An approach to characterise spatio-temporal drought dynamics

Vitali Diaz [a,b,*], Gerald A. Corzo Perez [a], Henny A.J. Van Lanen [c], Dimitri Solomatine [a,b,d], Emmanouil A. Varouchakis [e]

[a] IHE Delft Institute for Water Education, Hydroinformatics Chair group, Delft, 2601 DA, the Netherlands
[b] Delft University of Technology, Water Resources Section, Delft, the Netherlands
[c] Hydrology and Quantitative Water Management Group, Wageningen University, Wageningen, the Netherlands
[d] Water Problems Institute of the Russian Academy of Sciences, Moscow, Russia
[e] School of Environmental Engineering, Technical University of Crete, Chania, Greece

A B S T R A C T

The spatiotemporal monitoring of droughts is a complex task. In the past decades, drought monitoring has been increasingly developed, while the consideration of its spatio-temporal dynamics is still a challenge. This study proposes a method to build the spatial tracks and paths of drought, which can enhance its monitoring. The steps for the drought tracks calculation are (1) identification of spatial units (areas), (2) centroids localisation, and (3) centroids linkage. The spatio-temporal analysis performed here to extract the areas and centroids builds upon the Contiguous Drought Area (CDA) analysis. The potential of the proposed methodology is illustrated using grid data from the Standardized Precipitation Evaporation Index (SPEI) Global Drought Monitor over India (1901-2013), as an example. The method to calculate the drought tracks allows for identification of drought paths delineated by an onset and an end in space and time. Tracks, severity and duration of the drought are identified, as well as localisation (onset and end position), and rotation. The response of the drought tracking method to different combinations of parameters is also analysed. Further research is in progress to set up a model to predict the drought tracks for particular regions across the world, including India (https://www.researchgate.net/project/STAND-Spatio-Temporal-ANalysis-of-Drought).

1. Introduction

Drought is a regional phenomenon that often covers large territorial extensions (World Meteorological Organization WMO, 2006). It can occur anywhere in the world with severe consequences (impacts) in water resources and socioeconomic activities (Below et al., 2007; Sheffield and Wood, 2011; Tallaksen and Van Lanen, 2004; Wilhite, 2000). WMO stresses that to improve drought impacts mitigation, it is necessary to develop and implement national policies based on the best description and characterisation of drought (World Meteorological Organization WMO, 2006).

There is no unique definition of drought. However, there is an agreement that it is an anomaly in precipitation and temperature that when extended over a region causes a lack of soil moisture, runoff and groundwater (Mishra and Singh, 2010; Van Loon, 2015). This lack of water is expressed by a drought indicator, which transforms the hydrometeorological variable into a value that is related to such a water anomaly (Mishra and Singh, 2011; Wanders et al., 2010). In drought monitoring, the drought indicators are generally used to identify the lack of water.

Regarding drought monitoring, Hao et al. (2017) provide an overview of its status for regional and global applications. They report as an essential advance the integration of more data resources to feed drought indicators, allowing for a better description of hydrometeorological and vegetation condition. This integration includes the use of hydrological simulations, as well as remote sensing, and forecasting data. For instance, the European Drought Observatory (Sepulcre-Cantó et al., 2012) provides the condition of drought evolution (development) in Europe based on satellite observations and modelled soil moisture. On the one hand, current drought monitoring allows for following drought development for a specific location or a given region, mainly through the visualisation and analysis of time series of drought indicators. On the other hand, the spatial condition of drought, including its extent, is monitored with the help of time snapshots, which provide qualitative information on the spatial behaviour of the phenomenon.

In terms of the spatial development of the drought, nowadays, the available drought monitors deliver information about the spatial extent of droughts (i.e. snapshots). However, consistent procedures for tracking of drought areas are lacking, not allowing for assessing temporal variations that form its spatio-temporal dynamics (Hao et al., 2017). In

addition, the spatial distribution of drought at a specific time does not give information about the spatial pathway of the droughts. Implementing the data analysis and hydroinformatics technologies to trace drought in space and in time on drought monitors can enhance its spatial tracking and prediction.

The necessity to increase our understanding of the spatio-temporal development of drought has motivated the studies where drought is considered as a phenomenon that has at least the following main characteristics: duration, intensity (magnitude), and spatial extent (area) (Andreadis et al., 2005; Corzo Perez et al., 2011; Diaz et al., 2018; Herrera-Estrada et al., 2017; Lloyd-Hughes, 2012; Sheffield et al., 2009; Tallaksen et al., 2009; Van Huijgevoort et al., 2013; Vernieuwe et al., 2019). A general framework for carrying out spatio-temporal analysis of drought can be formulated based on these studies, and it can be briefly described as follows. First, a given drought indicator is used to transform the hydro-meteorological variable into water anomalies. The drought indicator is computed in a spatial context, where the study region is embedded in a grid. Then, by establishing a threshold on the drought indicator, the condition of non-drought/drought is identified in each of the cells of the grid. This condition can be expressed in a binary way, i.e. using 0 s and 1 s. Finally, neighbouring cells showing the same drought condition are aggregated into regions (clusters) by applying a clustering technique. In this way, drought is defined in space and in time, with a spatial extent and duration.

The spatio-temporal analysis of drought that would also include the spatial drought tracking explicitly is however limited to a few studies such as Diaz et al. (2018), Herrera-Estrada et al. (2017), and Zhou et al. (2019). The first two address the analysis for large-scale studies and the latter presents a basin-scale application. Although there are other publications that consider the study of drought extent locations, they miss the explicit calculation of spatial drought tracks. Following the framework mentioned in the previous paragraph, after the extraction of drought extents (areas), it is possible to identify their location and further construction of the spatial tracks (defined by the linkage between consecutive centroids in time). The calculation and further analysis of these tracks, along with outcome on drought areas, may help to answer the following questions regarding drought dynamics. What are the main places where drought remains? Are there predominant routes followed by drought? How fast does drought change (its extent and location) along its spatial path? Literature review shows that the development of methodologies to describe drought dynamics is still in progress, therefore more research is needed in this regard (e.g. Herrera-Estrada et al., 2017; Vernieuwe et al., 2019; Zhou et al., 2019).

This study aims to explain the main principles of a new method that complement current drought monitoring by tracking the spatial extent of drought (referred to in this document as area, or cluster). In this study, the description and the application of the methodology to calculate drought tracks are presented in detail. The proposed method is accompanied by an algorithm to calculate the drought characteristics. Both methods are described after this introduction section. The spatio-temporal Contiguous Drought Area (CDA) analysis (Corzo Perez et al., 2011) is used as the basis for the development of the tracking method. The CDA is applied to identify the neighbouring cells that form the drought clusters. A drought is defined by an onset location, pathway over time, and an end location based on the built tracks. A new drought characteristic is introduced in this study, namely rotation (Sect. 2.2), a feature often used when tracking objects in space (details in Sect. 2.2). The application of drought tracking method is performed over the country of India for the period 1901-2013.

2. Methods

2.1. S-TRACK: spatial tracking of drought

The spatial identification of drought tracks is firstly introduced by Diaz et al. (2018) and further developed in this research. S-TRACK con-
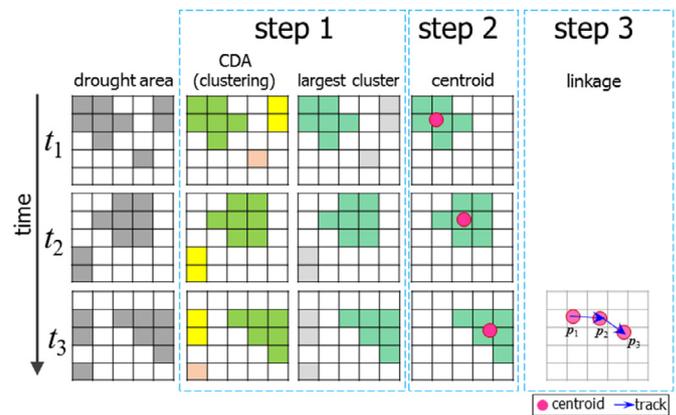

Fig. 1. Schematic overview of S-TRACK method for spatial drought tracking which involves: (step 1) spatial drought units (clusters) computation, (step 2) centroids localisation, and (step 3) centroids linkage (see Sect. 2.1). An example is presented for the case of three times steps: from t_1 to t_3. Columns in the diagram show the sequence of the steps. Coloured cells in the first column indicate all cells in drought. Colours in the second column point out different clusters identified. In the third column, the largest contiguous area in drought is presented with a different colour. Only the largest cluster is shown in the fourth column and its centroid (p) is indicated by a point. Subscripts indicate time steps.

sists of the three main steps: (1) calculation of the spatial drought units (referred to here also as areas or clusters); (2) localisation of centroids; and (3) linkage of centroids (Fig. 1).

Step 1. *Spatial drought units computation*

In the spatial context, drought units are identified by means of the Contiguous Drought Area (CDA) analysis (Corzo Perez et al., 2011). A CDA is composed of neighbouring cells in drought. These cells in drought are identified in each time step. When the drought indicator is below or equal to the selected threshold, the value of 1 is used to indicate that the cell is in drought, otherwise, the value of 0 is used, indicating non-drought. Drought indicators (DIs) are mathematical representations of a water anomaly (see Sect. 2.3.1). In general, CDA can be applied over any DI that is in a grid form. Following the CDA methodology, in each time step, the CDAs are computed.

CDA analysis follows a connected-component labelling approach to cluster the cells in drought (Haralick and Shapiro, 1992). In this approach, a two-scan algorithm is applied. Firstly, each cell is numbered for location issues. Then, the first run is performed where the binary grid is explored and connected (contiguous) components (cells) are assigned with provisional labels. These labels point out the connection of every cell with its 8 nearest neighbours. Within the grid, in a section of 3×3 cells, 9 cells in total, the central cell has 8 surroundings. In this first run, the cell's label does not refer to the number of cluster yet but to the cells with which the given cell is connected. Finally, a 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. CDA analysis is conducted in each time step over the whole grid. For more details on CDA analysis refer to Corzo Perez et al. (2011).

The use of CDA relies on the assumption that the binary description of drought condition (0 s and 1 s) is homogeneous over the whole grid. Thus, if two or more cells denote drought conditions (value of 1), and are contiguous in space, it is assumed that all of them are part of the same drought unit. In this respect, it is recommended to choose a drought indicator that considers the normalisation of the values in the spatial domain. In this study, a standardised drought indicator is applied as mentioned afterwards, which allows the clustering of neighbouring cells in drought (cells with 1 s).

After clusters (areas in drought) are identified, the major (largest) one is identified in each time step t (Fig. 1). As the tracking algorithm
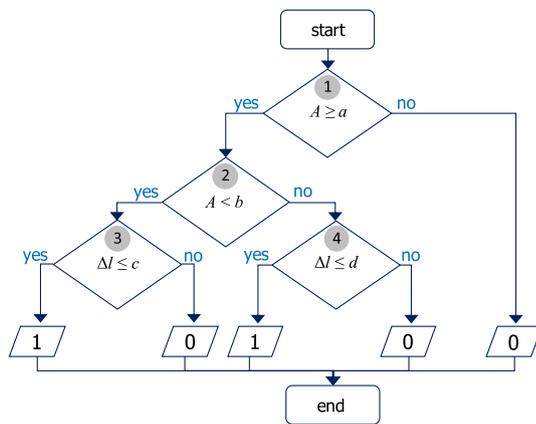
Fig. 2. Flowchart showing the rules for linking drought areas (clusters) in time. Numbers in the boxes indicate the sequence of rules 1 to 4. The output of 1 is used to point out that the drought area A at time t joins its predecessor at time t–1, otherwise 0 is retrieved. The distance between the centroids at times t and t–1 is represented by Δl. The linking algorithm has the following parameters: a, b, c and d. The first two used to control drought area A, and the last two, to check distance Δl.

focuses on the calculation of the major spatial drought extent in each time step, small or one-cell units are discriminated with the selection of the largest one, allowing the elimination of possible artefact drought areas.

Step 2. *Centroids localisation*

After identification of the major (largest) drought cluster, its centroid (p) is calculated in each time step. This feature is used as the location of the cluster in a similar way as Corzo Perez et al. (2011) and Lloyd-Hughes (2012) present. The way in which the clusters are joined in time is explained in the following step. Step 2 and 3 presented in this document, are an extension of the CDA analysis of Corzo Perez et al. (2011). Another possibility to indicate the location of a given cluster is, for instance, to use the point with the lowest drought indicator value (Andreadis et al., 2005; Herrera-Estrada et al., 2017). In this research, we chose the centroid since we already reduce the spatial representation of drought indicator by using only 1 s and 0 s, i.e. drought and non-drought condition, respectively.

Step 3. *Centroids linkage*

The algorithm to link centroids of consecutive clusters in time is a set of rules to separate or join the sequence in time (Fig. 2). The rules consider two types of threshold parameters: (1) two that control the magnitude (size) of the cluster (A, with dimensions L^2), and (2) two that constrain the Euclidean distance between consecutive clusters (Δl, with dimensions L) (Fig. 2). The parameters are denoted as follows: a, b, c and d. The first two are used to the drought area A, and the last two to the distance Δl. The output in this step is the time series of 0 s and 1 s, denoted by S(t). Here, the value of 1 indicates the linkage of clusters in time. If the cluster at time t is not connected with the cluster at time t–1, the value of 0 is used instead. Consecutive values of 1 s in the time series S show the occurrence of what is defined as a drought track. The flowchart of the rules for linking the centroids is presented in Fig. 2, and below these rules are explained.

Centroids linkage starts by identifying if the cluster area A is higher than a (Fig. 2, rule 1). This first comparison helps to discriminate small clusters. If A is below a, there is no connection between consecutive clusters and this procedure finalises, retrieving 0. Before comparing the distance between areas (Δl), the second comparison of A is applied to identify if it is a "very large" area (Fig. 2, rule 2). Parameter b is proposed to consider these large areas. When A is below b, the parameter c is used to compare distances between clusters (Fig. 2, rule 3). Otherwise,

when A is above b ("very large" area), to restrict the distances, parameter d is considered instead (Fig. 2, rule 4). The reason of the second comparison of cluster areas and the use of parameter d is because centroids of clusters with a considerable size may be located farther away from each other and then the distance Δl could fall outside of the limit indicated by parameter c (Fig. 3).

Another parameter that could be included in this linkage algorithm is the degree of overlap between consecutive clusters in time. This way of intersection is not considered directly in our linkage algorithm as a parameter (e.g. percentage of overlapping). The overlap is contemplated in the use of the parameters that control the distance between clusters. An intersection may occur when the distance between centroids is short (Fig. 3).

2.2. Calculation of drought characteristics

The methodology to build drought tracks allows for the identification of paths with an onset and an end location. The information calculated along the paths can help to describe the occurrence of drought. Particularly, it is possible to extract information regarding the duration, severity, as well as rotation. In the following analysis of the spatio-temporal drought dynamics, severity has a different meaning compared to on-site analysis or CDA studies. In the latter, it expresses a certain degree of water missing, an anomaly compared to normal conditions. Herein, severity has a spatial meaning, it is connected to the temporal evolution of the size of the area in drought, irrespective of the strength of the drought. In the following paragraphs, the procedure to calculate drought characteristics is presented. The proposed approach is called DDRASTIC-spatial (Drought DuRAtion, SeveriTy and Intensity Computing-spatial events). DDRASTIC-spatial is applied after drought tracks are identified by the S-TRACK algorithm. This approach has as a predecessor (Diaz et al., 2019), which however does not consider the elements related to the spatial domain, such as clusters, locations and paths.

For the calculation of the drought duration, firstly the onset and the end are obtained: the time series S(t) of 1 s and 0 s calculated with S-TRACK method is analysed to do so. As mentioned, the consecutive sequence of 1 s in the time series S, indicates the occurrence of a drought track. One isolated value of 1 shows the linking of two clusters in time. Two consecutive values of 1 show the linkage of three clusters in time, and so on. In a sequence of 1 s, the time of the first value of 1 (t_{first}) is the time step at which the second and first cluster are connected. The time step of the last value of 1 (t_{last}) is the one when the last and the penultimate clusters are linked. The onset ti is defined as $ti = t_{\text{first}} - 1$, while the end tf as $tf = t_{\text{last}}$. The duration (dd) is calculated with Eq. (1).

$$dd = \sum_{t=ti}^{tf} S(t) \tag{1}$$

The magnitudes of areas of the largest clusters calculated in each time step with S-TRACK method are saved in the time series DA (drought area). The drought area is used as the measure of the drought severity (ds), which is computed as the sum of drought areas of the period defined by the onset (ti) and the end (tf) (Eq. (2)). Drought intensity (di) is defined as the ratio between drought severity and duration (Eq. (3)).

$$ds = \sum_{t=ti}^{tf} DA(t) \tag{2}$$

$$di = ds/dd \tag{3}$$

Identification of locations where a drought path starts and ends can provide its main direction. The initial and final locations are identified using the centroids of the first and last cluster, respectively. The location is a relative position in the spatial domain of the study region. It refers to a point in the axes south-north (S-N) and west-east (W-E) (Fig. 4). The origin of the axes is assigned arbitrarily, here it is proposed to place this origin in the centroid of the study region. The centroid of a particular
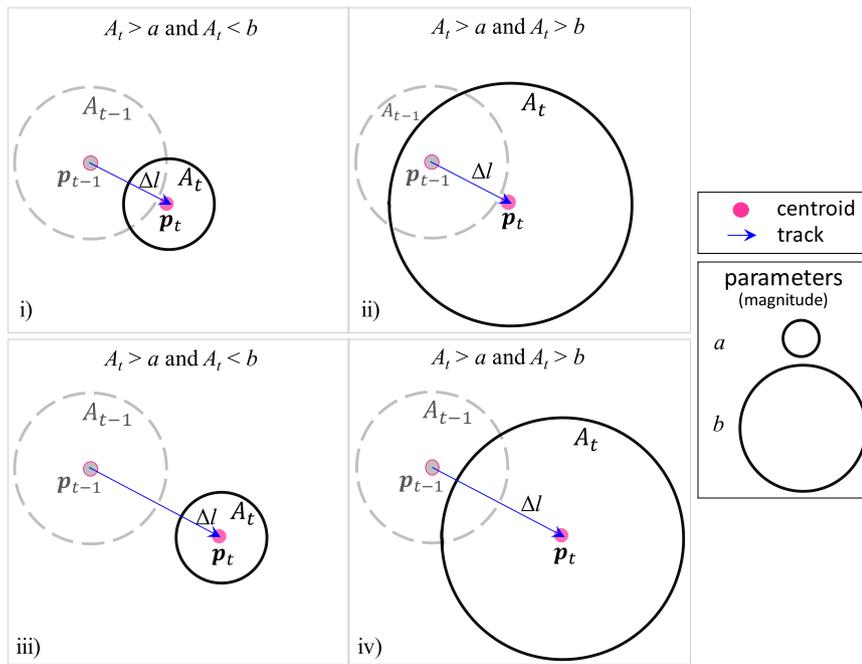
Fig. 3. Schematic overview of the four cases of linking clusters (drought areas) in time. Area at time t is indicated by A_t (bold circle) and its predecessor at time t–1 by A_{t-1} (dashed circle). Centroids of areas A_t and A_{t-1} are denoted by p_t and p_{t-1} (points), respectively. Distance between centroids is represented by Δl (arrow). An example of the size of parameters a and b is represented by the circles shown on the right. Centroids in both (i) and (ii) have the same location, in the same way, the centroids in both (iii) and (iv). Areas A_t in (i) and (iii) are of similar size and between the parameters a and b. On the other hand, in (ii) and (iv), areas A_t are also equal but above those parameters (case of a "very large" area). Only the parameters of drought area are represented in this figure. Schemes (i) to (iv) help to illustrate the relevance of using parameters that consider not only the magnitude of areas but also the distance between them within the linking algorithm. As a distance limit that helps in linking large areas may not be adequate in connecting smaller ones, as shown in (iv) and (iii), the two distances parameters are proposed in the linking algorithm (see Sect. 2.1 for details).
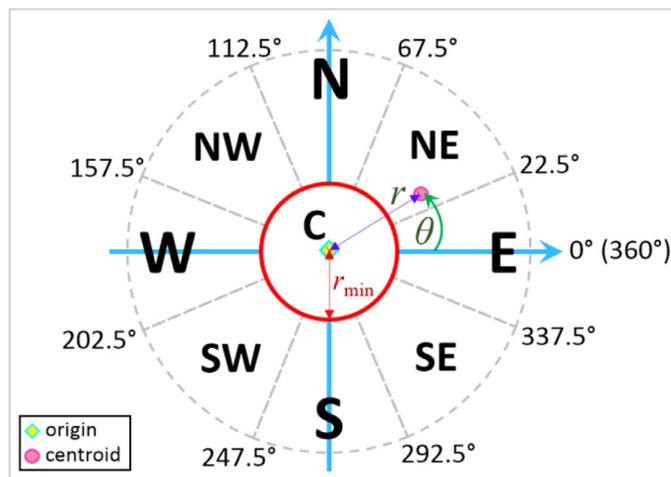

Fig. 4. Schematic overview of the procedure to define centroid's location of a cluster. A centroid can be located in one of nine positions: centre (C), east (E), northeast (NE), north (N), northwest (NW), west (W), southwest (SW), south (S) and southeast (SE). The symbol r stands for the distance between the cluster's centroid and the one of the region. The angle between the W-E axis and the line defined by centroid's cluster is indicated by θ. The radius to define if a cluster is located in the centre (C) of the region is pointed out by $r_{\min}$.

Table 1
Rules to define the location of a centroid's cluster. Nine positions are proposed: centre (C), east (E), northeast (NE), north (N), northwest (NW), west (W), southwest (SW), south (S) and southeast (SE).

Id	Rule				Location
0	$r \leq r_{\min}$				C
1	$r > r_{\min}$ and	0°	$\leq \theta <$	22.5° or 337.5° $\leq \theta <$ 360°	E
2	$r > r_{\min}$ and	22.5°	$\leq \theta <$	67.5°	NE
3	$r > r_{\min}$ and	67.5°	$\leq \theta <$	112.5°	N
4	$r > r_{\min}$ and	112.5°	$\leq \theta <$	157.5°	NW
5	$r > r_{\min}$ and	157.5°	$\leq \theta <$	202.5°	W
6	$r > r_{\min}$ and	202.5°	$\leq \theta <$	247.5°	SW
7	$r > r_{\min}$ and	247.5°	$\leq \theta <$	292.5°	S
8	$r > r_{\min}$ and	292.5°	$\leq \theta <$	337.5°	SE

r = distance between centroid's cluster and the one of the region;
θ = angle between W-E axis and the line defined by centroid's cluster;
$r_{\min}$ = limit distance to consider the location in the centre (C) of the region.

cluster can be located in one of the nine proposed positions: centre (C), east (E), northeast (NE), north (N), northwest (NW), west (W), southwest (SW), south (S), and southeast (SE) (Fig. 4). Centre (C) is situated in the centroid of the study region (Fig. 4). A point (centroid) is in the centre if the distance (r) between such point and the origin is within the $r_{\min}$ radius (Fig. 4). If distance r is out of the $r_{\min}$ radius, the location is assigned based on the angle θ. This angle is calculated between the W-E axis and the line defined between the centroid and origin (Fig. 4). All the rules to identify the centroid's location are presented in Table 1. Within the algorithm, instead of letters, locations are denoted by means of numerical identifiers (Ids) as presented in the first column of Table 1.

Drought tracks provide the visual overview of how drought moves in the spatial domain. Initial and end location (initial and end point of the track) help to identify the direction followed by a given drought

cluster. Yet another characteristic that complements the description of the drought dynamics is its rotation. This characteristic is defined as the circular orientation followed by the spatial extent of drought. Rotation is a feature commonly attributed to objects that experience changes in space. It is an essential characteristic analysed in other weather-related phenomena such as cyclones (e.g. Chavas et al., 2017; Rahman et al., 2018) but that has not been used and explored much in droughts so far. This characteristic is included because it is foreseen that it can help to analyse the impact of the drought drivers, such as the climate and land surface control factors, on the spatial development of droughts. The drought rotation patterns are expected to be different for each combination of the aforementioned factors. We see this study as an initial step towards developing a technological framework for identifying and interpreting the drought rotation.

As the drought track can switch between clockwise and counterclockwise along the pathway, we propose to classify the rotation in a more general way as (1) mostly clockwise (cw), or (2) mostly counterclockwise (ccw) (Fig. 5). To determine the rotation, a procedure is suggested which makes use of the centroids' coordinates. The algorithm is based on computing a polygon's area (A) from the vector with the coordinates x and y representing the vertices (Eq. (4)). In this algorithm, firstly the sum of products between the coordinates x and y, denoted
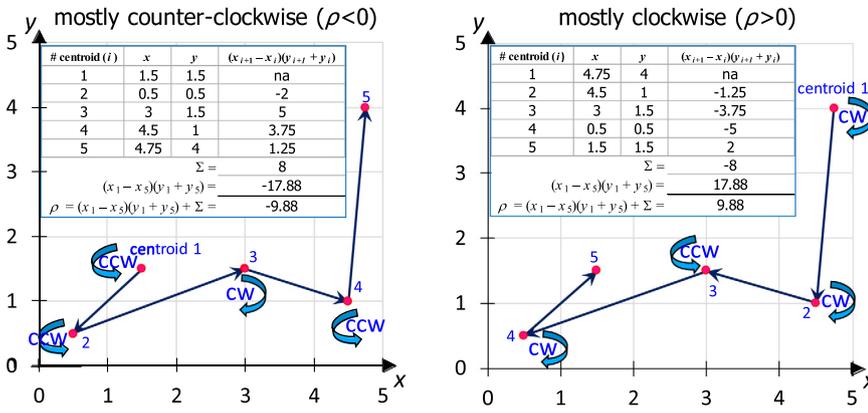
Fig. 5. Example of rotation calculation. Two types are considered: (1) mostly counter-clockwise when $\rho<0$ (left); and (2) mostly clockwise when $\rho>0$ (right). The number in each centroid (point) indicates the tracking sequence. Arrows show the track direction and the rotation. Rotation of each line segment is also pointed out by cw and ccw that stand for clockwise and counter-clockwise, respectively.

by ρ (Eq. (5)), is calculated. Then, ρ is applied to define the rotation direction (Eq. (6)). The coordinates x and y are taken from the ones of centroids' clusters. When there are only two points (two clusters), or when the track is horizontal or vertical, the rotation is not defined, because ρ takes the value of zero. In Fig. 5, two examples of the calculation of rotation are shown by way of illustration. One example is presented for mostly counter-clockwise (Fig. 5 (left)) and one for mostly clockwise (Fig. 5 (right)). We chose this approach to compute rotation because it distinguishes between "big" and "small" turns in the calculation (Eq. (5)). The fourth column in both tables presented in Fig. 5 provides examples of how the magnitude of each turn is considered differently in the rotation algorithm.

$$A = \frac{1}{2}|\rho| \tag{4}$$

$$\rho = (x_1 - x_n)(y_1 + y_n) + \sum_{i=1}^{n-1} (x_{i+1} - x_i)(y_{i+1} + y_i) \tag{5}$$

$$\omega = \begin{cases} \text{cw (mostly clockwise) if } \rho > 0 \\ \text{ccw (mostly couter} - \text{clockwise) if } \rho < 0 \\ \text{nan (not defined) if } \rho = 0 \end{cases} \tag{6}$$

2.3. Experimental setup

2.3.1. Drought indicator data

Drought tracks were calculated with S-TRACK algorithm for the period 1901 to 2013 (113 years). The analysis was conducted for India, on a monthly basis. Data from the Standardized Precipitation Evaporation Index (SPEI) Global Drought Monitor (http://spei.csic.es/) was used (Beguería et al., 2014) to test the proposed methodology for drought tracking and characterisation. The procedure to calculate SPEI (Vicente-Serrano et al., 2010) is similar to the one used to compute the Standardized Precipitation Index (SPI) (Mckee et al., 1993), but taking into account precipitation (P) minus potential evaporation (E) instead of only P. SPEI data from the drought monitor are in a grid form for different temporal aggregation periods. In this study, we used SPEI-6, which corresponds to anomalies of the six-month accumulation of $P - E$. This aggregation usually refers to extended periods of lack of water availability, therefore consequences of what is commonly called meteorological drought are closer to that caused by the so-called hydrological drought (World Meteorological Organization WMO, 2012).

2.3.2. Drought areas and centroids

Before the application of the drought tracking algorithm, the size of the largest clusters and the distances between the centroids of consecutive clusters in time were calculated. This calculation was performed, on the one hand, to understand the order of their magnitude and frequency, and on the other hand, to set the values of the tracking algorithm parameters.

For the definition of drought areas, usually, the threshold of -1 is used to indicate drought condition in the drought indicators that follow a similar methodology than SPI, also referred to as standardised ones. In this research, the same threshold (SPEI = -1) was selected to define drought condition in each cell of the grid in each time step. When SPEI was below -1, with 1 s the drought condition was indicated, in another case, with 0 s the non-drought status was pointed out. This binary representation allowed the identification of spatial drought units (clusters) through the application of the spatio-temporal analysis of Contiguous Drought Area (CDA) (Sect. 2.1).

The largest clusters in each time step were then identified. The area of the largest cluster was compared with the total one to identify the similarity in size between them. It is assumed that the more similar the larger area to the total one, the better the identification of the drought tracks will be. This stands because the tracking algorithm focuses on only one area per time step. For the comparison, the area of all clusters (DA_total) and the area of the largest one (DA_largest) were calculated. Both areas were expressed as percentages calculated as the ratio between the number of cells in drought and the total number of cells. The total number of cells considered for India was 1,173.

Once the centroids were identified, the distances between consecutive centroids were calculated over time (Sect 2.1). Both the clusters and the distances were calculated for the entire period of analysis on a monthly basis.

2.3.3. Tracking algorithm calibration and evaluation

S-TRACK uses four parameters and they have to be user-defined, or, better, calibrated. The problem of calibrating this algorithm is that there is no ground-truth data on the drought tracks, hence, some aliases should be used. A full-fledged calibration procedure can be applied (e.g. one of the randomised search algorithms, like an evolutionary algorithm). The optimal parameters should be selected based on information of reported drought. In the absence of drought tracks, it is necessary to have data at least on the onset and end month of the reported droughts. The near-optimal parameters are those that provide the best match between the observed and calculated onsets/ends.

However, in this paper, we applied a simplified procedure. Considering that there is no available information to compare the calculated drought paths in the study area, we limited the procedure to a qualitative analysis of the paths of the most severe droughts reported in the analysis period. The droughts of 1905, 1942, 1965, 1972, 1987, 2000, and 2002 were considered because their severe impacts were referenced (Guha-Sapir, 2018). The qualitative evaluation was focused on the analysis of the extreme incidences using a combination of parameters. From the whole set of combinations, we have chosen three: the one that produces the smallest number of droughts paths (combination_1), the one that yields the largest number of droughts paths (combination_3), as
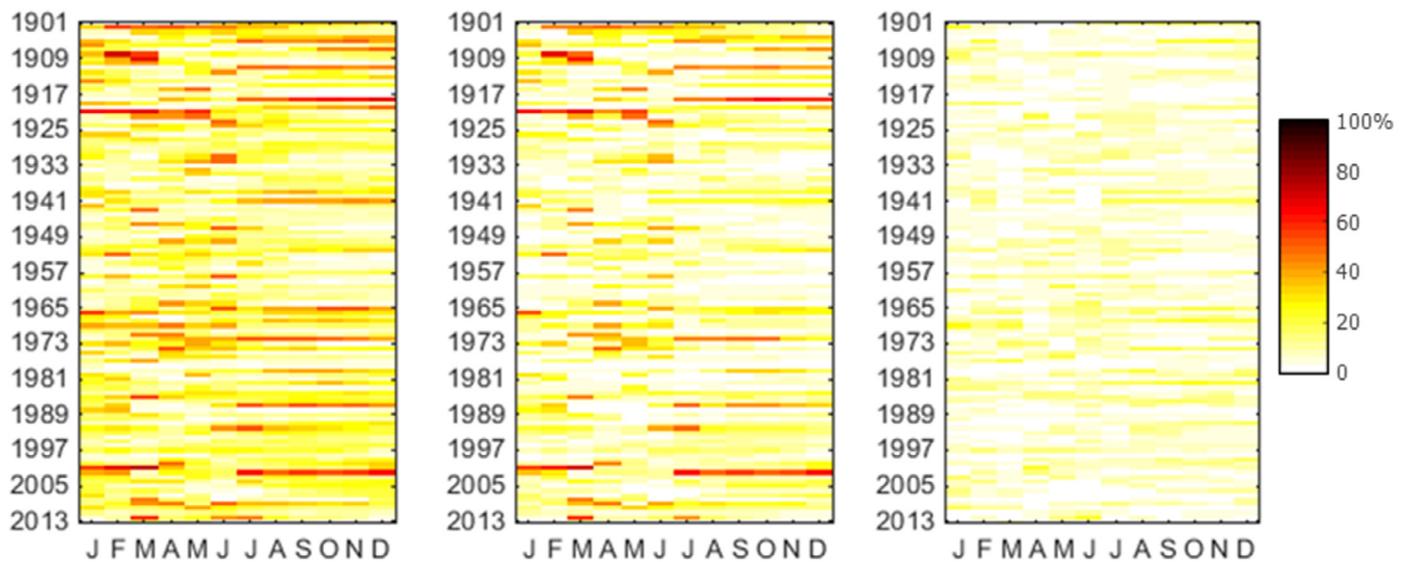
Fig. 6. Percentage of drought area considering all clusters (DA_total, left), and considering only the largest one (DA_largest, centre). Right panel shows the difference between DA_total and DA_largest.

well as the one that produces the number of drought paths similar to the number of years of the analysis period (combination_2).

Yet another important part of the algorithm evaluation is its sensitivity analysis. It allows for assessing the robustness of the method through the analysis of the outputs under the variation of parameters (Pannell, 1997). Sensitivity analysis generally allows answering the following questions when evaluating an algorithm. How parameters and output are related? What level of accuracy in the parameters is required? Which parameters are more sensitive, and what drought characteristics do they influence most? What are the consequences of varying the parameters?

In principle, calibration and sensitivity analysis steps have to be coordinated, e.g. allowing for removal of less sensitive parameters from the set of the parameters to be calibrated (e.g. to speed up calibration). In this work, as the algorithm is not computationally complex, this approach was not followed. The sensitivity analysis was performed to assess the effect of parameters over the identification of droughts tracks and characteristics. The questions mentioned in the previous paragraph were used as a guideline to perform such an analysis.

3. Results

3.1. Drought areas and centroids

Drought areas and centroids were computed for the period 1901 to 2013. With respect to the areas, firstly the comparison between the area of all clusters (DA_total) and area of the largest one (DA_largest) was performed. Fig. 6 shows the monthly values of both DA_total and DA_largest arranged in matrices. Columns indicate months from January (J) to December (D), while rows point out the year from 1901 to 2013. Drought area magnitude is indicated with a colour scale, where the darker the colour, the higher the drought area is. The white colour denotes months with small drought areas (less than 10%). It is observed that for almost all months DA_total and DA_largest have similar values, and this agreement is especially high for the largest values. The drought area average for the period was 17.4% for DA_total, and 11.5% for DA_largest. Fig. 6 (right) presents the difference between DA_total and DA_largest. Across the whole period, the average of the differences was 5.9%. As DA_largest and DA_total were very similar, it can be considered that the largest cluster is a good proxy to analyse how drought changes in the region without considering the occurrence of two consecutive drought tracks.
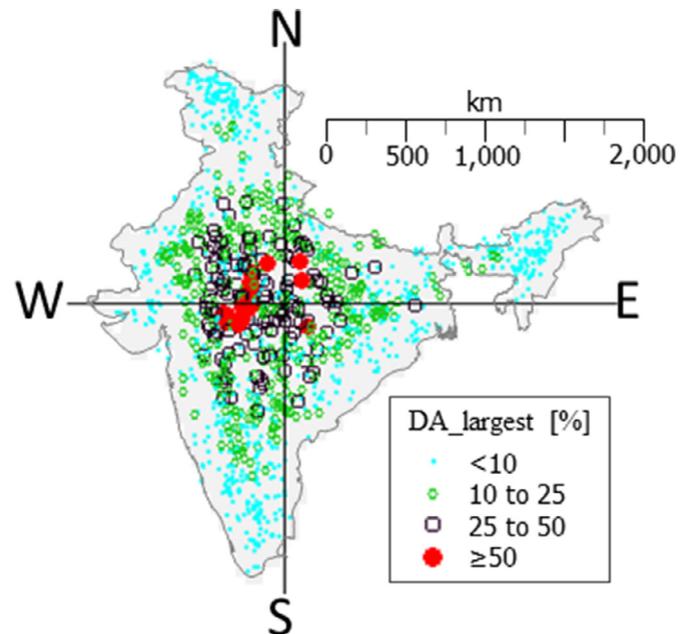
Fig. 7. Centroids of the largest clusters (DA_largest) identified on a monthly basis. Spatial drought extent is schematized by four symbols pointing out the drought area. The origin of the axes is placed in the centre of the country.

The centroids of the largest clusters are presented in Fig. 7. The spatial drought extent is shown schematically with symbols that indicate four intervals of the percentage of drought area with respect to the country extent. The origin of the axes is placed in the centre of the country. It is observed that the spatial distribution of the centroids is almost uniformly distributed over India. However, a higher density of the areas with a considerable extent can be seen in the central region.

The distances between consecutive clusters in time were calculated also for the whole period. Fig. A1 (Appendix A) presents the area of the largest cluster (DA_largest) and the distance (Δl) between consecutive clusters in time. It can be observed that the occurrence of DA_largest is greater than 25% during all decades of the analysis period. A pattern is observed between DA_largest and Δl: when DA_largest increases, Δl usually decreases. This behaviour was expected, because the more the area increases, the smaller the distance between centroids becomes.
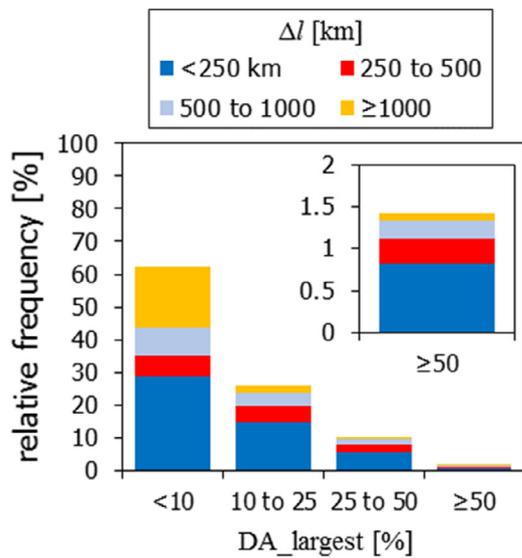
Fig. 8. Relative frequency of the largest cluster area (DA_largest) and distances (Δl) between consecutive clusters in time.

This means that the location of the consecutive clusters is becoming the same. When Δl does not follow this behaviour, it might be because the consecutive areas in time are very far each other, i.e. they are part of different drought paths.

Fig. 8 shows the frequency of both the largest cluster area (DA_largest) and the distance (Δl) between consecutive clusters in time. For both variables, results are displayed in four intervals. It was observed that as the area increases, the frequency of long distances between these areas decreases, while the frequency of small distances increases. For the DA_largest the intervals of 25-50% and ≥ 50%, the frequency of the small distances (Δl < 250 km) was slightly greater than half of all the distances. This results of DA_largest and the distances Δl confirm quantitatively what is observed in Fig. A1: in general when the area grows, the distances between the centroids tend to decrease. On the other hand, the small value of the frequency of large distances in large areas (intervals 25-50% and ≥ 50%) indicates that there are large consecutive areas in time that are not necessarily connected to each other.

3.2. Sensitivity of S-TRACK results to the choice of parameters

S-TRACK algorithm has a number of parameters. For the reasons mentioned above (Sect. 2.3.3), it is useful to study the sensitivity of its outputs to these parameters. Based on the results of areas and distances between clusters (Sect. 3.1), the S-TRACK algorithm was set to take parameters values within the following ranges: $a \leq 50$, $b \geq 50$, $c \geq 50$, and $d \geq 50$th percentile (median). As mentioned, a and b are parameters that control the size of clusters (areas), and c and d are parameters that constrain the distances between consecutive clusters in time. The average duration, average severity, onset location, as well as end location, were calculated for the different combinations of parameters. Results for a (30, 40, and 50), b (50, 70, and 90), c (50, 60, 70, 80, and 90), and d (50, 60, 70, 80, and 90) are presented in Figs. 9 and A2 to A6 (Appendix A). The a and b parameters are expressed as percentage of drought area and c and d as km. At the end of this section, a summary of the results is presented.

Fig. 9 shows the number of drought paths (combination of tracks linked in time). It is observed that the number of drought paths, in general, increased when a decreased. This is expected since parameter a is the one that determines if a cluster joins the consecutive clusters in each time step. When a is small, more clusters are expected to be connected in each time step and therefore more drought paths can be identified. The value of b (used for "very large" areas) influenced the number of paths

less than a, e.g. so that when b increased, there was a small proportional increase in the number of paths for all combinations of parameters. The combined variation of b and c influenced more the number of paths for small values of d. It was observed that in general, the number of paths drops when a increases and both b, c, and d decrease. In general, the number of drought paths was more sensitive to the changes in parameter a.

In Fig. A2 the average duration of drought paths is presented. Although the variation of average duration was small to the changes of parameters, a slight increase was observed, as a decreased and both b, c and d increased. The average duration was more sensitive to the increase in c and d that are the parameters that control the distance between consecutive clusters in time.

Regarding the severity, it was smaller when a increased and both b, c, and d decreased (Fig. A3). Severity is calculated as the ratio between the total sum of drought areas and duration (number of months), so it is getting lower as duration increases (see Eqs. (1), (2), and –(3)). Similarly to the number of drought paths, the average severity was also sensitive to changes in parameter a. It was observed that when the number of paths decreased, the average severity increased (Figs. A2 and A3). This behaviour in severity is the effect of the selection of a that controls the size of the areas that are joined in each instant of time. If a is small, more areas can be joined and severity may decrease due to the effect it produces the pooling of more areas of small sizes divided by a longer duration (see Eqs. (1), (2), and –(3)).

Figs. A4 and A5 show the mode of onset and end location of drought paths, respectively. In Fig. A4, not many changes were observed in the onset location. East was the most common onset location in most combinations of parameters, followed by South. On the other hand, Fig. A5 shows the end locations that in most combinations the South, followed by East were the most common. When both a decrease and b increased, the South was the most common end location.

Fig. A6 shows the mode of rotation. It was observed in most cases that mostly clockwise (cw) was the common rotation in the drought paths. When a decreased and b increased, the mostly clockwise rotation was the most common rotation. This was the case when more drought paths were obtained. It was observed that rotation was most sensitive to the variations of c and d that are the parameters which control the distance between consecutive clusters in time.

Summary of results

Table 2 shows a summary of how the tracking algorithm responds to different combinations of parameters. In particular, the behaviour of the number of paths, duration, severity, onset and end location, as well as rotation, is indicated. The combinations where it was observed that the values of these characteristics tend to increase or decrease is presented. In general, the most sensitive parameter (important) is the one that controls the minimum area (parameter a). Changes in this parameter have more influence on the result of the number of drought paths and duration. Regarding duration and severity, it is observed that as the paths last longer the severity decreases. This may apply because the severity is calculated as the sum of the areas of clusters that belong to the drought duration. Thus, while the duration increases, the areas that are added tend to be smaller and then the sum does not increase significantly.

The combination 11 (Table 2) refers to the identification of paths of "very large" areas. In this combination, it is expected that the initial and final locations will be in the centre. Centroids of these cluster areas tend to be identical to that of the region. For these paths, it is also observed that the rotation tends to be clockwise.

In combinations 6, 7 and 14 (Table 2), by decreasing the parameter that controls the minimum area (parameter a), more drought paths are identified, with the characteristic of being long and with a small severity (formed by a number of smaller areas). In these combinations, drought paths usually start in the East and end in the South, with a clockwise rotation.

a = 50

b = 50

		d: 50	60	70	80	90
c	50	79	89	98	101	99
	60	79	89	98	101	99
	70	79	89	98	101	99
	80	79	89	98	101	99
	90	79	89	98	101	99

b = 70

		d: 50	60	70	80	90
c	50	79	81	89	92	92
	60	85	89	96	99	101
	70	91	94	98	101	101
	80	99	100	103	101	101
	90	103	105	106	100	99

b = 90

		d: 50	60	70	80	90
c	50	79	79	86	86	87
	60	90	89	94	95	96
	70	101	98	98	99	99
	80	108	103	101	101	101
	90	109	105	100	99	99

a = 40

b = 50

		d: 50	60	70	80	90
c	50	95	105	110	111	111
	60	95	108	112	112	112
	70	97	109	112	110	110
	80	100	109	112	109	109
	90	100	109	113	110	110

b = 70

		d: 50	60	70	80	90
c	50	95	98	103	106	107
	60	101	108	111	114	117
	70	109	112	112	113	113
	80	119	115	113	109	109
	90	125	122	119	111	110

b = 90

		d: 50	60	70	80	90
c	50	95	95	101	101	103
	60	109	108	112	113	115
	70	116	113	112	113	113
	80	118	112	109	109	109
	90	122	117	111	110	110

Legend: 79 – 150

a = 30

b = 50

		d: 50	60	70	80	90
c	50	112	120	124	123	118
	60	114	125	128	125	120
	70	121	128	129	123	118
	80	125	126	126	118	114
	90	135	133	131	121	116

b = 70

		d: 50	60	70	80	90
c	50	112	115	120	121	122
	60	119	125	128	129	131
	70	130	129	129	127	126
	80	139	129	125	118	118
	90	150	140	132	119	116

b = 90

		d: 50	60	70	80	90
c	50	112	112	118	117	119
	60	126	125	129	129	131
	70	135	131	129	128	128
	80	134	127	121	118	118
	90	138	132	121	116	116

Fig. 9. Number of drought paths obtained with different combinations of parameters.

Table 2

Summary of drought characteristics obtained with different combinations of parameters. Numbers in parentheses indicate the location as presented in Figs. A4 and A5. Abbreviations ccw and cw stand for counter-clockwise and clockwise, respectively.

#	a	b	c	d	Number of paths		Drought characteristics — Duration		Severity	Onset location	End location	Rotation
1	↑	↓	↓	↓	decreases	tends to decrease	decreases	tends to decrease	increases			
2	↑	↓	↑	↑	decreases		decreases		increases			
3	↑	↑	↑	↑	decreases		decreases		increases			
4	↑	↑	↓	↓	decreases		decreases		increases			
5	↓	↓	↓	↓	increases		increases		decreases			
6	↓	↓	↑	↑	increases		increases		decreases		tends to the south (7)	tends to cw
7	↓	↑	↑	↑	increases		increases	tends to increase	decreases	tends to the east (1)		tends to cw
8	↓	↑	↓	↓	increases		increases		decreases			
9	↑	↓	↓	↑	decreases		decreases		increases			
10	↑	↓	↑	↓	decreases		decreases		increases			
11	↑	↑	↓	↑	decreases		decreases		increases	tends to increase	tends to the centre (0)	tends to the centre (0)	tends to cw
12	↑	↑	↑	↓	decreases		decreases		increases			
13	↓	↓	↓	↑	increases		increases		decreases			
14	↓	↓	↑	↓	increases		increases		decreases		tends to the south (7)	tends to the east (1)	tends to ccw
15	↓	↑	↓	↑	increases		increases		decreases			
16	↓	↑	↑	↓	increases	tends to increase	increases		decreases	tends to decrease		
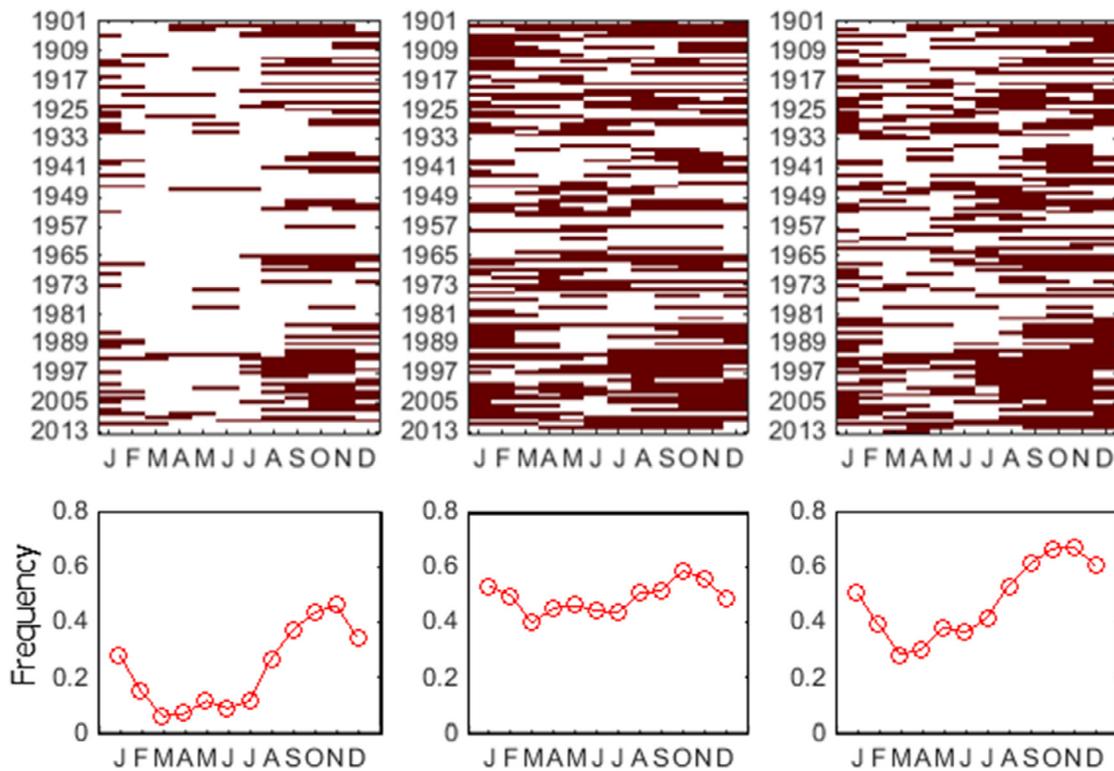
Fig. 10. Occurrence of drought paths calculated with three combination of parameters: (left) combination_1 ($a = 50$, $b = 50$, $c = 50$, $d = 50$), (centre) combination_2 ($a = 40$, $b = 50$, $c = 70$, $d = 80$), and (right) combination_3 ($a = 30$, $b = 70$, $c = 90$, $d = 50$). Consecutive cells in colour indicate the occurrence of a drought path (top). Frequency is calculated per column from January (J) to December (D) (bottom).

If the drought path starts in the South, it usually ends in the East, and in this case, the rotation is counter-clockwise, i.e. the rotation follows the minor turn (Table 2 (combination 14)). In other words, if the path starts in the South and ends in the East, it is more likely to be directed towards the East showing a counter-clockwise rotation, instead of going firstly to the West, then North and finally East, showing a clockwise rotation in this case.

3.3. Qualitative evaluation of drought paths

Seven of the most extreme droughts reported during the analysis period were selected for testing S-TRACK. These droughts, as it was mentioned earlier, correspond to the following years: 1905, 1942, 1965, 1972, 1987, 2000, and 2002. In the absence of information regarding the dynamics of the droughts, such as trajectories, our validation focused on the analysis of the calculated tracks in the period when the droughts occurred.

From the set of parameter combinations shown in the previous section, three were selected to analyse the calculated drought tracks. For the first combination (combination_1, $a = 50$, $b = 50$, $c = 50$, $d = 50$), the number of drought paths obtained was the lowest. For the second combination (combination_2, $a = 40$, $b = 50$, $c = 70$, $d = 80$), the number of drought paths was similar to the number of years of the analysis period, i.e. there was approximately one drought path per year. Finally, in the third combination (combination_3, $a = 30$, $b = 70$, $c = 90$, $d = 50$), the highest number of drought paths was identified.

Fig. 10 presents the occurrence of drought paths calculated for the three combinations of parameters. Columns indicate the months from January (J) to December (D) and the rows show the years. Consecutive cells in colour indicate the occurrence of a drought path (Fig. 10 (top)). The frequency per month was calculated to analyse the distribution of the tracks over the months (Fig. 10 (bottom)). In general, the month with the less frequency of drought tracks was March. From January to July, the first part of the year, the frequency was fewer than from August

to December. It was observed that when the number of drought paths increased (Fig. 10 (top, from left to right)), the frequency of drought tracks in each month increased as well (Fig. 10 (bottom)).

Fig. 11 shows the results from the calculation of clusters and distances between centroids to the construction of drought paths for the drought of 1987-1988. In Appendix A, one can see the other six droughts (Figs. A7, A8, A9, A10, A11, and A12). In Fig. 11 (top) clusters and centroids are presented. Areas of largest cluster (DA_largest) and distances between consecutive areas in time (Δl) are shown for the period from 1987/1 to 1988/6 (Fig. 11 (centre)). Duration of the drought paths is indicated in a schematic way with a horizontal line for each combination of parameters. Drought tracks calculated with the three combinations of parameters are also presented (Fig. 11 (bottom)). In most of the seven droughts, the maximum areas of the largest clusters were in the second half of the year and the first half of the following one (e.g. Fig. 11 (centre)). It was observed that, in general, when DA_largest increased, Δl usually tended to decrease (e.g. Fig. 11 (centre)). This relationship can be explored in further research to define quantitatively the onset and end of the droughts.

Table 3 presents a summary of the duration of the selected droughts. It was observed that although the number of drought paths increases from the combination_1 (Fig. 10 (left)) to the combination_3 (Fig. 10 (right)), in terms of the most severe droughts, the durations remain almost similar (Table 3 (column 2 and 4) and Fig. 11 (bottom)). More drought tracks were identified in the first part of the year in combination_3. If the parameters c and d that control the distance between centroids are more flexible, i.e. consider longer distances, drought tracks of the second part of the year are more likely to join those of the first part of the next year, as occurs in combination_2. In the combination_2, the drought paths showed the longest durations (Table 3 (column 3) and Fig. 11 (bottom, centre)).

In all the selected droughts (Figs. 11 and A7 to A12), it was observed that consecutive clusters in time overlap considerably, which suggests that the spatial extent after reaching a considerable size, it remains in
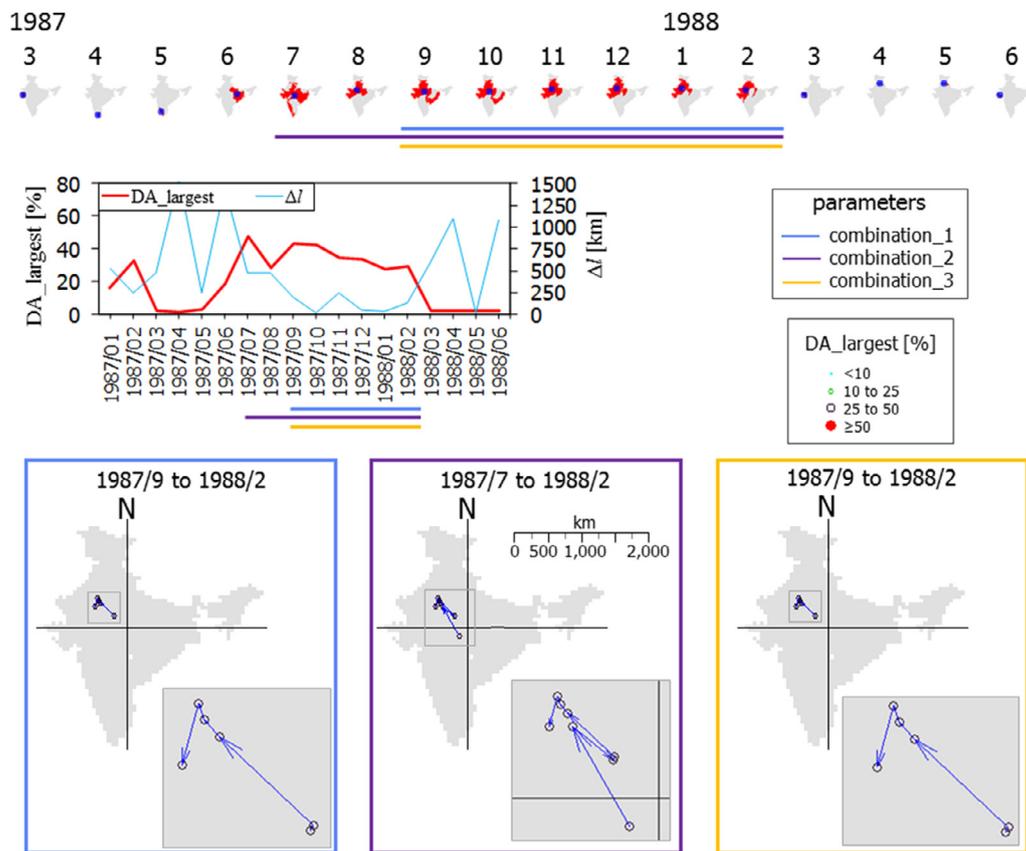
Fig. 11. Results for the drought of 1987. Largest clusters and centroids are indicated from 1987/3 to 1988/6 (top). Area of largest cluster (DA_largest) and distance between consecutive clusters in time (Δl) are displayed from 1987/1 to 1988/6 (centre). The drought duration is pointed out schematically with a horizontal line for each combination of parameters. Drought tracks calculated with the three combinations of parameters are also presented (bottom). Spatial drought extent is schematised by four symbols pointing out the size of area. The origin of the axes is placed in the centre of the country. Arrows point out the direction of each track segment. Insets show zoomed-in views.

Table 3
Duration of selected droughts calculated with three combinations of parameters. In parentheses, the period is indicated.

Drought	Duration [months]		
	Combination_1 $a = 50, b = 50, c = 50, d = 50$	Combination_2 $a = 40, b = 50, c = 70, d = 80$	Combination_3 $a = 30, b = 70, c = 90, d = 50$
1	6 (1905/7 to 1905/12)	12 (1905/6 to 1906/5)	6 (1905/7 to 1905/12)
2	5 (1942/10 to 1943/2)	6 (1942/10 to 1943/2)	5 (1942/10 to 1943/2)
3	6 (1965/7 to 1965/12)	22 (1965/5 to 1967/2)	6 (1965/7 to 1965/12)
4	3 (1972/8 to 1972/10)	16 (1972/4 to 1973/7)	3 (1972/8 to 1972/10)
5	6 (1987/9 to 1988/2)	8 (1987/7 to 1988/2)	6 (1987/9 to 1988/2)
6	5 (2000/8 to 2000/12)	11 (2000/8 to 2001/6)	6 (2000/7 to 2000/12)
7	5 (2002/8 to 2002/12)	12 (2002/4 to 2003/3)	6 (2002/8 to 2003/1)

the same region. This presence of large drought areas in the same region over time may explain the severity of drought events in those droughts. There was no predominant pathway followed by droughts in those years. In terms of spatial extent, 2000 and 2002 events were the largest as shown in Figs. A11 and A12, respectively. The drought with the longest duration was that of 1965 (Table 3), which is consistent with the reported in (Guha-Sapir, 2018).

4. Discussion

4.1. Drought indicator and areas

In the presented version of the tracking method, we used a unique threshold over the drought indicator to indicate drought and non-drought conditions in each grid cell (1 s and 0 s). This threshold is one of the most common used in drought studies when considering stan-

dardised drought indicators. SPEI was applied in this research, but it is possible to use any other, including threshold approach (Wanders et al., 2010), with the condition of being spatially distributed. The effects of other drought indicator thresholds over the cluster size were not assessed because the scope of this study was limited to testing the drought tracking algorithm.

On the other hand, the clustering algorithm used in this study assumes that all cell values in the space domain are homogeneous. To ensure that this assumption is correct, it is recommended the selection of a drought indicator that uses a normalization procedure into its calculation. In addition, our clustering approach is based only on drought indicator values and does not consider others aspects that can influence the delimitation of the spatial extent of drought, such as topography, land use, and climate regions. In further studies, it is recommended to incorporate other elements to make the clustering method more general. Another way of considering the factors mentioned above, without

modifying/changing the clustering algorithm, is the use of a drought indicator that takes into account variables such as soil moisture or runoff.

4.2. Drought tracking method

S-TRACK algorithm is an extension of the Contiguous Drought Area (CDA) analysis of Corzo Perez et al. (2011). This drought tracking algorithm was firstly introduced in Diaz et al. (2018) and further developed in this research. The current version of S-TRACK focuses on the largest drought areas. In this way, areas with a considerable territorial extent are identified. We are aware that smaller, intense droughts would not be captured by this tracking algorithm. Also, that mild droughts over large areas obtained by the algorithm would overshadow smaller, intense droughts.

Although S-TRACK makes use of CDA analysis for the extraction of drought clusters, other algorithms used for the same purpose can also be considered. These algorithms include the recursion-based approach (Andreadis et al., 2005; Herrera-Estrada et al., 2017; Lloyd-Hughes, 2012; Sheffield et al., 2009), and variations of the connected-component labelling approach (Van Huijgevoort et al., 2013; Vernieuwe et al., 2019). The composition of drought clusters extracted with any of these algorithms should be similar. The main difference between the algorithms is in the computational efficiency and processing time, which is an important element to consider when processing a large amount of data. In this sense, algorithms based on connected-component labelling are considered to be more efficient (He et al., 2009).

To connect two consecutive clusters in time and ensure that they are not far in space, the length between centroids of the clusters is taken into account, similar to Herrera-Estrada et al. (2017) and Zhou et al. (2019). The degree of the overlap between these two clusters can be another way to handle the connection between them. Yet another, and more comprehensive way of joining clusters in time, is through the use of the CDA approach but extended to the time domain, i.e. to connect 26 nearest neighbour cells, a forming a cube in space-time domain, as shown in Corzo Perez et al. (2011), Lloyd-Hughes (2012), and Herrera-Estrada et al. (2017).

In cases when more than one drought track occurs at the same time, the algorithm will aim to identify the one that is composed of the largest areas. In its current version, the algorithm neither detects simultaneous drought tracks nor merges the areas of the same time step into a single one.

In this research, we compared the area of all clusters and the area of the largest one in each time step, to see if the presence of more than one large area is predominant or not. We found that difference between DA_total − DA_largest was, in most of the cases, close to zero (Fig. 6). This difference between DA_total and DA_largest indicates that the size of the area of the largest cluster is very similar to the total one. Based on the latter, it is assumed that the presence of more than one large cluster at the same time step, is not dominant. Then the research was focused on testing the tracking algorithm, without considering the effect of the presence of more than one simultaneous drought track.

If the presence of more than one consecutive drought track is suspected, an option to perform this algorithm is to carry our tracking for different sub-regions of the study area (analyse it by parts) and then superimpose the drought tracks. In this way, one would expect to identify more than one track, if any. In future versions of the tracking algorithm, it is recommended to include the identification of more than one simultaneous drought tracks.

The use of CDA approach can retrieve areas with "islands" of non-drought cells (0 s). In this research, we do not consider the possible effects of these holes over the drought tracks construction (the centroid could be located in one of these holes). We assume that centroid is a good spot to locate the contiguous drought area.

In largest clusters, the centroid approaches the centroid of the analysis region. This is an expected outcome because if the cluster covers the entire region of analysis, the centroid will be similar to that one

of the region. In our case, the maximum DA_largest was 70.7%, therefore in the period of analysis, no cluster covered the entire territory. In addition, two simultaneous large clusters are not expected.

Although the drought tracks that occurred near the boundaries of the domain could not be considered appropriately, i.e. the tracks could be miscalculated, it is assumed that these boundary tracks do not significantly impact the region. To improve the calculation in such cases, it is recommended to increase the size of the analysed region.

5. Summary and conclusions

In this study, a method that allows the construction of drought tracks in space is introduced. The onset and end of drought paths (combination of linked drought tracks) are used to compute the drought duration. The information obtained during the path calculation is employed to compute the severity, as well as the onset and end location, direction, and rotation. All these features have been identified as drought characteristics and are framed within the DDRASTIC-spatial methodology, also presented in this paper. Outputs of the tracking algorithm S-TRACK and the method for drought characterisation DDRASTIC-spatial help to describe the dynamics of droughts.

S-TRACK has four parameters. Parameters *a* and *b* control the size of the cluster (area) to be included in the drought tracks. Parameters *c* and *d* limit the distances between consecutive clusters in time (Sect. 2.1). In this paper, S-TRACK is used to construct the drought tracks in space.

From the application of S-TRACK, some key findings are presented:

- The number of drought paths, duration, and severity are more sensitive to the change of the parameter that limits the minimum drought area (parameter *a*) (Sect. 2.1).
- If the duration of the drought paths increases, severity does not necessarily do so, because the longer the duration, the areas that make up the path tend to be smaller (Sect. 3.2).
- To obtain drought paths with longer durations, it is important to be flexible with the parameters that control the distance between areas (parameters *c* and *d*), i.e. to consider larger distances.

The outcome of the approach presented in this paper is relevant for (i) drought forecasting, i.e. drought tracks can help to predict how drought moves over a particular region, and (ii) for improving knowledge on drought-generating processes. The first item is more for operational purposes (short term) and the second item for scientific research (long term).

Regarding the improvement of knowledge on drought-generating processes, i.e. the interaction between climate and land surface characteristics, a new drought characteristic is introduced in this research, the rotation (Sect. 2.2). This feature is used in the study of other weather-related phenomena such as cyclones because it helps in the description/identification of forcing mechanisms behind their spatial development (e.g. Chavas et al., 2017; Rahman et al., 2018). We are of the opinion that this drought characteristic can also help in the identification and description of climate and land surface control factors that drive the spatial behaviour of droughts.

For the considered case study in India, we found that consecutive clusters in time overlap considerably in the droughts selected (1905, 1942, 1965, 1972, 1987, 2000, and 2002), which suggests that the spatial extent of drought, after reaching a considerable size, remains in the same region. This presence of large drought areas in the same region over time may explain the severity of droughts in those years. There is no predominant pathway followed by droughts in those years. In terms of spatial extent, 2000 and 2002 events are the largest. The drought with the longest duration is that of 1965. A paper was prepared where the parameters of the tracking algorithm were calibrated based on the information of droughts reported. In that document, a description of droughts is presented based on the drought paths and characteristics (Diaz et al., 2019).

Further research is aimed at trying to develop an approach to predict the subsequent development of tracks identified by S-TRACK. These progress of these developments and other aspects of the study can be found at www.researchgate.net/project/STAND-Spatio-Temporal-ANalysis-of-Drought.

Author contribution statement

VD, GACP, HvL, DS, and EAV contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

Appendix A

Fig. A1. Area of the largest cluster (DA_largest) and distances between consecutive centroids in time (Δl) for the period 1901-2013.

a = 50

b = 50

c \ d	50	60	70	80	90
50	3.9	4.5	5.1	5.7	6.2
60	3.9	4.5	5.1	5.7	6.2
70	3.9	4.5	5.1	5.7	6.2
80	3.9	4.5	5.1	5.7	6.2
90	3.9	4.5	5.1	5.7	6.2

b = 70

c \ d	50	60	70	80	90
50	3.9	4.6	5	5.2	5.3
60	3.9	4.5	4.9	5.2	5.2
70	3.9	4.6	5.1	5.4	5.5
80	3.9	4.6	5.2	5.7	5.9
90	3.9	4.6	5.3	6	6.2

b = 90

c \ d	50	60	70	80	90
50	3.9	4.2	4.2	4.3	4.2
60	4.2	4.5	4.6	4.7	4.7
70	4.4	4.8	5.1	5.2	5.2
80	4.6	5.1	5.6	5.7	5.7
90	4.9	5.4	6	6.2	6.2

a = 40

b = 50

c \ d	50	60	70	80	90
50	3.9	4.5	5.1	5.8	6.2
60	4	4.4	5.1	5.8	6.2
70	4	4.5	5.2	6	6.4
80	4	4.6	5.4	6.2	6.6
90	4.1	4.7	5.4	6.3	6.7

b = 70

c \ d	50	60	70	80	90
50	3.9	4.5	4.9	5.1	5.2
60	4	4.4	4.9	5.2	5.2
70	4	4.6	5.2	5.6	5.7
80	4.1	4.9	5.6	6.2	6.4
90	4.1	4.9	5.6	6.4	6.7

b = 90

c \ d	50	60	70	80	90
50	3.9	4.1	4.2	4.2	4.2
60	4.2	4.4	4.6	4.6	4.6
70	4.5	4.9	5.2	5.3	5.3
80	5.1	5.6	6.1	6.2	6.3
90	5.3	5.8	6.4	6.6	6.7

Legend: 3.8 – 7.5

a = 30

b = 50

c \ d	50	60	70	80	90
50	3.9	4.4	5.1	5.8	6.4
60	3.9	4.4	5.1	5.9	6.5
70	3.8	4.4	5.2	6.1	6.8
80	3.9	4.7	5.5	6.7	7.3
90	3.9	4.7	5.6	6.8	7.5

b = 70

c \ d	50	60	70	80	90
50	3.9	4.4	4.7	5	5.1
60	3.9	4.4	4.8	5.1	5.2
70	4	4.6	5.2	5.7	5.9
80	4.2	5.1	5.9	6.7	6.9
90	4.3	5.2	6.1	7.1	7.5

b = 90

c \ d	50	60	70	80	90
50	3.9	4	4.1	4.2	4.2
60	4.2	4.4	4.5	4.6	4.6
70	4.5	4.9	5.2	5.4	5.4
80	5.3	5.8	6.4	6.7	6.7
90	5.7	6.2	7.1	7.5	7.5

Fig. A2. Average duration (months) of drought paths obtained with different combinations of parameters.

a = 50

b = 50

c \ d	50	60	70	80	90
50	225.4	219.4	222.3	219.1	215
60	225.4	219.4	222.3	219.1	215
70	225.4	219.4	222.3	219.1	215
80	225.4	219.4	222.3	219.1	215
90	225.4	219.4	222.3	219.1	215

b = 70

c \ d	50	60	70	80	90
50	225.4	228.8	237.1	240.2	241
60	217.4	219.4	229.1	231.9	231.1
70	211	212.6	222.3	225.8	224.8
80	204.3	203.8	215.8	219.1	217.9
90	200.5	200.6	210.1	215.1	215

b = 90

c \ d	50	60	70	80	90
50	225.4	234.7	249.8	253.7	254.4
60	211.6	219.4	232.9	237.2	238.1
70	205.1	212.9	222.3	228	228.5
80	199.5	206.1	214	219.1	219.6
90	198	203.5	208.9	214.4	215

a = 40

b = 50

c \ d	50	60	70	80	90
50	197.7	195.7	198.8	197.9	192.4
60	197.5	193.6	197.1	196	190.6
70	193.7	190.6	194.2	192.8	187
80	192	189.7	193.4	191	185.3
90	191.7	189.3	192.7	190.4	184.7

b = 70

c \ d	50	60	70	80	90
50	197.7	200.6	208.5	211.3	213.5
60	192.4	193.6	202.2	204.1	205.4
70	184.9	186.7	194.2	196.2	194.8
80	179.1	180	189.7	191	189.7
90	175.1	176.4	183.7	185.3	184.7

b = 90

c \ d	50	60	70	80	90
50	197.7	205.5	218.8	220.5	222.9
60	187.4	193.6	204.3	206.2	208.5
70	181.8	188.4	194.2	197.6	197.9
80	176.5	182.4	188.2	191	191.3
90	173.3	178.3	181.5	184.4	184.7

Color scale: 150.5 – 254.4

a = 30

b = 50

c \ d	50	60	70	80	90
50	176.1	176.7	179.3	179.2	176.9
60	174.1	174.1	177.1	177.1	174.5
70	167.7	167.9	171.5	171.7	168.2
80	165.7	167.5	171.5	168.4	164.5
90	156.2	160.3	163.2	160.4	157.5

b = 70

c \ d	50	60	70	80	90
50	176.1	179.1	186	189.1	191.4
60	171.6	174.1	181.3	183.6	185.7
70	164	166	171.5	174.5	173.8
80	160	160.8	167.9	168.4	167.3
90	152.1	154.1	157.9	157.1	157.5

b = 90

c \ d	50	60	70	80	90
50	176.1	182.7	195.2	197.7	200.2
60	168.9	174.1	184	186.7	189
70	162.6	167.2	171.5	175.4	175.6
80	157.5	161.6	165	168.4	168.6
90	150.5	153.8	155	157.2	157.5

Fig. A3. Average severity of drought paths obtained with different combinations of parameters. Severity is expressed as the ratio between the total sum of areas (in percentage) and duration (months).

a = 50

b = 50

		d				
		50	60	70	80	90
	50	0	7	0	1	1
	60	0	7	0	1	1
c	70	0	7	0	1	1
	80	0	7	0	1	1
	90	0	7	0	1	1

b = 70

		d				
		50	60	70	80	90
	50	0	1	0	1	1
	60	0	7	7	1	1
c	70	0	1	0	1	1
	80	0	1	0	1	1
	90	0	1	1	1	1

b = 90

		d				
		50	60	70	80	90
	50	0	0	0	0	0
	60	0	7	0	0	0
c	70	0	0	0	0	0
	80	1	1	1	1	1
	90	1	1	1	1	1

a = 40

b = 50

		d				
		50	60	70	80	90
	50	1	1	1	1	1
	60	7	7	7	1	1
c	70	0	7	1	1	1
	80	7	7	7	1	1
	90	7	7	7	1	1

b = 70

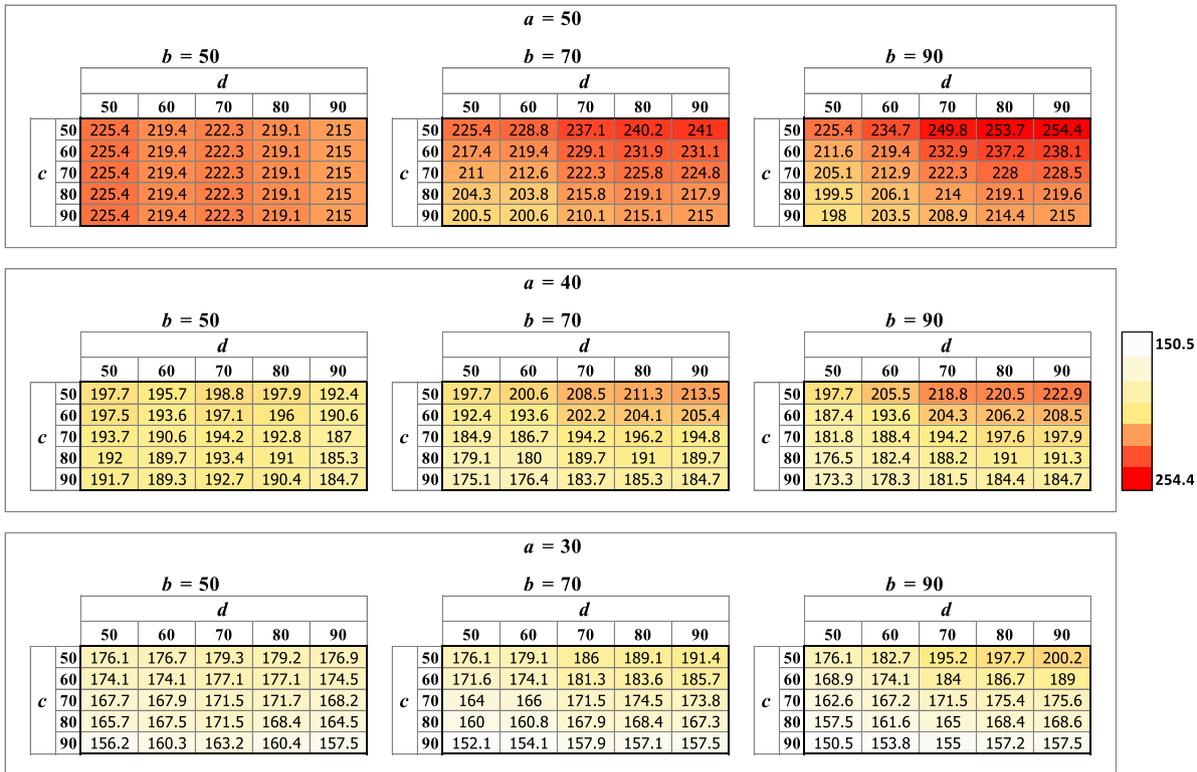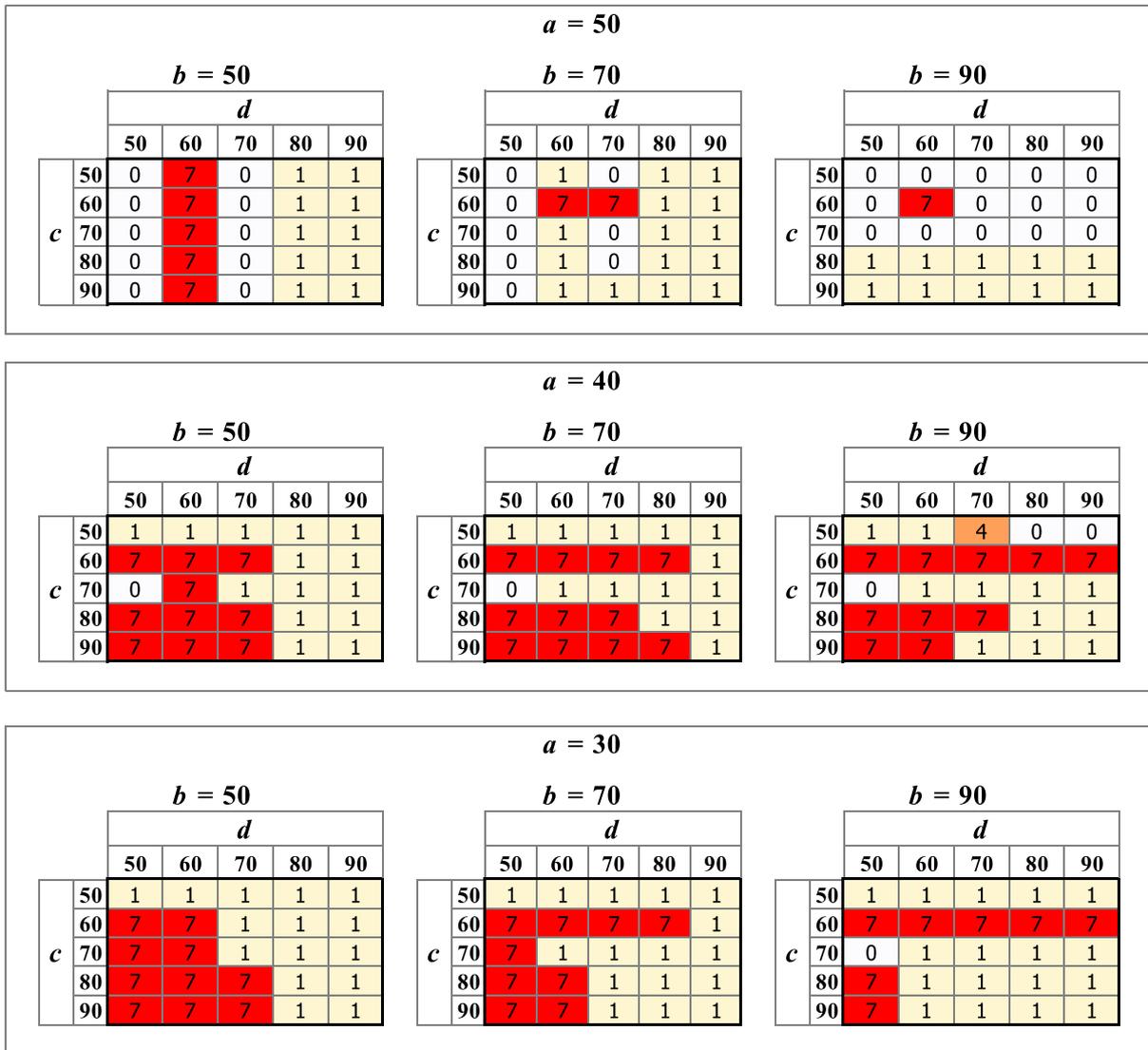|		d				
		50	60	70	80	90
	50	1	1	1	1	1
	60	7	7	7	7	1
c	70	0	1	1	1	1
	80	7	7	7	1	1
	90	7	7	7	7	1

b = 90

		d				
		50	60	70	80	90
	50	1	1	4	0	0
	60	7	7	7	7	7
c	70	0	1	1	1	1
	80	7	7	7	1	1
	90	7	7	1	1	1

a = 30

b = 50

		d				
		50	60	70	80	90
	50	1	1	1	1	1
	60	7	7	1	1	1
c	70	7	7	1	1	1
	80	7	7	7	1	1
	90	7	7	7	1	1

b = 70

		d				
		50	60	70	80	90
	50	1	1	1	1	1
	60	7	7	7	7	1
c	70	7	1	1	1	1
	80	7	7	1	1	1
	90	7	7	1	1	1

b = 90

		d				
		50	60	70	80	90
	50	1	1	1	1	1
	60	7	7	7	7	7
c	70	0	1	1	1	1
	80	7	1	1	1	1
	90	7	1	1	1	1

Fig. A4. Mode of onset location of drought paths obtained with different combinations of parameters. Locations: centre (0), east (1), northwest (4), and south (7).
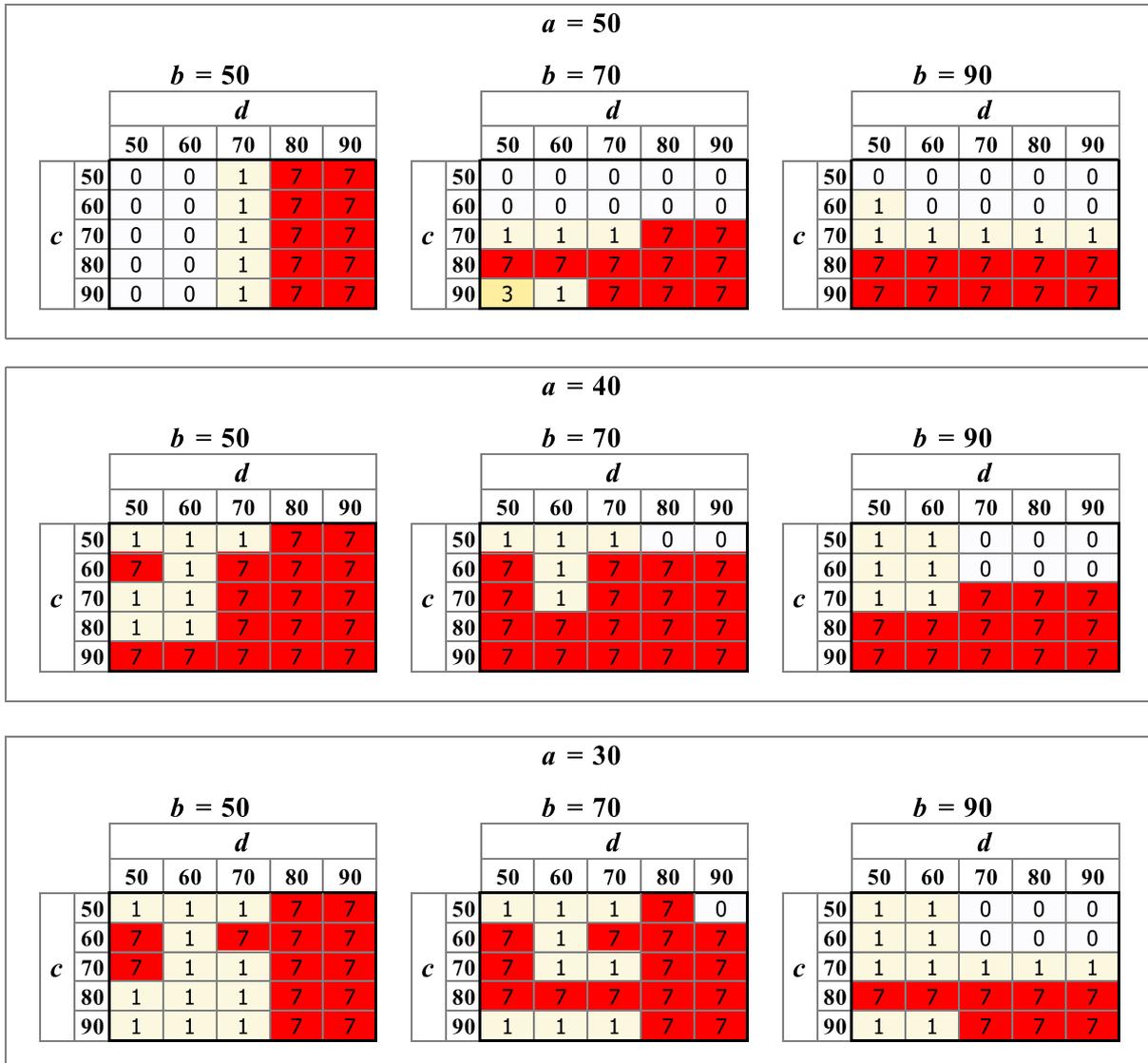
a = 50

b = 50

c \ d	50	60	70	80	90
50	0	0	1	7	7
60	0	0	1	7	7
70	0	0	1	7	7
80	0	0	1	7	7
90	0	0	1	7	7

b = 70

c \ d	50	60	70	80	90
50	0	0	0	0	0
60	0	0	0	0	0
70	1	1	1	7	7
80	7	7	7	7	7
90	3	1	7	7	7

b = 90

c \ d	50	60	70	80	90
50	0	0	0	0	0
60	1	0	0	0	0
70	1	1	1	1	1
80	7	7	7	7	7
90	7	7	7	7	7

a = 40

b = 50

c \ d	50	60	70	80	90
50	1	1	1	7	7
60	7	1	7	7	7
70	1	1	7	7	7
80	1	1	7	7	7
90	7	7	7	7	7

b = 70

c \ d	50	60	70	80	90
50	1	1	1	0	0
60	7	1	7	7	7
70	7	1	7	7	7
80	7	7	7	7	7
90	7	7	7	7	7

b = 90

c \ d	50	60	70	80	90
50	1	1	0	0	0
60	1	1	0	0	0
70	1	1	7	7	7
80	7	7	7	7	7
90	7	7	7	7	7

a = 30

b = 50

c \ d	50	60	70	80	90
50	1	1	1	7	7
60	7	1	7	7	7
70	7	1	1	7	7
80	1	1	1	7	7
90	1	1	1	7	7

b = 70

c \ d	50	60	70	80	90
50	1	1	1	7	0
60	7	1	7	7	7
70	7	1	1	7	7
80	7	7	7	7	7
90	1	1	1	7	7

b = 90

c \ d	50	60	70	80	90
50	1	1	0	0	0
60	1	1	0	0	0
70	1	1	1	1	1
80	7	7	7	7	7
90	1	1	7	7	7

Fig. A5. Mode of end location of drought paths obtained with different combinations of parameters. Locations: centre (0), east (1), north (3), and south (7).
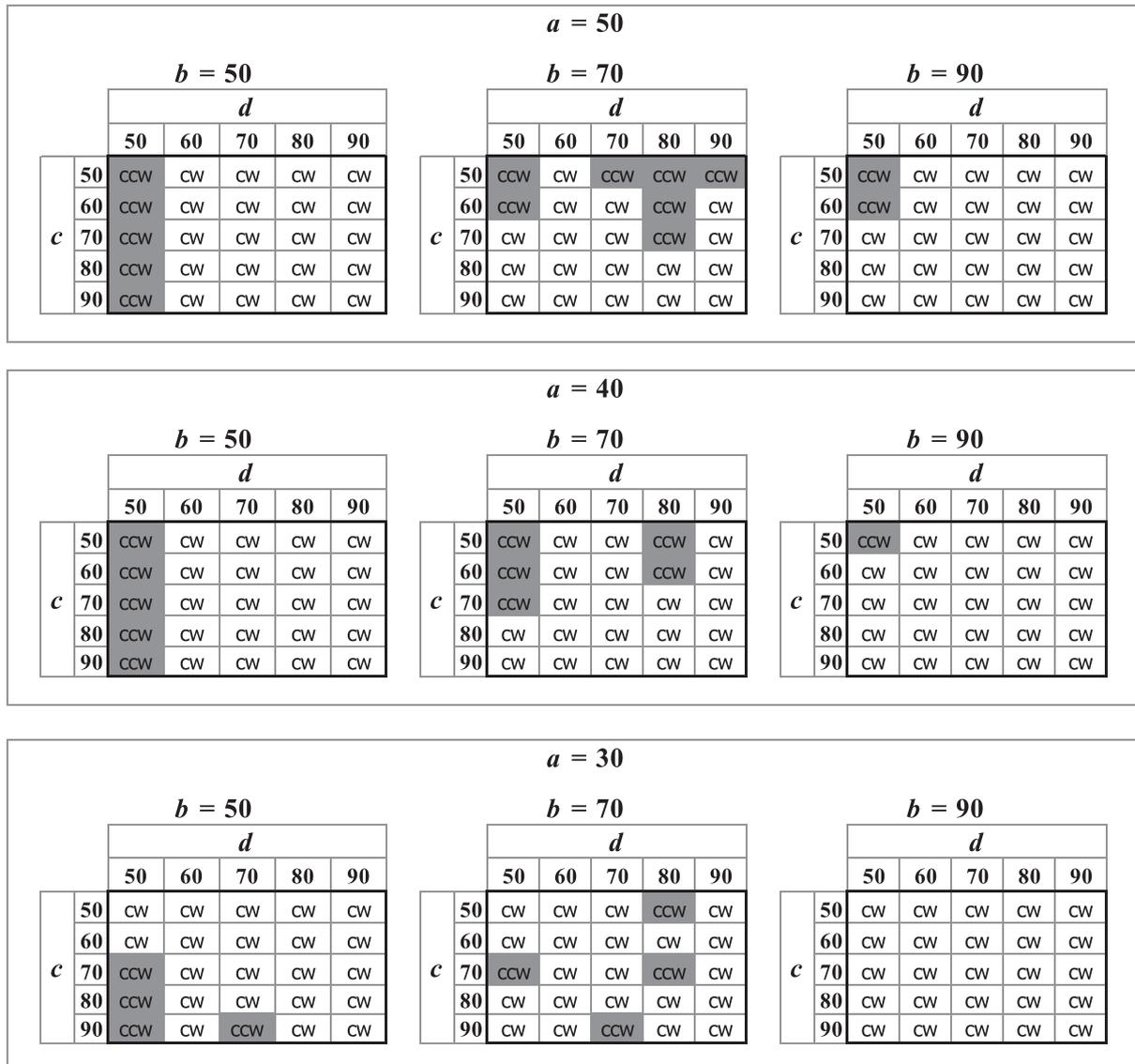
a = 50

b = 50

c \ d	50	60	70	80	90
50	CCW	CW	CW	CW	CW
60	CCW	CW	CW	CW	CW
70	CCW	CW	CW	CW	CW
80	CCW	CW	CW	CW	CW
90	CCW	CW	CW	CW	CW

b = 70

c \ d	50	60	70	80	90
50	CCW	CW	CCW	CCW	CCW
60	CCW	CW	CW	CCW	CW
70	CW	CW	CW	CCW	CW
80	CW	CW	CW	CW	CW
90	CW	CW	CW	CW	CW

b = 90

c \ d	50	60	70	80	90
50	CCW	CW	CW	CW	CW
60	CCW	CW	CW	CW	CW
70	CW	CW	CW	CW	CW
80	CW	CW	CW	CW	CW
90	CW	CW	CW	CW	CW

a = 40

b = 50

c \ d	50	60	70	80	90
50	CCW	CW	CW	CW	CW
60	CCW	CW	CW	CW	CW
70	CCW	CW	CW	CW	CW
80	CCW	CW	CW	CW	CW
90	CCW	CW	CW	CW	CW

b = 70

c \ d	50	60	70	80	90
50	CCW	CW	CW	CCW	CW
60	CCW	CW	CW	CCW	CW
70	CCW	CW	CW	CW	CW
80	CW	CW	CW	CW	CW
90	CW	CW	CW	CW	CW

b = 90

c \ d	50	60	70	80	90
50	CCW	CW	CW	CW	CW
60	CW	CW	CW	CW	CW
70	CW	CW	CW	CW	CW
80	CW	CW	CW	CW	CW
90	CW	CW	CW	CW	CW

a = 30

b = 50

c \ d	50	60	70	80	90
50	CW	CW	CW	CW	CW
60	CW	CW	CW	CW	CW
70	CCW	CW	CW	CW	CW
80	CCW	CW	CW	CW	CW
90	CCW	CW	CCW	CW	CW

b = 70

c \ d	50	60	70	80	90
50	CW	CW	CW	CCW	CW
60	CW	CW	CW	CW	CW
70	CCW	CW	CW	CCW	CW
80	CW	CW	CW	CW	CW
90	CW	CW	CCW	CW	CW

b = 90

c \ d	50	60	70	80	90
50	CW	CW	CW	CW	CW
60	CW	CW	CW	CW	CW
70	CW	CW	CW	CW	CW
80	CW	CW	CW	CW	CW
90	CW	CW	CW	CW	CW

Fig. A6. Mode of rotation of drought paths obtained with different combinations of parameters. Rotation is indicated by ccw and cw that sand for mostly counter-clockwise, and mostly clockwise, respectively.

Fig. A7. Results of the drought of 1905. Largest clusters and centroids are indicated from 1905/3 to 1906/6 (top). Area of largest cluster (DA_largest) and distance between consecutive clusters in time (Δl) are displayed from 1905/1 to 1906/6 (centre). The drought duration is pointed out schematically with a horizontal line for each combination of parameters. Drought tracks calculated with the three combinations of parameters are also presented (bottom). Spatial drought extent is schematised by four symbols pointing out the size of area. The origin of the axes is placed in the centre of the country. Arrows point out the direction of each track segment. Insets show zoomed-in views.
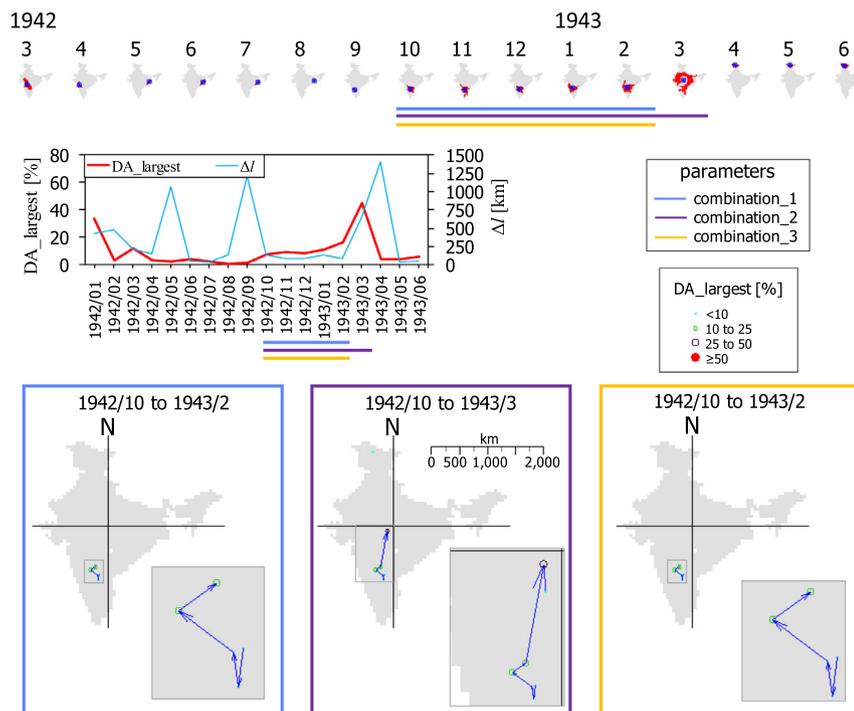
Fig. A8. Same as Fig. A7 but for the drought of 1942.
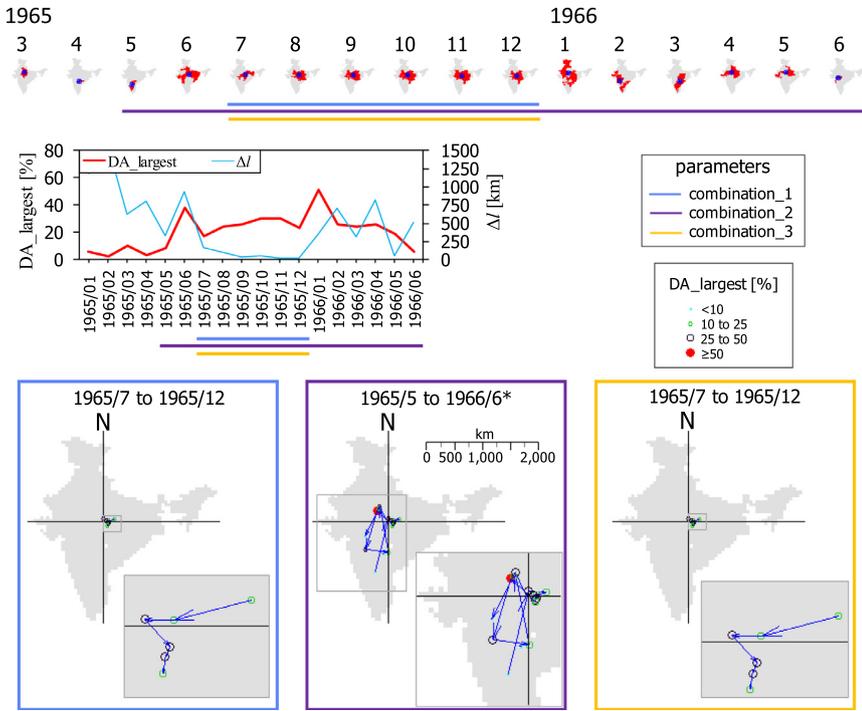
Fig. A9. Same as Fig. A7 but for the drought of 1965. *In the figure only it is shown the tracks until 1966/6 but they end in 1967/2.
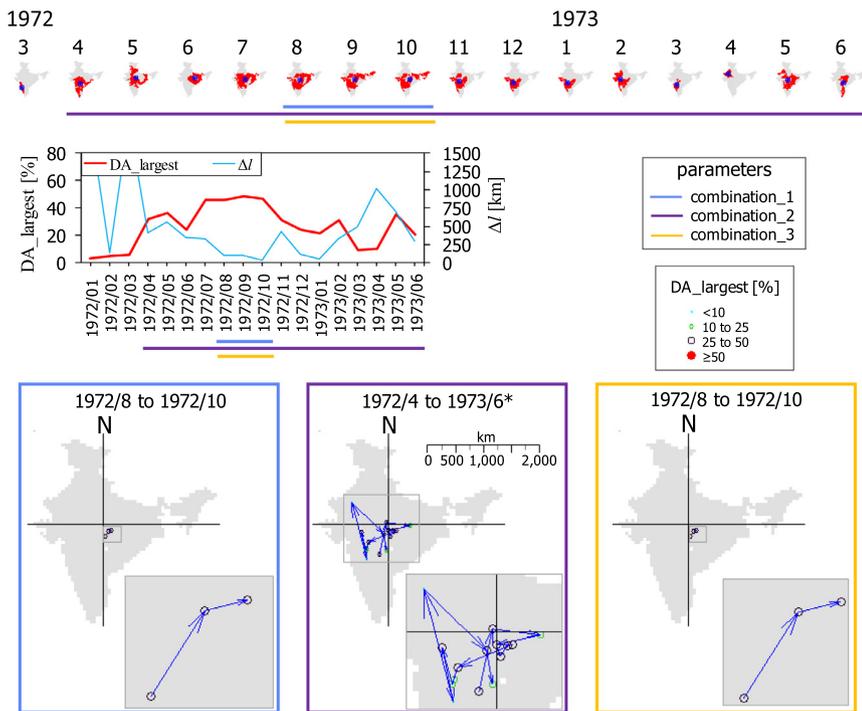
Fig. A10. Same as Fig. A7 but for the drought of 1972. *In the figure only it is shown the tracks until 1973/6 but they end in 1973/7.
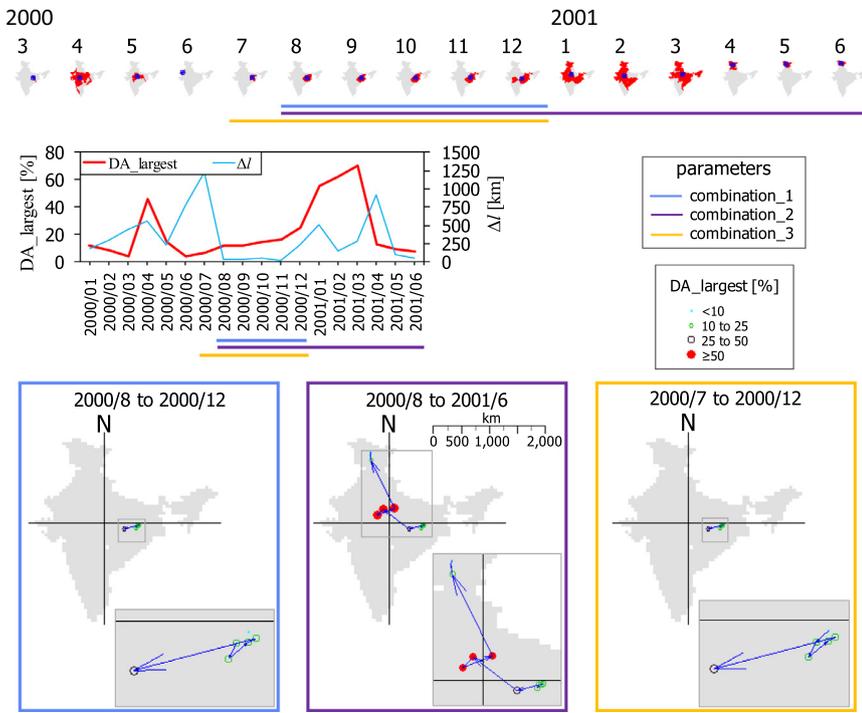
Fig. A11. Same as Fig. A7 but for the drought of 2000.
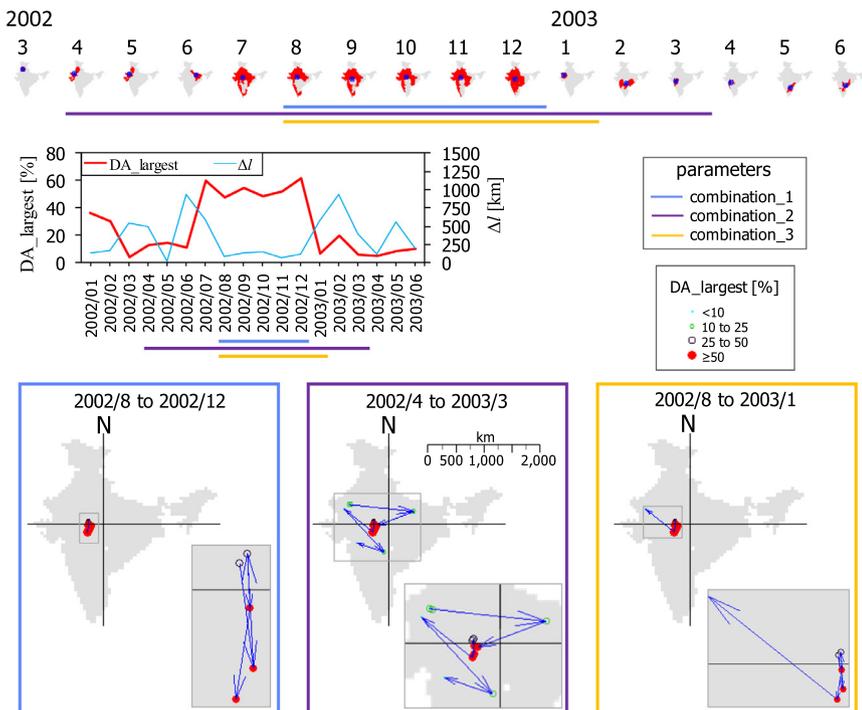
Fig. A12. Same as Fig. A7 but for the drought of 2002.

Declaration of Competing Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2020.103512.

References

Andreadis, K.M., Clark, E.A., Wood, A.W., Hamlet, A.F., Lettenmaier, D.P., 2005. Twentieth-century drought in the conterminous United States. J. Hydrometeorol. 6 (6), 985–1001. https://doi.org/10.1175/JHM450.1.

Beguería, S., Vicente-Serrano, S.M., Reig, F., Latorre, B., 2014. Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. International Journal of Climatology 34 (10), 3001–3023. https://doi.org/10.1002/joc.3887.

Below, R., Grover-Kopec, E., Dilley, M., 2007. Documenting drought-related disasters: a global reassessment. J. Environ. Dev. 16 (3), 328–344. https://doi.org/10.1177/1070496507306222.

Chavas, D.R., Reed, K.A., Knaff, J.A., 2017. Physical understanding of the tropical cyclone wind-pressure relationship. Nat. Commun. 8 (1), 1360. https://doi.org/10.1038/s41467-017-01546-9.

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

Diaz, V., Corzo, G., Lanen, H.A.J.V., Solomatine, D.P., 2019a. 4 - Spatiotemporal drought analysis at country scale through the application of the STAND toolbox. In: Corzo, G., Varouchakis, E.A. (Eds.), Spatiotemporal Analysis of Extreme Hydrological Events. Elsevier, pp. 77–93. https://doi.org/10.1016/B978-0-12-811689-0.00004-5.

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

Guha-Sapir, D. (2018). EM-DAT: the emergency events database - Université catholique de Louvain (UCL) - CRED. Retrieved from emdat.be

Diaz, V., Corzo Perez, G.A., Van Lanen, H.A.J., Solomatine, D., Varouchakis, E.A., 2019b. Characterisation of the dynamics of past droughts. Sci. Total Environ., 134588 https://doi.org/10.1016/j.scitotenv.2019.134588.

Hao, Z., Yuan, X., Xia, Y., Hao, F., Singh, V.P., 2017. An overview of drought monitoring and prediction systems at regional and global scales. Bull. Am. Meteorol. Soc. 98 (9), 1879–1896. https://doi.org/10.1175/BAMS-D-15-00149.1.

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

He, L., Chao, Y., Suzuki, K., Wu, K., 2009. Fast Connected-component labeling. Pattern Recogn. 42 (9), 1977–1987. https://doi.org/10.1016/j.patcog.2008.10.013.

Herrera-Estrada, J.E., Satoh, Y., Sheffield, J., 2017. Spatio-temporal dynamics of global drought. Geophys. Res. Lett. 2254–2263. https://doi.org/10.1002/2016GL071768.

Lloyd-Hughes, B., 2012. A spatio-temporal structure-based approach to drought characterisation. Int. J. Climatol. 32 (3), 406–418. https://doi.org/10.1002/joc.2280.

Mckee, T.B., Doesken, N.J., Kleist, J., 1993. The relationship of drought frequency and duration to time scales. AMS 8th Conference on Applied Climatology January, 179–184. https://doi.org/10.1002/10490403.

Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391 (1–2), 202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012.

Mishra, A.K., Singh, V.P., 2011. Drought modeling - a review. J. Hydrol. 403 (1–2), 157–175. https://doi.org/10.1016/j.jhydrol.2011.03.049.

Pannell, D.J., 1997. Sensitivity analysis of normative economic models: theoretical framework and practical strategies. Agric. Econ. 16 (2), 139–152. https://doi.org/10.1016/S0169-5150(96)01217-0.

Rahman, S.M., Yang, R., Di, L., 2018. Clustering Indian ocean tropical cyclone tracks by the standard deviational ellipse. Climate 6 (2). https://doi.org/10.3390/cli6020039.

Sepulcre-Cantó, G., Horion, S., Singleton, A., Carrao, H., Vogt, J., 2012. Development of a combined drought indicator to detect agricultural drought in Europe. Nat. Hazard. Earth Syst. Sci. 12 (11), 3519–3531. https://doi.org/10.5194/nhess-12-3519-2012.

Sheffield, J., Wood, E.F., 2011. Drought: Past problems and future scenarios. P. Earthscan, Ed..

Sheffield, J., Andreadis, K.M., Wood, E.F., Lettenmaier, D.P., 2009. Global and continental drought in the second half of the twentieth century: Severity-area-duration analysis and temporal variability of large-scale events. J. Climate 22 (8), 1962–1981. https://doi.org/10.1175/2008JCLI2722.1.

Tallaksen, L.M., Van Lanen, H.A.J., 2004. Hydrological drought - processes and estimation methods for streamflow and groundwater. In: Tallaksen, L.M., Van Lanen, H.A.J. (Eds.). Developments in Water Sciences, 48. Elsevier B.V., The Netherlands.

Tallaksen, L.M., Hisdal, H., Lanen, H.A.J.V., 2009. Space-time modelling of catchment scale drought characteristics. J. Hydrol. 375 (3–4), 363–372. https://doi.org/10.1016/j.jhydrol.2009.06.032.

van Huijgevoort, M.H.J., Hazenberg, P., van Lanen, H., Teuling, A.J., Clark, A.J., Folwell, D.B., Uijlenhoet R, S., 2013. Global multimodel analysis of drought in runoff for the second half of the twentieth century. J. Hydrometeorol. 14 (5), 1535–1552. https://doi.org/10.1175/JHM-D-12-0186.1.

Van Loon, A.F., 2015. Hydrological drought explained. Wiley Interdiscip. Rev. 2 (4), 359–392. https://doi.org/10.1002/wat2.1085.

Vernieuwe, H., De Baets, B., Verhoest, N.E.C., 2019. A mathematical morphology approach for a qualitative exploration of drought events in space and time. Int. J. Climatol. 1–14. https://doi.org/10.1002/joc.6226.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. Journal of Climate 23 (7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1.

Wanders, N., van Lanen, H.A.J., van Loon, A.F., 2010. Indicators for drought characterization on a global scale. WATCH Technical Report No 24. Retrieved from http://www.eu-watch.org/media/default.aspx/emma/org/10646416/WATCH+Technical+Report+Number+24+Indicators+For+Drought+Characterization+on+a+Global+Scale.pdf.

Wilhite, D.A., 2000. Drought as a natural hazard: concepts and definitions. DROUGHT, A Global Assessment, Vol I and II. Routledge Hazards and Disasters Series.

World Meteorological Organization (WMO), 2006. Drought monitoring and early warning: concepts, progress and future challenges. progress and future challenges. WMO-No, WMO-No. 1006. WMO, Geneva, Switzerland Retrieved from http://www.droughtmanagement.info/literature/WMO_drought_monitoring_early_warning_2006.pdf.

Zhou, H., Liu, Y., Liu, Y., 2019. An approach to tracking meteorological drought migration. Water Resour. Res. 55 (4), 3266–3284. https://doi.org/10.1029/2018WR023311.

World Meteorological Organization (WMO). 2012. Standardized Precipitation Index user guide. WMO-No.1090. Geneva, Switzerland. Retrieved from http://library.wmo.int/pmb_ged/wmo_1090_en.pdf