3_7/view.php for days 3-7, and https://origin.wpc.ncep.noaa.gov/wpc_ensemble_clusters/day_8_10/view.php for days 8-10. Clusters are now generated for 4 regions (East, Central, West, and Alaska twice daily (0000 and 1200 UTC) out to day 10 using the GEFS, CMC, and EC ensembles.

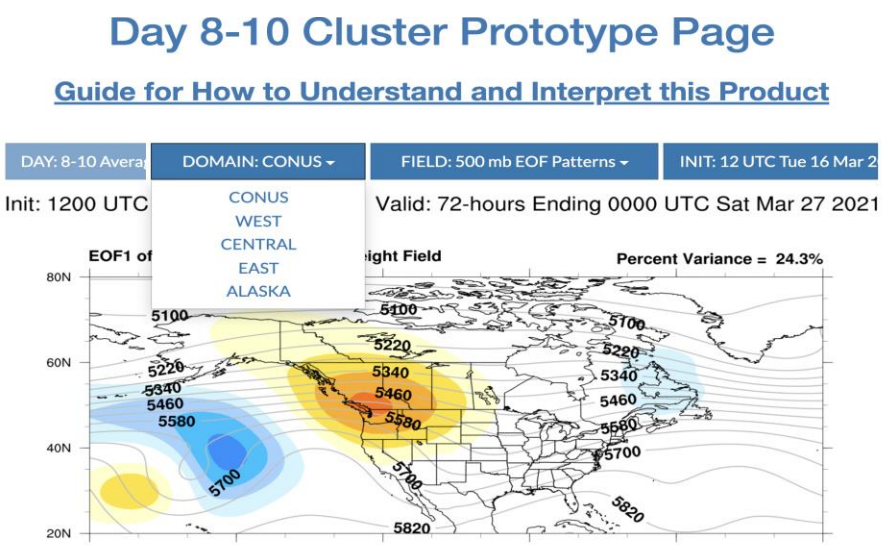


Figure 18. The regional cluster web page showing the pulldown regional options.

Figure 19 shows the new DESI (Dynamic Ensemble-based Scenarios for IDSS) software interface for NOAA clustering on a workstation (funding and effort for this was from a NOAA JTTI) for an event on 18-19 February 2022. It illustrates how the user can select the region of interest, the clustering period (day 9 in this example), the meteogram location (ISP), and the initialization date (0000 UTC 10 Feb 2022). The user can group by clusters or ensemble system (GEFS, EC, and CMC). Figure 19 also shows the three 500 hPa clusters at day 9 and the height anomaly calculated with respect to the full (100 member) ensemble mean, which illustrates three scenarios with the mid-level trough over the eastern U.S, with cluster 3 (39% of member) having a deep trough.

This DESI is now operational at over 40 NWS forecast offices in August 2022.

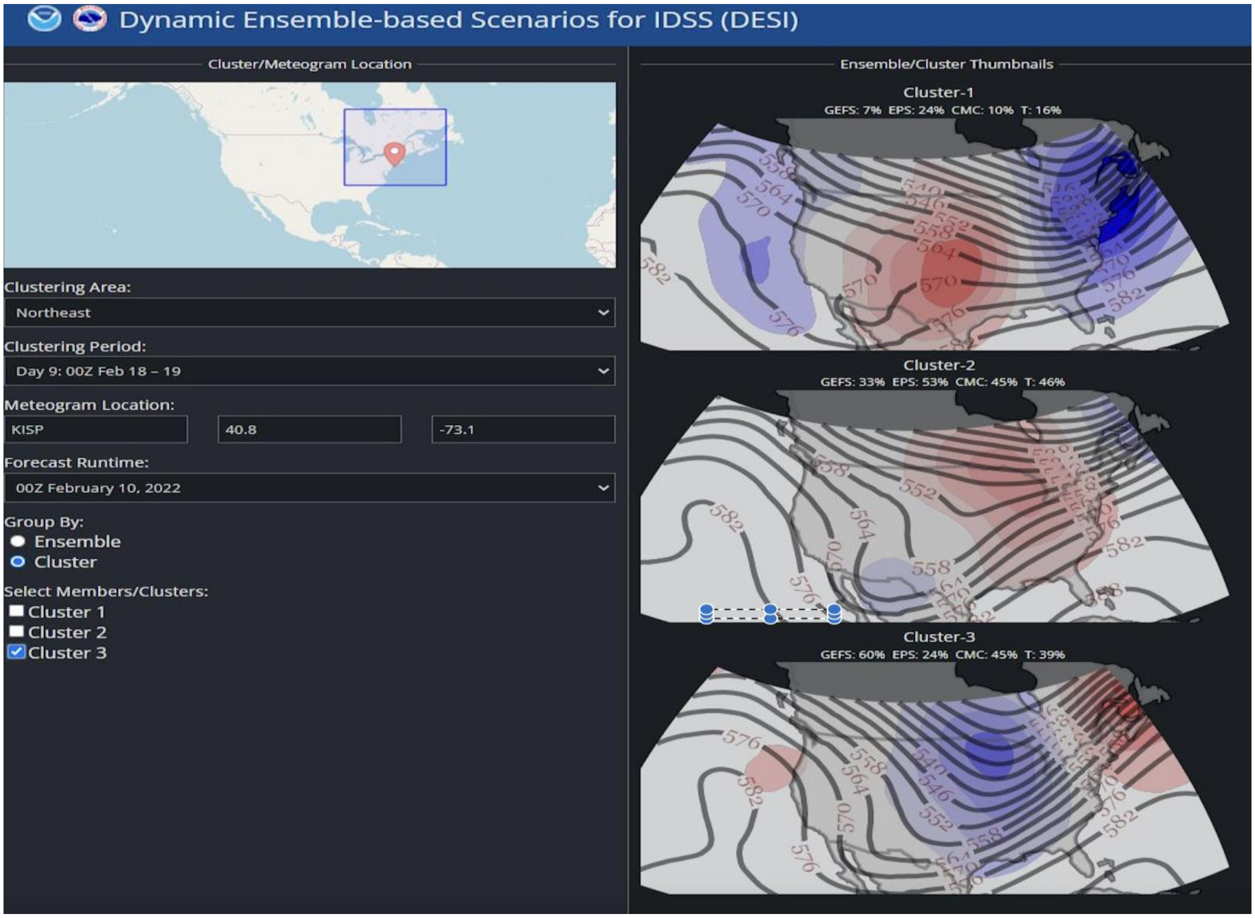


Figure 20. DESI interface showing how forecasters can choose domain, time period, cluster(s) for viewing, and the three 500-hPa cluster patterns with the anomaly calculated with respect to the full ensemble mean.

c. Formal Papers

Zheng, M., E. Chang, and B.A. Colle: 2022: Downstream predictability associated with Rossby wave packets for East coast winter storms In revision for *Mon. Wea. Rev*.

Kiel, B, and B. Colle, 2022: Comparison of Clustering Approaches in a Multimodel Ensemble for U.S. East Coast Winter Storms. In revision to *Wea. Forecasting.*

Colle, B.A., R. Auld, K. Johnson, C. O'Connell, T.G. Taylor, and J. Rice, 2021: Improving communication of uncertainty and risk of high-impact weather through innovative forecaster workshops. *Bull. Amer. Meteor. Society*, **102**, 1424-1430. https://doi.org/10.1175/BAMS-D-20-0108.1.

Zheng, M., E. Chang, and B.A. Colle: 2019: Evaluation of a multi-model ensemble for extratropical cyclones using a fuzzy clustering approach. *Wea. Forecasting*, **147**, 1967-1987.

Wirth, V., M. Riemer, E. K. M. Chang, and O. Martius, 2018: Rossby wave packets on the mid-latitude Rossby waveguide, *Mon. Wea. Rev*., **146**, 1965-2001.

Mandelbaum T., B.A. Colle, 2021: Assessing the spread-error relationship for East Coast winter storms. To be submitted to *Wea. Forecasting*.

Zheng, et al, 2022: Ensemble sensitivity of U.S. East Coast winter storms: the multi-model climatology and paths of forecast uncertainty in medium range. In preparation.

d. Stony Brook CSTAR graduates (alum)/students:

David Stark (M.S., 2012) – NWS General Forecaster at Upton, NY

Matthew Souders (M.S., 2013) –Weather Analytics, New Hampshire

Michael Layer (M.S., 2014) – Weatherworks, Hackettstown, NJ

Michael Erickson (Ph.D., 2015) – NOAA Contractor (Weather Prediction Center)

Minghua Zheng (Ph.D. -2016, Post-doc at Scripps)

Nathan Korfe (M.S. 2016) – Research Meteorologist at WindLogics, MN)

Taylor Mandelbaum (M.S. 2018)—Meteorologist and Data Analyst at NY Power Authority

Benjamin Kiel– M.S. 2021) – Forecasting, coding and instrumentation related to missile testing at Regan Test Site on Kwajalein Atoll in the Marshall Islands.