

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

Vitali Diaz<sup>1,1</sup>, Gerald Augusto Corzo Perez<sup>2,2</sup>, Henny Van Lanen<sup>3,3</sup>, and Dimitri Solomatine<sup>1,1</sup>

<sup>1</sup>IHE Delft Institute for Water Education

<sup>2</sup>UNESCO-IHE Institute for Water Education

<sup>3</sup>Wageningen University

November 30, 2022

## 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.

# 1 **Three-dimensional clustering in the characterization of spatiotemporal** 2 **drought dynamics: cluster size filter and drought indicator threshold** 3 **optimization**

4 Vitali Diaz<sup>1,2</sup>, Gerald A. Corzo Perez<sup>1</sup>, Henny A.J. Van Lanen<sup>3</sup>, Dimitri Solomatine<sup>1,2,4</sup>

5 <sup>1</sup>IHE Delft Institute for Water Education, Delft, 2601 DA, the Netherlands

6 <sup>2</sup>Delft University of Technology, Delft, the Netherlands

7 <sup>3</sup>Hydrology and Quantitative Water Management Group, Wageningen University,  
8 Wageningen, the Netherlands

9 <sup>4</sup>Water Problems Institute of the Russian Academy of Sciences, Moscow, Russia

10

## 11 **Abstract**

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

## 27 **Keywords**

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

## 30 **1 Introduction**

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

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

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

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

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

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

## 72 **2 Methods and data**

### 73 **2.1 Drought calculation**

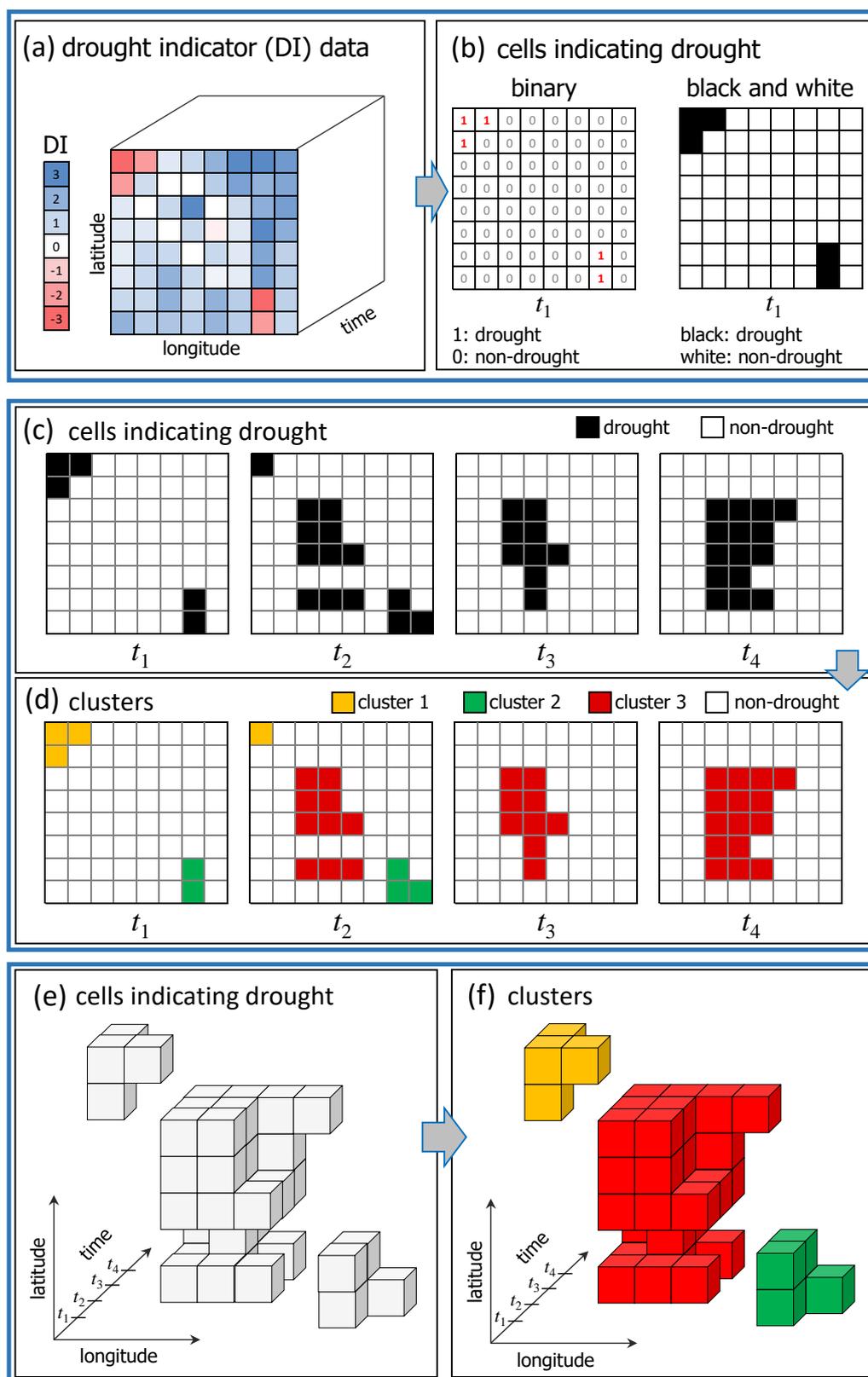
74 In this methodology, drought is represented by a 3-D cluster. The data must be organized in a  
75 grid system where each cell is geographically referenced (latitude and longitude). The  
76 arrangement of these geospatial data cells can be schematized as a cube with latitude, longitude,  
77 and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

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

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

$$96 \quad D_s(t) = \begin{cases} 1 & \text{if } DI(t) \leq T \\ 0 & \text{if } DI(t) > T \end{cases} \quad (\text{Eq. 1})$$

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



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).

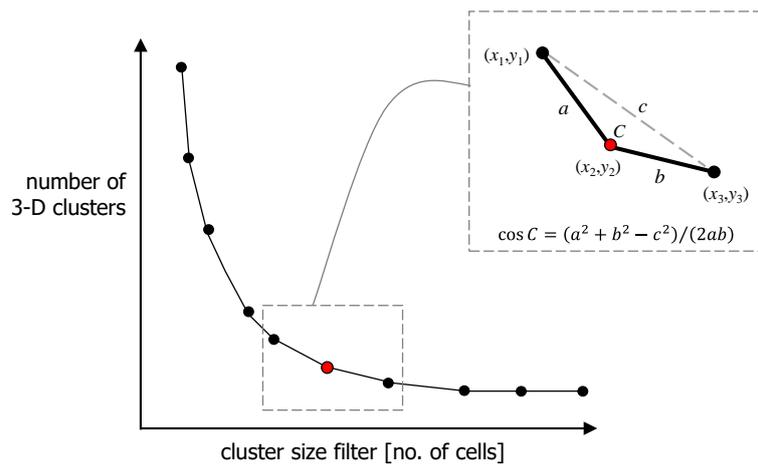
106 The third step is the calculation of the 3-D clusters. A clustering method to extract the drought  
107 events is applied following [Corzo Perez et al. \(2011\)](#). The drought events correspond to the 3-  
108 D clusters calculated with the data of 0s and 1s (Figure 1d and 1f). The following describes the  
109 unsupervised machine learning-based method to extract the clusters. The method follows the  
110 connected-component labeling approach to cluster the cells in drought ([Haralick and Shapiro,  
111 1992](#)). In this method, a two-scan algorithm is applied. First, each cell is numbered. Then, the  
112 first run is carried out, in which the binary grid is explored and provisional labels are assigned  
113 to connected (contiguous) components (cells). These labels identify the connection of every  
114 cell with its nearest neighbors. In this first run, the cell's label does not yet refer to the cluster  
115 number but to the cells with which the given cell is connected. Finally, the second scan is  
116 carried out to find similar cell connections, i.e., clusters, which are given a unique label.  
117 Examination of the grid can be performed by columns or by rows. The clustering method is  
118 conducted over the whole binary data (Figure 1d and 1f).

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

134 For the identification of 3-D drought clusters, we also considered a filter of duration. All the  
135 3-D clusters of one-month duration were excluded from further analysis. We made sure that  
136 none of these clusters was of a large number of cells, thus avoiding the risk of removing severe  
137 events. The 3-D clusters of one-month duration were removed for each case of cluster size filter  
138 and drought indicator threshold.

139 Drought calculation concluded with the selection of the cluster size filter (Figure 2). The  
 140 optimal cluster size was defined as the value in the curve ‘cluster size filter vs. the number of  
 141 3-D clusters,’ at which the number of clusters stabilizes, i.e., does not undergo considerable  
 142 changes for a constant increase of the cluster size filter (Figure 2). This point is defined as the  
 143 vertex of the curve (Figure 2). To identify this point, we developed a method based on the angle  
 144  $C$  formed between two continuous segments of the curve ‘cluster size filter vs. number of 3-D  
 145 clusters’ (Figure 2). We calculated this angle  $C$  using the law of cosines (Eq. 2), which  
 146 considers the length of the sides of the triangle  $a$ ,  $b$ , and  $c$  formed by the coordinates of three  
 147 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  
 148 points, the angle  $C$  was calculated. The vertex corresponds to the smallest angle of all the  
 149 calculated  $C$ s. The optimal cluster size filter was calculated for each drought indicator  
 150 threshold.

151 
$$\cos C = \frac{(a^2 + b^2 + c^2)}{(2ab)} \quad (\text{Eq. 2})$$



152  
 153 **Figure 2.** Scheme of the method to calculate the optimal cluster size filter. The angle  $C$  is calculated  
 154 with the sides  $a$ ,  $b$ , and  $c$  of the triangle formed by the three subsequent points (zoomed-in view).

155 **2.2 Drought characterization**

156 After identifying the droughts, the onset and end in time, duration, and severity (magnitude)  
 157 were calculated for each drought (3-D cluster). Drought duration ( $dd$ ) and magnitude ( $ds$ ) were  
 158 obtained with Eqs. 3 and 4, respectively. The times  $t_i$  and  $t_f$  are the onset and end of each  
 159 drought, respectively.  $DA$  is the drought area (number of cells) at each time step  $t$ . Eqs. 3 and  
 160 4 were applied for each 3-D cluster.

161 
$$dd = t_i - t_f + 1 \quad (\text{Eq. 3})$$

162 
$$ds = \sum_{t=t_i}^{t_f} DA(t) \quad (\text{Eq. 4})$$

163 For each drought, the spatial trajectory was also calculated. These trajectories were built with  
164 the union of the centroids of the drought areas at each time step ( $t$ ), following Diaz et al.  
165 (2020a). Trajectories of the largest events were analyzed.

## 166 **2.3 Data**

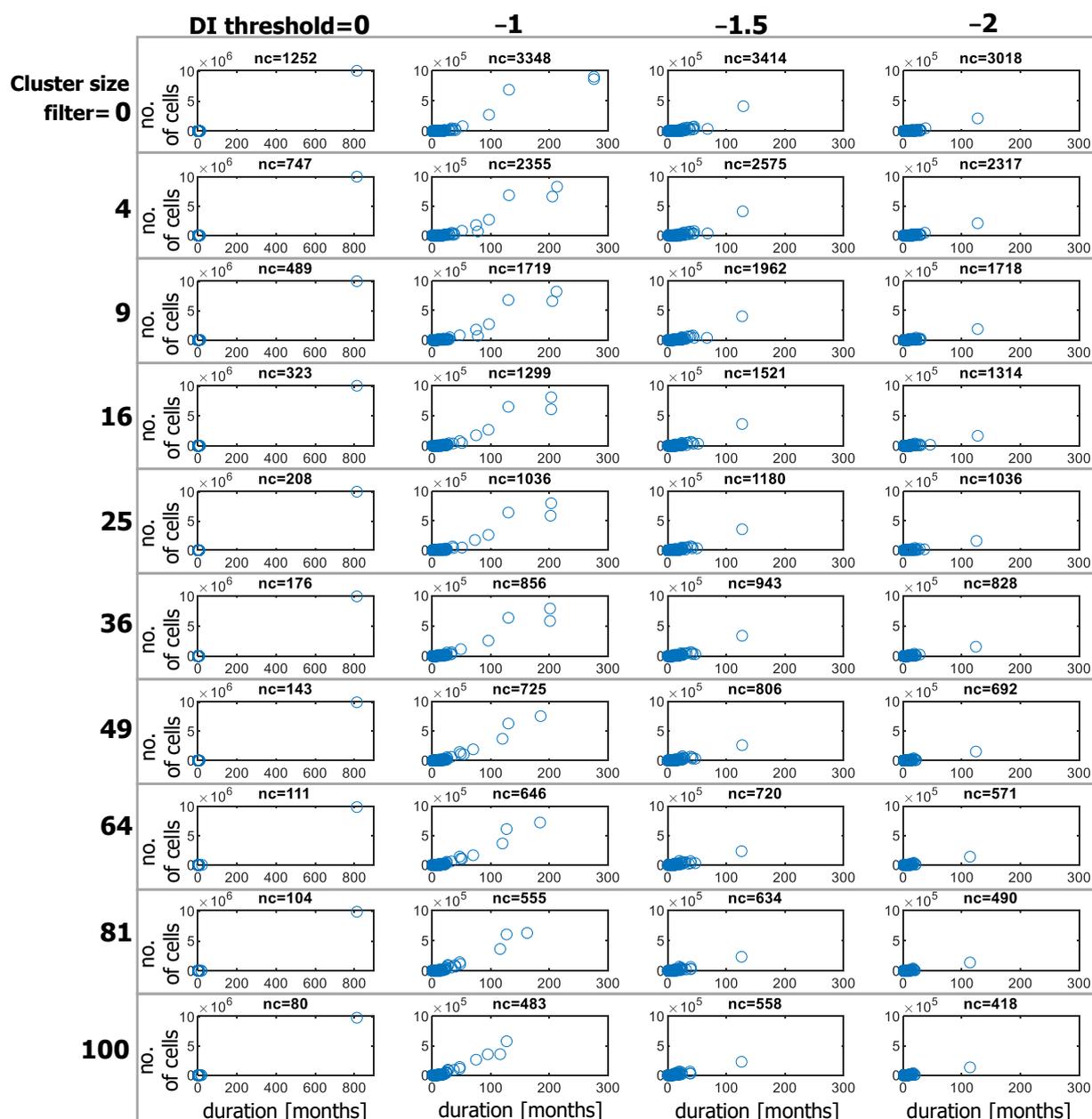
167 For the identification of droughts, we used data of the Standardized Precipitation Index (SPI)  
168 retrieved from the Latin American Flood and Drought Monitor (LAFDM)  
169 ([https://platform.princetonclimate.com/PCA\\_Platform/lafdmLanding.html](https://platform.princetonclimate.com/PCA_Platform/lafdmLanding.html)).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

## 178 **3 Results and discussion**

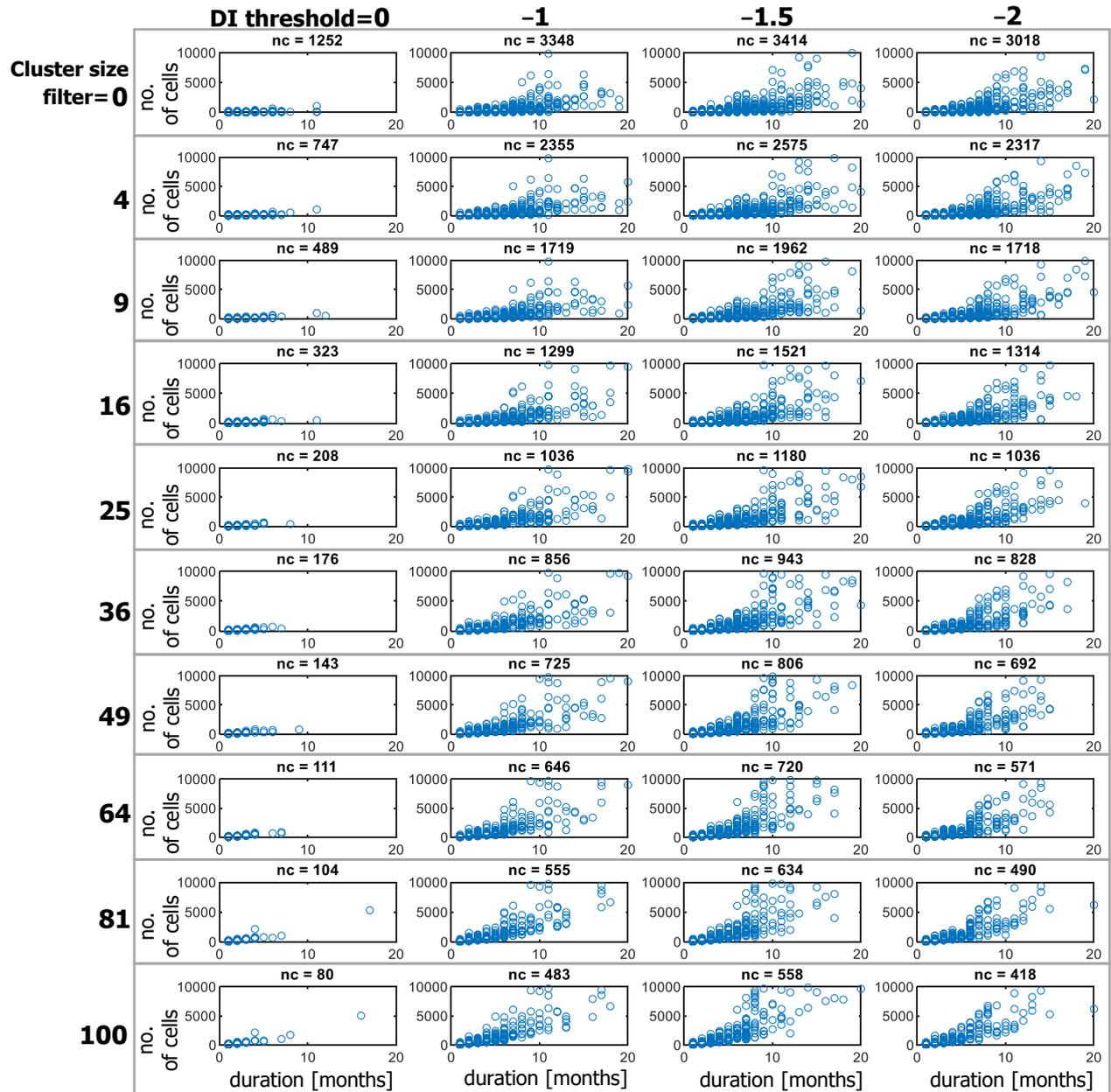
### 179 **3.1 Drought calculation**

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

194 Figures 3 and 4 display the clusters of one-month duration. Although these clusters have a  
 195 number of cells greater than the cluster size filter in each case, they are small compared to the  
 196 rest of the clusters in each case (Figures 3 and 4); for this reason, they were also removed.  
 197 Figures 5 and 6 show the final results after removing the one-month duration clusters. Figure  
 198 5 shows the droughts for all durations and magnitudes. In Figure 5, scarce events with durations  
 199 greater than 100 months and a considerable number of cells are observed; most droughts have  
 200 durations of less than 20 months. Figure 6 shows the results for droughts of up to 20 months  
 201 and 10,000 cells. A direct relationship is observed between duration and magnitude.



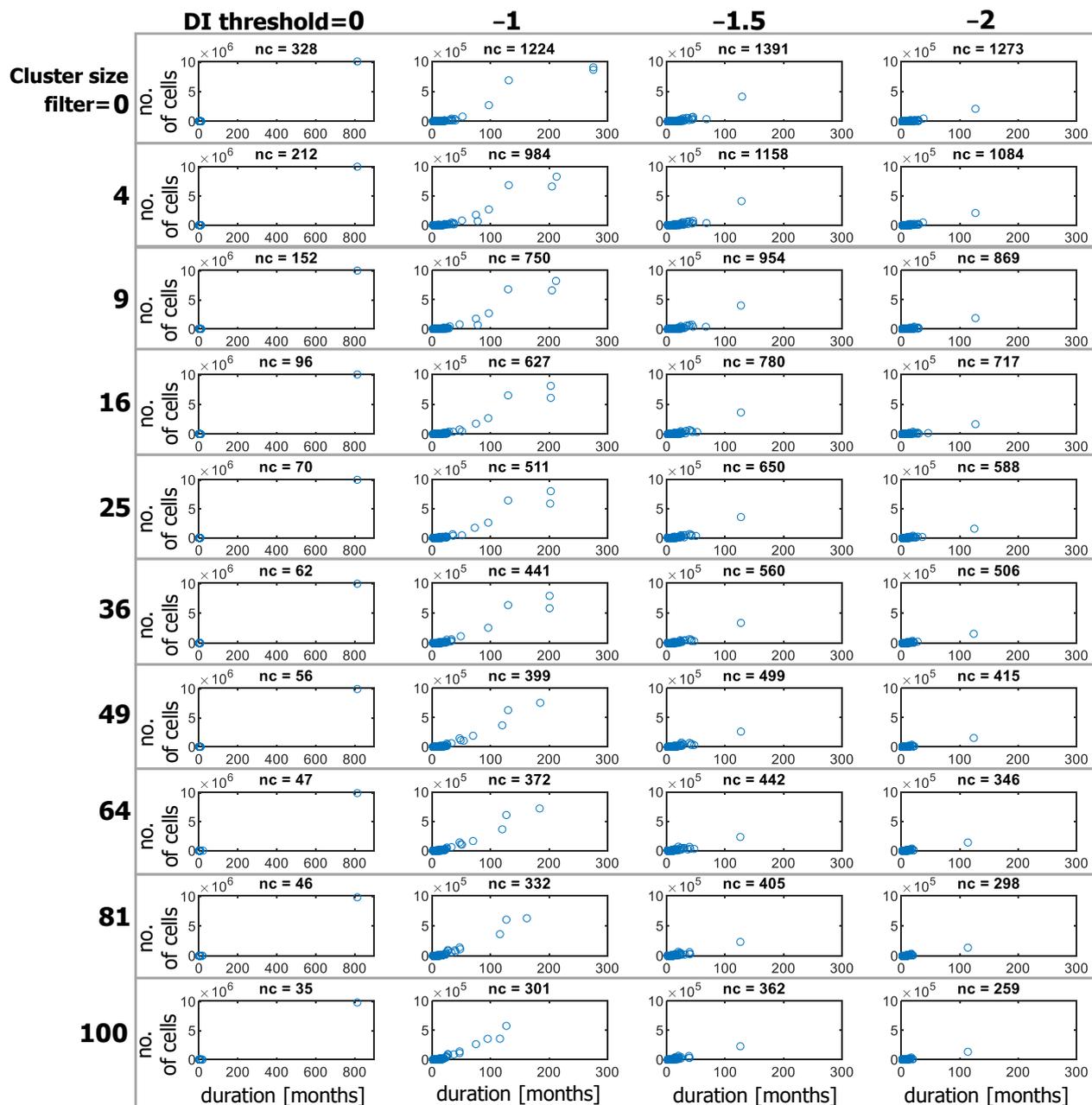
202  
 203 **Figure 3.** Number of 3-D clusters (nc) calculated for different drought indicator (DI) thresholds and  
 204 cluster size filters. The duration (months) and magnitude (number of cells) are indicated. Note: The  
 205 number of cells for the first threshold is up to  $10 \times 10^6$ , while for the rest it is up to  $10 \times 10^5$ .



206

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

208



209

210

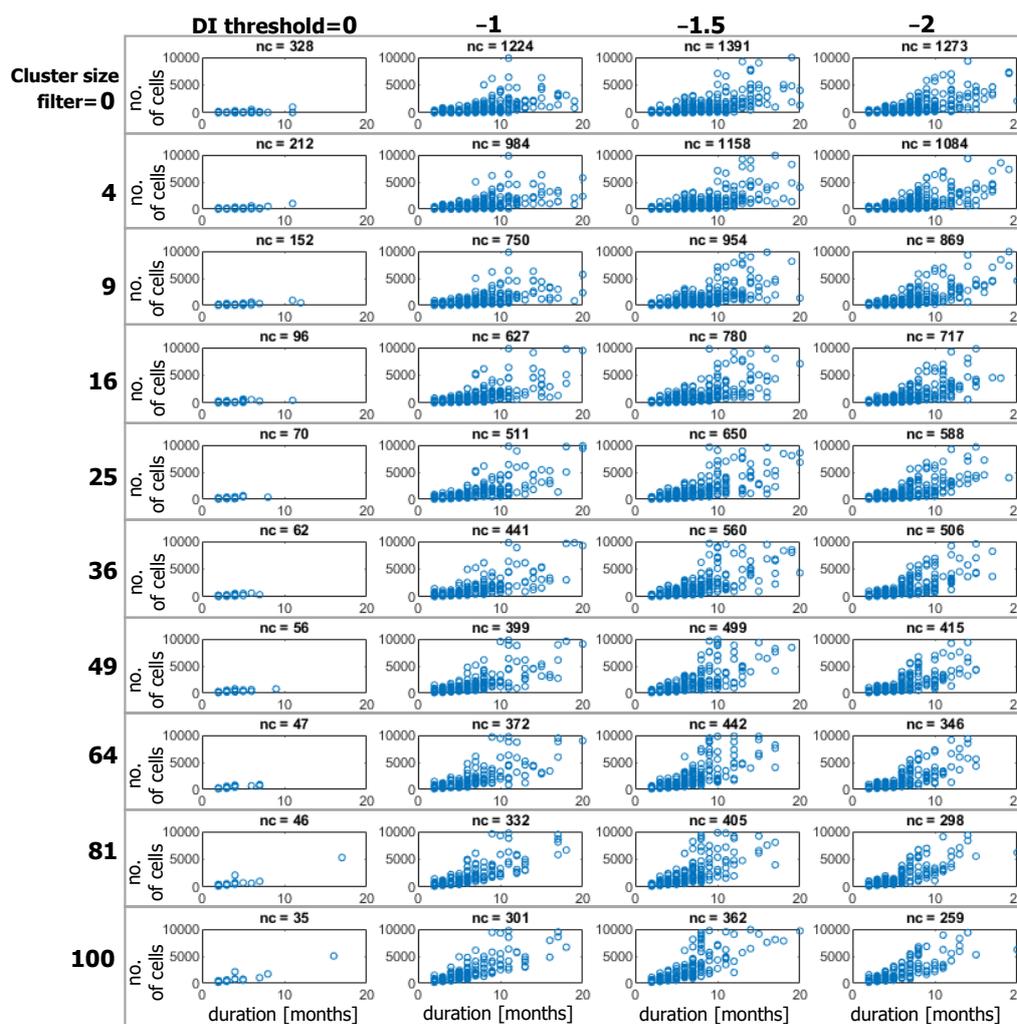
211

212

213

214

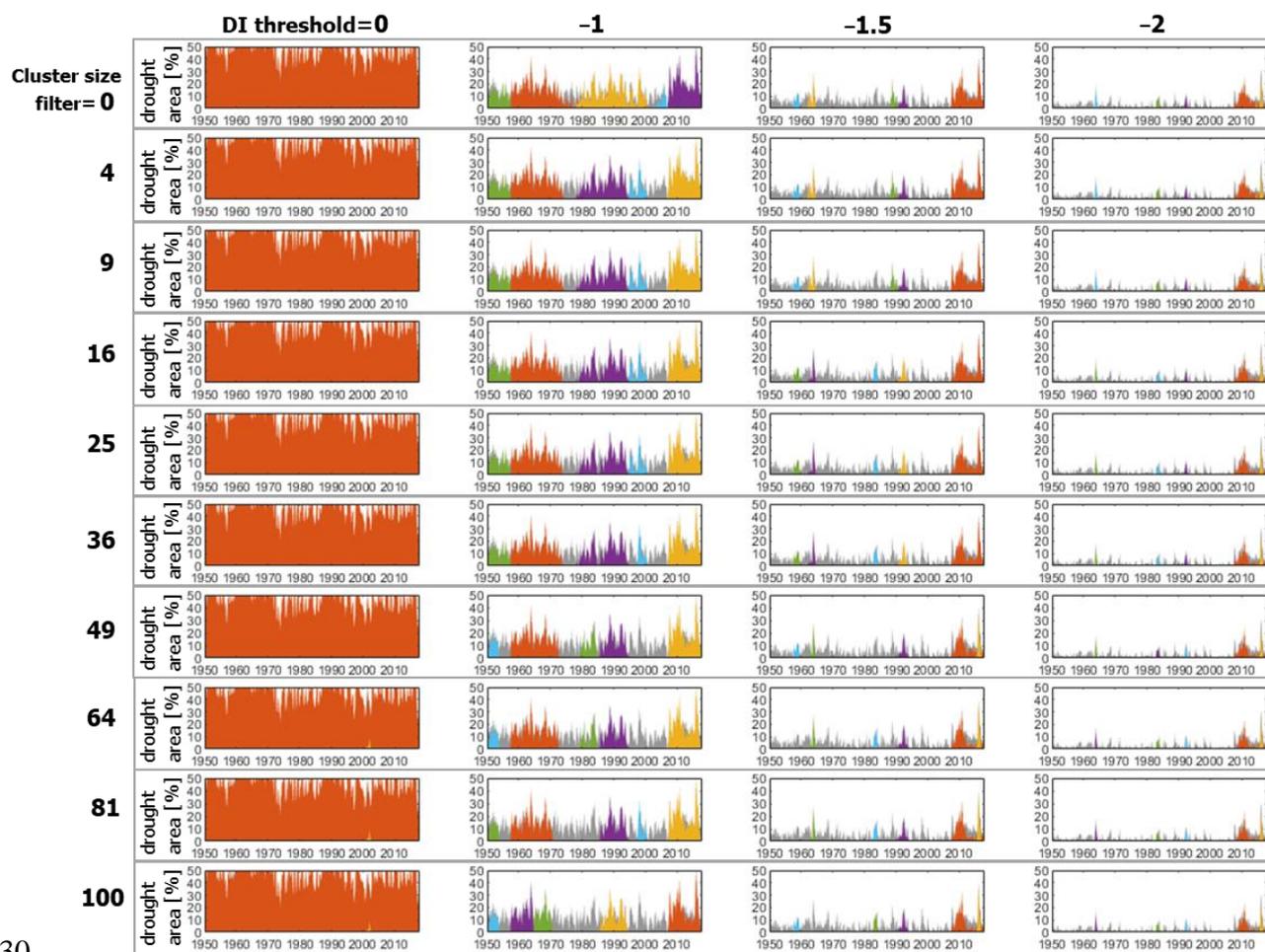
**Figure 5.** Number of 3-D clusters (nc) calculated for different drought indicator (DI) thresholds and cluster size filters. The duration (months) and magnitude (number of cells) are indicated. Note: The number of cells for the first threshold is up to  $10 \times 10^6$ , while for the rest it is up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



215

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

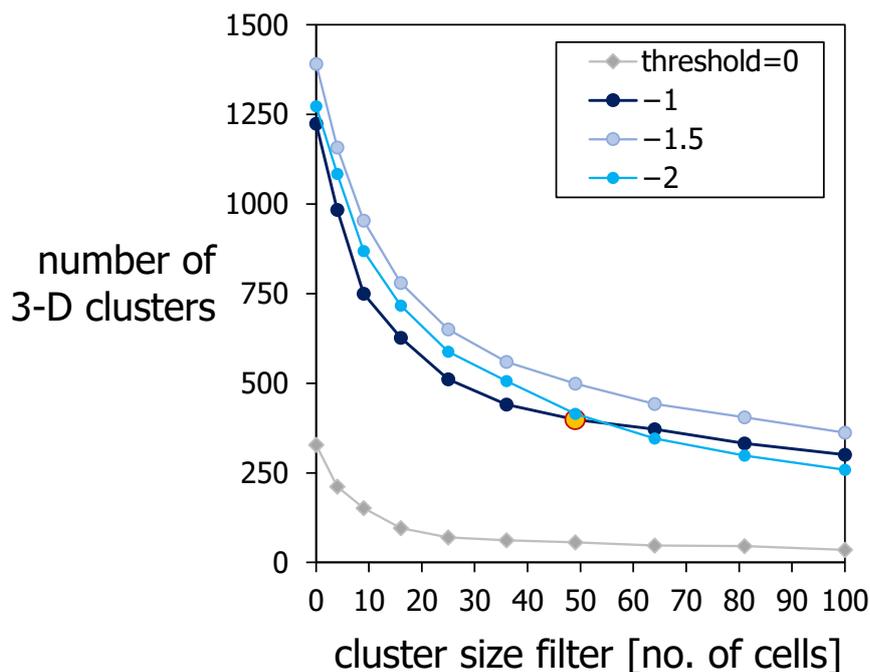
217 Figure 7 shows the percentages of drought area (PDAs) without considering the one-month  
 218 duration clusters. PDAs were calculated for each 3-D cluster as the number of cells of each  
 219 drought area at each time step divided by the total number of cells of the region (24,877). This  
 220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



230

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

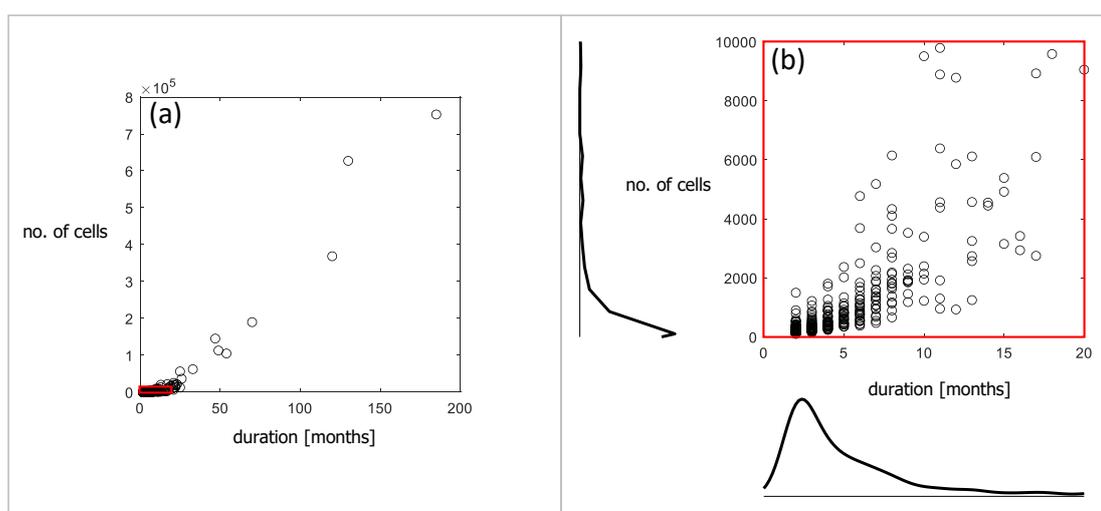
234 The number of 3-D clusters for each drought indicator threshold and cluster size filter is  
 235 presented in Figure 8. The method developed to find the optimal cluster size filter (Sect. 2.1)  
 236 shows, for instance, that, for the threshold of  $-1$ , the cluster size filter is 49 cells. Table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



243  
 244 **Figure 8.** Identification of the optimal cleaning filter size. The result for the threshold of -1 is  
 245 highlighted.

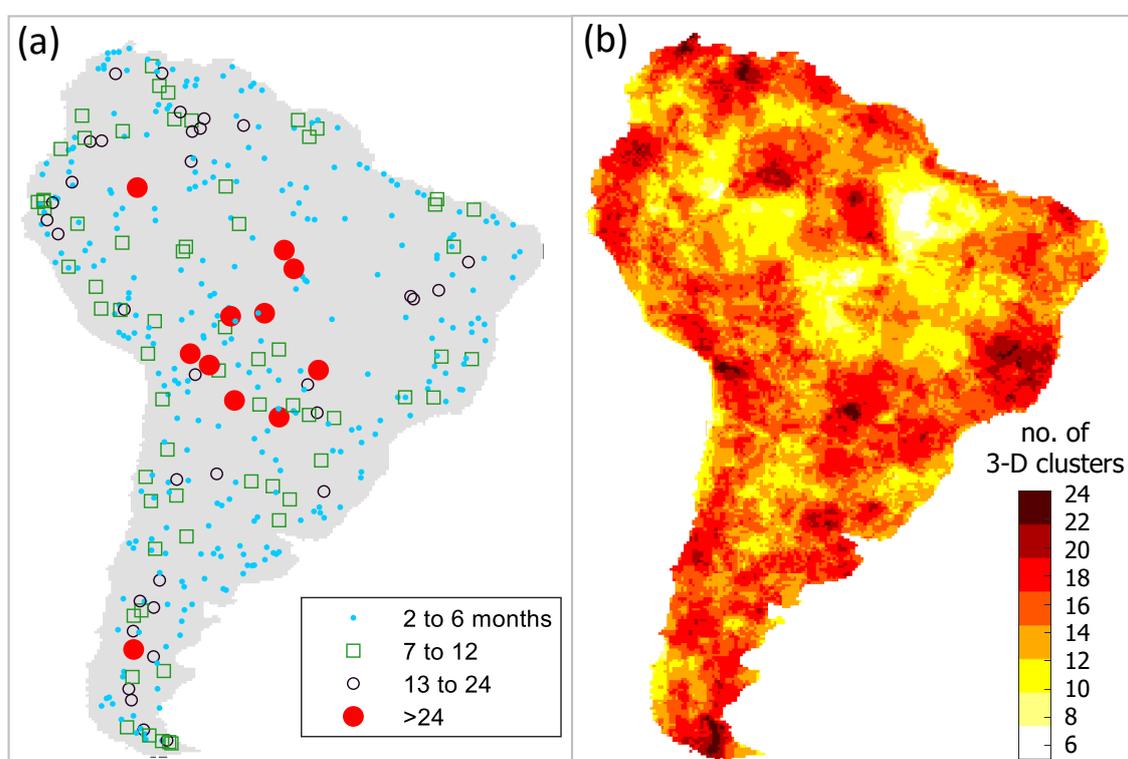
### 246 3.2 Drought characterization

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



252  
 253 **Figure 9.** (a) Duration (months) and magnitude (number of cells) of the droughts for the period 1950–  
 254 2017. (b) Detail for drought durations up to 20 months and 10,000 cells.

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

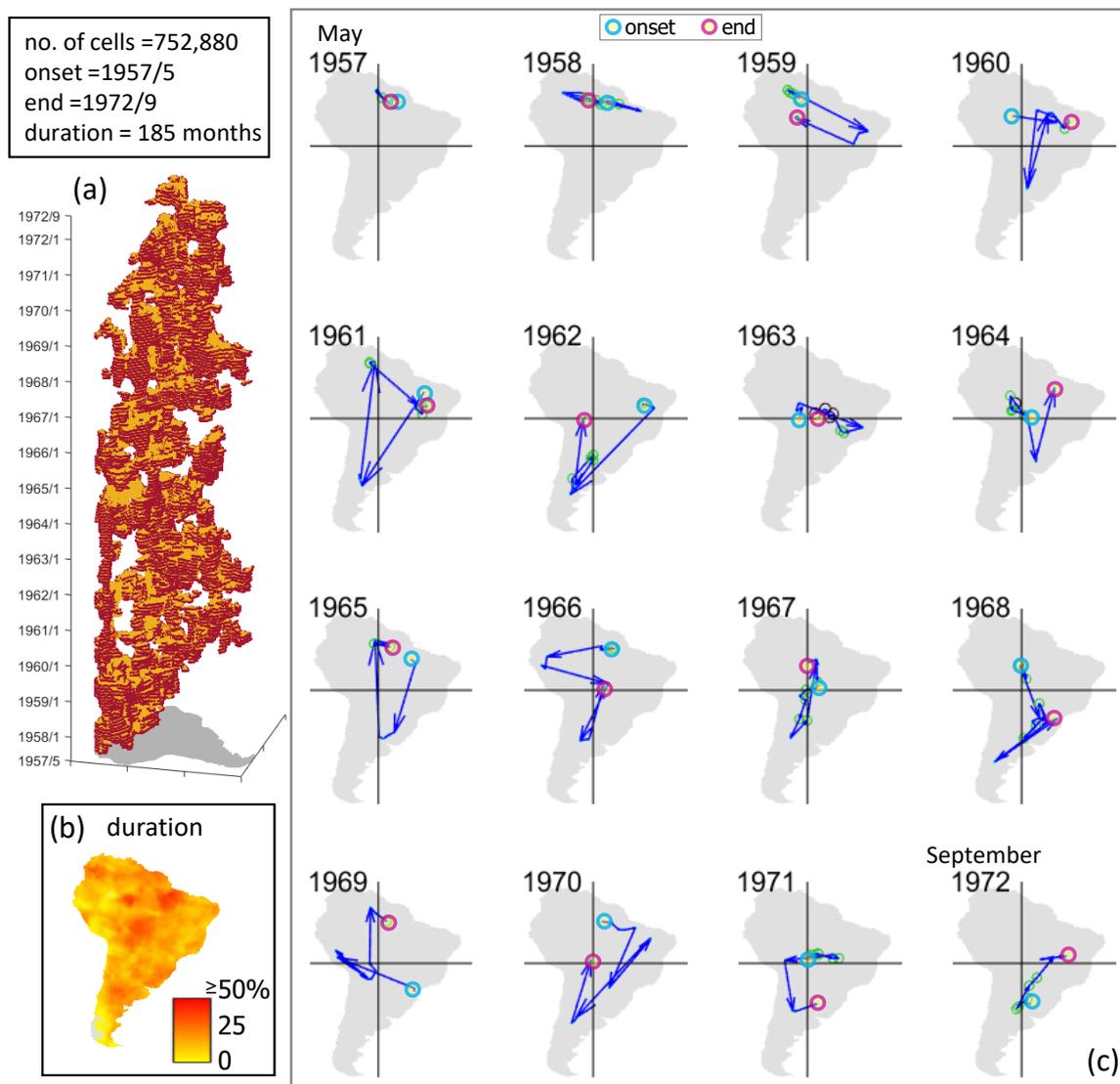


266  
267 **Figure 10.** (a) Centroids of the 3-D clusters for the period 1950–2017. Centroids are classified by cluster  
268 duration [months]. (b) Occurrence of the 3-D clusters: each cell indicates the number of 3-D clusters  
269 that took place there in the period 1950–2017.

270 The following paragraphs present the results of three of the most extreme droughts: those of  
271 1957–1972, 2006–2017, and 1984–1994.

272 The drought with the longest duration is shown in Figure 11. The drought lasted from May  
273 1957 to September 1972 (185 months). The 3-D cluster is shown in Figure 11a. It is observed

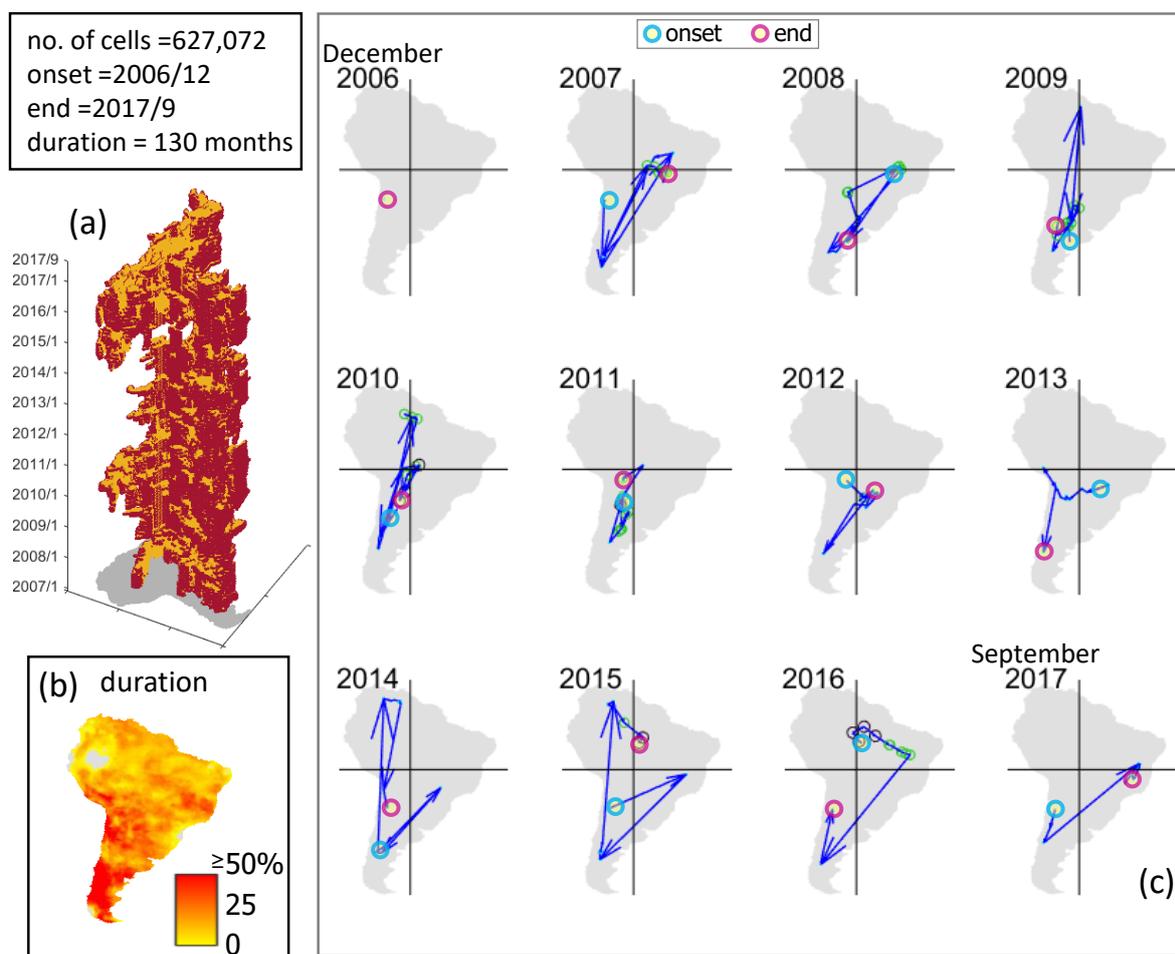
274 that this cluster completely covered the region (Figure 11a and 11b). However, its spatial  
 275 distribution over time varied significantly, concentrating mainly in the north and south of the  
 276 subcontinent. Figure 11b shows the distribution of the duration of each cell in the study area.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



283

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

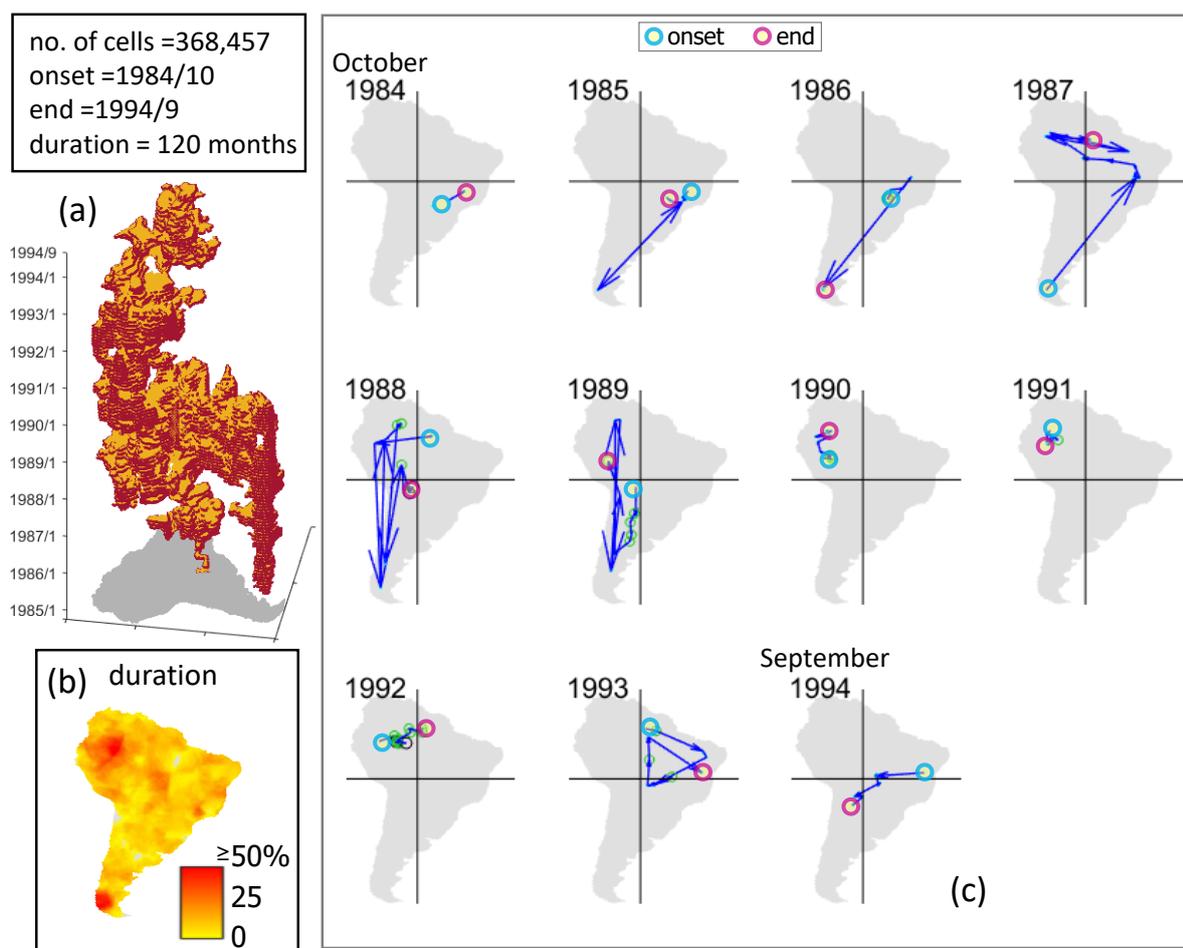
288 The second-longest drought is shown in Figure 12. This drought lasted from December 2006  
 289 to September 2017 (130 months). The 3-D cluster is shown in Figure 12a. The event was mainly  
 290 concentrated in the southwestern and southern coast of South America and some areas of Brazil  
 291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



296  
 297 **Figure 12.** (a) Drought from December 2006 to September 2017. (b) Percentage of duration: each cell  
 298 shows the percentage of time in drought with respect to the duration of the indicated drought. (c)  
 299 Monthly drought trajectories per year; the onset and the end of the drought trajectory are indicated for  
 300 each year by blue and red circles.

301  
 302

303 The third-longest drought is shown in Figure 13. The 3-D cluster spans almost the entire study  
 304 area, although it was mostly concentrated in the northwest and south of the subcontinent  
 305 (Figure 13a and 13b). The largest extent was observed in the period from 1988 to 1991 (Figure  
 306 13a). The trajectories of this drought show three main patterns (Figure 13c): they run from  
 307 northeast to south in the first years, they then change from north to south, and finally they are  
 308 located in the north.

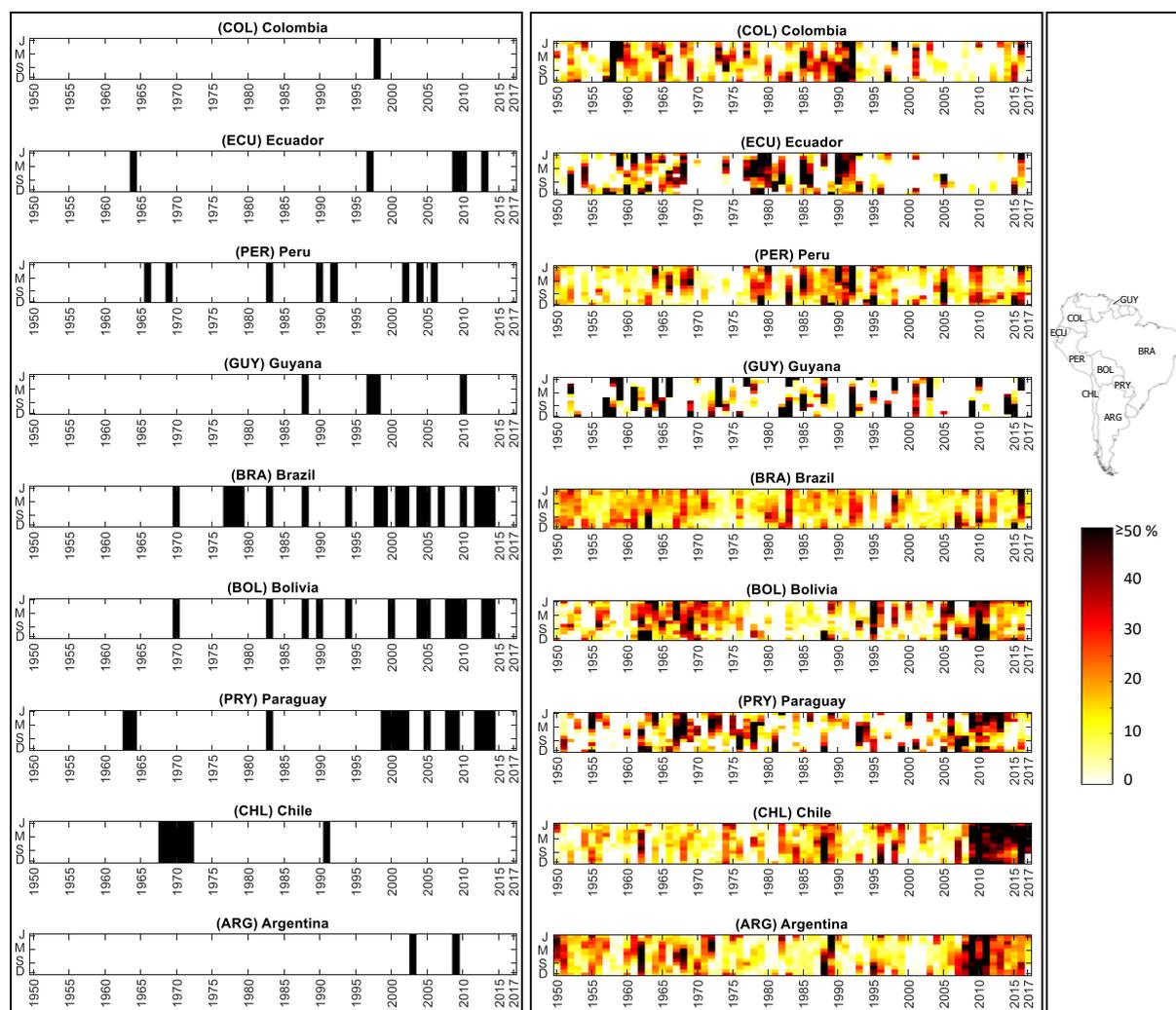


309 **Figure 13.** (a) Drought from October 1984 to September 1994. (b) Percentage of duration: each cell  
 310 shows the percentage of time in drought with respect to the duration of the indicated drought. (c)  
 311 Monthly drought trajectories per year; the onset and the end of the drought trajectory are indicated for  
 312 each year by blue and red circles.  
 313

### 314 3.3 Comparison with the reported droughts

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

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



329 **Figure 14.** (a) Reported droughts in the EMDAT database. (b) Percentage of drought area calculated  
 330 for each country.  
 331

## 332 **4 Summary and conclusions**

333 This research introduces a method of calculating the optimal cluster size filter for the 3-D  
334 characterization of drought. The combined effect of the cluster size filter and different drought  
335 indicator thresholds to calculate drought is also presented. The methodology was tested in  
336 South America with data from the Latin American Flood and Drought Monitor (LAFDM) for  
337 1950–2017.

338 The following conclusions are drawn:

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

349 The main findings for South America are as follows:

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

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

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

### 369 **5 Acknowledgments**

370 VD thanks the Mexican National Council for Science and Technology (CONACYT) and  
 371 Alianza FiiDEM for the study grand 217776/382365. HvL is supported by the H2020  
 372 ANYWHERE project (Grant Agreement No. 700099). DS acknowledges the grant No. 17-77-  
 373 30006 of the Russian Science Foundation, and the Hydroinformatics research fund of IHE Delft  
 374 in whose framework some research ideas and components were developed. The study is also a  
 375 contribution to the UNESCO IHP-VII programme (Euro FRIEND-Water project) and the Panta  
 376 Rhei Initiative on Drought in the Antropocene of the International Association of Hydrological  
 377 Sciences (IAHS).

### 378 **Appendix A**

379 **Table A1** Angle  $C$  [degrees] for each drought indicator (DI) threshold and cluster size filter. The lowest  
 380  $C$  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	<b>168.1*</b>	177.3	175.9
64	152.4	174.0	170.1*	172.8*
81	153.3	171.5	179.2	173.5
100	-	-	-	-

381

382

383

384

385 **References**

- 386 Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F., and Lettenmaier, D. P. (2005).  
387 Twentieth-Century Drought in the Conterminous United States. *Journal of*  
388 *Hydrometeorology*, 6(6), 985–1001. <https://doi.org/10.1175/JHM450.1>
- 389 Corzo Perez, G. A., van Huijgevoort, M. H. J., Voß, F., and van Lanen, H. A. J. (2011). On  
390 the spatio-temporal analysis of hydrological droughts from global hydrological models.  
391 *Hydrology and Earth System Sciences*, 15(9), 2963–2978. [https://doi.org/10.5194/hess-](https://doi.org/10.5194/hess-15-2963-2011)  
392 15-2963-2011
- 393 Diaz, V., Corzo Perez, G. A., Van Lanen, H. A. J., and Solomatine, D. (2018). *Intelligent*  
394 *drought tracking for its use in Machine Learning: implementation and first results*. (G.  
395 La Loggia, G. Freni, V. Puleo, and M. De Marchis, Eds.), *HIC 2018. 13th International*  
396 *Conference on Hydroinformatics* (Vol. 3). Palermo: EasyChair.  
397 <https://doi.org/10.29007/klgg>
- 398 Diaz, V., Corzo Perez, G. A., Van Lanen, H. A. J., Solomatine, D., and Varouchakis, E. A.  
399 (2020a). An approach to characterise spatio-temporal drought dynamics. *Advances in*  
400 *Water Resources*, 137, 103512. <https://doi.org/10.1016/j.advwatres.2020.103512>
- 401 Diaz, V., Corzo Perez, G. A., Van Lanen, H. A. J., Solomatine, D., and Varouchakis, E. A.  
402 (2020b). Characterisation of the dynamics of past droughts. *Science of The Total*  
403 *Environment*, 718, 134588. <https://doi.org/10.1016/j.scitotenv.2019.134588>
- 404 Guha-Sapir, D. (2019). EM-DAT: The Emergency Events Database - Université catholique  
405 de Louvain (UCL) - CRED. Retrieved from [www.emdat.be](http://www.emdat.be)
- 406 Haralick, R. M., and Shapiro, L. G. (1992). *Computer and Robot Vision, Volume I*. Addison-  
407 Wesley.
- 408 Herrera-Estrada, J. E., and Diffenbaugh, N. S. (2020). Landfalling Droughts: Global Tracking  
409 of Moisture Deficits From the Oceans Onto Land. *Water Resources Research*, 56(9).  
410 <https://doi.org/10.1029/2019WR026877>
- 411 Herrera-Estrada, J. E., Satoh, Y., and Sheffield, J. (2017). Spatio-Temporal Dynamics of  
412 Global Drought. *Geophysical Research Letters*, 2254–2263.  
413 <https://doi.org/10.1002/2016GL071768>
- 414 Lloyd-Hughes, B. (2012). A spatio-temporal structure-based approach to drought  
415 characterisation. *International Journal of Climatology*, 32(3), 406–418.  
416 <https://doi.org/10.1002/joc.2280>
- 417 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](https://www.cazalac.org/mwar_lac/fileadmin/imagenes2/Remote_Sensing/Princeton_Da)  
442 [y1\\_Overview\\_and\\_Station\\_Merging\\_light.pdf](https://www.cazalac.org/mwar_lac/fileadmin/imagenes2/Remote_Sensing/Princeton_Da)
- 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](http://www.droughtmanagement.info/literature/WMO_drought_monitoring_early_warni)  
447 [ng\\_2006.pdf](http://www.droughtmanagement.info/literature/WMO_drought_monitoring_early_warni)
- 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](https://library.wmo.int/doc_num.php?explnum_id=7768)
- 451 Yevjevich, V. (1967). *An objective approach to definitions and investigations of continental*

452        *hydrologic droughts. Hydrology Paper 23* (Vol. 23). Fort Collins, Colorado. Retrieved  
453        from  
454        [https://dspace.library.colostate.edu/bitstream/handle/10217/61303/HydrologyPapers\\_n2](https://dspace.library.colostate.edu/bitstream/handle/10217/61303/HydrologyPapers_n2)  
455        3.pdf  
456