

# Identification of large-scale meteorological patterns associated with extreme precipitation in the US northeast

Laurie Agel<sup>1,2</sup>  · Mathew Barlow<sup>1,3</sup> · Steven B. Feldstein<sup>4</sup> · William J. Gutowski Jr.<sup>5</sup>

Received: 26 October 2016 / Accepted: 4 May 2017 / Published online: 13 May 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** Patterns of daily large-scale circulation associated with Northeast US extreme precipitation are identified using both k-means clustering (KMC) and Self-Organizing Maps (SOM) applied to tropopause height. The tropopause height provides a compact representation of the upper-tropospheric potential vorticity, which is closely related to the overall evolution and intensity of weather systems. Extreme precipitation is defined as the top 1% of daily wet-day observations at 35 Northeast stations, 1979–2008. KMC is applied on extreme precipitation days only, while the SOM algorithm is applied to all days in order to place the extreme results into the overall context of patterns for all days. Six tropopause patterns are identified through KMC for extreme day precipitation: a summertime tropopause ridge, a summertime shallow trough/ridge, a summertime shallow eastern US trough, a deeper wintertime eastern US trough, and two versions of a deep cold-weather trough

located across the east-central US. Thirty SOM patterns for all days are identified. Results for all days show that 6 SOM patterns account for almost half of the extreme days, although extreme precipitation occurs in all SOM patterns. The same SOM patterns associated with extreme precipitation also routinely produce non-extreme precipitation; however, on extreme precipitation days the troughs, on average, are deeper and the downstream ridges more pronounced. Analysis of other fields associated with the large-scale patterns show various degrees of anomalously strong moisture transport preceding, and upward motion during, extreme precipitation events.

## 1 Introduction

Extreme precipitation is an important concern, in terms of societal and economic impacts, with adverse impacts due to global warming expected to increase (Melillo et al. 2014; Kunkel et al. 2013). The intensity and frequency of precipitation in the Northeast US (NE), particularly for the heaviest events, has increased in recent decades. Groisman et al. (2013) identified a 74% increase in the frequency of top 1% events from 1958 to 2010, while Griffiths and Bradley (2007) found, at 35% of their regional stations, a precipitation increase of ~20–50% per decade (1961–2000) in the top 5% wettest days, and Douglas and Fairbank (2011) found an increase ranging between 0.9 and 2.0 in. (22.86–50.8 mm) per decade in the annual coastal precipitation maxima from 1970 to 2008. These increases have been accompanied by an increase in NE flooding (Peterson et al. 2013), where some stream gauges have shown up to a ~9–12% increase in flooding events per decade for various record lengths of 85–127 years.

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00382-017-3724-8) contains supplementary material, which is available to authorized users.

---

✉ Laurie Agel  
Laurie\_Agel@student.uml.edu

<sup>1</sup> Department of Environmental, Earth, and Atmospheric Sciences, University of Massachusetts Lowell, One University Avenue, Lowell, MA 01854, USA

<sup>2</sup> Intercampus Marine Science Graduate Program, University of Massachusetts Lowell, Lowell, MA, USA

<sup>3</sup> Climate Change Initiative, University of Massachusetts Lowell, Lowell, MA, USA

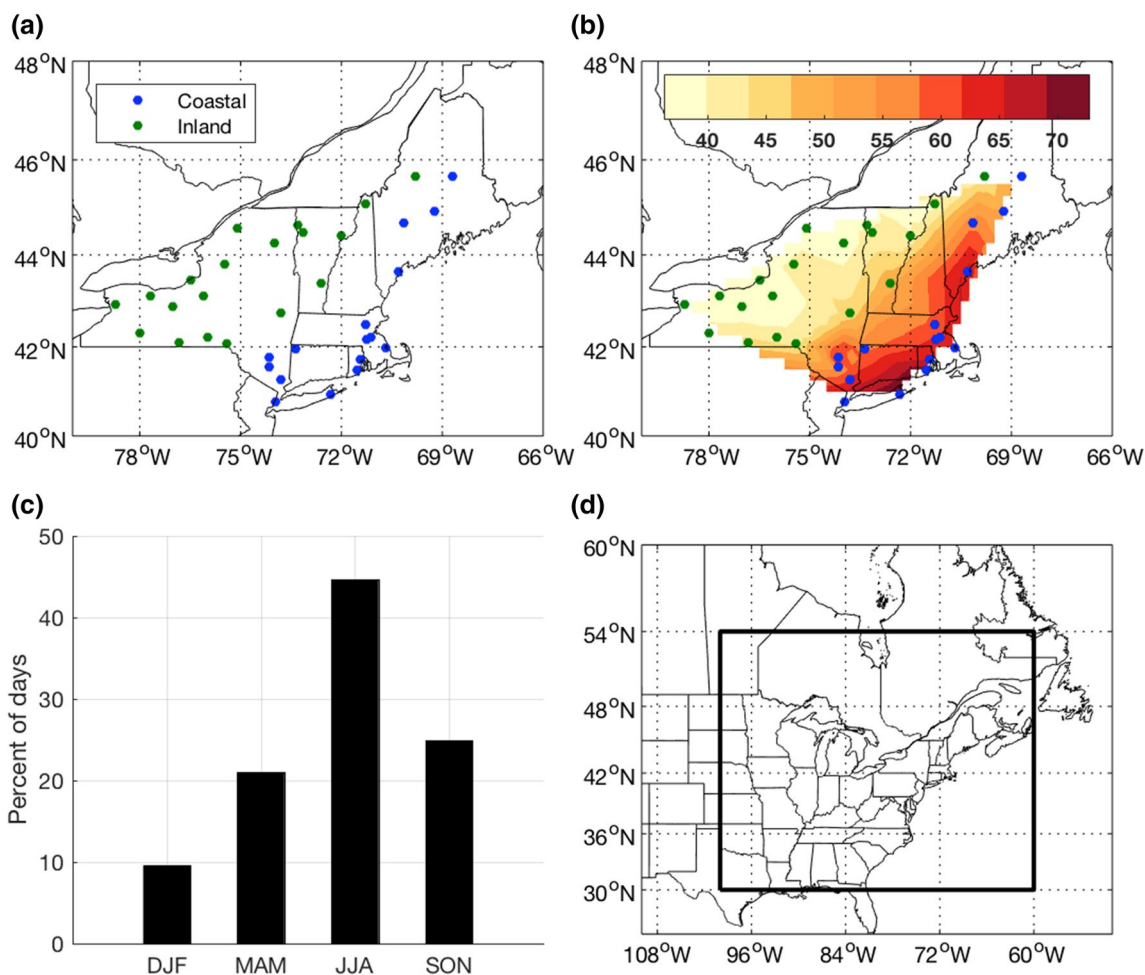
<sup>4</sup> Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA

<sup>5</sup> Department of Geological and Atmospheric Sciences, Iowa State University of Science and Technology, Ames, IA, USA

Because of the severe impacts associated with these extreme precipitation events, it is important to understand the dynamics, both large-scale and regional, that lead to extreme precipitation, in order to prepare for and better predict extreme precipitation events in both the current and changing climate. This study is the second part of a three-part effort to better understand the underlying dynamics that are associated with extreme precipitation in the NE. The first part examined the climatology of NE extreme precipitation in the context of overall precipitation (Agel et al. 2015). This second study seeks to identify the large-scale meteorological patterns (LSMPs; Grotjahn et al. 2016) that are associated with extreme NE precipitation. LSMPs describe the synoptic-scale circulation that can be attributed to specific phenomena, in this case, extreme precipitation. Both non-hierarchical k-means clustering (KMC; Diday and Simon 1976; Michelangeli et al. 1995) and Self-Organizing Maps (SOMS; Kohonen 2001) are used on an

estimate of the dynamic tropopause (DT) pressure field to identify the key LSMPs. While both techniques are effective at identifying circulation patterns, the combination of techniques is used to extract robust features identified by both techniques, and to gain additional insights that cannot be achieved by either technique alone. This combination of KMC and SOM typing may also be useful in identifying and examining the LSMPs associated with extremes for other regions. A third (ongoing) effort will examine in detail the dynamical structures and key ingredients for extreme precipitation such as moisture and instability that are linked to each of the identified LSMPs.

The climatology of extreme and overall NE precipitation is explored in Agel et al. (2015). The climatology can be divided into coastal and inland regimes (Fig. 1a). At coastal locations, precipitation intensity is greater than at inland locations, with peaks in precipitation in the spring and fall; while for inland locations, precipitation is more frequent



**Fig. 1** Plots of **a** USHCN station locations, with inland locations shown as *green dots* and coastal locations shown as *blue dots*, **b** 1979–2008 top 1% precipitation thresholds, interpolated from station locations (*shaded*, in mm), **c** seasonal frequency of 1979–2008

top 1% extreme precipitation days excluding those due to tropical cyclones (shown as a percent of 691 days), and **d** geographic domain (enclosed by *black box*) used for KMC and SOM typing

but less intense, with a single peak in total precipitation in the summer. Extreme precipitation is identified as the top 1% of daily precipitation observations from 1979 to 2008, with 66% of the extreme precipitation days associated with nearby synoptic-scale storms (that is, extreme precipitation falling within 1000 km of a low-pressure center), 19% associated with tropical cyclones, and the remainder (15%) assumed to be associated with processes such as convection, fronts, or mesoscale dynamics far removed from low-pressure centers. Here, we take the second step towards understanding the underlying dynamics of these particular extreme precipitation events by identifying the LSMPs associated with the non-tropical cyclone-related processes that lead to NE extreme precipitation. The large-scale links between tropical cyclones and NE precipitation extremes are explored elsewhere (e.g. Barlow 2011).

There is limited previous literature tying LSMPs to extreme precipitation in the NE. Roller et al. (2016) used KMC to identify wintertime LSMPs in the NE, and noted that extreme precipitation occurs preferentially in several of the patterns, although no specific links between LSMPs and extreme precipitation were established. Collow et al. (2016) investigated circulation features for summertime extreme precipitation in the NE, and found anomalously lower surface pressure and greater divergence aloft on extreme days. However, there are many studies that identify LSMPs for extreme precipitation for regions other than NE, including Jones et al. (2004), Santos et al. (2007), Favre and Gershunov (2009), Milrad et al. (2010), Higgins et al. (2011), Kunkel et al. (2012), and Glisan and Gutowski (2014a, b), although these studies do not use objective pattern-identification techniques such as KMC or SOMs to identify the LSMPs. Alternatively, several studies do use KMCs or SOMs to identify LSMPs (Cavazos 1999; Hewitson and Crane 2002; Gutowski et al. 2004; Cassano et al. 2007, 2015; Riddle et al. 2013; Glisan et al. 2016), although not necessarily to address precipitation extremes. In particular, Cavazos (1999) used SOMs of 500–700-hPa thickness, mean sea-level pressure (MSLP) tendency and specific humidity to identify LSMPs related to precipitation extremes in southeastern Texas and northeastern Mexico. Cassano et al. (2015) used SOMs of MSLP to identify patterns related to temperature extremes in the Arctic and northern Canada. Riddle et al. (2013) used KMC on 500-hPa geopotential heights over North America to identify wintertime circulation patterns associated with all 8 MJO phases (Wheeler and Hendon 2004). In addition, there are several non-pattern-based studies that tie extreme precipitation to specific mechanisms: Hawcroft et al. (2012) and Pfahl and Wernli (2012) investigate the relationship between extreme precipitation and extratropical cyclones and find that extratropical cyclones account for up to 90% of precipitation extremes in certain regions; Catto and Pfahl

(2013) find that globally 75% of midlatitude precipitation extremes are related to fronts associated with extratropical cyclones; and Pfahl et al. (2014) quantifies the role of warm conveyor belts in producing extreme precipitation.

Here, we use both KMC and SOM techniques together to identify robust LSMPs associated with NE extreme precipitation. KMC is used to separate extreme precipitation days into a number of circulation patterns, while SOMs are used to separate all days in the study period into representative circulation patterns. By examining which SOM patterns are tied to extreme precipitation days, we gain secondary validation of the number of KMC clusters chosen and the KMC patterns themselves, as well as information to place the extreme precipitation day patterns within the context of the SOM pattern-space representing all days. This additional information includes evaluation of pattern duration and transitions to other patterns, as well as perspective into whether extreme-producing SOM pattern days are different, or in any way extreme, from non-extreme-producing pattern days.

Our LSMPs are identified by applying KMC and SOMs to an estimate of the DT pressure field (details in Sect. 2). The DT provides a compact representation of the upper-tropospheric potential vorticity (PV), as isentropic gradients of PV are concentrated at the tropopause distribution (Morgan and Nielsen-Gammon 1998). Since upper-tropospheric PV is closely related to surface development, mid-level vertical motion, and the overall evolution and intensity of weather systems (e.g., Hoskins et al. 1985; Davis and Emanuel 1991; Bosart 1999), consideration of the DT provides a convenient perspective on the synoptic-scale dynamics we can expect to play an important role in many (though certainly not all) extreme precipitation events. Our future work will build on this upper-level dynamical perspective to include consideration of moisture availability, diabatic heating, low-level circulation, and stability. It also has the advantage of less seasonal dependency, in contrast to a single pressure-level field such as 200-hPa geopotential height, which is not fully representative of the tropopause and the upper-level jet in all seasons (particularly the warm season). This allows us to use typing for all calendar days at once, which increases the sample size, and to use a single annual threshold for extremes. Analysis of the DT is also useful for identifying areas of deep convection (Schumann and Roebber 2010), strong jet-stream waveguides (Martius et al. 2010), wave-breaking events, tropopause folds and PV streamers (Nielsen-Gammon 2001).

The balance of this paper is organized as follows. The data used for the study is described in Sect. 2. The SOM and KMC techniques are described in Sect. 3. Results are given in Sect. 4, with separate subsections covering the KMC results and sensitivity analysis, the SOM results, extreme pattern persistence and transitioning using the

KMC/SOM results together, and a brief overview of the three-dimensional circulation related to the KMC tropopause patterns. The paper concludes with a summary and discussion in Sect. 5.

## 2 Data

### 2.1 Precipitation

To identify extreme precipitation days, daily precipitation records over the period 1979–2008, from 35 United States Historical Climatology Network (USHCN; Williams et al. 2004) stations located in the NE (Fig. 1a), are sorted according to daily accumulation, and the top 1% of days with precipitation at each station are selected to represent extreme precipitation events, as in Agel et al. (2015). The thresholds for extreme precipitation vary from inland locations to coastal locations, as shown in Fig. 1b. Only stations missing less than 1% of data for the time period are included. This results in 1563 extreme precipitation events (where an event represents extreme precipitation at a particular station, noting that extreme precipitation can occur at multiple stations on the same day). The events are filtered to remove all days where extreme precipitation occurs at stations within 1000 km of a tropical cyclone center, as identified in the National Hurricane Center revised Atlantic hurricane database (HURDAT2; Landsea and Franklin 2013). This results in 691 unique extreme precipitation days, where non-tropical cyclone-related extreme precipitation events occur at one or more stations in the region. The seasonal frequency of the extreme days is shown in Fig. 1c.

Interquartile (25–75%) precipitation days are also identified, for comparison to extreme precipitation typing results. Since it is difficult to identify unique interquartile days using station precipitation, which can vary considerably from station to station, we use the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC)  $0.25^\circ \times 0.25^\circ$  Daily U.S. Unified Precipitation (CPCU; Chen et al. 2008), provided by NOAA/OAR/ESRL PSD ([http://www.esrl.noaa.gov/psd/data/gridded/data.unified.daily.conus.html](http://www.esrl.noaa.gov/psd/http://www.esrl.noaa.gov/psd/data/gridded/data.unified.daily.conus.html)). The daily grid values are averaged over the region bounded by the thick blue line in Fig. 1d (identical to the domain used for the typing analysis), and the 25th–75th percentile values are chosen to represent the interquartile precipitation. For the study area and period, the gridded product has good spatial correlation to daily station data (mean correlation 0.82 and median correlation 0.86), although a direct comparison of CPCU precipitation (accumulation from 12 UTC of the day before to 12 UTC of the day) to USHCN precipitation is complicated by the varied station 24-h accumulation periods.

### 2.2 Typing field

Dynamic tropopause (DT) pressure anomalies are used as input to both the SOM and KMC typing algorithms. We use the National Aeronautics and Space Administration (NASA) Modern Era Retrospective Reanalysis for Research and Application (MERRA, Rienecker et al. 2011) blended tropopause pressure field as a proxy for DT pressure, available at hourly intervals and  $1/2^\circ \times 2/3^\circ$  resolution, for the years 1979–2008. The blended field is a mix of lower-latitude thermal-based tropopause pressure and higher-latitude PV-based tropopause pressure, which provides a smooth transition from the tropics to the mid-latitudes. As these fields are very similar within our domain (e.g., the spatial correlation between the purely PV-based DT field and the blended field averages 0.98 for our domain, based on Pearson correlation between grid box values for each day of a single year, 1980), either the blended or PV-based field can be used interchangeably as the DT for this application. The region used for the typing algorithms is shown bounded by the thick blue line in Fig. 1d ( $30^\circ\text{N}$ – $54^\circ\text{N}$  and  $100^\circ\text{W}$ – $60^\circ\text{W}$ ), and encompasses the eastern third of the US and parts of southern Canada.

We calculate DT pressure anomalies by first computing the daily mean (MERRA provides 24 fields per day). The long-term daily mean is then calculated by taking the 30-year average of daily means for each unique calendar day (e.g. 1 January, 2 January, etc., through 31 December), resulting in a 366-day dataset. We then smooth this 366-day dataset with a 21-day running mean. The smoothed long-term daily means are then subtracted from the daily means to create daily DT pressure anomalies (e.g. the 1 January long-term daily mean is subtracted from each 1 January daily mean). Although we use daily mean anomalies for the typing application, all composite figures are shown using the 12 UTC field, unless otherwise noted.

### 2.3 Other

Additional MERRA fields are used to analyze the key meteorological ingredients leading to extreme precipitation for each pattern, including 500-hPa geopotential height, MSLP, integrated vapor transport (IVT), and vertical velocity. IVT is available on the same grid and timescale as the DT pressure; all other fields are available at 3-hr intervals on a  $1.25^\circ \times 1.25^\circ$  grid.

## 3 Methodology

### 3.1 K-means clustering (KMC)

In this study, we use the MATLAB built-in “kmeans” algorithm to separate the DT pressure anomalies on extreme



precipitation days into a number of representative patterns. Before inputting, the DT pressure anomalies are area-weighted based on latitude (each value is multiplied by the cosine of its latitude), and then standardized at each grid-point by removing the grid-point mean (based on all days 1979–2008) and dividing by the grid-point standard deviation (based on all days 1979–2008). The field is then reduced through EOF filtering to retain 90% of variance (this reduces our variable space from 2989 grids to 21 dimensions, which significantly reduces the KMC computation time).

KMC is a process that separates input data into a pre-selected number of non-overlapping groups. Each group, or “cluster”, is defined by its “centroid” (the mean of the inputs assigned to that cluster). Initial centroids are usually determined by a random or semi-random selection of input data. All input data is then assigned to the nearest centroid (here, based on the squared Euclidean point-to-centroid distance), after which the centroids are recalculated. This process is repeated (input data assigned to nearest centroid) until further iterations no longer reduce the sum of the intra-cluster variances. The final centroids are then representative patterns of the input data. In practice, many  $k$ -means separations (each initialized independently) are run on the input data, and an objective technique is used to select the “best” separation, or partition. Often these techniques involve selecting the partition with the smallest intra-cluster and/or largest inter-cluster variance. Here we use method of Michelangeli et al. (1995) to ascertain the “best” partitioning into clusters. We run 1000 independent trials, and for each partition pair  $P$  and  $Q$  we compute the anomaly correlation coefficients  $C(P_i, Q_j)$  between clusters  $P_{i, i=1..k}$  and  $Q_{j, j=1..k}$ , where  $k$  represents the pre-selected number of clusters. For each cluster  $P_i$ , the best analog to a cluster in  $Q$  is defined as  $A_i = \text{maximum}(C(P_i, Q_{j, j=1..k}))$ , and the overall correlation between partitions  $P$  and  $Q$  is defined as  $\text{minimum}(A_{i, i=1..k})$ . This is repeated for every partition pairing, and the “best” partition is that with the highest mean overall correlation to all other partitions.

One of the difficulties of the  $k$ -means approach is that the number of clusters desired must be specified a priori. There are a number of techniques to infer a reasonable number of clusters for a dataset. Here, we use two objective tests to establish the best value for  $k$ . The first test involves computing the Classifiability Index (CI; Michelangeli et al. 1995) for each partitioning given  $k$  varying from 1 to 10. The CI gives a measure of how consistently the data is separated into representative patterns. For each value of  $k$ , 1000 partitionings are created, and the overall correlation of each partition  $P$  to each other partition  $Q$  is calculated as described above. The CI is defined as the mean of these correlations, and varies from 0 to 1. Using  $k=1$  would naturally result in a CI equal to 1.0 since there

is only one way to separate the data into a single pattern; while using values larger than  $k=1$  would likely result in smaller values of CI, as randomized initializations would produce at least slightly different, and often very different, final partitionings. The optimum  $k$  is determined by creating 100  $k$ -means separations of random red-noise datasets (generated from the original data) for each value of  $k=n$ ,  $n=1.0.10$ , and calculating the corresponding random-noise CIs. The smallest  $k$  that generates a CI that is higher than the 90th percentile random-noise CI for the same  $k$  is then considered the “optimum” choice for the number of clusters.

The second technique is based on finding the first local minimum of the “Volume Ratio Index” (VRI; Riddle et al. 2013), where VRI for any particular  $k=n$  is the sum of the ratios of the 90th percentile of Euclidean distances-to-centroid for each cluster to the 90th percentile of Euclidean distances-to-centroid for the  $k=1$  partition. The 90th percentile is based on the 1000 randomly initialized partitionings, as discussed above for the CI. A minimum VRI represents the solution that covers the range of data most efficiently. Riddle et al. (2013) note that this measure is notably unstable for large values of  $k$ , and should be used with caution until the method is perfected. In this case, results are stable up to  $k=8$ .

### 3.2 Self-organizing maps (SOMs)

SOMs are gaining traction in the climate research community as a powerful tool to analyze meteorological phenomena (Feldstein and Lee 2014; Cassano et al. 2015). SOMs use neural network classification and unsupervised learning to iteratively separate input data along discrete “nodes” of the pattern-space represented by the input field. Each node is defined by a set of weights equal to the input data dimension. The process creates more (fewer) nodes where data density is high (low). The resulting patterns are similar to those derived through other typing means, such as KMC; however SOMs are unique in that each piece of input data affects the weights of not only the node it is assigned to, but also neighboring nodes. For atmospheric applications, it is useful to arrange the nodes in a rectangular pattern-space, where the number of rows and columns are specified beforehand. The nodes represent a continuum of the pattern-space the input field occupies, where similar nodes are near each other in the two-dimensional mapping, and dissimilar nodes are farther apart. This mapping has an advantage over KMC, where the relationship between clusters is not easily inferred. Care must be taken to choose the appropriate dimensions, as too many nodes may dilute the significant physical differences in the patterns, while too few nodes may not effectively isolate extreme-producing (or other targeted phenomena) patterns.

In this study, the SOM algorithm is used to separate DT pressure anomalies for all days, 1979–2008. The field is pre-processed as for the KMC approach, except the dimensionality is not reduced with EOF. The MATLAB SOM Toolbox 2.0 from <http://www.cis.hut.fi/somtoolbox/> is used, with parameters set for linear initialization, 200 initial training iterations, 1200 secondary training iterations, and a two-dimensional rectangular pattern space. There are few objective tests for optimum pattern space dimensions (Grotjahn et al. 2016). Cassano et al. (2015) recommend choosing a pattern dimension that minimizes the root-mean-square deviation (RMSD) of the input fields to the assigned SOM patterns, while also minimizing the distortion, based on Sammon (1969) mapping, between adjacent nodes. Here, we use a trial-and-error approach to identify a pattern space that is just large enough to (1) provide a sufficient range of patterns to fully represent important circulation features, (2) provide a reasonable spatial correlation of individual DT pressure fields to assigned patterns, and (3) isolate extreme days to a subset of patterns. Pattern spaces tested include  $3 \times 4$ ,  $3 \times 5$ ,  $4 \times 5$ ,  $4 \times 6$ ,  $5 \times 6$ , and  $5 \times 7$ . We ultimately choose the  $5 \times 6$  pattern space to represent the range of tropopause patterns and to otherwise meet the requirements listed above. It is important to note that our goals here justify this approach: we are searching for additional corroboration of our KMC clusters, and separation of extreme precipitation-producing patterns within the total pattern-space. This trial-and-error method may not be useful for all applications, particularly when there is no prior knowledge of the dataset.

### 3.3 Combining approaches

We use a combination of KMC and SOM analysis to identify the LSMPs associated with extreme precipitation. The KMC algorithm is used to separate DT pressure anomalies on extreme precipitation days into characteristic patterns. A SOM analysis of DT pressure anomalies is then performed on all days in the time series, and the SOM patterns that most frequently represent extreme precipitation days are compared to those derived from KMC. This has the dual advantage of verifying the KMC patterns for extreme days, and to place the KMC patterns into the context of the larger SOM pattern-space. It also provides a third independent assessment of the number of clusters chosen for KMC, as we anticipate distinct extreme precipitation-producing patterns to be separated in the SOM pattern-space. Additionally, the SOM results (all days) give insight into how the KMC (extremes) patterns transition to/from other patterns, or persist in the same pattern.

### 3.4 Significance testing

In several cases, we evaluate whether frequency analysis results are significantly different from background frequency (e.g. whether the frequency of extreme precipitation days assigned to a SOM node is higher or lower than the frequency of all days assigned to the node, or whether the frequency of DJF days in a KMC pattern is higher or lower than the overall frequency of DJF extreme precipitation days) using a Monte Carlo approach with 1000 trials. We illustrate the approach with a hypothetical situation. First, consider a case in which there are 1000 days in the study, of which 20 are extreme precipitation days (a background frequency of 2%). If 50 days are assigned to a certain SOM node and of these 5 days are extreme precipitation days (10%), then for each trial we would randomly select 50 days from the set of all days (without replacement) and we would calculate the percent of those days that are extreme. Over the course of many trials, the distribution of the percent of extreme days should peak close to the background frequency (2%), and we can determine if the actual frequency of extreme days for the SOM (10%) is statistically different from the background frequency, by sorting the frequency results of the 1000 trials from smallest to largest for each pattern, and choosing the top and bottom 2.5% as the upper and lower bounds. Frequencies higher (lower) than the upper (lower) bound would be significant at the 0.05 level. As a second example, consider the case where 100 extreme precipitation days are assigned to a KMC cluster, with 10 of those days occurring in DJF (10%), while 15% of all extreme precipitation days occur during DJF. To determine if the percent of DJF days (10%) assigned to the cluster is significantly lower than the background frequency of 15%, we would randomly select 100 days from the set of all days (without replacement) and we would calculate the percent of those days that occur in DJF, repeating 1000 times and determining the significance at the 0.05 level as before. A similar Monte Carlo approach is used to evaluate the statistical significance of pattern correlation and duration results.

### 3.5 Sensitivity to changing time and spatial domain

For our study of NE extreme precipitation, we used a typing domain for the dynamic tropopause field that is much larger than the area where the stations are located (seven NE states as shown in Fig. 1a). This allows us to capture large-scale ridge and trough circulations that likely lead to precipitation. To test the sensitivity of the results to domain size, we also performed the KMC analysis using a larger (smaller) domain, by reducing (increasing) the domain by  $2^\circ\text{N}$  and  $2^\circ\text{W}$ . The smaller domain results are very similar, and the larger domain results in similar patterns but with

one pattern acquiring some dates from other clusters. We also perform the analysis on half the dataset (randomly selected), and get nearly identical results for  $k=3$  but not for larger  $k$ . This may be due to the small sample size—if we randomly select 75% of the days we get nearly identical results as those using all the days. As a final note, using a single field time (12 UTC) as opposed to the daily mean does not appreciably change the results.

## 4 Results

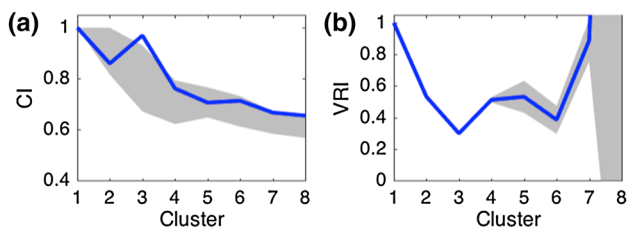
### 4.1 K-means analysis

In this study, as stated in Sect. 2, the top 1% of daily precipitation at each of 35 NE stations, 1979–2008, is used to identify extreme precipitation days, which results in 691 non-tropical cyclone-related days. The DT pressure anomalies on these days are typed according to the KMC algorithm, with  $k$  varying from 1 to 10. Two objective measures to identify the optimum separation of patterns, CI and VRI, both suggest using  $k=3$ , with a possible secondary suggestion of  $k=6$  (Fig. 2). The 3-cluster solution includes a ridge pattern, an eastern US trough, and a deep Ohio Valley trough (Patterns 1–3 in Fig. 3a, where here and throughout the paper, “pattern” refers to the cluster composite). The ridge pattern is slightly more common than either trough pattern, while the Ohio Valley trough is the least common. Seasonally, there is some variation in the spatial location and strength of the patterns (Fig. 3b), as well as the frequency of the patterns (Fig. 3c). Monte Carlo analysis is used to determine whether the seasonal frequency is more or less likely to occur than the background frequency (as denoted by bar color in Fig. 1c). The ridge pattern is more likely to occur during JJA and less likely to occur in other

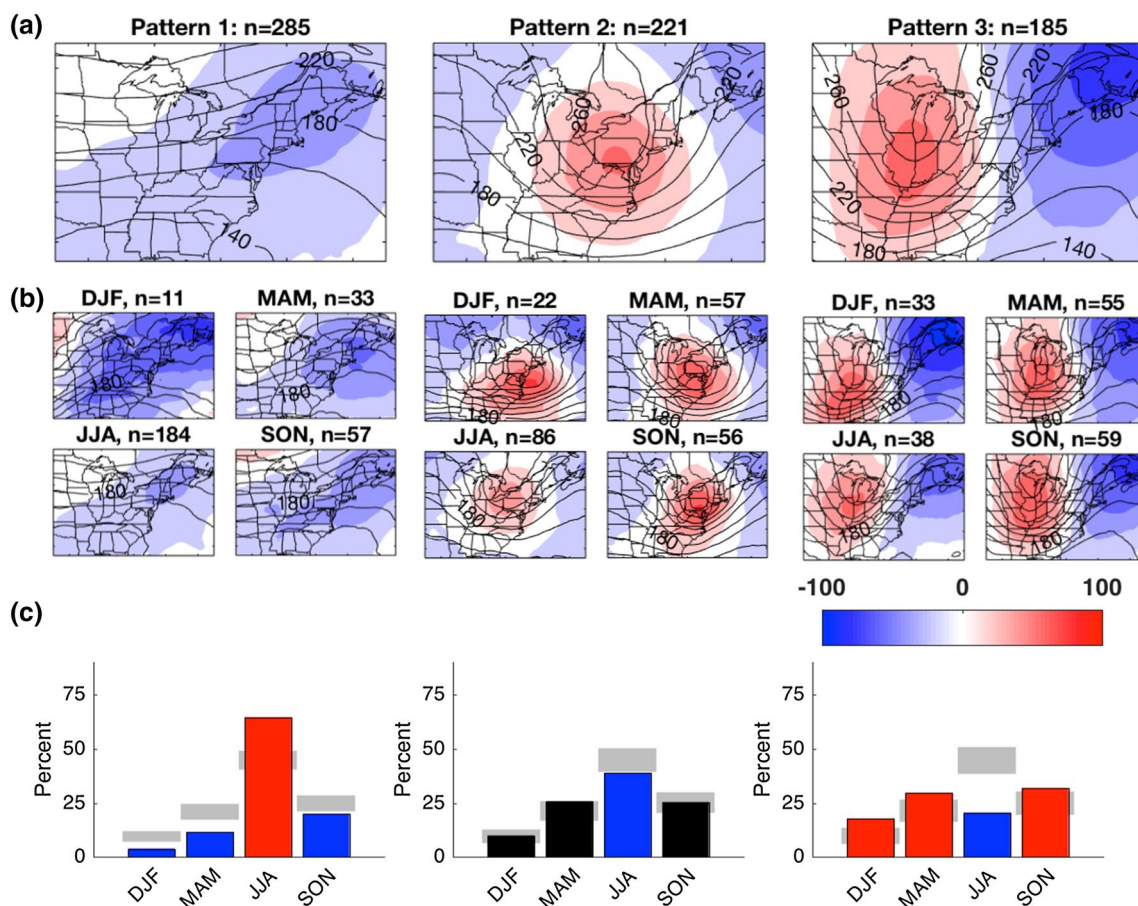
seasons, while the eastern trough pattern is less likely to occur during JJA. The Ohio Valley trough is less likely to occur during JJA but more likely to occur during any other season.

Because of seasonal frequency differences, KMC is also performed on the top 1% of precipitation extremes per season (corresponding figures not included). This results in 66 DJF extreme days, where CI and VRI both suggest  $k=4$ , although the CI result is not statistically outside the 90% confidence interval for red noise. The four patterns include a deep Ohio Valley trough, two versions of eastern troughs—one farther east and one farther north, and a trough/ridge combination. Only the Ohio Valley trough shares consistent dates with the annual solution Ohio Valley trough. For MAM, 145 extreme days are identified, and CI and VRI suggest using  $k=3$ . The three resulting patterns are very similar to the annual solution for  $k=3$ , with excellent consistency for MAM dates. For JJA, 308 extreme days are identified, and CI and VRI both suggest  $k=3$ . The three patterns include a trough/ridge combination, a ridge, and a shallow eastern trough, with excellent date agreement for the latter two with the annual solution (the trough/ridge days are a combination of the first solution’s ridge and Ohio Valley trough days). For SON, 172 extreme days are identified. The CI measure does not show a preferred partitioning, but VRI suggests  $k=3$ : a ridge, an eastern trough, and an Ohio Valley trough, with excellent agreement of SON dates with the annual solution. Based on these results, we find that seasonal application of KMC to DT pressure anomalies results in similar patterns to the annual solution, but with an additional trough/ridge pattern in DJF and JJA, and some slight adjustments to trough locations based on the season.

We also investigate the secondary  $k=6$  annual solution suggested by CI and VRI (C1–C6 in Fig. 4a, with seasonal frequency in Fig. 4b). Importantly, this solution comprises each of the seasonal patterns, including the variations in trough locations and the DJF/JJA trough/ridge pattern. The six patterns include two versions of the  $k=3$  ridge: an expansive ridge over the entire eastern US during JJA (C1), and the trough/ridge combination (C4). The patterns also include two versions of the  $k=3$  eastern trough: a deeper trough that is more likely to occur during cold seasons (C2), and a shallower trough that is more likely to occur during JJA (C5). Lastly, the patterns include two versions of the  $k=3$  Ohio Valley trough: a slightly shallower trough located farther north and more likely to occur during MAM or SON (C3), and a very deep trough located farther southwest and more likely to occur during DJF (C6). Not only are each of the 6 seasonal patterns represented, there is also an excellent separation of the  $k=3$  dates into the two versions of each of the  $k=3$  patterns (Fig. 5). Thus the  $k=6$  solution offers an elegant refinement of the  $k=3$  solution,



**Fig. 2** Two objective measures of k-means “optimum”  $k$ , for k-means separation of MERRA daily mean DT pressure anomalies, using 691 top 1% extreme precipitation dates from 1979 to 2008. The first measure is **a** the “Classifiability Index” (CI) from Michelangeli et al. (1995), where values outside the 90% confidence interval of results for random red-noise simulations (*grey shading*) indicate optimum separation. The second measure is **b** the “Volume Ratio Index” (VRI) from Riddle et al. (2013), where optimal separation occurs at the first local minimum. *Grey shading* in the VRI plot shows the standard error of the means of 1000 simulations



**Fig. 3** Results of k-means separation of MERRA daily mean DT pressure anomalies on 691 extreme precipitation days, for  $k=3$  patterns, showing **a** composites of the three patterns (contours, in 20-hPa intervals, and anomalies, shaded), **b** seasonal composites of the three patterns, and **c** seasonal frequency (expressed as a percent) of each

pattern (black bars, except red if greater than that expected, and blue if less than that expected, due to chance). The text above each pattern indicates the number of dates in each composite. The three patterns, left to right, represent a “ridge”, an “eastern trough”, and an “Ohio Valley trough”, respectively

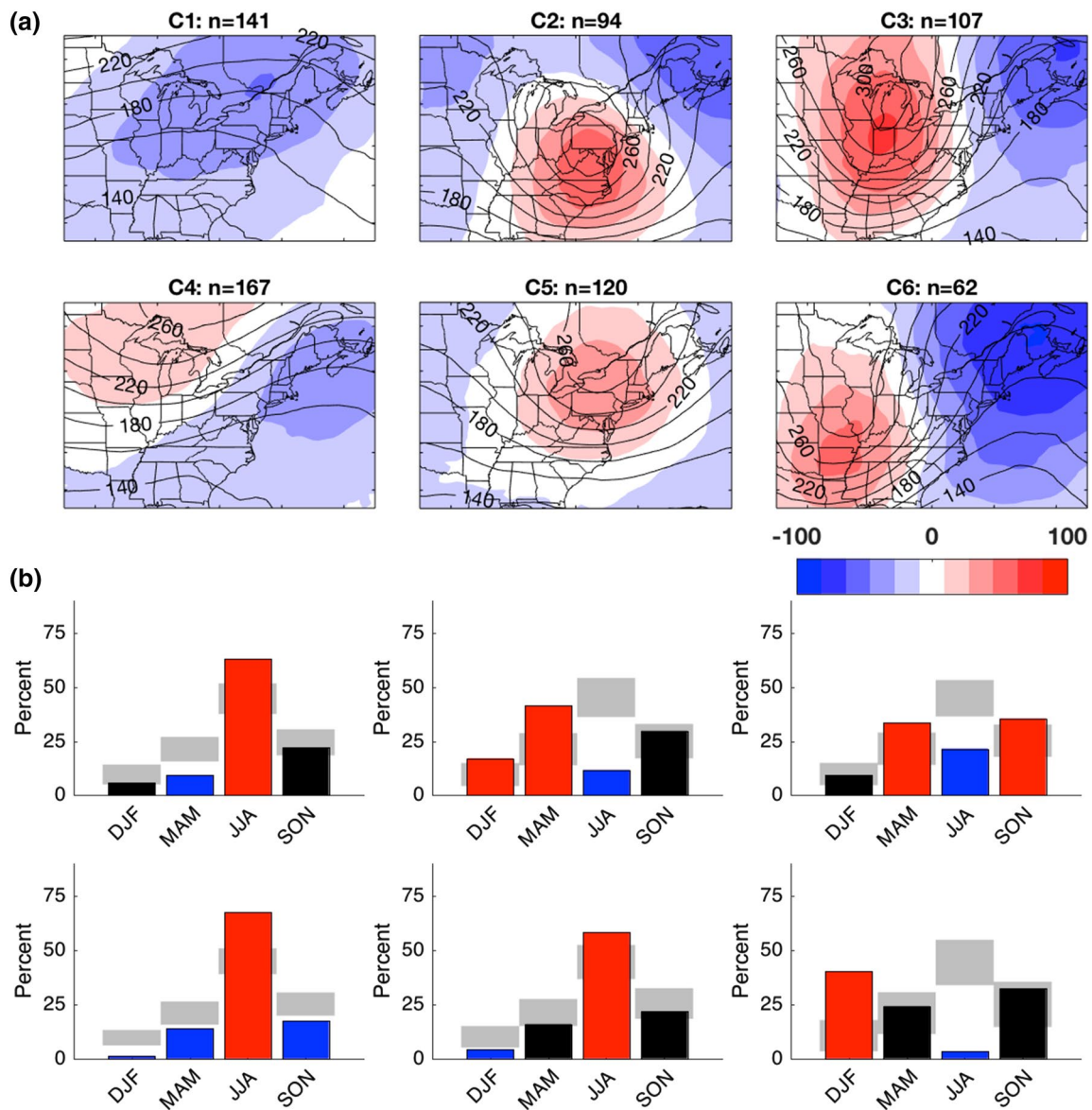
eliminating the need for separate seasonal typing analyses. A comma-separated text file showing cluster assignments by date, for both the  $k=3$  and  $k=6$  solutions, is available as Online Resource 1.

Differences in the location of the stations that experience extreme precipitation are also well represented by the separation into three (Patterns 1–3, Fig. 6a) and six patterns (C1–C6, Fig. 6b), with the 6-pattern solution once again giving finer detail. Composites of CPCU gridded daily precipitation are also shown for each cluster in Fig. 6. The ridge patterns (C1/C4) are more likely to cause extremes at stations in the western portion of the domain. The wintertime eastern trough (C2) is more likely to generate extremes at southern New England stations, while the summertime eastern trough (C5) affects northern New England stations. The DJF Ohio Valley trough (C6) generates more extremes at southwestern New England stations, while the wintertime Ohio Valley trough (C3) does not appear to favor a particular region.

For further insight into the appropriateness of the patterns, we look at how well individual days are represented by each pattern. Figure 7 shows mean spatial correlation for each of the KMC patterns, for both the  $k=3$  and  $k=6$  solutions, while Online Resource 2 (Figs. S1–S6) show the DT pressure for the top 6 correlated days to each of the  $k=6$  patterns. The correlation is moderate for the KMC patterns, considering there are only a few patterns accounting for each of the 691 extreme precipitation days. However, increasing  $k=3$  to  $k=6$  tends to increase the correlation of the individual days, with the deep troughs showing the best improvement in correlation.

As noted, we performed the KMC analysis using DT pressure anomalies. However, we also performed the same analysis using 500-hPa geopotential height anomalies (results not shown), in order to test the pattern sensitivity to input field. For the 500-hPa geopotential height anomalies, four patterns are suggested, including a ridge, two versions of the eastern US trough, and a deeper Ohio Valley trough

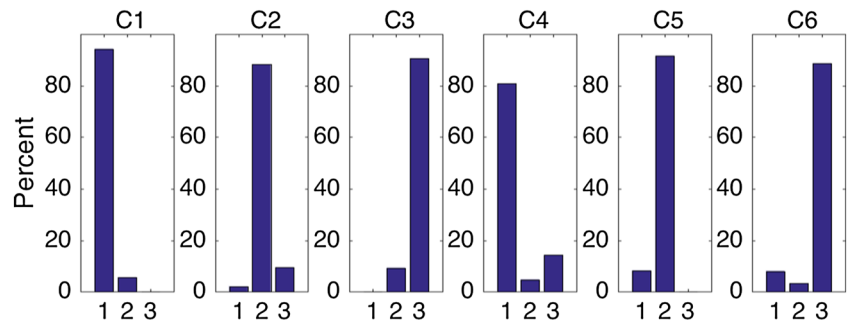


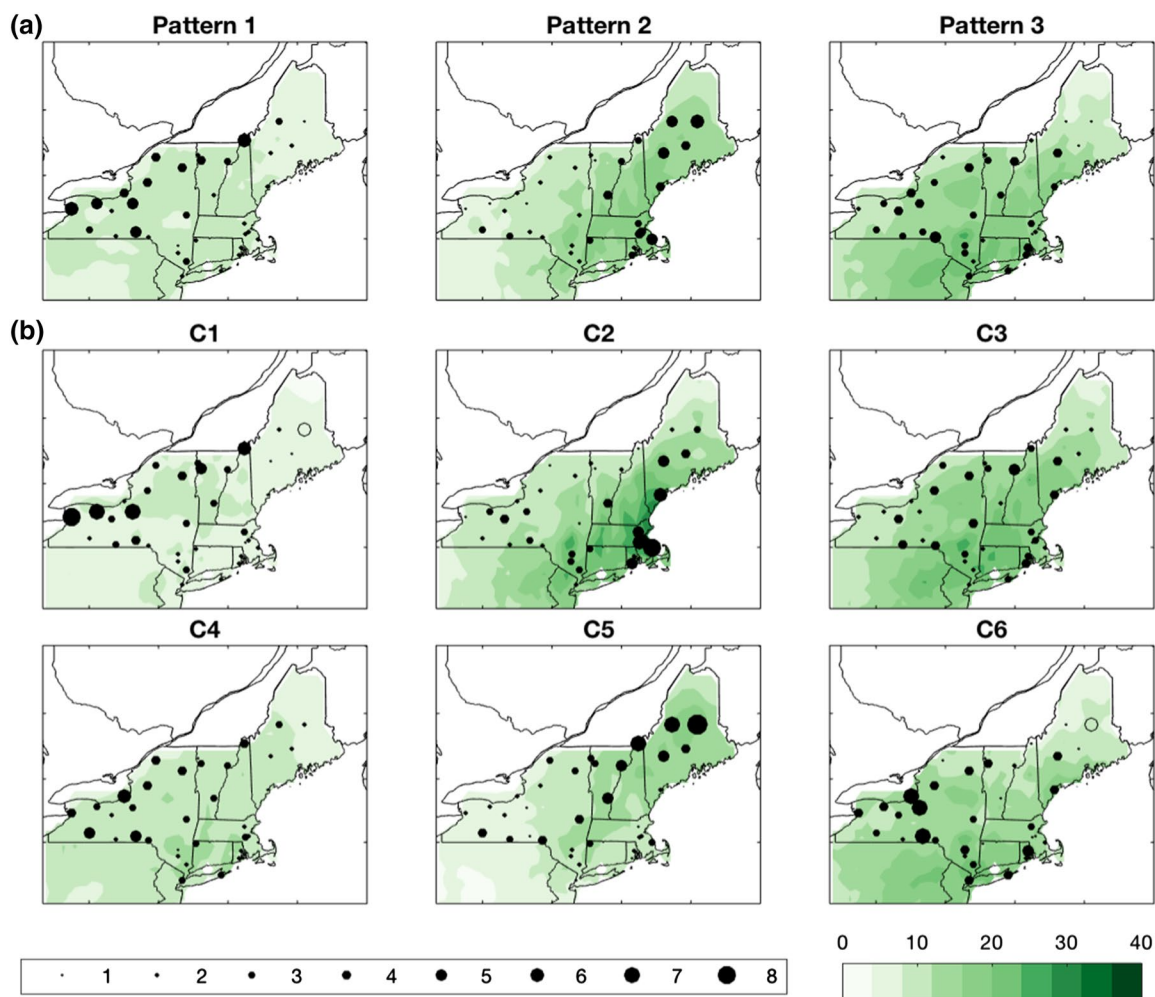


**Fig. 4** Results of k-means separation of MERRA daily mean DT pressure anomalies on 691 extreme precipitation days, for k=6 patterns named C1–C6, showing **a** composites of the six patterns (anomalies, shaded, and contours, in 20-hPa intervals), and **b** seasonal frequency (expressed as a percent) of each pattern (black bars, except

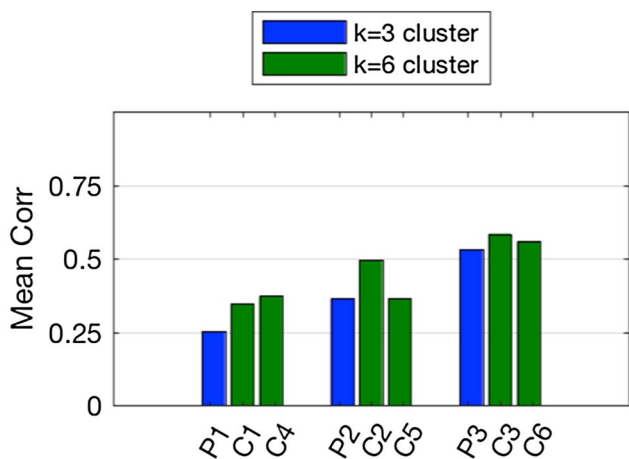
red if greater than that expected, and blue if less than that expected, due to chance). The text above each pattern indicates the number of dates in each composite. The left patterns represent two types of “ridges”, the center patterns two “eastern troughs”, and the right patterns two “Ohio Valley troughs”

**Fig. 5** The frequency of cluster days for each of k=6 clusters C1–C6, that fall into each of k=3 clusters (1–3 on the x-axis). The graphs show that C1 and C4 coincide with the first k=3 cluster, C2 and C5 coincide with the second k=3 cluster, and C3 and C6 coincide with the third k=3 cluster, with very little exception





**Fig. 6** Composite gridded daily CPCU precipitation (mm, shaded), and percent of extreme pattern-days assigned to each station within a pattern 1–3 for  $k=3$  and **b** C1–C6 for  $k=6$ , with the size of the dots proportional to the percentage of extreme pattern-days



**Fig. 7** Mean spatial correlation of individual days to patterns for KMC  $k=3$  solution (P1–P3) and  $k=6$  solution (C1–C6)

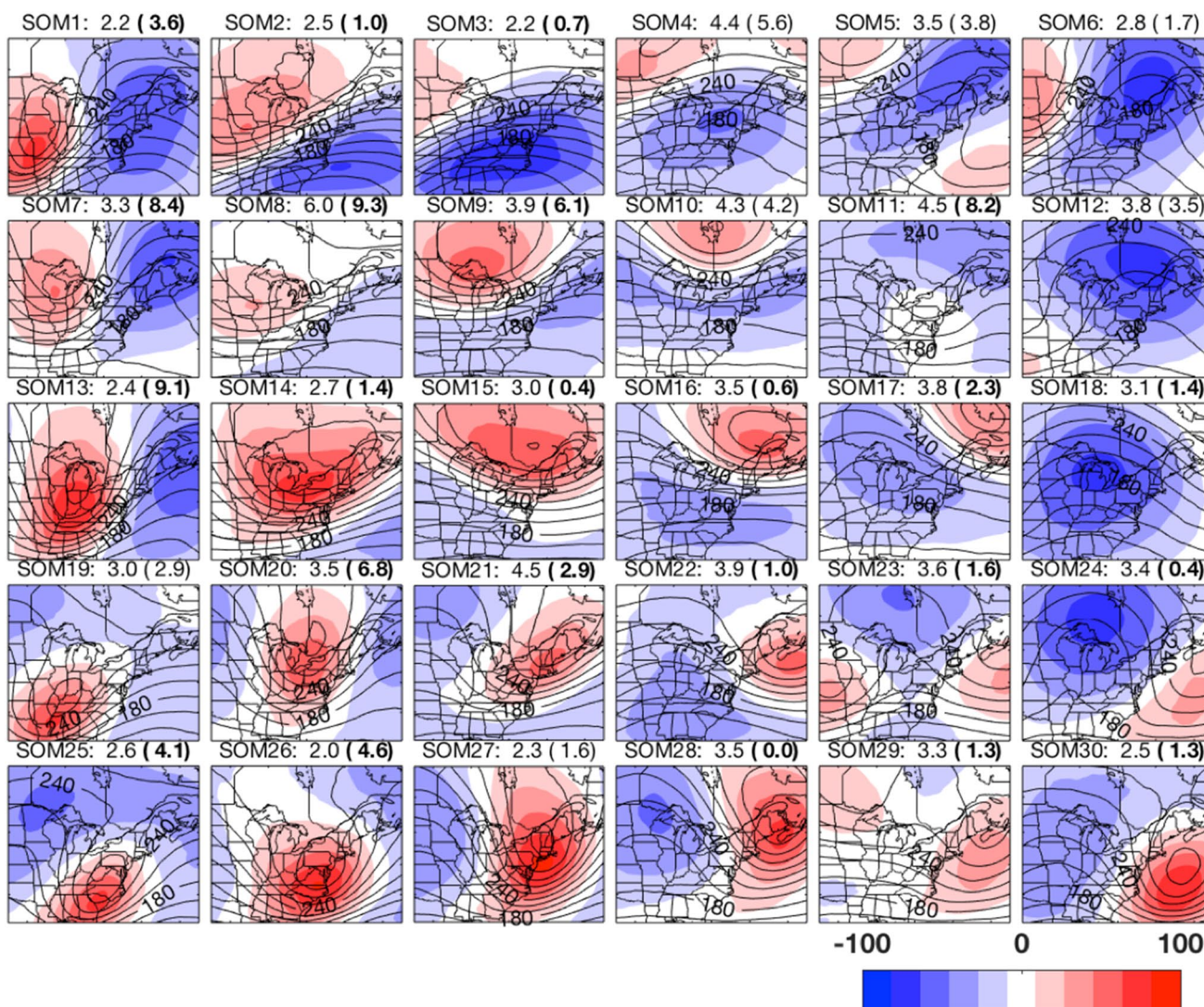
(all very similar to the patterns generated using the DT pressure anomalies), with a seasonal separation of the eastern US troughs. A secondary CI/VRI suggestion of  $k=8$  identifies a third type of shallow eastern US trough active in MAM, again confirming that using a secondary optimal separation can isolate seasonality in patterns, without performing typing techniques on separate seasons. There is good overlap between the six tropopause patterns and the eight 500-hPa patterns, with five patterns of each matching ~70% or greater of the dates. The sixth tropopause pattern, the summertime eastern US trough (C5), is split among two 500-hPa patterns; while several of the wintertime tropopause trough dates (C2, C3) combine for the eighth 500-hPa pattern. Hence, despite using a different typing field, we find that the ridge pattern and the seasonal variations of the eastern US trough and deeper Ohio Valley trough are robust patterns of the large-scale circulation associated with NE extreme precipitation.



An important result of this KMC analysis is that the minimal cluster solution indicated by objective techniques such as CI and VRI may at times mask out important pattern variations, especially those due to seasonality; and secondary solutions offered by these techniques should not be ignored, as they may more fully address the space/time complexity of the patterns representing some phenomena. Here we prefer the secondary solution of  $k=6$  to efficiently isolate seasonal variations in the patterns.

### 4.2 SOM analysis

To place the KMC results into a larger context, and as a secondary verification of the extreme precipitation tropopause patterns, we also conduct a SOM analysis of the DT pressure anomalies on all days, 1979–2008, using a  $5 \times 6$  rectangular pattern space (Fig. 8). The resulting SOM patterns are labeled 1–30, with the frequency of pattern days indicated above each pattern, along with the relative frequency of extreme precipitation days shown in parenthesis (bolded if significantly different from the pattern frequency, as determined by Monte Carlo analysis). Each of the 30 patterns represents from 2.0 to 6.0% of the days in the time



**Fig. 8** Pattern-space resulting from a  $5 \times 6$  SOM separation of the MERRA daily mean DT pressure anomalies into 30 patterns (contours, in 20-hPa intervals, and anomalies, shaded). The text above each pattern indicates the SOM node, the percent of days accounted for by each SOM node, and the percent of extreme precipitation

days (top 1%, excluding tropical-cyclone-related extremes, shown in parentheses) accounted for by each SOM node. The percent of extreme days is bold if the value is lower or higher than that expected due to chance

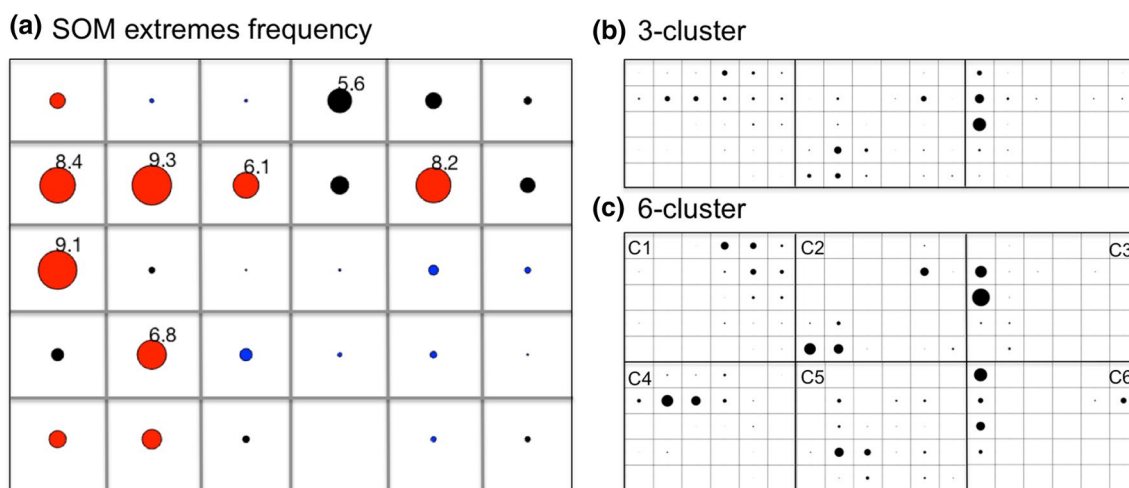
series (Individual SOM assignments by date are detailed in Online Resource 1). The pattern-space separates into deep troughs toward the left of the pattern-space, shallower troughs along the top of pattern-space, and ridge-like patterns to the upper right of the pattern-space. Certain SOM nodes are more representative of extreme precipitation days (red dots in Fig. 9a), particularly along the left side (SOMs 1, 7, 8, 9, 13, 20, 25, 26) and upper right (SOMs 4 and 11) of the pattern-space. Of these, 7 SOMs (those with percentages noted above the red dots in Fig. 9a) represent 53.5% of the extreme days. Several of these nodes (SOMs 7, 13, 20, 26) feature nearly or more than double the number of extreme days than expected due to chance. Conversely, few extreme precipitation days are represented by the patterns in the lower right of the pattern-space (blue dots in Fig. 9a). However, it is important to note that extreme precipitation can and does occur in any SOM node, with the exception of SOM 28.

To investigate how the SOM extreme precipitation patterns relate to the KMC patterns, the percent of each KMC cluster’s dates that maps into the SOM pattern-space is evaluated. For the  $k=3$  solution (Fig. 9b), the ridge pattern is more frequently represented by a number of upper SOM nodes, the eastern trough by four lower left SOM nodes, and the central trough by three upper left SOM nodes. For the  $k=6$  solution (Fig. 9c), the ridge (C1) is represented by the upper right of the pattern-space, while the trough/ridge combination (C4) is represented by the upper left of the pattern-space. The wintertime eastern trough (C2) is represented by two of the four SOMs associated with the eastern trough for the  $k=3$  solution, while the summertime eastern trough (C5) is associated with the other two. Similarly, the

spring/fall Ohio Valley trough (C3) is represented by the lower two of the three SOMs associated with the  $k=3$  Ohio Valley trough, while the DJF version is most represented by the uppermost SOM. Again, the KMC  $k=6$  solution provides a more nuanced separation of the patterns, which is reflected in the SOM pattern-space, and provides insight into how the circulation associated with extreme precipitation may be different than that for non-extreme precipitation days.

### 4.3 Further insights using KMC and SOM analysis together

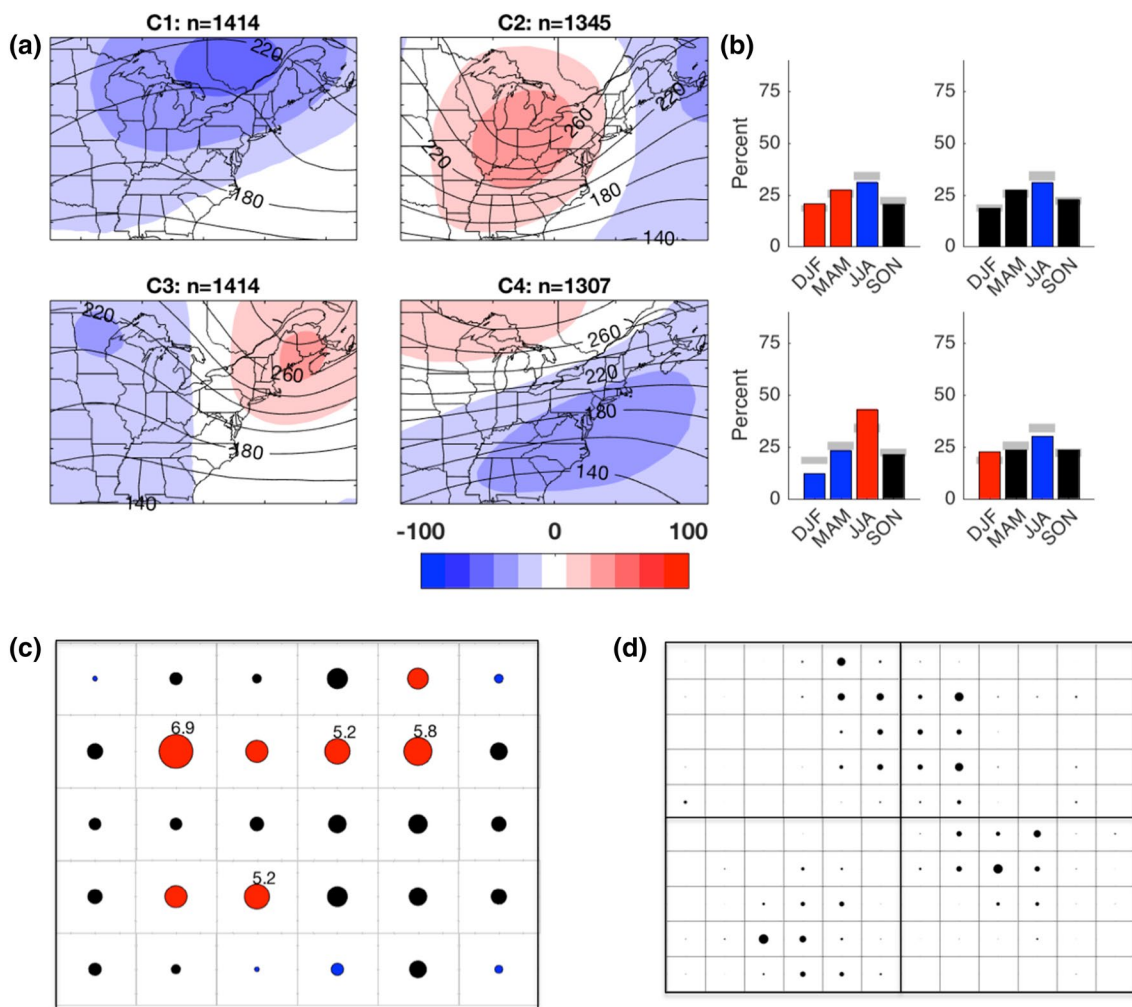
To see whether the LSMPs associated with extreme precipitation are significantly different than ordinary precipitation circulation patterns, we also conduct KMC analysis on interquartile, or “typical” precipitation days. The optimal KMC solution involves four patterns (Fig. 10a, b): a wintertime ridge, a shallow Ohio Valley trough, a summertime ridge/trough combination, and a wintertime very shallow southern Canadian trough. Based on the relative frequency of interquartile precipitation days within the SOM pattern-space, we find that interquartile precipitation is more frequent in the central patterns (Fig. 10c), as opposed to edge patterns more associated with extreme precipitation. However, the mapping of the KMC interquartile patterns into the SOM pattern-space (Fig. 10d) shows a broad but non-overlapping coverage of the entire pattern-space, with the exception of the corner SOMs (SOMs 1, 6, 19, 30). Hence there is nothing qualitatively or markedly different about the KMC patterns associated with extremes; indeed these same patterns reflected in the SOM pattern-space also



**Fig. 9** a Frequency of top 1% extreme precipitation days accounted for by each SOM pattern, expressed as dots with radius proportional to value, where red dots show values higher than that expected due to chance, and blue dots show lower values than that expected due

to chance. Values over 5% are indicated. The right panel b shows the SOM frequency (blue dots, with radius proportional to value) of the top 1% extreme precipitation dates for each of the  $k=3$  k-means results, and c shows the same for the  $k=6$  k-means solution





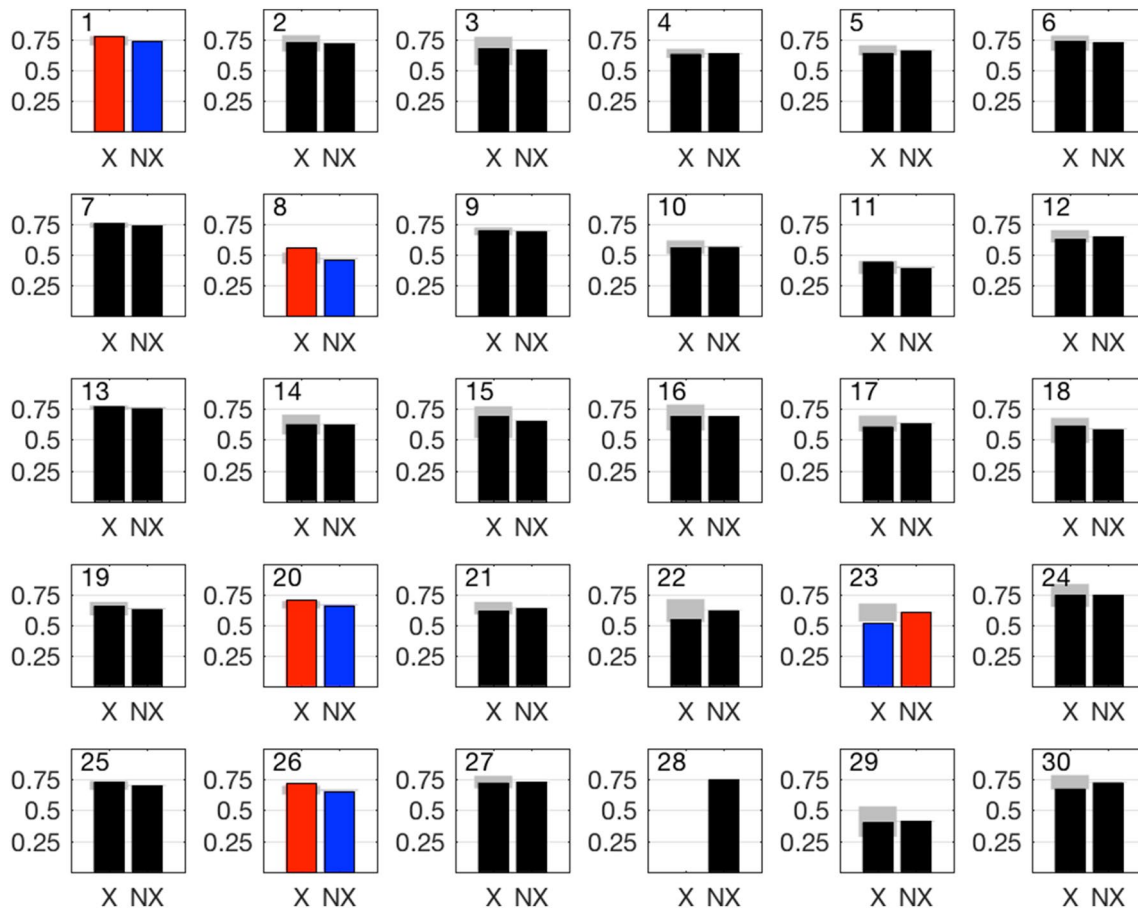
**Fig. 10** Similar to Figs. 4 and 9, except for k-means separation of MERRA daily mean DT pressure anomalies for interquartile precipitation dates (1979–2008) into four clusters. **a** The composite DT pressure patterns for each cluster, **b** the seasonal frequency (expressed as

a percent) of each cluster **c** the frequency (expressed as a percent) of interquartile days associated with each SOM pattern, and **d** the frequency (expressed as a percent) of each cluster’s dates accounted for by each SOM pattern

generate ordinary precipitation; however not all ordinary precipitation-generating patterns regularly produce extreme precipitation.

This motivates our next question: is there anything particularly “extreme” about the SOM patterns themselves on extreme days versus ordinary precipitation days? That is, to the extent that LSMPs can “explain” extreme precipitation, are there distinguishing features of each SOM pattern that occur on extreme-precipitation days and do not occur on ordinary precipitation days? To see how well the extreme precipitation patterns represent individual extreme days, we calculate the mean spatial correlation of individual days to their assigned SOM pattern (Fig. 11), for both extreme and non-extreme days. For many of the trough patterns (left of pattern-space), the spatial correlation is slightly higher on extreme days than for non-extreme days, although this

result is only significant at the 0.05 level for SOMs 1, 8, 20, and 26, which correspond to KMC trough patterns C2, C4, C5, and C6. Many of the ridge patterns (upper right of pattern space) that correspond to KMC pattern C1 also show slightly higher correlations on extreme days, although these results are not significant at the 0.05 level. For the lower right of the pattern space, which represents fewer extreme precipitation days, correlations are smaller on extreme precipitation days (significant for SOM 23). This indicates that the pattern features (troughs/ridges/zonal flow) are disproportionately influenced by the majority of extreme precipitation days, despite the significantly smaller number of those days. To explore this further, Fig. 12 shows the extreme precipitation day tropopause mean, non-extreme precipitation day tropopause mean, and difference of tropopause means (extreme day mean minus non-extreme day



**Fig. 11** Mean spatial correlation of individual days to corresponding SOM patterns for extreme precipitation days (X) and non-extreme precipitation days (NX). Bars are shaded red (blue) if the mean corre-

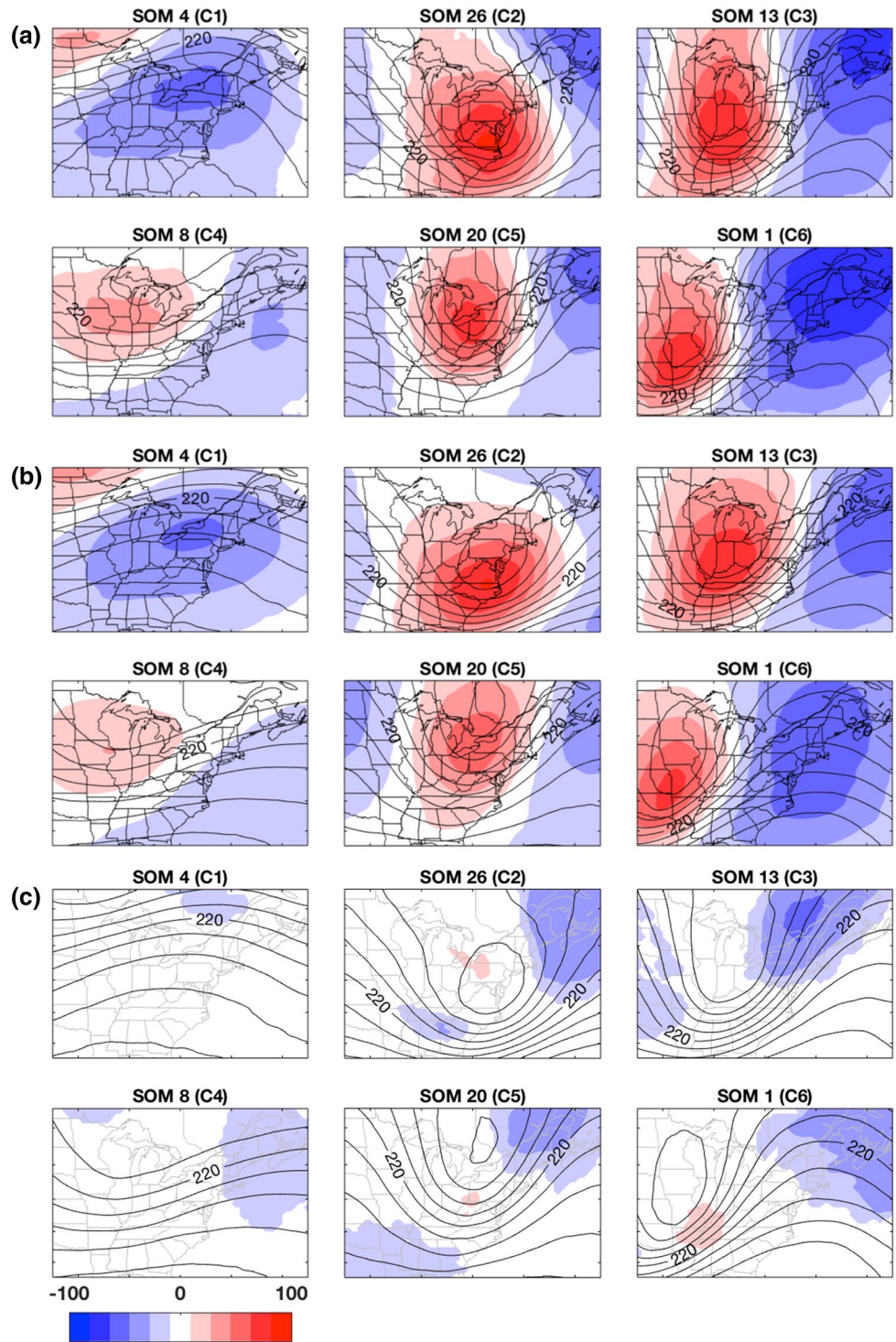
lation is significantly greater (less) than that expected due to chance, at the 0.05 level, using Monte Carlo analysis. The grey shading indicates the 95% confidence interval of means due to chance

mean, with only differences significant to the 0.05 level from Monte Carlo analysis shown) for six SOM nodes most associated with the KMC clusters C1–C6. For the eastern and Ohio Valley troughs, the patterns appear enhanced on extreme days (i.e. deeper troughs, more amplified ridges). For SOM 26, related to KMC C2, the trough is also more regressive on extreme days, which may cause longer duration precipitation. For the ridge patterns, there is enhanced ridging over NE on extreme days, particularly for SOM 8 (KMC C4) across northern New York State. From this we can argue that there are in fact “extreme” spatial distinctions, at least in the composites, of LSMPs for extreme precipitation days. However, the nature of the causal relationship is not clear. Does extreme precipitation lead to these pattern distinctions, or do the pattern distinctions provide a favorable environment for extreme precipitation, or does some combination of both occur? In a purely dynamical sense, the enhanced ridges/troughs could be a sign of blocking or other large-scale circulation factors that provide a favorable environment for extreme precipitation, or

long-duration precipitation. However, the enhanced ridging could actually be due to thermodynamic factors—for example, low-level diabatic heating can lead to destruction of tropopause potential vorticity, and hence tropopause ridging. Additionally, for some extreme precipitation cases, the LSMP may in fact be quite “ordinary”, while the thermodynamics (moisture and instability) may be “extraordinary”. The extent to which variations in the tropopause pattern itself or variations in other related thermodynamic features that may be inherent to the patterns are responsible for extreme precipitation is the subject of ongoing research, and is also briefly discussed in Sect. 4.4.

Another advantage to using KMC and SOM analysis in conjunction is that it allows an assessment of whether the LSMPs that generate extremes tend to be persistent or transient (that is, whether extreme precipitation is related to pattern duration). In most cases, the circulation remains in a single SOM pattern for a single day only (on average over 80% of the time), with an average duration of 1.28 days. Figure 13 shows the mean duration of

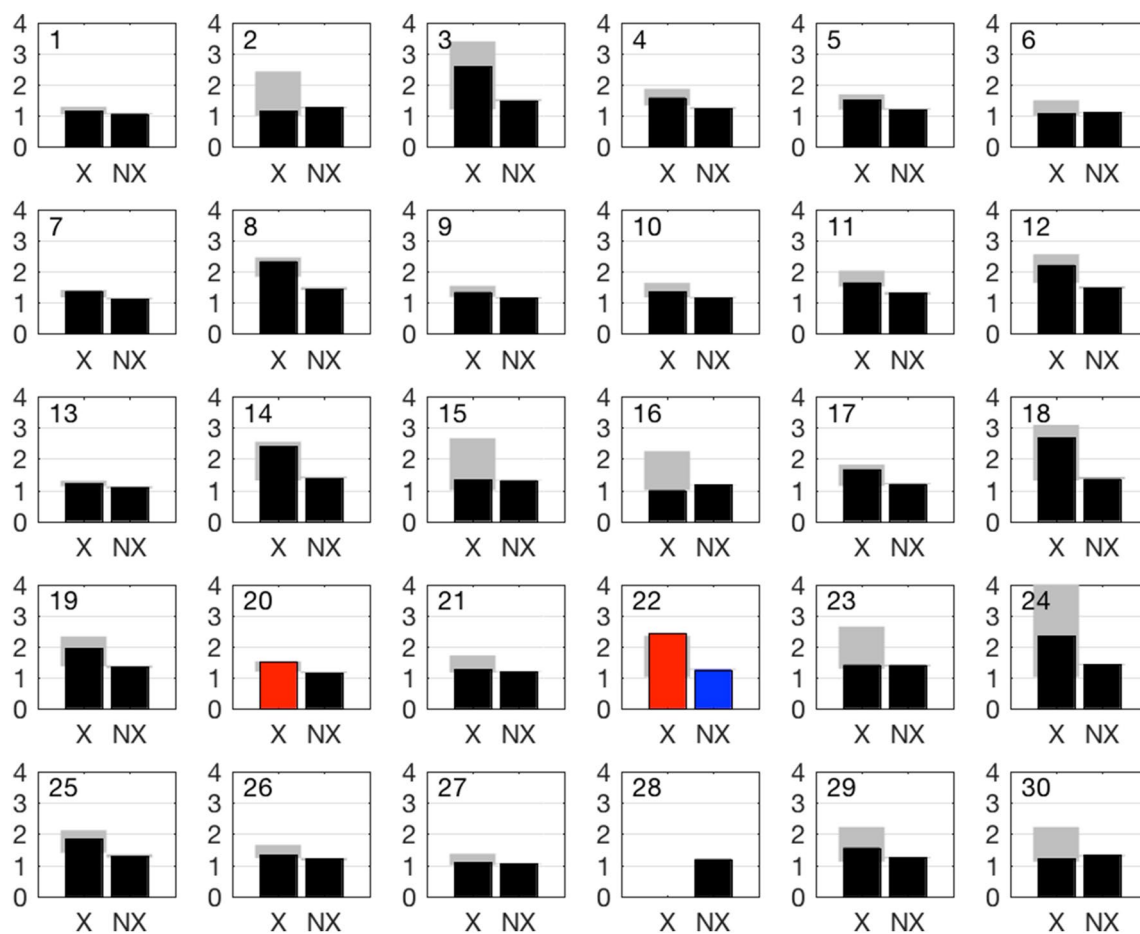
**Fig. 12** DT pressure (contours, in 20-hPa increments) and anomalies (shaded, hPa) for **a** extreme days, **b** non-extreme days, and **c** difference between extreme and non-extreme days for specific SOM nodes. For **c**, the SOM node with the highest number of extreme days for each of the KMC patterns is used (the associated KMC node is noted in panel title), and only differences significant at the 0.05 level based on Monte Carlo analysis are shown



extreme and non-extreme days for each of the SOM patterns. The Ohio Valley trough patterns (SOMs 1, 7, 13) show some of the shortest durations in the pattern-space, while the ridges show some of the longest durations. Duration tends to be longer for events involving extreme days, although this result is only significant for SOM 20 (the summertime eastern trough) and SOM 22. Based on

this, extreme precipitation may be related to the persistence of a circulation regime (longer duration events, as opposed to more intense rainfall). This is consistent with the findings of Gutowski et al. (2008), in which extreme precipitation in the upper Midwest occurs more often with slower-moving troughs.





**Fig. 13** Mean pattern duration (in days) for each SOM, for strings of extreme precipitation days (X) and non-extreme precipitation days (NX). Bars are shaded red (blue) if the mean duration is significantly greater (less) than that expected due to chance, at the 0.05 level, using Monte Carlo analysis. Grey shading indicates the 95% confidence

interval of mean duration due to chance. A string of consecutive days in a pattern is counted in the X category if at least one of the consecutive days features extreme precipitation; otherwise it is counted in the NX category

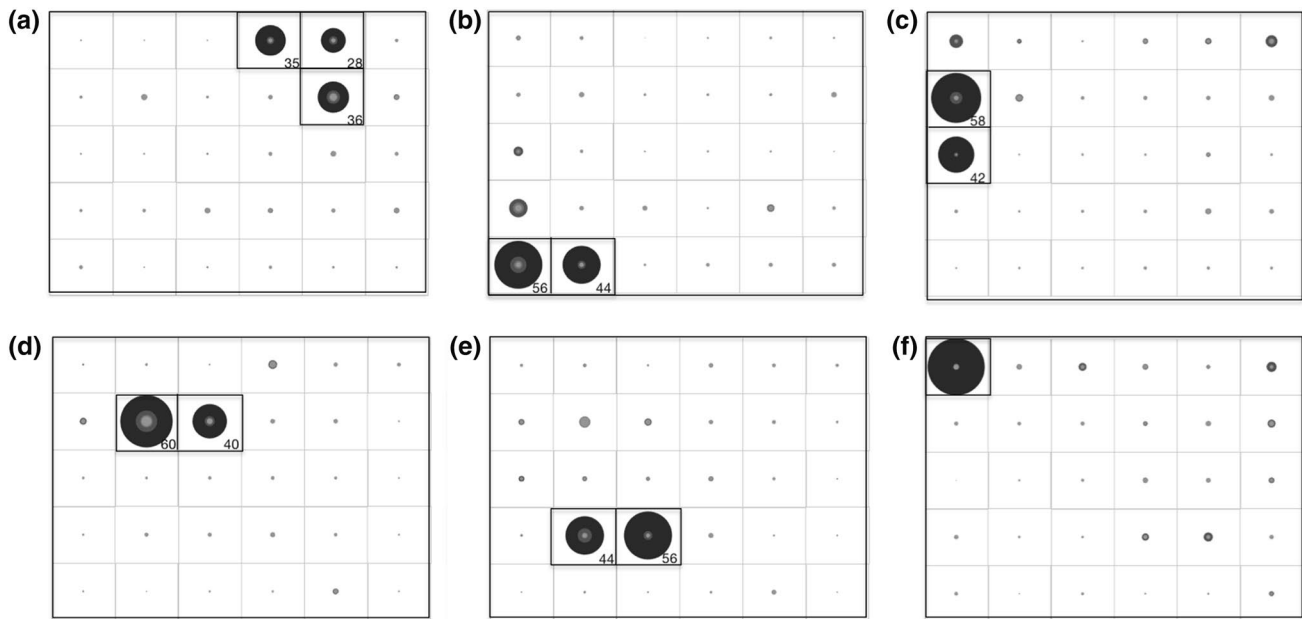
SOMs can also be useful in gauging whether extreme precipitation days uniquely precede or follow other patterns. They can also be used to determine whether extreme days tend to transition to other extreme days (that is, to see how chronologically separated are the extremes). Figure 14 illustrates this approach for each of the  $k=6$  clusters of extreme precipitation days. The starting (background) frequency of extreme days assigned to each of the most extreme-generating SOMs (thick black outlines) is shown as a dark circle with radius proportional to the frequency. The SOM patterns assigned to these extreme dates 24 h earlier are shown as lighter circles with radius proportional to the frequency. Since in each of these cases the frequency will be less than that of the day before, the  $t-24$  h circles appear to the inside of the first circles. Similarly, the  $t-48$  h and  $t-72$  h SOM frequencies are shown as progressively lighter circles. From this figure it can be inferred that in general, extreme day patterns can originate in almost

any part of the pattern-space within 72 h of the extreme, but broad ridges (Fig. 14a) tend to originate in the right-most pattern space, wintertime eastern troughs (Fig. 14b) tend to transition from deep Ohio Valley troughs, Ohio Valley troughs (Fig. 14c, f) tend to originate from the rightmost part of the SOM pattern-space (ridges), and trough/ridge combinations (Fig. 14d) and summertime eastern troughs (Fig. 14e) can begin in almost any part of the SOM pattern space up to 72 h ahead of an extreme. These figures also show that the deep Ohio Valley troughs tend to transition quickly away from the original pattern (as noted in Fig. 13), while the other patterns show a tendency to persist.

#### 4.4 Brief overview of the 3D circulation

Here we provide a brief overview of 500-hPa geopotential heights and MSLP, vertical velocity anomalies, and integrated vapor transport (IVT) associated with each of the six





**Fig. 14** SOM node pattern transitions for 72, 48, 24, and 0 h before extreme precipitation days, shown as percentages of extreme days for **a** C1, **b** C2, **c** C3, **d** C4, **e** C5, and **f** C6. Circle size is proportional to percent of cluster extreme days. The primary SOM nodes correspond-

ing to C1–C6 are outlined, with the starting proportion (at  $t=0$ ) indicated for each node if there is more than one primary node. Shades of grey indicate the temporal relation to the extreme day, from the darkest shade for  $t=0$  to the lightest shade for  $t=72$  h

tropopause patterns (Fig. 15). We anticipate that in many cases, the LSMPs themselves may not directly influence the generation of extreme precipitation, but rather provide a favorable environment for other processes such as moisture advection, diabatic heating, convection, and mesoscale lift to generate extremes. As expected, the 500-hPa height patterns closely resemble the tropopause patterns, although the circulation for the eastern wintertime trough C2 is not closed as it is for the tropopause. Surface lows are evident to the east of the upper-level troughs for each of the trough patterns C2, C3, C5, and C6. For the ridge pattern C1, there is a surface disturbance eastward of the Great Lakes that may be related to air mass differences (thermal contrasts). For each pattern type there is an area of anomalously high upward motion over the NE, which is more intense and localized for the wintertime troughs, likely cause by synoptic-scale quasi-geostrophic (QG) forcing. However, even under the tropopause ridge, there is anomalous upward motion. Although synoptic-scale QG forcing cannot be ruled out, the upward motion may also be related to widespread convection, or frontal dynamics. Moisture availability and transport is indicated by the IVT composites. For the trough patterns, there is ocean-enhanced water vapor transport into the NE from the south or southeast, related to the warm fronts associated with the surface lows. For the ridge patterns, there is moisture transport from the west and southwest, related to high pressure in place to the southeast. Extreme precipitation occurs due to a combination of

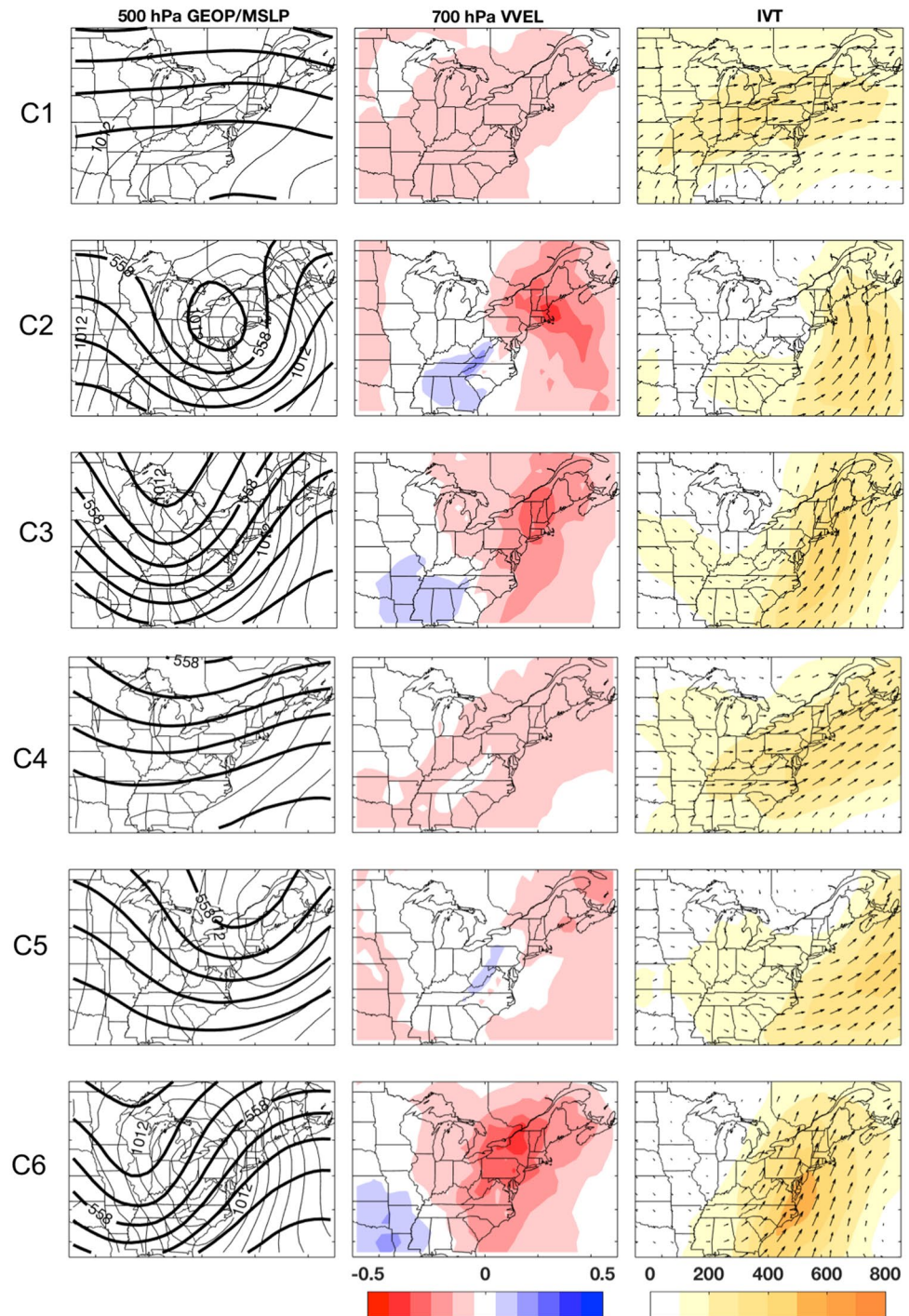
variations in upward motion and moisture availability. For each of these six tropopause patterns, ongoing research will identify the specific mechanisms and key ingredients that support extreme precipitation, whether related to synoptic forcing, frontal dynamics, moisture availability and delivery, or air mass stability.

### 5 Summary and discussion

In this study, we combine both k-means clustering (KMC) and Self-Organizing Maps (SOMs) to separate DT pressure anomalies into a small functional set of patterns, or LSMPs, that describe the large-scale circulation on non-tropical cyclone-related extreme precipitation days in the Northeast US. We use DT pressure anomalies for typing, since the dynamic tropopause captures both upper-level movements and is dynamically related to lower-level features, such as mid-level troughs and deepening cyclones. We achieve consistent results using this field, even when typing seasons separately, or modifying the time and space domains. The results of the KMC and SOM analyses are available in Online Resource 1.

We identify six LSMPs related to extreme precipitation. The first encompasses a strong ridge of high pressure that is most frequent in JJA. A second weaker ridge pattern is part of a shallow trough/ridge extending across southern Canada and northern New England, and is also active during

**Fig. 15** Composites of 500-hPa geopotential heights (*left column, thick contours, in 6-dm increments*) and MSLP (*left column, thin contours, in 2-hPa increments*), vertical velocity anomalies (*middle column, shaded,  $\text{Pa s}^{-1}$* ), and integrated moisture transport (*right column, shaded,  $\text{kg m}^{-1} \text{s}^{-1}$* ) for KMC clusters C1–C6



JJA. The third pattern features a deep, negatively-tilted wintertime trough across the eastern US; and the fourth pattern represents the JJA version of this trough (shallower and more progressive). The fifth and sixth patterns feature deep cold-season troughs extending from the Great Lakes and Ohio Valley to the southern tier states: the deepest trough is more active during spring and fall, and the weaker trough (representing the fewest extremes of all the patterns) is more active during DJF.

Four of the six extreme precipitation day patterns feature an upper-level trough over or near the NE. These patterns are associated with the passage of synoptic storms, and the resulting extreme precipitation is related to strong vertical ascent coupled with an abundant moisture feed from the south and southwest. Although each of these patterns feature troughs, they differ in location, strength, and seasonality, as well as amount of moisture availability and lift. Interestingly, 43.7% of the extreme

precipitation days are associated with the two summertime ridge patterns. Upon initial consideration, this may seem a surprising result. However, it is partly related to the annual top 1% definition of extreme precipitation—there are simply more warm-weather extreme precipitation days (in large part due to increased moisture capacity in the warmer environment). These extreme precipitation days may be related to convective activity associated with warm, moist unstable air (e.g. localized thunderstorms), or be linked to frontal dynamics not directly associated with nearby extratropical storms or tropical cyclones, such as elongated cold fronts and mesoscale convective complexes (Kunkel et al. 2012). In the case of frontal dynamics, extreme precipitation can occur on the cold side of mesoscale convective complex outflow boundaries or the cold side of stationary thermal fronts (Schumacher and Johnson 2005). Shortwave vorticity maxima propagating through longwave ridges may also provide a trigger for convection within these ridge patterns (Milrad et al. 2014). A detailed dynamical analysis of each pattern is the subject of ongoing research, including investigations into circulation features at other levels, vertical motion, moisture transport and moisture availability, stability, association with synoptic storms, and frontal processes.

We end with a few words regarding our methodology of combining both KMC and SOM analyses. While either algorithm by itself can be a useful tool, here we use KMC to identify the patterns associated with extreme precipitation days, and SOM analysis to place these patterns into the larger context of all days. This allows us also to test the sensitivity of the identified patterns to the analysis technique. We find that optimal measures of the lowest  $k$  for the KMC technique may not provide enough fine detail, especially when performing annual analyses; secondary measures of optimum  $k$  may be more useful for finding meaningful pattern separations that compactly take into account seasonality and spatial distinctions. SOM analysis (using a  $5 \times 6$  pattern-space) on all days confirms the six patterns related to extreme precipitation. We then use the SOM analysis in conjunction with the KMC results to determine how extreme precipitation large-scale circulation differs from that of other days, and whether there are distinct markers for antecedent or subsequent patterns. We can also ascertain whether the patterns themselves are extreme in any way on extreme precipitation days. Our primary results from this combined KMC/SOM methodology are:

- There are only a few patterns that are frequently associated with extreme precipitation (7 SOM patterns explain 53.5% of the extreme precipitation day KMC patterns).
- The patterns that are related to extreme precipitation can also occur for ordinary precipitation; however, the patterns most associated with ordinary precipitation rarely produce extreme precipitation.
- Many of the patterns related to extreme precipitation days persist longer for extreme precipitation than for ordinary precipitation (although this result is only significant for two of the patterns), suggesting duration as a factor in rainfall totals, even for ridge patterns.
- For the SOM trough patterns most associated with the KMC patterns, the trough patterns tend to have deeper troughs and stronger downstream ridging on extreme precipitation days than for non-extreme precipitation days. Correlations on extreme days are slightly higher than for non-extreme days, despite contributing only around 6% of the days, confirming the more exaggerated DT troughs/ridges on extreme days.

Although the primary purpose of this study is to identify LSMPs associated with extreme precipitation in the NE, which will allow us to further understand and identify the key ingredients that lead to extreme precipitation in the NE, we anticipate that the techniques used here can be applied to other regions and definitions of extreme events.

**Funding** Funding provided by National Science Foundation (NSF Project #1623912).

## References

- Agel L, Barlow M, Qian J-H, Colby F, Douglas E, Eichler T (2015) Climatology of daily precipitation and extreme precipitation events in the northeast United States. *J Hydrometeorol* 16:2537–2557. doi:10.1175/JHM-D-14-0147.1
- Barlow M (2011) Influence of hurricane-related activity on North American extreme precipitation. *Geophys Res Lett* 38:L04705. doi:10.1029/2010gl046258
- Bosart LF (1999) Observed cyclone life cycles. In: Shapiro MA, Grønås S (eds) *The Life cycles of extratropical cyclones*. American Meteorological Society, Boston, pp 187–213. doi:10.1007/978-1-935704-09-6\_15
- Cassano JJ, Uotila P, Lynch AH, Cassano EN (2007) Predicted changes in synoptic forcing of net precipitation in large Arctic river basins during the 21st century. *J Geophys Res Biogeos* 112:G04549. doi:10.1029/2006JG000332
- Cassano EN, Glisan JM, Cassano JJ, Gutowski WJ, Jr., Seefeldt MW (2015) Self-organizing map analysis of widespread temperature extremes in Alaska and Canada. *Clim Res* 62:199–218
- Catto JL, Pfahl S (2013) The importance of fronts for extreme precipitation. *J Geophys Res Atmos* 118:10791–10801. doi:10.1002/jgrd.50852
- Cavazos T (1999) Large-scale circulation anomalies conducive to extreme precipitation events and derivation of daily rainfall in northeastern Mexico and Southeastern Texas. *J Clim* 12:1506–1523
- Chen M, Xie P, Co-authors (2008) CPC unified gauge-based analysis of global daily precipitation, Western Pacific Geophysics Meeting, Cairns, Australia, 29 July–1 August, 2008. From



- NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. <http://www.esrl.noaa.gov/psd>
- Collow ABM, Bosilovich MG, Koster RD (2016) Large-scale influences on summertime extreme precipitation in the northeastern United States. *J Hydrometeorol* 17:3045–3061. doi:[10.1175/jhm-d-16-0091.1](https://doi.org/10.1175/jhm-d-16-0091.1)
- Davis CA, Emanuel KA (1991) Potential vorticity diagnostics of cyclogenesis. *Mon Weather Rev* 119:1929–1953. doi:[10.1175/1520-0493](https://doi.org/10.1175/1520-0493)
- Diday E, Simon JC (1976) Clustering analysis. In: Fu KS (ed) *Digital pattern recognition*. Springer, Berlin, pp 47–94. doi:[10.1007/978-3-642-96303-2\\_3](https://doi.org/10.1007/978-3-642-96303-2_3)
- Douglas EM, Fairbank CA (2011) Is precipitation in northern New England becoming more extreme? Statistical analysis of extreme rainfall in Massachusetts, New Hampshire, and Maine and updated estimates of the 100-Year Storm. *J Hydrol Eng* 16:203–217. doi:[10.1061/\(asce\)he.1943-5584.0000303](https://doi.org/10.1061/(asce)he.1943-5584.0000303)
- Favre A, Gershunov A (2009) North Pacific cyclonic and anticyclonic transients in a global warming context: possible consequences for Western North American daily precipitation and temperature extremes. *Clim Dyn* 32:969–987. doi:[10.1007/s00382-008-0417-3](https://doi.org/10.1007/s00382-008-0417-3)
- Feldstein SB, Lee S (2014) Intraseasonal and interdecadal jet shifts in the northern hemisphere: the role of warm pool tropical convection and sea ice. *J Clim* 27:6497–6518. doi:[10.1175/JCLI-D-14-00057.1](https://doi.org/10.1175/JCLI-D-14-00057.1)
- Glisan JM, Gutowski WJ (2014a) WRF winter extreme daily precipitation over the North American CORDEX Arctic. *J Geophys Res Atm* 119:10, 738–710, 748. doi:[10.1002/2014JD021676](https://doi.org/10.1002/2014JD021676)
- Glisan JM, Gutowski WJ (2014b) WRF summer extreme daily precipitation over the CORDEX Arctic. *J Geophys Res Atm* 119:1720–1732. doi:[10.1002/2013JD020697](https://doi.org/10.1002/2013JD020697)
- Glisan JM, Gutowski WJ, Cassano JJ, Cassano EN, Seefeldt MW (2016) Analysis of WRF extreme daily precipitation over Alaska using self-organizing maps. *J Geophys Res Atm* 121:7746–7761. doi:[10.1002/2016JD024822](https://doi.org/10.1002/2016JD024822)
- Griffiths ML, Bradley RS (2007) Variations of twentieth-century temperature and precipitation extreme indicators in the northeast United States. *J Clim* 20:5401–5417
- Groisman PY, Knight RW, Zolina OG (2013) Recent trends in regional and global intense precipitation patterns. In: Sr. RP, Hossain F (eds) *Climate vulnerability: understanding and addressing threats to essential resources*. Volume 5, vulnerability of water resources to climate. Elsevier Publishing House, Amsterdam, pp 25–55
- Grotjahn R et al (2016) North American extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Clim Dyn* 46:1151–1184. doi:[10.1007/s00382-015-2638-6](https://doi.org/10.1007/s00382-015-2638-6)
- Gutowski WJ, Otieno FO, Arritt RW, Takle ES, Pan Z (2004) Diagnosis and attribution of a seasonal precipitation deficit in a U.S. regional climate simulation. *J Hydrometeorol* 5:230–242. doi:[10.1175/1525-7541\(2004\)005<0230:DAOAS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0230:DAOAS>2.0.CO;2)
- Gutowski WJ, Willis SS, Patton JC, Schwedler BRJ, Arritt RW, Takle ES (2008) Changes in extreme, cold-season synoptic precipitation events under global warming. *Geophys Res Lett* 35:L20710. doi:[10.1029/2008GL035516](https://doi.org/10.1029/2008GL035516)
- Hawcroft MK, Shaffrey LC, Hodges KI, Dacre HF (2012) How much Northern Hemisphere precipitation is associated with extratropical cyclones? *Geophys Res Lett* 39:L24809. doi:[10.1029/2012gl053866](https://doi.org/10.1029/2012gl053866)
- Hewitson BC, Crane RG (2002) Self-organizing maps: applications to synoptic climatology. *Clim Res* 22:13–26
- Higgins RW, Kousky VE, Xie P (2011) Extreme precipitation events in the south-central United States during May and June 2010: historical perspective, role of ENSO, and trends. *J Hydrometeorol* 12:1056–1070. doi:[10.1175/jhm-d-10-05039.1](https://doi.org/10.1175/jhm-d-10-05039.1)
- Hoskins BJ, McIntyre ME, Robertson AW (1985) On the use and significance of isentropic potential vorticity maps. *Q J R Meteorol Soc* 111:877–946. doi:[10.1002/qj.49711147002](https://doi.org/10.1002/qj.49711147002)
- Jones C, Waliser DE, Lau KM, Stern W (2004) Global occurrences of extreme precipitation and the Madden–Julian oscillation: observations and predictability. *J Clim* 17:4575–4589
- Kohonen T (2001) *Self-organizing maps*. Springer, New York
- Kunkel KE, Easterling DR, Kristovich DAR, Gleason B, Stoecker L, Smith R (2012) Meteorological causes of the secular variations in observed extreme precipitation events for the Conterminous United States. *J Hydrometeorol* 13:1131–1141. doi:[10.1175/jhm-d-11-0108.1](https://doi.org/10.1175/jhm-d-11-0108.1)
- Kunkel K et al (2013) *Regional Climate Trends and Scenarios for the U.S. National Climate Assessment. Part 1. Climate of the Northeast U.S.*
- Landsea CW, Franklin JL (2013) Atlantic hurricane database uncertainty and presentation of a new database format. *Mon Weather Rev* 141:3576–3592. doi:[10.1175/MWR-D-12-00254.1](https://doi.org/10.1175/MWR-D-12-00254.1)
- Martius O, Schwierz C, Davies HC (2010) Tropopause-Level waveguides. *J Atmos Sci* 67:866–879. doi:[10.1175/2009JAS2995.1](https://doi.org/10.1175/2009JAS2995.1)
- Melillo JM, Richmond TC, Yohe GW (2014) Climate change impacts in the United States: the third national climate. Assessment. doi:[10.7930/J0Z31WJ2](https://doi.org/10.7930/J0Z31WJ2)
- Michelangeli P-A, Vautard R, Legras B (1995) Weather regimes: recurrence and quasi stationarity. *J Atmos Sci* 52:1237–1256. doi:[10.1175/1520-0469\(1995\)052<1237:WRRASQ>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1237:WRRASQ>2.0.CO;2)
- Milrad SM, Atallah EH, Gyakum JR (2010) Synoptic typing of extreme cool-season precipitation events at St. John's, Newfoundland, 1979–2005. *Weather Forecast* 25:562–586
- Milrad SM, Atallah EH, Gyakum JR, Dookhie G (2014) Synoptic typing and precursors of heavy warm-season precipitation events at Montreal, Québec. *Weather Forecast* 29:419–444 doi:[10.1175/WAF-D-13-00030.1](https://doi.org/10.1175/WAF-D-13-00030.1)
- Morgan MC, Nielsen-Gammon JW (1998) Using tropopause maps to diagnose midlatitude weather systems. *Mon Weather Rev* 126:2555–2579. doi:[10.1175/1520-0493\(1998\)126<2555:UTMTDM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<2555:UTMTDM>2.0.CO;2)
- Nielsen-Gammon JW (2001) A visualization of the global dynamic tropopause. *Bull Am Meteorol Soc* 82:1151–1167. doi:[10.1175/1520-0477\(2001\)082<1151:AVOTGD>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<1151:AVOTGD>2.3.CO;2)
- Peterson TC et al (2013) Monitoring and understanding changes in heat waves, cold waves, floods, and droughts in the United States: state of knowledge. *Bull Am Meteorol Soc* 94:821–834. doi:[10.1175/BAMS-D-12-00066.1](https://doi.org/10.1175/BAMS-D-12-00066.1)
- Pfahl S, Wernli H (2012) Quantifying the relevance of cyclones for precipitation extremes. *J Climate* 25:6770–6780. doi:[10.1175/jcli-d-11-00705.1](https://doi.org/10.1175/jcli-d-11-00705.1)
- Pfahl S, Madonna E, Boettcher M, Joos H, Wernli H (2014) Warm Conveyor belts in the ERA-interim dataset (1979–2010). Part II: moisture origin and relevance for precipitation. *J Climate* 27:27–40. doi:[10.1175/JCLI-D-13-00223.1](https://doi.org/10.1175/JCLI-D-13-00223.1)
- Riddle EE, Stoner MB, Johnson NC, L'Heureux ML, Collins DC, Feldstein SB (2013) The impact of the MJO on clusters of wintertime circulation anomalies over the North American region. *Clim Dyn* 40:1749–1766. doi:[10.1007/s00382-012-1493-y](https://doi.org/10.1007/s00382-012-1493-y)
- Rienecker MM et al (2011) MERRA: NASA's Modern-Era retrospective analysis for research and applications. *J Clim* 24:3624–3648. doi:[10.1175/jcli-d-11-00015.1](https://doi.org/10.1175/jcli-d-11-00015.1)
- Roller CD, Qian J-H, Agel L, Barlow M, Moron V (2016) Winter weather regimes in the northeast United States. *J Clim* 29:2963–2980. doi:[10.1175/JCLI-D-15-0274.1](https://doi.org/10.1175/JCLI-D-15-0274.1)
- Sammon JW (1969) A nonlinear mapping for data structure analysis. *IEEE Trans Comput* C-18:401–409. doi:[10.1109/T-C.1969.222678](https://doi.org/10.1109/T-C.1969.222678)



- Santos JA, Corte-Real J, Ulbrich U, Palutikof J (2007) European winter precipitation extremes and large-scale circulation: a coupled model and its scenarios. *Theor Appl Climatol* 87:85–102
- Schumacher RS, Johnson RH (2005) Organization and environmental properties of extreme-rain-producing mesoscale convective systems. *Mon Weather Rev* 133:961–976. doi:[10.1175/MWR2899.1](https://doi.org/10.1175/MWR2899.1)
- Schumann MR, Roebber PJ (2010) The influence of upper-tropospheric potential vorticity on convective morphology. *Mon Weather Rev* 138:463–474. doi:[10.1175/2009MWR3091.1](https://doi.org/10.1175/2009MWR3091.1)
- Wheeler MC, Hendon HH (2004) An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Mon Weather Rev* 132:1917–1932. doi:[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)
- Williams CN, Vose RS, Easterling DR, Menne MJ (2004) United States historical climatology network daily temperature, precipitation, and snow data. ORNL/CDIAC-118, NDP-070. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee. [http://cdiac.ornl.gov/ftp/ushcn\\_daily](http://cdiac.ornl.gov/ftp/ushcn_daily)