A spatio-temporal structure-based approach to drought characterisation

Article

Accepted Version

A spatio-temporal structure based approach to drought characterisation

BENJAMIN LLOYD-HUGHES

Walker Institute, Department of Meteorology,

University of Reading, Earley Gate, Reading, RG6 6AR, U.K.

(Email: b.lloydhughes@reading.ac.uk; Tel: 0118 378 6592)

Submitted to the International Journal of Climatology

2 November 2009

Revision 1: 25 January 2010

Abstract

Drought characterisation is an intrinsically spatio-temporal problem. A limitation of previous approaches to characterisation is that they discard much of the spatio-temporal information by reducing events to a lower order subspace. To address this, an explicit 3-dimensional (longitude, latitude, time) structure based method is described in which drought events are defined by a spatially and temporarily coherent set of points displaying standardized precipitation below a given threshold. Geometric methods can then be used to measure similarity between individual drought structures. Groupings of these similarities provide an alternative to traditional methods for extracting recurrent space-time signals from geophysical data. The explicit consideration of structure encourages the construction of summary statistics which relate to the event geometry. Example measures considered are the event volume, centroid, and aspect ratio. The utility of a 3-dimensional approach is demonstrated by application to the analysis of European droughts (15^oW to 35^oE and 35^oN to 70^oN) for the period 1901-2006. Large scale structure is found to be abundant with 75 events identified lasting for more than 3 months and spanning at least $0.5 \times 10^6 km^2$. Near complete dissimilarity is seen between the individual drought structures and little or no regularity is found in the time evolution of even the most spatially similar drought events. The spatial distribution of the event centroids and the time evolution of the geographic cross-sectional areas strongly suggest that large area, sustained droughts result from the combination of multiple small area ($\sim 10^6 km^2$) short duration ($\sim$3 months) events. The small events are not found to occur independently in space. This leads to the hypothesis that local water feedbacks play an important role in the aggregation process.

Keywords: Drought, climatology, structure, classification, standardized precipitation index, SPI

1 Introduction

Drought has claimed over one million lives and $60 billion in losses globally since 1974 (UN 2008, MunichRe 2008). Modern water supply infrastructure can eliminate direct mortalities yet the societal impacts of water scarcity remain and cannot be overstated. A thorough understanding of the peril is essential for mitigating against the risk as it stands and for preparedness in the face of climate change.

The majority of notable high precipitation events are characterised by highly localised, short lived, heavy downbursts. The same is not true of the most notable drought events. These typically last for several months or even years and span thousands of square kilometres. Previous authors have considered the spatial and temporal nature of drought at regional (Livada & Assimakopoulos 2007, Vicente-Serrano 2006), continental (Briffa *et al.* 1994, Lloyd-Hughes & Saunders 2002, van der Schrier *et al.* 2006) and global scales (Sheffield *et al.* 2009). A limitation of these studies is that they reduce the full 3-dimensional space-time drought structure to a lower order subspace. The analysis is limited to the time evolution of fixed areal quantities, for example, the global area average; and/or the time dependent amplitude of fixed spatial patterns, such as loadings onto empirical orthogonal functions. The block neighbourhood used in the severity-area-duration (SAD) approach (Andreadis *et al.* 2005) is a true space-time representation; however, the focus on areal extent obscures much of the time evolution of the drought structure. Consideration of the space-time structure is fundamental to the understanding of the role of external and internal forcings on drought development. Improved appreciation of these drivers is crucial for the development of effective monitoring and early warning systems.

This paper describes a method for characterizing coupled space-time deficits from gridded precipitation data. The emphasis is placed on the explicit representation of the intrinsic 3-dimensional space-time deficits that constitute drought.

2 Methodology

2.1 Drought definition

The term 'drought' is frequently used to refer to the adverse impacts of the lack of precipitation rather than the lack of precipitation as a meteorological event (Smakhtin & Schipper 2008) and this can present difficulties with respect to event definition. In this paper, drought is defined in a strictly meteorological sense. Specifically, it is defined by negative values of the Standardized Precipitation Index (SPI) (McKee *et al.* 1995) at the 3-monthly time scale (SPI_3). This definition represents a good proxy for large scale stream flow drought in Europe (Lloyd-Hughes *et al.* 2009, Szalai & Szinell 2000) but since an objective definition of drought remains elusive it should be remembered that other definitions may be more appropriate for other applications.

2.2 Feature extraction

Drought data on a regular grid, either observed or modelled, can be represented as a set of maps of values at fixed geographical locations stacked sequentially in time order as a 3 dimensional array (henceforth data stack) with dimensions representing longitude, latitude, and time. A variety of techniques exist for the extraction of recurrent spatial patterns from such stacks of data *e.g.* principal component analysis (PCA), empirical orthogonal teleconnections, cluster analysis, *etc.* (see Hannachi *et al.* (2007) for a review). A common feature of these methods is invariance to the time ordering of the grids *i.e.* the results remain the same for any ordering of the maps within the stack. Similarly, a wide range of time series techniques are available to extract temporal patterns from the data (see von Storch & Zwiers (1999) or Wilks (1995)) *e.g.* t-mode PCA, Fourier decomposition, wavelet analysis, *etc.*. A common feature of these methods is restriction to a particular spatial domain, be this a fixed areal quantity or the amplitude of fixed spatial weights such as a principal component.

The choice of technique for the extraction of common space-time features is

limited. Methods commonly applied in the atmospheric sciences are restricted to principal oscillation patterns, extended empirical orthogonal functions (eEOF), and Hilbert space decomposition. Unfortunately, these methods will not reliably detect irregularly occurring features (Horel 1984) and interpretation of the resultant patterns is often difficult without prior knowledge of the signal to be retrieved. In light of these difficulties, this paper proposes a simple agglomerative technique to identify large scale coherent space-time drought events.

Andreadis *et al.* (2005) describe a spatial identification procedure in which all pixels that have a soil moisture (or runoff) percentile value below 20% are considered as being under drought. Those pixels are then classified into drought classes using a simple clustering algorithm. The first pixel under drought is assigned to the first class. Then, the 3×3 neighborhood of this pixel is searched for pixels under drought that are classified in the same drought cluster. This procedure is repeated until no pixels in the 3×3 neighborhood of the current pixel are under drought, and a new cluster is created for the next pixel below the drought threshold. A natural extension of this method is to extend the clustering from the 3×3 spatial domain to $3 \times 3 \times 3$ space-time domain. Such an extension is described below.

The aim of the new method is to extract coherent space-time structures from within the data. A starting point is to identify those cells within the stack of grids that are in drought. This is achieved by defining drought at a cell if the SPI_3 value is $\leq$ -1 (below one standard deviation below expected for the 3 monthly period at that time of year at that location). The next step is to locate any neighbouring cells that are also in drought. The definition of neighbour presents several possibilities. At a simple level, Figure 1 (a), neighbouring cells in two dimensions are taken to be those which share a common vertex. However, if the data contain gaps, as is common with observational data, then it is useful to extend this concept to cells that share a common vertex at some radius R cells away. The concept is illustrated in Figure 1 (b) for $R = 1$ where the two groups of cells on the left hand side (originally coloured orange and purple) are now considered to form a single unit (coloured orange). Coherency in three dimensions follows similarly by the consideration of common vertices within the data stack in the planes of cells above and below the

cell of interest.

It is important to note that this treatment of the data implicitly equates the length scales (longitude $\equiv$ latitude $\equiv$ time) of the individual cubes (voxels) of data that comprise the stack. The general case requires a separate scaling for each dimension (longitude, latitude, time) with radii (R_{lon}, R_{lat}, R_{time}). Here it is sufficient to consider R_{lon} and R_{lat} to be comparable and equal. The explicit time averaging applied in the construction of the SPI provides direct control over the temporal scaling and it is appropriate to set $R_{time} = 0$ and only consider immediate neighbours in time.

2.3 Data

The proposed methodology is tested over Europe on SPI_3 values computed from the CRU TS3 0.5^o monthly data set 1901-2007 (Mitchell & Jones 2005). A useful feature of the TS3 dataset is the provision of auxiliary grids which supply a count of the stations contributing to each monthly grid cell value. The counts drop off sharply and their temporal variation becomes larger in the estimates of precipitation toward the east of Europe. Such changes in the network of contributing stations are likely to introduce inhomogeneities in the spatio-temporal correlation structure of the data. This motivates a truncation of the European domain at 35^oE. The study region is thus defined to be the land area 15^oW to 35^oE and 35^oN to 70^oN.

The focus of this preliminary analysis is on large scale events and the thus the raw SPI data are filtered to eliminate spatially coherent events smaller than 500,000 km^2. Such an area is of the order typical of the extra-tropical cyclones that dominate the European climate (Barry & Chorley 2003) and is in agreement with the practice advocated by Sheffield *et al.* (2009) for eliminating tenuous spatial connections. In order to further focus on the most temporally coherent events, the constraint is extended to the degree of spatial overlap between successive time slices of each particular event. Event structures in which the overlap falls below 500,000 km^2 are considered to be incoherent and are split into separate events.

2.4 Similarity

Once a set of coherent space-time events has been identified it is possible to test for similarities in their structure. The comparison of 3-dimensional shapes remains an active topic in computer science (Lmaati *et al.* 2009). The approach employed here is to construct a dissimilarity matrix of rank n where n is the n^{th} largest event as defined by the n neighbouring voxels of which it is comprised. The geographical extent of each event (longitude, latitude) is bounded by the study region, namely, 15^oW to 35^oE and 35^oN to 70^oN. The temporal extent of the event is bounded by the first and last voxels in the structure along the time direction. Events occurring at different times can be centred in a common space by aligning the time components of the centroids of each structure. Thus it is possible to embed all possible events within a subspace of the space-time domain dimensioned $Range_{lon}$, $Range_{lat}$, and $Range_t$ where the first two are defined by the study region and $Range_t$ is defined by the data as the longest duration event.

A simple measure of similarity is the Jaccard coefficient S_j (Jaccard 1901, Gower & Legendre 1986) where

$$S_j = \frac{a}{a+b+c} \tag{1}$$

and a is the number of elements shared between the two events, that is the number of common voxels. b is the number of voxels in event one which do not coincide with voxels in event two and c is the number of voxels within event two which do not coincide with voxels in event one. S_j has the desirable property of normalising the event volume but many other choices are possible (see Legendre & Legendre (1998)). Dissimilarity is then defined as (*ibid.*)

$$D_j = \sqrt{1 - S_j} \tag{2}$$

The resulting dissimilarity matrix can be decomposed by a variety of clustering techniques (Kaufman & Rousseeuw 1990). This analysis uses partitioning about medoids (PAM) (*ibid.*) which is a robust alternative to the traditional k-means

clustering method. An inherent problem with clustering analysis is the choice of the number of clusters k into which the data are partitioned. The silhouette width (Rousseeuw 1987) provides a comparison of the tightness of the groupings of events within each cluster to the separation between clusters. Taking the i^{th} object in the data set and denoting the cluster to which it has been assigned as A, then when the cluster contains other objects in addition to i it is possible to compute

$$a(i) = \text{average dissimilarity of } i \text{ to all other objects of } A \tag{3}$$

Considering any cluster C which is different from A, then

$$d(i, C) = \text{average dissimilarity of } i \text{ to all objects of } C \tag{4}$$

Assuming that $k > 1$, that is there is more than one cluster, then it is possible to locate the cluster B which contains the minimum

$$b(i) = \min\{d(i, C)\} \tag{5}$$

from any $C \neq A$. The cluster B is the second best choice for the assignment of object i. The silhouette width is defined as

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{6}$$

The best case is a value close to $+1$ which implies that the within cluster dissimilarity $a(i)$ is much smaller than the smallest between cluster dissimilarity $b(i)$, that is the second best choice of cluster B is not nearly as close as the actual choice A. When $s(i)$ is close to zero $a(i)$ and $b(i)$ must be nearly equal and it not clear whether i should have been assigned to A or B. If $s(i)$ is close to -1 then $a(i)$ is much larger than $b(i)$ and it would be more natural to classify i as belonging to B, under these circumstances it is likely that i has been misclassified.

2.5 Summary statistics

The 3-dimensional nature of the proposed event definition means that, in addition to the structural classification, it is possible to compute summary statistics which relate to the event geometry. The specific choice of statistic will depend upon the task in hand *e.g.* climate model validation, visualization, or trend detection. Example measures considered here are the volume, centroid, and aspect ratio. The event volume is important since this represents a measure of the absolute severity in combined terms of extent and duration. The aspect ratio discriminates between large volumes arising from short period deficits over a wide area and those accrued from sustained deficits over limited areas. Finally, the event centroids can be used to explore the distribution of events in space and time.

A complicating factor is the dependence of the grid cell area (and hence cell volume) on latitude. Taking the Earth to be spherical with radius $R_e = 6371$km the grid cell area between longitudes λ_1 and λ_2 and latitudes ϕ_1 and ϕ_2 is given by

$$Area = \int_{\phi_1}^{\phi_2} R_e \int_{\lambda_1}^{\lambda_2} R_e \cos\phi \, d\lambda \, d\phi = R_e^2(\lambda_2 - \lambda_1)(\sin\phi_2 - \sin\phi_1) \tag{7}$$

which is used to transform from event volumes in grid cell units to deficits measured in km^{2}months.

3 Results

3.1 Event extraction: 1976

The well known European drought of 1976 (Zaidman *et al.* 2002, Doornkamp *et al.* 1980) provides an interesting example to illustrate the output from the extraction procedure as applied to a single event. Figure 2 (a) is an isometric projection of the coherent voxels of $\text{SPI}_3 \leq -1$ captured by $R_{lon} = R_{lat} = 0$ grid cell units of 0.5^o with $R_{time} = 1$ month. The view is from the southwest looking backward in time from December to January 1976. Whilst the image serves the pedagogical purpose of illustrating the 3 dimensional coherency of the data, it does little to reveal the

spatio-temporal characteristics of the drought. To address this panel (b) provides a Hovmöller type plot of the number of voxels in the drought volume counted north-south through time. This view is made up of pixels which are shaded according to the number of cells within the event volume counted along each meridional band. Data for a particular month form a row which completely spans the study region from west to east. The coloured elements represent the shadow, or 'footprint', of the event as projected onto the back plane of the bounding box shown in panel (a). The depth of colour represents the integrated thickness of the event in this direction. Since the number of missing values is a function of the coastline, this representation is limited by the fact that the cell count varies along an particular longitude and hence the maximum possible event depth varies across the domain. Thus, in contrast to colour variations along a column which can readily be interpreted as changes in the scale of the drought with time, variations along a row need to be interpreted with caution. A further limitation of the 2-dimensional projection is that contributions are confounded along the direction of the summation. In particular, we lose all sense of where a drought might break along any longitudinal band. Panel (c) provides a similar view but counts cells east-west through time. Finally, panel (d) maps cell counts by location to represent the drought severity and maximum spatial extent. This view is analogous to the drought severity maps introduced by Andreadis *et al.* (2005) except that here the map represents severity aggregated throughout the lifetime of the event rather than for a snapshot of a particular time. Whilst this remains an imperfect representation of the full 3-dimensional event structure, the combined views capture the essence of the event. Importantly, we see the core of the drought centred over mid-western Europe around the summer of 1976 and the coherent evolution from earlier deficits in the south and west to later water shortages over Scandinavia.

Since the choice of coherency radii ($R_{lat}, R_{lon}, R_{time}$) are likely to exert a significant influence on the form of the event, the extraction was repeated with $R_{lon} = R_{lat} = 2$ and $R_{time} = 1$ grid cell units. The resultant strucutre (not shown) was found to be very similar to the previous result. The differences are minor extensions of the event to the south into Spain and Turkey. The sensitivities of smaller drought

events to changes in coherency radii are considered in Section 3.3 below.

3.2 Spatio-temporal characterisation

Coherent space-time drought events were extracted from the European SPI_3 dataset 1901-2006 for coherency radii $R_{lon} = R_{lat} = 0, 1, 2$ grid cell units of 0.5^o. Droughts lasting less than 3 months were discarded to leave a total of 75, 89, and 96 events respectively. Silhouette widths were computed for clusters of volumes from each of the event sets defined by $R_{lat} = R_{lon} = 0, 1, 2$. The number of clusters k was taken from a minimum of 2 to a maximum defined by the number of events within the set. The results are shown in Figure 3(a). The average silhouette width remains below 0.05 for all R and k considered. The standard interpretation of silhouette widths Makra & Sümeghy (2007) is that average values below 0.25 indicate the absence of any substantial structure within the dissimilarity matrix. The conclusion is that there is no significant commonality of form between the individual drought events. This result is confirmed by an alternative set of silhouette widths constructed by aligning each event by start date rather than the temporal component of the centroid. These are shown in Figure 3(b) where again the average widths are seen to remain below 0.05.

The origin of the near complete dissimilarity between events is immediately evident from Figure 4. This is a map of the geographical location of the event centroids of the 75 events captured by $R = 1$ and $SPI_3 \leq -1$ and lasting for at least 3 months. With the possible exception of the clusters of events over Bellarus and southern Finland, the majority of drought events are well separated. There is a noticeable tendency for the centroids of the largest and longest droughts to cluster around the central and eastern section of the study region. This is an example of the 'mid domain effect' (Colwell & Lees 2000) whereby geometric constraints, in this case missing data over the oceans to the north, south and west, lead to an apparent geographical bias in the analyzed data.

Further dissimilarity is revealed by Figure 5 which maps the time evolution of the locations of the event centroids. The tail of each arrow indicates the mean location of the centroid during the first half of the event. The head of each arrow points

to the mean location of the centroid during the later half of the event. The rose diagram and histogram to the left of the map show a clear tendency for drought to develop preferentially in either direction along the east-west axis. The distribution mirrors the geographical bias seen in Figure 4 and arises from the greater availability of grid cells along the east-west axis for any given drought to grow into. Likewise the increase in land area north-south toward the east of the study region accounts for the bias in west-east over east-west development. Whilst beyond the scope of this paper, it would be an interesting exercise to construct a geographically constrained null hypothesis of random orientation and test for any residual directionality in drought development. The reason that such a test is not developed here is that the construction of the appropriate null events would require complete data coverage across the study region. A partial solution might be to apply the present method to reanalysis data such as ERA40 from the European Centre for Medium-range Weather Forecasting. Events could first be extracted from the full precipitation field across the study region and then again with data over water masked out. Comparison between the two events sets should provide information on the influence of data availability on the orientation of the 3-dimensional drought structures. However, in order to compare results from the reanalysis with observations (such as CRU TS3), great care would have to be taken to ensure consistency in the autocorrelations (see *e.g.* Beale *et al.* (2008), Legendre (1993)).

3.3 Summary statistics

The previous section considered the spatial characteristics of the centroids of the events defined by $R_{lon} = R_{lat} = 0$. Here consideration is given to the distribution of these events in time. Figure 6 plots each event (coloured by duration) on the time line 1901-2007 and charts the evolution of the geographic cross-sectional areas. The events have been coloured according to duration to aid comparison with Figures 4 and 5. There are three main points of interest. Firstly, with the exception of two small events in the winter 1984-85, the events are well separated in time. The absence of overlap between events indicates that large scale drought events within the European sector tend to appear in isolation. Secondly, with the possible exception

of 1976, all events which last for 9 months or longer exhibit a sharp contraction in their extent at least once during their lifetimes. Similar contractions are seen in the majority of events which last for 6 months or longer. This may indicate that longer period droughts result from the merger of two or more shorter period events. Thirdly, the rate of occurrence of drought exhibits variability at the decadal scale. This includes extended periods such as 1934-1938, 1954-1959 and 1964-1969 where large scale droughts are almost completely absent. It is interesting to note that each of these periods corresponds with the negative phase of the North Atlantic Oscillation (NAO). However, this observation is tempered by the strongly negative NAO period 1939-1944 which is seen to be relatively drought rich.

The time series representation provided by Figure 6 permits a direct comparison of the approximate timing and severity of the major events identified here with those reported in previous drought catalogues (Briffa *et al.* 1994, Lloyd-Hughes & Saunders 2002, van der Schrier *et al.* 2006, Sheffield *et al.* 2009, Lloyd-Hughes *et al.* 2009). Perhaps unsurprisingly given the general similarities of the drought indices used, there is close correspondence between the event sets.

Moisture deficits in spring and early summer, in particular over the Mediterranean region, have been shown to contribute to the development of extremely high summer temperatures in continental Europe (Zampieri *et al.* 2009). If these results extend to summer drought then it is reasonable to expect a seasonal bias in the onset of the events in Figure 6. The histograms (inset Figure 6) show the seasonal distribution of the dates in which the droughts begin and end. The counts are stacked by event duration and coloured according to the same scale as the main figure. χ^2 tests against the null hypothesis of a uniform distribution of starts and finishes between seasons gave p-values of 0.74 and 0.96 respectively and no significant seasonal bias can be detected. However, it is interesting to note that 7 out the 8 longest droughts began in either winter or spring.

To further test the sensitivity of the results to the choice of coherency radii, a subset of the summary statistics described in Section 2.5 were computed for $R_{lon} = R_{lat} = 0$ and are displayed in Figure 7. Namely by row, these are the event volumes, durations, maximum spatial areas, and aspect ratios. Each column represents a

specific choice of radius. Collectively, these confirm the tentative conclusion drawn from analysis of the 1976 event in Section 3.1; that the extracted drought structures are insensitive to small changes in R_{lon} and R_{lat}. Note that the absence of events lasting for less than 3 months or with maximum areas less than 500,000km^2 which were excluded prior to the analysis. The distributions of event volume, duration, and to a lesser extent the maximum extent, exhibit strong positive skew. This supports the hypothesis that the larger scale droughts are aggregates of smaller ones.

4 Discussion

4.1 Event definition

It has been necessary to impose many subjective choices on the selection of parameters that define the event structure, all of which have the potential to exert a significant influence on the geometric form of the event. The parameter space for event definition within the proposed framework spans:

(a) Spatial resolution
(b) Temporal resolution
(c) Proportion and distribution of data gaps
(d) SPI averaging period
(e) SPI event threshold
(f) Spatial coherency radii
(g) Minimum event area
(h) Minimum event overlap

Typically the spatio-temporal characteristics of the data can be fixed prior to analysis and intercomparisons will be valid if the data are regridded onto the same basis and data gaps are homogeneous in time. The SPI averaging period and threshold are features of the drought metric which, in common with the standard application of the SPI, can be tuned to the specific scales (time period and probability) of interest. In fact, drought definition need not be restricted to the SPI for drought

definition. It is easy to see how the 3-dimensional techniques described here could be applied to alternative drought definitions *e.g.* the Palmer Drought Severity Index. The only real limitation is that the data are represented on a space-time grid. Thus parameters (a)-(e) are not specific to the proposed methodology, but relate to the difficulties in defining drought in general.

The choice of spatial coherency radii, minimum event area and overlap are specific to the proposed methodology. These control the level of aggregation of the data and can be used to shift the focus from tightly bounded local events to more diffuse continental scale phenomena. The preliminary analysis presented here indicates that the results are robust for small changes in R_{lon} and R_{lat}. Experiments with different values of minimum event area and overlap show a similar degree of insensitivity.

4.2 Dissimilarity of European droughts 1901-2006

The complete lack of similarity between the event volumes is surprising since previous work on the spatial distribution of European drought (Briffa *et al.* 1994, Lloyd-Hughes & Saunders 2002, van der Schrier *et al.* 2006) indicate several preferred modes of variability. Spatial similarity between drought events is easily revealed by searching for maxima in the Jaccard coefficients between the individual monthly slices of the 3-dimensional events. Figure 8 illustrates the striking similarities between slices of events during (a) March 1963 and (b) January 1996. Panels (c) and (d) show the raw SPI_3 data from which these slices of the events were extracted. The pattern correlation is 0.78. Near perfect agreement is seen across the entire study domain including the excess precipitation to the south. Since the SPI fields were selected solely on the correspondence between slices of the drought events defined over the Baltic, the exceptional pan-European agreement is a strong indication that similar climate dynamics are responsible for driving both events.

Figure 9 compares the events within the 3-dimensional spatio-temporal context. The severity-extent maps in panels (a) and (b) reveal quite different pictures. The 1963 event is seen to be centred to the east of the Baltic whilst the 1996 event is centred over Denmark. The corresponding Hövmoller plots in panels (c) to (f), aside

from March 1963 and January 1996 (shown by horizontal black lines), are also very different. March 1963 is located in the middle of an event of 8-months duration that started over northern central Europe and migrated east, whereas January 1996 was toward the beginning of a 5-month event that remained static over northern central Europe. It is interesting to note that if the Hövmoller plots for the 1962/3 event are truncated at April 1963 then there is much greater similarity with the 1995/6 event. Reference to Figure 6 reveals a contraction in the drought area in April 1963. Thus, whilst clearly spatio-temporally coherent, the 1962/3 event might be considered to be composed of two distinct sub-droughts, the first being very similar to the 1995/6 event.

This result has strong implications for analogue forecast methods, that is, those which aim to predict the future from the past evolution of systems which resemble the present. A spatial snapshot of drought is a very poor guide to its future development.

4.3 Drought as a spatial process

There is evidence for dynamically driven spatial similarity between droughts at a monthly timescale. There is also evidence for randomness in the space-time character of European droughts (manifest in the complete dissimilarity between their large scale structures). The spatial distribution of the event centroids and the time evolution of the geographic cross-sectional areas both strongly suggest that large area sustained droughts result from the combination of smaller shorter lasting events. However, if this is the case, more spatially isolated but temporally coincident events are expected than are observed in Figure 6. One possible explanation for this is that the underlying generating process is one of relatively small events ($\sim 10^6 \mathrm{km}^2$ lasting $\sim$ 3 months) that can occur anywhere in the domain but that feedbacks exist that tend to favour co-location in space and time. This is supported by model studies of the prolonged large scale drought and summer heatwave of 2003 (Fischer *et al.* 2007) which found local water feedbacks to be an essential feature required to sustain the event. Whilst local effects are important, the multiannual variability seen in the rate of drought occurrence indicates that larger scale forcings such as changes in ocean

circulation might also play an important role in controlling drought occurrence.

An alternative hypothesis is that the observed results are predetermined by the use of SPI_3. This measure of drought retains considerable short-time scale variablity, reflecting contributions from individual weather systems to the monthly precipitation totals, that will (by definition) emerge at the 3-monthly time scale. It is possible to discriminate between the two hypotheses by applying the new method to a drought index with greater memory. To this end, the analysis was repeated using SPI_{12}. Figure 10 illustrates the time evolution of the cross-sectional areas of the events defined by $SPI_{12} \leq -1$ and $R_{lon} = R_{lat} = 0$. The longer memory index, as might be expected, shifts the emphasis onto longer lasting events. However, the time series retains a close correpondence to that shown in Figure 6 for SPI_3. Of particular interest to this discussion, is that the longest lasting events, *e.g.* 1941-1944 and 1920-1922, are again seen to arise from the aggregation of several shorter period deficits. It is also worth noting that the SPI12 events are again seen to be well separated in time. This again suggests the existence of feedbacks that to favour the co-location of drought developments in space and time. With respect to seasonality, the SPI_{12} events display a bias toward an Autumn start. However, χ^2 tests against the null hypothesis of a uniform distribution of starts and finishes between seasons gave p-values of 0.64 and 0.85 respectively which can not be considered to be significant.

The small-area short duration hypothesis is further supported by the summary statistics for the SPI_{12} events. These are presented in Figure 11. The majority of evetns are again seen to be of small area ($\sim 10^6 \text{km}^2$) and short duration ($\sim$3 months).

For completeness, silhouette widths (not shown) were computed for the SPI_{12} events. The move to a longer memory index, with the implicit assumption of reducing high frequnecy varibability, was not found to increase the degree of 3-dimensional similarity between the events and the average silhouette width remainded below 0.05.

5 Conclusions

An approach to drought characterisation based on spatio-temporal structure has been shown to provide valuable additional information not immediately available from traditional analyses. Firstly, individual time slices of events such as 1962/63 and 1995/96 which can be shown to be remarkably similar (*e.g.* March 1963 and January 1996) can be seen to have very different time evolutions. Secondly, explicit consideration of the time evolution of the spatial structure highlights the role of external and internal forcings on drought development. The time evolution of the geographic cross-sectional areas of both short-period (SPI_3) and long-period (SPI_{12}) strongly suggest that large area, sustained droughts result from the combination of multiple small area ($\sim 10^6 km^2$) short duration ($\sim$3 months) events. Consideration of summary statistics such as the distribution of maximum event cross-sections and event centroids leads to the hypothesis that drought may beget drought, and thus that local water feedbacks may play an important role in the aggregation process.

The 3-dimensional view is unamenable to simple summary representation; to the contrary, it encourages consideration from multiple perspectives. However, since complexity across dimension and scale is intrinsic to deficits of precipitation, it is hoped that this broader view will deepen our understanding of the space-time nature of drought.

Acknowledgements

I thank Jean-Philippe Vidal, Christel Prudhomme and Jamie Hannaford for helpful comments on the text. I thank Deloitte for supporting the Deloitte-Walker Institute Research Fellowship at the University of Reading.

References

Andreadis KM, Clark EA, Wood AW, Hamlet AF, Lettenmaier DP. 2005. Twentieth-century drought in the conterminous united states. *Journal of Hydrometeorology*. **6(6)**:985–1001.

Barry RG, Chorley RJ. 2003. *Atmosphere, weather & climate*. Routledge, 8th edition.

Beale CM, Lennon JJ, Gimona A. 2008. Opening the climate envelope reveals no macroscale associations with climate in european birds. *Proceedings of the National Academy of Sciences of the United States of America*. **105(39)**:14908–14912.

Briffa KR, Jones PD, Hulme M. 1994. Summer moisture variability across Europe, 1892-1991: An analysis based on the Palmer Drought Severity Index. *International Journal of Climatology*. **14**:475–506.

Colwell RK, Lees DC. 2000. The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology & Evolution*. **15(2)**:70–76.

Doornkamp JC, Gregory KJ, of British Geographers. I. 1980. *Atlas of drought in Britain, 1975-76 / edited by J.C. Doornkamp and K.J. Gregory ; cartographic advisor, A.S. Burn ; forward by Denis Howell*. Institute of British Geographers, London :. ISBN 0901989312.

Fischer EM, Seneviratne SI, Luthi D, Schar C. 2007. Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophysical Research Letters*. **34(6)**:L06707.

Gower JC, Legendre P. 1986. Metric and euclidean properties of dissimilarity coefficients. *Journal of classification*. **3**:5–48.

Hannachi A, Jolliffe IT, Stephenson DB. 2007. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*. **27(9)**:1119–1152.

Horel JD. 1984. Complex principal component analysis: Theory and examples. *Journal of Applied Meteorology*. **12**:1660–1673.

Jaccard P. 1901. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*. **37**:547–579.

Kaufman L, Rousseeuw P. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.

Legendre P. 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology*. **74(6)**:1659–1673.

Legendre P, Legendre L. 1998. *Numerical Ecology*. Developments in Environmental Modelling, 20. Elsevier.

Livada I, Assimakopoulos VD. 2007. Spatial and temporal analysis of drought in Greece using the Standardized Precipitation Index (SPI). *Theoretical and Applied*

Climatology. **89(3-4)**:143–153.

Lloyd-Hughes B, Hannaford J, Parry S, Keef C, Prudhomme C, Rees G. 2009. UK and European drought catalogue. Technical report, UK Environment Agency.

Lloyd-Hughes B, Saunders MA. 2002. A drought climatology for Europe. *International Journal Of Climatology*. **22(13)**:1571–1592.

Lmaati EA, Oirrak AE, Kaddioui MN. 2009. A visual similarity-based 3D search engine. *Data Science Journal*. **8**:78–87.

Makra L, Sümeghy Z. 2007. Objective analysis and ranking of Hungarian cities, with different classification techniques, part 1: Methodology. *Acta Climatologica et Chorologica*. **40-41**:79–89.

McKee TB, Doesken NJ, Kleist J. 1995. Drought monitoring with multiple time scales. *Ninth conf. on Applied Climatology*. 233–236.

Mitchell TD, Jones PD. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal Of Climatology*. **25(6)**:693–712.

MunichRe. 2008. Great weather disasters 1950-2007. Technical report, NatCatSERVICE.

Rousseeuw P. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. **20(1)**:53–65.

Sheffield J, Andreadis KM, Wood EF, Lettenmaier DP. 2009. Global and continental drought in the second half of the twentieth century: Severity-area-duration analysis and temporal variability of large-scale events. *Journal of Climate*. **22(8)**:1962–1981.

Smakhtin VU, Schipper ELF. 2008. Droughts: The impact of semantics and perceptions. *Water Policy*. **10(2)**:131–143.

Szalai S, Szinell C. 2000. Comparison of two drought indices for drought monitoring in Hungary - a case study. In JV Vogt, F Somma, editors, *Drought and Drought Mitigation in Europe*, 161–166. Kluwer: Dordrecht. 325pp.

UN. 2008. Trends in sustainable development 2008-2009. United Nations Division for Sustainable Development.

van der Schrier G, Briffa KR, Jones PD, Osborn TJ. 2006. Summer moisture variability across europe. *Journal Of Climate*. **19(12)**:2818–2834.

Vicente-Serrano S. 2006. Differences in spatial patterns of drought on different time scales: An analysis of the Iberian Peninsula. *Water Resources Management*. **20(1)**:37–60.

von Storch H, Zwiers FW. 1999. *Statistical Analysis in Climate Research*. Cambridge University Press: Cambridge.

Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press: London. 467pp.

Zaidman MD, Rees HG, Young AR. 2002. Spatio-temporal development of streamflow droughts in north-west Europe. *Hydrology and Earth System Sciences*. **6(4)**:733–751.

Zampieri M, D'Andrea F, Vautard R, Ciais P, de Noblet-Ducoudré N, Yiou P. 2009. Hot european summers and the role of soil moisture in the propagation of mediterranean drought. *Journal of Climate*. **22(18)**:4747–4758.

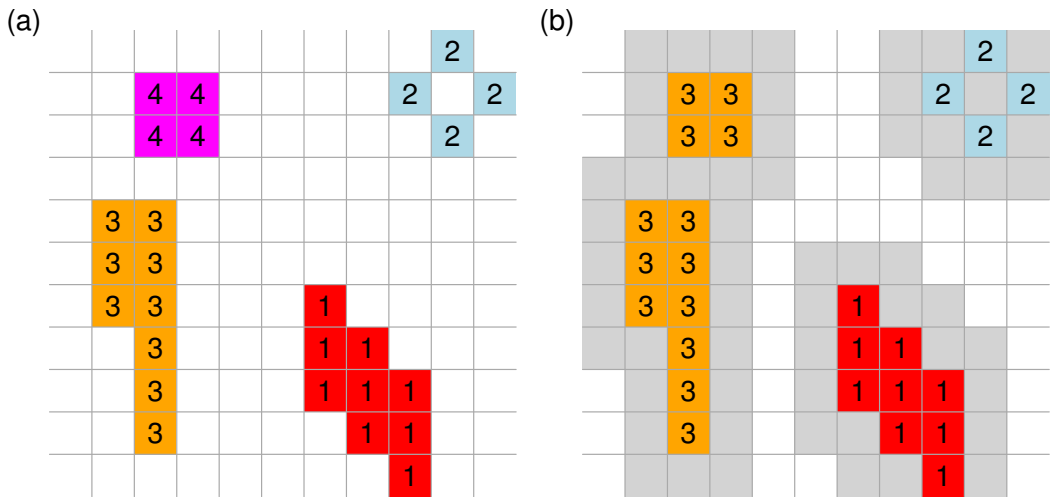

Figure 1: Definition of spatial coherency in 2 dimensions. Groups of coherent cells are shaded in a single colour. Panel (a) is for $R = 0$ and shows four groups. Panel (b) is for $R = 1$ and shows three groups. The gray shading illustrates the effective size of each group.

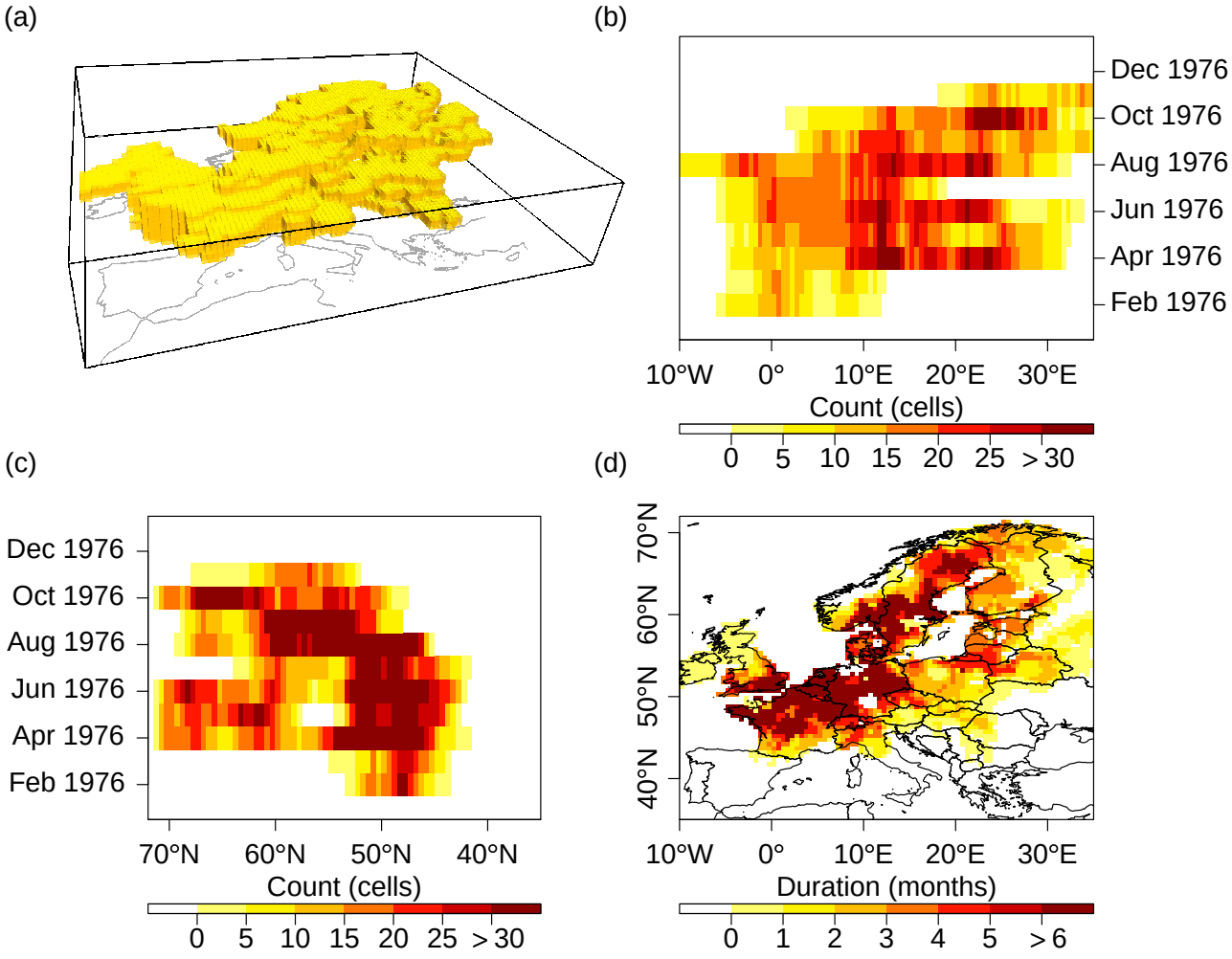

Figure 2: (a) Isometric view of the 1976 drought event defined by $R_{lon} = R_{lat} = 0$ and $\text{SPI}_3 \leq -1$. (b) Hovmöller type plot of the number of grid cells in drought counted north-south through time. (c) East-west Hovmöller. (d) The view from above shows the drought severity and maximum spatial extent.

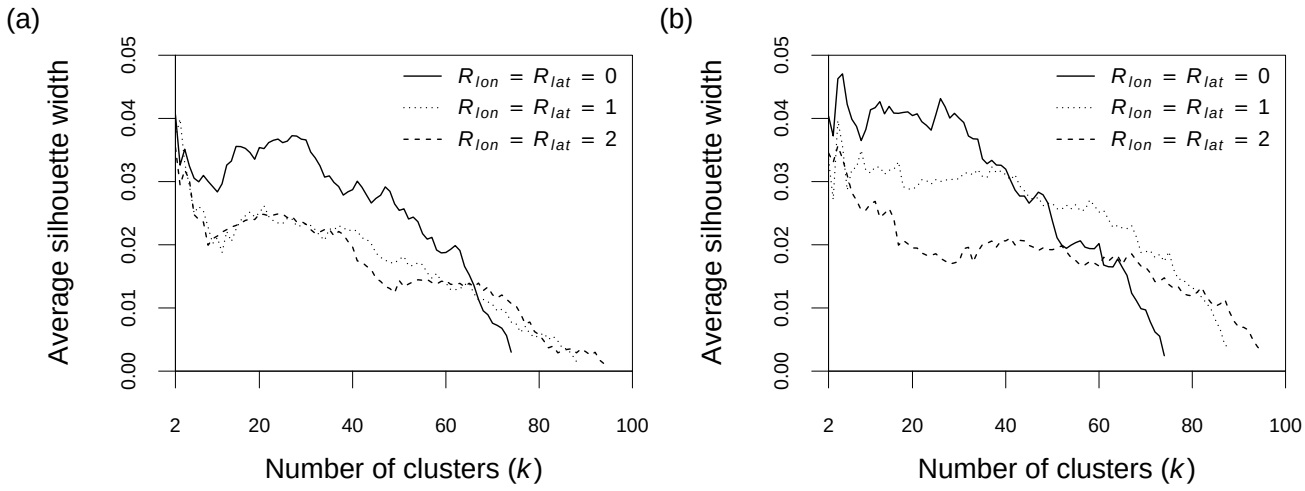

Figure 3: Average silhouette widths for clusters of droughts defined by $R_{lon} = R_{lat} = 0, 1, 2$ with the SPI_3 threshold ≤ -1 and lasting for at least 3 months. (a) Events aligned by the time component of the event centroid. (b) Events aligned by start date.

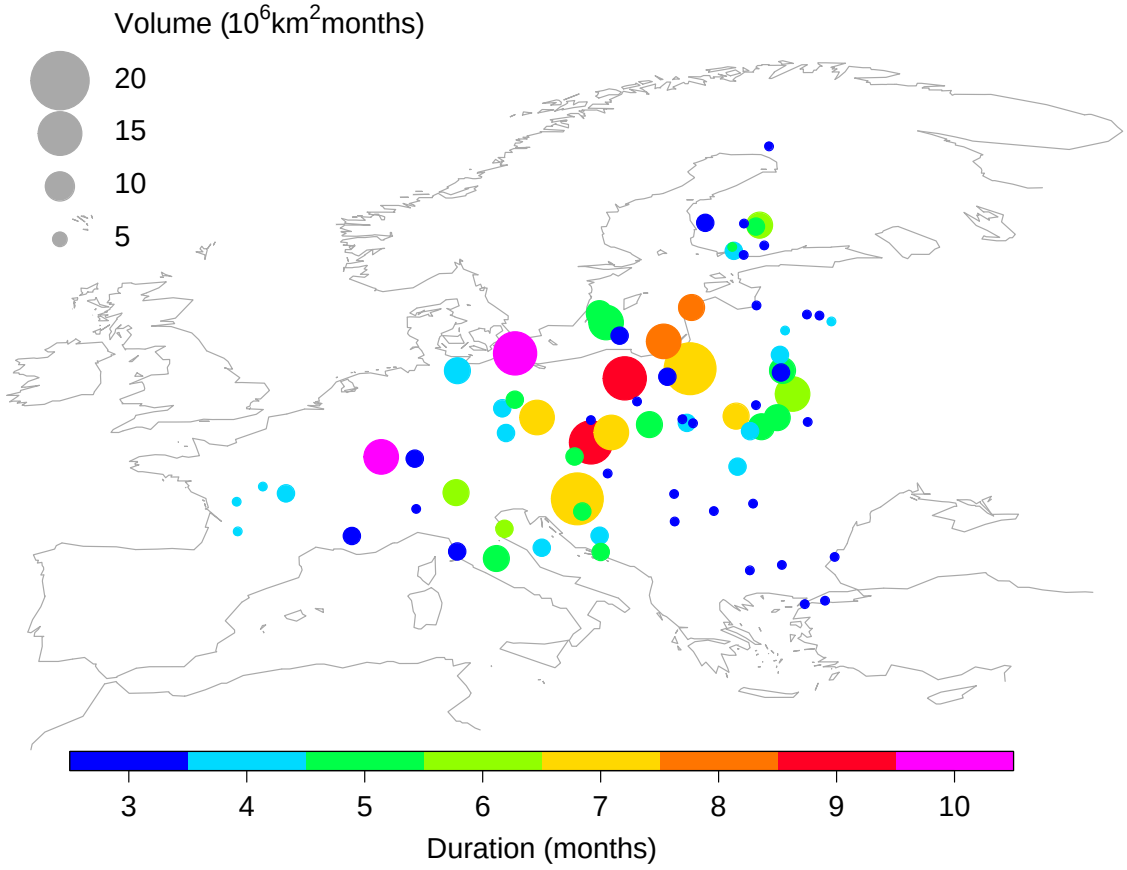

Figure 4: Distribution of event centroids for droughts defined by $R_{lon} = R_{lat} = 0$ with the SPI_3 threshold ≤ -1 and lasting for at least 3 months.

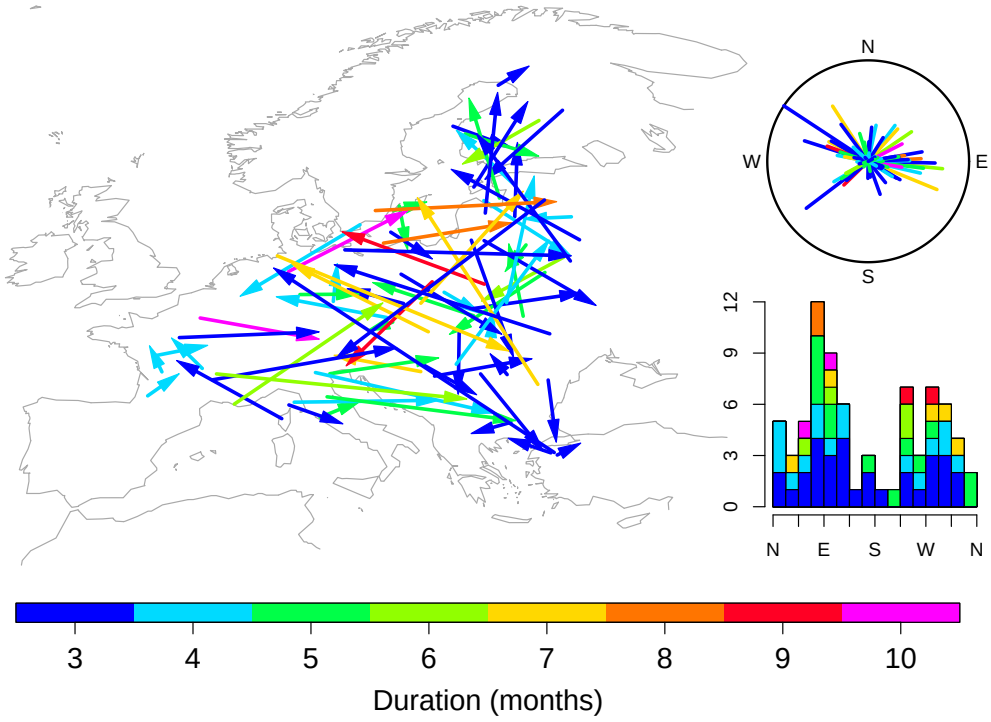

Figure 5: Time evolution of the locations of the event centroids for droughts defined by $R_{lon} = R_{lat} = 0$ with the SPI_3 threshold ≤ -1 and lasting for at least 3 months. The tail of each arrow indicates the mean location of the centroid during the first half of the event. The head of each arrow points to the mean location of the centroid during the later half of the event. The rose diagram and histogram (stacked by duration) illustrate the directional distribution of the shifts in the event centroids.

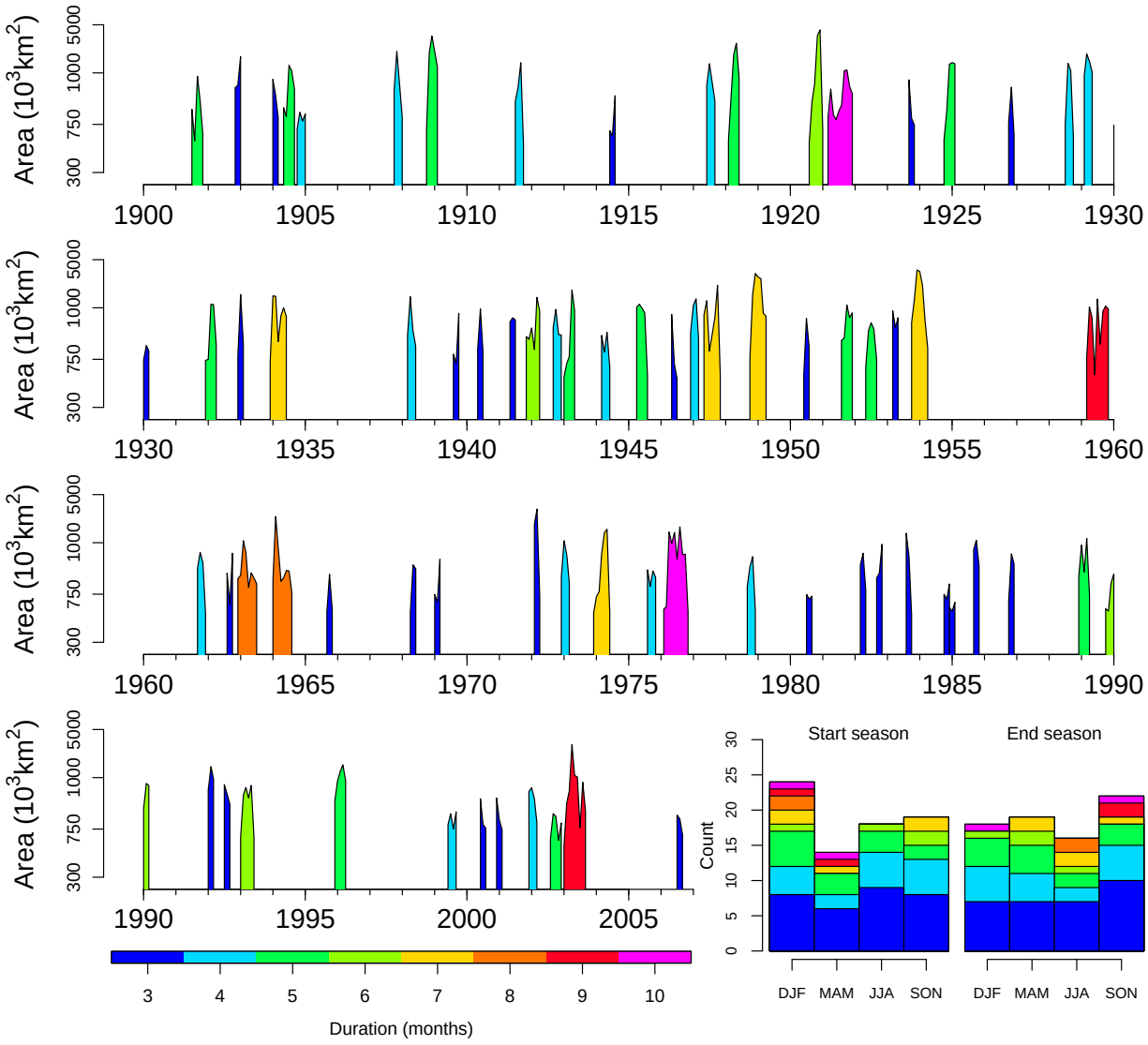

Figure 6: Time evolution of the geographical cross-sectional areas of the 75 droughts events defined by $R_{lon} = R_{lat} = 0$ with the SPI_3 threshold ≤ -1 and lasting for at least 3 months. The histograms (inset) show the seasonal distribution of the dates in which the droughts begin and end. The counts are stacked by event duration.

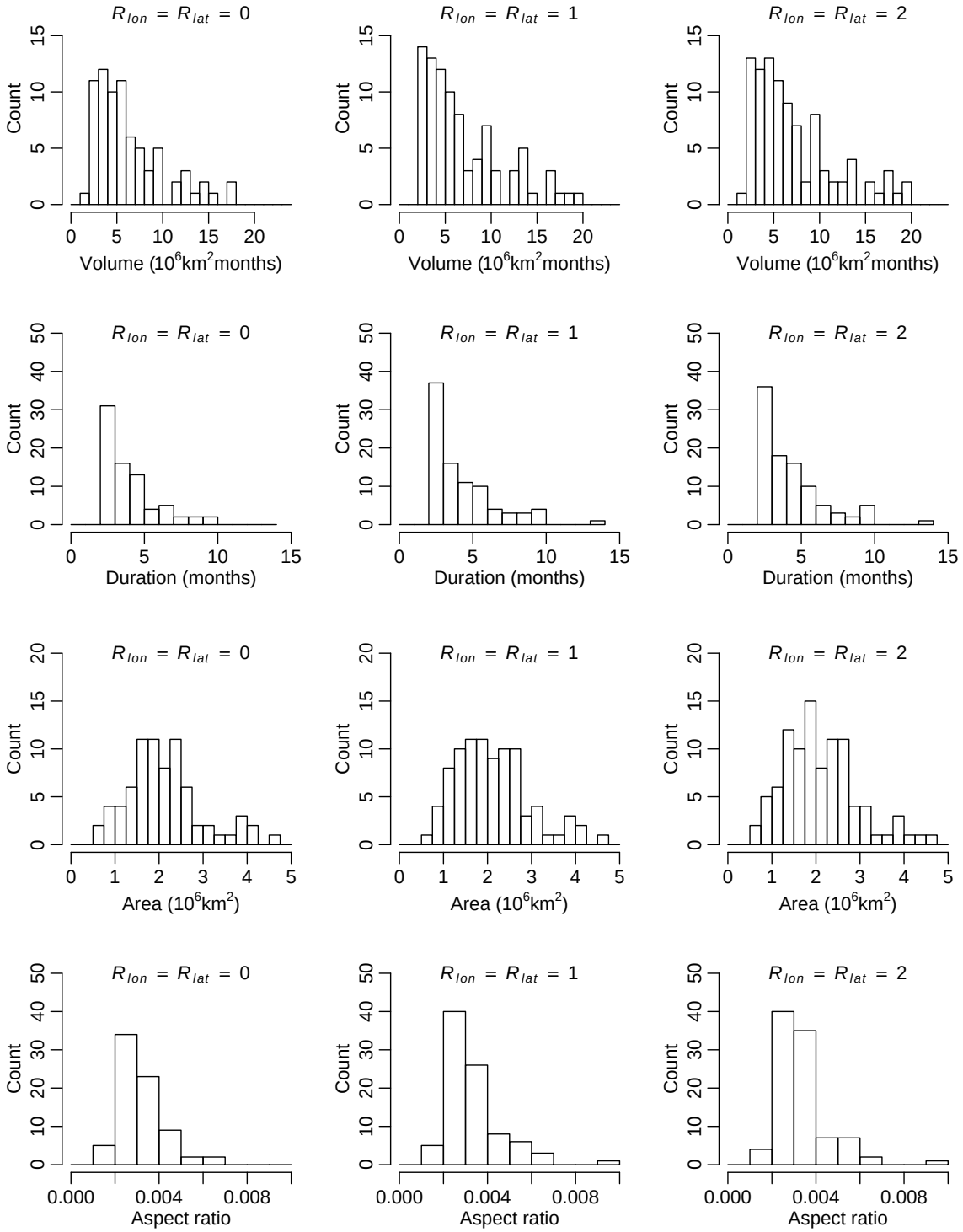

Figure 7: Summary statistics described in Section 2.5 computed for $R_{lon} = R_{lat} = 0, 1, 2$ and SPI_3 threshold ≤ -1. By row, these are the event volumes, durations, maximum spatial areas, and aspect ratios. Each column represents a specific choice of radius.

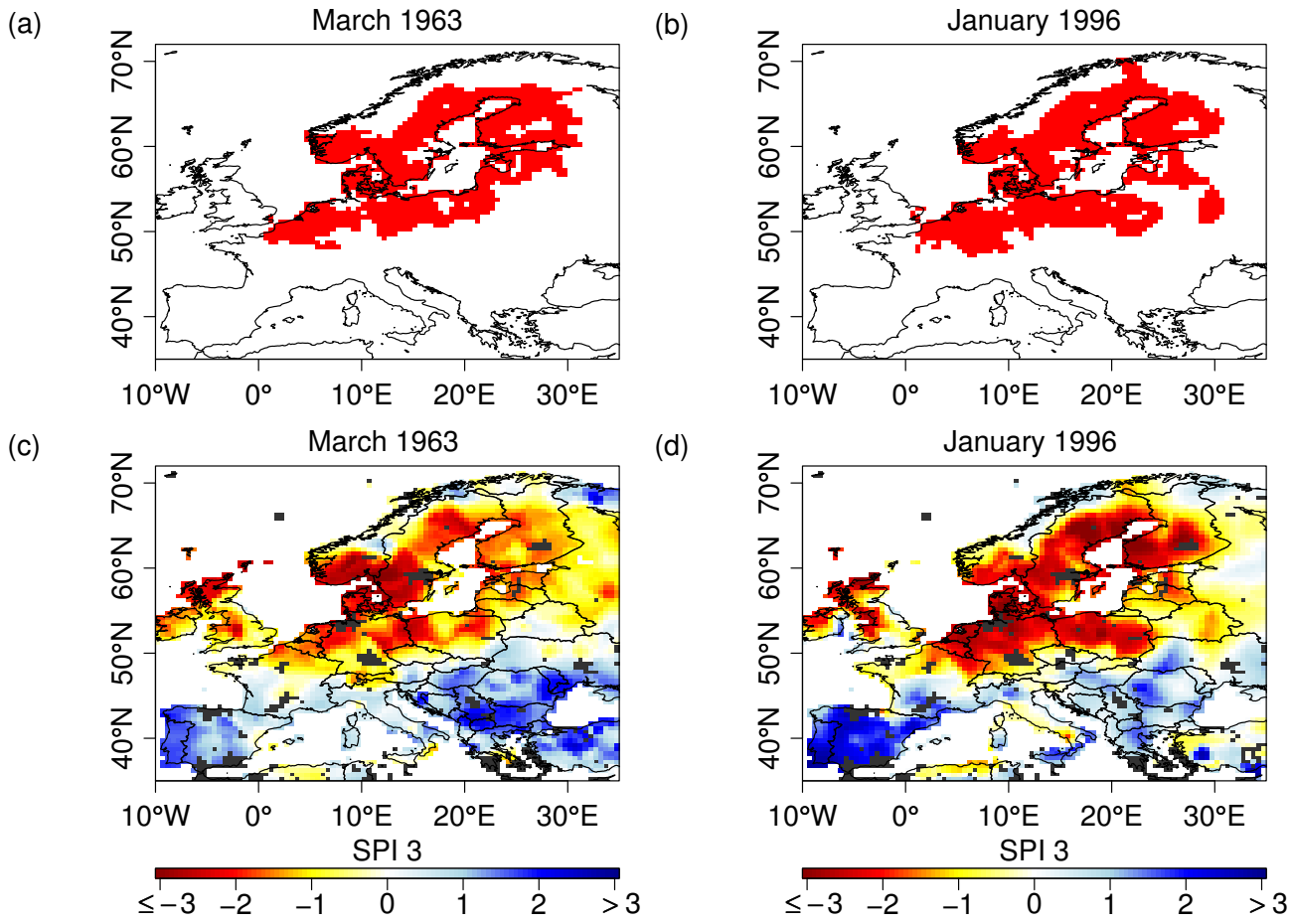

Figure 8: (a) Single month slices through events for in March 1963 and January 1996 respectively. These correspond to the most similar 2-dimensional segments from the volumetric events defined by $R_{lon} = R_{lat} = 0$ with the SPI_3 threshold ≤ -1. (b) The corresponding SPI_3 fields from which the events were extracted. Areas of dark gray shading indicate regions where the gamma distribution used in the construction of the SPI is inadequate to standardise the precipitation.

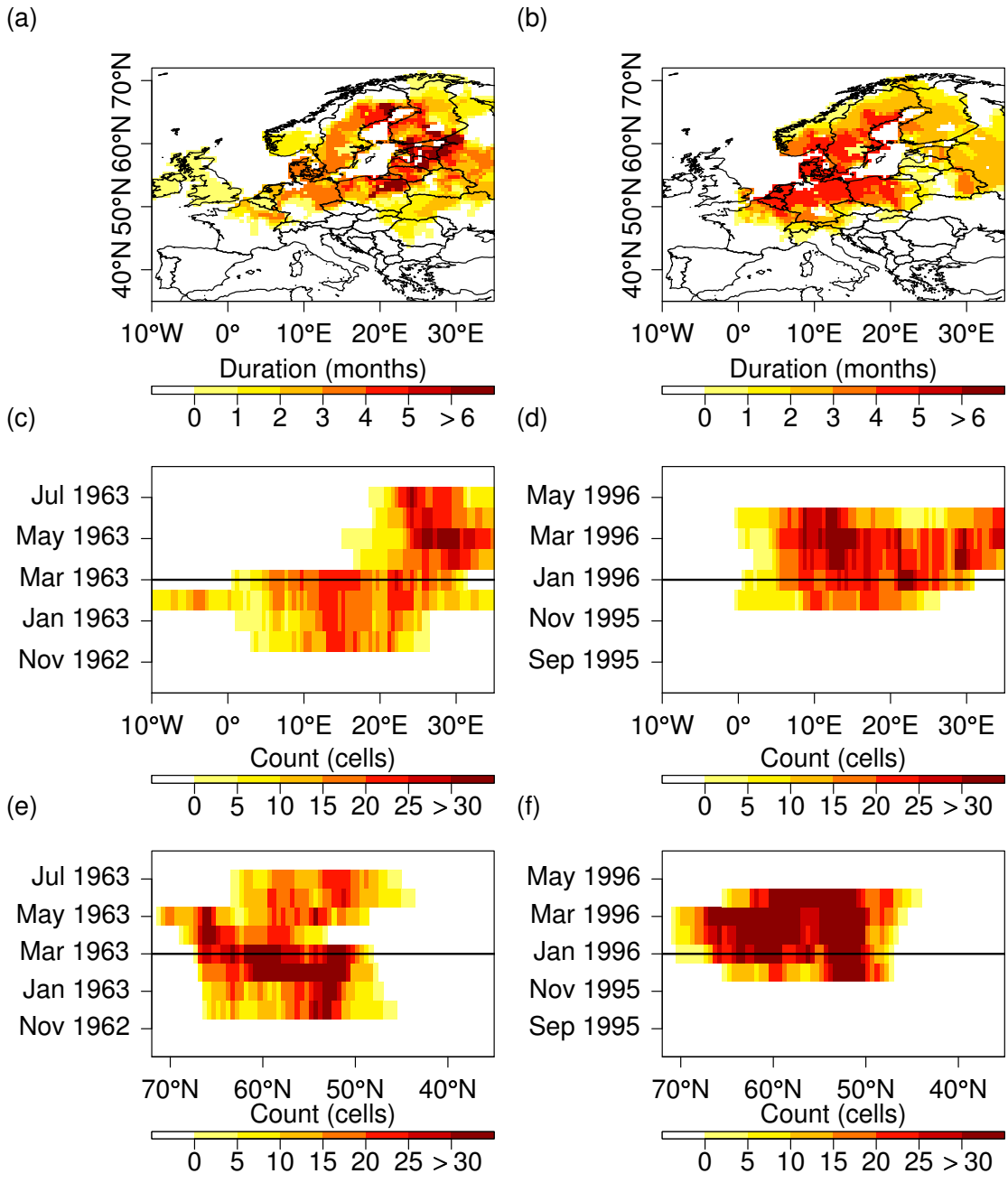

Figure 9: As Figure 2 but comparing the spatio-temporal characteristics of the 1963 and 1996 events.

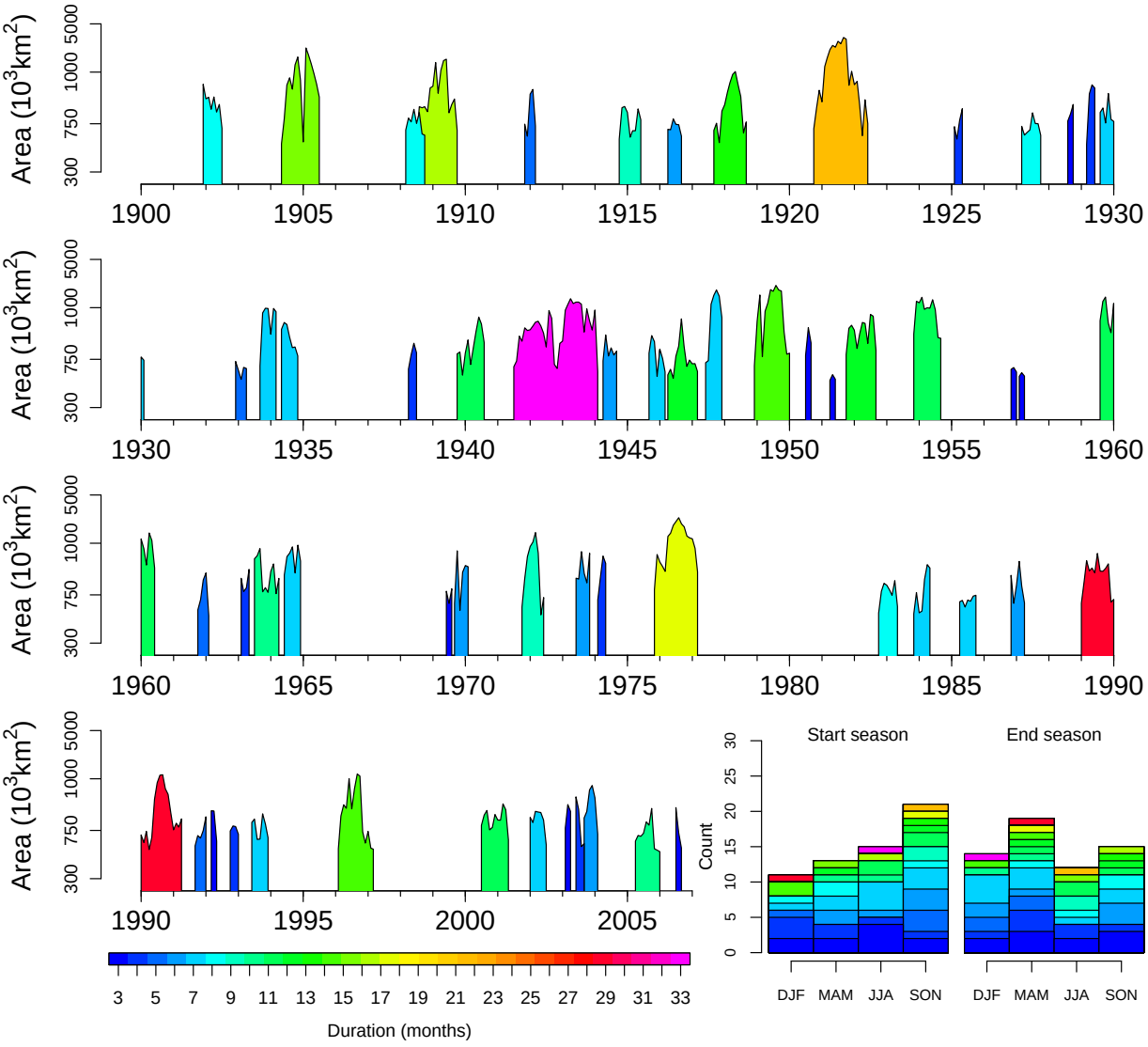

Figure 10: As Figure 6 but for SPI_{12}.

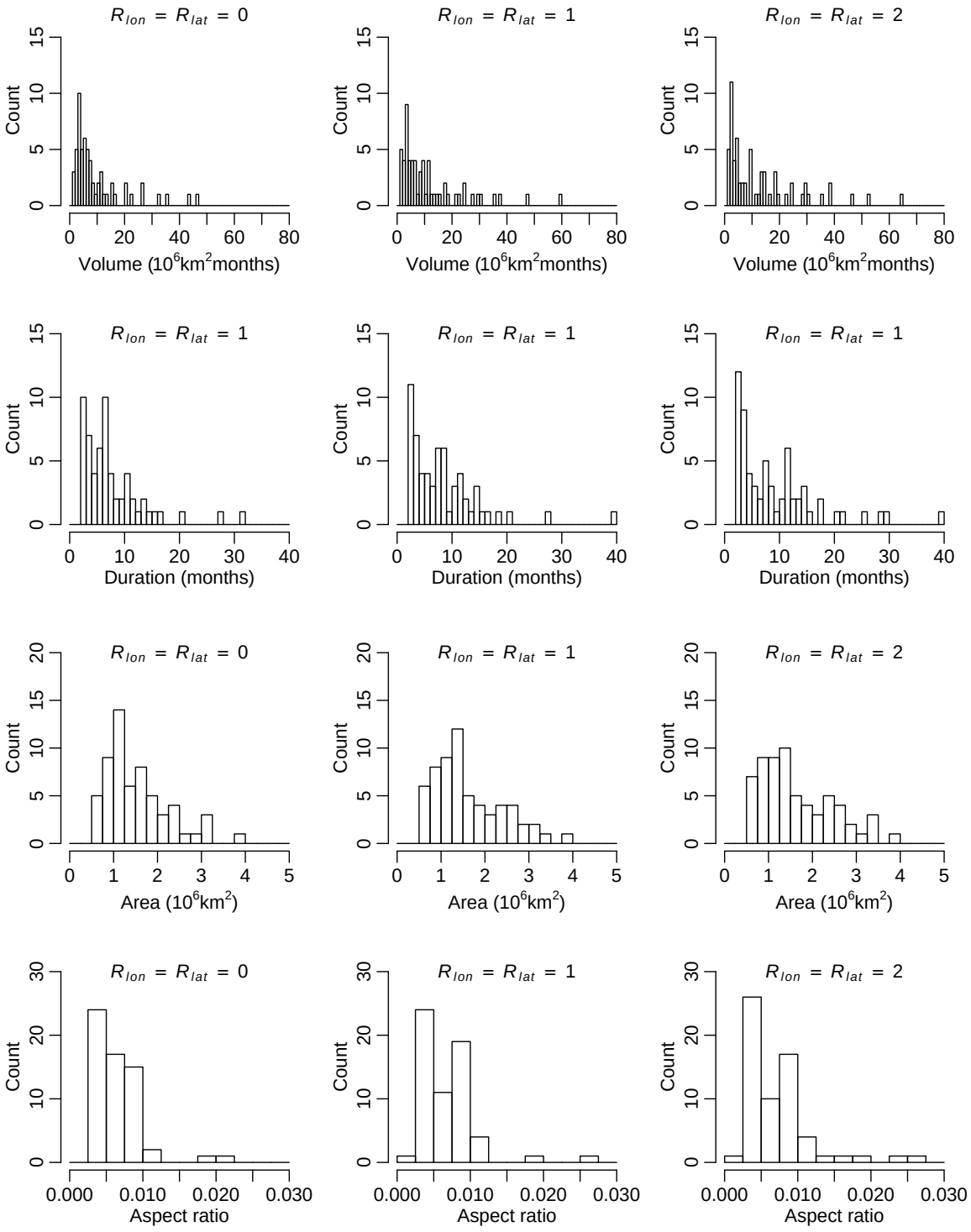

Figure 11: As Figure 7 but for SPI_{12}