Three-dimensional clustering in the characterization of spatiotemporal drought dynamics: cluster size filter and drought indicator threshold optimization

Vitali Diaz[1,2], Gerald A. Corzo Perez[1], Henny A.J. Van Lanen[3], Dimitri Solomatine[1,2,4]
[1]IHE Delft Institute for Water Education, Delft, 2601 DA, the Netherlands
[2]Delft University of Technology, Delft, the Netherlands
[3]Hydrology and Quantitative Water Management Group, Wageningen University, Wageningen, the Netherlands
[4]Water Problems Institute of the Russian Academy of Sciences, Moscow, Russia

Abstract

In its three-dimensional (3-D) characterization, drought is approached as an event whose spatial extent changes over time. Each drought event has an onset and end time, a location, a magnitude, and a spatial trajectory. These characteristics help to analyze and describe how drought develops in space and time, i.e., drought dynamics. Methodologies for 3-D characterization of drought include a 3-D clustering technique to extract the drought events from the hydrometeorological data. The application of the clustering method yields small 'artifact' droughts. These small clusters are removed from the analysis with the use of a cluster size filter. However, according to the literature, the filter parameters are usually set arbitrarily, so this study concentrated on a method to calculate the optimal cluster size filter for the 3-D characterization of drought. The effect of different drought indicator thresholds to calculate drought is also analyzed. The approach was tested in South America with data from the Latin American Flood and Drought Monitor (LAFDM) for 1950–2017. Analysis of the spatial trajectories and characteristics of the most extreme droughts is also included. Calculated droughts are compared with information reported at a country scale and a reasonably good match is found.

Keywords

spatiotemporal drought analysis, drought tracking, drought dynamics, drought characterization, drought clustering

1 Introduction

In recent decades, methods for drought calculation have increasingly treated this phenomenon as an event characterized in space and time (Andreadis et al., 2005; Corzo Perez et al., 2011; Diaz et al., 2020a, 2020b; Herrera-Estrada and Diffenbaugh, 2020; Lloyd-Hughes, 2012; Sheffield et al., 2009; van Huijgevoort et al., 2013; Vernieuwe et al., 2020). Each drought event

is calculated considering different characteristics, such as duration, spatial extent, and location. A better characterization of drought improves the analysis of its behavior and its possible effects on different economic and environmental sectors (World Meteorological Organization (WMO), 2006). Treating drought as an event with spatial extent has also allowed new methods for its monitoring and prediction to be proposed (Diaz et al., 2018, 2020a).

The approach to calculating drought in three dimensions (longitude, latitude, and time) has followed a gradual process. One of the first works to consider drought with a spatial extent (area) was Yevjevich (1967). He used spatially distributed synthetic precipitation data to define the drought areas. The time series of drought areas allowed him to calculate the onset, end, duration, and magnitude of drought. Another key work that marked a before and after in calculating drought was the research of Andreadis et al. (2005). They introduced a methodology for the calculation of drought areas. In this methodology, the areas are contiguous in space. By using a clustering technique, the different areas (2-D clusters) are calculated. Later, this method was expanded to three dimensions. The works of Corzo Perez et al. (2011) and Lloyd-Hughes (2012) are some of the first examples in which drought is calculated as a 3-D object. Drought characterization includes new features such as volume (number of cells) and location. The location is defined as the centroid of the 3-D cluster. Subsequently, the centroids of the drought are used to monitor its spatial trajectory, i.e., drought tracking. In works such as those of Diaz et al. (2020a, 2020b), Herrera-Estrada et al. (2017), and Herrera-Estrada and Diffenbaugh (2020), spatial trajectories of drought are calculated, serving to analyze the dynamics of drought. The 3-D characterization of drought opens up other possibilities for drought monitoring and prediction. In addition to predicting drought duration and magnitude, its spatial extent, location, and trajectory could also be predicted (Diaz et al., 2018, 2020a).

Research gaps remain in the topic of 3-D drought characterization. When calculating drought, it is common practice to apply a cluster size filter to remove small clusters resulting from the methodology and not from the droughts themselves. In most cases, the cluster size filter is chosen arbitrarily, or such a choice is generally driven by the past work or experience. Another parameter commonly chosen arbitrarily is the drought indicator threshold (Sect. 2.1), i.e., the value for which a value of the hydrometeorological variable is considered a drought. Although there is extensive research on the latter, its combined effect with the cluster size filter has not been fully analyzed.

This research proposes a method to calculate the optimal cluster size filter for the 3-D characterization of drought. Droughts are calculated with a 3-D clustering technique using different cluster size filters and drought indicator thresholds. The analysis of the most extreme

droughts in the analysis period is also presented. The methodology was tested in South America, using data from the Latin American Flood and Drought Monitor (LAFDM) (Sheffield et al., 2014; Wood et al., 2016) for 1950–2017.

2 Methods and data

2.1 Drought calculation

In this methodology, drought is represented by a 3-D cluster. The data must be organized in a grid system where each cell is geographically referenced (latitude and longitude). The arrangement of these geospatial data cells can be schematized as a cube with latitude, longitude, and time as its sides (Figure 1). Each 3-D drought cluster is made up of cells that indicate drought (Figure 1). The methodology considers three main steps: (1) calculation of the drought indicator, (2) identification of the cells in drought, and (3) calculation of the 3-D clusters (Figure 1).

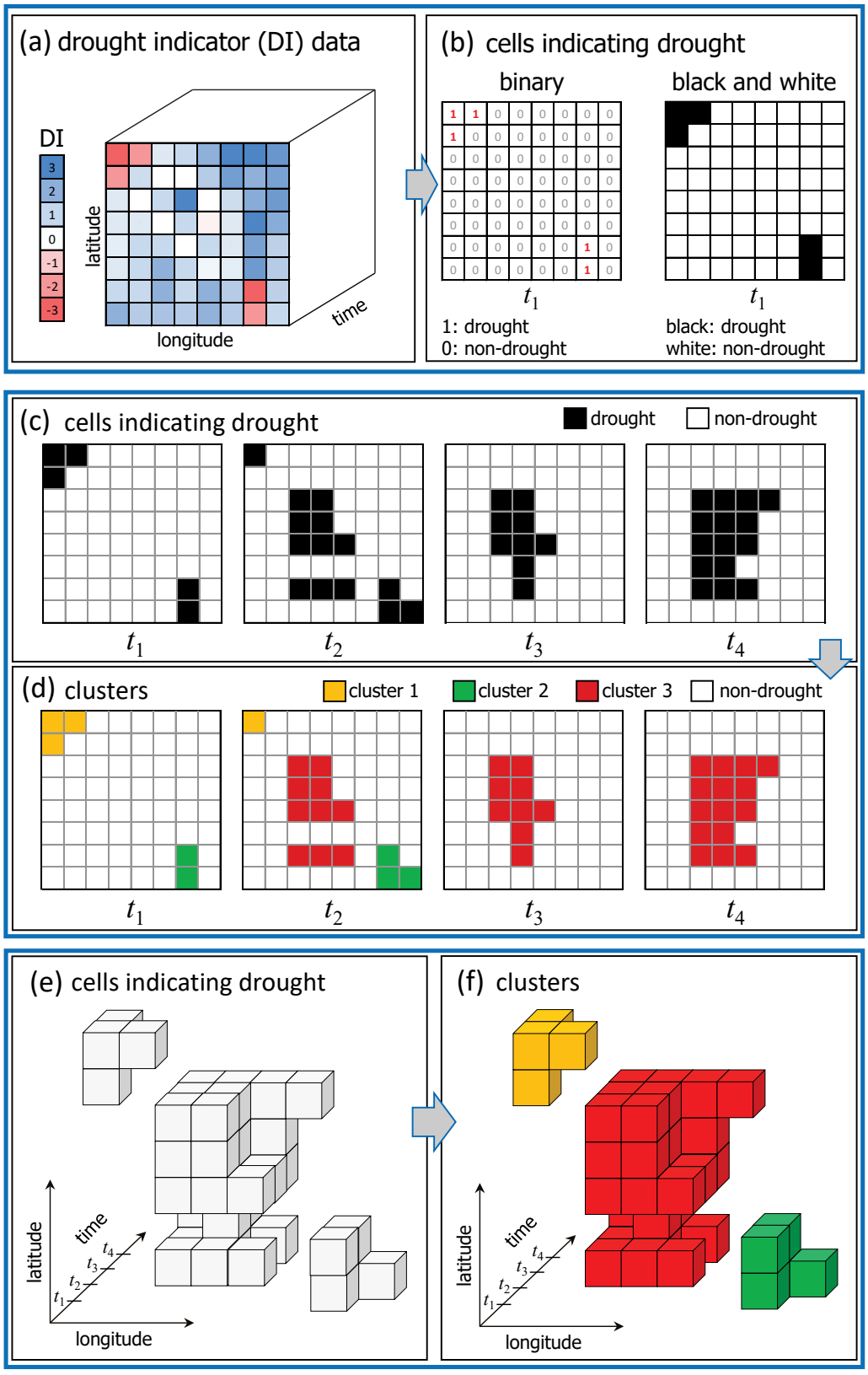
First, the drought indicator is calculated with the hydrometeorological data. This indicator makes it possible to define anomalies in the data by using a statistic (Figure 1a). The standardized drought indicators, for example, standardize the values of the hydrometeorological variable. Using these standardized values, the drought is identified by means of a threshold associated with a statistical value that defines an anomaly as drought. Values lower than the threshold are considered as drought. Other indicators use a threshold value directly applied to the hydrometeorological variable for the entire analysis period or a moving time window. In this second method, the values that are below the threshold are also considered as drought.

After drought indicator calculation, the indication that a cell is in drought is carried out with the drought indicator data (second step). The binary classification is used to indicate drought, i.e., the use of 1s and 0s (Figure 1b, 1c, and 1e). In this way, drought is identified with 1s and non-drought with 0s, as indicated in Eq. 1. When the drought indicator (DI) is below a selected threshold (T), a cell is in drought. In Eq. 1, D_S stands for drought state, i.e., drought (1) or non-drought (0).

$$D_S(t) = \begin{cases} 1 & \text{if } \mathrm{DI}(t) \le T \\ 0 & \text{if } \mathrm{DI}(t) > T \end{cases} \quad \text{(Eq. 1)}$$

This research also analyzes the effect of the threshold (T) on the calculation of droughts (3-D clusters). The thresholds of 0, −1, −1.5, and −2 are tested. It is noted that, in standardized drought indices, the values equal to or below zero indicate ‘drought’. Eq. 1 is applied in each cell in each time step (t).

101



102

103 **Figure 1.** Schematic overview of the methodology for 3-D clusters calculation: (1) calculation of the
104 drought indicator (a), (2) identification of cells in drought (b, c, d), and (3) calculation of 3-D clusters
105 (d and f). 2-D view in each time step (c and d) and 3-D view (a, e, and f).

The third step is the calculation of the 3-D clusters. A clustering method to extract the drought events is applied following Corzo Perez et al. (2011). The drought events correspond to the 3-D clusters calculated with the data of 0s and 1s (Figure 1d and 1f). The following describes the unsupervised machine learning-based method to extract the clusters. The method follows the connected-component labeling approach to cluster the cells in drought (Haralick and Shapiro, 1992). In this method, a two-scan algorithm is applied. First, each cell is numbered. Then, the first run is carried out, in which the binary grid is explored and provisional labels are assigned to connected (contiguous) components (cells). These labels identify the connection of every cell with its nearest neighbors. In this first run, the cell's label does not yet refer to the cluster number but to the cells with which the given cell is connected. Finally, the second scan is carried out to find similar cell connections, i.e., clusters, which are given a unique label. Examination of the grid can be performed by columns or by rows. The clustering method is conducted over the whole binary data (Figure 1d and 1f).

As mentioned, it is a common practice to remove small 3-D clusters that constitute 'artifact' droughts produced by the clustering technique. This task uses a cluster size filter to clean the number of calculated 3-D clusters. This cluster size filter is usually set based on similar studies or experiences, or left at the default value. In this research, we propose a method to calculate the optimal cluster size filter introduced in the following. To test this method, we used the cluster size filters of 0, 4, 9, 16, 25, 36, 49, 64, 81, and 100 cells to remove clusters. The value of 0 indicates no cluster cleaning. The procedure is as follows. For each time step (t), clustering was carried out to identify 2-D clusters. Afterward, the 2-D clusters below each cluster size filter were removed. Finally, the 3-D clusters were identified for each cleaned sample data. We carried out the procedure in this way because, owing to the large size of this region and the resolution of the data, extremely large events were identified when applying the 3-D clustering. The subsequent filtering of 3-D clusters was not practical, i.e., there were no or few small 3-D clusters to remove. By applying the cluster cleaning to the 2-D clusters in each time step, we removed small isolated areas (2-D clusters) throughout the region, which was found a more effective means of cleaning the 3-D cluster data.

For the identification of 3-D drought clusters, we also considered a filter of duration. All the 3-D clusters of one-month duration were excluded from further analysis. We made sure that none of these clusters was of a large number of cells, thus avoiding the risk of removing severe events. The 3-D clusters of one-month duration were removed for each case of cluster size filter and drought indicator threshold.

Drought calculation concluded with the selection of the cluster size filter (Figure 2). The optimal cluster size was defined as the value in the curve 'cluster size filter vs. the number of 3-D clusters,' at which the number of clusters stabilizes, i.e., does not undergo considerable changes for a constant increase of the cluster size filter (Figure 2). This point is defined as the vertex of the curve (Figure 2). To identify this point, we developed a method based on the angle *C* formed between two continuous segments of the curve 'cluster size filter vs. number of 3-D clusters' (Figure 2). We calculated this angle *C* using the law of cosines (Eq. 2), which considers the length of the sides of the triangle *a*, *b*, and *c* formed by the coordinates of three subsequent points, i.e., the points (x_1, y_1), (x_2, y_2), and (x_3, y_3) (Figure 2). For each triad of points, the angle *C* was calculated. The vertex corresponds to the smallest angle of all the calculated *C*s. The optimal cluster size filter was calculated for each drought indicator threshold.

$$\cos C = \left(a^2 + b^2 + c^2\right) \Big/ \left(2ab\right) \quad \text{(Eq. 2)}$$

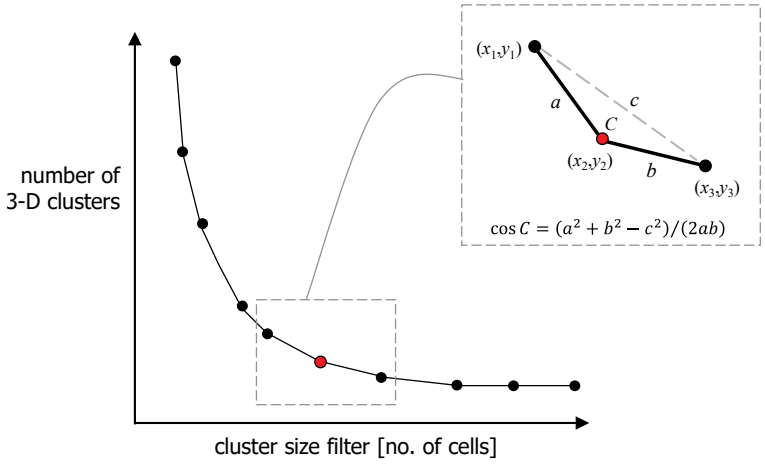


Figure 2. Scheme of the method to calculate the optimal cluster size filter. The angle *C* is calculated with the sides *a*, *b*, and *c* of the triangle formed by the three subsequent points (zoomed-in view).

2.2 Drought characterization

After identifying the droughts, the onset and end in time, duration, and severity (magnitude) were calculated for each drought (3-D cluster). Drought duration (*dd*) and magnitude (*ds*) were obtained with Eqs. 3 and 4, respectively. The times *ti* and *tf* are the onset and end of each drought, respectively. DA is the drought area (number of cells) at each time step *t*. Eqs. 3 and 4 were applied for each 3-D cluster.

$$dd = ti - tf + 1 \quad \text{(Eq. 3)}$$

$$ds = \sum_{t=ti}^{tf} \mathrm{DA}\left(t\right) \quad \text{(Eq. 4)}$$

For each drought, the spatial trajectory was also calculated. These trajectories were built with the union of the centroids of the drought areas at each time step (t), following Diaz et al. (2020a). Trajectories of the largest events were analyzed.

2.3 Data

For the identification of droughts, we used data of the Standardized Precipitation Index (SPI) retrieved from the Latin American Flood and Drought Monitor (LAFDM) (https://platform.princetonclimate.com/PCA_Platform/lafdmLanding.html). Within LAFDM, the SPI is calculated with precipitation and numerically indicates the magnitude of the water anomaly, and runs from -3 to 3. Being below zero, SPI shows a dryness condition, whereas above zero, it indicates a wetness condition (McKee et al., 1993). The aggregation period of six months in the SPI, denoted by SPI6, is considered a good proxy for monitoring drought condition on the surface, i.e., on runoff and soil moisture, which is more relevant to the potential impact on agricultural activities (WMO, 2012). Thus, SPI6 data was considered for the calculation of droughts. The period of the analysis was 1950–2017 (816 months) on a monthly basis. The spatial resolution was 0.25 deg.

3 Results and discussion

3.1 Drought calculation

Figures 3 and 4 show the duration, magnitude (number of cells), and number of 3-D drought clusters (nc) for each drought indicator threshold and cluster size filter. Figure 3 presents the results for all the durations and magnitudes. For better visualization of the results, Figure 4 was prepared, which deploys the droughts for durations up to 20 months and 10,000 cells. Results show that, for the threshold of 0, the least number of droughts is achieved (Figures 3 and 4). These clusters are smaller in quantity, but their structure is made up of more cells than those of the other thresholds, as shown also in Figure 7. As the drought indicator threshold decreases (from 0 to −2), more clusters are identified, although they have fewer cells (Figures 3 and 4). This increase in the number of clusters is not constant; there is a decrease when the threshold has a small value such as drought indicator ≤ -2. The latter indicates that the number of clusters with extreme drought (drought indicator ≤ -2) tends to be smaller. In the case of the cluster size filter (number of cells), the decrease in the number of clusters is more evident than in the drought indicator threshold (Figures 3 and 4). Results show that when the cluster size filter increases, it separates large clusters into smaller clusters with shorter durations.

Figures 3 and 4 display the clusters of one-month duration. Although these clusters have a number of cells greater than the cluster size filter in each case, they are small compared to the rest of the clusters in each case (Figures 3 and 4); for this reason, they were also removed. Figures 5 and 6 show the final results after removing the one-month duration clusters. Figure 5 shows the droughts for all durations and magnitudes. In Figure 5, scarce events with durations greater than 100 months and a considerable number of cells are observed; most droughts have durations of less than 20 months. Figure 6 shows the results for droughts of up to 20 months and 10,000 cells. A direct relationship is observed between duration and magnitude.

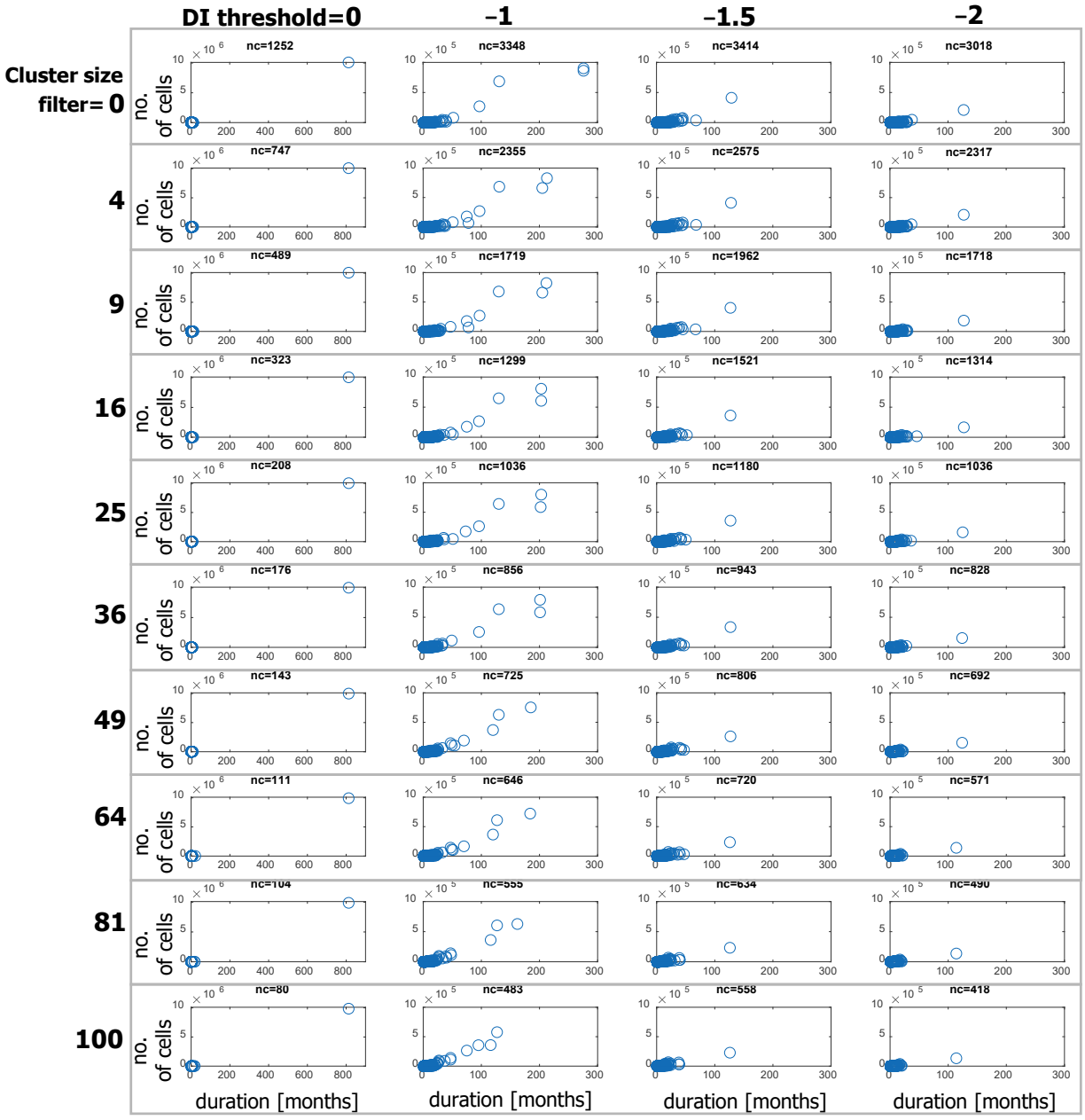


Figure 3. Number of 3-D clusters (nc) calculated for different drought indicator (DI) thresholds and cluster size filters. The duration (months) and magnitude (number of cells) are indicated. Note: The number of cells for the first threshold is up to 10×10^6, while for the rest it is up to 10×10^5.

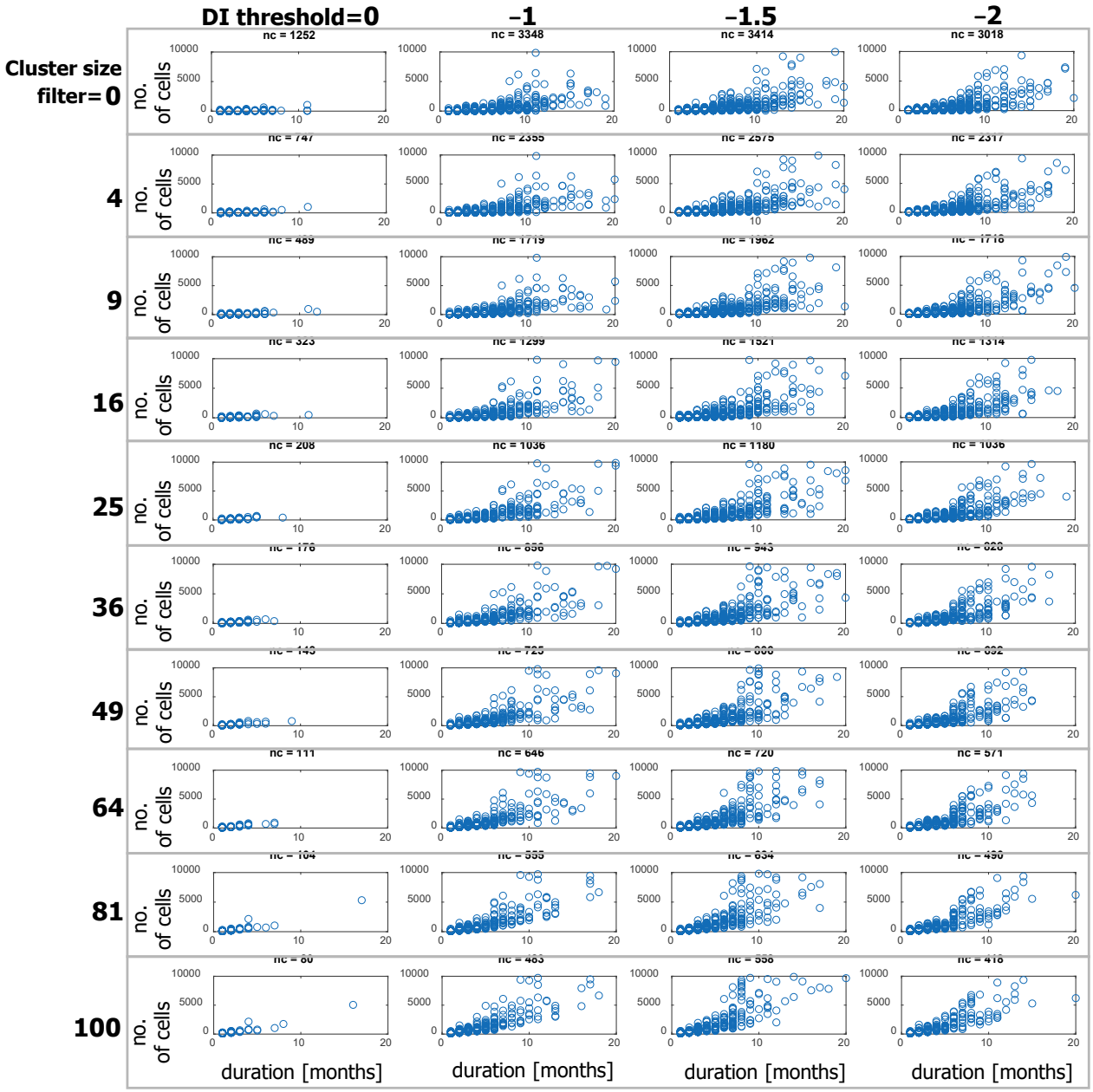


Figure 4. As Figure 3 but for durations up to 20 months and sizes up to 10,000 cells.

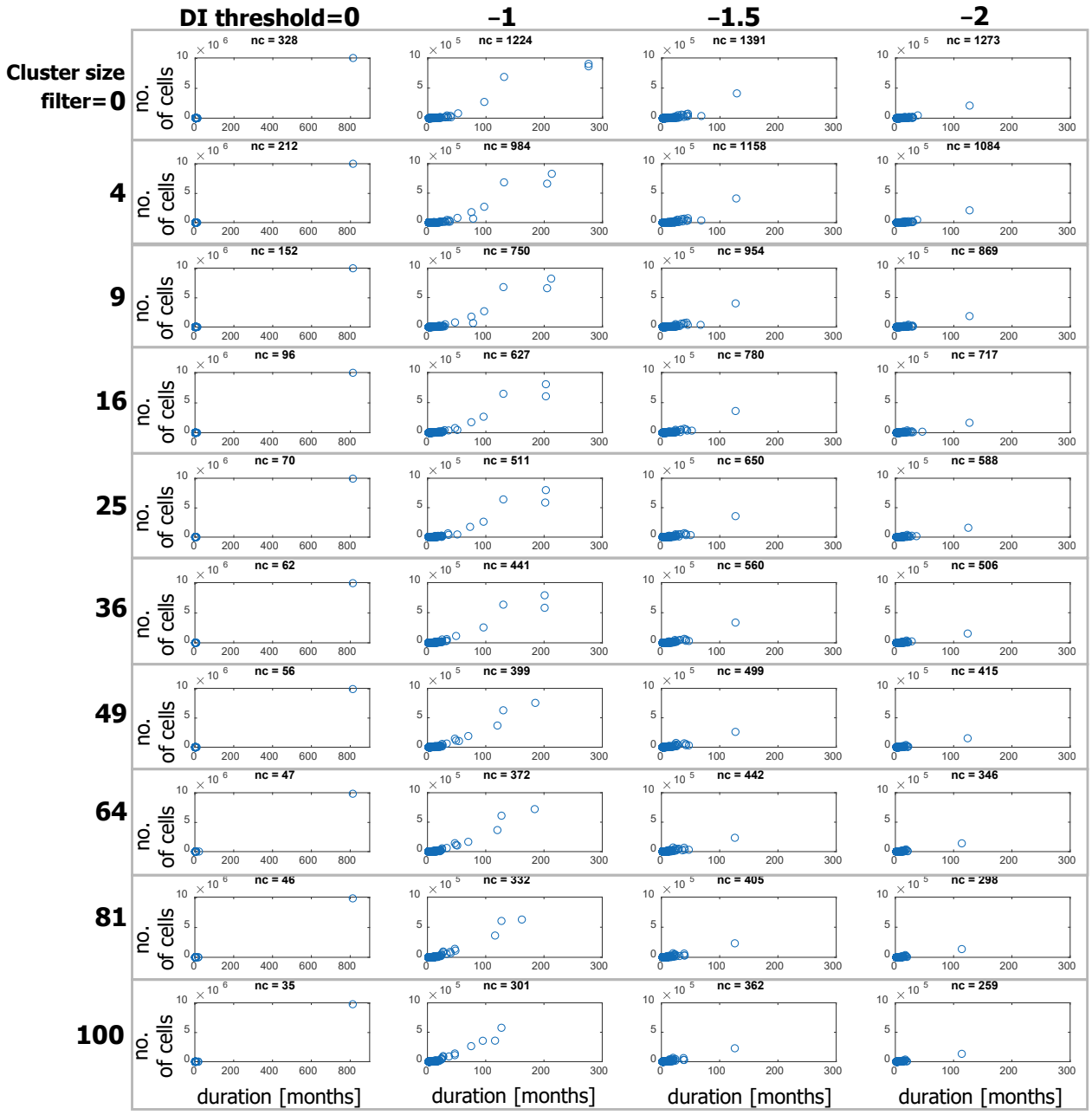


Figure 5. Number of 3-D clusters (nc) calculated for different drought indicator (DI) thresholds and cluster size filters. The duration (months) and magnitude (number of cells) are indicated. Note: The number of cells for the first threshold is up to 10×10^6, while for the rest it is up to 10×10^5. In these results, the 3-D clusters of one-month duration were excluded.

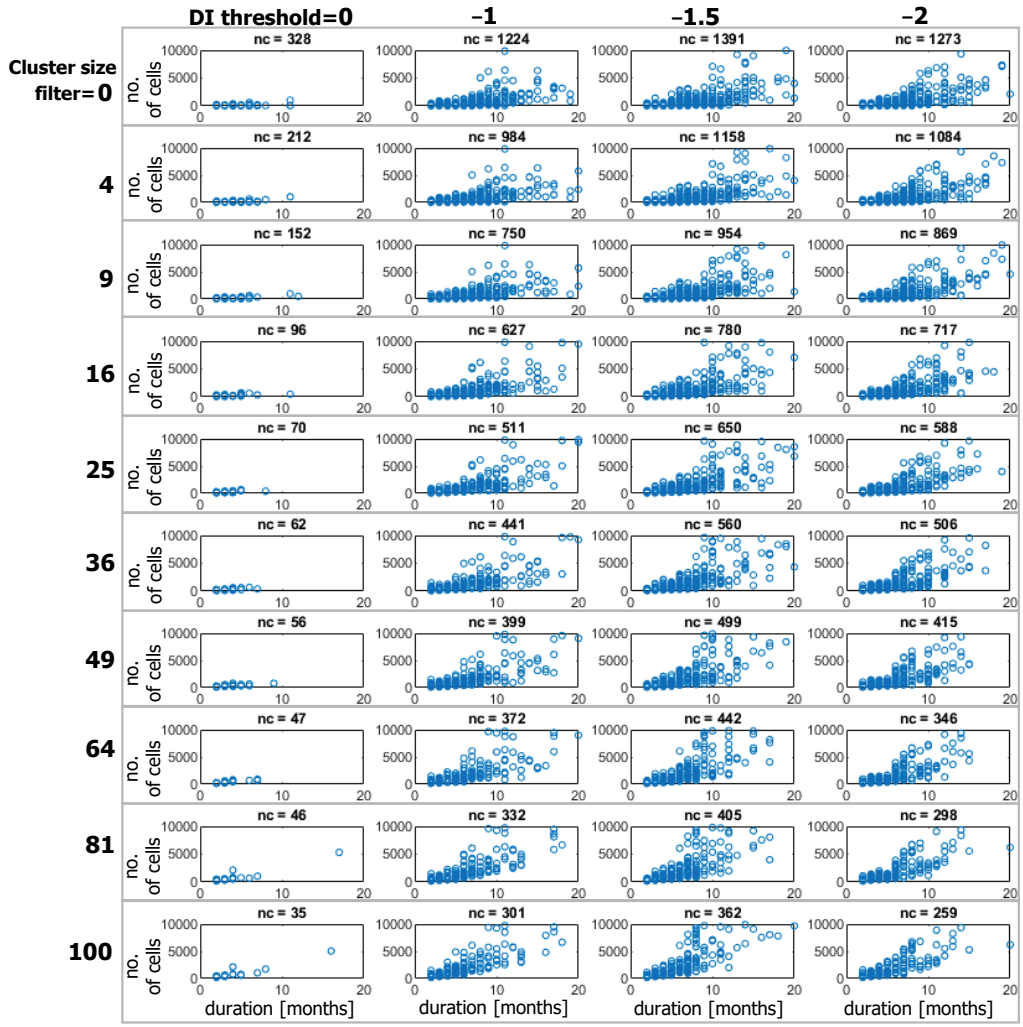


Figure 6. As Figure 5 but for durations up to 20 months and sizes up to 10,000 cells.

Figure 7 shows the percentages of drought area (PDAs) without considering the one-month duration clusters. PDAs were calculated for each 3-D cluster as the number of cells of each drought area at each time step divided by the total number of cells of the region (24,877). This figure helps visually compare the durations and magnitudes of the events for each threshold. Figure 7 shows a drought event with a duration almost equal to the analysis period when the threshold is 0. This result shows that, even in this large study area, there are consecutive drought areas in time with at least 20% of the total study area that are connected to each other and form the long-lasting 3-D cluster. For the threshold of 0, the use of the cluster size filter does not show any significant difference. In the other thresholds, more events with shorter durations and less extensive areas are observed. In general, as the threshold decreases, indicating a more severe drought, the events are smaller in magnitude (size) and duration. The results also show that the 2010–2020 decade had more events with considerable magnitudes that indicated severe and extreme drought.

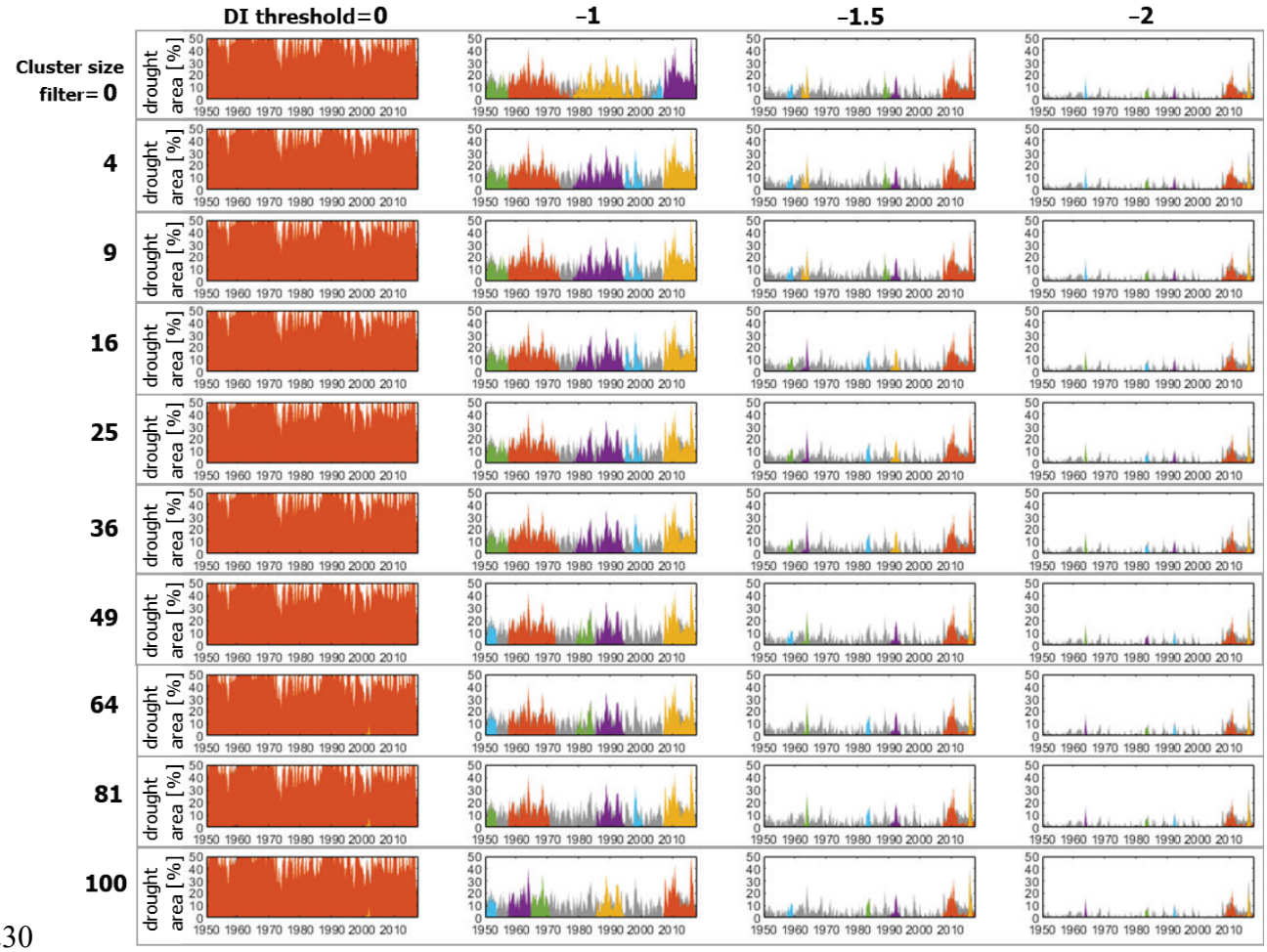


Figure 7. Percentage of drought area calculated for each 3-D cluster. Results for each drought indicator (DI) threshold and cluster size filter. Drought areas are shown for the period 1950–2017. Drought events are indicated in different colors.

The number of 3-D clusters for each drought indicator threshold and cluster size filter is presented in Figure 8. The method developed to find the optimal cluster size filter (Sect. 2.1) shows, for instance, that, for the threshold of −1, the cluster size filter is 49 cells. Table A1 shows the results for each of the thresholds. In the following sections, the results of the drought characterization are shown for this optimal cluster size filter. For the case of the drought indicator threshold, we focused our analysis on the threshold of −1. In general, the −1.5 and −2 thresholds produced more events but with shorter durations and magnitudes (sizes) (Figure 7). The threshold of 0 produced a long-lasting event and some small events. Based on the results, this threshold of 0 is not recommended for drought calculation.

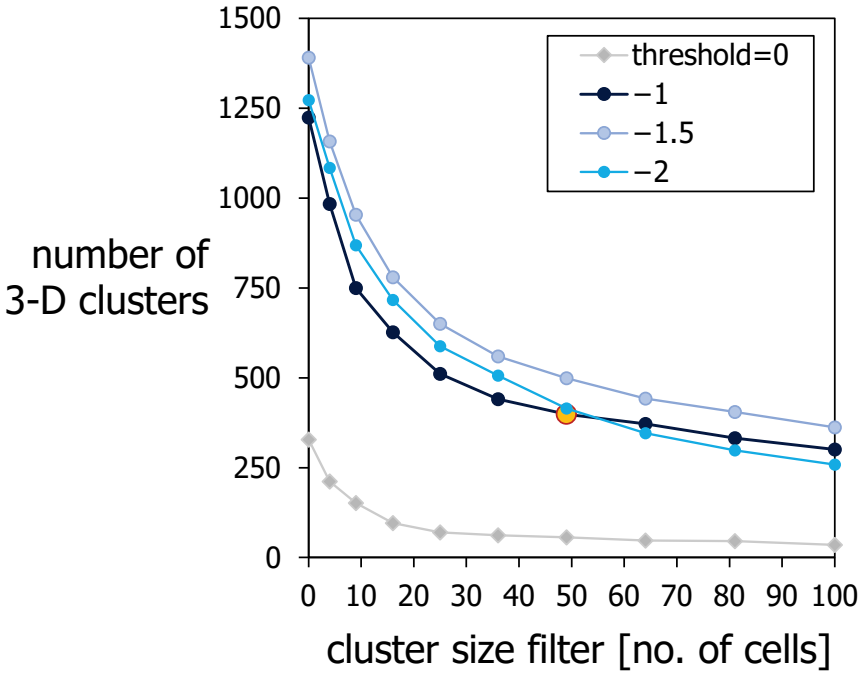


Figure 8. Identification of the optimal cleaning filter size. The result for the threshold of −1 is highlighted.

3.2 Drought characterization

A total of 399 drought events were identified. Figure 9 shows that the distribution of the duration and magnitude (number of cells) is mainly concentrated in durations of less than 20 months and 10,000 cells. Few events lasting around 24 months or longer are observed. The linear correlation coefficient between the duration and magnitude is $R^2 = 0.89$, indicating an almost linear relationship between these two characteristics.

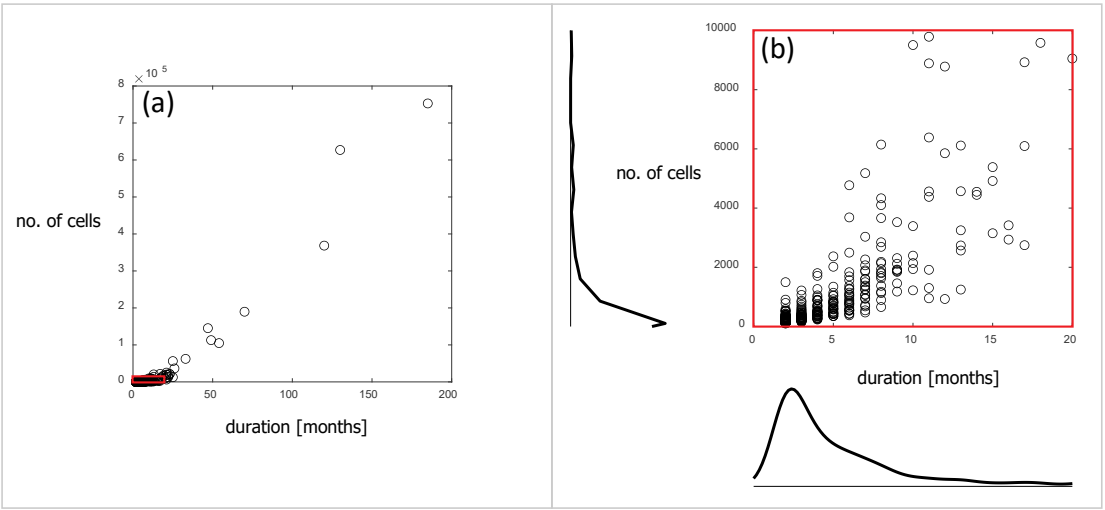


Figure 9. (a) Duration (months) and magnitude (number of cells) of the droughts for the period 1950–2017. (b) Detail for drought durations up to 20 months and 10,000 cells.

The centroids of each of the 3-D clusters are shown in Figure 10a. The centroids are classified by their duration in four intervals: 2 to 6, 7 to 12, 13 to 24, and 24 or more months. The centroids of the clusters of two to six months are observed practically throughout the entire study area, as well as the centroids of the seven- to 12-month events. Those of durations of 13 to 24 months are also seen, although with less density. Most of the centroids of 3-D clusters with durations greater than 24 months are located in the central region of the study area, although two centroids are located outside this region, one in the south and one in the north. Figure 10b shows the number of 3-D clusters counted in each cell. In general, 14 or more clusters are observed over the study area, except in some areas of the Amazon basin, upper Magdalena River basin in Colombia, the north of Uruguay, the northwest of Argentina, and the south of Chile. One of the areas with the least occurrence is the lower Amazon basin, near the discharge.

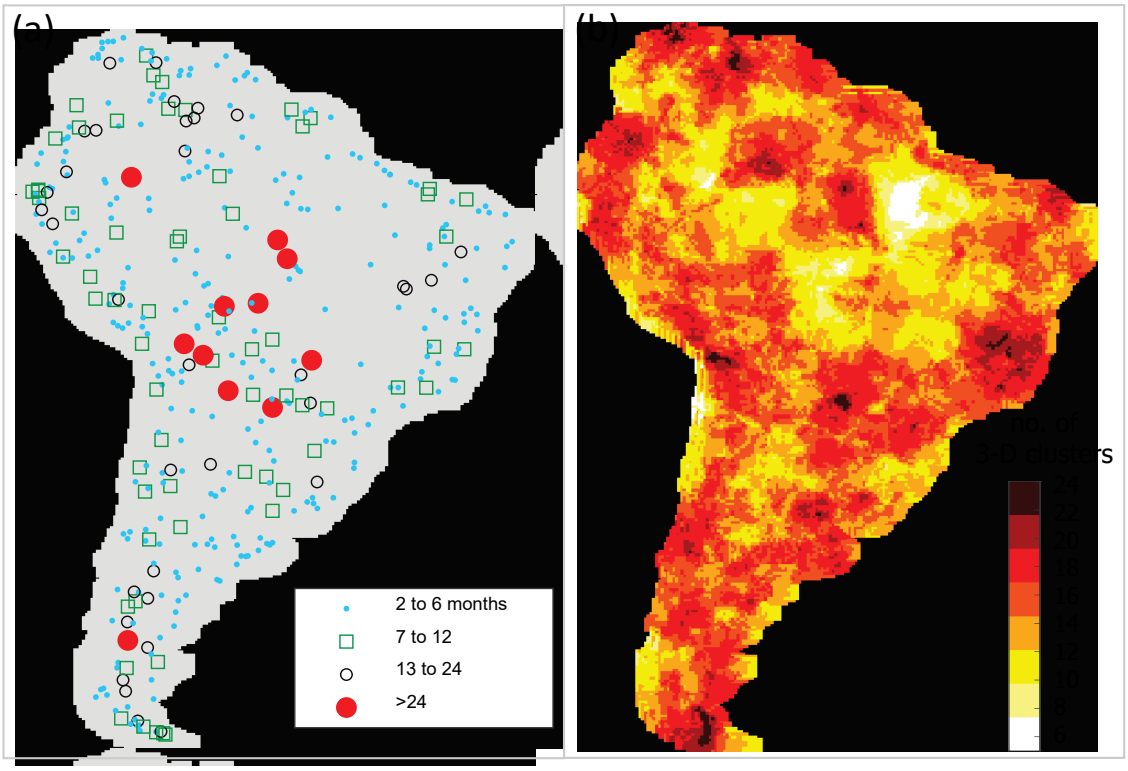


Figure 10. (a) Centroids of the 3-D clusters for the period 1950–2017. Centroids are classified by cluster duration [months]. (b) Occurrence of the 3-D clusters: each cell indicates the number of 3-D clusters that took place there in the period 1950–2017.

The following paragraphs present the results of three of the most extreme droughts: those of 1957–1972, 2006–2017, and 1984-1994.

The drought with the longest duration is shown in Figure 11. The drought lasted from May 1957 to September 1972 (185 months). The 3-D cluster is shown in Figure 11a. It is observed

that this cluster completely covered the region (Figure 11a and 11b). However, its spatial distribution over time varied significantly, concentrating mainly in the north and south of the subcontinent. Figure 11b shows the distribution of the duration of each cell in the study area. This figure shows the amount of time each cell was in drought for this particular drought. In general, the central, north–central, and south–central regions had the longest amount of time in drought. The spatial trajectories of this drought are shown in Figure 11c. It is observed that the centroids of each drought area follow a pattern in which the trajectories go from northeast to south, northwest, and back to northeast, mostly in a clockwise direction, although the direction is counterclockwise in the last two years.

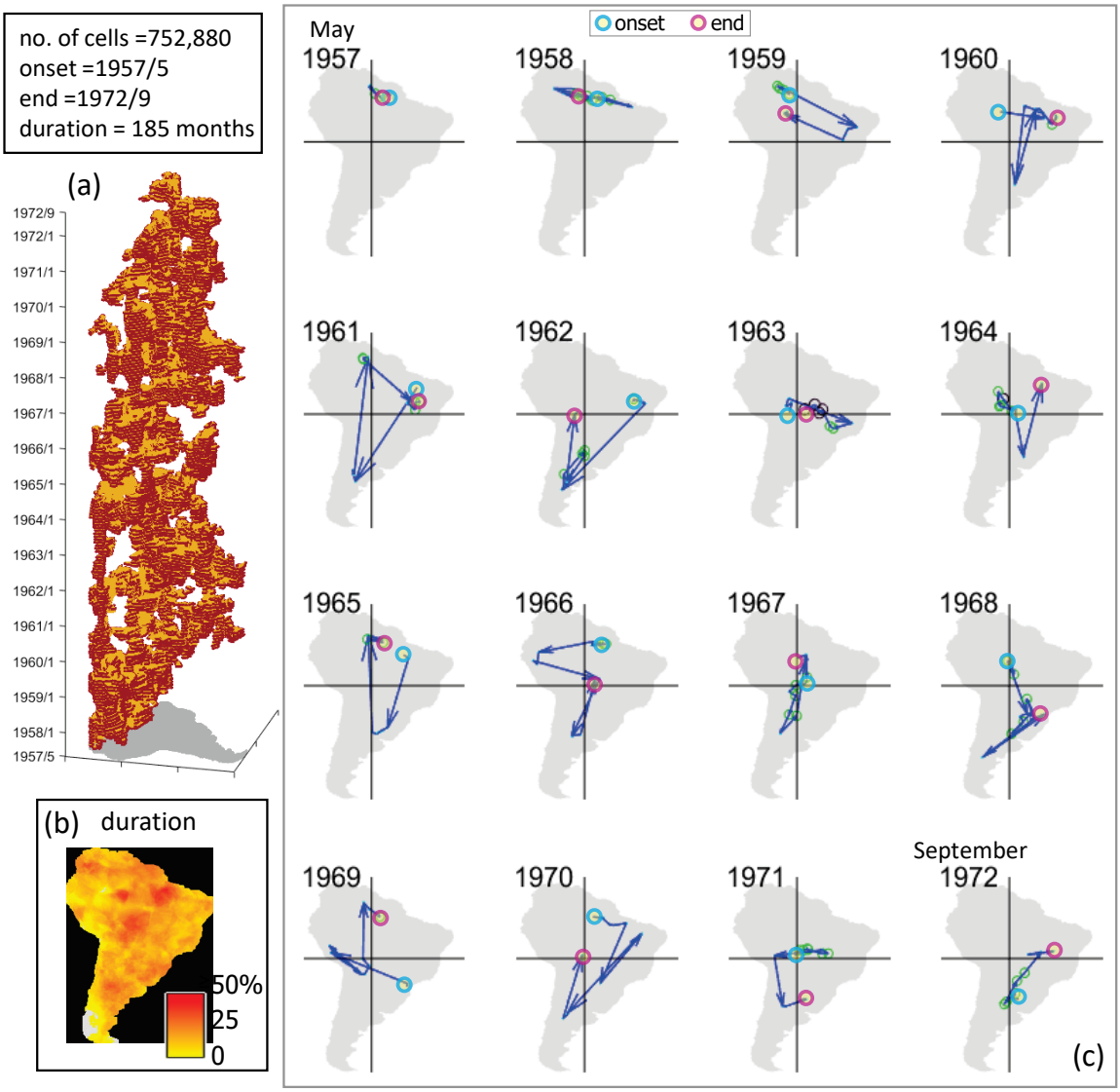


Figure 11. (a) Drought from May 1957 to September 1972. (b) Percentage of duration: each cell shows the percentage of time in drought with respect to the duration of the indicated drought. (c) Monthly drought trajectories per year; the onset and the end of the drought trajectory are indicated for each year by blue and red circles.

The second-longest drought is shown in Figure 12. This drought lasted from December 2006 to September 2017 (130 months). The 3-D cluster is shown in Figure 12a. The event was mainly concentrated in the southwestern and southern coast of South America and some areas of Brazil (Figure 12a and 12b). The trajectories of this drought presented in Figure 12c show a different dynamic from the drought shown previously (Figure 11c). The trajectories in the first years are almost diagonal, from southwest to northeast. Later, the trajectories are more concentrated in the center and then again extend to the southwest and northeast. At the end of the duration, more trajectories are shown on the coasts of the region.

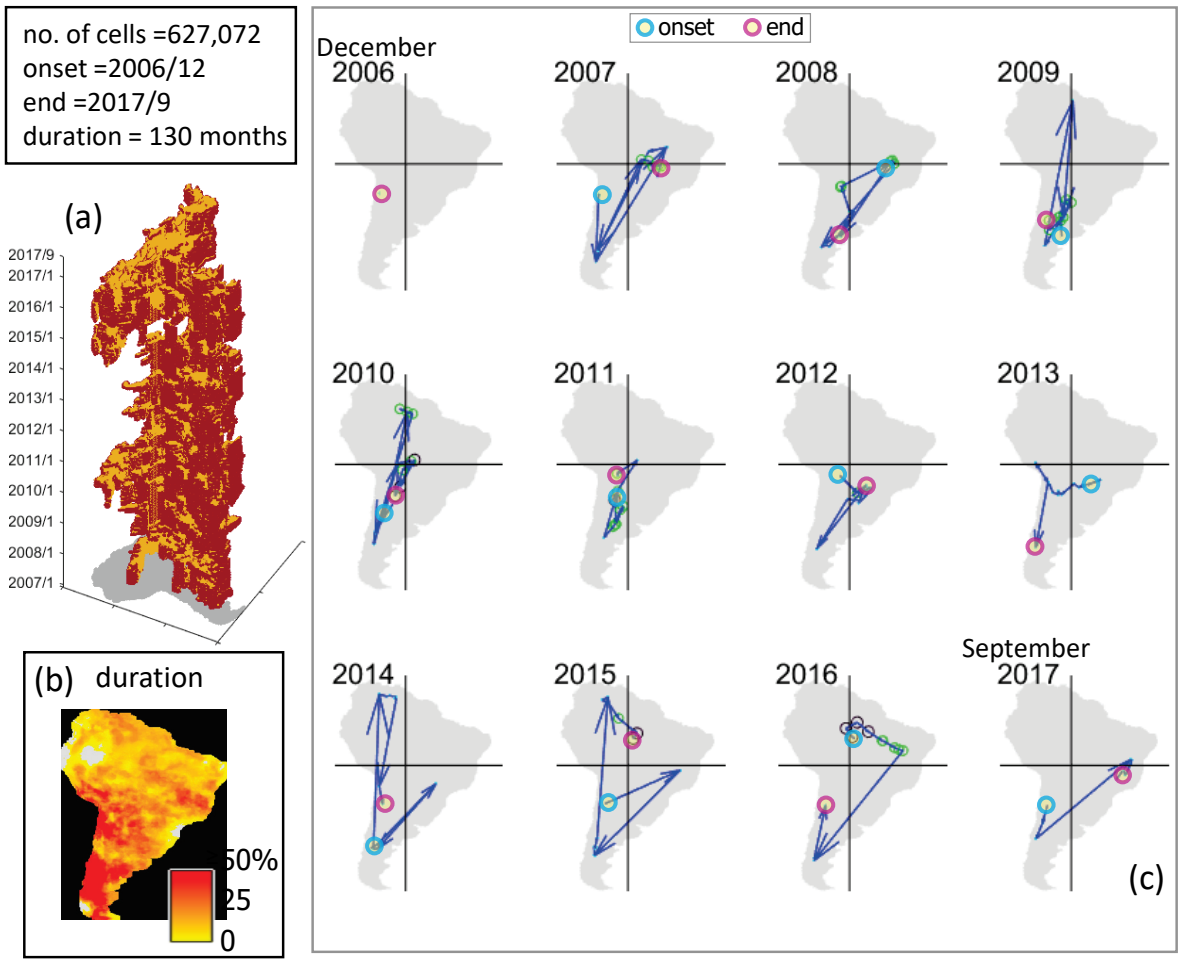


Figure 12. (a) Drought from December 2006 to September 2017. (b) Percentage of duration: each cell shows the percentage of time in drought with respect to the duration of the indicated drought. (c) Monthly drought trajectories per year; the onset and the end of the drought trajectory are indicated for each year by blue and red circles.

The third-longest drought is shown in Figure 13. The 3-D cluster spans almost the entire study area, although it was mostly concentrated in the northwest and south of the subcontinent (Figure 13a and 13b). The largest extent was observed in the period from 1988 to 1991 (Figure 13a). The trajectories of this drought show three main patterns (Figure 13c): they run from northeast to south in the first years, they then change from north to south, and finally they are located in the north.

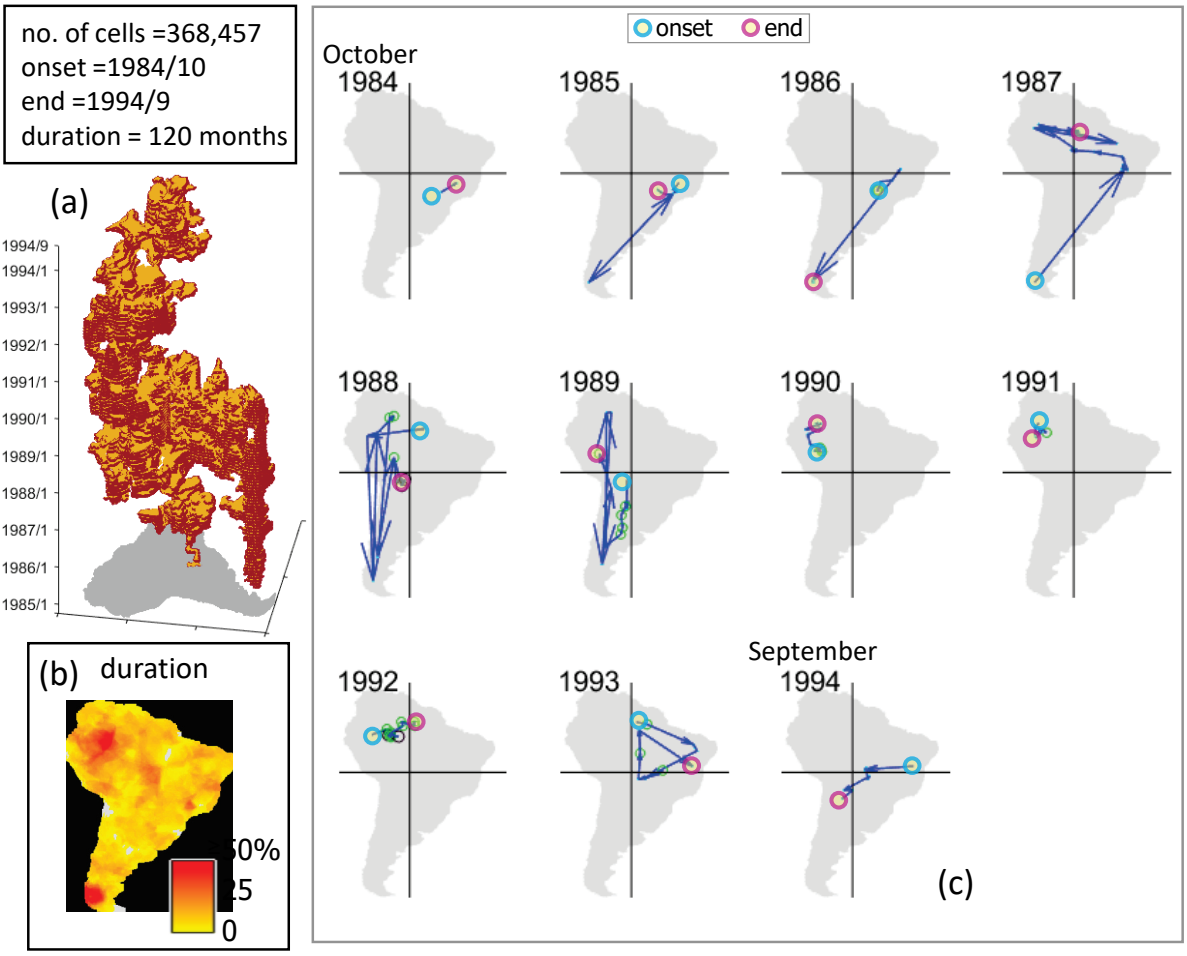


Figure 13. (a) Drought from October 1984 to September 1994. (b) Percentage of duration: each cell shows the percentage of time in drought with respect to the duration of the indicated drought. (c) Monthly drought trajectories per year; the onset and the end of the drought trajectory are indicated for each year by blue and red circles.

3.3 Comparison with the reported droughts

Figure 14a shows the droughts reported in the emergency events database (EMDAT) (Guha-Sapir, 2019) for some of the region's countries. The percentages of drought area calculated for each of the countries are shown in Figure 14b. The results show that, in general, the occurrence of the calculated droughts coincides with the information reported. The results indicate some

periods with important drought events in the 1960s, 1980s, 1990s, 2000s, and 2010s. In the most recent decades (the 2000s and 2010s), Brazil, Bolivia, Paraguay, Chile, and Argentina have experienced more droughts than in the rest of the period. The percentage of drought area shows that drought extent increases in the second semester of the year and concludes in the first semester of the following year, coinciding with the growing period of various crops of the region. These droughts can compromise the optimal soil moisture conditions necessary for the crops. According to EMDAT, Brazil is one of the countries that faced the most economic losses due to droughts in the period of analysis. The drought event presented in Figure 12 that lasted from 2006 to 2017 mainly encompassed Bolivia, Paraguay, Chile, Argentina, and part of Brazil, as shown by the percentages of drought area in Figure 14b.

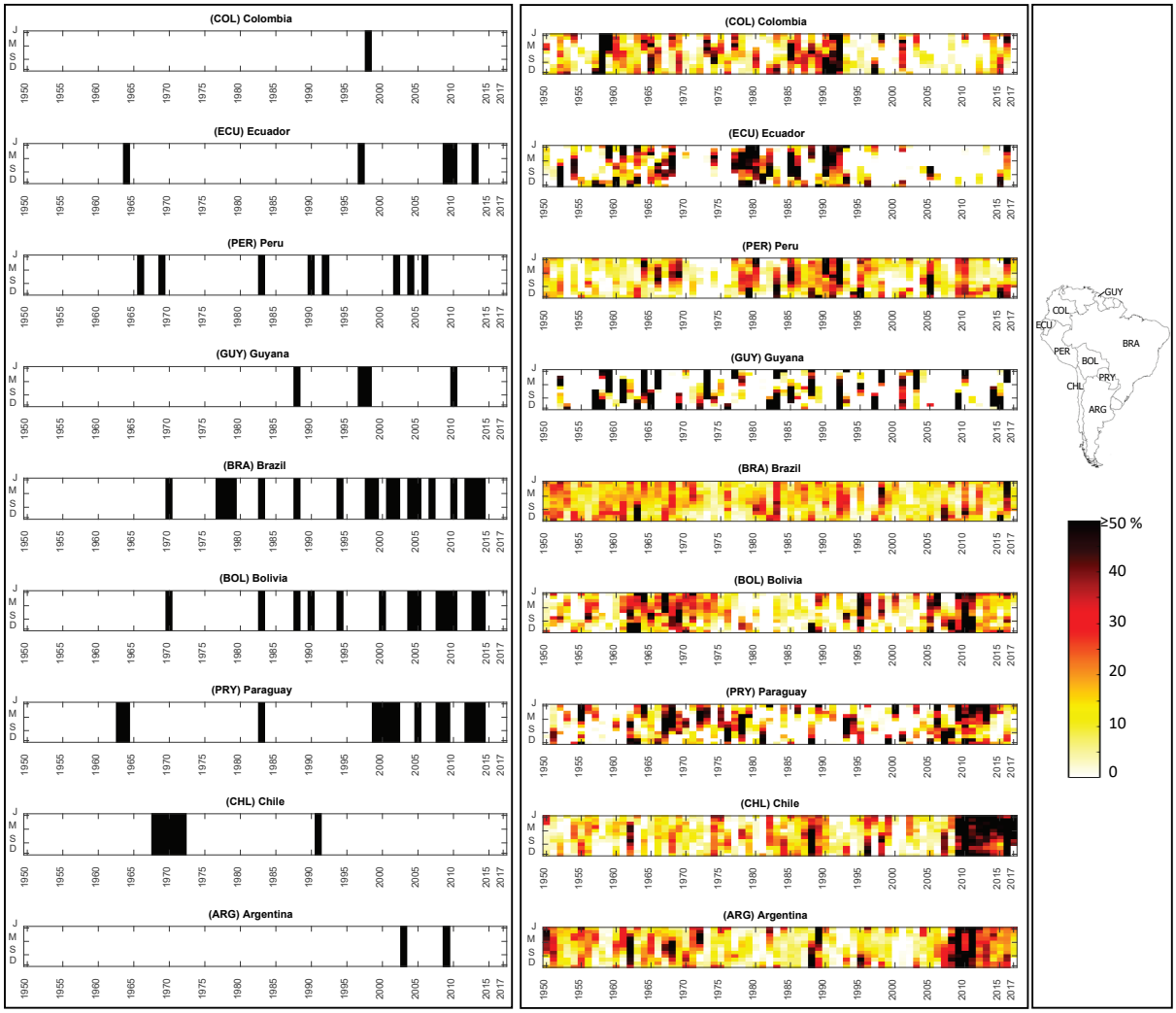


Figure 14. (a) Reported droughts in the EMDAT database. (b) Percentage of drought area calculated for each country.

4 Summary and conclusions

This research introduces a method of calculating the optimal cluster size filter for the 3-D characterization of drought. The combined effect of the cluster size filter and different drought indicator thresholds to calculate drought is also presented. The methodology was tested in South America with data from the Latin American Flood and Drought Monitor (LAFDM) for 1950–2017.

The following conclusions are drawn:

- The drought indicator thresholds of 0, −1, − .5, and −2 were tested. In general, the −1.5 and −2 thresholds produce more drought events with shorter durations and smaller magnitudes (sizes) than the threshold of −1. The threshold of 0 produces a long-lasting drought event and some small events. Based on the results, this threshold is not recommended as a method of calculating 3-D drought clusters.
- The optimal cluster size filter depends on the spatial resolution of the data and the threshold used.
- Durations and magnitude (number of cells) are mainly concentrated in less than 20 months and 10,000 cells. Few events lasting around 24 months or longer are observed. A linear relationship between these two characteristics is found ($R^2 = 0.89$).

The main findings for South America are as follows:

- Droughts of two to six months are observed practically throughout the entire study area, as well as 12-month droughts.
- Some regions show little occurrence of droughts, such as the Amazon basin, upper Magdalena River basin in Colombia, the north of Uruguay, the northwest of Argentina, and the south of Chile. One of the areas with the least occurrence is the lower Amazon basin, near the discharge.
- The 1957–1972, 2006–2017, and 1984–1994 droughts were the most extreme.
- In general, the occurrence of the calculated droughts coincides with the information reported. The 1960s, 1980s, 1990s, 2000s, and 2010s were the periods with more droughts.
- In the most recent decades (the 2000s and 2010s), Brazil, Bolivia, Paraguay, Chile, and Argentina have shown a greater occurrence of droughts than in the rest of the period.

Further research may include extension and testing of this methodology on other types of drought indicators. A more detailed study of the characteristics of droughts is also

recommended. The relationship between drought and the South American Low-Level Jet phenomenon (Montini et al., 2019), the moisture transporter from the Amazon to the subtropics, is another interesting topic that could be explored in further studies. The results of this study are important for the calculation and characterization of drought and better monitoring and the construction of future drought forecasting systems in the region.

5 Acknowledgments

VD thanks the Mexican National Council for Science and Technology (CONACYT) and Alianza FiiDEM for the study grand 217776/382365. HvL is supported by the H2020 ANYWHERE project (Grant Agreement No. 700099). DS acknowledges the grant No. 17-77-30006 of the Russian Science Foundation, and the Hydroinformatics research fund of IHE Delft in whose framework some research ideas and components were developed. The study is also a contribution to the UNESCO IHP-VII programme (Euro FRIEND-Water project) and the Panta Rhei Initiative on Drought in the Antropocene of the International Association of Hydrological Sciences (IAHS).

Appendix A

Table A1 Angle *C* [degrees] for each drought indicator (DI) threshold and cluster size filter. The lowest *C* angle is indicated with an asterisk. For the case of the threshold of −1, the result is indicated in bold.

Cluster size filter [number of cells]	DI threshold = 0	−1	−1.5	−2
0	-	-	-	-
4	177.2	179.7	179.6	179.9
9	177.6	178.0	179.1	178.7
16	168.0	178.8	178.3	178.6
25	145.1*	175.5	177.0	176.4
36	168.7	171.7	174.9	179.5
49	173.8	**168.1***	177.3	175.9
64	152.4	174.0	170.1*	172.8*
81	153.3	171.5	179.2	173.5
100	-	-	-	-

References

Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F., and Lettenmaier, D. P. (2005). Twentieth-Century Drought in the Conterminous United States. *Journal of Hydrometeorology*, *6*(6), 985–1001. https://doi.org/10.1175/JHM450.1

Corzo Perez, G. A., van Huijgevoort, M. H. J., Voß, F., and van Lanen, H. A. J. (2011). On the spatio-temporal analysis of hydrological droughts from global hydrological models. *Hydrology and Earth System Sciences*, *15*(9), 2963–2978. https://doi.org/10.5194/hess-15-2963-2011

Diaz, V., Corzo Perez, G. A., Van Lanen, H. A. J., and Solomatine, D. (2018). *Intelligent drought tracking for its use in Machine Learning: implementation and first results*. (G. La Loggia, G. Freni, V. Puleo, and M. De Marchis, Eds.), *HIC 2018. 13th International Conference on Hydroinformatics* (Vol. 3). Palermo: EasyChair. https://doi.org/10.29007/klgg

Diaz, V., Corzo Perez, G. A., Van Lanen, H. A. J., Solomatine, D., and Varouchakis, E. A. (2020a). An approach to characterise spatio-temporal drought dynamics. *Advances in Water Resources*, *137*, 103512. https://doi.org/10.1016/j.advwatres.2020.103512

Diaz, V., Corzo Perez, G. A., Van Lanen, H. A. J., Solomatine, D., and Varouchakis, E. A. (2020b). Characterisation of the dynamics of past droughts. *Science of The Total Environment*, *718*, 134588. https://doi.org/10.1016/j.scitotenv.2019.134588

Guha-Sapir, D. (2019). EM-DAT: The Emergency Events Database - Université catholique de Louvain (UCL) - CRED. Retrieved from www.emdat.be

Haralick, R. M., and Shapiro, L. G. (1992). *Computer and Robot Vision, Volume I*. Addison-Wesley.

Herrera-Estrada, J. E., and Diffenbaugh, N. S. (2020). Landfalling Droughts: Global Tracking of Moisture Deficits From the Oceans Onto Land. *Water Resources Research*, *56*(9). https://doi.org/10.1029/2019WR026877

Herrera-Estrada, J. E., Satoh, Y., and Sheffield, J. (2017). Spatio-Temporal Dynamics of Global Drought. *Geophysical Research Letters*, 2254–2263. https://doi.org/10.1002/2016GL071768

Lloyd-Hughes, B. (2012). A spatio-temporal structure-based approach to drought characterisation. *International Journal of Climatology*, *32*(3), 406–418. https://doi.org/10.1002/joc.2280

Mckee, T. B., Doesken, N. J., and Kleist, J. (1993). The relationship of drought frequency

418 and duration to time scales. *AMS 8th Conf. Appl. Climatol.*, *179*(January), 179–184.
419 https://doi.org/citeulike-article-id:10490403
420 Montini, T. L., Jones, C., and Carvalho, L. M. V. (2019). The South American Low-Level
421 Jet: A New Climatology, Variability, and Changes. *Journal of Geophysical Research:*
422 *Atmospheres*, *124*(3), 1200–1218. https://doi.org/10.1029/2018JD029634
423 Sheffield, J., Andreadis, K. M., Wood, E. F., and Lettenmaier, D. P. (2009). Global and
424 continental drought in the second half of the twentieth century: Severity-area-duration
425 analysis and temporal variability of large-scale events. *J. Clim.*, *22*(8), 1962–1981.
426 https://doi.org/10.1175/2008JCLI2722.1
427 Sheffield, J., Wood, E. F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Ogallo, L. (2014). A
428 Drought Monitoring and Forecasting System for Sub-Sahara African Water Resources
429 and Food Security. *Bulletin of the American Meteorological Society*, *95*(6), 861–882.
430 https://doi.org/10.1175/BAMS-D-12-00124.1
431 van Huijgevoort, M. H. J., Hazenberg, P., van Lanen, H. A. J., Teuling, A. J., Clark, D. B.,
432 Folwell, S., … Uijlenhoet, R. (2013). Global Multimodel Analysis of Drought in Runoff
433 for the Second Half of the Twentieth Century. *Journal of Hydrometeorology*, *14*(5),
434 1535–1552. https://doi.org/10.1175/JHM-D-12-0186.1
435 Vernieuwe, H., De Baets, B., and Verhoest, N. E. C. (2020). A mathematical morphology
436 approach for a qualitative exploration of drought events in space and time. *International*
437 *Journal of Climatology*, *40*(1), 530–543. https://doi.org/10.1002/joc.6226
438 Wood, E. F., Fisher, C., Sheffield, J., and Chaney, N. (2016). *Overview of the Latin American*
439 *and Caribbean Flood and Drought Monitor (LACFDM)*. Foz de Iguazú, Brazil.
440 Retrieved from
441 https://www.cazalac.org/mwar_lac/fileadmin/imagenes2/Remote_Sensing/Princeton_Da
442 y1_Overview_and_Station_Merging_light.pdf
443 World Meteorological Organization (WMO). (2006). *Drought monitoring and early warning:*
444 *concepts, progress and future challenges*. *WMO-No. 1006*. Geneva, Switzerland.
445 Retrieved from
446 http://www.droughtmanagement.info/literature/WMO_drought_monitoring_early_warni
447 ng_2006.pdf
448 World Meteorological Organization (WMO). (2012). *Standardized Precipitation Index user*
449 *guide*. *WMO-No. 1090*. Geneva, Switzerland. Retrieved from
450 https://library.wmo.int/doc_num.php?explnum_id=7768
451 Yevjevich, V. (1967). *An objective approach to definitions and investigations of continental*

hydrologic droughts. Hydrology Paper 23 (Vol. 23). Fort Collins, Colorado. Retrieved from https://dspace.library.colostate.edu/bitstream/handle/10217/61303/HydrologyPapers_n23.pdf