Application of Clustering Algorithms to TRMM Precipitation over the Tropical and South Pacific Ocean

MAXWELL PIKE

Department of Environmental Sciences, Rutgers, The State University of New Jersey, New Brunswick, New Jersey

BENJAMIN R. LINTNER

Department of Environmental Sciences, Rutgers, The State University of New Jersey, and Rutgers Institute of Earth, Ocean, and Atmospheric Sciences, New Brunswick, New Jersey

(Manuscript received 16 July 2019, in final form 27 January 2020)

ABSTRACT

Understanding multiscale rainfall variability in the South Pacific convergence zone (SPCZ), a southeastward-oriented band of precipitating deep convection in the South Pacific, is critical for both the human and natural systems dependent on its rainfall, and for interpreting similar off-equatorial diagonal convection zones around the globe. A k -means clustering method is applied to daily austral summer (December–February) Tropical Rainfall Measuring Mission (TRMM) satellite rainfall to extract representative spatial patterns of rainfall over the SPCZ region for the period 1998–2013. For a $k = 4$ clustering, pairs of clusters differ predominantly via spatial translation of the SPCZ diagonal, reflecting either warm or cool phases of El Niño–Southern Oscillation (ENSO). Within each of these ENSO phase pairs, one cluster exhibits intense precipitation along the SPCZ while the other features weakened rainfall. Cluster temporal behavior is analyzed to investigate higher-frequency forcings (e.g., the Madden–Julian oscillation and synoptic-scale disturbances) that trigger deep convection where SSTs are sufficiently warm. Pressure-level winds and specific humidity from the Climate Forecast System Reanalysis are composited with respect to daily cluster assignment to investigate differences between active and quiescent SPCZ conditions to reveal the conditions supporting enhanced or suppressed SPCZ precipitation, such as low-level poleward moisture transport from the equator. Empirical orthogonal functions (EOFs) of TRMM precipitation are computed to relate the “modal view” of SPCZ variability associated with the EOFs to the “state view” associated with the clusters. Finally, the cluster number is increased to illustrate the change in TRMM rainfall patterns as additional degrees of freedom are permitted.

1. Introduction

Precipitating deep convection over the tropical ocean is observed to organize into large-scale regional rainbands, including zonally oriented bands such as the Pacific intertropical convergence zone (ITCZ) and tilted or diagonal bands such as the South Pacific convergence zone (SPCZ). Oriented along a northwest-to-southeast axis, the SPCZ extends from the equatorial western Pacific warm pool to the midlatitudes of the south central Pacific (Trenberth 1976; Vincent 1994). Although some dynamic and thermodynamic controls on the occurrence and organization of the SPCZ have been identified, the multiscale nature of the processes governing

precipitating deep convection in the SPCZ is complex (Power 2011). For example, synoptic-scale variability of the SPCZ associated with midlatitude wave activity and transient disturbances has been described (Widlansky et al. 2011), but this high-frequency activity occurs in conjunction with intraseasonal variability such as the Madden–Julian oscillation (MJO; Matthews et al. 1996; Matthews 2012; Haffke and Magnusdottir 2013), interannual variability such as El Niño–Southern Oscillation (ENSO; Trenberth 1976; Folland et al. 2002; Vincent et al. 2011), and lower-frequency variability such as the interdecadal Pacific oscillation (IPO; Salinger et al. 2001; Folland et al. 2002; Linsley et al. 2008).

The ENSO influence on the SPCZ has been described primarily in terms of a spatial displacement of the principal diagonal axis of SPCZ precipitation: under

Corresponding author: Maxwell Pike, mpik1@rutgers.edu

ENSO warm phase, or El Niño, conditions, the SPCZ diagonal shifts northeastward toward the anomalously warm central/eastern Pacific, whereas during ENSO cool phase, or La Niña, conditions, the main SPCZ diagonal shifts to the southwest (Folland et al. 2002). For ENSO-neutral (“normal”) conditions, the SPCZ is located between these extreme positions. Lintner and Boos (2019) relate the SPCZ shifts from ENSO forcing to atmospheric energetics constraints, as anomalous divergent column moist static energy (MSE) flux out of the ENSO source region during El Niño induces a shift of the SPCZ toward this MSE source region. Of course, the intensity and location of the SST anomalies vary from one ENSO event to another, as do the spatial shifts of the SPCZ diagonal. Moreover, the SPCZ may respond distinctly to different flavors of ENSO (Capotondi et al. 2015). For example, during some El Niño events, the SPCZ and ITCZ have been observed to merge as a single convection zone along the equatorial central to eastern Pacific (Vincent et al. 2011; Borlace et al. 2014).

For the MJO influence, during Wheeler–Hendon RMM index phases 6, 7, and to a lesser extent, 8, precipitation in the SPCZ is enhanced, whereas during phases 1 and 2 rainfall is suppressed (Wheeler and Hendon 2004). There is significant variation in the location and intensity of precipitation across the SPCZ region, especially given that various forcing mechanisms may act together. For instance, ENSO largely dictates where thermodynamic conditions favorable to deep convective precipitation will occur, through the effect of anomalous SST on surface energy fluxes (Allen and Mapes 2017), and the MJO may enhance or suppress rainfall with conditions favoring large-scale ascent or descent and associated moistening or drying depending on MJO phase.

Prior work summarizes the mechanisms for interpreting SPCZ behavior by organizing these into two categories defined geographically, focusing either on processes originating to the east or west of the SPCZ. In terms of processes originating to the east, Takahashi and Battisti (2007) examine the role of orography over South America on the SPCZ. They hypothesize that a large area of orographically forced subsidence downstream of the Andes over the southeastern Pacific ultimately influences the location and orientation of the SPCZ by constraining the thermodynamic conditions under which precipitating deep convection develops. Lintner and Neelin (2008) describe how anomalous low-level advection of low MSE air into the SPCZ region through increasing trade wind strength in the presence of the mean large-scale tropospheric moisture gradient may suppress precipitation along the easternmost SPCZ.

By contrast, other studies point to processes affecting the SPCZ originating to its west. Trenberth (1976) identifies a “graveyard” for midlatitude fronts collocated with the SPCZ and emphasizes how Rossby wave activity directed into the SPCZ region advects positive vorticity, which in turn may trigger deep convection in thermodynamically favorable regions of elevated low-level MSE. Matthews (2012) also suggests that pulses of enhanced precipitation in the SPCZ are associated with midlatitude wave activity propagating along the subtropical jet; upon reaching the jet exit east of Australia, waves experience refraction in the basic state flow toward the westerly duct over the equatorial central Pacific. Cyclonic vorticity centers experience shearing associated with the equatorward decrease of zonal winds away from the axis of the jet, inducing a northwest to southeast tilt that may explain the diagonal orientation of the SPCZ. In a similar vein, van der Wiel et al. (2015, 2016a) examine the effects of Rossby waves advecting positive vorticity into the SPCZ region, triggering deep convective occurrence over sufficiently warm SSTs.

SPCZ convection shows more tropical characteristics in its western, more equatorward region and a mix of tropical and extratropical characteristics in its eastern, more poleward region (Kiladis et al. 1989). Kiladis et al. (1989) also indicate that the shape and orientation of the SPCZ are not solely tied to the distribution of Pacific SSTs, as the axis of maximum precipitation in the SPCZ does not align precisely with the axis of warmest SSTs extending southeastward from the western Pacific warm pool. Kiladis et al. (1989) present a series of sensitivity experiments exploring the influence of the Australian and South American landmasses on the SPCZ in which each of these landmasses were removed separately as well as in combination: while some changes in SPCZ strength were evident—a weaker SPCZ without Australia or slightly stronger without South America—the SPCZ persisted in all cases. This behavior supports influences of both eastern and western origin.

In the present study, we analyze daily precipitation data from the NASA Tropical Rainfall Measuring Mission (TRMM; Adler et al. 2003) for the period 1998 through 2013, focusing on the austral summer months, December, January, and February (DJF), when the SPCZ is climatologically most intense and spatially extensive (Meehl 1987; Vincent 1994). This dataset in particular was selected to establish a baseline to compare model performance against in future work. We opted to use TRMM over other, longer observational products such as GPCP and CMAP as these are not available at daily temporal resolution. We employ *k*-means clustering to extract representative patterns from daily TRMM data with an aim of advancing

mechanistic understanding of the controls on the SPCZ convection. Unlike empirical orthogonal function (EOF) analysis, which provides a “modal view” of variability, the use of clustering algorithms provides a discrete set of patterns in which the daily data are assigned, thus affording a “state view.” However, as we demonstrate explicitly below, EOF analysis of daily TRMM rainfall complements our k -means cluster results.

Our analysis bears some conceptual similarity to Haffke and Magnusdottir (2013), in which a Markov random field statistical model is applied to GridSat infrared retrievals to construct binary labels of SPCZ presence or absence. From these labels, metrics based on three aspects of seasonal-mean SPCZ shape were used to assign the SPCZ into eight categories. In a more recent study, Lorrey and Fauchereau (2017) apply k -means clustering to extract modes of geopotential height anomalies in the southwestern Pacific region. For a $k = 6$ clustering of surface level geopotential height from the NCEP–NCAR1 reanalysis for the period January 1950–December 2014, Lorrey and Fauchereau (2017) construct composites of zonal and meridional winds and precipitation to investigate the dominant synoptic types in the southwestern Pacific. Since Lorrey and Fauchereau (2017) do not explicitly focus on the DJF season in which SPCZ precipitation is most active, direct comparison of their results to ours is not possible, although as we highlight further below, there is some consistency between rainfall composites derived from their circulation-based clusters and our precipitation clusters. In this study, we explore a plausible range of k (from $k = 4$ to $k = 8$), guided by both physical considerations and a quantitative metric for estimating an upper bound on cluster number.

Questions remain about what dynamic and thermodynamic factors generate or control SPCZ rainfall, particularly with regard to such factors reflecting contributions associated with variability acting across multiple time scales. Specifically, while prior work has identified forcings affecting SPCZ intensity and location from synoptic through interdecadal time scales, how such forcings interact with one another to produce observed SPCZ behavior remains largely unexplored. Thus, one objective of the present study is to assess how SPCZ region precipitation responds in the presence of both low-frequency variability such as ENSO as well as higher-frequency variability including the MJO and synoptic time scale disturbances. To do this, we examine the temporal characteristics of the clusters, in particular aspects such as the characteristic persistence of each cluster and the day-to-day transitions between clusters.

We further seek to diagnose the thermodynamic and dynamic mechanisms of SPCZ region rainfall regimes identified by k -means cluster analysis through construction of composites of wind, specific humidity, and vorticity from the Climate Forecast System Reanalysis (CFSR; NCAR Staff 2007; Saha et al. 2010). One aim of our analysis is to frame the interpretation of SPCZ in a different light than the eastern versus western control paradigm of Takahashi and Battisti (2007), namely what we term lower- versus upper-level control, which we see as providing another useful framework for diagnosing SPCZ behavior. Since we plan to apply clustering approaches as tools for evaluating the SPCZ in climate models in future work, the present study also comprises a comprehensive observational baseline for evaluating simulations.

2. Datasets and methods

a. TRMM and CFSR

The TRMM 3B42 dataset provides four times daily (6 hourly) rainfall estimates at 0.25° resolution (Adler et al. 2003; see <https://gpm.nasa.gov/TRMM>). For our purposes, we aggregate these data to daily averages and to a $2.5^\circ \times 2.5^\circ$ spatial grid; differences in the clustering output at 0.25° versus 2.5° resolution are found to be negligible, especially in light of interest in relating the clusters to synoptic-scale variability, while the computational requirements are reduced for the coarser data. The TRMM data for DJF are analyzed over a domain spanning from 120°E to 90°W and from 35°S to 10°N . Leap days have been excluded so that each DJF season comprises 90 days, for a total of 1350 days over the entire period of 1998–2013. Aggregation of the TRMM data allows for comparison to CFSR reanalysis available at the same spatial resolution ($2.5^\circ \times 2.5^\circ$), while the four times daily CFSR data are also averaged to daily values.

b. Overview of k -means clustering and linear unidimensional scaling

The k -means clustering method is an unsupervised learning approach that seeks to classify data into k representative groupings (Abbas 2008). Considered in a representative multidimensional space, the data may be regarded as comprising a cloud of varying density throughout; k -means clustering seeks to identify regions of higher density within this space, with such regions corresponding to the clusters. Clusters are constructed by iterative assignment of input data based on the minimization of a user-specified distance metric between a given datum and the cluster centroid.

For k -means clustering, it is necessary to specify several parameters, including 1) the desired number of clusters, 2) the distance metric, 3) the number of sample iterations, and 4) the initial seeding of clusters. We use an approach employed for estimating the number of clusters to use for a given dataset commonly referred to as the elbow method (Kodinariya and Makwana 2013). For this method, a plausible range of k is selected and clustering is performed for each value of k within this range. For a given value of k , the intracluster variance of each cluster is computed, and the sum of these variances over all clusters, known as the summed square of errors (SSE), is obtained. SSE is expected to decrease with increasing k , as increasing the cluster number should lead to reduced intracluster variance. Where the SSE slope as a function of k changes rapidly indicates a point beyond which the addition of clusters does little in which the isolating further representative patterns inherent to the data. Thus, the presence of slope changes provides a heuristic estimate of the upper bound on the appropriate cluster number of a given input dataset.

We use simple Euclidean distance as the metric, with random seeding of the initial clusters/clusters. The algorithm is then run until convergence is reached. Although convergence is not guaranteed, the number of sample iterations should be chosen so that convergence is possible; for our application, we chose 1000 iterations. To assess robustness, we ran both algorithms 50 times with the same inputs but with random initial seeding; the number of days that moved from one cluster to another from run to run was negligible.

In our analysis, we consider a range of clusters from $k = 4$ to $k = 8$. This range of k was guided in part by results found in Haffke and Magnusdottir (2013), who determined qualitatively that eight modes effectively capture the daily OLR variability over a similar domain. We explore the sensitivity of our results to increasing the number of clusters in section 3e. In particular, we are interested in exploring how the clusters we have analyzed for $k = 4$ change as k is increased and new clusters emerge.

The output of k -means clustering is not arranged in any systematic way (i.e., the ordering of clusters is random). However, clusters can be arranged a posteriori using an approach known as linear unidimensional scaling (LUS; Hubert et al. 2002). LUS orders input objects along a single scaling axis by applying a linear least squares minimization procedure to a matrix of distances between every pair of objects, which is referred to as the proximity matrix. In short, objects that are more similar lie closer together along the scaling axis. In what follows, when referring to k -means, we

mean k -means with the application of LUS ordering, unless otherwise stated.

3. Results

a. $k = 4$ clustering of TRMM total rainfall

We begin here by depicting results for a $k = 4$ clustering of TRMM precipitation (Fig. 1), leaving aside for now considerations of what the “proper” number of clusters should be. As we demonstrate, $k = 4$ provides a readily interpretable set of clusters. Note that in Fig. 1, and in subsequent figures, clusters are ordered clockwise from the top left (i.e., the upper-left panel depicts cluster 1 while the lower-left panel depicts cluster 4).

To enable more quantitative comparison among the clusters, we calculate the centroids of the SPCZ regional precipitation and slopes of the principal SPCZ diagonal over the portion of the domain encompassing 160°E – 130°W , 30°S – 0° . The centroids are weighted by both the precipitation rate at a given gridpoint and the cosine of latitude to account for decreasing grid point area moving poleward. Centroid displacements relative to the climatological DJF centroid are computed for each cluster. The SPCZ slopes are estimated via a linear regressive best fit through the points of maximum precipitation at each longitude within the longitude interval specified above. For comparative purposes, we compute the meridional translation of each cluster’s best-fit line (at 170°W) relative to the climatological DJF best fit line, as well as the angle of each cluster’s best-fit line relative to the DJF climatology (Table 1). Further, we qualitatively compare shifts in the 4 mm day^{-1} contours along the northern and southern boundaries of both the cluster’s precipitation and the mean precipitation to gauge how the rainfall distribution has moved.

The centroids of both clusters 1 and 2 exhibit pronounced northward departures from the climatological DJF centroid, with a slight eastward shift in cluster 1 (for a total departure of approximately 464.4 km) and a slight westward shift in cluster 2 (with a total departure of approximately 563.5 km). Similarly, the midpoints of the axis of maximum SPCZ rainfall for clusters 1 and 2 occur to the north of the climatological axis midpoint by 83.8 and 308.9 km, respectively, which underscores the translation of the entire rainband. The principal SPCZ axes for clusters 1 and 2 exhibit steeper meridional tilts relative to the climatological mean (by -3.9° and -5.4° longitude per degree latitude). The 4 mm day^{-1} contours in clusters 1 and 2 also support the overall northward and eastward translation of SPCZ region rainfall compared to the average DJF distribution.

FIG. 1. The k -means + LUS output ($k = 4$) on TRMM DJF 1999–2013 total precipitation (filled contours; mm day^{-1}). Clusters are ordered by LUS clockwise from upper left, with each cluster labeled according to the shift from mean position and SPCZ activity as described in the text. In each panel, the 4 mm day^{-1} isohyet, centroid, and best-fit line through the longitudes of maximum SPCZ rainfall at each latitude are depicted for the cluster (in black) and for the climatological DJF mean (teal). SST anomalies generated from Optimum Interpolation Sea Surface Temperature $0.25^\circ \times 0.25^\circ$ (OISST) (red + blue contours; 0.5° increments; $^\circ\text{C}$) (Banzon et al. 2016).

In both clusters 3 and 4, the cluster centroids lie to the west of the climatological DJF centroid (692.9 km in cluster 3 and 424.7 km in cluster 4), while the mid-points of the best fit lines in clusters 3 and 4 shift south of the climatological mean axis midpoint, by 295.4 and 506.4 km respectively. The SPCZ axis manifests a steeper meridional tilt in cluster 4 (-6.0° longitude per degree latitude relative to the mean), while the axis in cluster 3 becomes nearly zonal (7.9° longitude per degree latitude relative to the mean). The displacements of the 4 mm day^{-1} contours along both the northern and southern margins of the SPCZ region suggest strengthening of the equatorial dry slot, shifting precipitation to the south and west.

We note that within each pair of shifted clusters (1 and 2, and 3 and 4) there exists one cluster with enhanced or “active” SPCZ precipitation, and another cluster with a relatively weak or “quiescent” SPCZ. Under both active and quiescent conditions, the SPCZ is evident as a region of daily precipitation in excess of 4 mm day^{-1} . However, the SPCZ-active clusters have a large core region of rainfall rates of $10\text{--}20 \text{ mm day}^{-1}$, while the quiescent clusters are typically less than 10 mm day^{-1} across the entire SPCZ region. For the rest of this paper, we will refer to clusters based on the overall behavior of SPCZ region rainfall intensity and location; that is, cluster 1 is northeastern-displaced, SPCZ active (hence, NE SA); cluster 2 is northeastern-displaced, SPCZ

TABLE 1. North/south displacement of mean precipitation axis and rotation of axis for total and deseasonalized clusters.

Cluster	Average north/south slope displacement at 190° lon from mean (km)	Angular change of slope at 190° lon
Total precipitation		
NE SA	83.8	-3.9°
NE SQ	308.9	-5.4°
SW SQ	-295.4	7.9°
SW SA	-506.4	-6.0°
Deseasonalized precipitation		
NE SA	321.0	-5.3°
Unshifted SA	-111.1	-3.8°
Unshifted SQ	31.3	5.0°
SW SA	-695.2	2.5°

TABLE 2. Percentage of days for each ENSO phase assigned to the $k = 4$ clusters. Chi-squared test score for cluster assignment: 286.25, critical value for $p = 0.05$ is 12.59.

		$k = 4$ clusters			
		NE SA	NE SQ	SW SQ	SW SA
ENSO phase	Days in cluster	260	323	496	271
	Warm	29.2%	55.6%	8.3%	6.9%
	Neutral	24.8%	36.7%	24.1%	14.4%
	Cold	12.2%	3.3%	55.7%	28.8%

quiescent (NE SQ); cluster 3 is southwestern-displaced, SPCZ quiescent (SW SQ); and cluster 4 is southwestern-displaced, SPCZ active (SW SA).

b. Temporal behavior of clusters and their connection to time scales of variability

1) CLUSTER ASSIGNMENT AND RELATIONSHIP TO SST AND ENSO

In Fig. 1, we composite the anomalous mean sea surface temperatures associated with each cluster, using daily data from the Optimum Interpolation Sea Surface Temperature (OISST) (Banzon et al. 2016) dataset. In these composites, the cluster-mean SSTAs for the SW active/quiescent pair for $k = 4$ show little significant difference (Fig. 1), with cold anomalies occurring along the equatorial central and eastern Pacific. However, for the cluster-mean SSTAs of the NE cluster pair, we note that while both clusters exhibit warm anomalies, their intensity and location are asymmetric: positive SSTAs of 1° – 1.5°C are evident for the NE SQ cluster, and less intense (0.5° – 1°C) warm anomalies are roughly coincident with the SPCZ in the NE SA cluster. In general, however, the SSTA and precipitation distribution in Fig. 1 are suggestive of ENSO.

We explicitly verify the connection to ENSO by computing the percentage of days for each phase of ENSO, as defined by the seasonal-mean oceanic Niño index (ONI), assigned to the $k = 4$ clusters over the entire 15-yr analysis period (Table 2). Together, over 80% of days with $\text{ONI} \geq 0.5^{\circ}\text{C}$ are assigned to the NE SA and NE SQ clusters, and a similar percentage of $\text{ONI} \leq -0.5^{\circ}\text{C}$ days are assigned to the SW SQ and SW SA clusters. Neutral phase days manifest a more uniform distribution across clusters, albeit with a modest preference for the NE SQ cluster. As k is increased, neutral phase conditions are removed from the El Niño or La Niña clusters, as discussed further in section 3e. Considering the distribution of cluster assignments by year provides further evidence of ENSO influence on the clustering (Fig. 2). The daily assignments of DJFs with El Niño conditions are dominated by the NE SA and NE SQ clusters. By contrast, for La Niña years, the

SW SA and SW SQ clusters dominate the daily cluster assignments by percentage, and several cool ONI years (1999, 2000, and 2011) are notable for having no NE SQ days. This time behavior supports the influence of ENSO phase and associated spatial shifts within the TRMM data. Neutral DJFs reflect more uniform daily assignments across the four clusters, suggesting that the precipitation distribution on a given day in a neutral DJF is approximately equally likely to be similar to any of the four clusters.

Returning to the asymmetry seen in the northeastern-shifted clusters, while a northern shift of the SPCZ is expected during warm events (Borlace et al. 2014; Choi et al. 2015), one would expect that the region of intense precipitation would be associated with higher SST anomalies but we see the opposite occurs in the warm SST composites. Of the four warm ONI events between December 1998 and February 2013, only 2010 had a magnitude exceeding 1°C , and the average warm anomaly during this time period is only $+0.9^{\circ}\text{C}$; by contrast, the extremely intense El Niño events, such as

FIG. 2. Percentage distribution of daily cluster assignments by year. The color coding of the year labels reflects the Niño-3.4 SST anomaly for that year's DJF season, with cool, neutral, and warm events indicated in blue, purple, and red, respectively.

FIG. 3. Distribution of the persistence of cluster assignments, represented by the number of consecutive daily assignments to the same cluster (y axis) of interval lengths indicated on the x axis (in days). Note that because 1-day interval lengths are much more common than larger interval lengths, these are not included in order to reduce the scale on the y axis.

1998, have much larger amplitudes. Further, two of these warm events occurred in the central Pacific (2005, 2007); the remaining two (2003, 2010) were eastern Pacific events. These factors could serve to dilute the intensity of the warm SST anomaly composites. However, this does not explain why quiescent northeastern-shifted conditions occur with warmer SSTs. During the warm events (2003, 2005, 2007, and 2010), we note that the ONI warm anomaly typically peaks in December as the ENSO event evolves. Examining the seasonal distribution of days within the NE SQ cluster (not shown), we find a greater incidence of early season days assigned to the NE SQ cluster compared to NE SA. Over the four warm events, the ONI region SSTs for NE SQ days average 0.25° – 0.5° warmer than for NE SA days and, as noted in Fig. 2, three DJFs had no days assigned to NE SQ. Averaged over all years, this yields warmer cluster-mean SSTs for NE SQ compared to NE SA.

2) CLUSTER PERSISTENCE AND TRANSITIONS

For further insight into the temporal behavior of the clusters, now we consider their persistence and transitions. For cluster persistence, we construct the histogram of time intervals of consecutive days occurring in the same cluster (Fig. 3). While there are a few instances of lengthy persistence of 10 days or more within each cluster, the average residence time and standard deviation in days for each of the four clusters are NE SA, 3.5 ± 4.9 days; NE SQ, 3.8 ± 3.4 days; SW SQ, 5.8 ± 5.6 days; and SW SA, 4.6 ± 5.0 days. The subweekly mean cluster residence times point to high-frequency, synoptic-scale forcing acting on the precipitation distribution (Kiladis and Weickmann 1992; Niznik et al. 2015; van der Wiel et al. 2015).

To investigate the temporal connections between clusters, we assess the day-to-day transitions of cluster pairs. Figure 4 provides a summary of the cluster transition statistics, conditioned on ONI phase. During warm DJFs, the day-to-day transitions are dominated by the two NE-shifted clusters, either in the form of “self-transitions” (i.e., no day-to-day changes in cluster assignment) or to the other cluster exhibiting a similar precipitation spatial shift (i.e., NE SA $\rightarrow$ NE SQ or NE SQ $\rightarrow$ NE SA). Analogous behavior is observed for transitions during cold anomalies between the two southwestern-shifted clusters. Figure 4, together with the SST composites and year-to-year variation in cluster assignments, indicates that while ENSO dictates the location of favorable SST conditions for deep convection to occur, a characteristic rainfall regime typically persists for 3–6 days before transitioning, typically in intensity rather than in overall location, and that further controls exist on precipitation intensity within regions of favorable SSTs.

3) CLUSTER ASSIGNMENT AND THE MJO

We suggest that the MJO may account for some of the larger persistence times seen in Fig. 3, given the MJO’s characteristic propagation speed ($3\text{--}5\text{ ms}^{-1}$) and the typical width (several thousand kilometers) of its “envelope” of convection (Waliser et al. 2009). Moreover, the MJO’s propagation across the domain may account for some of the transitions between clusters. To isolate the MJO forcing in the $k = 4$ clustering, we compare the daily cluster assignments to observed MJO phase, via the phase space defined by the daily Wheeler–Hendon RMM index (Kiladis et al. 2014) (Fig. 5). Note that only days for which the amplitude of the RMM index exceeds unity, which is taken to indicate the presence of an MJO event, are plotted. Consecutive days within a given cluster are connected with line segments. Table 3 summarizes the distribution of days per cluster according to MJO phase. These results support an MJO contribution to the $k = 4$ cluster assignments. For example, the NE SA cluster (black symbols) is favored during MJO phases 6, 7, and 8, as 48.1% of NE SA days fall in these phases. Phases 6 and 7 also favor SW SA conditions (blue symbols), as 38.7% of SW SA days fall in these phases. Conversely, NE SQ maps to phase 3 (13.0% of NE SQ days fall in this phase) and SW SQ maps onto phases 3, 4, and 5 (42.1% of SW SQ days fall in these phases). Additionally, a slightly larger percentage (40.2%) of days with no active MJO influence (as defined by $\text{RMM} < 1$) is assigned to NE SQ than to other clusters, for which $\text{RMM} < 1$ occurs between 30.6% and 32.7% of days in these clusters. In RMM space, phases 3–5

FIG. 4. Summary of the day-to-day transitions between clusters. For example, during a warm event, 4.2% of the transitions were from cluster 1 on day 1 to cluster 2 on day 2. The color coding reflects the Niño-3.4 region SST anomaly for that year's DJF season (blue for a cold event, purple for neutral, and red for warm).

generally see reduced rainfall over the central Pacific. The uneven distribution of days in each cluster by phase suggests that the precipitation distribution and intensity depicted in these clusters is influenced in part by the MJO's convective enhancement, and that this forcing is at least partly responsible for daily cluster assignments.

4) DESEASONALIZED CLUSTERING

We have thus far not directly addressed the potential sensitivity of our clustering to the seasonal evolution of SPCZ region rainfall over the course of DJF. However, as noted in the previous discussion of cluster-mean SSTAs, the cluster assignments do exhibit some preferential occurrence within the DJF season. With the "phase-locking" of ENSO (Chen and Jin 2020) to the annual cycle, the seasonal evolution of the daily cluster assignments could affect the interpretation of the results. To explore this further, we repeat the $k = 4$ clustering on deseasonalized daily precipitation in which each day's 15-yr climatological daily is subtracted prior to clustering. The resultant deseasonalized clusters for $k = 4$ are illustrated in Fig. 6. Note that a weighted

climatology corresponding to the days in each cluster has been restored to facilitate comparison with the $k = 4$ clustering of total rainfall in Fig. 1.

Two of the deseasonalized clusters feature rainfall displaced to the northeast and southwest, respectively, of the principal SPCZ axis, and the 4 mm day^{-1} contours follow these shifts in the axis, consistent with El Niño and La Niña phase forcing. We term these clusters NE SA and SW SA. The rainfall in the remaining two clusters is largely unshifted, with one of the clusters featuring a more intense SPCZ and the other a less intense one, but with axes and 4 mm day^{-1} contours that are essentially collocated with the climatological SPCZ. We term these unshifted SPCZ active (U SA) and unshifted SPCZ quiescent (U SQ), respectively.

More pronounced shifts occur in deseasonalized clusters relative to the analogous clusters in Fig. 1: in particular, for the northeastern- and southwestern-shifted deseasonalized clusters, we find respective axis shifts of 321.0 and 695.2 km, respectively. Further, the northern and southern margins of the deseasonalized NE SA cluster 4 mm day^{-1} contour extend farther north and east, compared to the total, whereas in the deseasonalized SW SA cluster the

FIG. 5. Daily $k = 4$ cluster assignments and MJO phase. Values plotted correspond to RMM1 (x axis) vs RMM2 (y axis), color coded by cluster assignment. Note that only values with an amplitude greater than 1, indicating the presence of an MJO, are shown. The geographic sectors corresponding to the eight conventional MJO phases are highlighted.

opposite holds, and the northern and southern edges of the 4 mm day^{-1} contour extend farther to the southwest than in the total case. In the deseasonalized NE SA cluster, the slope becomes slightly more meridional with a difference of -2.9° longitude per degree latitude relative to the mean, while the deseasonalized SW SA cluster's slope becomes more zonal with a difference of 6.3° longitude per degree latitude relative to the mean.

Exploring the sensitivity of cluster assignment and ONI via linear regression, we find that NE SA is approximately 3 times as sensitive to changes in SST anomaly than U SA and SW SA. This means, for example, that for a DJF El Niño event with an $\text{ONI} = 1.0^\circ\text{C}$, $\sim 2/3$ of the days in that DJF most strongly resemble the precipitation in NE SA, while the remaining days more strongly resemble U SA and U SQ at 15%–20% each. Conversely, for a La Niña of equal magnitude, only $\sim 25\%$ – 33% of the days in that season resemble a precipitation distribution like that evident in SW SA, with another 25%–33% like in U SA, and the rest, approximately 30%–50%, in U SQ. Together, the behavior of the deseasonalized clusters indicates that El Niño shifts the rainfall in the SPCZ to the north and east of its mean position and increases its intensity, whereas La Niña shifts the SPCZ rainfall south and west and decreases its intensity. This difference in SPCZ intensity between El Niño and La Niña was also pointed out by Lintner and Boos (2019).

TABLE 3. Percentage of days assigned to each cluster ($k = 4$) by MJO phase. Chi-squared test score for cluster assignment: 276.07, critical value for $p = 0.05$ is 32.67.

		$k = 4$ clusters			
		NE SA	NE SQ	SW SQ	SW SA
MJO phase	Days in cluster	260	323	496	271
	1	6.5%	7.7%	2.8%	2.2%
	2	5.4%	7.4%	6.0%	5.2%
	3	3.5%	13.0%	8.9%	5.2%
	4	2.3%	4.3%	15.5%	3.7%
	5	1.5%	4.3%	17.7%	7.4%
	6	11.9%	6.2%	7.9%	21.0%
	7	23.1%	9.0%	4.0%	17.7%
	8	13.1%	7.7%	3.8%	7.0%
No active MJO		32.7%	40.2%	31.0%	30.6%

Similar to our analysis of the total precipitation, we quantify the cluster assignment of the deseasonalized data in the context of its relationship to ENSO via seasonal-mean ONI. The NE SA assignments show a positive correlation with $\text{ONI} \geq 0.5^\circ\text{C}$: 73% of days in that cluster occur with an ONI above this threshold, with higher ONI values associated with a greater daily incidence of NE SA assignment. Conversely, only 5 days with seasonal mean $\text{ONI} \leq -0.5^\circ\text{C}$ are assigned to this cluster. By contrast, SW SA occurrences are greatest with ONI indices less than or equal to -0.5° , with 81% of days at or below this threshold in the cluster. Only 9 SW SA days have an $\text{ONI} \geq 0.5^\circ\text{C}$. This suggests that the shifts are strongly dictated by the phase of ENSO, something that was not as clear in the total precipitation case.

Like SW SA above, the largest fraction of days assigned to U SA corresponds to anomalously cool ONI conditions, with 67% of days occurring with an $\text{ONI} \leq -0.5^\circ\text{C}$. U SQ contains a mix of days above and below the ENSO event onset thresholds, with 20% of the days occurring with $\text{ONI} \geq 0.5^\circ\text{C}$, and 58% occurring with $\text{ONI} \leq -0.5^\circ\text{C}$. The centroids of U SA and U SQ shift ~ 370 and 320 km , respectively, with respect to the climatological DJF precipitation centroid, less than any of the four total precipitation clusters, or the other two deseasonalized clusters. This suggests an asymmetrical response by the SPCZ to SST anomalies in terms of both the axis of most intense rainfall, as well as the intensity of the rainfall itself.

c. Composite analysis

While ENSO phase clearly dictates the location of the SPCZ diagonal, it does not account for the occurrence of an active or quiescent SPCZ. To investigate the dynamic and thermodynamic signatures of an active versus quiescent SPCZ, we construct composites of CSFR circulation and moisture fields for the days assigned to each of the $k = 4$ clusters (for total rainfall). Figures 7 and 8

FIG. 6. As in Fig. 1, but using deseasonalized TRMM data.

depict composites of anomalous precipitation, zonal and meridional wind, and specific humidity at 925 and 500 mb (1 mb = 1 hPa), respectively. Here, anomalies are defined as differences between field averages for all days within a given cluster relative to daily DJF

climatologies. Statistical significance of composite anomalies is assessed via a Welch's t test.

At both pressure levels, the NE SA cluster exhibits pronounced westerly wind anomalies to the south of the equator across the western and central Pacific.

FIG. 7. Longitude-latitude composites by the $k = 4$ clusters of 925-mb wind anomaly (barbs; m s^{-1} ; referenced to the 5.0 m s^{-1} arrow at the bottom of each panel), specific humidity anomaly (contours; g kg^{-1}), and precipitation anomaly (filled contours; mm day^{-1}).

FIG. 8. As in Fig. 7, but for winds and specific humidity at 500 mb.

These westerly wind anomalies are accompanied by converging meridional flow at approximately 5°S; conceptually, it is consistent with weakening of the Walker circulation experienced during El Niño conditions. Positive specific humidity anomalies at both 925 and 500 mb align closely with the region of anomalous precipitation but also extend to the east where precipitation anomalies are small; here, the climatological mean SSTs are cooler than those to the west, and less conducive to supporting precipitating deep convection. During El Niño, the warming of SSTs increases low-level specific humidity through enhanced evaporation (Neelin and Held 1987). The NE SQ cluster manifests a similar, albeit weaker, anomalous zonal wind pattern compared to the NE SA cluster. Further, as is evident at both pressure levels, NE SQ exhibits a zonal specific humidity anomaly primarily concentrated along the ITCZ. At 500 mb, the region of (weak) SPCZ rainfall is separated from the ITCZ by anomalously dry air.

The SW clusters manifest strengthening of easterly winds, consistent with an enhanced Walker circulation present during La Niña. The signature of the strengthened Walker circulation is further evident in the spatial distribution of anomalous moisture, with positive specific humidity anomalies largely confined to the extreme western Pacific, and anomalously dry air predominating across much of the central and eastern Pacific. A pronounced anomalous moisture gradient is evident in both

SW clusters, and at both pressure levels examined, the axis of which is seen to shift slightly to the west in SW SQ, likely associated with the stronger easterly anomalous winds characteristic of this cluster.

An especially prominent feature in both the SPCZ active clusters is the anomalous low-level poleward flow and enhanced moisture along the axis of the SPCZ. Anomalous moist poleward flow with enhanced precipitating deep convection has been noted for other off-equatorial convection zones, such as the mei-yu-baiu front in the northwestern Pacific (Sampe and Xie 2010), as well as in the Caribbean springtime rainband (Allen and Mapes 2017).

The anomalous dipole of precipitation oriented along the northwest to southeast axis of the SPCZ is also consistent with composited precipitation based on the geopotential height clusters in Lorrey and Fauchereau (2017), where similar diagonally oriented positive and negative precipitation anomalies are shown to set up on either side of an axis running from northwest to southeast in the SPCZ region. In our study, positive moisture anomalies lie to the north and east of this axis and dry anomalies to the south and west during warm events, and vice versa during cold events. Also similar to Lorrey and Fauchereau (2017), we see the presence of a cyclonic low-level circulation in the area of the positive rainfall anomalies, with poleward flow along the axis of heaviest precipitation ahead of this cyclonic feature. Since the region of warmest SSTs fluctuates with ENSO,

to the extent that analogous dynamical triggers are responsible for the SPCZ's observed intensity fluctuations, the region in which this anomalous poleward flow, high specific humidity values, and associated rainfall occur is expected to shift accordingly. Our composites indicate similar relationships among winds, moisture, and precipitation, suggesting a similar mechanistic connection.

We also computed time-lagged upper-level vorticity composites in order to isolate signatures of transient, propagating upper-level waves originating in midlatitudes as triggers for an active SPCZ, following [van der Wiel et al. \(2015\)](#). These composites (not shown) do not show clear indications of propagating upper-level forcing in the cluster assignments or the transitions between clusters, although our sample size is smaller than in [van der Wiel et al. \(2015\)](#). We revisit the potential role of upper-level forcing mechanisms in the EOF analysis presented in [section 3d](#).

In addition to compositing variables over longitude and latitude, we examine vertical cross sections along transects sampled across the SPCZ ([Fig. 9](#)). We select a southwest to northeast transect that passes through the region of most intense climatological DJF precipitation in the SPCZ (line A–B in the top panel of [Fig. 9](#)), as well as the axes of heaviest precipitation in both the NE SA and SW SA clusters. Overall, these transects depict dipoles of specific humidity anomalies that are antisymmetric with ENSO phase, as in the longitude/latitude view. Positive specific humidity anomalies, extending vertically to between 600 and 400 mb in the SQ clusters, and to between 300 and 200 mb in the SA clusters, align with the heaviest precipitation along the transect (shown by the black curves under each cross section).

What is somewhat unexpected, however, are the differences in anomalous zonal winds across the four clusters. While the moisture fields for the northeast- and southwest-shifted clusters are largely antisymmetric, the zonal wind fields are not. Considering the active clusters, NE SA is largely dominated by anomalous westerlies in the region of greatest precipitation and specific humidity through most of the troposphere, with easterly flow above 200 mb. By contrast, for SW SA, we again see the presence of anomalous westerly flow in the region of greatest precipitation and positive anomalous specific humidity, but there are near-surface easterlies over the dry, weaker precipitating region. Thus, while the dynamics appear distinct in each panel, similarly intense rainfall maxima, significantly higher than the maxima along this transect in either of the SQ panels, are observed.

Comparing the anomalous zonal wind distribution in the SA and SQ clusters, antisymmetry is also apparent, although perhaps not in the sense expected. The analog to NE SA's vertical wind structure appears to be that of

SW SQ, which differs in both moisture distribution and precipitation intensity. However, instead of strong, anomalous westerlies dominating the transect here, we see strong easterlies in the same region, although in conjunction with a negative moisture anomaly. Similarly, but to a lesser degree, comparing the NE SQ composite to the SW SA composite, we see a dipolar wind distribution, although reversed because of the influence of ENSO. These differences underscore some dynamic asymmetries between ENSO phases as well as between SPCZ active and quiescent conditions.

These transects highlight differences in the dynamics and thermodynamics present in the SPCZ. We have demonstrated that SPCZ location is largely determined by ENSO phase and its control on where the warmest SSTs occur. This ENSO influence is also apparent in the transects, for which we see expected shifts or reversals of the zonal winds associated with ENSO. However, we note asymmetries present in the upper versus lower troposphere controls across clusters including the intrusion of dry, near-surface air in the SW SA cluster not present in the NE SA cluster, despite both clusters having similar moisture structures and precipitation intensity (albeit in different locations because of ENSO). Understanding the transport of moisture into and throughout the SPCZ is critical to improving the overall understanding of the SPCZ's behavior and is something that will be pursued in future work.

d. EOF analysis

To complement the cluster analysis, we further apply EOF analysis on both the total and deseasonalized precipitation. [Figure 10](#) depicts the spatial patterns of the first two leading modes of each of total and deseasonalized precipitation. For the total precipitation, modes 1 and 2 of account for $\sim 4.7\%$ and $\sim 3.5\%$ of the total variance, respectively.

Mode 1 exhibits a spatial pattern with a dipolar structure in the SPCZ, with a nodal line roughly parallel to the principal diagonal of most intense SPCZ rainfall, while mode 2 is dominated by anomalies of one sign along the SPCZ diagonal. The rainfall distribution of mode 1 is consistent with the $k = 4$ active SPCZ clusters, with either northeastward or southwestward displacements occurring during El Niño and La Niña phases, respectively. However, there is no clear analog in the clusters of total rainfall for the mode 2 behavior in which an essentially unshifted SPCZ is either enhanced or suppressed. Meanwhile modes 3 and 4 (not shown) are characterized by a wavelike structure along the poleward margin of the SPCZ. That the spatial patterns of modes 3 and 4 are in quadrature (i.e., they have similar explained variance) further suggests propagating behavior.

SPCZ vertical transects (S.Hum anomaly, total wind anomaly, precip distribution)

FIG. 9. Vertical composites by the $k = 4$ clusters along the cross-SPCZ axis between points A and B depicted in the top panel. Shown here are anomalous specific humidity (filled contours; g kg^{-1}), and zonal wind (contours; m s^{-1}). Below each transect, the precipitation along the transect is shown for both the cluster (black) and climatological DJF mean (teal).

Total and Deseasonalized EOF 1 and EOF 2

FIG. 10. (left) EOF1 and (right) EOF2 on (top) total and (bottom) deseasonalized precipitation. The slope, centroid, and 4 mm day^{-1} contour of mean total precipitation are plotted for reference.

Previously, [van der Wiel et al. \(2015\)](#) had demonstrated that upper-level midlatitude wave activity entering the SPCZ region can excite SPCZ region deep convection, although it is unclear whether the propagating behavior suggested by modes 3 and 4 is related to the mechanism of [van der Wiel et al. \(2015\)](#).

For the deseasonalized precipitation, the two leading modes account for similar amounts of variance as the leading modes for the total precipitation (4.9% and 3.4%). Comparing the deseasonalized EOF1 and EOF2 to the deseasonalized clusters, EOF1 captures the northeast or southwest-shifted precipitation displacement, much like that seen in the NE SA and SW SA clusters. Deseasonalized EOF2 strongly resembles EOF2 for total rainfall, as well as the precipitation pattern of the deseasonalized U SA cluster, with a region of enhanced rainfall along the climatological SPCZ. Overall, differences between the total and deseasonalized EOFs are minor; the largest differences occur near the equator in mode 2, where the deseasonalized spatial pattern shows a reduction in amplitude along the ITCZ. This is likely a product of the gradual northward seasonal displacement of the ITCZ over the course of austral summer, which is not present in the deseasonalized data.

We briefly remark on the resemblance of the spatial structures of the EOFs to those found by [Matthews \(2012\)](#) using the the daily mean OLR dataset of

[Liebmann and Smith \(1996\)](#) as a proxy for deep convection. Our leading modes explain less overall variance (by a factor of 2) compared to the percentages found by [Matthews \(2012\)](#). The larger explained variances for OLR relative to precipitation are likely the result of the former being spatially smoother than the latter; we note, too, that the OLR data analyzed by [Matthews \(2012\)](#) exceed the length of TRMM data analyzed here by a factor of 2.

[Figure 11](#) presents a scatterplot of the principal components of modes 1 and 2 of both total and deseasonalized precipitation, with values color-coded according to daily cluster assignments for $k = 4$. The centroids of each cluster have been plotted for reference. Looking first at the total precipitation plot (top), although this scatterplot indicates some separation of clusters, the axes of variation represented by PC1 and PC2 do not distinguish the clusters in a straightforward way. Significant mixing occurs along the edges of each cluster, and NE SA (black) and SW SA (blue) appear quite diffuse. By contrast, the scatterplot of the first two deseasonalized PCs (bottom) colored by their $k = 4$ cluster assignments shows more distinct separation in two dimensions than is evident for total precipitation. Despite this, the lines that best separate the clusters are not perpendicular, suggesting that the clusters reflect behavior that is not captured by the linear and orthogonal EOF modes.

FIG. 11. Scatterplot of days in (top) total and (bottom) deseasonalized EOFs according to PC1 (x axis) and PC2 (y axis) values, color-coded by cluster assignment. Centroids for each cluster are indicated by a star.

e. Sensitivity of clustering results to varying k

Our analysis has so far considered $k = 4$, which, given our focus on the SPCZ, has yielded a readily interpretable set of rainfall patterns in terms of the location and intensity of SPCZ rainfall. Of course, there is no reason, *a priori*, for selecting four clusters. Thus, we now present results for increasing k , considering both the stability of the clusters we have thus far identified and the emergence of new ones. The SSE (section 2b) is illustrated in Fig. 12 for the total rainfall for the range of k from 1 to 10. Ideally, the SSE would exhibit a single elbow that we could interpret as an upper bound on cluster number; however, we note from Fig. 12 that there are multiple elbows associated with changes in SSE slope at $k = 4, 6$, and 8 . In what follows, we consider $k = 8$ as an upper bound, even though a smaller k is certainly justifiable based on this heuristic approach.

To visualize the sensitivity to increasing k , we compare the root-mean-square (RMS) error of the clusters for a value of k to those for $k + 1$. Results of this cluster tracking are depicted in Fig. 13. Here, cluster pairs from k and $k + 1$ with the smallest RMS error are identified by matching colored boxes, while the cluster in $k + 1$

FIG. 12. Sum of squared errors (SSE; y axis, normalized as a percentage relative to $k = 1$) across the range of cluster numbers spanning from $k = 1$ to 10.

without an associated cluster in k is placed in a unique colored box. The emergent clusters in $k + 1$ tend to be located (according to LUS ordering) adjacent to or between those clusters in k that show the largest decrease in days assigned. In general, when increasing from k to $k + 1$ clusters, each of the clusters in k “lose” data to their analogs in $k + 1$, as outliers within the k clusters appear in an emergent cluster. This can be seen in the comparison of $k = 4$ to $k = 5$, for which the number of days in the cluster at position 4 (271) decreases by almost 50% for the cluster at position 9 (151). Similar behavior is evident for the other k to $k + 1$ transitions, apart from $k = 6$ to $k = 7$. For this transition, a cluster resembling the $k = 6$ cluster at position 14 occurs at position 17 with $k = 7$, while in the “expected” position (position 21) peak precipitation is centered over the monsoon region of northern Australia and adjacent seas and over the western Pacific warm pool. The overall precipitation distribution in this cluster resembles that of La Niña-like conditions, with an extensive region of low precipitation south of the ITCZ relative to the climatological DJF mean. This is not entirely unexpected, as the warmest SSTs tend to occur over the warm pool, creating a region favorable for deep convection there. Returning to the cluster in position 14 in $k = 6$, its analog for $k = 7$, cluster 17, appears at the opposite end of the LUS ordering because the left side of the LUS ordering for $k = 7$ depicts an unshifted, neutral SPCZ precipitation distribution, and the clusters on the right of the ordering, 21 and 22, depict strongly western Pacific, warm-pool centered patterns.

For $k = 8$, there are four clusters with enhanced SPCZ rainfall (23, 24, 25, and 30) and four with weakened rainfall in the mean SPCZ region (26–29). Clusters 23 and 30 are most similar in their intensity and rainfall

FIG. 13. Relationships among clusters for varying cluster numbers between 4 and 8. Results from $k = 4$ through $k = 8$, arranged by LUS. The upper-right values indicate the number of days assigned to each cluster (top value) and LUS score (bottom value). The LUS score is computed with all 30 clusters included, and the upper-left number is used to designate the position of the cluster. Like-colored boxes around clusters in k and $k + 1$ indicate the lowest RMS score (most analogous cluster), while uniquely colored boxes in $k + 1$ indicate clusters without a smallest RMS analog in k .

distribution to the NE SA and SW SA clusters for $k = 4$. The remaining two active clusters potentially point to dynamics distinct from the SPCZ shifts in NE SA and SW SA. Cluster 25's centroid is displaced south of the DJF climatological mean, although looking at previous analogs of cluster 25 in $k = 5$ through $k = 7$ (clusters 8, 14, and 17 in purple) where the centroid shift is even less pronounced, the precipitation distribution very closely resembles that of the climatological mean or ENSO neutral conditions. Cluster 24 shows a more easterly shifted active SPCZ region, with a rainfall maximum that is separated spatially from the deep tropics.

The remaining clusters in $k = 8$, clusters 26–29, depict weakened SPCZ rainfall, although some have enhanced rainfall in other regions of the study domain. Cluster 27 has generally weak rainfall across the domain, while cluster 26 has an enhanced central/eastern Pacific ITCZ. Meanwhile, clusters 28 and 29 exhibit the most intense rainfall over the western Pacific, with the former showing the greatest intensity north of the equator near the Philippines and the latter showing the greatest intensity south of the equator near the northern coast of Australia. These clusters may reflect rainfall behavior associated with the Australian monsoon wet season as mentioned in the discussion of cluster 21.

Our exploration of increasing cluster number across a plausible range of k is useful from what it indicates both about how k -means clustering partitions the TRMM rainfall and also about additional forcing mechanisms or processes impacting regional precipitation distribution

and intensity that are not evident at lower values of k such as clusters in $k = 8$ that appear to be dynamically distinct from the clusters generated by lower values of k . Furthermore, it reaffirms the complex dynamics present in the equatorial Pacific/SPCZ region, and the potential for interactions among mechanisms or processes operating on distinct time scales to shape rainfall in the region.

4. Summary and conclusions

In this study, we apply clustering methods to TRMM rainfall data over the SPCZ region to identify a small set of representative spatial patterns. For k -means clustering with four clusters, the rainfall is found to be partitioned according to the position of SPCZ rainfall as well as its intensity, with position largely following from ENSO phase and thus largely interannual in nature. Examination of the temporal persistence and transitions between clusters points to modulation of spatial distribution and intensity of SPCZ region precipitation on time scales ranging from synoptic to intraseasonal.

Compositing CFSR reanalysis wind and moisture fields by cluster assignment gives some mechanistic insight into the rainfall patterns. Active SPCZ conditions during either ENSO phase are associated with pronounced poleward flow at low levels along the SPCZ, with moistening throughout the lower troposphere. Allen and Mapes (2017) have connected low-level poleward flow over the Atlantic with intense precipitation events, which they suggest are triggered by

upper air troughs advecting plumes of high precipitable water out of the deep tropics. These plumes acquire a zonal tilt because of increasing westerlies with latitude. The occurrence of quasi-random deep convection within these bands accounts for high precipitation totals compared to surrounding areas. We posit that a similar mechanism operates in the SPCZ.

Distinct asymmetries with regard to moisture and wind vertical structure are found, despite similar precipitation intensities. For specific humidity, the NE SA cluster exhibits considerably larger positive lower to midtroposphere anomalies coincident with enhanced rainfall relative to the SW SA cluster. Anomalous westerlies coincide with the positive moisture anomalies for both the NE SA and SW SA clusters; on the other hand, the SW SA composite is marked by much stronger easterlies in the region of reduced rainfall (and moisture) relative to the NE SA composite. In fact, vertical wind and moisture fields for the NE SA and SW SQ cluster pairs strongly resemble one another (with signs reversed), as do the fields for the NE SQ and SW SA cluster pairs.

The $k = 4$ clustering results underscore the importance of interactions occurring across different temporal scales. For example, the year-to-year variability associated with ENSO and the attendant changes in SST determine the geographic region in which the SPCZ is likely to occur, while higher-frequency variability determines whether rainfall is actually realized. While the influence of ENSO on the SPCZ is now fairly well understood, the influences of higher-frequency forcings including the MJO and synoptic time scale forcings are less clear. By clustering on rainfall after removing the influence of the mean seasonal evolution of SPCZ regional rainfall, we find a somewhat different behavior evident in the clustering. In particular, the deseasonalized clustering yields a pair of clusters with enhanced SPCZ rainfall displaced to the northeast or southwest (NE SA and SW SA), while the other pair effectively maintains the climatological (i.e., unshifted) position of the SPCZ diagonal but with either enhanced or suppressed rainfall (U SA and U SQ, respectively). Examining the year-to-year behavior of the deseasonalized clusters shows that El Niño shifts the SPCZ to the northeast and enhances rainfall, while La Niña shifts it to southwest and decreases rainfall.

As a complement to the cluster-based analyses, we calculate EOFs of both the total and deseasonalized precipitation. Mode 1 depicts enhanced precipitation shifted either northeast or southwest of the climatological mean: a signal also supported in the total and deseasonalized cluster analysis output. However, mode 2, which depicts an unshifted, active, or quiescent region

of precipitation, is more consistent with deseasonalized clusters U SA and U SQ. These differences further highlight the influence of the seasonal cycle on the SPCZ, as well as the interplay among forcings on different time scales. Although we can analyze the how variability on distinct time scales modulates the location and intensity of rainfall, it is also important to consider how interactions may occur among these different time scales of influence.

We further performed clustering over the range from $k = 4$ to $k = 8$ to examine both the robustness of the $k = 4$ patterns we have emphasized as well as the emergence of other features when the degrees of freedom in the clustering are increased. The general SPCZ region behavior of the clusters for $k = 4$ is preserved across the studied range of increasing k , although we find patterns corresponding to rainfall centers located farther eastward and extending more poleward than in the $k = 4$ case. Additional regional centers of behavior also appear at larger k , such as those related to strengthening or weakening of precipitation in the Australian monsoon region.

As we noted in the introduction, part of the motivation for the present study is our interest in applying clustering-based techniques as an approach for evaluating climate models. Significant model errors, biases, and spread are evident in simulated SPCZ region rainfall, including the well-known double ITCZ bias (Lin 2007; Bellucci et al. 2010; Brown et al. 2013) and an SPCZ that is too zonal (Brown et al. 2013). In future work, we plan to use both the cluster spatial patterns and their frequencies of occurrence as targets for model evaluation. We expect that this will provide constructive insights into the nature of SPCZ region simulation deficiencies. For example, we can explore whether a model featuring a too zonal climatological mean SPCZ shows this as a common feature across all clusters or as a problematic aspect of a subset of clusters. A further consideration is that the optimal number of clusters may vary from model to model, which we suggest may be used as a further diagnostic of model evaluation.

Acknowledgments. The authors thank Anthony Broccoli, Paul Loikith, and James Miller for useful discussions about the analysis and results and Bryan Raney for computational assistance and support. The authors acknowledge funding support from National Science Foundation EAGER Grant AGS-1842543.

REFERENCES

- Abbas, O. A., 2008: Comparisons between data clustering algorithms. *Int. Arab J. Inf. Technol.*, **5**, 320–325.

- Adler, R. F., and Coauthors, 2003: The version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167, [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).
- Allen, T., and B. E. Mapes, 2017: The late spring Caribbean rain-belt: Climatology and dynamics. *Int. J. Climatol.*, **37**, 4981–4993, <https://doi.org/10.1002/joc.5136>.
- Banzon, V., T. M. Smith, T. M. Chin, C. Liu, and W. Hankins, 2016: A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies. *Earth Syst. Sci. Data*, **8**, 165–176, <https://doi.org/10.5194/essd-8-165-2016>.
- Bellucci, A., S. Gualdi, and A. Navarra, 2010: The double-ITCZ syndrome in coupled general circulation models: The role of large-scale vertical circulation regimes. *J. Climate*, **23**, 1127–1145, <https://doi.org/10.1175/2009JCLI3002.1>.
- Borlace, S., A. Santoso, W. Cai, and M. Collins, 2014: Extreme swings of the South Pacific convergence zone and the different types of El Niño events. *Geophys. Res. Lett.*, **41**, 4695–4703, <https://doi.org/10.1002/2014GL060551>.
- Brown, J. N., and Coauthors, 2013: Implications of CMIP3 model biases and uncertainties for climate projections in the western tropical Pacific. *Climatic Change*, **119**, 147–161, <https://doi.org/10.1007/s10584-012-0603-5>.
- Capotondi, A., and Coauthors, 2015: Understanding ENSO diversity. *Bull. Amer. Meteor. Soc.*, **96**, 921–938, <https://doi.org/10.1175/BAMS-D-13-00117.1>.
- Chen, H., and F. Jin, 2020: Fundamental behavior of ENSO phase-locking. *J. Climate*, **33**, 1953–1968, <https://doi.org/10.1175/JCLI-D-19-0264.1>.
- Choi, K.-Y., G. A. Vecchi, and A. T. Wittenberg, 2015: Nonlinear zonal wind response to ENSO in the CMIP5 models: Roles of the zonal and meridional shift of the ITCZ/SPCZ and the simulated climatological precipitation. *J. Climate*, **28**, 8556–8573, <https://doi.org/10.1175/JCLI-D-15-0211.1>.
- Folland, C. K., J. A. Renwick, M. J. Salinger, and A. B. Mullan, 2002: Relative influences of the Interdecadal Pacific Oscillation and ENSO on the South Pacific convergence zone. *Geophys. Res. Lett.*, **29**, 1643, <https://doi.org/10.1029/2001GL014201>.
- Haffke, C., and G. Magnusdottir, 2013: The South Pacific Convergence Zone in three decades of satellite images. *J. Geophys. Res. Atmos.*, **118**, 10 839–10 849, <https://doi.org/10.1002/jgrd.50838>.
- Hubert, L. J., P. Arabie, and J. J. Meulman, 2002: Linear unidimensional scaling in the L2-norm: Basic optimization methods using MATLAB. *J. Classif.*, **19**, 303–328, <https://doi.org/10.1007/s00357-001-0047-5>.
- Kiladis, G. N., and K. M. Weickmann, 1992: Extratropical forcing of tropical Pacific convection during northern winter. *Mon. Wea. Rev.*, **120**, 1924–1938, [https://doi.org/10.1175/1520-0493\(1992\)120<1924:EFOTPC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<1924:EFOTPC>2.0.CO;2).
- , H. von Storch, and H. van Loon, 1989: Origin of the South Pacific convergence zone. *J. Climate*, **2**, 1185–1195, [https://doi.org/10.1175/1520-0442\(1989\)002<1185:OOTSPC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1989)002<1185:OOTSPC>2.0.CO;2).
- , J. Dias, K. H. Straub, M. C. Wheeler, S. N. Tulich, K. Kikuchi, K. M. Weickmann, and M. J. Ventrice, 2014: A comparison of OLR and circulation based indices for tracking the MJO. *Mon. Wea. Rev.*, **142**, 1697–1715, <https://doi.org/10.1175/MWR-D-13-00301.1>.
- Kodinariya, T. M., and P. R. Makwana, 2013: Review on determining number of cluster in K-means clustering. *Int. J. Adv. Res. Comput. Sci. Manage. Stud.*, **1**, 90–95, <http://www.ijarcsms.com/docs/paper/volume1/issue6/V1I6-0015.pdf>.
- Liebmann, B., and C. A. Smith, 1996: Description of a complete (interpolated) OLR dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277.
- Lin, J.-L., 2007: The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean–atmosphere feedback analysis. *J. Climate*, **20**, 4497–4525, <https://doi.org/10.1175/JCLI4272.1>.
- Linsley, B. K., P. Zhang, A. Kaplan, S. S. Howe, and G. M. Wellington, 2008: Interdecadal–decadal climate variability from multicoral oxygen isotope records in the South Pacific convergence zone region since 1650 A.D. *Paleoceanography*, **23**, PA2219, <https://doi.org/10.1029/2007PA001539>.
- Lintner, B. R., and J. D. Neelin, 2008: Eastern margin variability of the South Pacific convergence zone. *Geophys. Res. Lett.*, **35**, L16701, <https://doi.org/10.1029/2008GL034298>.
- , and W. R. Boos, 2019: Using atmospheric energy transport to quantitatively constrain South Pacific convergence zone shifts during ENSO. *J. Climate*, **32**, 1839–1855, <https://doi.org/10.1175/JCLI-D-18-0151.1>.
- Lorrey, A. M., and N. C. Fauchereau, 2017: Southwest Pacific atmospheric weather regimes: Linkages to ENSO and extratropical teleconnections. *Int. J. Climatol.*, **38**, 1893–1909, <https://doi.org/10.1002/joc.5304>.
- Matthews, A. J., 2012: A multiscale framework for the origin and variability of the South Pacific Convergence Zone. *Quart. J. Roy. Meteor. Soc.*, **138**, 1165–1178, <https://doi.org/10.1002/qj.1870>.
- , B. J. Hoskins, J. M. Slingo, and M. Blackburn, 1996: Development of convection along the SPCZ within a Madden–Julian oscillation. *Quart. J. Roy. Meteor. Soc.*, **122**, 669–688, <https://doi.org/10.1002/qj.49712253106>.
- Meehl, G. A., 1987: The annual cycle and interannual variability in the tropical Pacific and Indian Ocean regions. *Mon. Wea. Rev.*, **115**, 486–495, [https://doi.org/10.1175/1520-0493\(1987\)115<0027:TACAIV>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0027:TACAIV>2.0.CO;2).
- NCAR Staff, Eds., 2007: The Climate Data Guide: Climate Forecast System Reanalysis (CFSR). Accessed 8 November 2017, <https://climatedataguide.ucar.edu/climate-data/climate-forecast-system-reanalysis-cfsr>.
- Neelin, J. D., and I. M. Held, 1987: Modeling tropical convergence based on the moist static energy budget. *Mon. Wea. Rev.*, **115**, 3–12, [https://doi.org/10.1175/1520-0493\(1987\)115<0003:MTCBOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0003:MTCBOT>2.0.CO;2).
- Niznik, M. J., B. R. Lintner, A. J. Matthews, and M. J. Widlansky, 2015: The role of tropical–extratropical interaction and synoptic variability in maintaining the South Pacific convergence zone in CMIP5 models. *J. Climate*, **28**, 3353–3374, <https://doi.org/10.1175/JCLI-D-14-00527.1>.
- Power, S., 2011: Understanding the South Pacific convergence zone and its impacts. *Eos, Trans. Amer. Geophys. Union*, **92**, 55, <https://doi.org/10.1029/2011EO070006>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- Salinger, M. J., J. A. Renwick, and A. B. Mullan, 2001: Interdecadal Pacific Oscillation and South Pacific climate. *Int. J. Climatol.*, **21**, 1705–1721, <https://doi.org/10.1002/joc.691>.
- Sampe, T., and S.-P. Xie, 2010: Large-scale dynamics of the meiyu-baiu rain band: Environmental forcing by the westerly jet. *J. Climate*, **23**, 113–134, <https://doi.org/10.1175/2009JCLI3128.1>.
- Takahashi, K., and D. S. Battisti, 2007: Processes controlling the mean tropical Pacific precipitation pattern. Part II: The SPCZ

- and the southeast Pacific dry zone. *J. Climate*, **20**, 5696–5706, <https://doi.org/10.1175/2007JCLI1656.1>.
- Trenberth, K. E., 1976: Spatial and temporal variations of the Southern Oscillation. *Quart. J. Roy. Meteor. Soc.*, **102**, 639–653, <https://doi.org/10.1002/qj.49710243310>.
- van der Wiel, K., A. J. Matthews, D. P. Stevens, and M. M. Joshi, 2015: A dynamical framework for the origin of the diagonal South Pacific and South Atlantic convergence zones. *Quart. J. Roy. Meteor. Soc.*, **141**, 1997–2010, <https://doi.org/10.1002/qj.2508>.
- , —, M. M. Joshi, and D. P. Stevens, 2016: Why the South Pacific convergence zone is diagonal. *Climate Dyn.*, **46**, 1683–1698, <https://doi.org/10.1007/s00382-015-2668-0>.
- Vincent, D. G., 1994: The South Pacific Convergence Zone (SPCZ): A review. *Mon. Wea. Rev.*, **122**, 1949–1970, [https://doi.org/10.1175/1520-0493\(1994\)122<1949:TSPCZA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<1949:TSPCZA>2.0.CO;2).
- Vincent, E. M., M. Lengaigne, C. E. Menkes, N. C. Jourdain, P. Marchesiello, and G. Madec, 2011: Interannual variability of the South Pacific Convergence Zone and implications for tropical cyclone genesis. *Climate Dyn.*, **36**, 1881–1896, <https://doi.org/10.1007/s00382-009-0716-3>.
- Waliser, D., and Coauthors, 2009: MJO simulation diagnostics. *J. Climate*, **22**, 3006–3030, <https://doi.org/10.1175/2008JCLI2731.1>.
- Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2).
- Widlansky, M. J., P. J. Webster, and C. D. Hoyos, 2011: On the location and orientation of the South Pacific convergence zone. *Climate Dyn.*, **36**, 561–578, <https://doi.org/10.1007/s00382-010-0871-6>.