JGR Atmospheres

RESEARCH ARTICLE
10.1029/2019JD031460

Large-Scale Atmospheric Circulation Patterns Associated With U.S. Great Plains Warm Season Droughts Revealed by Self-Organizing Maps

Yizhou Zhuang[1], **Rong Fu**[1], **and Hongqing Wang**[2]

[1]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA, USA, [2]Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China

Abstract This study uses a multivariate self-organizing map approach to diagnose precipitation anomalies over the United States' Great Plains during the warm season (April–August) and the associated anomalous large-scale atmospheric patterns, as represented by standardized anomalies of 500 hPa geopotential (Z500′), integrated vapor transport (IVT′), and convective inhibition index (CINi′). Circulation patterns favoring dryness identified by the method are generally consistent with those shown in previous studies, but this study provides a more comprehensive and probabilistic characterization of those that favor drought over the Southern Great Plains (SGP) and the Central Great Plains (CGP) and their temporal evolutions. Six circulation types that are associated with warm season rainfall variability over the Great Plains are identified. The SGP droughts are attributable to more frequent and persistent northern low-southern high as well as dominant high circulation types and are connected to larger negative CINi′. In contrast, CGP droughts are attributable to more frequent and persistent western low-eastern high, or northern high-southern low, or dominant high patterns, and are linked to a larger negative IVT′, but not larger CINi′. Thus, these results suggest that land surface dryness and a stable atmospheric boundary layer may play a more important role over the SGP than reduced moisture transport in warm season droughts, but reduced moisture transport may play a more important role than thermodynamic stability in droughts over the CGP.

Key Points:
- Multivariate self-organizing map reveals six circulation types related to warm season rainfall variability over the U.S. Great Plains
- SGP drought is linked to higher geopotentials either over the entire or southern Great Plains with strong negative convective inhibition energy
- CGP drought is attributable to higher geopotentials either over the entire, or northern, or western Great Plains and weak moisture transport

Supporting Information:
- Supporting Information S1

Correspondence to:
Y. Zhuang,
zhuangyz@atmos.ucla.edu

Citation:
Zhuang, Y., Fu, R., & Wang, H. (2020). Large-scale atmospheric circulation patterns associated with U.S. Great Plains warm season droughts revealed by self-organizing maps. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD031460. https://doi.org/10.1029/2019JD031460

Received 6 AUG 2019
Accepted 15 FEB 2020
Accepted article online 19 FEB 2020

1. Introduction

The United States' Great Plains (hereafter the Great Plains) covers a large portion of the interior flat North America and is a significant source of grain for the United States and global food markets and so is central for global food security (Kotkin et al., 2012). It is prone to extreme droughts, for example, the 1930s Dust Bowl, the 2011 drought over the Southern Great Plains (SGP), and the 2012 drought over the Central Great Plains (CGP) and Northern Great Plains (e.g., Basara et al., 2013; Hoerling et al., 2009; Hoerling et al., 2014; Seager & Hoerling, 2014). Most dynamic models show virtually no skill for seasonal prediction of precipitation during the warm season over the Great Plains (Quan et al., 2012) in part because the predictability provided by El Niño–Southern Oscillation and Atlantic sea surface temperature anomalies is relatively weak during summer compared to during winter (Roundy et al., 2014). Many previous studies have identified factors influencing precipitation, including oceanic variability, increase of Convective Inhibition energy (CIN), soil moisture feedback, boundary layer/lower tropospheric circulation anomalies, and anomalous anticyclonic flows/high pressure systems (e.g., Bronnimann et al., 2009; Fernando et al., 2016; Pu et al., 2016; Schubert et al., 2004; Trenberth et al., 1988). These studies have shown that the formation and evolution of anomalous circulation patterns and thermodynamic conditions responsible for droughts can vary between different events. The statistical distribution and evolution of these anomalous patterns and their relative influence on summer drought have not been clearly characterized, in part because conventional statistical approaches such as correlation, regression, or composite analysis are only suitable for identifying patterns based on correlations that dominate the entire analysis period or sample domain. They are not effective in identifying multiple patterns that represent only some of the analysis period or some samples, or multiple anomalous circulation patterns that cause similar precipitation anomalies. Thus, these widely used conventional statistical analyses are unable to represent the rich variety of anomalous circulation patterns associated with summer droughts.

Self-organizing map (SOM), also called Kohonen Map, is an unsupervised neural network first introduced by Kohonen (1990) and then applied to various disciplines. Since the late 1990s, SOM has been increasingly applied in atmospheric sciences (e.g., Ambroise et al., 2000; Cavazos, 2000; Cavazos et al., 2002; Hewitson & Crane, 1994; Hewitson & Crane, 2002; Malmgren & Winter, 1999; Mattingly et al., 2016). Reusch et al. (2005) compared SOM with principal component analysis (PCA) (or empirical orthogonal functions, EOFs) using synthetic North Atlantic sea level pressure data, showing that SOM is more robust and more capable of isolating all predefined patterns than PCA. Johnson et al. (2008) showed that SOM can reveal a continuum of sea level pressure patterns that correspond to the well-known teleconnection patterns. Swales et al. (2016) performed SOM analysis on the vertically integrated vapor transport (IVT) field and identified two synoptic settings related to extreme precipitation in the U.S. West. Rodriguez-Morata et al. (2018) characterized synoptic patterns related to hydrogeomorphic processes in Peru with different El Niño types as well as neutral El Niño–Southern Oscillation conditions using a SOM constructed from 200 hPa geopotential and 850 hPa specific humidity fields. Huang et al. (2017) constructed a SOM from sea level pressure, sea surface temperature, and surface temperature to investigate the combined effect of Arctic Oscillation and El Niño–Southern Oscillation on the wintertime surface temperature anomalies over East Asia. Further applications of SOM in atmospheric and oceanic sciences can be found in Liu and Weisberg (2011). However, to our knowledge, there is no study that has explored this approach to characterize drought related atmospheric conditions over the Great Plains.

Building upon previous drought related studies and applications of SOM in meteorology, we apply a multivariate SOM approach to clarify the following questions: (1) Can SOM capture not only the anomalous large-scale atmospheric patterns associated with the Great Plains summer droughts and wet anomalies that have been identified by previous studies but also those not discovered previously? (2) How do different patterns of anomalous circulation and thermodynamic conditions contribute to the anomalously dry and wet summers? (3) How do these patterns evolve over time?

The data and methods used in this work are described in section 2. The results of SOM analysis of the precipitation fields and classification of geographic regions are described in section 3. The anomalous atmospheric patterns related to warm season drought and anomalous wet conditions over the Great Plains identified by multivariate SOM and temporal evolution of these patterns are reported in section 4. The main conclusions and a brief discussion of the limitation of the SOM method and implication of our results are provided in section 5.

2. Data and Methods

2.1. Data Sets

The Climate Prediction Center (CPC) global unified gauge-based analysis of daily precipitation with 0.5° resolution and the European Centre for Medium-Range Weather Forecast Reanalysis-Interim (ERAI) 6-hourly outputs with 1.0° resolution are used in this study. The 2 m dew point temperature T_d(2 m) and 700 hPa temperature T(700 hPa) from the ERAI reanalysis are used to calculate the CIN index (CINi): T_d(2 m) − T(700 hPa) (Myoung & Nielsen-Gammon, 2010a). CINi can be regarded as the sum of two terms: surface dew point depression T_d(2 m) − T(2 m) and low-level lapse rate T(2 m) − T(700 hPa); this definition of CINi is opposite in sign to that used in Myoung and Nielsen-Gammon (2010a) to ensure that CINi is negative to match CIN's definition of negative buoyancy energy. Thus, a more negative CINi indicates a more stable lower troposphere. Previous studies (e.g., Fernando et al., 2016; Myoung & Nielsen-Gammon, 2010a) suggest that CINi is an approximation to the actual CIN value as calculated from sounding profiles and the parcel model method (Zhuang et al., 2018). Zonal wind (u), meridional wind (v), and specific humidity (q) from 300 to 1,000 hPa (20 vertical levels) are also obtained from the ERAI to calculate the IVT: $g^{-1}\int_{1000\text{hPa}}^{300\text{hPa}} q\sqrt{u^2+v^2}\,dp$; the last atmospheric variable from the ERAI is the 500 hPa geopotential (Z500).

In order to have consistent spatial and temporal resolution in both the CPC precipitation and ERAI reanalysis, we average the 0.5° CPC daily precipitation data to 1.0° resolution using the bilinear interpolation method and calculate daily means from the 6-hourly ERAI reanalysis data (00 to 18 UTC). A simple 7 day moving average filter is applied to all data sets to reduce high-frequency weather noise. To account for

dependence of grid point distance on latitude, all data are weighted by the square root of cosine of latitude before use of other data processing procedures. All data are transformed to standardized anomalies before entering the SOM analysis, which will be introduced later. Anomalies are defined as departures from daily climatology for the 1981–2010 period, and standardized anomalies are calculated by dividing the anomalies by the standard deviation of their daily average for the same period. The daily climatological mean and standard deviation used to calculate (standardized) anomalies are computed using samples within 15 days before/after the specific day of year, for example, the climatology of 1 April value is calculated using daily data from 17 March to 15 April during 1981–2010 ($31 \times 30 = 930$ days). This running mean approach reduces large daily fluctuations and so provides smooth daily climatological values. In addition, only the data during the warm season of the Great Plains (April to August, 153 days each year) in the 1979–2017 period (39 years) is used, resulting in a total of $153 \times 39 = 5{,}967$ daily samples. The target domain for this study is 25–53°N and 67–125°W, which covers the whole Contiguous United States (CONUS).

2.2. Methods

The analysis in this study is largely based on the SOM, or Kohonen Map (Kohonen, 1990). SOM is a powerful neural network approach that can effectively classify complex data matrices even when the data elements are related to each other nonlinearly. Details of the theoretical framework are documented in Kohonen (1990). For our application in this study, the basic idea of SOM is to find a prescribed number of patterns called nodes, such that on any given day, one of the patterns is a good fit to what is observed in that day. These nodes are found by starting with some first guess as to the nodes and iteratively improving that guess by comparison with observations. The algorithm of the SOM has been implemented in many computer languages. Here we only cover the basic steps of SOM and settings used for SOM training in MATLAB. In short, the training of SOM map occurs in the following steps and is repeated over many iterations.

1. Weights for all SOM nodes are initialized. A node is a vector that describes the data structure. The weight vector for node k is $W_k = \{w_{kj}\},\ k = 1,2,...,K,\ j = 1,2,...,M_v$, where K is the total node number and M_v is the number of elements of the SOM node, for example, the number of the grids in a spatial map.
2. A data vector $X_i = \{x_{ij}\}$ $(i = 1,2,...,N_t,\ j = 1,2,...,M_v)$ is selected at random from the input training data set, where N_t is the number of observations (time length in days). The Euclidian distance between this data vector and each node's weights are calculated as $D_{ik} = \sqrt{\sum_{j=1}^{N_v} \left(x_{ij} - w_{kj}\right)^2}$. The node with smallest distance is determined as the Best Matching Unit (BMU) for the data vector.
3. The neighborhood nodes of the BMU are calculated. The number of neighbors decreases over time as the training iteration step is increased.
4. Node weight for the BMU is adjusted so that the BMU becomes more like the input data vector. Weights of the BMU's neighbors are also adjusted, following the rule that the closer a neighbor is to the BMU, the more its weight is changed.
5. Step 2 to 4 are repeated for a number of times.

Several quantities are used to assess the representation of the SOM map for the input data sets. One is quantization error (QE), defined as average Euclidian distance between each data vector and its BMU node,

$$\mathrm{QE} = \frac{1}{N_t} \sum_i \min_k \{D_{ik}\}$$

Assuming that the Pearson correlation coefficient between the data vector X_i (e.g., a map at ith time step) and its BMU is ρ_i, then the mean correlation coefficient is defined as

$$\bar{\rho}^* = \frac{1}{N_t} \sum_{i=1}^{N_t} \rho_i$$

and the mean correlation coefficient for node k is defined as

$$\bar{\rho}_k = \sum_{\mathrm{BMU}(i)=k} \rho_i \Big/ \sum_{\mathrm{BMU}(i)=k} 1$$

Another quantity CC_{jk} is the correlation coefficient between node j and node k and the maximum correlation coefficient between the two nodes is defined as

$$CC_{max} = \max_{i \neq k} \{CC_{ik}\}$$

In this study, the SOM analysis is performed mainly using the "selforgmap" function in the Deep Learning Toolbox from MATLAB. Topology function "gridtop" and distance function "dist" are used throughout all SOM analysis in this study, with the initial neighborhood radius also set to default value 3. The principal component (PC) initialization function "initsompc" is used to initialize node weights so that the initial weights are distributed across the space spanned by the top two PCs of the inputs. This node weight initializing approach enables the SOM map to start out with a more reasonable pattern than other approaches like random initialization. As a result, the training of the SOM map can be finished in fewer training steps. We tested training steps of 200 to 2,000 (with increment of 50) for each SOM used below, and the result show that the differences of performance statistics introduced in above paragraphs for different training steps are much smaller than the difference of performance statistics for different SOM node numbers. Therefore, the training steps is set to a fixed value of 1,000.

For the SOM training algorithm, the batch learning algorithm is used so that the node weights are adjusted with the use of all input data vectors, instead of learning with one data vector at a time. In the "learnsomb" function, which is the MATLAB version of batch learning, there is an activation stage where only 90% of all input data vectors will be randomly selected and used to calculate the node weight changes. This process can reduce influence of outliers on SOM training. However, this activation process in batch learning also results in some randomness to the result; that is, the resulting SOM map will be slightly different for each time it is trained.

Based on the basic SOM approach, the analysis in this study includes the following major steps:

1. Identify geographic regions using SOM that share similar temporal patterns of rainfall variability over the CONUS. These precipitation anomalies over these regions are used to represent their dryness/wetness conditions. This step is done by using precipitation time series at grid points as data vectors for SOM training (spatial grid points are classified).
2. Identify general combinations of the standardized anomalies of Z500, IVT, and CINi (defined as Z500′, IVT′, and CINi′) to represent the large-scale atmospheric patterns over the CONUS. Once those patterns are classified, analyze how they are related to dryness/wetness conditions over the regions identified from Step (1). This step is done by using combined observation of Z500, IVT, and CINi of different days for SOM training (temporal points are classified). This step also includes a PCA-preprocessing technique to reduce the input data dimension and thus considerably speeds up the training process of SOM.
3. Identify the pattern transitions in general and during drought events. This step mainly analyzes the pattern transition frequency from the SOM result in Step (2).

2.2.1. Identify Geographic Regions Representing the Great Plains

Because precipitation variability can vary significantly over the target domain, and our first objective is to identify the regions with similar temporal variations of the warm season rainfall, we perform SOM analysis on the time series of the warm season precipitation anomaly $P_i = \{p_{ij}\}$ of all grid points ($i = 1, 2, \ldots, N_{lon} \times N_{lat}$) over the CONUS, where p_{ij} is the precipitation anomaly of grid point i at time j ($j = 1, 2, \ldots, N_t$). The input data matrix for the SOM has a size of (N_t, $N_{lon} \times N_{lat}$). In our case, $N_{lon} = 59$, $N_{lat} = 29$, and $N_t = 153 \times 39 = 5967$.

A K node SOM is trained to classify grid points in the spatial dimension. Node weight series

$$W_k = \{w_{kj}\}, \quad k = 1, 2, \ldots, K$$

or the average precipitation anomaly time series

$$P_k^* = \sum_{BMU(i)=k} P_i / \sum_{BMU(i)=k} 1 = \left\{ \sum_{BMU(i)=k} p_{ij} / \sum_{BMU(i)=k} 1 \right\}$$

can then be used to represent the general precipitation status. The nodes located near the Great Plains region

are used for further analysis. Our following analysis in section 3.1 will show that W_k and P_k^* are very similar, and either one of them can be used to represent regional dryness/wetness condition. In this study, we use the node weight time series W_k for the following analysis.

2.2.2. Multivariate SOM Analyses for Large-Scale Atmospheric Variables

A_i (i = 1,2,3) represents the three atmospheric variables used in this study, that is, Z500′, IVT′, and CINi′, each of which has a size of ($N_{lon} \times N_{lat}$, N_t). As this step is to identify general atmospheric pattern settings, SOM is used to classify temporal points (days) with similar combination of these three variables. To do this, data matrices of these three variables are concatenated in the spatial dimension; the input data matrix for the SOM has a size of ($3 \times N_{lon} \times N_{lat}$, N_t). Note that the representation of rows (grid points) and columns (days) of the input data matrix is reversed compared to that of section 2.2.1. In this study, the specific input size is (5,133, 5,967). Our experiments show that SOM training with such input size and 1,000 epochs will cost over half an hour on a PC with a two-core CPU, or about 8 min if using a powerful server with double 20-core CPUs. Although a single training does not take much time, many simulations are needed for the trial-and-error process in testing parameters and with more data samples and variables, the training time can increase tremendously.

In order to cut down training time for using multivariate SOM, we introduce a PCA preprocessing technique to reduce the dimension of the input matrix size and remove the synoptic noise. PCA is a widely used method for dimension reduction or extracting individual modes of variability. It takes the input of high-dimensional data (multiple variables) and generates output of low-dimensional data (principle components) and the corresponding loading vectors. In the field of atmospheric and oceanic science, PCA is also known as EOF. Detailed description of PCA/EOF can be found in numerous books and studies (e.g., Bretherton et al., 1992; Jolliffe, 2002; Van den Dool, 2007).

The PCA preprocessing starts by applying PCA to each of the three variables A_i (i = 1, 2, 3). This yields three new matrices $A_i^* = E_i A_i$ with a reduced size of (M_i, N_t), where E_i is the loading vector matrix (EOF modes) and has the size of (M_i, $N_{lon} \times N_{lat}$). Each row in A_i^* represent a PC that explains certain fraction of variability in the original data A_i. To properly represent the original data, for each new variable A_i^*, the top M_i^* ($M_i^* \le M_i$) PCs explaining most data variability are selected. Assuming the eigenvalues for the top M_i^* PCs are λ_{ij} (j = 1, 2, ... ,M_i^*). We can scale variable A_i^* by dividing it by the squared sum of all its eigenvalues ($\lambda_i^* = \sqrt{\sum_{j=1}^{M_i^*} \lambda_{ij}^2}$); this step is to guarantee that the variabilities of all three variables are equally important for the pattern classification, an assumption of this study. These three variables are concatenated along the spatial dimension to form a new data matrix $B = \begin{bmatrix} A_1^*/\lambda_1^* \\ A_2^*/\lambda_2^* \\ A_3^*/\lambda_3^* \end{bmatrix}$ with a size of (M^*, N_t), where $M^* = M_1^* + M_2^* + M_3^*$. The SOM pattern classification is then performed on the concatenated data matrix B along the temporal dimension. The resultant weight matrix of each SOM node ($W^* = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{M^*} \end{bmatrix}$) is separated into three parts ($W_1^* = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{M_1^*} \end{bmatrix}$, $W_2^* = \begin{bmatrix} w_{M_1^*+1} \\ w_{M_1^*+2} \\ \vdots \\ w_{M_1^*+M_2^*} \end{bmatrix}$, $W_3^* = \begin{bmatrix} w_{M_1^*+M_2^*+1} \\ w_{M_1^*+M_2^*+2} \\ \vdots \\ w_{M^*} \end{bmatrix}$) and then scaled back by multiplying by λ_i^* ($W_i = W_i^* \cdot \lambda_i^*$). The loading vectors from the previous PCA analysis are used to reconstruct the SOM node matrix to the original spatial grids ($\mathrm{SOM}_i = W_i^T E_i = \lambda_i^* W_i^{*T} E_i$, where T means transpose of the matrix).

This PCA preprocessing procedure described above can greatly reduce input data size and thus reduce training time. For example, if only using PCs explaining 90% variance, SOM is applied to an 84 ($M_1^* = 37, M_2^* = 8, M_3^* = 39$) × 5,967 ($N_t$) matrix instead of the original data with the size of 5,133 × 5,967. Table S1 in the supporting information further lists the difference of training time, QE, $\overline{\rho}^*$, and $\mathrm{CC}_{\max}$ between SOMs trained

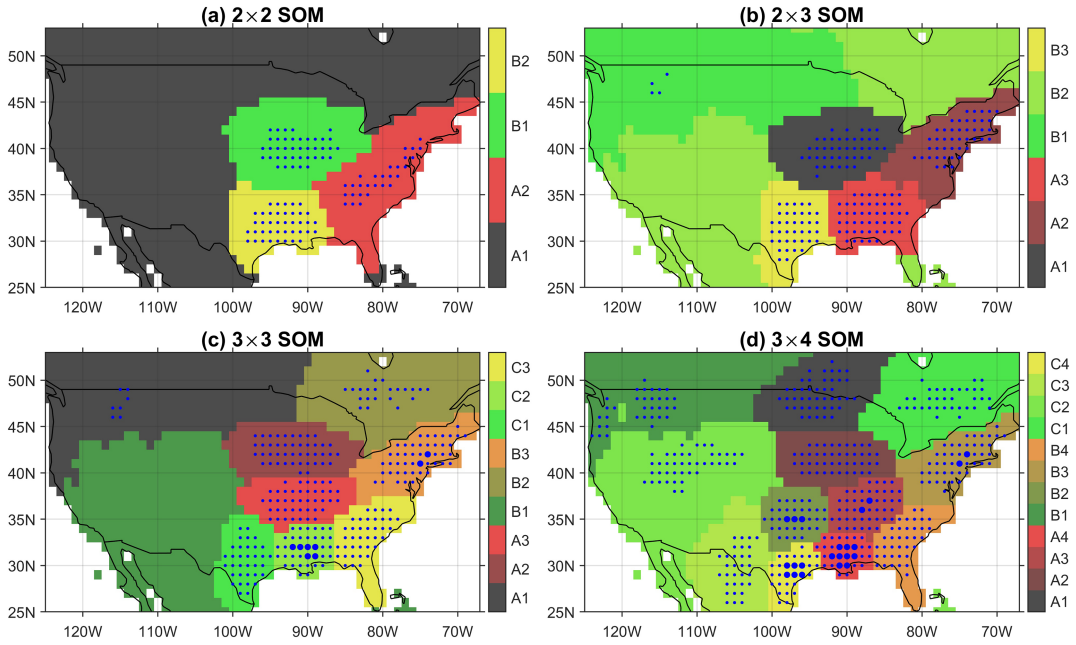

Figure 1. Precipitation region classification with different SOM map size: (a) 2 × 2; (b) 2 × 3; (c) 3 × 3; (d) 3 × 4. The small (large) dot indicates the grid point where the correlation coefficient between the precipitation anomaly time series and its corresponding SOM node weight series is larger than 0.6 (0.8) with a *p* value smaller than 0.05.

without and with PCA preprocessing (keeping PCs explaining 99%, 90%, 80%, and 70% variability), and each scheme is run for 30 times in order to calculate mean value and standard deviation of the performance statistics. The tested SOM node scheme is 5 × 5, and training epoch is 1,000. The SOM training program is run under a Linux server with two E5-2698v4 20-core CPU and 256 GB RAM. The training time is greatly reduced after training with PCA processing (~60 times faster for 90% PCA). As we reduce the threshold for percentage of PC explained variability from 99% to 70%, input data size and training time both decrease, but the SOM performance indicated by QE, $\overline{\rho}^{*}$ and CC_{max} also degrades as well. In this study, we use the PCA preprocessing scheme with 90% threshold for the following reasons: (1) it maintains most data structure from the original data while only some small scale spatial features are removed; (2) it reduces the input data matrix size very significantly (~1/61 of the original size); and (3) performance statistics are very similar to that without PCA preprocessing as shown in Table S1.

3. Precipitation Region Partition

In this section, we apply SOM to classify all available grid points from the CONUS domain into several representative regions. The purpose of this process is to acquire several precipitation indices to represent a number of regions instead of calculating the average precipitation anomaly over a manually selected region. Because the SOM results are sensitive to the number of nodes chosen subjectively, we first evaluate the sensitivity of our results to the number of the SOM nodes. Figure 1 shows four different node number schemes ranging from 2 × 2 to 3 × 4. The node connection, organization, corresponding color for each node, and Euclidean distance between neighboring nodes are shown in Figure S1. As indicated by the results of all four schemes, precipitation variability is quite different between the two regions separated by the ~101°W longitude line due to the difference of orography and surface altitude between them. In the 2 × 2 scheme (Figure 1a), the west half of CONUS and all the area to the north of ~45°N are categorized into the same node; most of the Great Plains area and the area spanning from east coast to the southeast United States is classified as the other three nodes. As we increase node number (from Figures 1a to 1d), more detailed classification of the precipitation nodes appears in the western and northern United States with increased correlation between the SOM node and its corresponding grid points, indicating more robust classification. In this study, we will use Nodes A1 and B3 in the 2 × 3 scheme (Figure 1b) as these two nodes largely overlap

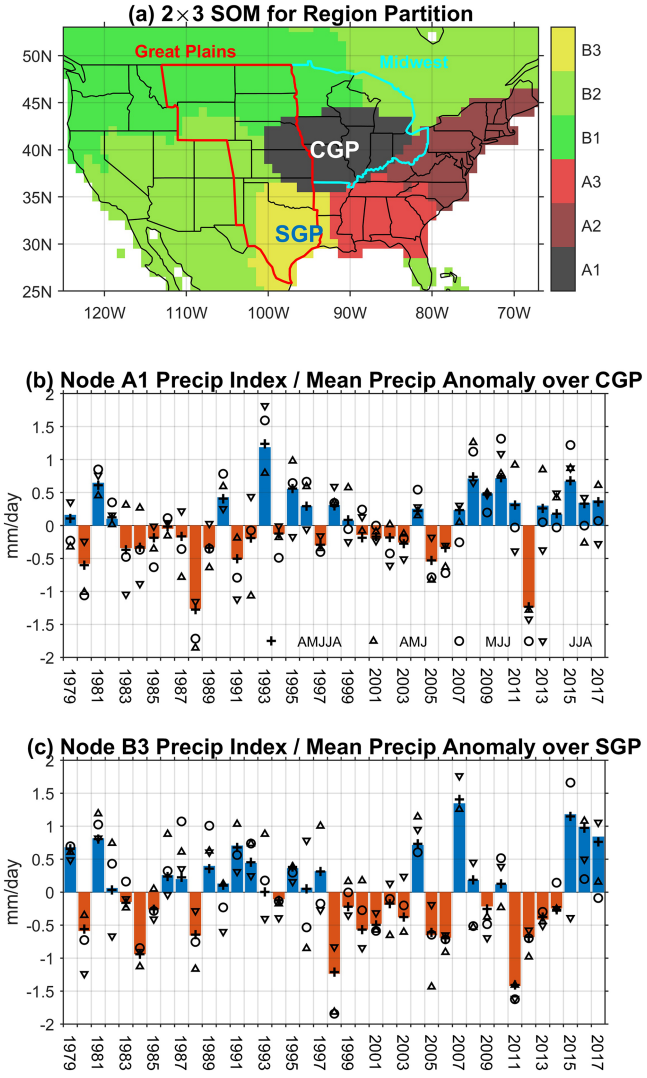

Figure 2. Precipitation region classification with a 2 × 3 SOM (a) and the corresponding precipitation index and average precipitation anomaly for (b) Node A1 (CGP) and (c) Node B3 (SGP). Bars represent the average node weight through April to August; the positive (negative) mean value is plotted as blue (orange) color to represent a wetter (drier) condition for each year while the cross, upward pointing triangle, circle, and downward pointing triangle represent average precipitation for the period of April to August, April to June, May to July, and June to August, respectively.

with the traditionally recognized Central Great Plains/Midwest and SGP (Figure 2a), and there are ample grid points with large correlation coefficient within these two nodes. We will refer to the area delineated by Nodes A1 and B3 in Figure 1b as CGP and SGP, respectively. We can also use Nodes B1 and B2 in Figure 1a, or Nodes A2, A3, and C1 in Figure 1c, for the same analysis as described below.

The node weight time series for a specific node, representing the average precipitation anomaly for this node, can be used to indicate the general precipitation variation over the region occupied by this node as described in section 2.2.1. In Figures 2b and 2c, the bars show the average node weight for April–August of each year, which are very close to the mean precipitation anomaly indicated by the "+" sign. Major drought years over SGP/CGP can be easily seen in Figures 2b and 2c as well, with the most severe and long-lasting ones being the 1988 (CGP), 1998 (SGP), 2011 (SGP), and 2012 (CGP) droughts. Other droughts with shorter durations can also be seen with the mean precipitation anomalies of various 3-month periods marked in Figures 2b and 2c, for example, the 2005 SGP drought that has an early break time during the summer.

4. Atmospheric Patterns Related to SGP and CGP Droughts

4.1. SOM Node Patterns for All Warm Season Days

The SOM pattern classification that is applied to the concatenated field data consists of standardized anomalies of Z500, IVT, and CINi following the steps described in section 2.2.2. As there is no ideal way to determine the best SOM size, this study uses the trial-and-error approach to test several combinations of node number scheme (from 2 × 2 to 5 × 6, 7 schemes in total) and training epochs (from 200 to 2,000 by increments of 50). Our experiment shows that the training result is not sensitive to training epochs, possibly due to the use of the PCA initialization approach, which leads to faster convergence, and this study will use the training epoch 1,000 for all SOM mentioned below. SOM performance statistics including QE, $\overline{\rho}^*$, and CC_{max} for the seven node schemes tested are shown in Table S2. As the node number increases, QE generally decreases and $\overline{\rho}^*$ increases, indicating the SOM node's representativeness for the data improves. However, CC_{max} also increases, which indicates that the similarity between nodes increases and thus results in redundancy. According to Huang et al. (2017) and Fettweis et al. (2011), CC_{max} larger than 0.7 means the similarity between at least two nodes is high. Based on these considerations, the 5 × 5 SOM scheme is chosen for this section, and the result is shown in Figure 3 (Z500′ as contours and IVT′ as colors) and Figure 4 (Z500′ as contours and CINi′ as colors).

In Figure 3, due to the geostrophic balance relation between geopotential and wind field, east or southeast displaced geopotential high anomalies with centers in the eastern or southeastern United States are commonly related to northward geostrophic wind anomalies, thus enhancing the moisture transport from the Gulf of Mexico, and positive IVT′ often occurs along the western or northern edge of the anomalous geopotential high where the geopotential gradient is the largest (e.g., Nodes A1, C1, D1, E1, C2, and D2). In contrast, north or west displaced geopotential high centers generally correspond to negative IVT′ over the Great Plains (e.g., Nodes A4, A5, B3, B5, C5, D4, D5, E4, and E5). In Figure 4, we can also see that negative CINi′ values are distributed near the geopotential high center over the CONUS (e.g., Nodes A2, A3, A4, A5, B3, B4, B5, C2, C5, D3, D5, E2, E3, E4, and E5), as strong anomalous geopotential high and negative CINi′ both can

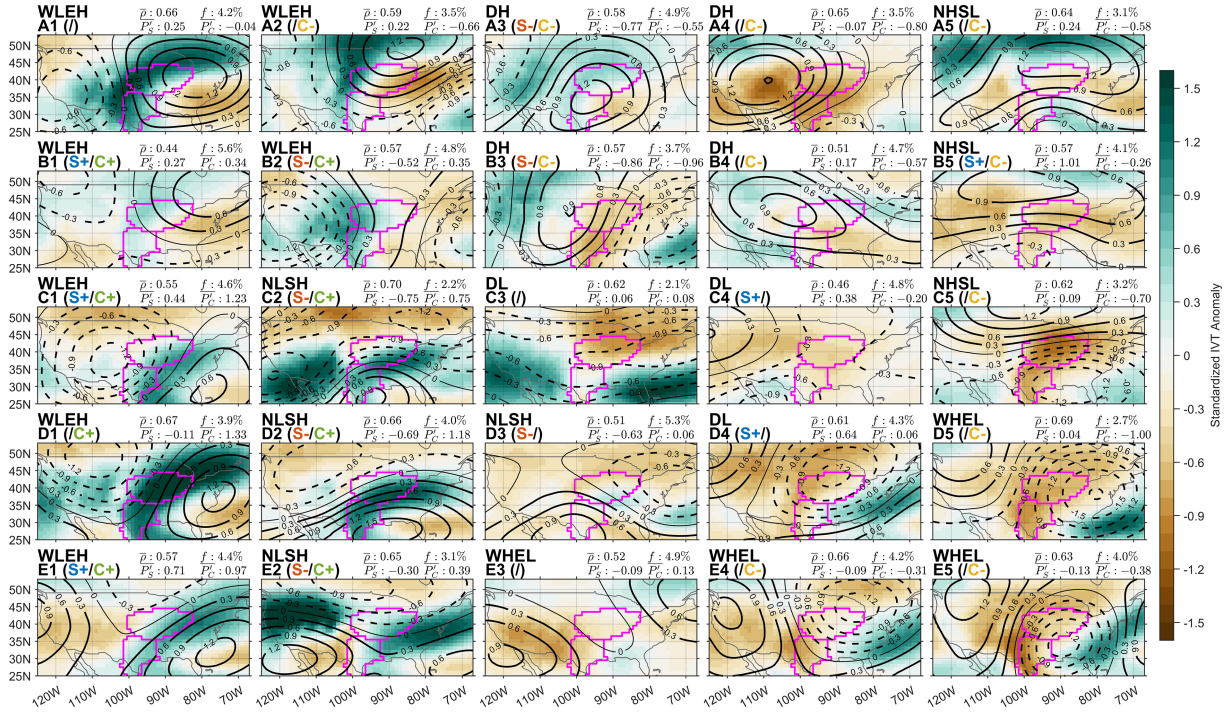

Figure 3. The 5 × 5 SOM patterns of standardized Z500 anomaly (Z500′) and standardized IVT anomaly (IVT′). Positive (negative) Z500′ is plotted as solid (dashed) contours. Positive (negative) IVT′ is marked as green (yellow) color, and a deeper color means a larger IVT′ value. For each node pattern, the corresponding circulation type (i.e., DH, etc.), the associated domain mean standardized precipitation anomalies for the CGP and SGP ($\overline{P'_C}$ and $\overline{P'_S}$), the frequency of occurrence of this node pattern *f*, and the mean of the pattern correlation coefficients for each node $\overline{\rho}$ are indicated above the corresponding panel. Nodes with significant positive (negative) $\overline{P'_C}$ or $\overline{P'_S}$ are labeled with a "C+ (−)" or "S+ (−)", respectively.

limit upward disturbances and suppress convection development and together they have positive feedback on each other and work to trigger and intensify rainfall deficits.

Frequency of occurrence for each node (*f*), as well as average standardized precipitation anomalies over the SGP and CGP ($\overline{P'_S}$ and $\overline{P'_C}$), respectively, are also shown besides the node pattern in Figures 3 and 4. We use a Monte Carlo test, which randomly shuffles the order of year in the data for 10,000 realizations, to determine whether the standardized precipitation anomaly ($\overline{P'_S}$ and $\overline{P'_C}$) is significantly positive or negative for all the SOM nodes. The actual standardized precipitation anomaly is considered significantly positive if it is larger than 95% of the highest mean standardized precipitation anomalies of the randomly sampled data and considered significantly negative if it is smaller than the lowest 5% of them. Patterns with significant positive (negative) $\overline{P'_C}$ or $\overline{P'_S}$ are labeled with a "C+" ("C−") or "S+" ("S−") in the title of each subfigure, respectively. It should be noted that dry nodes for SGP or CGP are generally adjacent in the SOM grid. Among these dry nodes, A3 and B3 are related to decreased precipitation in both SGP and CGP, indicating that the patterns associated with these nodes representing SGP and CGP droughts are related.

Depending on their center location and coverage of geopotential high and low anomalies, we further manually classify these nodes into six circulation types. These six circulation types are (1) western low-eastern high (WLEH, Nodes A1/A2/B1/B2/C1/D1/E1), which is mostly associated with wetter precipitation anomalies over either the SGP or the CGP; (2) western high-eastern low (WHEL, Nodes D5/E3/E4/E5), associated with less water vapor transport over a vast region of the Great Plains and Midwest, and often linked to significant drier precipitation anomalies over the CGP. (3) northern low-southern high (NLSH, Nodes C2/D2/D3/E2), largely corresponding to drier precipitation anomalies over the SGP and wetter precipitation anomalies over the CGP; (4) northern high-southern low (NHSL, Nodes A5/B5/C5), corresponding with the drier precipitation anomalies over the CGP, and higher or normal precipitation over the SGP; (5) dominant low (DL, Nodes C3/C4/D4), associated with normal or positive precipitation anomalies over the SGP

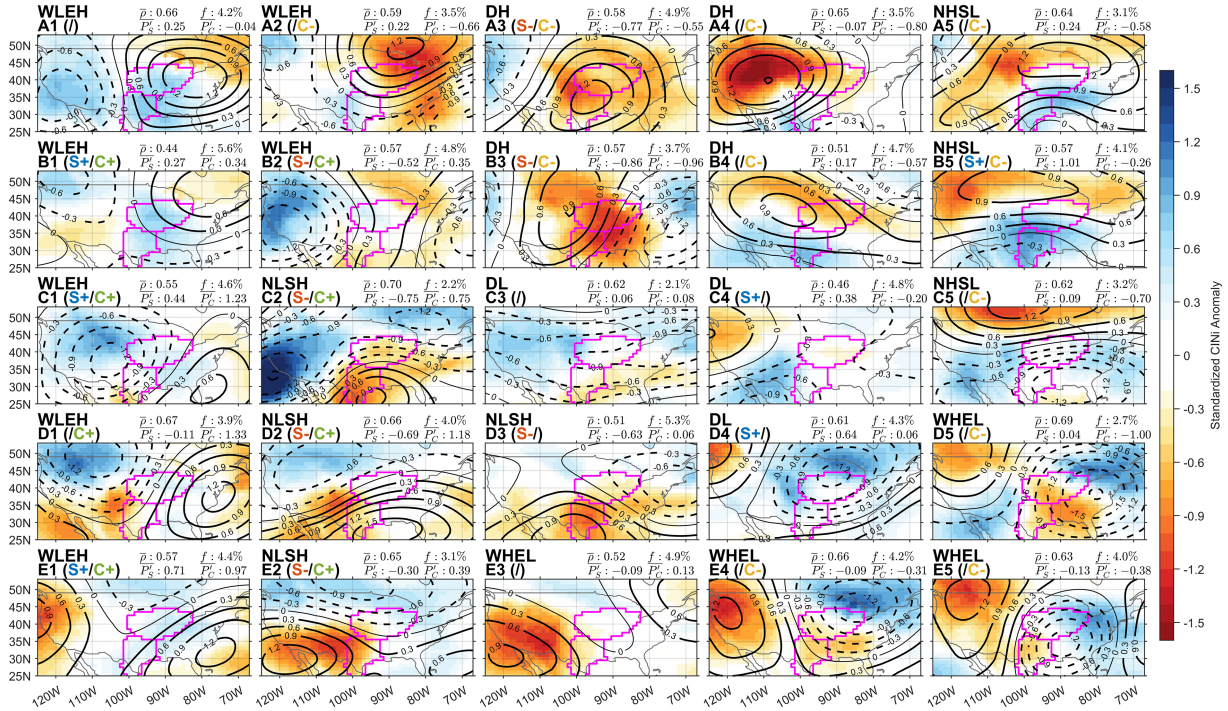

Figure 4. Same as Figure 3, but for the standardized Z500 anomaly (Z500′) and standardized CINi anomaly (CINi′).

and normal over the CGP; (6) dominant high (DH, Nodes A3/A4/B3/B4) generally associated with drier conditions over the SGP (except for A4/B4) and the CGP.

In Figure 5, we further evaluate the impacts of the large-scale circulation patterns represented by the SOM nodes on rainfall anomalies. We define five precipitation categories based on the average standardized precipitation anomaly over SGP and CGP, respectively: (1) very dry ($P < P_{16.7\%}$); (2) moderately dry ($P_{16.7\%} \le P < P_{33.3\%}$); (3) normal ($P_{33.3\%} \le P < P_{66.7\%}$); (4) moderately wet ($P_{66.7\%} \le P < P_{83.3\%}$); and (5) very wet ($P \ge P_{83.3\%}$). The dry (wet) category includes both the very dry (wet) and the moderately dry (wet) categories. Figures 5a and 5b also show the percentages of all precipitation categories (very dry, moderately dry, normal, moderately wet, and very wet) for each node for the SGP and the CGP, respectively. This information is used to estimate the probability of a specific precipitation state when the large-scale atmospheric pattern is classified into a given SOM node. Since the main focus of this study is the Great Plains' drought, the dry and very dry categories are labeled with specific percentage numbers in Figures 5a and 5b, shown as black and red numbers, respectively. Larger percentages of the dry to very dry precipitation categories are associated with the nodes that represent the anomalous circulation favorable for drier conditions (on the left side of the red dashed line). For example, for the seven nodes favorable for dry conditions over the SGP (marked as "S–" in Figures 3 and 4), the percentage of dry days assigned to the node ranges from 73% (B3) to 41% (E2) of the days when each of these nodes occurred, that is, the circulation patterns represented by that nodes occurred. The percentage of very dry days ranges from 45% (Node A3) to 24% (Node E2) of the days when each of these nodes occurred. Over the CGP, for the 11 nodes marked with "C-", the percentage of the dry days ranges from 79% (D5) to 38% (B5), and the percentage of very dry category ranges from 54% (D5) to 16% (E4) of the days when each of these nodes occurred.

The WLEH patterns correspond to strong, anomalous low (high) Z500′ dominant over the western (eastern) United States. Such a circulation pattern enhances northward transport of moisture from the Gulf of Mexico to central United States. Hence, most of these nodes correspond to higher precipitation over the CGP or SGP (positive $\overline{P_C'}$ or $\overline{P_S'}$, and low probability of dry conditions). The WHEL patterns are mostly (except E3) associated with a strong ridge over the western United States, and the southward circulation anomalies caused by this west-east geopotential anomaly contrast weaken the Great Plains low level jet and result in a smaller

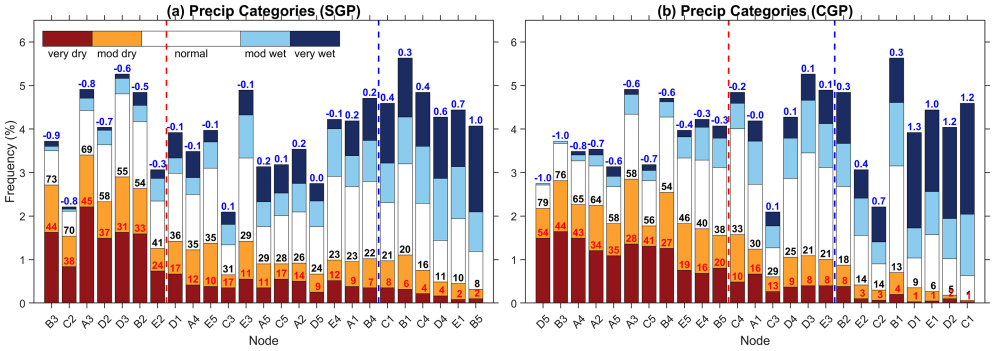

Figure 5. Percentage of each precipitation category in each node of the 5 × 5 SOM: (a) and (b) for the SGP and the CGP, respectively. The blue number above the bar is the mean standardized precipitation anomaly of the node; the black (red) number represents the percentage of the dry (very dry) days in the node. Nodes to the left of red dashed line and those to the right of blue dashed line are dry and wet nodes labeled as "S−"/"C−" and "S+"/"C+" in Figures 3 and 4.

IVT′ over the Great Plains. Three out of the four WHEL nodes (D5, E4, and E5) are significantly related to dry conditions over the CGP ($\overline{P_C'}$ from −0.31 to −1.00 in Figures 3 and 4, dry days percentage from 40% to 79% in Figure 5b). As for the SGP, although Figure 3 shows that SGP receives even less IVT′ in WHEL nodes, none of these nodes are significantly related to dry condition over the SGP ($\overline{P_S'}$ from −0.09 to 0.04, dry days percentage from 23% to 35%), which may indicate moisture transport is a less important factor for SGP drought than for CGP. For NLSH patterns, three of them (C2, D2, and E2) are associated with more moisture transport over different regions of the CONUS, a result of displacements of the subtropical high toward north to south-central, southeastern, and southwestern United States, respectively. Given the relationship between anomalous geopotential and wind field, regions that show most IVT′ increase are those with largest geopotential anomaly gradient (strong geostrophic wind), which can explain much of the wet condition over the CGP. D3, which also belongs to the NLSH type, does not show increased IVT′ due to its relatively weak geopotential high and smaller geopotential gradient. Although most of these NLSH nodes show increased moisture transport over the Great Plains, the strong anomalous geopotential high suppresses upward motion and increases surface temperature, as well as has a larger negative CINi′ over the SGP. These NLSH nodes are related to much of the dry conditions over the SGP (41% to 70% in Figure 5a), but not that of the CGP (5% to 21%, Figure 5b). In contrast, the NHSL patterns mostly (B5 and C5) lead to lower IVT′ (< −0.5) over a vast area of the northern CONUS and correspond to a significant fraction of dry conditions over the CGP (38% to 56%), and high or normal precipitation over the SGP. The DH patterns (A3, A4, B3, and B4) feature an extensive area of positive Z500′ over CONUS, leading to high probabilities of dry condition over the CGP (54% to 76%). The DH patterns A3 and B3 are also associated with negative $\overline{P_S'}$ and high probability of dryness (69% to 73%) over the SGP, whereas other SOM nodes (A4 and B4) have smaller probability of dryness there (22% to 35%). The DL patterns (Nodes C3/C4/D4) are associated with negative Z500′ over much of the CONUS, and low probability of dryness of the SGP (11% to 31%) and CGP (25% to 33%).

4.2. SOM Node Patterns for Very Dry Days

Using all data for pattern classification will potentially give a better view of pattern transitions. However, this can also smooth out some of the signals most related to extreme precipitation events. Although the 16 labeled dry nodes discussed above in Figures 3 and 4 are related to a significantly larger negative precipitation anomaly and have higher probabilities of causing dry conditions than other nodes, there are still a large portion of samples assigned to these nodes that belong to the normal or wet condition. In this section, in order to retrieve atmospheric patterns related to extreme dry conditions, we further select the very dry days over the SGP or the CGP or both, which consist of ~30% of the total samples used in the previous section, and apply a new SOM analysis to them. The reason that we apply the SOM for very dry days either over the SGP or the CGP, instead of for very dry days over each region separately, is to capture the circulation patterns related to the extreme dry condition affecting both two regions.

Similar as done in section 4.1, we choose a 4 × 4 node scheme after testing several SOM node schemes (performance statistics for seven node schemes tested are shown in Table S3). The result is shown in Figure 6

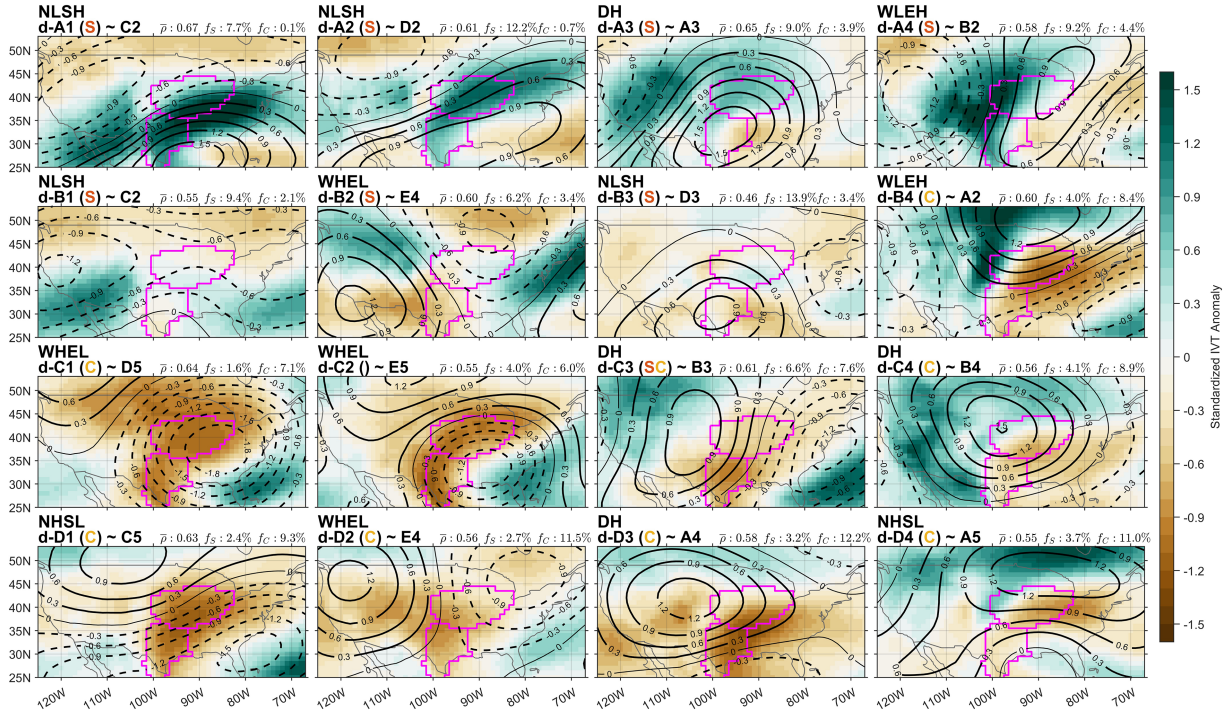

Figure 6. Same as Figure 3, but for the very dry days over the CGP or the SGP. For each node pattern, f_C (f_S) represents the percentage of very dry days over the CGP (SGP) categorized as this node pattern. Nodes with significant negative $\overline{P'_C}$ ($\overline{P'_S}$) are labeled with a "C" or "S," respectively.

(Z500′ as contours and IVT′ as colors) and Figure 7 (Z500′ as contours and CINi′ as colors). This node scheme also yields the large-scale patterns consistent with those associated with the dry conditions shown in Figures 3 and 4, but mostly with stronger amplitudes. A prefix "d-" is added to the node names in the new 4 × 4 SOM to distinguish them from those based on all the warm season days (Figures 3 and 4). Table S4 lists two nodes in the 5 × 5 SOM that are most correlated with each node in the 4 × 4 SOM here and the corresponding correlation coefficients are also displayed. Bold font for the correlation coefficient indicates that the correlation is significant at a 0.05 level. Most nodes in the 4 × 4 SOM are significantly correlated with one or two nodes in the 5 × 5 SOM with a correlation coefficient larger than 0.8, except for d-B2 and d-C4. For each node in this 4 × 4 SOM, the node in the 5 × 5 SOM that has the largest correlation coefficient is referred to as its "counterpart" hereafter. Difference of standardized Z500′, IVT′, and CINi′ between nodes in this 4 × 4 SOM and its counterpart are also shown in Figure S2 (Z500′ as contours and IVT′ as colors) and Figure S3 (Z500′ as contours and CINi′ as colors).

Results in Figures 7 and S3 show that six out of eight "S" labeled very dry nodes (except d-A4 and d-B1) for the SGP show larger negative CIN′ values over more extensive areas over the SGP and its vicinity than their counterparts. In addition to enhanced stability, there are six of these "S" nodes (except d-A2 and d-B1) that show positive Z500′ either with stronger amplitude or with larger coverage than their counterparts. In contrast, only three nodes show substantially larger negative IVT′ (d-A2, d-B1, and d-C3) than their counterparts over the SGP (Figures 6 and S2).

As for the CGP, six out of eight "C" labeled nodes show larger negative IVT′ over the CGP than their counterparts (except d-C3 and d-D3), indicating more severe moisture deficit during very dry days (Figures 6 and S2). On the other hand, although seven nodes show larger negative CINi′ over the CGP (except d-C1), only d-C3 and d-C4 shows larger Z500′ near the anomalous geopotential high center (Figures 7 and S3).

These results are consistent with findings in previous process studies that strong negative CINi′, due to land surface dryness and low-level warm air advection are the main environmental conditions that suppresses rainfall over the SGP (e.g., Fernando et al., 2016; Myoung & Nielsen-Gammon, 2010a). These results are also consistent with previous process studies that weaker moisture transport by the low-level jets and westerly

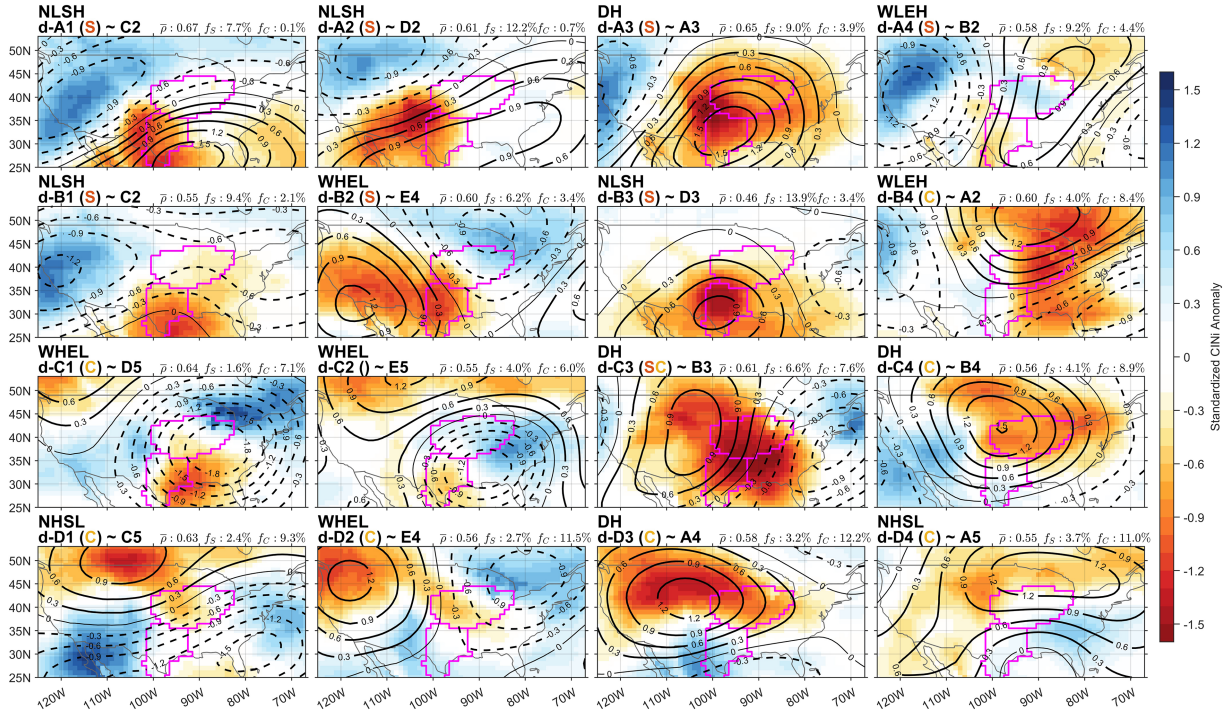

Figure 7. Same as Figure 6, but for the standardized Z500 anomaly (Z500′) and standardized CINi anomaly (CINi′).

advection of drier air mainly contributes to the drought conditions in the CGP (Weaver et al., 2009). In particular, Weaver et al. (2009) found that the Great Plains low level jet anomalies leads the precipitation and moisture flux in the northern Great Plains (similar to the CGP here). Erfanian and Fu (2019) also found that the moisture transport deficit is a main contributor to the extreme drought over the northern Great Plains in 2012, due to the anomalously dry zonal advection in spring and dry meridional advection in summer.

Similar to Figure 5 in section 4.1, percentages of precipitation categories of each node in the 4 × 4 SOM for very dry days of the SGP and the CGP are shown in Figures 8a and 8b, respectively. Because of the input data selection criterion (very dry days of the SGP or the CGP), most daily samples assigned to the nodes are from the very dry days. However, a few samples in Figure 8a (Figure 8b) are from the other four precipitation categories as they correspond to very dry days of the CGP (SGP) but not the SGP (CGP). Node orders are sorted based on the number of very dry days explained by these nodes. The first 8 nodes in both Figures 8a and 8b consist of most very dry days (~75%) in the SGP and CGP, respectively. They are labeled with a "S" or "C" in the node titles of Figures 6 and 7, and S2 and S3, thus representing the circulation patterns associated with the majority of very dry days over the SGP or the CGP. Blue numbers above bars indicate average precipitation anomalies for the very dry days from the nodes. Over the SGP (Figure 8a), the very dry days are linked to circulation patterns represented by eight SOM nodes, mostly consisting of the d-A and d-B nodes. Among them, d-B3 and d-A2 with strong negative CIN′ occur most frequently. The mean standardized precipitation anomalies for most nodes are −1.3 (−1.2 to −1.4). Over the CGP (Figure 8b), the very dry days are linked to circulation patterns represented by another eight SOM nodes, mostly the d-D and d-C nodes. Among these, d-D3, d-D2, and d-D4 occur most frequently. The SOM nodes for CGP have stronger precipitation deficits for the very dry days than those at the SGP, as well as larger precipitation differences among different nodes (−1.1 to −1.6 in Figure 8b).

To see which nodes and node types are mainly responsible for the past severe drought years, Figures 9a and 9b show the number of very dry days contributed by each SOM node for several years with largest total number of very dry days over the SGP and CGP, respectively. Node pattern type for each node is also displayed as color in Figure 9. The total number of very dry days in a year can also be regarded as an indicator of drought intensity and persistence. It can be seen that in the SGP, 2011, 1998, and 1980 had substantially more very

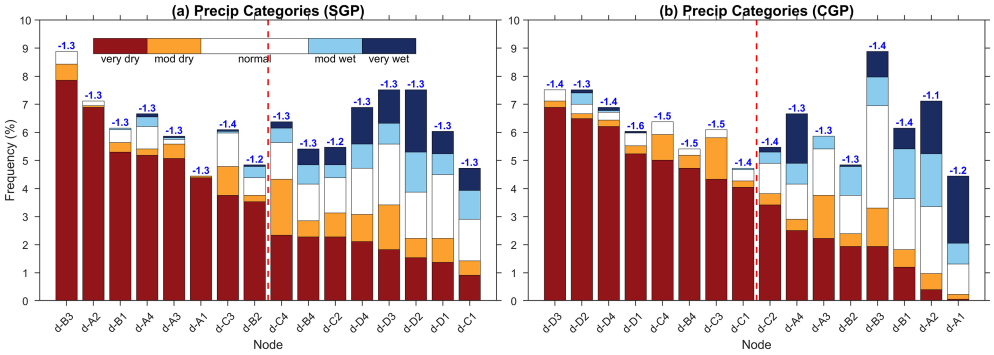

Figure 8. Same as Figure 5, but for the 4 × 4 SOM. The blue number above the bar is the mean standardized precipitation anomaly of only the very dry days in the node; nodes to the left of red dashed line are very dry nodes labeled with "S"/"C" in Figures 6 and 7.

dry days than other years since 1979. For the CGP, 1988 and 2012 have more very dry days than other years. For the 2011 SGP drought, ~45% of the very dry days are related to NLSH type node d-B3, which is also the leading node that accounts for most of the total very dry days during 1979 to 2017 (~14%, not shown). In addition, the NLSH type that includes node d-B3, d-A2, d-A1, and d-B1, accounts for ~71% of very dry days over SGP in 2011. Similarly, in 1998, NLSH is also most related to very dry days (~58%). In the 1980 SGP drought, though NLSH is still important (~37%), the dominant node for very dry days becomes the DH type node d-A3 (~39%). Over the CGP, DH type (e.g., Nodes d-C3, d-C4, and d-D3) accounts for most very dry days during both the two most severe drought years 1988 and 2012 (~39% and ~69%), as well as some other drought years. The other anomalous circulation types for very dry days of the CGP, WHEL and NHSL, are scattered among less severe drought years.

4.3. Temporal Evolution of SOM Node Patterns During Drought Events

In this section, we investigate the temporal evolution of SOM nodes during long drought periods. Since we have applied a 7-day moving average filter to the original precipitation data, we analyze the 7-day lag node-to-node transition frequency both for all samples and for days during the drought events. In addition, because transition analysis is on a continuity basis, the analysis in this section is based on the result of 5 × 5 SOM for all the days during our analysis period.

Here, a drought event is defined as (1) a time period that consists of consecutive days of negative precipitation anomalies and has at least 1 day that meets the criterion of the dry day category defined in section 4.1 over the SGP or the CGP; (2) if the interval between two events is less than 7 days and these days are dry or normal days, then the two events and the interval are considered one single event; and (3) the duration of the event is at least 39 days for the SGP or 28 days for the CGP. Different duration thresholds are used for the

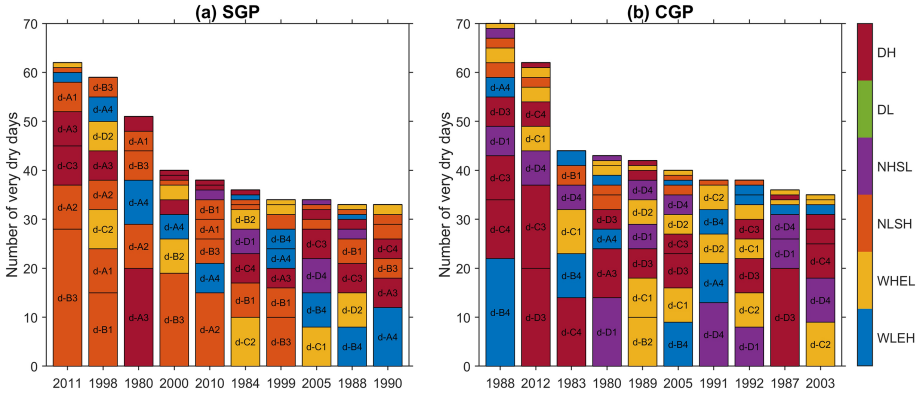

Figure 9. Numbers of very dry days contributed by each SOM node for top 10 years with the largest number of very dry days over the (a) SGP and (b) CGP, respectively. Nodes belong to the same circulation type are marked with the same color.

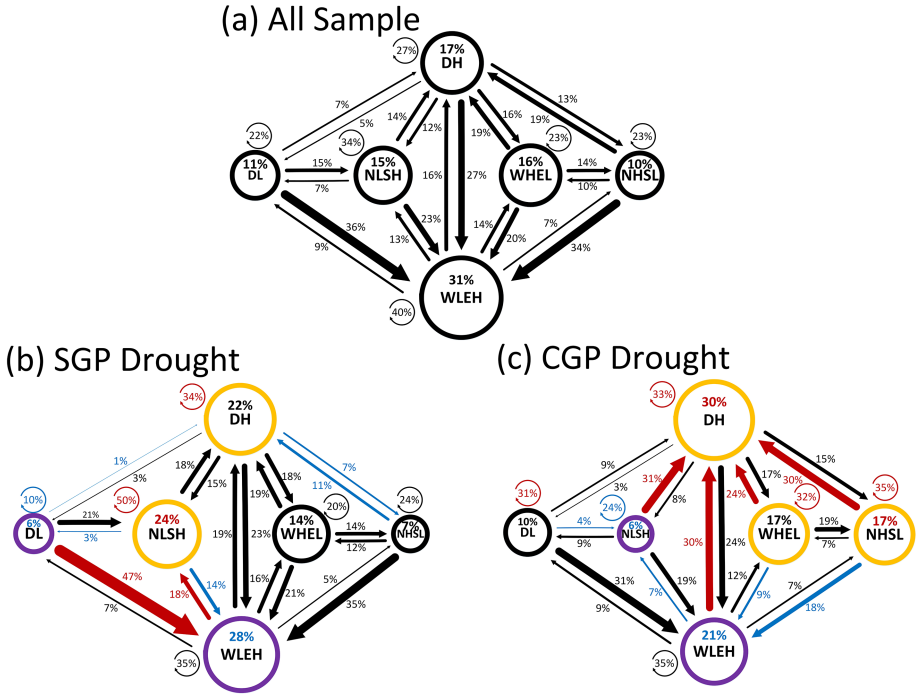

Figure 10. Frequencies of transition between different circulation types for (a) all samples, (b) SGP drought events, and (c) CGP drought events. The frequency of the transition is proportional to the arrow width and its value is labeled next to the arrow. The frequency of persistence of each circulation type is listed next to the circle of circulation pattern type and surrounded by a curly arrow. The number inside the circle is the percentage of the frequency of occurrence of that given circulation type. Red color of the arrow/number in (b) and (c) indicates that the changes of the frequency during SGP or CGP drought events from the climatological values shown in (a) is significantly greater ($p < 0.1$). The blue color indicates that the frequency is significantly less than that of climatology shown in (a). Orange (purple) color of the type circle indicates that days assigned to that type are mostly linked to negative (positive) precipitation anomaly.

SGP and the CGP because the SGP has more drought events with longer duration than the CGP, and using these two thresholds results in similar sample numbers selected for the SGP and the CGP. Fifty drought events are identified using the above criteria, 22 for the SGP and 28 for the CGP, which account for 21.0% and 20.7% of all daily samples, respectively, for these two regions. Most of these events (42 out of 50) have dry days for over 50% of their daily samples. The top eight drought years with the largest cumulative precipitation anomalies of drought events in each of the SGP and the CGP are listed in Table S5. Out of the 8 years listed for each region in Table S5, the top 2 years (1998 and 2011 for the SGP, 1988 and 2012 for the CGP) stand out as their durations and accumulative precipitation deficit are much greater than those of the other years. In particular, the 1998 SGP drought lasted 124 days with a cumulative rainfall deficit of 209.14 mm. The 2011 SGP drought consisted of two drought events that lasted 117 days in total with a cumulative rainfall deficit of 195.24 mm. The 1988 CGP drought lasted 97 days with a cumulative rainfall deficit of 196.49 mm, and the 2012 CGP drought lasted 109 days with a cumulative rainfall deficit of 192.99 mm.

We calculate the 7-day lag node transition frequency for all daily samples as well as for samples only from the SGP and CGP drought events. We choose to analyze 7-day lag transitions instead of 1-day lags because the original data has been preprocessed using a 7-day moving average. Hence, patterns within a day or two will be very similar and using 1-day lag would result in very few frequency transitions to other nodes. Note that although the 7-day moving window smooths out most of the weather fluctuations, it does not guarantee day $x + 1$ to $x + 6$ always be the same when day $x + 0$ and $x + 7$ are assigned to the same node.

Because many node-to-node transitions have few occurrences, we combine these nodes into the six aforementioned pattern types, that is, WHEL, WLEH, NHSL, NLSH, DL, and DH. We then calculate the frequencies of type-to-type pattern transition instead of node-to-node transition. The climatological statistical distributions of the anomalous circulation pattern types and transition frequencies among them are characterized and compared to those during the SGP and CGP drought periods in Figure 10. Whether the transition frequency between certain pattern types during the SGP or CGP drought periods is significantly larger or

smaller than that of all samples is tested using a Monte Carlo test similar to that of section 4.1, where the order of all years (1979–2017) is shuffled and the node frequency and transition frequency are recalculated 10,000 times (e.g., after shuffling the years, the SGP drought event during 1 May to 19 June 2011 is resampled for the same dates 1 May to 19 June, but in different year like 2000). If over 90% of the frequencies of the randomly shuffled data are less (greater) than the actual frequency during the SGP or CGP drought events, the actual value is considered significantly more (less) frequent than the climatological value in Figure 10a and is marked by a red (blue) number.

Figure 10a shows that a pattern type has a frequency of 22–40% of maintaining itself for 7 days climatologically for our analysis period. The top cell is the DH circulation type, which consists of four SOM nodes and occurs ~17% of the time during our analysis period. DH is responsible for dry spells (7–14 days) over both the SGP and CGP. We can see that it has a relatively small frequency of maintaining itself (~27%) and has similar frequencies of occurrence in the transition to and from other circulation pattern types such as DL, NLSH, and WHEL. The bottom cell is the WLEH type, which brings the most moisture to the Great Plains region and is most related to the positive precipitation anomaly over the SGP or CGP. It consists of seven SOM nodes and occurs ~31% of the time during our analysis period, about 2 to 3 times as often as other circulation types, and it also has the largest frequency of maintaining itself (40%). The other circulation types also have high frequencies of transition to WLEH.

NLSH and DH types are the most frequent circulation patterns of SGP drought events (Figure 10b). During long drought events over the SGP, the frequency of the NLSH type increases significantly ($p < 0.001$) by 9%, while the increase of the frequency of DH type (5%) is not statistically significant ($p \approx 0.16$). Their frequency of sustaining themselves also increases significantly by 16% and 7%, respectively. Such an increase of persistence indicates that a more stationary positive geopotential anomaly, presumably reinforced by land-atmosphere interaction, can play an important role in drought development. In particular, Fernando et al. (2016) found that soil moisture was significantly correlated to Z500′ lagged by 2–3 weeks during the 2011 SGP drought and concluded that intensified surface dryness due to cumulative precipitation deficit from early months could intensify the high midtropospheric geopotential anomalies induced by drought; Myoung and Nielsen-Gammon (2010b) also pointed out that a persistent anomalous high in the midtroposphere can enhance negative CINi′ through increasing warm air advection and suppressing rainfall.

It is also shown that, during the SGP drought events, NLSH and DH types have fewer frequencies of transition to some other types, such as WLEH, NHSL, and DL. Meanwhile, the two main wet circulation types, that is, WLEH and DL, both show decreased frequency of occurrence by 3% and 5%, respectively, and decreased frequency of persistence by 5% (not significant) and 12%, respectively.

For the CGP droughts (Figure 10c), frequencies of occurrence for two of the three main dry circulation types, that is, DH and NHSL, increase significantly to 30% and 17%, respectively, about 13% and 7% higher than those of climatological values (Figure 10a). Their frequencies of persistence also increase by 6% and 12%, respectively, relative to their climatological values. The transition from NLSH, WLEH, NHSL, and WHEL to DH increases 17%, 14%, 5%, and 11%, respectively, as well. The stronger increase of DH and of the transitions from other types to DH circulation patterns over the CGP than those over SGP are consistent with the active role of anomalously high pressure system and IVT in determining the CGP drought as suggested by Figures 3 and 6.

In summary, Figure 10 shows that higher DH and lower WLEH frequencies contribute to drought events over both the SGP and the CGP. This is mainly due to an increase of persistence of the DH and increased frequencies of transitions from WLEH to DH. In addition to higher DH and lower WLEH frequencies, drought events over the SGP are also related a higher NLSH frequency and a lower DL frequency. This is mainly due to the increase of frequency of NLSH persistence, and the increase of frequencies of transitions from DL and WLEH to NLSH. For the CGP, drought events are also associated with higher frequencies of WHEL and NHSL types in addition to DH, and a lower frequency of NLSH in addition to WLEH, mainly due to an increase of persistence of WHEL and NHSL, and the transition from NLSH to DH.

5. Conclusions and Discussions

Droughts can be attributed to a variety of anomalous circulation patterns. Conventional statistical analyses, such as correlation, regression and composite methods are limited in providing comprehensive and objective

characterization of these patterns. To address this challenge, we have applied multivariate SOM to characterize large-scale atmospheric patterns that are related to droughts over the SGP and the CGP.

The results in this study have shown a rich tapestry of anomalous atmospheric circulation patterns associated with dry and wet conditions during the warm season over the SGP and the CGP. In particular, we have identified six anomalous patterns associated with anomalously dry and wet conditions over the Great Plains. They are the WLEH, the WHEL, the NLSH, the NHSL, the DL, and the DH. Dry conditions over the SGP are associated with a higher probability of NLSH or DH type, whereas dry conditions over the CGP is associated with higher probability for WHEL, NHSL, or DH types. Node patterns favoring SGP and CGP dryness are well separated by the SOM: Nodes related to SGP dryness mostly indicate increased stability of the lower troposphere (larger negative CIN′) near the high center of Z500′, consistent with a stronger influence of land-atmospheric coupling (e.g., Myoung & Nielsen-Gammon, 2010a); in contrast, the nodes related to CGP dryness are associated with weaker moisture transport. A separate SOM analysis only for very dry days also confirms the different contributions of midtropospheric high, enhanced low-level stability, and reduced moisture transport in strengthening drought condition between the SGP and the CGP. Anomalous Z500 patterns responsible for very dry days are mostly very similar to patterns associated with dry days, but with an increased amplitude. Most very dry nodes for the SGP (in the 4×4 SOM) have larger negative CIN′ near the anomalous high center of the Z500′ than the SGP dry node (in the 5×5 SOM), while almost all the very dry nodes for the CGP have larger IVT deficits than the CGP dry nodes.

These results have shown that our multivariate SOM analysis is able to capture the key anomalous circulation pattern and thermodynamic conditions associated with Great Plains droughts identified in previous studies. For example, the DH pattern features an extensive area of positive geopotential anomalies over the central CONUS, which is consistent with the anomalous ridge over the Great Plains identified by Namias (1983) and Trenberth et al. (1988). The reduced IVT are consistent with the weakened Great Plains low level jet shown by Schubert et al. (2004) and Bronnimann et al. (2009), and stronger CINi′ and impact of land surface feedbacks found by Schubert et al. (2004), Myoung and Nielsen-Gammon (2010a), and Fernando et al. (2016). More importantly, the SOM analysis has identified another three circulation patterns associated with the dry conditions over the SGP (NLSH), and the CGP (WHEL and NHSL) that were overlooked previously. The analysis of the SOM node transitions is also able to provide information that is challenging to obtain by commonly used conventional statistical analyses, for example, a probabilistic characterization of the changes of the multiple anomalous circulation patterns that contribute to Great Plains droughts. In particular, our results show that higher frequencies of NLSH and DH and lower frequencies of WLEH and DL contribute to drought events over the SGP, mainly due to an increase of NLSH and DH persistence, and the higher frequencies of transitions from DL and WLEH to NLSH. Drought events over the CGP are associated with higher WHEL, NHSL, and DH frequencies, and lower WLEH and NLSH frequencies, mainly because of an increase in both persistence of WHEL, NHSL, and DH and transitions from WLEH and NLSH to DH.

This work demonstrates the utility of the SOM in identifying multiple patterns of large-scale circulation and thermodynamic conditions associated with summer rainfall anomalies over the Great Plains by examining the coupling and the transitions among different patterns. However, the SOM method has some limitations. First, as noted in previous sections, there is no clear way to best determine the optimal setting to perform SOM analysis, except to subjectively determine it based on a trial-and-error approach for different combination of parameters (e.g., SOM map shape, node number, topology function, distance function, training epoch, and initial neighborhood radius) and on examination of the quantization error and correlation coefficient. Generally, a larger node number would be better for type assignments (section 4.1) and capturing smoother pattern transition characteristics (section 4.3). Incorporating more variables into the SOM analysis could also be helpful in providing a more complete picture of drought related fields. For example, 250 hPa wind speed and sea level pressure, as representations for circulation patterns of the upper troposphere and surface, respectively (Loikith et al., 2017); soil moisture, as an important variable for depicting drought and its long term memory; precipitable water vapor, similar to IVT, can provide information on moisture source, as well as convective instability (Schiro et al., 2016); outgoing longwave radiation, as proxy for deep convective activities (Rodriguez-Morata et al., 2018). However, adding more variables would require a larger sample number for the SOM clustering to converge properly, which can be limited by the availability of the data samples. Therefore, for the limited 39 year daily data, we only focused on three atmospheric variables and chose a node number of 25 at maximum for this preliminary analysis. One direction of our future work

is to investigate if it is possible to extend the analysis to more variables that have potential influence on precipitation variability of the Great Plains, especially land surface variables, for example, soil moisture and surface fluxes, as long-lasting droughts are demonstrated to be a result of strong land-atmospheric coupling. Another possible direction is to apply this analysis to model output data and evaluate the model's ability to adequately capture the anomalous circulation patterns and thermodynamic conditions that cause drought over the Great Plains. If a model is unable to capture these patterns and conditions, we can then examine how the model's error influences drought predictability over the Great Plains.

Acknowledgments
The authors thank Robert E. Dickinson and Sarah Worden for their very helpful comments and improvement of English. This work was supported by the MAPP Drought Task Force FY17 Competition (NOAA-OAR-CPO-2017-2004896), by the MAPP Climate Test Bed FY16 Competition (NOAA-OAR-CPO-2016-2004413), by the startup fund provided to Rong Fu by the University of California, Los Angeles, and by the National Natural Science Foundation of China (41275112). The ERA Interim data were retrieved using the ECMWF WebAPI services (https://confluence.ecmwf.int/display/WEBAPI/). The CPC Global Unified Precipitation data provided by the NOAA/OAR/ESRL PSD are available from their website (https://www.esrl.noaa.gov/psd). The authors also thank the three anonymous reviewers for their thorough and helpful comments and suggestions.

References

Ambroise, C., Seze, G., Badran, F., & Thiria, S. (2000). Hierarchical clustering of self-organizing maps for cloud classification. *Neurocomputing*, *30*(1-4), 47–52. https://doi.org/10.1016/S0925-2312(99)00141-1

Basara, J. B., Maybourn, J. N., Peirano, C. M., Tate, J. E., Brown, P. J., Hoey, J. D., & Smith, B. R. (2013). Drought and associated impacts in the Great Plains of the United States—A review. *International Journal of Geosciences*, *4*, 72–81.

Bretherton, C. S., Smith, C., & Wallace, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, *5*(6), 541–560. https://doi.org/10.1175/1520-0442(1992)005<0541:Aiomff>2.0.Co;2

Bronnimann, S., Stickler, A., Griesser, T., Ewen, T., Grant, A. N., Fischer, A. M., et al. (2009). Exceptional atmospheric circulation during the "Dust Bowl". *Geophysical Research Letters*, *36*, L08802. https://doi.org/10.1029/2009GL037612

Cavazos, T. (2000). Using self-organizing maps to investigate extreme climate events: An application to wintertime precipitation in the Balkans. *Journal of Climate*, *13*(10), 1718–1732. https://doi.org/10.1175/1520-0442(2000)013<1718:Usomti>2.0.Co;2

Cavazos, T., Comrie, A. C., & Liverman, D. M. (2002). Intraseasonal variability associated with wet monsoons in southeast Arizona. *Journal of Climate*, *15*(17), 2477–2490. https://doi.org/10.1175/1520-0442(2002)015<2477:Ivawwm>2.0.Co;2

Erfanian, A., & Fu, R. (2019). Drier spring over the US Southwest as an important precursor of summer droughts over the US Great Plains. *Atmospheric Chemistry and Physics*, *19*, 15,199–15,216. https://doi.org/10.5194/acp-19-15199-2019

Fernando, D. N., Mo, K. C., Fu, R., Pu, B., Bowerman, A., Scanlon, B. R., et al. (2016). What caused the spring intensification and winter demise of the 2011 drought over Texas? *Climate Dynamics*, *47*(9-10), 3077–3090. https://doi.org/10.1007/s00382-016-3014-x

Fettweis, X., Mabille, G., Erpicum, M., Nicolay, S., & van den Broeke, M. (2011). The 1958–2009 Greenland ice sheet surface melt and the mid-tropospheric atmospheric circulation. *Climate Dynamics*, *36*(1-2), 139–159. https://doi.org/10.1007/s00382-010-0772-8

Hewitson, B. C., & Crane, R. G. (1994). *Neural nets: Applications in geography*. New York: Springer.

Hewitson, B. C., & Crane, R. G. (2002). Self-organizing maps: Applications to synoptic climatology. *Climate Research*, *22*(1), 13–26. https://doi.org/10.3354/cr022013

Hoerling, M., Eischeid, J., Kumar, A., Leung, R., Mariotti, A., Mo, K., et al. (2014). Causes and predictability of the 2012 Great Plains drought. *Bulletin of the American Meteorological Society*, *95*(2), 269–282. https://doi.org/10.1175/Bams-D-13-00055.1

Hoerling, M., Quan, X. W., & Eischeid, J. (2009). Distinct causes for two principal US droughts of the 20th century. *Geophysical Research Letters*, *36*, L19708. https://doi.org/10.1029/2009GL039860

Huang, W. Y., Chen, R. Y., Yang, Z. F., Wang, B., & Ma, W. Q. (2017). Exploring the combined effects of the Arctic Oscillation and ENSO on the wintertime climate over East Asia using self-organizing maps. *Journal of Geophysical Research: Atmospheres*, *122*, 9107–9129. https://doi.org/10.1002/2017JD026812

Johnson, N. C., Feldstein, S. B., & Tremblay, B. (2008). The continuum of Northern Hemisphere teleconnection patterns and a description of the NAO shift with the use of self-organizing maps. *Journal of Climate*, *21*(23), 6354–6371. https://doi.org/10.1175/2008jcli2380.1

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.., xxixp. 487). New York: Springer.

Kohonen, T. (1990, The self-organizing map). *Proceedings of the IEEE*, *78*(9), 1464–1480. https://doi.org/10.1109/5.58325

Kotkin, J., Zimmerman, D., & Mulligan, K. (2012). The rise of the Great Plains: Regional opportunity in the 21st centuryRep., Texas Tech University.

Liu, Y., & Weisberg, R. H. (2011). A review of self-organizing map applications in meteorology and oceanography. In *Self organizing maps—Applications and novel algorithm design*, edited, InTech. https://doi.org/10.5772/13146

Loikith, P. C., Lintner, B. R., & Sweeney, A. (2017). Characterizing large-scale meteorological patterns and associated temperature and precipitation extremes over the Northwestern United States using self-organizing maps. *Journal of Climate*, *30*(8), 2829–2847. https://doi.org/10.1175/Jcli-D-16-0670.1

Malmgren, B. A., & Winter, A. (1999). Climate zonation in Puerto Rico based on principal components analysis and an artificial neural network. *Journal of Climate*, *12*(4), 977–985. https://doi.org/10.1175/1520-0442(1999)012<0977:Cziprb>2.0.Co;2

Mattingly, K. S., Ramseyer, C. A., Rosen, J. J., Mote, T. L., & Muthyala, R. (2016). Increasing water vapor transport to the Greenland Ice Sheet revealed using self-organizing maps. *Geophysical Research Letters*, *43*, 9250–9258. https://doi.org/10.1002/2016GL070424

Myoung, B., & Nielsen-Gammon, J. W. (2010a). The convective instability pathway to warm season drought in Texas. Part I: The role of convective inhibition and its modulation by soil moisture. *Journal of Climate*, *23*(17), 4461–4473. https://doi.org/10.1175/2010jcli2946.1

Myoung, B., & Nielsen-Gammon, J. W. (2010b). The convective instability pathway to warm season drought in Texas. Part II: Free-tropospheric modulation of convective inhibition. *Journal of Climate*, *23*(17), 4474–4488. https://doi.org/10.1175/2010jcli2947.1

Namias, J. (1983). Some causes of United-States drought. *Journal of Climate and Applied Meteorology*, *22*(1), 30–39. https://doi.org/10.1175/1520-0450(1983)022<0030:Scousd>2.0.Co;2

Pu, B., Dickinson, R. E., & Fu, R. (2016). Dynamical connection between Great Plains low-level winds and variability of central Gulf States precipitation. *Journal of Geophysical Research: Atmospheres*, *121*, 3421–3434. https://doi.org/10.1002/2015JD024045

Quan, X. W., Hoerling, M. P., Lyon, B., Kumar, A., Bell, M. A., Tippett, M. K., & Wang, H. (2012). Prospects for dynamical prediction of meteorological drought. *Journal of Applied Meteorology and Climatology*, *51*(7), 1238–1252. https://doi.org/10.1175/Jamc-D-11-0194.1

Reusch, D. B., Alley, B. A., & Hewitson, B. C. (2005). Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. *Polar Geography*, *29*(3), 188–212. https://doi.org/10.1080/789610199

Rodriguez-Morata, C., Ballesteros-Canovas, J. A., Rohrer, M., Espinoza, J. C., Beniston, M., & Stoffel, M. (2018). Linking atmospheric circulation patterns with hydro-geomorphic disasters in Peru. *International Journal of Climatology*, *38*(8), 3388–3404. https://doi.org/10.1002/joc.5507

Roundy, J. K., Ferguson, C. R., & Wood, E. F. (2014). Impact of land-atmospheric coupling in CFSv2 on drought prediction. *Climate Dynamics*, *43*(1-2), 421–434. https://doi.org/10.1007/s00382-013-1982-7

Schiro, K. A., Neelin, J. D., Adams, D. K., & Lintner, B. R. (2016). Deep convection and column water vapor over tropical land versus tropical ocean: A comparison between the Amazon and the tropical Western Pacific. *Journal of the Atmospheric Sciences*, *73*(10), 4043–4063. https://doi.org/10.1175/Jas-D-16-0119.1

Schubert, S. D., Suarez, M. J., Pegion, P. J., Koster, R. D., & Bacmeister, J. T. (2004). On the cause of the 1930s Dust Bowl. *Science*, *303*(5665), 1855–1859. https://doi.org/10.1126/science.1095048

Seager, R., & Hoerling, M. (2014). Atmosphere and ocean origins of North American droughts. *Journal of Climate*, *27*(12), 4581–4606. https://doi.org/10.1175/Jcli-D-13-00329.1

Swales, D., Alexander, M., & Hughes, M. (2016). Examining moisture pathways and extreme precipitation in the US Intermountain West using self-organizing maps. *Geophysical Research Letters*, *43*, 1727–1735. https://doi.org/10.1002/2015GL067478

Trenberth, K. E., Branstator, G. W., & Arkin, P. A. (1988). Origins of the 1988 North-American drought. *Science*, *242*(4886), 1640–1645. https://doi.org/10.1126/science.242.4886.1640

Van den Dool, H. M. (2007). *Empirical methods in short-term climate prediction*, xxiii (p. 215). Oxford, New York: Oxford University Press.

Weaver, S. J., Ruiz-Barradas, A., & Nigam, S. (2009). Pentad evolution of the 1988 drought and 1993 flood over the Great Plains: An NARR perspective on the atmospheric and terrestrial water balance. *Journal of Climate*, *22*(20), 5366–5384. https://doi.org/10.1175/2009jcli2684.1

Zhuang, Y. Z., Fu, R., & Wang, H. Q. (2018). How do environmental conditions influence vertical buoyancy structure and shallow-to-deep convection transition across different climate regimes? *Journal of the Atmospheric Sciences*, *75*(6), 1909–1932. https://doi.org/10.1175/Jas-D-17-0284.1