Application of Unsupervised Learning Techniques to Identify Atlantic Tropical Cyclone Rapid Intensification Environments

ANDREW E. MERCER, ALEXANDRIA D. GRIMES, AND KIMBERLY M. WOOD

(Manuscript received 30 April 2020, in final form 20 November 2020)

ABSTRACT: Tropical cyclone (TC) track forecasts have improved in recent decades while intensity forecasts, particularly predictions of rapid intensification (RI), continue to show low skill. Many statistical methods have shown promise in predicting RI using environmental fields, although these methods rely heavily upon supervised learning techniques such as classification. Advances in unsupervised learning techniques, particularly those that integrate nonlinearity into the class separation problem, can improve discrimination ability for difficult tasks such as RI prediction. This study quantifies separability between RI and non-RI environments for 2004–16 Atlantic Ocean TCs using an unsupervised learning method that blends principal component analysis with $k$-means cluster analysis. Input fields consisted of TC-centered 1° Global Forecast System analysis (GFSA) grids (170 different variables and isobaric levels) for 3605 TC samples and five domain sizes. Results are directly compared with separability offered by operational RI forecast predictors for eight RI definitions. The unsupervised learning procedure produced improved separability over operational predictors for all eight RI definitions, five of which showed statistically significant improvement. Composites from these best-separating GFSA fields highlighted the importance of mid- and upper-level relative humidity in identifying the onset of short-term RI, whereas long-term, higher-magnitude RI was generally associated with weaker absolute vorticity. Other useful predictors included optimal thermodynamic RI ingredients along the mean trajectory of the TC. The results suggest that the orientation of a more favorable thermodynamic environment relative to the TC and midlevel vorticity magnitudes could be useful predictors for RI.

KEYWORDS: Tropical cyclones; Principal components analysis; Statistical techniques; Machine learning

1. Introduction

Tropical cyclones (TCs) can cause numerous societal hazards, including storm surge, high winds, and flooding from excessive rainfall. Although these hazards are primarily associated with coastal locations, the strongest TCs often produce significant impacts far inland from the coast, particularly with regard to flooding. Prior knowledge of TC intensity is essential for resource deployment ahead of a landfalling system, yet TC intensity forecasts remain challenging due to the numerous processes that contribute to TC intensification. Many of these processes are inherently thermodynamic (such as latent heat flux from high sea surface temperatures; Kaplan and DeMaria 2003) and not well represented by the dynamic weather models used in operational forecasts (Kotroni and Lagouvardos 2004; Klemp 2006; Mercer et al. 2013). Kinematic factors relevant to TC intensity change are frequently noisy (e.g., 200-hPa divergence; Leroux 2016) or require sufficient vertical resolution [e.g., vertical wind shear (Wang et al. 2015; Shu et al. 2013)] to represent processes important for intensification (Leroux 2016). Recent developments in numerical weather prediction, most notably the Hurricane Weather and Research Forecasting model (HWRF; Tallapragada et al. 2014), have improved intensity forecasts within dynamic model systems, yet statistical models [e.g., the Statistical Hurricane Intensity Prediction Scheme (SHIPS); DeMaria et al. 2005] remain popular tools for operational TC intensity forecasting.

Further complicating this issue is the difficulty in predicting rapid intensification (RI), a large increase in a TC’s maximum, 1-min sustained 10-m wind speed (hereinafter referred to as the maximum wind speed) over a short time period (typically 24 h). The National Hurricane Center (NHC) operationally defines RI as a 30-kt (15.4 m s$^{-1}$; 1 kt $\sim 0.51$ m s$^{-1}$) increase in peak wind speed in 24 h, an inherently rare event as roughly 6% of 24-h TC intensity change episodes meet this threshold (Kaplan et al. 2010). Additionally, the exact physical processes governing RI remain poorly understood (Wang and Wu 2004; Grimes and Mercer 2014), an issue compounded by the relative lack of boundary layer observations within the TC environment and heavy reliance on global operational dynamic forecast models to fill these observational gaps.

Recent work has improved our understanding of processes governing the RI of Atlantic Ocean TCs. The probability of RI increases when a deep layer of warm ocean water is available to enhance surface heat fluxes, and increased instability during nighttime can be more conducive for RI (Holliday and Thompson 1979). On the synoptic scale, TC–trough interactions can support RI through enhanced upper-level divergence (e.g., Molinari et al. 1995), although the benefits of such interactions vary and may depend on the TC–trough configuration (e.g., Hanley et al. 2001; Fischer et al. 2017). It may be the nature and proximity of these upper-level disturbances that impact RI: shorter-wavelength troughs tend to be more favorable, and cutoff lows 500–1000 km southwest of the TC are generally more supportive of RI (Fischer et al. 2019). Bosart
et al. (2000) focused heavily on synoptic-scale kinematic conditions when studying RI in Hurricane Opal (1995), noting that enhanced upper-level divergence coupled with low vertical wind shear were present (in addition to enhanced surface sensible and latent heat fluxes).

In terms of smaller-scale processes, Willoughby et al. (1982) focused on the mesoscale, demonstrating the importance of eyewall replacement cycles in enhancing intensification rates, particularly as the diameter of the outer eyewall contracted. Willoughby (1998) expanded that work by identifying a mid-level inversion in the eye whose elevation controls the rate of pressure falls within TCs (which is strongly related to RI). More recent work by Plotkin et al. (2019) supported work from Chen and Zhang (2013) and Tao and Jiang (2015) suggesting that shallow to moderate depth convection could help to trigger RI, while deep convection was a response to RI.

These recent studies still do not fully encompass all RI situations as evidenced by the ongoing challenge of improving RI forecasts. These limitations force investigators to identify relevant TC environmental features associated with RI using an event-centric approach [e.g., Hurricane Opal 1995 (Bosart et al. 2000); Hurricane Felix 2007 (Liu and Xie 2012); Hurricane Michael 2012 (Alaka et al. 2017)]. Such case studies have revealed important details for specific TCs that underwent RI, but their results do not necessarily translate to all RI events. Recent work by Mercer and Grimes (2015) assessed RI environments using a climatological approach with a large database (5409) of TC track points, but their study only considered the first RI occurrence within a TC’s life cycle and relied on reanalysis fields instead of an operational analysis product, limiting applicability of their results to real-time forecasting.

Rather than focus on processes exclusively, previous studies (e.g., Kaplan et al. 2015) showed some success employing statistical hypothesis testing to identify relevant diagnostic predictors (derived from previous physical understanding) that can distinguish between RI and non-RI environments. Predictors that showed the greatest discrimination capability are currently used operationally by the NHC in a statistical RI prediction scheme known as the SHIPS Rapid Intensification Index (SHIPS-RII; Kaplan et al. 2015). Most SHIPS-RII predictors rely upon statistical moments (usually the mean and standard deviation) of TC-centric atmospheric quantities measured spatially over hundreds of kilometers to obtain single values relevant for a TC environment. This is an important limitation of the SHIPS-RII system as these global statistics lose any relevant spatial pattern information, which can contribute to the poor RI/non-RI separability currently seen operationally (e.g., Fig. 1, although this is observed with many other predictors as well). Grimes and Mercer (2016) incorporated this spatial variability by applying rotated principal component analysis (RPCA; Richman 1986) to TC-relevant reanalysis fields from the Global Ensemble Forecast System reforecast database (Hamill et al. 2013). Low-level (850 hPa) equivalent potential temperature ($\theta_e$; Tsonis 2007) profiles coupled with midlevel static stability (consistent with Willoughby 1998) were useful in distinguishing RI and non-RI events, although Grimes and Mercer (2016) never formally quantified RI/non-RI classification ability of these diagnostic variables. The lack of spatial information in current SHIPS-RII predictors, coupled with continued low RI forecast skill in operational statistical models (Kaplan et al. 2015), suggests that further investigation could offer additional insight into spatial patterns associated with RI.

Most previous research into RI/non-RI classification has applied techniques known as supervised learning methods, where the user oversees many aspects of the study, including the case selection, the predictor selection, the classification criteria and method, and traditional cross-validation techniques (R Russell and Norvig 2010). Techniques such as regression and classification fall under the broader heading of supervised learning methods. These methods have value as evidenced by the improvements seen in RI forecasts in recent years (Kaplan et al. 2015). However, supervised learning methods are limited by the knowledge and biases of prior understanding, which can exacerbate classification difficulty in rare event scenarios such as the RI problem.

Recently, unsupervised learning methods (such as k-means cluster analysis and principal component analysis; Wilks 2011) have gained popularity with difficult classification problems. Unsupervised learning methods enable the data to drive classification, which can reveal previously unseen relationships among covariates and improve physical understanding of existing problem-relevant processes (Hinton et al. 1999). Poor classification results caused by class imbalances, multiple resolutions of space and time, varying degrees of noise, and other uncertainties have benefitted from unsupervised techniques
(Karpatne et al. 2018). Although unsupervised learning is more challenging, these methods are powerful as a pattern mining tool, often revealing previously unseen patterns that supervised techniques are unable to depict. A newer unsupervised clustering technique, kernel principal component analysis (KPCA; Schölkopf et al. 1998), can increase cluster separability over traditional linear clustering methods such as hierarchical clustering, $k$-means cluster analysis, or PCA (Mercer et al. 2011; Mercer and Grimes 2015). However, the use of KPCA as a clustering technique has been primarily limited to the medical industry (e.g., Shiokawa et al. 2018) or remote sensing (e.g., Binol 2018). Atmospheric science studies have benefited from KPCA in forming composites; for example, Richman and Adrianto (2010) obtained more physically realistic mean sea level pressure fields using KPCA over RPCA. Mercer et al. (2011) and Mercer and Grimes (2015) also identified new covariability patterns within severe weather outbreak classes and TC environments, respectively, that better matched the underlying physical structures of each. However, those studies did not exclusively quantify class separability, which is a more common KPCA application in other disciplines and is directly applicable to the challenge of identifying RI environments diagnostically and eventually prognostically. The primary objective of this study is to identify diagnostic fields that best separate RI and non-RI TC environments, based on conditions present at the beginning of the intensity change period, using unsupervised learning methods (specifically cluster analysis coupled with RPCA and KPCA). Specifically, we address separability for eight RI categories, seven of which were included in Kaplan et al. (2015) with an eighth category that defines RI as a 65-kt intensity increase in 72 h (65 kt/72 h). We do not provide any prognostic applications of our methods at this stage; instead, we focus on identifying and describing the fields that best separate RI and non-RI cases at RI onset to provide updated insight into diagnostic variables important for RI. Specifically, we identify spatial fields in from operational model analysis output that provide improved RI/non-RI separability over currently utilized predictors in the operational SHIPS-RII. We then develop TC-centric composites for RI and non-RI environments from these fields to address the physical characteristics portrayed by each diagnostic variable. This approach yields options for new predictors to be added in an operational statistical modeling scheme for Atlantic RI. Section 2 gives an overview of the method of the unsupervised learning, and section 3 provides details on the unsupervised learning results. Section 4 details the physical structures underlying the best-separating environments between RI and non-RI TC cases. Last, section 5 summarizes the results and outlines future directions.

2. Data and methods

a. Datasets and domains

Unsupervised learning methods require large databases of cases from which the classes may be derived. Additionally, atmospheric data for each case must be available continuously (no missing points) in both space and time, requiring a global gridded analysis dataset with a long period of record. We selected Atlantic TC cases from the NHC’s “best track” hurricane database 2 (HURDAT2; Landsea and Franklin 2013), which provides postseason reanalyzed best tracks spanning 1851–2019 at 6-h intervals with occasional additional data points for instances such as landfall. We restricted our sample to times at which a TC was labeled as a tropical depression, tropical storm, or hurricane during the 2004–16 Atlantic hurricane seasons for a total of 183 TCs that include 3605 six-hour HURDAT2 entries (hereinafter referred to as cases). Only TC locations over water were retained.

Since this study explores a range of RI definitions that span from 12 to 72 h, the cases evaluated per definition must vary (e.g., short-lived TCs that did not have 72 h of track points were excluded for the 72-h RI definition). Cases were also excluded if the intensity change period overlapped a TC’s landfall. The RI/non-RI class was established by computing the change in peak wind speed between the start and end of the defined intensity change period (Kaplan et al. 2015). Cases meeting the RI criterion for the given RI definition were labeled as RI, while all other included cases were labeled as non-RI. This reduced case sample sizes for each of the RI definitions described herein (Table 1). We evaluated eight RI definitions: seven from Kaplan et al. (2015) as well as the 65-kt/72-h RI definition currently used operationally by the SHIPS-RII. Each definition roughly constituted the 90th–95th percentiles of intensification rate (see Table 3 in Kaplan et al. 2015). Note that we used HURDAT2 only to label cases either RI or non-RI; no HURDAT2 information was included in the unsupervised learning portion of this study.

The unsupervised learning procedure required a spatially continuous atmospheric dataset describing atmospheric conditions that make up each case to quantify separability in RI/non-RI environments. Since an eventual goal of this work is the integration of these results into an operational forecast product, we relied on operational numerical weather prediction analysis fields to quantify TC environments. Specifically, the 1° Global Forecast System Analysis (GFS) fields were used to represent the three-dimensional TC environment because 0.5° GFS data availability begins in 2015 but 1° fields are available starting in 2004. By selecting 1° GFS fields, our ability to render features smaller than roughly 300 km in size was limited, so our study focuses on larger-scale (>300 km) TC

<table>
<thead>
<tr>
<th>RI definition</th>
<th>Temporal interval</th>
<th>$N^*$</th>
<th>RI frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 kt</td>
<td>12 h</td>
<td>3257</td>
<td>5.74%</td>
</tr>
<tr>
<td>25 kt</td>
<td>24 h</td>
<td>2913</td>
<td>11.64%</td>
</tr>
<tr>
<td>30 kt</td>
<td>24 h</td>
<td>2913</td>
<td>7.45%</td>
</tr>
<tr>
<td>35 kt</td>
<td>24 h</td>
<td>2913</td>
<td>4.53%</td>
</tr>
<tr>
<td>40 kt</td>
<td>24 h</td>
<td>2913</td>
<td>2.88%</td>
</tr>
<tr>
<td>45 kt</td>
<td>36 h</td>
<td>2583</td>
<td>5.42%</td>
</tr>
<tr>
<td>55 kt</td>
<td>48 h</td>
<td>2279</td>
<td>5.53%</td>
</tr>
<tr>
<td>65 kt</td>
<td>72 h</td>
<td>1760</td>
<td>7.96%</td>
</tr>
</tbody>
</table>
characteristics associated with Atlantic RI. This is an important issue as our work is limited to characterizing the environmental factors related to RI, not the TC’s internal dynamics, which have also been shown to be critical in diagnosing RI (Hendricks et al. 2010; Judt and Chen 2016; Rios-Berrios and Torn 2017). The GFSA fields are available at 21 isobaric levels, of which we retained 11 (which included the mandatory levels): 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, and 100 hPa. We used the other 10 levels when calculating vertical derivatives but did not directly include them in the unsupervised learning procedure.

From GFSA, we obtained base-state meteorological characteristics as well as numerous derived fields (Table 2). Additionally, the static stability parameter $\sigma$ from Bluestein (1993) was included:

$$\sigma = -\frac{R_d T}{p} \frac{d \ln \theta}{dp},$$

(1)

where $T$ is temperature (K) at the level of interest, $R_d$ is the dry air gas constant (287 J kg$^{-1}$ K$^{-1}$), $p$ is pressure (Pa), and $\theta$ is potential temperature (K). The static stability parameter was computed at all levels except 1000 and 100 hPa because of data limitations. The vertical derivative in calculating $\sigma$ was estimated using the isobaric level above and below the GFSA level of interest via centered finite differencing. Seven surface-based fields were also included (Table 2) in the analysis. In total, 170 GFSA fields were retained for each TC case. To center these fields on each case, we located the minimum mean sea level pressure within 500 km of the HURDAT2 center to ensure grids were centered on the GFSA estimated TC center location.

![Fig. 2. An illustration of the five TC domains overlaid on the average 925-hPa specific humidity (kg kg$^{-1}$) field for all TC cases. Each inner domain is reduced in size by 2$^\circ$ on both sides.](image)

The unsupervised techniques employed herein utilize a similarity matrix (e.g., the correlation matrix in RPCA) describing spatial relationships between GFSA grid points. As such, any changes to the size of the TC-centric GFSA domains will lead to differences in the unsupervised learning results. That is, a spatial domain that focuses on the TC center (such as a 2$^\circ$ × 2$^\circ$ TC-centric domain) will have a different clustering outcome than a larger domain that spanned the typical spatial range over which SHIPS-RII predictors are computed (an 18$^\circ$ × 18$^\circ$ domain or an approximate radius of 1000 km). To address this issue, five separate spatial domain sizes were tested, starting with an 18$^\circ$ × 18$^\circ$ (~2000 km × 2000 km) domain and reducing subsequent domains by 2$^\circ$ on all sides to the smallest size of 2$^\circ$ × 2$^\circ$ (~220 km × 220 km; Fig. 2). All 170 GFSA fields were retained for each of the five domain sizes listed above, yielding a total of 850 separate fields that were individually tested in the unsupervised learning methods for each RI definition. This reduction was done to emulate the distances used for computing field means with the SHIPS-RII predictors, which are based on sizes of approximately 200 km (e.g., 0–200-km precipitable water; Kaplan et al. 2015).

### b. Unsupervised learning methods

For the unsupervised learning method (summarized in the flowchart in Fig. 3), we employed a combination of RPCA, KPCA, and cluster analysis to quantify each GFSA field’s RI/non-RI separability. The analysis was done in two steps. First, the GFSA spatial dimension was reduced to a subset of representative basis vectors (principal components; Wilks 2011) through eigen decomposition of a similarity matrix. Traditional RPCA (Richman 1986; Wilks 2011) is a well-known eigen decomposition technique and utilizes the correlation matrix as a measure of similarity among variables, rendering the resulting relationships linearly. That is, RPCA is not able to quantify nonlinear relationships between covariates (Richman and Adrianto 2010). KPCA addresses this issue by...
using the kernel similarity matrix from support vector machines (SVMs; Cristinini and Shawe-Taylor 2000) to measure nonlinear relationships among covariates. Mathematically, in KPCA the RPCA correlation matrix $R$, given by

$$ R = \frac{1}{n-1} Z^T Z $$  \tag{2}$$

is updated to the kernel matrix $K$

$$ K = \frac{1}{n-1} \varphi(Z^T) \varphi(Z), $$  \tag{3}$$

where $\varphi$ is a kernel map function that projects the matrix $Z$ of $n$ standard anomalies ($z$ scores) into a higher dimensional hyperspace where linear separability is better represented. One major advantage of kernel methods such as KPCA is that prior knowledge of the kernel map function $\varphi$ is unnecessary. Mathematically, this is a result of Mercer’s theorem (Cristinini and Shawe-Taylor 2000), which states that the application of a kernel function $K(x, y)$ that results from an unknown map function $\varphi$, applied to any finite set of vectors $x$ and $y$, will yield a positive definite, nonsingular matrix $K$. Conversely, if the matrix $K$ can be computed, there must be a map function $\varphi$ that exists and needs not be known. As a result, $K$ retains all necessary mathematical requirements to serve as a similarity matrix in a PCA, with the only limitation that relationships given by $K$ are not initially centered and are projected into a higher dimensional nonlinear hyperspace, making direct interpretation of loadings from the $K$ basis vectors more challenging. Since $K$ must be centered prior to eigen decomposition, it may be centered by

$$ K_{\text{center}} = K - 21_{l_{m}}K + 1_{l_{n}}K1_{l_{n}}, $$  \tag{4}$$

where the $1_{l_{m}}$ is a matrix for which all elements are $1/n$ and $n$ is the number of columns of $Z$. Once centered, the $K_{\text{center}}$ matrix is used to solve the eigenvalue problem given in Eq. (5). This is analogous to the eigenvalue problem for RPCA:

$$ K_{\text{center}} \alpha_i = \lambda_i \alpha_i, $$  \tag{5}$$

where the $\alpha_i$ and $\lambda_i$ are eigenvectors and eigenvalues of $K_{\text{center}}$, respectively. Resulting eigenvectors and eigenvalues are then used to compute KPC loadings by scaling the eigenvectors by the square root of their associated eigenvalue. KPC scores are computed by projecting the original anomaly matrix $Z$ onto the KPC loadings. This approach mirrors the traditional RPCA approach except for the computation of the similarity matrix $K_{\text{center}}$.

As expected, many possible kernel functions can yield a $K$ matrix that is suitable for PCA. In this study, we tested the two most utilized kernel functions (Schölkopf et al. 1998):

- **polynomial kernel**
  $$ K(x, y) = (x^T y + 1)^d $$  \tag{6}$

- **radial basis function kernel**
  $$ K(x, y) = \exp \left( -\frac{1}{2\sigma^2} \|x - y\|^2 \right). $$  \tag{7}$

The variable $d$ represents the degree of the polynomial kernel and $\sigma$ measures spread in the radial basis function (RBF) kernel. These parameters are tunable, meaning that tuning experiments are required to identify the best separability (typically done through trial and error). In our study, $d$ values between 2 and 10 were considered for polynomial kernels, as well as $\sigma$ values of 5, 10, 25, 50, 75, 100, 200, 500, and 1000 for RBF kernels (a total of 18 configurations for $K$). Afterward, both RPCA and each of the 18 KPCA configurations were candidate methods for reducing the spatial dimension of the given GSFA field. Because traditional techniques (such as the scree test; Wilks 2011) for determining the optimal number of PCs to retain are not possible with KPCA (Mercer et al. 2011), we retained between 3 and 9 PCs for each of the configurations, keeping the number that yielded the greatest separability (described below). The PCA configuration (either RPCA or one of the 18 KPCA methods with the optimal number of PCs retained) that maximized the separability in RI and non-RI cases was retained for each GSFA field. This first phase of the method yielded 850 individual dimension-reduced PC score matrices (one for each GSFA field) from which clustering was employed. This approach was repeated for all eight RI definitions.

Once the optimal PCA configuration was identified, a $k$-means cluster analysis was done on each individual PC score matrix (850 total analyses), retaining two clusters each time (intended to represent RI and non-RI classes). This $k$-means cluster analysis was fully driven by the data; no prior knowledge of RI/non-RI class was provided (as is typical in unsupervised learning). Each $k$-means clustering was configured with 100 random restarts (Wilks 2011) to maximize the repeatability of the results, where the most frequent cluster center obtained from these 100 restarts was retained. Once the two clusters were computed, RI/non-RI cases from the HURDAT2 data were counted within each cluster. The cluster with the larger number of RI cases was deemed the RI cluster, and the other cluster was forced to be the non-RI cluster regardless of its count of non-RI cases (to obtain separated classes). Once the RI and non-RI clusters were identified, each
clustered case was given one of 4 possible outcomes, either RIs in the RI cluster (hereinafter referred to as true positives), RIs in the non-RI cluster (false negatives), non-RIs in the RI cluster (false positives), and non-RIs in the non-RI cluster (true negatives). We then scaled each count by the total number of RIs (for true positive and false negatives) or the total number of non-RIs (for true negatives and false positives) to obtain the true positive rate (TPR; the percentage of RIs in the RI cluster), false positive rate (FPR; the percentage of non-RIs in the RI cluster), false negative rate (FNR; the percentage of RIs in the non-RI cluster), and true negative rate (TNR; the percentage of non-RIs in the non-RI cluster). Last, we computed Youden’s J index (Youden 1950) as a measure of separability:

\[ J = (TPR + TNR) - 1. \]

A perfect clustering of cases would yield a J value of 1, and J values could be negative in instances where the majority of RIs and non-RIs were in the same cluster (an undesirable result). Note that these are not truly hits or misses; no forecast was performed by the unsupervised learning. Instead, our separation approach identified RI cases that resembled non-RIs (and vice versa), helping to reveal features that make RI/non-RI discrimination more difficult.

Once the J metric was computed for each individual GFSA field and RI definition, the GFSA variable with the maximum J for each RI definition was deemed the best-separating field. Once the optimal separating field and associated unsupervised learning technique was found for each RI definition, composites of these best fields were formulated by averaging all cases in each of the four outcome categories listed above. These composites allowed us to identify physical differences in the TC environments for the best-separating variable for each RI definition, which may provide insight to existing predictors in baseline SHIPS-RII predictors. We also included J computed directly from a k-means clustering on the raw GFSA anomaly fields (the Z matrix prior to PCA preprocessing) to show the benefit (if any) of the PCA preprocessing in improving separability in the categories (section 4).

c. Comparison with SHIPS-RII predictors

The methods described above helped to identify the GFSA fields that best separate RI and non-RI environments. However, given the research goal of identifying fields that improve separability over current operational methods, it was also necessary to quantify improvements in separability relative to predictors used operationally in the SHIPS-RII. For this, J was computed using the full set of SHIPS-RII predictors (obtained from the SHIPS data archive) for each of the eight RI definitions valid at the onset of the intensity change episode. This was done by performing a k-means cluster analysis on the SHIPS-RII predictors valid at the onset of the intensity change period (Table 3). For six of the eight RI definitions, the SHIPS-RII data grouped over 50% of the non-RI cases with the RI cluster, demonstrating a lack of separability. Interestingly, the 65-kt/72-h SHIPS-RII cluster analysis grouped almost all RIs into one cluster. However, over 60% of non-RIs were included in that RI cluster, meaning separability was still limited in this instance (despite a relatively high J = 0.33). SHIPS-RII predictors showed the best overall separability for the 20-kt/12-h (J = 0.24) and 40-kt/24-h (J = 0.25) definitions, although the RI

<table>
<thead>
<tr>
<th>RI definition</th>
<th>TPR</th>
<th>TNR</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 kt</td>
<td>55.1%</td>
<td>68.8%</td>
<td>0.240</td>
</tr>
<tr>
<td>25 kt</td>
<td>65.2%</td>
<td>32.6%</td>
<td>-0.021</td>
</tr>
<tr>
<td>30 kt</td>
<td>61.8%</td>
<td>32.4%</td>
<td>-0.058</td>
</tr>
<tr>
<td>35 kt</td>
<td>50.8%</td>
<td>32.1%</td>
<td>-0.172</td>
</tr>
<tr>
<td>40 kt</td>
<td>57.1%</td>
<td>67.8%</td>
<td>0.250</td>
</tr>
<tr>
<td>45 kt</td>
<td>73.6%</td>
<td>32.9%</td>
<td>0.065</td>
</tr>
<tr>
<td>55 kt</td>
<td>86.5%</td>
<td>33.9%</td>
<td>0.204</td>
</tr>
<tr>
<td>65 kt</td>
<td>97.1%</td>
<td>35.4%</td>
<td>0.325</td>
</tr>
</tbody>
</table>

3. Unsupervised learning results

Table 4 outlines the PCA configuration (either RPCA or one of the 18 KPCA configurations tested) that yielded maximum separability J for each RI definition, as well as the TPR, TNR, and J statistic. Table 5 lists the associated GFSA field and domain size configuration (Fig. 2) that provided the maximum separability for each RI definition. These results show the benefits of nonlinear separability introduced by the matrix K, as six of the eight RI definitions were best separated when utilizing KPCA, typically with either a high degree polynomial or a radial basis function. Only the 20-kt/12-h and 35-kt/24-h categories showed better separability with RPCA, although the performance differences among the RPCA results and the next highest ranked KPCA results were only slightly lower (<10%; not shown).

When comparing the unsupervised learning results with those results that did not include the PCA preprocessing (Fig. 4, blue points), the benefits of PCA were mixed. The PCA
Table 4. True positive rate and true negative rate for each RI definition using our unsupervised learning approach with the GFSA. The first column gives the RI definition, the second column is the optimal PCA preprocessing method, the third column is the kernel for KPCA (if any), the fourth column is the number of PCs retained, the fifth column is the TPR, the sixth column is the TNR, and the last column is the $J$ statistic. Note that the false positive rate and false negative rate can be computed from the table (e.g., $\text{FNR} = 100\% - \text{TPR}$).

<table>
<thead>
<tr>
<th>RI definition</th>
<th>KPCA or RPCA</th>
<th>KPC kernel</th>
<th>No. PCs</th>
<th>TPR</th>
<th>TNR</th>
<th>$J$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 kt/12 h</td>
<td>RPCA</td>
<td>—</td>
<td>3</td>
<td>78.6%</td>
<td>51.8%</td>
<td>0.304</td>
</tr>
<tr>
<td>25 kt/24 h</td>
<td>KPCA</td>
<td>RBF ($\sigma = 1000$)</td>
<td>5</td>
<td>74.9%</td>
<td>53.9%</td>
<td>0.288</td>
</tr>
<tr>
<td>30 kt/24 h</td>
<td>KPCA</td>
<td>Polynomial ($d = 9$)</td>
<td>7</td>
<td>57.1%</td>
<td>74.1%</td>
<td>0.313</td>
</tr>
<tr>
<td>35 kt/24 h</td>
<td>RPCA</td>
<td>—</td>
<td>5</td>
<td>81.1%</td>
<td>53.5%</td>
<td>0.346</td>
</tr>
<tr>
<td>40 kt/24 h</td>
<td>KPCA</td>
<td>RBF ($\sigma = 200$)</td>
<td>6</td>
<td>84.5%</td>
<td>55.4%</td>
<td>0.400</td>
</tr>
<tr>
<td>45 kt/36 h</td>
<td>KPCA</td>
<td>Polynomial ($d = 10$)</td>
<td>3</td>
<td>77.1%</td>
<td>57.1%</td>
<td>0.343</td>
</tr>
<tr>
<td>55 kt/48 h</td>
<td>KPCA</td>
<td>RBF ($\sigma = 25$)</td>
<td>7</td>
<td>86.5%</td>
<td>50.0%</td>
<td>0.366</td>
</tr>
<tr>
<td>65 kt/72 h</td>
<td>KPCA</td>
<td>Polynomial ($d = 2$)</td>
<td>8</td>
<td>89.3%</td>
<td>55.2%</td>
<td>0.446</td>
</tr>
</tbody>
</table>

Some preprocessing did not offer statistically significant improvements in $J$ despite showing improvement in six of the eight RI definitions. With most RI definitions, the $J$ value for the unsupervised approach with no PCA preprocessing fell within the middle 50% of the bootstrap replicates of the PCA preprocessed $J$ results. Given the lack of significant improvements, the results were essentially indistinguishable for the shorter duration (<36 h) RI definitions, and the longer RI definitions offered improvements in $J$ of 15%–20% when PCA preprocessing was included. Given the trend of improving separability when using PCA preprocessing, we elected to keep the preprocessing step to maximize separability despite the lack of significance in the results.

Although the PCA preprocessing added little additional benefit, the unsupervised learning method increased separability over the SHIPS-RII predictors (Table 3) as the bootstrap replicates of $J$ (boxplots in Fig. 4) showed statistically significant improved separability over SHIPS-RII for all but three RI definitions. Additionally, despite the results not showing statistical significance, improvements in $J$ from the unsupervised learning method (relative to SHIPS-RII) ranged from 30% to 40% for all RI thresholds except the 20-kt/12-h definition. It is likely that RI occurring at the 12-h time scale is strongly influenced by internal processes (e.g., Judt and Chen 2016) that cannot be well represented by the 1° GFSA fields used in this study. The limited number of RI cases (2.88%) associated with the 40-kt/24-h RI definition was likely the primary reason for the nonsignificant improvements for that definition, since unsupervised learning techniques benefit from large datasets (Dolnicar et al. 2013). Although the 65-kt/72-h RI definition is relatively new (not included in Kaplan et al. 2015), the relatively small sample size of 72-h periods ($N = 1760$) likely affected our ability to discriminate RI and non-RI environments. More work is needed to establish the causes for the lack of significant improvements fully, but it is still important to note that $J$ increased for all RI definitions when using unsupervised learning, and with increasing RI database sizes, significant differences are likely to result in the future.

Most separability improvements offered by the unsupervised learning approach resulted from increasing the TNR. For six of eight RI definitions, the SHIPS-RII predictors grouped a majority of RIs and non-RIs in the same cluster, which increased the likelihood of false negatives and false positives, producing negative $J$ values for three of the eight RI definitions. Our unsupervised learning approach grouped at least 50% of the RIs in one cluster and 50% of the non-RIs in the other cluster for every RI definition, an improvement over the SHIPS-RII that could reduce the false positive rate when applied to a statistical classification algorithm (although such work is outside the scope of this study). Through a reduction in the misclustering rate in both the RI and non-RI clusters via unsupervised learning, the $J$ statistic improved for all eight RI definitions even though only five of those improvements demonstrated statistical significance. Overall, the unsupervised learning techniques provided improved separability relative to SHIPS-RII predictors. Next, we constructed composites for these best-separating fields to ascertain the distinctions in each variable for RI/non-RI classification (section 4).

4. Composite results

As stated previously, cases within each cluster analysis were grouped into one of four possible categories using the best-separating GFSA field for each RI definition: true positives, false negatives, false positives, and true negatives (defined in section 2b). The results highlighted some generally consistent trends among the best-separating fields for each RI definition. First, the large number of non-RI cases (90%–97% of $N$ from Table 1) for each RI definition resulted in highly negatively correlated false positive and true negative composites ($r \sim -0.99$). Although the non-RI composites were spatially

Table 5. GFSA fields and domain sizes associated with the optimal KPCA/RPCA configurations outlined in Table 4. Column 3 is the vertical level of the level of interest.

<table>
<thead>
<tr>
<th>RI definition</th>
<th>Variable name</th>
<th>Vertical level</th>
<th>Domain size</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 kt/12 h</td>
<td>Specific humidity</td>
<td>300 hPa</td>
<td>$2^\circ \times 2^\circ$</td>
</tr>
<tr>
<td>25 kt/24 h</td>
<td>Relative humidity</td>
<td>500 hPa</td>
<td>$10^\circ \times 10^\circ$</td>
</tr>
<tr>
<td>30 kt/24 h</td>
<td>Temperature</td>
<td>1000 hPa</td>
<td>$18^\circ \times 18^\circ$</td>
</tr>
<tr>
<td>35 kt/24 h</td>
<td>Absolute vorticity</td>
<td>400 hPa</td>
<td>$10^\circ \times 10^\circ$</td>
</tr>
<tr>
<td>40 kt/24 h</td>
<td>CAPE</td>
<td>Surface</td>
<td>$14^\circ \times 14^\circ$</td>
</tr>
<tr>
<td>45 kt/36 h</td>
<td>Relative humidity</td>
<td>400 hPa</td>
<td>$18^\circ \times 18^\circ$</td>
</tr>
<tr>
<td>55 kt/48 h</td>
<td>Specific humidity</td>
<td>925 hPa</td>
<td>$6^\circ \times 6^\circ$</td>
</tr>
<tr>
<td>65 kt/72 h</td>
<td>Absolute vorticity</td>
<td>850 hPa</td>
<td>$18^\circ \times 18^\circ$</td>
</tr>
</tbody>
</table>
similar, the clustering algorithm’s separating of the non-RI events into these two groups was a consequence of important magnitude differences between the non-RI groups and is therefore reliable. Second, we computed average storm motion for each composite and found that all four groups showed a general west-northwest trajectory with minimal variability. Given this consistency in storm trajectory among the eight RI categories, we expect our composites would strongly resemble those from a storm-relative framework that is frequently used for spatial TC analyses (e.g., Sun et al. 2019). Finally, we observed a general tendency for RI cases to be farther south and west in the study domain. This result suggests latitude and longitude could be useful predictors in a RI/non-RI classification model such as SHIPS-RII (although such development is outside the scope of this work). These consistencies among definitions should be considered when interpreting the results presented below.

In addition to generating the spatial composites, we computed permutation tests (Efron and Tibshirani 1993) comparing the GFSA field magnitudes between the composites in the RI cluster (true positive and false negative) and composites in the non-RI cluster (true negative and false positive). These tests were performed on each grid point in the composite domain using the raw GFSA values for each constituent case in each cluster. All grid points significant at $\alpha = 0.05$ were shaded in the composite figures to identify regions of statistical significance.

a. 20-kt/12-h RI definition

The unsupervised learning results for the 20-kt/12-h RI definition found that the 300-hPa specific humidity GFSA field on the $2^\circ \times 2^\circ$ spatial domain provided the maximum separability (Fig. 5), implying inner-core upper-level moisture may help distinguish RI from non-RI. Despite the coarse nature of the 1° GFSA input, the unsupervised learning approach still highlighted the inner-core region, where smaller-scale processes previously noted to be important for short-term RI would occur (e.g., Judt and Chen 2016).

Our results indicate above-normal moisture is present at 300 hPa for true positive and false positive (Figs. 5a,b) cases relative to the total sample mean, which contrasts with below-normal moisture for true negative and false negative composites (Figs. 5c,d). Greater moisture at upper levels may indicate a well-moistened atmospheric column that can better support convection and thus be more conducive to short-term RI, which may be why so few RI events (19) were associated with below-average moisture (Fig. 5c). It also implies that lower-entropy environmental air has not infiltrated the upper-level inner-core region for most RI cases. Conversely, the below-normal moisture signal for the non-RI composites may signify an adverse influence from the environment that inhibits short-term RI. Since we only show the best-separating field, it is likely that other factors besides 300-hPa moisture are present in the false negative cases that support RI, and likewise that other RI-necessary environmental factors are lacking in the false positive composite (Fig. 5b).

Since our study approach enables spatial analysis, we can detect a moisture gradient in all four composites oriented either northeast to southwest (Figs. 5a,b) or southwest to northeast (Figs. 5c.d). The true positive composite (Fig. 5a) showed an area of slightly drier air northeast of the TC center, potentially a consequence of the Bermuda high that is frequently present to the north and east of a westward-advancing Atlantic TC. Further, those RI cases in the false negative composite showed the same southwest-to-northeast moisture gradient characteristic of the true negative composite (Fig. 5d), which contrasts the RI environment in the true positive cases.

These results were valid on the smallest TC domain size considered ($2^\circ \times 2^\circ$), suggesting environmental factors (including moisture) farther from the TC center were less important for 20-kt/12-h RI according to the unsupervised learning method. Such a result corroborates work that shows inner-core, smaller-scale processes dominate for RI (e.g., Hendricks et al. 2010), such as the distribution of diabatic heating (e.g., Nolan et al. 2007). Note that this RI definition is typically the most difficult to classify (Kaplan et al. 2015), and our study relies on data too coarse to resolve small-scale features, so the lack of evident RI/non-RI distinguishing characteristics in these composites was anticipated. Importantly, these results were all statistically significant, implying that the orientation of the moisture gradient coupled with better-resolved small-scale features near the TC center may help distinguish RI from non-RI environments.

b. 25-kt/24-h RI definition

Our analysis found that midlevel (500 hPa) relative humidity on a $10^\circ \times 10^\circ$ domain was the best-separating field for RI and non-RI environments for 25-kt/24-h RI. The true positive (Fig. 6a) and true negative (Fig. 6d) composites were strongly negatively correlated (coefficient $r = -0.93$), with a maximum southwest of the TC center in the true positive composite, suggesting that the orientation of the midlevel relative
humidity field can help identify RI likelihood. We found that this elevated moisture in the true positive composite was upshear of the TC center (not shown), which has been shown in previous studies to be important for intensification (Bhalachandran et al. 2019). The false positive cluster (Fig. 5b) showed a similar orientation and maximum of relative humidity southwest of the TC. However, the positive anomaly magnitudes were smaller and the negative anomalies to the northeast were larger, implying that the enhanced moisture to the southwest could not be fully utilized for RI once it was advected around the center due to relatively drier air in the northeast quadrant. The false negative composite (Fig. 5c) showed an orientation like the true negative non-RI group (Fig. 6d), although the magnitude of the relative humidity gradient was larger and the maximum was nearer the TC center. Considering the largely west-northwestward motion of these

300 hPa Specific Humidity

Fig. 5. Composite 300-hPa specific humidity on the $2^\circ \times 2^\circ$ domain for the 20-kt/12-h RI definition: (a) the true positive composite, (b) the false positive composite, (c) the false negative composite, and (d) the true negative composite (organized like a traditional contingency table; Wilks 2011), where $N$ indicates the number of included cases. Blue shading represents standardized anomalies below the mean; red shading represents anomalies above the mean. The hatched regions show significant differences at $\alpha < 0.05$ comparing the RI composites [in (a) and (c)] and non-RI composites [in (b) and (d)].
cases, it is possible that the positive anomaly in the false negative composite (Fig. 6c) is near a region of low-level convergence in the right-front quadrant and thus supports deepening convection in this region (e.g., Frank and Ritchie 1999).

Given the relatively small fraction of RI events in the false negative composite (~25%), our results imply a good environmental RI indicator may be a strong maximum of midlevel relative humidity southwest of the TC center, which decreases the likelihood of midlevel ventilation and increases instability for deepening convection, both of which should increase the probability of RI (Tang and Emanuel 2012). In these composites, the relative humidity maxima and minima in each region were statistically significantly different among the RI composites (Figs. 6a,c) and non-RI composites (Figs. 6b,d), further demonstrating the potential applicability of our results.

c. 30-kt/24-h definition

The 30-kt/24-h RI definition results showed that 1000-hPa temperature on the largest $18^\circ \times 18^\circ$ domain (Fig. 7) was best for distinguishing RI and non-RI environments. These results were surprising given the generally barotropic conditions in the tropics, but the true positive composite

---

**Fig. 6.** As in Fig. 5, but for 500-hPa relative humidity on a $10^\circ \times 10^\circ$ TC-centric grid valid for the 25-kt/24-h RI definition. Wind vectors represent the average background flow for all cases in the given cluster.
(Fig. 7a) revealed a northward-oriented temperature gradient present in most RI cases (57%). Note that the 1000-hPa temperature standard deviations range between 1 and 3 K at each grid point in the RI hit cluster, meaning that the average along-gradient temperature change is only 1–2 K across the entire domain.

Since the average storm motion associated with the true positive composite was toward the west-northwest, the temperature gradient (Fig. 7a) suggests the composite TC may be moving around the periphery of the Bermuda high. Similarly, the true negative composite (Fig. 7d) shows a progression into relatively cooler air following the same storm motion. The pattern for the false negative composite (Fig. 7c) resembles the true negative composite but with higher temperatures, particularly on the equatorward side of the composite TC, although temperatures were slightly above average in the false negative composite. Similarly, the false positive composite (Fig. 7b) was nearly identical to the true positive map ($r = 0.91$), although the temperature gradient magnitude in the RI composite was slightly greater. As was the case with the 25-kt/24-h RI definition, statistically significant differences were observed in the warm and
cool maxima in each composite domain. These results reinforce the importance of the magnitudes of these features. For example, a warm maximum on the northeast side of the true positive composite is a good indication of RI that results in a relatively small percentage of misclustered non-RIs (748 of 2781, or ~27%). Both the false and true negative composites (Figs. 7c,d) show a generally northerly temperature gradient, which suggests the TC is moving into a region with reduced atmospheric instability and thus less low-level support for convection. The orientation of the temperature gradient, when coupled with the average west-northwest storm trajectory, is clearly isolated by the clustering algorithm as important for identifying RI environments at the 30-kt/24-h threshold.

d. 35-kt/24-h definition

For the 35-kt/24-h category, 400-hPa relative vorticity on the 10° × 10° GFSA domain (a radius of ~550 km from the GFSA estimated TC center) yielded the optimal separability between RI and non-RI events. Both the true and false positive composites (Figs. 8a,b) show a maximum positive anomaly near the TC center, which is stronger in the true positive composite. However, the positive relative vorticity maximum is better
aligned with the composite TC center in Fig. 8b, which may indicate already intense TCs. The average intensity in the false positive composite was roughly 68 kt for 1322 cases, while average intensity for the true positive composite was 65 kt for 68 cases, although the difference was not statistically significant \((p = 0.395\) from a permutation test). In the false and true negative composites, there is a relative vorticity maximum slightly northeast of the TC center in the false negative composite (Fig. 8c) that was absent from the true negative composite (Fig. 8d). Instead, the clustering algorithm focused on the elevated anticyclonic vorticity south-southwest of the TC in the true and false negative composites, a feature absent in either positive composite. Overall, these results suggest that TCs with stronger near-center relative vorticity magnitudes that appear slightly displaced from the TC center (as in Figs. 8a,c) are more likely to undergo 35-kt/24-h RI, a factor that may precede increasing vortex alignment and thus intensification (e.g., Tao and Zhang 2014). These regions also appeared statistically significantly different in the RI composites (Figs. 8a,c), further supporting this conclusion. However, additional work using a higher spatial resolution dataset would be necessary to determine the extent of this vortex tilt. The dramatic improvement in separability measure with the unsupervised learning technique over SHIPS-RJ (\(I\) increase from \(-0.172\) to \(0.350\)) suggests this field should not be ignored when evaluating the likelihood of 35-kt/24-h RI, despite the large percentage of non-RI cases in the false negative cluster (\(~47\%)\).

e. 40-kt/24-h definition

Surface-based CAPE fields best distinguished RI from non-RI cases for the 24-h RI category with the greatest magnitude of intensification (40 kt/24 h). The CAPE true positive pattern aligns with the 1000-hPa temperature field noted earlier (Fig. 7a; section 4.3), as higher CAPE appears associated with warmer air to the north in the 40-kt/24-h true positive composite (Fig. 9a). This pattern lends support to our earlier assumption that greater low-level thermodynamic energy was likely present at the onset of the intensification period. The composite wind field in Fig. 9a implies inflow of unstable air from the region of above-average CAPE located northwest of the TC center. When coupled with the northwestward storm motion discussed in section 4c, this pattern suggests the composite TC is moving into an environment suitable for deepening convection. The false positive composite (Fig. 9b) shows a similar pattern, although comparatively less CAPE wraps around the composite TC center and the maximum northwest of the TC center for true positive cases is not present for false positive cases (and this difference was statistically significantly different—not shown). Interestingly, the true negative composite (Fig. 9d) suggested the opposite pattern, namely that the composite TC was advancing into a region with weaker CAPE and therefore more stable air compared with the sample mean. False negative cases (Fig. 9c) did not show the maximum of CAPE northwest of the TC center (although only 6 cases, or \(~10\%)\% of RI events, are included), and this difference was statistically significant. This result suggests along-track surface CAPE maxima may be useful features for identifying RI/non-RI environments in future classification work.

f. 45-kt/36-h definition

For the 45-kt/36-h RI definition, relative humidity was once again an important measure for distinguishing RI and non-RI environments, this time at 400 hPa. This definition is associated with RI over a longer period, although the intensification rate is the same as the 30-kt/24-h definition previously examined, thus it is not surprising that our unsupervised learning approach again identified a thermodynamic variable. Both the true positive and false negative composites (Figs. 10a,c) showed a maximum of relative humidity collocated and slightly south-southwest of the TC center, features that were missing from the false positive and true negative composites (Figs. 10b,d). We expect the unsupervised learning method identified the mean moisture gradient as southerly in the false negative composite (Fig. 10c), resulting in those cases being separated from the true positive composite and again suggesting the moisture gradient may be a useful discriminator between RI and non-RI environments. The difference in moisture gradient was statistically significant between the two non-RI composites (Figs. 10b,d) and the two RI composites (Figs. 10a,c), further emphasizing its importance. This suggests that an abundance of relative humidity near the TC center compared with the sample mean, which was largely missing from both non-RI composites (Figs. 10b,d), may be useful when diagnosing RI onset. Given the proximity to the composite TC center, these relative humidity maxima in the RI cluster composites (Figs. 10a,c) suggest inner-core air is closer to saturation, reducing the effect of downdraft evaporative cooling (Braun et al. 2012), which may be an important indicator for RI. Similarly, the non-RI composites show relative humidity values below to slightly above average near the TC center, with the highest relative humidity anomalies displaced far from the TC center and likely not having a major influence on intensification. Higher values as well as stronger gradients of upper-level relative humidity have been associated with larger TC intensification rates (Wu et al. 2012), and our results further support the inclusion of this field in statistical RI prediction schemes.

g. 55-kt/48-h definition

Low-level moisture (925-hPa specific humidity on the \(6^\circ \times 6^\circ\) grid) near the TC center became an important indicator of RI at the 55-kt/48-h threshold. Given the average storm motion for all four composites remains west-northwestward, the true positive composite (Fig. 11a) suggests the TC is moving into an area with an abundance of low-level moisture. The false positive composite depicts a similar setup (Fig. 11b), although the magnitudes of the moisture maxima are much lower. The false negative pattern (Fig. 11c) shows a general maximum of moisture near the TC center, which is characteristic of intensifying TCs (Martinez et al. 2017), but given the average west-northwestward trajectory, the unsupervised learning method likely grouped these RIs with the non-RI cases as the associated TCs are moving into a lower moisture environment. However, this pattern was only shared among a small fraction (\(~12\%)\% of RI cases. Finally, the true negative composite TC (Fig. 11d) is moving into a comparatively dry low-level environment, which is less supportive of deep convection. Once again, the orientation of the moisture gradient was the primary distinguishing characteristic between the true
positive and true negative composites and resulted in statistically significant differences across most of both non-RI composites (Figs. 11b,d). Our results that highlight low-level specific humidity corroborates existing RI prediction schemes, which include an inner-core dry air predictor to capture the state of low-level moisture near the TC center (Kaplan et al. 2015).

h. 65-kt/72-h definition

The 65-kt/72-h RI definition included the smallest number of total TC samples ($N = 1760$), which likely affected the separation ability of the unsupervised learning method. Regardless, the algorithm revealed that 850-hPa absolute vorticity on the $18^\circ \times 18^\circ$ grid was the most useful GFSA field for isolating RI and non-RI environments. The true positive composite (Fig. 12a) showed a minimum in absolute vorticity near the TC center, suggesting the TC is likely weaker than a typical non-RI case (Fig. 12d) at the onset of this extended intensification period. The mean TC intensity in the true positive composite (38 kt for 101 samples) supported this conclusion when compared with the mean TC intensity for the true negative composite (67 kt for 898 samples), and this difference was statistically significant ($p = 0.001$). The elevated absolute vorticity values (and intensity) in the true negative composite (Fig. 12d) suggest the storm has little room to intensify 65 kt in

![Surface-Based CAPE](image-url)

Fig. 9. As in Fig. 6, but for surface-based CAPE on a $14^\circ \times 14^\circ$ TC-centric grid valid for the 40-kt/24-h RI definition.
The false negative composite was noisy (Fig. 12c) with no discernable patterns, a result of environmental variability within non-RI TCs at 72-h lead time and a small number (12) of RI events in this composite. Finally, the false positive composite (Fig. 12b) showed the inverse pattern ($r = -0.99$) to the true negative composite with slightly stronger vorticity magnitudes than the true positive composite (and these differences were statistically significant). The relative abundance of 850-hPa absolute vorticity suggests other factors are leading to TC intensity changes below a 65-kt/72-h rate in the false positive composite, such as the negative synoptic factors found by Hanley et al. (2001).

5. Discussion and conclusions

This study explored the utility of the unsupervised learning methods—namely, cluster analysis with PCA preprocessing—in identifying covariates useful for separating RI and non-RI environments in Atlantic TCs. The primary goal of the work was twofold. The first objective was to identify the optimal atmospheric variable and isobaric level that will yield improved
RI/non-RI separability relative to operationally used SHIPS-RII predictors for eight unique RI definitions. The second objective was to evaluate the physical relationships of these best-separating variables within their respective RI categories to demonstrate their potential for use as predictors in future classification schemes.

GFSA fields were used to quantify atmospheric characteristics underlying each TC case from 2004 to 2016 (a total of 3605 unique TC observations). A total of 170 GFSA atmospheric fields on five different domain sizes were tested for separability (a total of 850 individual GFSA fields), where separability was derived by a metric that quantified the percentages of RI and non-RI cases in the RI and non-RI clusters [Eq. (8)]. To accomplish this separation, we used a combination of RPCA/KPCA with cluster analysis (an unsupervised learning approach) to identify two clusters, one with a majority of RIs and another that had the minority of RIs (which typically also had the majority of non-RIs, but not always). Bootstrapped separability metric ($J$) values were used to assign statistical significance (or the lack thereof) to the increased separability offered by the unsupervised learning methods over SHIPS-RII predictors.

FIG. 11. As in Fig. 6, but for 925-hPa specific humidity on a 6° × 6° TC-centric grid valid for the 55-kt/48-h RI definition.
The resulting analyses (Fig. 4) showed that for five of the eight RI definitions, our unsupervised learning method improved separability relative to SHIPS-RII for all eight RI definitions considered, and that this improvement was statistically significantly better for five of the eight definitions. Additionally, these analyses revealed that the RPCA/KPCA preprocessing employed offered minimal additional separability benefit (<10%) for the shorter-duration RI categories, although the benefits of RPCA/KPCA were more evident for the 45-kt/36-h, 55-kt/48-h, and 65-kt/72-h rates. However, no statistically significant improvements in $J$ were seen when including the RPCA/KPCA preprocessing. After this initial assessment, four composites for the best-separating GFSA field for each RI definition were formulated based on true positives (RI cases in the RI cluster), true negatives (non-RI cases in the non-RI cluster), false negatives (RI cases in the non-RI cluster) and false positives (non-RI cases in the RI cluster).

Several separating GFSA fields were identified from the unsupervised learning methods, which were generally consistent across multiple RI definitions and can be summarized into six broad results:

**FIG. 12.** As in Fig. 6, but for 850-hPa absolute vorticity on an 18° × 18° TC-centric grid valid for the 65-kt/72-h RI definition.
1) True positive and false negative composites (which make up RI cases) consistently portrayed TCs moving into environments with thermodynamic ingredients conducive for RI, specifically regions with elevated relative and specific humidity coupled with above-average values of low-level instability (estimated by CAPE).

2) Mid- and upper-level relative humidity gradients frequently served as an important separating variable, specifically higher relative humidity located near the TC center in the true positive and false negative composites, suggesting a reduced likelihood of convective downdrafts within the TC core that is ideal for RI.

3) The most intense RI definition (40 kt/24 h) exhibited a true positive RI pattern associated with a composite inflow of CAPE west-northwest of the TC center in the direct path of the composite storm motion, which likely supported enhanced convection as the TC advanced into that favorable region (see point 1 above).

4) Abundant low-level moisture was important for the 55-kt/48-h category owing to a continued availability of fuel to aid in sustained intensification as the TC progressed through that environment.

5) For shorter-term RI (35 kt/24 h), relative vorticity magnitudes were higher near the centers of RI cases, suggesting TCs undergoing this rate of RI have strengthening midlevel circulations that may be aligning with the low-level center.

6) Despite the lower number of TC samples for the long duration 65-kt/72-h RI definition (N = 1760), which limited the separation ability, a relative lack of absolute vorticity in the true positive composite implies the TC is weaker and thus has time to deepen at least 65 kt over the extended time period (72 h), and these TCs were also in environments where surrounding absolute vorticity magnitudes were limited, suggesting these events were displaced from the influence of low-level cyclonic vorticity from the monsoon trough or the ITCZ, although true positive composites did show an area of anomalously large vorticity northeast of the TC center.

Several of these fields are not included in the currently operational version of SHIPS-RII nor the Bayesian and logistic regression RI models (Kaplan et al. 2015), particularly fields that incorporate midlevel relative vorticity or surface-based CAPE. Importantly, a majority of GFSA fields produced J values that exceeded the SHIPS-RII results for most definitions (although not always statistically significantly so), suggesting the unsupervised learning method could benefit from considering multiple processes simultaneously as opposed to the best-separating field presented here (a task outside of the scope of this project). Typically, the best-separating GFSA fields were consistent among multiple RI definitions, revealing humidity across a deep vertical layer, low-level instability, and midlevel vorticity as potentially important separating fields for RI/non-RI environments. Ultimately, the improved separability offered by the presented results reinforce existing knowledge of RI environment characteristics and offer an updated set of fields that could be used in a predictive scheme to increase RI/non-RI classification ability. Although many of these variables are included in the SHIPS-RII, the spatial information offered by the KPCA methods presented herein showed additional RI/non-RI separability and should be considered in future RI forecasting studies.

The above results should be considered in the context of their potential limitations. First, the composite approach only considered the best-separating GFSA field for each RI definition, which dramatically simplified the explanation of the complex processes involving RI. Since a goal of this work was to identify spatial characteristics of environmental fields that improve separability over SHIPS-RII, it was not possible to address covariability among all considered GFSA fields in this initial study, although covariability should be considered in future work. Further, using 1º GFSA TC locations and variables introduced undiagnosed error in the initial TC placement and data used to generate the composites. Additionally, a higher spatial resolution dataset with a long period of record could elucidate the importance of some of the smaller-scale processes for RI not characterized by the coarse GFSA. Although these are important considerations, we are better prepared to transfer the results of this study into an operational forecast by selecting GFSA as our meteorological dataset. Future work will address these limitations by identifying novel ways to quantify the covariability among the GFSA fields contributing most strongly to RI and then implementing those fields in a truly operational forecast.

These results highlight the benefit of unsupervised learning approaches in difficult classification scenarios such as the RI problem discussed herein and offer promise for future studies aimed at improving RI/non-RI separability with the ultimate goal of improving RI forecasts.

Acknowledgments. This project was funded by NOAA Grant NA17OAR4590140 as part of the Joint Hurricane Testbed program. We thank the reviewers for their insightful feedback that helped to improve this work.

Data availability statement. The data used in this study are available upon request from the corresponding author. The original GFSA fields were obtained online (https://www.ncdc.noaa.gov/nomads).

REFERENCES