Regression Mixture Model Clustering of Multimodel Ensemble Forecasts of Hurricane Sandy: Partition Characteristics

ALEX M. KOWALESKI AND JENNI L. EVANS

The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 11 March 2016, in final form 27 June 2016)

ABSTRACT

Track and cyclone phase space (CPS) forecasts of Hurricane Sandy from four global ensemble prediction systems are clustered using regression mixture models. Bayesian information criterion, cluster assignment strength, and mean-squared forecast error are used to select optimal model specifications. Fourth-order (third order) polynomials for 168-h forecasts (60-h forecast segments) and 5 (6) clusters for track (CPS) forecasts are selected.

Mean cluster paths from eight initialization times show that track and CPS clustering meaningfully partition potential tracks and structural evolutions, distilling a large number of ensemble members into several representative and distinct solutions. Rand index and adjusted Rand index calculations demonstrate a relationship between track and CPS cluster membership for both 168-h forecasts and 60-h forecast segments, indicating that certain tracks are preferentially associated with certain structural evolutions. These relationships are explained in greater detail using forecasts initialized at 0000 UTC 25 October.

Storm-centered cluster composite maps of 500-hPa geopotential height and 850-hPa equivalent potential temperature for the 120-h forecast valid at 0000 UTC 30 October (initialized at 0000 UTC 25 October) indicate that both track and CPS clustering successfully capture variations in the Sandy–trough interaction and the strength of the lower-troposphere warm core of Sandy at the time of observed landfall. Together, these results illustrate the relationship between the track and structural evolution of Sandy and suggest the potential of multiensemble mixture-model path clustering for tropical cyclone forecasting.

1. Introduction

Hurricane Sandy battered the U.S. mid-Atlantic coast on 29 and 30 October 2012, causing 72 fatalities and at least $50 billion (2012 U.S. dollars) in damage. (Blake et al. 2013). Because of a complex extratropical transition (ET) before landfall, Sandy became one of the largest Atlantic tropical cyclones ever observed, with a tropical storm–force wind field 930 km in radius. The left turn of Sandy, caused by interaction with a midlatitude trough, resulted in an unprecedented near-perpendicular landfall angle on the New Jersey coast (Hall and Sobel 2013). The massive wind field and the landfall angle produced a devastating storm surge along the New Jersey, New York, and Connecticut coasts.

Despite the complex track and structural evolution of Sandy, the European Centre for Medium-Range Weather Forecasts (ECMWF) model and its ensembles correctly forecasted the left turn up to a week before landfall (Magnusson et al. 2014). Other models, including the U.S. Global Forecast System (GFS), and the Canadian Global Environmental Model (GEM), produced less accurate medium-range forecasts in which Sandy moved out to sea.

Sensitivity studies of model forecasts of Sandy have shown that storm track was highly sensitive to conditions near Sandy early in the forecast. Munsell and Zhang (2014) used a 60-member Weather Research and Forecasting (WRF) Model ensemble to show that simulations in which Sandy made landfall had substantially greater southeasterly steering flow during the first 12 h (initialized at 0000 UTC 26 October) than simulations in which Sandy moved out to sea.

Sensitivity studies of model forecasts of Sandy have shown that storm track was highly sensitive to conditions near Sandy early in the forecast. Munsell and Zhang (2014) used a 60-member Weather Research and Forecasting (WRF) Model ensemble to show that simulations in which Sandy made landfall had substantially greater southeasterly steering flow during the first 12 h (initialized at 0000 UTC 26 October) than simulations in which Sandy moved out to sea. Torn et al. (2015) found that differences in upper-level divergence and midtroposphere moisture north of Sandy during the first 12 h of the simulation (initialized at 0000 UTC 25 October) critically influenced Sandy’s track. Using an 80-member experimental Global Ensemble Forecast System (GEFS), they showed that simulations with less upper-level divergence and midtroposphere moisture yielded a
weaker ridge north and northwest of Sandy. This induced a more westerly steering flow compared to simulations with greater upper-level divergence and midtroposphere moisture.

An analysis of ensemble prediction system (EPS) forecasts from multiple global models provides a framework for assessing uncertainties in tropical cyclone (TC) track forecasts and thermodynamic structure forecasts in cyclone phase space (CPS; Hart 2003) through ET (Evans and Hart 2003). Forecast skill of storm structure often decreases substantially during ET (Hart et al. 2006; Veren et al. 2009). Because storm structure affects storm impacts, correctly understanding ensemble uncertainty in both TC track and storm structure is critical to forecasting ET and its associated hazards. Furthermore, the interaction between a TC and the midlatitude flow during ET can generate or amplify Rossby waves (e.g., Riemer et al. 2008; Harr and Dea 2009; Riemer and Jones 2010). These Rossby wave perturbations are often associated with decreased synoptic forecast skill in downstream regions (Jones et al. 2003; Harr et al. 2008). Therefore, accurately forecasting storm track and structure during ET is critical for predicting storm to hemisphere scales.

Clustering of ensemble forecasts provides information beyond and superior to the traditional diagnostics of ensemble mean and ensemble spread. For example, if ensemble members are bifurcated, the ensemble mean may yield little useful information. Clustering can provide representative forecast outcomes (cluster mean values or trajectories) and the probability of each outcome (cluster populations). This additional information may aid forecasters in understanding potential outcomes and their probabilities of occurrence.

Clustering has been successfully applied to TC track and structure forecasts (e.g., Evans et al. 2006; Kuruppumullage Don et al. 2016), and to climatological storm tracks (Camargo et al. 2007a,b, 2008; Nakamura et al. 2009; Kim et al. 2011; Kozar et al. 2012; Paliwal and Patwardhan 2013). Many of these studies (e.g., Arnott et al. 2004; Nakamura et al. 2009; Kim et al. 2011) use nonhierarchical $k$-means or $c$-means clustering. Non-hierarchical clustering is also used on ensemble forecasts of synoptic conditions during and after ET (Harr et al. 2008; Anwender et al. 2008; Keller et al. 2011, 2014). These clustering techniques are well suited to clustering forecasts verifying at a single time; however, they require all clustered vectors to be of equal length, rendering them less useful for clustering time-varying paths such as TC tracks.

Regression mixture modeling (Gaffney et al. 2007) is well suited to clustering of time-varying paths because it can account for time-dependent variations. Camargo et al. (2007a,b, 2008) and Paliwal and Patwardhan (2013) use regression mixture models to partition historical TC tracks. These studies show how climate modes such as the El Niño–Southern Oscillation modulate storm tracks. Kuruppumullage Don et al. (2016) use regression mixture modeling to partition 120-h ECMWF Integrated Forecast System (IFS) ensemble track forecasts of three TCs, demonstrating the utility of mixture-model clustering of ensemble track forecasts.

In this study, regression mixture modeling is used to cluster track and CPS forecasts of Hurricane Sandy from a 117-member multimodel ensemble comprising four EPSs. Track and CPS clustering are performed on (up to) 168-h forecasts beginning at the ensemble initialization time and also for a constant 60-h forecast interval near the observed landfall time. First, optimal mixture-model specifications are determined. Then, cluster membership statistics among track and CPS clusters are analyzed to document relationships between track and thermodynamic structural evolution. Storm-centered cluster composite maps from 0000 UTC 30 October are generated to illustrate how each cluster captures the interaction between Sandy and the midlatitude trough near the observed landfall time.

This exploration of Hurricane Sandy ensemble forecasts proceeds as follows: overview of Hurricane Sandy, the dataset used, and the clustering methodology (section 2); description of statistics used to select the optimal mixture model specifications (section 3); documentation of track and CPS clustering results and relationships between track and CPS cluster membership (section 4); and illustration of how clustering captures the forecast variability in storm evolution via storm-centered synoptic composites (section 5). A summary and conclusions are presented in section 6.

2. Overview of Hurricane Sandy and methodology

a. Overview of Hurricane Sandy

The tropical depression that became Hurricane Sandy formed in the south-central Caribbean Sea at 1200 UTC 22 October. It became a tropical storm 6 h later and reached hurricane intensity near Jamaica at 1800 UTC 24 October. Early on 25 October Sandy made landfall in eastern Cuba as a category-3 hurricane. After crossing Cuba, Sandy turned northwestward, steered by a ridge to its northeast and a negatively tilted upper-level trough to its west. It weakened to a tropical storm at 0000 UTC 27 October due to increased vertical wind shear and interaction with a cooler and drier air mass. However, Sandy enlarged substantially through the trough interaction. Sandy regained hurricane intensity
at 1200 UTC 27 October, due to lower vertical wind shear, increased upper-level divergence, and a moister environment. On 27 October Sandy turned northeastward ahead of a larger, midtropospheric trough. The next day Sandy turned northward, blocked from moving eastward by an anomalous midtropospheric ridge to its northeast.

On 29 October, the negatively tilted midtropospheric trough over the southeastern United States steered and accelerated Sandy northwestward. Sandy reached its trough over the southeastern United States steered and northeast.

On 29 October, the negatively tilted midtropospheric trough over the southeastern United States steered and accelerated Sandy northwestward. Sandy reached its lowest central pressure (940 hPa) and largest spatial extent on 29 October as it acquired warm seclusion characteristics (Galmarino et al. 2013). After 1800 UTC 29 October Sandy weakened and lost tropical characteristics due to low sea surface temperatures and the cold continental air mass enveloping it. Sandy made landfall as a posttropical cyclone near Brigantine, New Jersey, at 2330 UTC 29 October, with maximum sustained winds of 236 m s\(^{-1}\) and a central pressure of 945 hPa. After landfall, Sandy turned west-northwestward and weakened, merging with a low pressure area over eastern Canada a few days later (Blake et al. 2013).

### Dataset

Forecast data for this study are obtained from the THORPEX Interactive Grand Global Ensemble (TIGGE) database at ECMWF (ECMWF 2016b; Bougeault et al. 2010). The data examined comprise four EPSs used in TC forecasting and available at 0000 and 1200 UTC: the ECMWF IFS, the National Centers for Environmental Prediction (NCEP) GEFS, the Canadian Meteorological Centre (CMC) Global Environmental Prediction System (GEPS), and the Met Office Global and Regional Ensemble Prediction System (UKMO MOGREPS; Table 1). These four EPSs comprise 117 forecasts (4 control forecasts and 113 perturbed ensemble forecasts). Forecasts from eight initialization times (every 12 h from 1200 UTC 23 October to 0000 UTC 27 October) are analyzed.

The actual evolution of Sandy is assessed via ERA-Interim reanalysis data, obtained from the ECMWF (ECMWF 2016a). Model resolution varies among the four EPSs (Table 1; Magnusson et al. 2014) and the ERA-Interim reanalysis, so all forecasts and reanalysis data are interpolated to a common grid with a 0.5° × 0.5° horizontal resolution.

Forecast storm positions every 6 h from each ensemble member are generated using the Geophysical Fluid Dynamics Laboratory (GFDL) Vortex Tracker, version 3.4a (Marchok 2002). After all storm tracks are determined, CPS values are calculated for each ensemble member every 6 h. The CPS describes the thermodynamic structure of the cyclone using three parameters: thickness asymmetry (\(B\)), lower-tropospheric thermal wind (\(V_L^U\)), and upper-tropospheric thermal wind (\(V_U^T\); Hart 2003). Higher \(B\) values signify greater differences in across-track 900–600-hPa geopotential thickness. The quantity \(V_L^T\) (\(V_U^T\)) describes the cyclone’s 900–600-hPa (600–300-hPa) thickness anomalies, and thus its warm-core or cold-core structure. Typically, \(V_L^T\) and \(V_U^T\) are reported as \(-V_L^T\) and \(-V_U^T\), with positive values indicating a warm core. We use the two-dimensional parameter space of \(B\) and \(-V_L^T\) to diagnose Sandy’s forecast structural evolution (Arnott et al. 2004; Evans et al. 2006). All \(B\) and \(-V_L^T\) values are smoothed using a 24-h running mean (Hart 2003).

c. Clustering methodology

The 117 forecast track and CPS paths from each initialization time are clustered using the regression mixture model of Gaffney et al. (2007). Forecasts are clustered using data out to a maximum of 168 h; however, if the GFDL Vortex Tracker stops tracking Sandy before 168 h in more than 10% of ensemble members, clustering is truncated at the last time step before this occurs. For the first five forecasts (1200 UTC 23 October–1200 UTC 25 October) less than 10% of ensemble members lose Sandy; for the final three forecasts 10% of ensemble members lose Sandy after 162, 156, and 144 h, respectively, and clustering is truncated at these times.

Here 60-h segments of forecast tracks and CPS paths between 0000 UTC 28 October and 1200 UTC 30 October are also clustered. This clustering is performed to compare with the 168-h clustering results and to show how track and CPS clustering results for a constant verification period near the landfall time change as the interval between initialization and verification shrinks. This 60-h interval is chosen because it contains the interactions between Sandy and both the blocking midlevel ridge and the negatively tilted trough, as well as Sandy’s final observed ET. Furthermore, this 60-h interval is included within the 168-h forecast for all initialization times.

In regression mixture model clustering, [see Gaffney et al. (2007) and Camargo et al. (2007a) for details], one or more variables (here latitude and longitude or \(B\) and \(-V_L^T\)) are conditioned onto a time variable to form time-dependent paths. For CPS clustering, \(B\) and \(-V_L^T\)
values are normalized using the method of Arnott et al. (2004) because \(-\frac{1}{2}V^T L_v \) values often exceed \(B\) values by an order of magnitude. Because each EPS represents vortex structure differently, \(B\) and \(-\frac{1}{2}V^T L_v \) from each EPS are normalized using only data from that EPS. We have found that this produces the best division of EPS forecasts among EPSs. CPS data for 60-h clustering are taken directly from the 168-h data; they are not re-normalized using only the 60-h interval.

Regression mixture-model clustering works by (i) calculating the mixture parameters (polynomial coefficients, covariance matrix, and membership weights) of each of \(K\) models and (ii) calculating the posterior probability of assignment of each path to each cluster given the shape parameters calculated in (i). Mixture parameters and posterior probabilities are iteratively calculated through an expectation–maximization (EM) algorithm, which converges on a likelihood maximum in parameter space, maximizing cluster assignment strength (e.g., Kuruppumullage Don et al. 2016). However, the EM algorithm sometimes converges on a local, rather than global likelihood maximum. Therefore, clustering is repeated 500 times with randomly generated initial membership weights used to calculate shape parameter values. The set of cluster assignments with the highest likelihood is taken as the final assignment. Each forecast path is then assigned to the model (cluster) with the highest probability of having generated that path.

3. Selection of optimal mixture-model specifications for each forecast

a. Overview

In regression mixture modeling, the mixture-model specifications (number of clusters and polynomial order) must be chosen prior to clustering. To determine the optimal specifications, clustering is performed on all combinations of polynomial order first through fifth and 2 through 7 clusters for each forecast type (e.g., 168-h track forecasts) at each initialization time. Values of Bayesian information criterion (BIC), mean-squared forecast error (MSFE), and fraction of clusters with probability of assignment below 0.95 \((F_{0.95})\) are used to determine the optimal model specifications for each forecast type.

The BIC is calculated from the maximum log-likelihood (MLL), with a penalty imposed according to the number of independent parameters \((k)\) and number of observations \((n)\) in the model. If \(L_n(m)\) is the MLL value, then

\[
\text{BIC}(m) = -2L_n(m) + k \log n. \tag{1}
\]

Unlike the MLL, which always favors more complex models, the BIC favors models that balance goodness-of-fit and simplicity. A smaller BIC value indicates greater support for a model (Kuruppumullage Don et al. 2016).

MSFE provides a means to estimate within-cluster spread and frequency of outliers. MSFE is similar to the “sum of squared distance” frequently used to determine optimal point clustering solutions (e.g., Arnott et al. 2004). However, it accounts for within-cluster spread at all times in the forecast period. MSFE is calculated by summing the squared distance between each cluster member and its cluster centroid at all times and then dividing by the total number of points in all paths. Smaller MSFE values indicate more compact clusters and fewer outliers, and thus a superior solution.

In mixture-model clustering, paths are assigned to clusters through posterior probabilities generated by the EM algorithm. Thus, each path has a probability of assignment to each cluster. The cluster assignment strength produced by a particular mixture model can be summarized by calculating the fraction of paths with maximum probability of assignment below a threshold value [chosen here as 0.95 following Kuruppumullage Don et al. (2016)]. A smaller fraction of paths with a maximum probability of assignment below 0.95 \((F_{0.95})\) indicates a superior model specification.

Optimal mixture model specifications must be determined for each forecast type (168-h forecasts and 60-h forecast segments of track and CPS). Chosen model specifications are allowed to vary among forecast types; however, a single optimal specification is ultimately sought for all forecasts of each type to facilitate comparisons among results from different initialization times.

b. Selection of optimal polynomial order

Kuruppumullage Don et al. (2016) showed that for mixture-model clustering of TC forecast tracks, selecting the optimal polynomial order is more straightforward than selecting the optimal number of clusters. Thus, for each forecast type polynomial order is selected first, by analyzing how BIC, MSFE, and \(F_{0.95}\) change with polynomial order using two–seven cluster solutions. Mean values of BIC, MSFE, and \(F_{0.95}\) with polynomial order and number of clusters are presented for each forecast type (Fig. 1).

For 168-h track forecasts, mean BIC and MSFE (calculated using two–seven clusters for all eight forecasts) decrease monotonically with polynomial order. BIC decreases by a mean of 2.9% from first- to second-order solutions, 2.7% from second- to third-order solutions, and 2.8% from third- to fourth-order solutions. However, BIC only decreases by a mean of 1.6% between fourth- and fifth-order solutions (Fig. 1a). MSFE shows similar qualitative results, decreasing by a mean of
20.1% between first- and second-order solutions, 11.2% between second- and third-order solutions, and 9.3% between third- and fourth-order solutions. However, a much smaller MSFE decrease (4.1%) is observed between fourth- and fifth-order solutions (Fig. 1b). Furthermore, fourth-order solutions produce relatively low $F_{0.95}$ values (Fig. 1c). Taken together, these diagnostics suggest that for 168-h track forecasts, a fourth-order solution balances goodness-of-fit with model simplicity.

For 168-h CPS forecasts, a fourth-order solution also appears optimal. BIC decreases by a mean of 15.4% between first- and second-order solutions, and decreases slightly (0.4%) when going to third order (though the final two forecasts have larger BIC decreases when going to third order; not shown). However, beyond third order BIC increases with higher polynomial orders (Fig. 1g). MSFE shows a qualitatively similar pattern is observed for MSFE, with MSFE decreasing by 34.2% between first- and second-order solutions, 8.0% between second- and third-order solutions, and 10.9% between third- to fourth-order solutions, compared to only 4.5% between fourth- and fifth-order solutions (Fig. 1f). We therefore, select a fourth-order solution for 168-h CPS forecasts.

Diagnostics suggest lower polynomial orders for 60-h track and CPS forecast segments. For 60-h track segments, mean BIC decreases by 6.1% between first- and second-order solutions, and decreases slightly (0.4%) when going to third order (though the final two forecasts have larger BIC decreases when going to third order; not shown). However, beyond third order BIC increases with higher polynomial orders (Fig. 1g). MSFE shows a
similar pattern, although MSFE decreases slightly between third- and fourth-order solutions (Fig. 1h). For 60-h track forecast segments, mean \(F_{0.95}\) also improves little beyond third order (Fig. 1i). Thus, we select a third-order solution, though we acknowledge that second order may suffice.

For 60-h CPS clustering, large mean decreases in BIC and MSFE are observed between first- and second-order solutions and between second- and third-order solutions (Figs. 1j and 1k). However, between third and fourth order BIC increases and MSFE improvement decreases dramatically. These results, along with near-constant \(F_{0.95}\) values between third and fifth order (Fig. 1l), suggest that a third-order solution is optimal for 60-h CPS clustering.

c. Selection of the optimal number of clusters

Having determined the optimal polynomial order for each forecast type being clustered, we next determine the optimal number of clusters for each forecast type. For 168-h track forecasts, percent change in BIC and MSFE with each additional cluster, along with values of \(F_{0.95}\) with number of clusters for (top to bottom) each initialization time with the bottom row being the mean values. Values are calculated using fourth-order polynomial mixture models.

For 168-h track forecasts, changes in BIC and MSFE with each additional cluster, along with values of \(F_{0.95}\) with cluster number, are plotted for each initialization time (Fig. 2). Mean statistics from the eight initialization times are presented for each forecast type (Fig. 3).

As in Kuruppumullage Don et al. (2016), the optimal number of clusters is challenging to determine, as BIC and MSFE decrease monotonically with each additional cluster. For 168-h track forecasts, mean MSFE decreases substantially between two- and three-cluster solutions (29.8%) and between three- and four-cluster solutions (26.0%). A smaller decrease is observed between four and five clusters (15.5%), and the decrease drops to 8.8% with the addition of cluster six (Fig. 3b). Going from five to six clusters also produces a substantially smaller mean BIC improvement than going from four to five clusters (Fig. 3a). Thus, it appears that a five-cluster solution is reasonable for 168-h track forecasts.
For 168-h CPS forecasts, the selection of an optimal number of clusters is even less straightforward, with no major signal in BIC, MSFE, or $F_{0.95}$ to indicate a single optimal number (Figs. 3d–f). However, for three of the eight initialization times (not shown), the fractional MSFE improvement going from five to six clusters is larger than the improvement from four to five clusters. Furthermore, the six-cluster solution has a 18% lower mean $F_{0.95}$ value than the next lowest specification (2.99 vs 3.63; Fig. 3f) Thus, we select a six-cluster solution for 168-h CPS forecasts.

Here 60-h track forecast segments show large MSFE decreases with the addition of cluster 3 (28.4%) and cluster 4 (27.7%). This decrease falls to 18.2% with the addition of cluster 5 and 13.9% with the addition of cluster 6 (Fig. 3h). For BIC, a similar pattern is observed, though the BIC improvement between two and three clusters is larger than that between three and four clusters (Fig. 3g). Although the metrics chosen do not decisively indicate whether four or five clusters are optimal, we select a five-cluster solution. Although this increases the likelihood of redundancy, it decreases the likelihood that the clustering solution will fail to partition meaningfully different track evolutions. Post-analysis of the five-cluster solutions confirms that the fifth cluster usually does not result from splitting a single cluster in the four-cluster partition.

Similar to 60-h track forecast segments, diagnostics of 60-h CPS forecast segments leave some ambiguity in the optimal number of clusters. Although the BIC improvement decreases monotonically with each additional cluster (Fig. 3j), MSFE shows a significant drop-off in
4. Track and CPS clustering results

a. General clustering results

Clustering of 168-h track forecasts initialized between 23 and 27 October produces clear partitions between tracks in which Sandy moves out to sea and those in which it makes landfall at varying locations along the Atlantic coast (Fig. 4). The spread among clusters slowly shrinks between 1200 UTC 23 October (Fig. 4a) and 0000 UTC 27 October (Fig. 4h). An east-tracking cluster that avoids landfall (green) is observed for all initialization times except the latest one; the population of this cluster decreases somewhat with time as fewer ensemble members predict Sandy to move out to sea (Table 3).

Clustering of 60-h track forecast segments (Table 4) show similar qualitative results; cluster spread decreases with time, a result expected as the interval between forecast initialization and verification shrinks (Fig. 5). At later initialization times, 60-h track clusters have slightly less spread than the 168-h clusters (cf. Figs. 5g,h and 4g,h). Nevertheless, even in the later forecasts 60-h clustering captures substantial variability in landfall timing and location.

The cluster solutions highlight the threat of Sandy to the mid-Atlantic coast even at 6–7-day lead times. For 168-h track clustering, the mean path of the most populous cluster makes landfall for all initialization times at and after 0000 UTC 24 October (and likely for 1200 UTC 23 October also, though the track terminates before landfall). At most initialization times both 168- and 60-h track clustering produce a most populous cluster that landfalls near the observed landfall location (Figs. 4, 5). Therefore, both 168- and 60-h track clustering demonstrate utility in partitioning multi-EPS track forecasts into a small number of representative outcomes and indicating their probability through cluster populations.

Clustering of 168-h CPS forecasts (Table 5) produces meaningful partitions among forecasts of the thermodynamic structural evolution of Sandy. Although the cluster solutions vary substantially among initialization times, we observe a trend of decreasing intercluster spread toward later initialization times (Fig. 6). For many initialization times, CPS clustering successfully partitions ensemble forecasts by the timing and completeness of ET, as well as by cyclone asymmetry ($B$) during ET. For every initialization time there is at least one cluster in which Sandy follows a structural evolution similar to the CPS path generated from the ERA-Interim reanalysis data (Fig. 6i). First Sandy becomes asymmetric ($B > 10$); then it experiences an additional asymmetry increase while becoming slightly more warm-core (increasing $-V_L^B$). Finally, both $B$ and $-V_L^B$ decrease as Sandy completes ET and weakens. For all initialization times at and after 0000 UTC 25 October, the most populous cluster captures this evolution. However, the ERA-Interim reanalysis data (as well other model- and satellite-based reanalyses; not shown) shows a decrease in $B$ and $-V_L^B$ between 27 and 28 October that the most populous cluster only captures in the final two forecasts (Figs. 6g and 6h). This indicates that CPS clustering captures the ET completion of Hurricane Sandy more effectively than the earlier decrease in $B$ and $-V_L^B$. This may be because the earlier structural changes occurred on a smaller scale than the final trough interaction, and thus were captured less effectively by many ensemble members.

Clustering results using 60-h CPS forecast segments (Table 6) show meaningful divisions among the structural evolutions of Sandy between 0000 UTC 28 October and 1200 UTC 30 October (Fig. 7). Forecast spread decreases substantially between early and late initialization times, an expected outcome as the interval between initialization and verification shrinks. However, even at later initialization times (such as 0000 and 1200 UTC 26 October; Figs. 7f and 8g), there is substantial intercluster spread in both 60-h CPS paths and CPS values at 0000 UTC 30 October. For most initialization times at least one cluster captures the qualitative characteristics of the structural evolution in the

### Table 2. Mixture-model specifications of each data type clustered.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Polynomial order</th>
<th>No. of clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>168-h track forecasts</td>
<td>Fourth</td>
<td>5</td>
</tr>
<tr>
<td>168-h CPS forecasts</td>
<td>Fourth</td>
<td>6</td>
</tr>
<tr>
<td>60-h track forecast segments</td>
<td>Third</td>
<td>5</td>
</tr>
<tr>
<td>60-h CPS forecast segments</td>
<td>Third</td>
<td>6</td>
</tr>
</tbody>
</table>

improvement between four and five clusters. However, a larger improvement is observed between five and six clusters, while MSFE increases with the addition of a seventh cluster. (Fig. 3k). Therefore, we select the six-cluster solution for 60-h CPS forecasts.

The diagnostics presented here give a clearer indication of optimal polynomial orders than of optimal numbers of clusters, consistent with the results of Kuruppumullage Don et al. (2016). However, our results do indicate reasonable model specifications for each forecast type. Thus, for 168-h track (CPS) forecasts we select fourth-order polynomials with five (six) clusters. For 60-h track (CPS) forecast segments we select third-order polynomials with five (six) clusters (Table 2). These selections provide meaningful partitions among track and CPS forecasts (section 4).
ERA-Interim data (Fig. 7i). However, in many cases the most populous cluster fails to capture this evolution, including in the final three forecasts (Figs. 7f–h). Therefore, while 60-h CPS clustering distills the wide-ranging ensemble outcomes into a small number of representative solutions, the most populous cluster is often less accurate than other clusters.

**b. Relationships between track and CPS cluster membership**

The Rand index (RI; Rand 1971) and adjusted Rand index (ARI; Hubert and Arabie 1985) measure the similarity between two clustering solutions (e.g., 168-h track clusters and 168-h CPS clusters; see the appendix).

**TABLE 3. Cluster membership for 168-h track forecasts. Clusters are arranged from 1 to 5 from west to east. The most populous cluster at each initialization time is set boldface.**

<table>
<thead>
<tr>
<th>Initialization time</th>
<th>1 (Red)</th>
<th>2 (Magenta)</th>
<th>3 (Blue)</th>
<th>4 (Cyan)</th>
<th>5 (Green)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1200 UTC 23 Oct</td>
<td>23</td>
<td>27</td>
<td>42</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>0000 UTC 24 Oct</td>
<td>36</td>
<td>34</td>
<td>26</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>1200 UTC 24 Oct</td>
<td>29</td>
<td>40</td>
<td>29</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>0000 UTC 25 Oct</td>
<td>12</td>
<td>28</td>
<td>50</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>1200 UTC 25 Oct</td>
<td>42</td>
<td>20</td>
<td>37</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>0000 UTC 26 Oct</td>
<td>22</td>
<td>49</td>
<td>21</td>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td>1200 UTC 26 Oct</td>
<td>21</td>
<td>42</td>
<td>45</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>0000 UTC 27 Oct</td>
<td>10</td>
<td>22</td>
<td>27</td>
<td>52</td>
<td>6</td>
</tr>
</tbody>
</table>
The RI varies from 0 to 1, while the ARI, which varies from \(-1\) to 1, corrects for similarities between clustering solutions produced by chance. An ARI value above 0 between two clustering solutions indicates a correspondence in cluster assignments above that expected from chance alone.

The RI between 168-h track and CPS forecasts varies from 0.650 to 0.738, with a mean of 0.706. The ARI varies between 0.127 and 0.208, with a mean of 0.170 (Table 7). The positive ARI values are statistically significant (99% confidence) for all eight initialization times. Both RI and ARI of 168-h forecasts decrease somewhat at later initialization times. The first five forecasts (1200 UTC 23 October–1200 UTC 25 October) have a mean RI (ARI) of 0.724 (0.188), while the final three forecasts (0000 UTC 26 October–0000 UTC 27 October) have a mean RI (ARI) of 0.676 (0.146). Although these are very small sample sizes, the decrease

<table>
<thead>
<tr>
<th>Initialization time</th>
<th>1 (Red)</th>
<th>2 (Magenta)</th>
<th>3 (Blue)</th>
<th>4 (Cyan)</th>
<th>5 (Green)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1200 UTC 23 Oct</td>
<td>24</td>
<td>26</td>
<td>41</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>0000 UTC 24 Oct</td>
<td>29</td>
<td>32</td>
<td>37</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>1200 UTC 24 Oct</td>
<td>20</td>
<td>41</td>
<td>25</td>
<td>25</td>
<td>6</td>
</tr>
<tr>
<td>0000 UTC 25 Oct</td>
<td>16</td>
<td>43</td>
<td>29</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>1200 UTC 25 Oct</td>
<td>13</td>
<td>29</td>
<td>40</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>0000 UTC 26 Oct</td>
<td>20</td>
<td>39</td>
<td>31</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>1200 UTC 26 Oct</td>
<td>9</td>
<td>26</td>
<td>42</td>
<td>29</td>
<td>11</td>
</tr>
<tr>
<td>0000 UTC 27 Oct</td>
<td>18</td>
<td>16</td>
<td>27</td>
<td>41</td>
<td>15</td>
</tr>
</tbody>
</table>

TABLE 4. As in Table 3, but for 60-h track forecast segments.

Fig. 5. As in Fig. 4, but for 60-h forecast segments. Cluster populations are recorded in Table 4.
in RI and ARI at later initialization times is consistent with the smaller spread among both track and CPS clusters (Figs. 4, 6). As track and CPS intercluster spreads decrease, associations of particular track clusters with particular CPS clusters weaken. Nevertheless, even for the final three initialization times, ARI values remain significantly positive, indicating that certain forecast tracks are more associated with specific structural evolutions.

The RI and ARI values between 60-h track and CPS forecast segments are somewhat higher than those for 168-h forecasts. The RI for 60-h forecast segments varies between 0.710 and 0.776, with a mean of 0.743, while the ARI varies between 0.130 and 0.308, with a mean of

<table>
<thead>
<tr>
<th>Initialization time</th>
<th>1 (Red)</th>
<th>2 (Magenta)</th>
<th>3 (Blue)</th>
<th>4 (Cyan)</th>
<th>5 (Green)</th>
<th>6 (Orange)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1200 UTC 23 Oct</td>
<td>16</td>
<td>15</td>
<td>17</td>
<td>19</td>
<td><strong>27</strong></td>
<td>23</td>
</tr>
<tr>
<td>0000 UTC 24 Oct</td>
<td>20</td>
<td>19</td>
<td><strong>27</strong></td>
<td>25</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>1200 UTC 24 Oct</td>
<td><strong>30</strong></td>
<td>18</td>
<td>26</td>
<td>11</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>0000 UTC 25 Oct</td>
<td>23</td>
<td>29</td>
<td>20</td>
<td>16</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>1200 UTC 25 Oct</td>
<td>21</td>
<td>10</td>
<td><strong>34</strong></td>
<td>14</td>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>0000 UTC 26 Oct</td>
<td><strong>36</strong></td>
<td>18</td>
<td>8</td>
<td>28</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>1200 UTC 26 Oct</td>
<td><strong>43</strong></td>
<td>26</td>
<td>27</td>
<td>9</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>0000 UTC 27 Oct</td>
<td><strong>26</strong></td>
<td>19</td>
<td>25</td>
<td>25</td>
<td>21</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 5. Cluster membership for 168-h CPS forecasts. Clusters are arranged from 1 to 6 from west to east. The most populous cluster at each initialization time is set boldface.

![Fig. 6](image-url) (a)–(h) Mean CPS path per CPS cluster for 168-h forecasts from each initialization time. Small circles indicate CPS values at 0000 UTC 30 Oct. The most populous cluster from each initialization time is indicated by the large colored circle in the top right of each panel. (i) CPS values from the ECMWF ERA reanalysis are shown for comparison.
0.227 (Table 7). Similar to 168-h forecasts, correspondence between track and CPS cluster membership decreases somewhat at later times.

The higher RI and ARI values produced using 60-h forecast segments suggest that there is greater relationship between track and CPS cluster membership for the 60-h forecast interval than for 168-h forecasts. This is likely because much of the track and structural divergence among ensemble members occurred during this interval. Therefore, clustering that focuses on this interval produces stronger relationships between track clusters and CPS path clusters than clustering that includes times with less intercluster spread in storm location and structure.

c. Relationships between EPS and track–CPS cluster membership

Rand index and ARI values calculated between EPS and track and CPS cluster assignments show some

![Fig. 7](image)

**Fig. 7.** As in Fig. 6, but for 60-h forecast segments and with (i) CPS values from the ECMWF ERA reanalysis between 0000 UTC 28 Oct and 1200 UTC 30 Oct.
relationship between EPS and cluster assignment for each forecast type. (Table 8). All ARI values are above zero, indicating that certain EPSs preferentially contribute members to specific track and CPS clusters. However, these RI and ARI values are lower than those between track and CPS cluster assignments (Table 7), indicating that there is weaker correspondence between EPS and track or CPS cluster membership than between track and CPS cluster membership.

Despite the generally modest correspondence between EPS and cluster membership, certain EPSs contribute noticeably more to specific clusters throughout the eight initialization times. For example, for 168-h track forecasts from all initialization times, the CMC GEPS contributes 58% (28 of 48) of the green cluster membership, despite comprising only 18% of all forecasts. This indicates that the GEPS had a tendency to erroneously forecast Sandy to move out to sea. In contrast, the ECMWF IFS contributes 54% (141 out of 262) of the magenta 168-h track cluster membership, an unsurprising result given the high skill of IFS Sandy track forecasts and the generally close agreement between the magenta cluster and the best track (Fig. 4).

It should be noted that the relatively low RI and ARI values between EPS and CPS cluster membership occur because CPS values from each ensemble forecast are normalized using data from only that EPS. When CPS values are normalized using data from all EPSs, there is a much stronger correspondence between EPS and cluster membership.

TABLE 7. Rand index (RI) and adjusted Rand index (ARI) values for track and CPS clustering of 168-h forecasts and 60-h forecast segments.

<table>
<thead>
<tr>
<th>Initialization time</th>
<th>168-h RI</th>
<th>168-h ARI</th>
<th>60-h RI</th>
<th>60-h ARI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1200 UTC 23 Oct</td>
<td>0.731</td>
<td>0.181</td>
<td>0.771</td>
<td>0.308</td>
</tr>
<tr>
<td>0000 UTC 24 Oct</td>
<td>0.738</td>
<td>0.206</td>
<td>0.722</td>
<td>0.192</td>
</tr>
<tr>
<td>1200 UTC 24 Oct</td>
<td>0.727</td>
<td>0.192</td>
<td>0.710</td>
<td>0.130</td>
</tr>
<tr>
<td>0000 UTC 25 Oct</td>
<td>0.705</td>
<td>0.157</td>
<td>0.748</td>
<td>0.257</td>
</tr>
<tr>
<td>1200 UTC 25 Oct</td>
<td>0.721</td>
<td>0.208</td>
<td>0.776</td>
<td>0.304</td>
</tr>
<tr>
<td>0000 UTC 26 Oct</td>
<td>0.691</td>
<td>0.153</td>
<td>0.763</td>
<td>0.290</td>
</tr>
<tr>
<td>1200 UTC 26 Oct</td>
<td>0.650</td>
<td>0.127</td>
<td>0.712</td>
<td>0.159</td>
</tr>
<tr>
<td>0000 UTC 27 Oct</td>
<td>0.688</td>
<td>0.159</td>
<td>0.738</td>
<td>0.174</td>
</tr>
<tr>
<td>Mean</td>
<td>0.706</td>
<td>0.173</td>
<td>0.743</td>
<td>0.227</td>
</tr>
</tbody>
</table>
relationship between EPS and CPS cluster assignments (not shown).

d. Individual forecast case: 0000 UTC 25 October

1) 168-H FORECASTS

For 168-h forecasts initialized at 0000 UTC 25 October, track clustering partitions ensemble tracks into five representative and meaningfully different solutions (Fig. 8a). In the red cluster mean track (12 members; Table 3) Sandy moves farthest to the west and rapidly up the east coast; it is already well inland by 0000 UTC 30 October. In the mean tracks of both the magenta cluster (28 members) and the blue cluster (most populous; 50 members), Sandy moves farther to the east before curving back to the northwest. The magenta and blue clusters follow similar tracks; the magenta mean track makes landfall in New Jersey, while the blue mean track landfalls on Long Island. However, the blue mean track is somewhat faster and thus closer to landfall at 0000 UTC 30 October than the magenta mean track. In the cyan mean track (16 members), Sandy moves much farther eastward and curves northwestward later, tracking toward Maine. In the green mean track (11 members), Sandy moves out to sea.

The divisions among track clusters indicate that 168-h track clustering produces meaningfully different representative forecast tracks (cluster mean paths) and probabilities of occurrence (cluster populations). For the 0000 UTC 25 October initialization time, the two most populous clusters (blue and magenta) follow similar trajectories, indicating that a track toward the mid-Atlantic is most probable. The other three clusters have smaller populations, and thus lower probabilities. Notably, the similarity between the blue and magenta clusters suggests that a four-cluster solution may be sufficient for this initialization time. However, the separation in landfall timing between the magenta and blue clusters shows that including the fifth cluster provides meaningful additional discrimination.

Clustering of 168-h CPS path forecasts initialized at 0000 UTC 25 October produces a meaningful partition among structural evolutions, though the variation is less straightforward than among track clusters (Fig. 8d). In the red mean CPS path (23 members) Sandy reaches its greatest \(-V_T^L\) value before it becomes asymmetric (\(B > 10\)); it then becomes asymmetric while \(-V_T^L\) decreases. Next, \(B\) increases substantially while \(-V_T^L\) increases slightly. Finally, both \(B\) and \(-V_T^L\) decrease as Sandy becomes more fully cold core and then weakens after landfall. The magenta mean CPS path (most populous; 29 members) is qualitatively similar to the red mean path. However, in the magenta mean path the final decrease in \(B\) and \(-V_T^L\) occurs later than the red mean path; at 0000 UTC 30 October the magenta cluster retains a much stronger warm core than the red cluster. The blue mean CPS path (20 members) is distinguished by a sharp \(-V_T^L\) increase on 29 October; it reaches a peak in warm core strength at 1800 UTC 29 October. Although the red, magenta, and blue clusters differ substantially in details, they follow similar qualitative evolutions: (i) increase in \(-V_T^L\), (ii) increase in \(B\) and decrease in \(-V_T^L\), (iii) increase in \(B\) and \(-V_T^L\), and (iv) decrease in \(B\) and \(-V_T^L\). The large combined population of these three clusters (72 members) suggests that this evolution had a high probability of occurrence; however, there are significant differences among these three clusters, especially in ET timing.

The cyan (16 members), green (16 members), and orange clusters (13 members) each undergo structural evolutions substantially different from the general evolution described above. In the cyan mean CPS path Sandy never becomes as asymmetric as in the red, magenta, or blue mean paths. The green mean CPS path is similar to the cyan path through 29 October; however, after 0000 UTC 30 October, \(B\) increases substantially, a
result not found in any other cluster. Finally, in the orange mean CPS path Sandy remains strongly warm core through the end of the forecast period (0000 UTC 1 November). Thus, mixture-model clustering of structural evolution by CPS path provides meaningful partitions in ET timing and storm structure during ET.

2) 60-H FORECAST SEGMENTS

Partitions of 60-h track and CPS forecast segments initialized at 0000 UTC 25 October and verifying between 0000 UTC 28 October and 1200 UTC 30 October provide complementary perspectives on the potential tracks and structural evolutions of Sandy. The 60-h track clustering produces a comparable partition as 168-h clustering; however, meaningful differences exist between the two partitions (Figs. 8a and 9a). In 60-h clustering, the magenta cluster (43 members), with a mean track that landfalls in New Jersey, is the most populous. The 60-h clustering also yields a greater difference between the magenta and blue cluster (29 members) mean tracks than yielded by 168-h clustering.

The 60-h CPS path clustering provides a complementary perspective to 168-h CPS path clustering. The mean magenta (most populous; 32 members) and blue (16 members) 60-h CPS paths (Fig. 9d) are quite similar to their corresponding 168-h paths (Fig. 8d) for the interval analyzed. However, 60-h CPS clustering produces a cyan cluster (14 members) that becomes quite asymmetric and has a mean $-V_L^T$ value of less than 50 at 0000 UTC 30 October, a result not produced by 168-h clustering. The 60-h CPS clustering also yields a second cluster (green; 21 members) that remains deeply warm core and relatively symmetric through 1200 UTC 30 October. The differences in representative tracks and CPS paths produced using 168-h forecasts and those produced using 60-h forecast segments indicate clustering using each interval provides unique and useful information about potential tracks and structural evolutions of Sandy and their probabilities of occurrence.

3) TRACK AND CPS MEMBERSHIP INTERCOMPARISON

For forecasts initialized at 0000 UTC 25 October, substantial correspondence exists between track and CPS cluster membership for both 168-h forecasts and 60-h forecast segments (ARI is 0.157 for 168-h forecasts and 0.257 for 60-h forecast segments). Many track (CPS) clusters are preferentially associated with one or two
CPS (track) clusters for both 168-h forecasts (Table 9) and 60-h forecast segments (Table 10).

Membership intercomparison can be described by the number of members of each track cluster that belong to each CPS cluster and the number of members of each CPS cluster that belong to each track cluster. For 168-h forecasts, 9 of 12 members of the red track cluster (Fig. 7a) belong to the red CPS cluster (Fig. 7d). A total of 22 of 27 members of the magenta track cluster belong to the magenta or blue CPS cluster, while 10 of 16 members of the cyan track cluster belong to the orange CPS cluster and 7 of 11 members of the green track cluster belong to the green CPS cluster (Table 9).

In terms of track membership per CPS cluster, 19 of 23 members of the red CPS cluster belong to the red or blue track cluster, while 28 of 29 members of the magenta CPS cluster belong to the magenta or blue track cluster. A total of 17 of 20 members of the blue CPS cluster also belong to the magenta or blue track cluster, 12 of 16 members of the cyan CPS cluster belong to the blue track cluster, 12 of 15 members of the green CPS cluster belong to the blue or green track cluster, and 10 of 13 members of the orange CPS cluster belong to the cyan track cluster (Table 9).

These results demonstrate that specific tracks (structural evolutions) are preferentially associated with specific structural evolutions (tracks). This is also shown by the spread in mean track per CPS cluster (Fig. 8b) and mean CPS path per track cluster (Fig. 8c). For example, the cyan track cluster (Fig. 8a) is associated with a less complete ET than the red, blue, or magenta track clusters (Fig. 8c). Similarly, the red CPS cluster, which has almost completed ET by 0000 UTC 30 October, is associated with the farthest west track and earliest landfall (Fig. 8b).

Substantial relationships are also observed between 60-h track and CPS clusters. The red track and red CPS clusters are strongly associated; 10 of 16 members of the red track cluster belong to the red CPS cluster and 10 of 15 members of the red CPS cluster belong to the red track cluster. There is also substantial association between the magenta track cluster and magenta CPS cluster; 25 of 43 members of the magenta track cluster belong to the magenta CPS cluster and 25 of 32 members of the magenta CPS cluster belong to the magenta track cluster. Other track (CPS) clusters are also primarily associated with two CPS (track) clusters. A total of 21 of 22 members of the cyan track cluster belong to the green or orange CPS cluster, and 4 of 6 members of the green track cluster belong to the cyan CPS cluster.

These results indicate that meaningful relationships exist between the track of Sandy between 0000 UTC 28 October and 1200 UTC 30 October and its structural evolution during this interval. For example, the 60-h red mean cluster track, which moves rapidly up the east coast (Fig. 9a), is associated with high B values and rapid ET, while the cyan cluster, in which Sandy moves much farther to the east, is associated with more symmetry and a stronger warm-core throughout the 60-h interval (Fig. 9c). Similarly, the orange and green CPS paths, in which Sandy remains relatively symmetric and warm core (Fig. 9d) correspond to more eastward tracks (Fig. 9b). These results are consistent with the synoptic environment around Sandy; more eastward motion is associated with delayed (or no) interaction with the trough over the eastern United States, and thus a delay in ET completion.

<table>
<thead>
<tr>
<th>CPS 1 (Red)</th>
<th>CPS 2 (Magenta)</th>
<th>CPS 3 (Blue)</th>
<th>CPS 4 (Cyan)</th>
<th>CPS 5 (Green)</th>
<th>CPS 6 (Orange)</th>
<th>Track 1 (Red)</th>
<th>Track 2 (Magenta)</th>
<th>Track 3 (Blue)</th>
<th>Track 4 (Cyan)</th>
<th>Track 5 (Green)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>25</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>18</td>
<td>5</td>
<td>12</td>
<td>5</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
5. Composites of storm structure at 0000 UTC 30 October

Variations among structural evolutions of Sandy and the relationship between track and structural evolution are illustrated using storm-relative composites of 500-hPa geopotential height (Fig. 10) and 850-hPa equivalent potential temperature ($\theta_e$; Fig. 11) at 0000 UTC 30 October from 168-h forecasts initialized at 0000 UTC 25 October. The structure of Sandy and its stage of

FIG. 10. Storm-centered composites of 500-hPa geopotential height (m, color shaded and contoured) for (a)–(e) track and (f)–(k) CPS clusters from forecasts initialized at 0000 UTC 25 Oct and verifying at 0000 UTC 30 Oct. The 5560-m geopotential isoline is contoured in red and the 992-hPa sea level pressure isobar is contoured in white. Tick marks indicate the number of degrees of latitude and longitude from the center of Sandy.
interaction with the midlatitude trough vary substantially among the track clusters (Figs. 10a–e) and among the CPS clusters (Figs. 10f–k). Clusters in which Sandy moves more westward and makes landfall earlier show more advanced stages of storm–trough interaction. This accords with the results of Munsell and Zhang (2014), who showed that WRF ensemble members in which Sandy tracked more westward interacted with the midlatitude trough earlier than those in which Sandy moved farther eastward before making landfall. In the farthest west track composite (Fig. 10a; corresponding to the red track in Fig. 8a) Sandy has already merged with the trough, which has broadened substantially. In the magenta and blue track composites in Fig. 8a (Figs. 10b and 10c), the Sandy–trough interaction and merger are well under way at 0000 UTC 30 October, but the trough remains narrower than in Fig. 10a. In the Fig. 8a cyan track composite (Fig. 10d), the Sandy–trough interaction is at a much earlier stage and Sandy remains a somewhat distinct 500-hPa geopotential feature. In the Fig. 8a green track composite (Fig. 10e) Sandy is well east of the trough and no interaction occurs.

The composite 500-hPa geopotential height maps of CPS clusters show a similar pattern (Figs. 10f–k). Clusters in which Sandy moves more westward (in Fig. 8b) show more advanced Sandy–trough interaction (e.g., Fig. 10f), while clusters in which Sandy moves more eastward show less advanced interaction (e.g., Figs. 10j,k). Notably, the variation in 500-hPa height patterns at 0000 UTC 30 October is greater among track clusters than among CPS clusters (cf. Figs. 10a,e and 10f,k). This supports the hypothesis that differences in Sandy’s structure at 0000 UTC 30 October were largely driven by differences in the storm track, and thus interaction with the midlatitude trough to its west.

Clusters in which Sandy moves more westward and makes landfall earlier are also associated with greater lower-tropospheric warm core decay (Fig. 11). In the farthest west track and CPS clusters (Figs. 11a and 11f) low-θ_e midlatitude air has already wrapped into the core of Sandy by 0000 UTC 30 October. Clusters in which Sandy moves farther eastward show higher θ_e values at its center and a stronger connection with high-θ_e air to its south (e.g., Figs. 11d,e and 11j,k), potentially due to a combination of less advanced trough interaction, later landfall, and less time over cold water north of the Gulf Stream. Similar to what is observed for 500-hPa geopotential height, there is greater variation in 850-hPa θ_e among track clusters than CPS clusters, indicating that track greatly influenced the thermodynamic structure of Sandy near the time of observed landfall.

The magenta and blue track clusters and magenta CPS cluster (Figs. 11b,c,g) demonstrate the warm seclusion characteristics diagnosed by Galarnneau et al. (2013), suggesting that these clusters capture aspects of the observed thermodynamic evolution of Sandy near the observed landfall time. Notably, though the magenta and blue track clusters take similar paths (Fig. 8a), the faster blue cluster is associated a weaker 850-hPa θ_e anomaly at 0000 UTC 30 October (cf. Figs. 11b,c). This may be due to both more advanced trough interaction in the blue cluster, as well as earlier landfall. By 0000 UTC 30 October 32% of blue cluster members have made landfall, compared with 19% of magenta cluster members.

6. Summary and conclusions

Regression mixture-model clustering is performed on 117 ensemble forecasts of Hurricane Sandy from four global ensemble prediction systems initialized at eight times between 1200 UTC 23 October and 0000 UTC 27 October. Track and cyclone phase space (CPS) forecasts are clustered using data out to 168 h and for a constant 60-h interval between 0000 UTC 28 October and 1200 UTC 30 October.

Optimal mixture model specifications for each forecast type are selected by determining how Bayesian information criterion, mean-squared forecast error, and fraction of ensemble members with probability of assignment below 0.95 change with polynomial order and number of clusters. We select a fourth-order (third order) polynomial with five clusters for 168-h track forecasts (60-h track forecast segments) and a fourth-order (third-order) polynomial with six clusters for 168-h CPS forecasts (60-h CPS forecast segments).

Both track and CPS clustering of 168-h forecasts and 60-h forecast segments produce meaningful divisions among clusters, distilling 117 ensemble forecasts into a small number of representative and distinct outcomes. Track clusters are divided among those in which Sandy goes out to sea and those in which Sandy landfalls at varying locations and times on the Atlantic coast. CPS clusters vary in the timing and progress through ET, as well as storm structure during ET. Notably, 168-h CPS clustering captures the final Sandy–trough interaction near landfall more accurately than it captures earlier thermodynamic changes on 27 and 28 October.

Rand index (RI) and adjusted Rand index (ARI) calculations show substantial correspondence between track cluster membership and CPS cluster membership for 168- and 60-h forecast clustering. Often, most members in a single track (CPS) cluster are found in one or two CPS (track) clusters. We illustrate this using forecasts initialized at 0000 UTC 25 October; however, it holds across all initialization times. Both RI and ARI
are generally lower at later initialization times, an expected result as the intercluster spread decreases. Throughout the eight initialization times, generally modest but nonnegligible correspondence is observed between EPS membership and cluster membership from each forecast type.

Cluster-mean 500-hPa geopotential height and 850-hPa \( \theta_e \) composites taken at 0000 UTC 30 October from forecasts initialized at 0000 UTC 25 October show substantially different stages of Sandy–trough interaction and differences in lower-troposphere thermodynamic storm structure among clusters. Clusters in
which Sandy moves more westward and makes landfall earlier and have more advanced trough interaction and greater low-\(\theta_c\) penetration into the inner core at 0000 UTC 30 October. Some cluster composites show lower-troposphere warm seclusion characteristics, indicating that these clusters predicted the structural evolution near landfall diagnosed by Galarneau et al. (2013). Clusters in which Sandy moves out to sea are associated with little trough interaction and a more intact warm core.

These results demonstrate that regression mixture-model clustering can distill multimodel ensemble track and CPS path forecasts into a small number of meaningfully distinct forecast outcomes. This result holds whether we consider 168-h forecasts or 60-h forecast segments near the observed landfall time. The substantial relationship between track and CPS cluster assignments provide evidence that the structural evolution of Sandy depended substantially on storm track.

Sandy was a well-forecast hurricane (Blake et al. 2013); however, the methodology presented here may be useful as an operational tool for forecasting other TCs. Regression mixture modeling of multi-EPS forecasts can provide forecasters with representative tracks and structural evolutions and their probabilities of occurrence. This may be especially useful for storms with substantial track forecast divergence [e.g., Sandy or 2015 Hurricane Joaquin; Berg (2016)]. Furthermore, relationships between track and CPS cluster membership can aid forecasters and emergency planners in determining potential landfall impacts. We are currently investigating the utility of multi-EPS regression mixture-model clustering as an operational forecast tool.

### Acknowledgments

We are grateful to Dr. Francesca Chiaromonte, Dr. Prabhani Kuruppuumullage Don, Dr. Robert Hart, Mr. Casey Webster, and Dr. Timothy Marchok for their assistance on this project. The ensemble forecasts were obtained via the ECMWF TIGGE database. We thank Dr. Ryan Torn and two anonymous reviewers for their insightful comments that helped improve this manuscript. This research was supported by the National Science Foundation under Grant ATM-1322532.

### APPENDIX

#### Rand Index and Adjusted Rand Index Calculations

Given a set of elements (here ensemble members) and two partitions (cluster solutions) of the set: \(X\) (say track) with \(r\) clusters and \(Y\) (say CPS path) with \(s\) clusters, so that \(X = \{X_1, X_2, \ldots, X_r\}\) and \(Y = \{Y_1, Y_2, \ldots, Y_s\}\), the following contingency table can thus be created (Table A1).

<table>
<thead>
<tr>
<th>Cluster</th>
<th>(Y_1)</th>
<th>(Y_2)</th>
<th>(\ldots)</th>
<th>(Y_s)</th>
<th>Sums</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1)</td>
<td>(n_{11})</td>
<td>(n_{12})</td>
<td>(\ldots)</td>
<td>(n_{1s})</td>
<td>(a_1)</td>
</tr>
<tr>
<td>(X_2)</td>
<td>(n_{21})</td>
<td>(n_{22})</td>
<td>(\ldots)</td>
<td>(n_{2s})</td>
<td>(a_2)</td>
</tr>
<tr>
<td>(\vdots)</td>
<td>(\vdots)</td>
<td>(\vdots)</td>
<td>(\ddots)</td>
<td>(\vdots)</td>
<td>(\vdots)</td>
</tr>
<tr>
<td>(X_r)</td>
<td>(n_{r1})</td>
<td>(n_{r2})</td>
<td>(\ldots)</td>
<td>(n_{rs})</td>
<td>(a_r)</td>
</tr>
<tr>
<td>Sums</td>
<td>(b_1)</td>
<td>(b_2)</td>
<td>(\ldots)</td>
<td>(b_s)</td>
<td>(\text{Sums})</td>
</tr>
</tbody>
</table>

For the calculation of the Rand index (RI; Rand 1971) and adjusted Rand index (ARI; Hubert and Arabie 1985), the following terms are defined:

\[
A = \text{Number of pairs of ensemble members that are in the same track cluster and the same CPS cluster}:
\]

\[
A = \sum_{ij} \left( \frac{n_{ij}}{2} \right).
\]

\[
B = \text{Number of pairs that are in the same track cluster, but different CPS clusters}:
\]

\[
B = \sum_{i} \left( \frac{n_{i}}{2} \right) - \sum_{ij} \left( \frac{n_{ij}}{2} \right).
\]

\[
C = \text{Number of pairs that are in the same CPS cluster, but different track clusters}:
\]

\[
C = \sum_{j} \left( \frac{n_{j}}{2} \right) - \sum_{ij} \left( \frac{n_{ij}}{2} \right).
\]

\[
D = \text{Number of pairs that are in different track clusters and different CPS clusters}:
\]

\[
D = \binom{n}{2} - A - B - C,
\]

\[
\text{RI} = \frac{A + D}{A + B + C + D} = \frac{A + D}{\binom{n}{2}}.
\]

The RI is the fraction of pairs that are in the same track and same CPS cluster plus the fraction of pairs that are in different track clusters and different CPS clusters.

For the ARI, the following terms are defined:

\[
\text{Index} = \sum_{ij} \left( \frac{n_{ij}}{2} \right) \quad (\text{Term } A \text{ from the RI calculation})
\]

\[
\text{ExpInd} = E \left[ \sum_{ij} \left( \frac{n_{ij}}{2} \right) \right] = \left[ \sum_{i} \left( \frac{n_{i}}{2} \right) \right] \left[ \sum_{j} \left( \frac{n_{j}}{2} \right) \right] / \binom{n}{2}.
\]

ExpInd is the expected agreement between two
TABLE A2. Contingency table of track and CPS cluster assignments for 60-h forecast segments initialized at 0000 UTC 25 Oct.  

<table>
<thead>
<tr>
<th></th>
<th>CPS 1</th>
<th>CPS 2</th>
<th>CPS 3</th>
<th>CPS 4</th>
<th>CPS 5</th>
<th>CPS 6</th>
<th>Sums</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track 1</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Track 2</td>
<td>4</td>
<td>25</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>43</td>
</tr>
<tr>
<td>Track 3</td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>0</td>
<td>8</td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>Track 4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>10</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Track 5</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Sums</td>
<td>15</td>
<td>32</td>
<td>16</td>
<td>14</td>
<td>21</td>
<td>18</td>
<td>116</td>
</tr>
</tbody>
</table>

MaxInd = \frac{1}{2} \left[ \sum \binom{n_i}{2} + \sum \binom{n_j}{2} \right]. \text{MaxInd is the maximum possible agreement between two partitions:}

\text{ARI} = \frac{\text{Index} - \text{ExpInd}}{\text{MaxInd} - \text{ExpInd}}

a. Example using 60-h forecast segment data initialized at 0000 UTC 25 October

Given the track and CPS cluster solutions from the 60-h forecast segment data initialized at 0000 UTC 25 October, the contingency table below is generated (Table A2).

b. Rand index calculation

\[ A = \sum_{i,j} \binom{n_{ij}}{2} = \binom{10}{2} + \binom{4}{2} + \binom{25}{2} + \binom{5}{2} + \binom{10}{2} \]
\[ + \binom{3}{2} + \binom{6}{2} + \binom{2}{2} + \binom{11}{2} + \binom{6}{2} + \binom{10}{2} = 585, \]

\[ B = \sum_{i} \binom{n_i}{2} - \sum_{i,j} \binom{n_{ij}}{2} = \binom{16}{2} + \binom{43}{2} + \binom{29}{2} + \binom{22}{2} + \binom{6}{2} - 585 = 1675 - 585 = 1090, \]

\[ C = \sum_{j} \binom{n_j}{2} - \sum_{i,j} \binom{n_{ij}}{2} = \binom{15}{2} + \binom{32}{2} + \binom{16}{2} + \binom{14}{2} + \binom{21}{2} + \binom{18}{2} - 585 = 1175 - 585 = 590, \]

\[ D = \binom{n}{2} - A - B - C = 6670 - 585 - 1090 - 590 = 4405, \]

\[ \text{RI} = \frac{A + D}{\binom{n}{2}} = \frac{4405 + 585}{6670} = 0.748. \]

c. Adjusted Rand index calculation

\[ \text{Index} = \sum_{i,j} \binom{n_{ij}}{2} = 585, \]

\[ \text{ExpInd} = \left[ \sum_{i} \binom{n_i}{2} \right] \frac{\binom{n}{2}}{6670} = 1675 \times 1175 = 295.07, \]

\[ \text{MaxInd} = \frac{1}{2} \left[ \sum_{i} \binom{n_i}{2} + \sum_{j} \binom{n_j}{2} \right] = \frac{1}{2} (1675 + 1175) = 1425, \]

\[ \text{ARI} = \frac{\text{Index} - \text{ExpInd}}{\text{MaxInd} - \text{ExpInd}} = \frac{585 - 295.07}{1425 - 295.07} = 0.257. \]

REFERENCES


