A regionalisation approach for rainfall based on extremal dependence

Saunders, K. R.; Stephenson, A. G.; Karoly, D. J.

DOI
10.1007/s10687-020-00395-y

Publication date
2020

Document Version
Final published version

Published in
Extremes

Extremes
https://doi.org/10.1007/s10687-020-00395-y

A regionalisation approach for rainfall based on extremal dependence

K. R. Saunders[1] · **A. G. Stephenson**[2] · **D. J. Karoly**[3,4]

Received: 15 June 2019 / Revised: 28 August 2020 / Accepted: 4 September 2020 /
Published online: 07 October 2020

Abstract

To mitigate the risk posed by extreme rainfall events, we require statistical models that reliably capture extremes in continuous space with dependence. However, assuming a stationary dependence structure in such models is often erroneous, particularly over large geographical domains. Furthermore, there are limitations on the ability to fit existing models, such as max-stable processes, to a large number of locations. To address these modelling challenges, we present a regionalisation method that partitions stations into regions of similar extremal dependence using clustering. To demonstrate our regionalisation approach, we consider a study region of Australia and discuss the results with respect to known climate and topographic features. To visualise and evaluate the effectiveness of the partitioning, we fit max-stable models to each of the regions. This work serves as a prelude to how one might consider undertaking a project where spatial dependence is non-stationary and is modelled on a large geographical scale.

Keywords Clustering · Climate extremes · Spatial dependence · Extremal dependence

AMS 2000 Subject Classifications 60G70 · 62P12 · 62G32 · 62D05

✉ K. R. Saunders
K.R.Saunders@tudelft.nl

[1] Delft Institute of Applied Mathematics, Delft University of Technology, Delft, Netherlands

[2] Data61, CSIRO, Clayton, Victoria, Australia

[3] School of Earth Sciences, The University of Melbourne, Parkville, Victoria, Australia

[4] NESP Earth Systems and Climate Change Hub, CSIRO, Aspendale, Victoria, Australia

1 Introduction

The impacts of extreme rainfall and associated flooding can be observed on a scale that covers hundreds of kilometres. For example, the 2011 floods in Australia affected an area the size of France and Germany (Queensland Floods Commission of Inquiry 2012). Flooding on this scale is also not unprecedented, with further evidence that extreme rainfall and associated flooding can occur across large geographical scales given in Fig. 1. These historical instances establish the need to understand the spatial range of potential impacts from extreme rainfall. However, for many countries this understanding is lacking, particularly on daily and sub-daily scales.

Statistical models can be used to assess the spatial range of dependence between rainfall extremes, with a summary of some common statistical methods given in Davison et al. (2012). Of particular interest are max-stable processes, which provide a natural extension of univariate extreme value theory to extremes in continuous space with dependence (de Haan 1984; Schlather 2002). Modelling rainfall extremes in continuous space is desirable as the risk at locations without stations can be assessed. Max-stable processes also have strong mathematical justification for extrapolating outside the range of the observed data. Given this, these processes have been used in several studies of extreme rainfall (Dombry and Eyi-Minko 2013; Saunders et al. 2017).

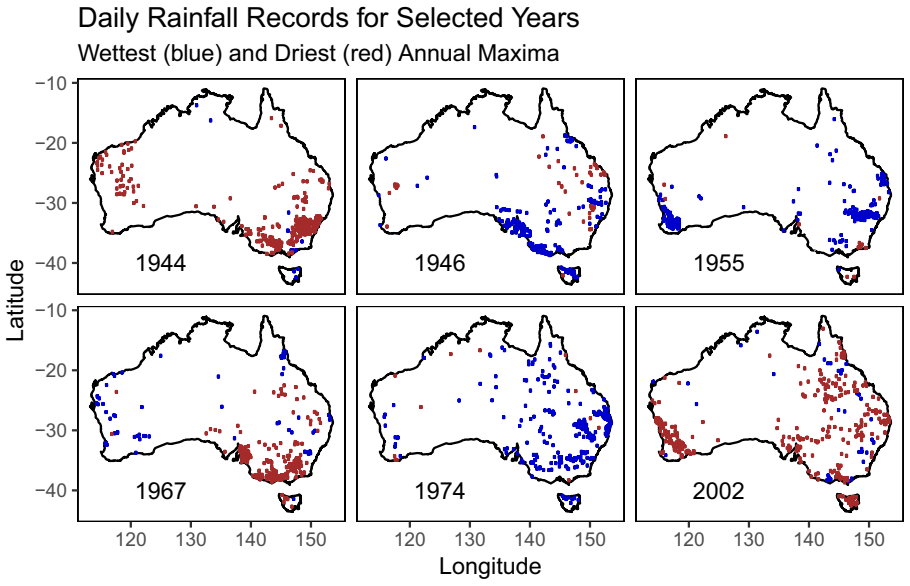

Fig. 1 For the given year, the plot shows the locations of stations at which the wettest annual maximum was observed (blue) and the driest (red). The years selected are the top three wettest (1946, 1955 and 1974) and top three driest (1944, 1967 and 2002) by proportion of stations. Note that observational periods do vary between stations. Stations are often clustered tightly in a given colour. These clusters can occur across large geographical scales

However, the parametric dependence structure of the max-stable process is often assumed fixed across a given domain for computational and mathematical simplicity (Oesting et al. 2017). Depending on the domain, a fixed dependence structure may be a reasonable modelling assumption. For a large geographical domain however, this assumption is likely to be poor. For example, Australia is one of the largest countries by area, with a diverse climate and complex topographic features (Stern et al. 2000; Risbey et al. 2009). Assuming a fixed parametric dependence structure is unlikely to yield meaningful results. This presents an obstacle to creating a parsimonious statistical model and reliably identifying which regions are likely to experience similar impacts from extreme rainfall.

Promising extreme value approaches are emerging that model non-stationarity within the dependence as a function of covariates (Huser and Genton 2016; Castro-Camilo et al. 2018, 2019). However, these methods are mathematically and computationally complex. As such they are prohibitive for many applied researchers in climatology and hydrology. To understand how the spatial range of dependence varies for rainfall extremes, a solution is therefore desired in which the method can be quickly implemented and in which the results lead to a simple interpretation.

To address this, we present a method for creating regionalisations of rainfall extremes, in which the regions are identified based on extremal dependence. Variations in the size and shape of these regions will indicate the spatial range of the dependence and whether the dependence behaviour is anisotropic. This knowledge can then be translated into insights for assessing and mitigating the potential impacts of extreme rainfall.

Regionalisations are common in flood frequency analysis and studies of hydrological extremes. Examples of different approaches to regionalisation based on extreme rainfall are given in Hosking and Wallis (1997), Carreau et al. (2017), Asadi et al. (2018) and Rohrbeck and Tawn (2020). For Australia, a regionalisation specific to rainfall extremes does not exist. However, there are regionalisations formed using topography and mean climate (Stern et al. 2000; CSIRO and Bureau of Meteorology 2015).

The regionalisation presented here is based on the clustering method presented in Bernard et al. (2013). In this method, a rank-based distance measure is used to cluster stations. This distance measure is related to bivariate extremal dependence via the F-madogram (Cooley et al. 2006). Using a rank-based distance is powerful, as no information about climate or topography is required to form spatially homogeneous clusters. This circumvents the challenge of variable selection. Additionally, we are free from distributional assumptions as the F-madogram can be estimated non-parametrically from raw maxima.

Where this paper extends the work of Bernard et al. (2013) is in the choice of unsupervised learning algorithm. In the original application, K-medoids was used for clustering. However, K-medoids is sensitive to point density. Additionally if there are too few clusters, K-medoids produces spurious clusters when used with the F-madogram distance. We demonstrate these undesirable features using simple examples. For station networks with varying point density, such as Australia, K-medoids is therefore ill-suited.

We propose using hierarchical clustering instead with the F-madogram distance. This ensures the clusters obtained are not affected by station density and are well informed by extremal dependence. The hierarchical nature of the algorithm also has an interpretation in terms of the changing strength of dependence. We demonstrate how the different clustering methods perform using daily rainfall stations in Australia. We show the serious consequences of incorrectly using K-medoids comparing with the results from the more robust hierarchical clustering. We also perform an additional classification step. This step converts the clusters from F-madogram space into a Euclidean space, giving a more intuitive spatial interpretation.

The resulting regionalisation generates valuable insights into the dependence of Australian rainfall extremes. We demonstrate this through a range of examples, highlighting features of climate and topography. We also show how the regions defined using a measure of partial dependence translate to the full dependence of spatial extremes. We achieve this by fitting max-stable models to the stations in each region. The results improve our understanding of the spatial range of extreme rainfall events, and how this range varies with increasing dependence strength.

The paper structure is as follows. In Section 2, the data are introduced. In Section 3, the clustering method is given, along with a discussion about how both the algorithm and dissimilarity affect the regionalisation. In Section 4, the classification step is outlined. In Section 5, max-stable processes are introduced and so are the steps needed to visualise the spatial range of extremal dependence. The regionalisation method is applied to Australian stations in Sections 6 and 7 highlights practical considerations for users.

2 Data

In this paper, we use the network of daily rainfall stations in Australia. These stations are mainly located near large cities and along the Eastern Australian coast, Fig. 2. In inland and more remote areas, there are far fewer stations. The station data are obtained from the quality controlled GHCN-Daily dataset (Durre et al. 2008, 2010 and can be accessed via the R package, rnoaa (Chamberlain 2017). However, we acknowledge the quality control, while thorough, is of a general design and is not targeted at identifying errors amongst extremal observations (Saunders 2018). For example, caution should be exercised when excluding observations flagged as outliers, as these observations may be extremes.

The Australian observations within GHCN-Daily are available via a reciprocal agreement with the Australian Bureau of Meteorology. The period we consider is restricted from 1910 to 2017. Prior to 1910 recording practices were not standardised throughout Australia.

The analysis is performed using the observed annual maximum rainfall. In extreme value approaches, this is referred to as block maxima. This is in contrast to peaks over threshold, where we are unconcerned with the date of the maxima within the yearly block. To ensure the quality of the observed maxima, we have restricted the data by only considering years which are 90% complete and stations at which there is a minimum of 20 years of observed maxima. This is necessary to ensure the quality

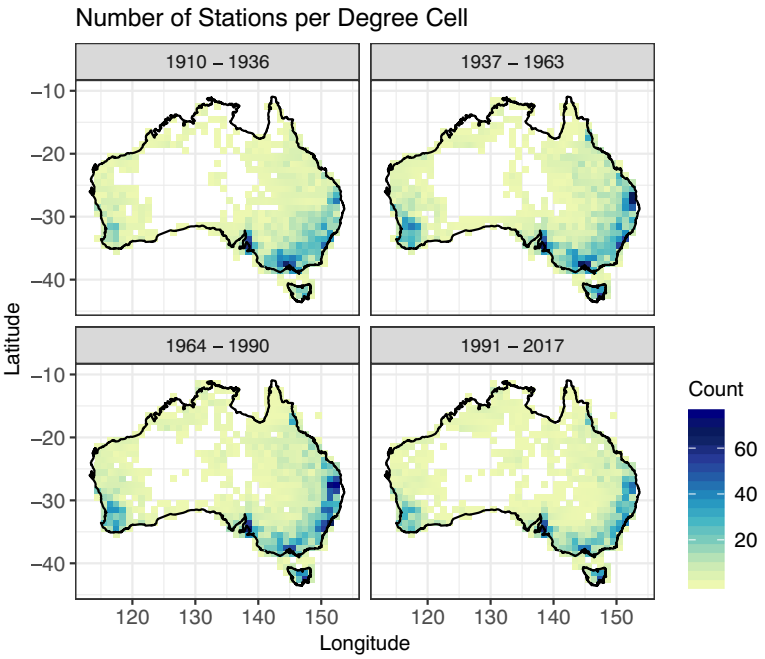

Fig. 2 The plot shows the number of stations within each one degree grid cell that have observations spanning the given time period

of any extreme value assumptions (eg. Coles 2001) and to limit the effects of missing maxima (eg. Haylock et al. 2000).

3 Clustering method

In the following section, we outline how to perform the clustering for the regionalistion. This includes describing the choice of the dissimilarity and choosing an appropriate clustering algorithm.

3.1 Clustering dissimilarity

A notion of dissimilarity (or similarity) between two points is required to apply clustering algorithms, with the type of dissimilarity chosen determining the cluster structure. For this application, following Bernard et al. (2013), we have chosen to use the F-madogram distance (Cooley et al. 2006)[1]. The F-madogram distance has an interpretation in terms of the pairwise dependence strength of extremes. The resulting cluster structure therefore inherits a meaningful, physical interpretation.

[1]The dissimilarity used in clustering can be a distance, but it does not necessarily need to satisfy the triangle inequality (Hastie et al. 2009).

3.1.1 F-madogram

The F-madogram (Cooley et al. 2006) links ideas of dependence in spatial statistics and dependence in extreme value theory. In spatial statistics a variogram (eg. Cressie 2015) is commonly used to understand the dependence between two locations in a stochastic process. However, for extremes the variogram is often undefined, as the distributions can be heavy-tailed and the variance is not finite. In contrast, the F-madogram, which is conceptually similar, is defined for heavy-tailed distributions.

Let $\mathcal{S} \subset \mathbb{R}^2$ and let $\{x_1, x_2, \ldots, x_n\}$ be the set of station locations for clustering. For $x_i \in \mathcal{S}$, define M_i as the random variable that represents the annual maximum of the daily rainfall at that station. Let the distribution function associated with M_i be $F_i(z)$. We can estimate $F_i(z)$ empirically via

$$\hat{F}_i(z) = \frac{1}{|\mathcal{Y}_i|} \sum_{y \in \mathcal{Y}_i} \mathbb{I}\left(M_i^{(y)} < z\right),$$

where $M_i^{(y)}$ is the annual maximum at station x_i in year y and $\mathcal{Y}_i$ is the set of all years for which there are annual maximum observations at x_i.

For stations $x_i \in \mathcal{S}$ and $x_j \in \mathcal{S}$, the F-madogram is given by the mean absolute difference (MAD) between two distribution functions and can be estimated non-parametrically using

$$\widehat{d}(x_i, x_j) = \frac{1}{2|\mathcal{Y}_{ij}|} \sum_{y \in \mathcal{Y}_{ij}} \left| \hat{F}_i\left(M_i^{(y)}\right) - \hat{F}_j\left(M_j^{(y)}\right) \right|,$$

where $\mathcal{Y}_{ij}$ is the set of years when both stations x_i and x_j have annual maximum observations. Note that $\mathcal{Y}_i$ and $\mathcal{Y}_{ij}$ may differ depending on missing observations.

Non-parametric estimation of the F-madogram avoids distributional assumptions and model fitting. This makes using this distance for clustering particularly powerful, as no external information about climate or topography is required and there is no need for variable selection. However, this assumes that annual maxima are stationary in time. It may be necessary to remove trends depending on the application, such as in the case of temperature extremes (Bador et al. 2015).

3.1.2 Bivariate extreme value distribution

The link between the F-madogram and extreme value theory provides the cluster structure with a physical interpretation in terms of the dependence of extremes. For any pair of stations, x_i and x_j, if the distribution of $(M_i,\ M_j)$ is well approximated by a bivariate extreme value distribution then

$$\mathbb{P}\left(M_i \leq z_i,\ M_j \leq z_j\right) = \exp\left\{-V_{ij}\left(\frac{-1}{\log F_i(z_i)},\ \frac{-1}{\log F_j(z_j)}\right)\right\}, \tag{1}$$

where the exponent measure $V_{ij}(a, b)$ is given by

$$V_{ij}(a, b) = 2\int_0^1 \max\left(\frac{w}{a}, \frac{1-w}{b}\right) \mathrm{d}H_{ij}(w),$$

and H_{ij} is any distribution function on $[0, 1]$ with expectation equal to 0.5 (eg. Resnick 1987; de Haan and Ferreira 2006).

In the special case where $z_i = z_j = z$, the bivariate extreme value distribution of Eq. 1 reduces to

$$\mathbb{P}(M_i \leq z, M_j \leq z) = \left[\mathbb{P}(M_i \leq z)\mathbb{P}(M_j \leq z)\right]^{V_{ij}(1,1)/2},$$

where

$$V_{ij}(1, 1) = \theta(h)$$

and $\theta(h)$ is the extremal coefficient, with $h = \|x_j - x_i\|$ (eg. Naveau et al. 2009). The range of $\theta(h)$ is $[1, 2]$, where the lower bound of the interval corresponds to dependence of M_i and M_j, and the upper bound conversely indicates independence. The value of $\theta(h)$ therefore provides an indication of the partial dependence between the maxima at the two locations x_i and x_j when $z_i = z_j = z$.

The F-madogram dissimilarity can be expressed as a function of the extremal coefficient (Cooley et al. 2006)

$$d(x_i, x_j) = \frac{\theta(h) - 1}{2(\theta(h) + 1)},$$

where the range of $d(x_i, x_j)$ is $[0, \frac{1}{6}]$. Therefore when it is suitable to approximate the pairwise distribution of annual maxima with bivariate extreme value distributions, clusters formed using the F-madogram distance will have an interpretation in terms of partial dependence of extremes.

Equally, we could have used $\theta(h)$ for the clustering dissimilarity. However, the F-madogram as a mathematical object can be estimated independently of distributional assumptions and therefore of extreme value assumptions. As such, it offers a more flexible choice for the dissimilarity.

3.1.3 Practicalities of missing dissimilarities

All pairwise dissimilarities are required for clustering. However, unlike gridded datasets, observational periods at two stations may not overlap due to missing data. Additionally, if the number of overlapping years is small, the F-madogram distance cannot be estimated reliably. Therefore to increase the amount of station data available, particularly in sparse regions, missing distances were interpolated.

At large Euclidean distances, we expect the maximum rainfall observed at pairs of stations to be close to independent. Given this, these missing dissimilarities were interpolated as $\frac{1}{6}$. This is a reasonable assumption and greatly reduces the missing dissimilarities. Also, for a station that has been renamed, the Euclidean distance between stations may be 0 and then the missing F-madogram distance is interpolated as 0.

For the remaining missing dissimilarities we fit regional linear models to the logarithm of the Euclidean distance. From this model we predicted missing distances, and while these predictions do not approximate local dependence well, they do serve as a reasonable approximation of overall dependence. At very small Euclidean distances predictions could take negative values, so the maximum of the predicted F-madogram

distance and zero was taken. Given the missingness present in our data and based on basic diagnostics, we were satisfied with the variance-bias trade off between including more stations using interpolation compared with only using F-madogram distances based on overlapping station data. We do acknowledge that more sophisticated interpolation methods are possible, but we do not expect them to change the overall result.

3.2 Clustering algorithm

In the previous section, we provided the necessary information about estimating the F-madogram distances and understanding the physical meaning behind the clustering structure. In the following section, we discuss the choice of clustering algorithm. We contrast cluster structures generated using K-medoids and hierarchical clustering, highlighting subtle features of these different algorithms. In particular, we discuss the suitability of these algorithms for our application.

3.2.1 K-medoids

In the clustering application of Bernard et al. (2013), K-medoids clustering was applied with the F-madogram distance. In K-medoids, the goal is to find K clusters such that the sum of dissimilarities relative to a representative point within each cluster is minimised. This representative point is known as the medoid. Denote the medoids $\{m_k \,|\, k = 1 \ldots, K\}$ and their associated clusters $\{C_k \,|\, k = 1 \ldots, K\}$, where $K \leq n$. To partition the points we can use the PAM algorithm (Kaufman and Rousseeuw 1990), see Algorithm 1.

Like many clustering algorithms, PAM converges to a local minimum, but not necessarily the global minimum. It is therefore advisable to repeat PAM with different initialisations of medoids to help ensure the consistency within the performance of the algorithm. We do not discuss how to select K optimally for K-medoids here, but implementations of various methods can be found in Charrad et al. (2014).

3.2.2 Implicit assumptions

Within unsupervised learning there is no true structure, however, we often still have implicit assumptions about the structure form. For our application, we have the expectation that two stations that are far away in Euclidean space will be clustered differently as the extremes at these stations are independent. We also have the expectation that stations that are geographically close will be clustered together as they are likely to be highly dependent.

Consider the two examples shown in Figs. 3 and 4. In each of these examples the structure is known and there are two groups of points. However, K-medoids clustering does not recover the two groups correctly. We have not used the F-madogram distance in these examples. Instead, it is more intuitive to think in Euclidean space, so the distance used is

$$d(x_i, x_j) = \max\left(\|x_j - x_i\|, 1\right),$$

Algorithm 1 K-medoids clustering.

procedure PARTITIONING AROUND MEDOIDS
  Choose the number of clusters, K
  Randomly select K points in $\mathcal{S}$ as the initial medoids, $\{m_k \mid k = 1 \dots K\}$
  Determine the closest medoid to each point
  Cluster points that share the same closest medoid
  for k in $1, \dots, K$ **do**
    Find the point within that cluster, C_k, such that

$$m_k^* = \underset{x_i \in C_k}{\operatorname{argmin}} \sum_{x_j \in C_k} \hat{d}(x_i, x_j).$$

    This point minimises the sum of
    dissimilarities within that cluster.
    if $m_k^* \neq m_k$ **then**
      Update the medoid so that $m_k = m_k^*$
    end if
  end for
  if Any of the medoids were updated **then**
    Repeat steps 4. – 12.
  end if
end procedure

where $\|\cdot\|$ is the Euclidean distance. Here the maximum value this distance can take is restricted to 1, in order to mimic the finite range of the F-madogram distance.

The example in Fig. 3 shows that K-medoids clustering is sensitive to the spatial density of points. The location of the medoids, the representative object within each cluster, is biased toward regions of higher spatial point density. This causes points in

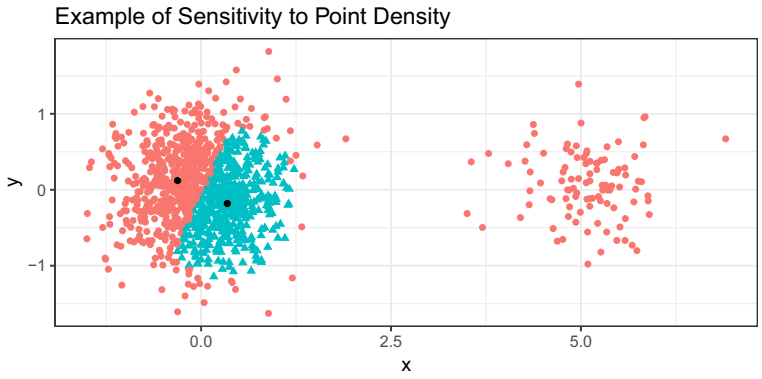

Fig. 3 Example of K-medoids clustering showing that the assignment of points to clusters is sensitive to the spatial density of points

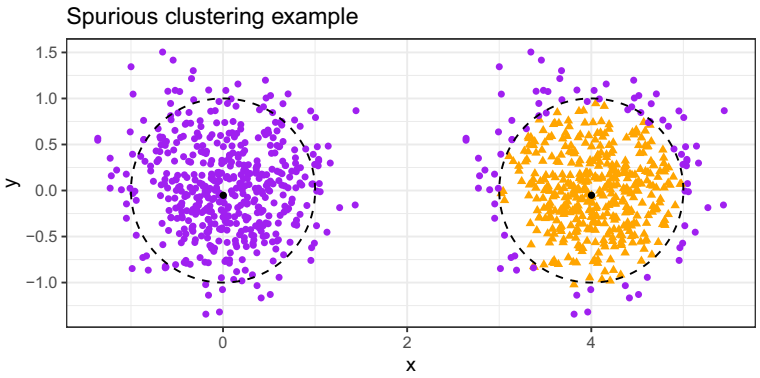

Fig. 4 Example of K-medoids clustering showing undesirable clustering behaviour when points are equidistant from all medoids

the smaller group to be clustered in an undesirable way, as our implicit assumptions is that points that are further away should be clustered differently. It is not until the value of K is increased to four or more that the second group is identified and separated. Further, intelligent initialisation of medoids did not recover the two groups. Under the optimisation this is not unexpected. If both medoids are posititioned in the denser region then the overall contribution of dissimilarities to the cost function is smaller than if a medoid was in each group. The points in the sparser region are therefore less important under the optimisation. This implies for datasets where the spatial density of points in F-madogram space varies, that the cluster structure will not have a meaningful interpretation in terms of extremal. Gridded datasets would be more resilient to this problem, provided proper consideration is given to land-sea and domain boundaries.

Of greater concern is that K-medoids can produce spurious clustering, as shown in the example of Fig. 4. Here, a circle of radius one is drawn around each medoid. Points outside of these circles are of distance 1 to either medoid. Under the optimisation, these points can be assigned randomly to either cluster without penalty. Insidiously, all these points are labeled the same due to a numeric ordering within the standard algorithm. Groups of points can therefore appear to be clustered meaningfully, even though they are not. Consequently, if there are too few medoids, points will be assigned randomly. It is tempting to introduce a distance penalty to help prevent this. Such a penalty would need to be considered relative to F-madogram space, which becomes tricky as we do not know the underlying structure and weak dependence can be present at large Euclidean distances. A distance penalty will also not fix the issue of medoid locations being biased towards regions of higher point density.

These examples demonstrate that the selection of the clustering method needs to be evaluated relative to the dataset to ensure the clustering is meaningful. Given that the Australian station network is highly variable in terms of spatial point density, it is highly unlikely that a cluster structure obtained using K-medoids and the

F-madogram distance will be informative in terms of extremal dependence. As such an alternative method is needed for clustering.

Two of the other most common methods are K-means and hierarchical clustering. K-means however is subject to the same failings demonstrated in Figs. 3 and 4. Further, K-means is also not an appropriate choice given Euclidean assumptions and a standard algorithm implementation in terms of points not distances (Hastie et al. 2009). Hierarchical clustering in contrast, can be used with an F-madogram distance to produce meaningful structures in terms of extremal dependence.

3.2.3 Hierarchical clustering

In hierarchical clustering an ordered sequence of partitions is created. This hierarchy of partitions has a natural intuition for our application, and can be interpreted as partitions of points based on strong dependence to weaker dependence. Graphically, this ordered sequence of partitions can be represented using a dendrogram. Let each point be its own cluster (leaf). Branches in the dendrogram are formed by successively combining leaves and other branches until all points are grouped together. For each merge, a new partition of the points is induced. The successive merging of branches therefore creates the ordered partition of points.

To decide how branches should be merged the definition of distance needs to be extended from between two points to include the distance between two groups of points. This is known as the linkage criterion (Murtagh 1983, 2014). Let C_k and $C_{k'}$ be two different clusters of points. We use the average linkage criterion

$$d(C_k, C_{k'}) = \frac{1}{|C_k|\,|C_{k'}|} \sum_{x_k \in C_k} \sum_{x_{k'} \in C_{k'}} d(x_k, x_{k'}).$$

Using the linkage criterion, we can construct an agglomerative dendrogram using Algorithm 2.

Algorithm 2 Hierarchical clustering.

1: **procedure** AGGLOMERATIVE
2: Let each point form its own cluster
3: Merge the pair of clusters with the smallest dissimilarity (ties are broken randomly)
4: Update the dissimilarities relative to the new cluster using the linkage criterion
5: Repeat steps 2–4, until all points are combined in a single cluster
6: **end procedure**

To determine an assignment of points into clusters, we need to select one of the partitions generated by the dendrogram. This can be done by cutting across the tree at a height h, and assigning the points in same branch to the same the cluster.

Equivalently, we can specify the number of clusters, K, and choose the cut height that corresponds to this number of clusters.

The height of the cut should be made with reference to the desired strength of association between the clusters, with the height at which the branches are fused determining the strength of association between two clusters. Therefore for two branches joined at the bottom of the tree, this suggests the points in these branches are strongly associated. For branches joined at the top of the tree, this suggests a much weaker association between the groups of points. Standard methods for choosing the cut height, or equivalently the number of clusters, include the gap statistic (Tibshirani et al. 2001). This method is not specific to hierarhcial clustering though and therefore should be used cautiously given the implicit clustering assumptions highlighted earlier. Equally valid, is choosing a cut height based on user knowledge. We chose the cut height by considering user knowledge in combination with visualising the full extremal dependence, as detailed later in Sections 5 and 6.5.

In hierarchical clustering, using a different linkage criterion will induce a different dendrogram and consequently different clusters. The average linkage criterion successfully recovers the two groups shown in Figs. 3 and 4. However, this is not the case for many standard linkage rules. Therefore a caveat of this method is that caution is needed in selecting an appropriate linkage criterion for the application.

4 Classification

Hierarchical clustering is performed in F-madogram space, however for most applications the regions need to be defined in Euclidean space. As such, an additional classification step is needed. This step is also necessary to classify locations without a station and to identify boundaries between two clusters for predictive purposes.

We have used a weighted k-nearest neighbour classifier (wk-NN) (Dudani 1976) to classify grid points covering our domain and to convert the clustering to a regionalisation. We chose the wk-NN method as it is non-parametric, based on minimal assumptions, and can form non-linear boundaries.

In standard k-nearest neighbour classification (k-NN) (eg Hastie et al. 2009), points are classified similarly to the majority of their k-nearest neighbours without using weights. However, the relationship between the F-madogram and Euclidean distance is not linear, so a weighted classifier is more appropriate for this application (Samworth 2012). Here we use an inverse weighted kernel. For classification details see Algorithm 3.

There is a variance bias trade-off when selecting the number of nearest neighbours, k_{nn}. However, when the clusters are well separated in Euclidean space there are a large range of suitable k_{nn} values. Considerations for this specific application are that we require k_{nn}, such that erroneously clustered stations do not impact the classification, and smaller clusters of only a few stations are not engulfed by a larger cluster and its label. It can be difficult to find an automated metric that will respect this latter criteria. Given the large range of suitable values, through visualisation and user knowledge, we used a value $k_{nn} = 15$.

Algorithm 3 Classification.

The stations, $\mathcal{S}$, from clustering will form the training points for the classification
For $x_i \in \mathcal{S}$, define $l(x_i)$ to be the label assigned with x_i
Grid the domain for classification
procedure WEIGHTED k NEAREST NEIGHBOURS
    Choose the number of nearest neighbours, k_{nn}, where $k_{nn} \leq n$
    for each grid point, g **do**
        According to Euclidean distance, get the $k_{nn} + 1$ nearest neighbours to g in $\mathcal{S}$
        Let the furthest of these neighbours be n_f
        Let the set, $\mathcal{N}$, contain the other nearest neighbours, $\{n_j \mid j = 1, \ldots k_{nn}\}$
        for each of the nearest neighbours, $n_j \in \mathcal{N}$ **do**
            Standardise the Euclidean distances between n_j and g

$$s(n_j) = \frac{\|g - n_j\|}{\|g - n_f\|}.$$

            We used an inverse weighted kernel to weight each neighbour.
            Get the associated weight for the neighbour, n_j,

$$w(n_j) = s(n_j)^{-1}.$$

        end for
        Let $\mathcal{C}$ be the set of labels associated with the neighbours in $\mathcal{N}$
        Determine the label of the majority of the weighted k_{nn} nearest neigh bours

$$l^* = \underset{l^* \in \mathcal{C}}{\operatorname{argmax}} \left(\sum_{l^* \in \mathcal{C}} \sum_{j=1}^{k_{nn}} w(n_j) \mathbb{I}(l(n_i) = l^*) \right),$$

        Classify $l(g)$ with the majority label, l^*
    end for
end procedure

5 Visualising dependence

Part of our motivation for creating this regionalisation was to understand the range of spatial dependence and scale of potential impacts from extreme rainfall. However, the distance used only partially reflects the full extremal dependence. Therefore to consider whether a partitioning forms an appropriate regional summary relative to the full dependence structure, we will fit max-stable processes to the stations in each region.

Max-stable process provide a natural extension from univariate extreme value theory and the GEV distribution, to models for extremes in continuous space with

dependence (de Haan 1984; Schlather 2002). The canonical example of these processes is the Smith model (Smith 1990). This model offers an intuitive storm shape interpretation, where a storm shape is scaled by a storm intensity and the pointwise maxima over infinitely many of these scaled-storms forms a realisation of the max-stable process. Mathematically

$$Z(x) \stackrel{d}{=} \max_{i \geq 1} \zeta_i W(x - U_i; 0, \Sigma), \quad x \in \mathcal{X} \subset \mathbb{R}^2,$$

where $\{\zeta_i : i \geq 1\}$ are points from a Poisson process on $(0, \infty)$ with intensity $\zeta^{-2} \mathrm{d}\zeta$ and $W(\cdot; 0, \Sigma)$ is a two-dimensional Gaussian density, with mean zero and covariance matrix Σ. Here, U_i are points of a homogeneous Poisson process defined on $\mathbb{R}^2$ that provide random translations of bivariate Gaussian density function. A visual representation of this process in 1-dimension is given in Fig. 5. The univariate marginals of this max-stable process are assumed to follow a standard Fréchet distribution.

The Smith model is used here due to its simplicity and as the dependence structure of this process is Gaussian. We can therefore visualise the dependence in two-dimensional

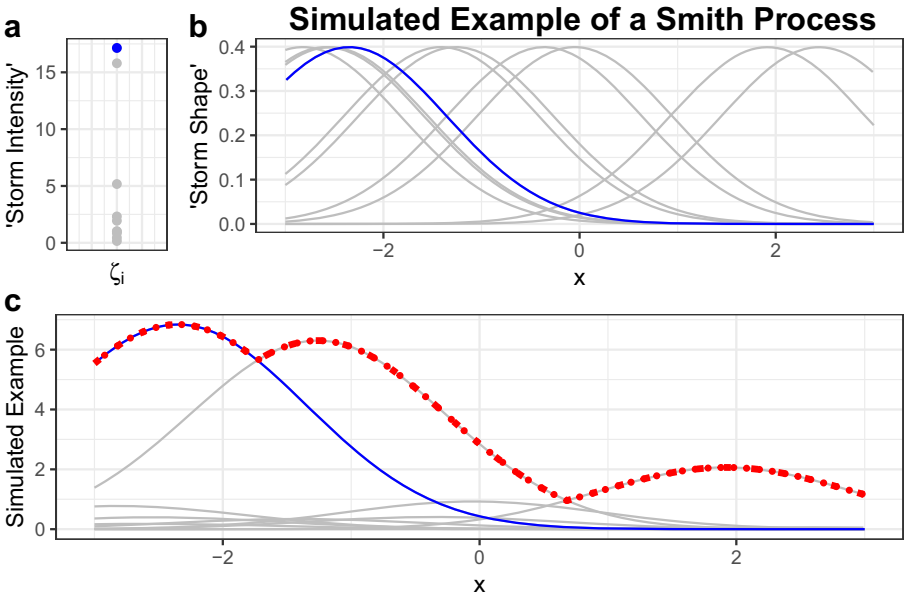

Fig. 5 This figure shows a visual example of each of the components that comprise the Smith model in one-dimension. Figure (**a**) shows points simulated from the inhomogeneous Poisson process, $\{\zeta_i\}$. Figure (**b**) shows a standard Gaussian density subject to random translations given by U_i. Figure (**c**) shows the product of figures (**a**) and (**b**), with an example carried through all figures shown in blue. The resulting simulation of a max-stable process is shown in red, and is given by the pointwise maxima over all scaled 'storm-shapes' shown in gray. Here only finitely many simulations of the index i are shown, however, this is all that is necessary to produce a simulated example from the Smith model (Schlather 2002)

Euclidean space using ellipses. The direction and size of these ellipses has a natural interpretation in terms of anisotropy and the range of the dependence.

For the Gaussian density, the probability of a point, x, lying within a radius, r, of the mean is given by the chi-squared distribution with two degrees of freedom

$$\mathbb{P}(\,||x-\mu|| < r) = 1 - \exp\left(\frac{-r^2}{2}\right).$$

For our elliptical curves, we have chosen r to correspond to the 1% level curve, for which $r \approx 3$. However, within the formulation of the Smith model the mean is zero as the Gaussian is subject to random spatial translations. Therefore to centre our the elliptical curves, we use the coordinate of the median longitude and median latitude of all suitable stations in the region, x_0. The parameterisation of the ellipses is then given by

$$x = x_0 + r(\cos\theta, \sin\theta)M,$$

where M is obtained from the Choleski decomposition of the covariance matrix, $\Sigma = M^T M$.

In general, if the partition is a good representative summary then we expect that the ellipses will have minimal overlap. If the ellipses were to overlap, this could indicate that points in the intersection could reasonably have been assigned to either cluster and there may be too many clusters. If we have too few clusters, then more ellipses could be added to summarise dependence.

To fit the Smith model we use composite likelihood, see Padoan et al. (2010) for details. In composite likelihood, the product over bivariate likelihood functions is optimised to obtain parameter estimates. Composite likelihood is used as it is not possible to optimise the full likelihood in higher dimensions where there are large numbers of stations (Castruccio et al. 2016; Huser et al. 2019). As we are primarily interested in the dependence parameters, we first fit the marginal distributions using standard maximum likelihood and standardise our marginals, prior to fitting the dependence parameters using composite likelihood.

We acknowledge that the Gaussian storm shape in the Smith model is a crude approximation of physical rainfall and there are other other max-stable processes we could have chosen (see Dey and Yan 2016). However, as we wish to visualise the full dependence, the Smith model serves as a useful exploratory tool. Additionally the code for fitting a Smith model with anisotropy is readily available in the SpatialExtremes package (Ribatet 2015), so the research and method is easily reproducible by others. However, we caution that appropriate starting values are often necessary to ensure convergence of the optimisation routine. We found fitting to different subsets of stations and then comparing parameter estimates was useful for identifying models that did not converge and could then be used to provide intelligent estimates of the starting values. In particular we were alerted to one of the dependence parameters being very small and the optimisation routine becoming stuck at the boundary of the parameter space.

6 Results

6.1 Hierarchical clustering compared with *K*-medoids

To highlight the impact of the choice of clustering algorithm, we have clustered stations in Southwest Western Australia using both hierarchical clustering and K-medoids, Fig. 6. Under hierarchical clustering, we observe clearer separation of the clusters in Euclidean space. This improved cluster cohesion is a benefit of the hierarchical algorithm having an agglomerative (bottom up) approach.

We also note that under hierarchical clustering, clusters can consist of a single station. Therefore to compare the clustering under the two different algorithms, we have chosen realisations where there are 8 core clusters that contain 10 or more stations. Single station clusters are an advantage of hierarchical clustering. In Fig. 6, clusters containing a single station strongly suggest an issue with the underlying quality of the station data. As these single station clusters can not be attributed to differences in local topography or to weak dependence relative to the cut height, as might be the case in sparse regions.

The ability of hierarchical clustering to have clusters of smaller size means that groups of stations with weaker dependence are not amalgamated into a larger groups at the expense of the overall cluster cohesion (see Fig. 3). It also prevents the occurrence of stations being clustered spuriously (see Fig. 4). In Fig. 6, we observe the effects of spurious K-medoids clustering as there is a large geographical separation

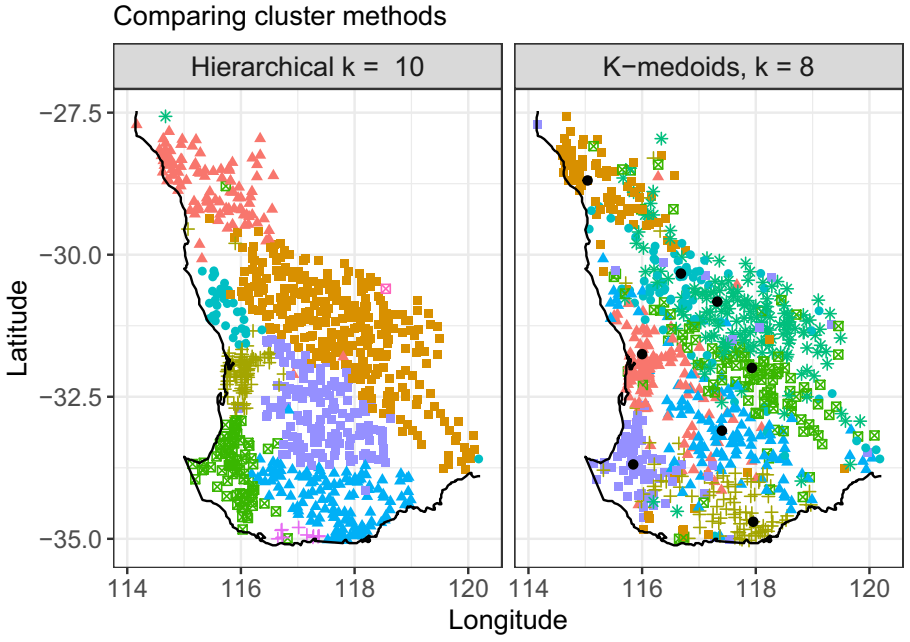

Fig. 6 Comparison of hierarchical clustering and K-medoids clustering for a set of stations in Southwest Western Australia

between some stations and their respective medoids. For example, the blue triangles at (116, -31) and the brown squares at (117, -35). For these reasons, we find hierarchical clustering superior for this application and use this method for the remainder of the paper.

6.2 Classification

Figure 7 shows the classification from the hierarchical clustering for a value of $k_{nn} = 15$. Due to the quality of the original clustering in F-madogram space and separation of clusters in Euclidean space there was very little difference for higher k_{nn} values. However, classification does offer the advantage that we have regional boundaries and do not need to visualise large numbers of points. The user may choose to omit single stations clusters at this step if the underlying case is bad data quality, although we have not done that here.

6.3 Ordered partitions

We mentioned earlier that one of the benefits of hierarchical clustering is that an ordered sequence of partitions is generated. In Figure 8, we show the evolution of these partitions for a range of cut heights for Southwest Western Australia. We observe that at the lower cuts heights that the regions are small in size. While at higher cut heights, where the dependence between clusters weakens, these smaller regions are amalgamated to form larger regions. Visualising the evolution of these ordered partitions helps our understanding of how the size of these regions changes with increasing strength of extremal dependence.

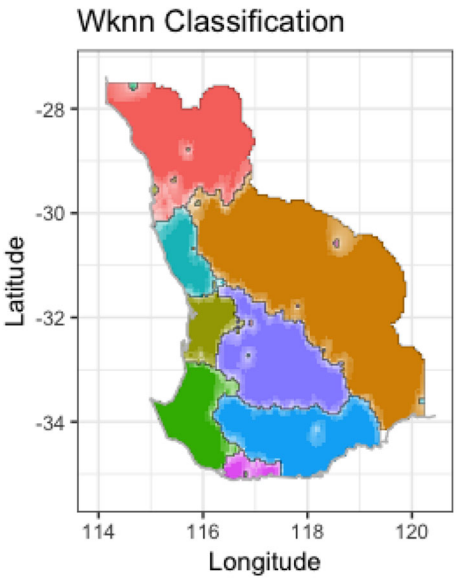

Fig. 7 Weighted k nearest neighbour classification showing cluster boundaries from the hierarchical clustering in Fig. 6

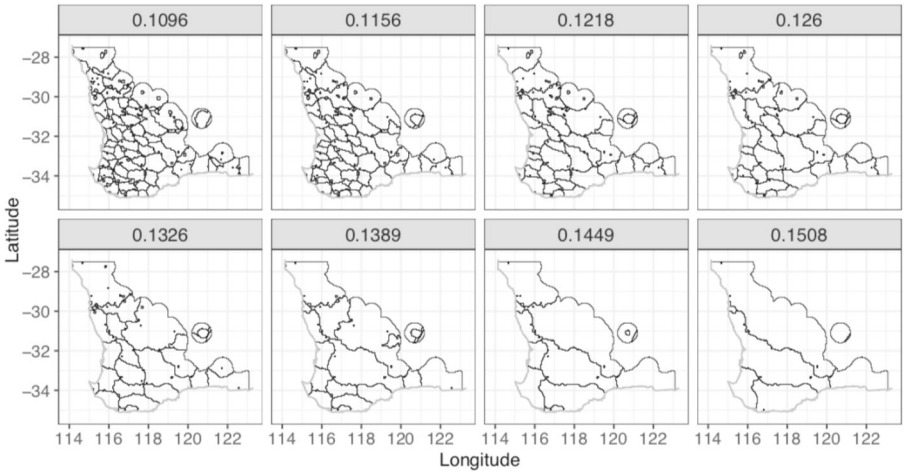

Fig. 8 Different regionalisations of Southwest Western Australian created using different cut heights in hierarchical clustering. The cut height is given in the facet label

Additionally the size and direction of the regions can then be interpreted relative to known climate or topography. We observe here that coastal clusters are generally smaller indicating that extreme rainfall is being driven by convective rainfall in these areas (Risbey et al. 2009). Whereas further inland, the size of clusters is larger, particularly as dependence weakens, and orientation of these clusters is consistent with the movement of frontal systems (Risbey et al. 2009).

6.4 Meaningful cut heights

While having the hierarchy of partitions is useful, often a single realisation of the clustering is desired. In this instance, it is important to consider how cut heights in F-madogram space translate to Euclidean space. Figure 9, shows a plot of Euclidean distance against the F-madogram distance for all pairs of stations in Southwest Western Australia. At low cut heights, the F-madogram distance changes rapidly relative to very small changes in Euclidean distance. At high cut heights, large changes in Euclidean distance are observed for small changes in F-madogram distance. Therefore there is a range of moderate cut heights that will translate into meaningful partitions of our stations in terms of extremal dependence in Euclidean space. For Fig. 9, suitable cut heights might be between 0.1 and 0.15. The cut height should therefore be chosen based on the desired application and the desired strength of extremal dependence.

6.5 Visualisation of full dependence

To understand how our regionalisation is related to the full extremal dependence, we have taken the additional step of fitting a Smith model. The full extremal dependence of each region can then be visualised using elliptical level curves. Similarity of the

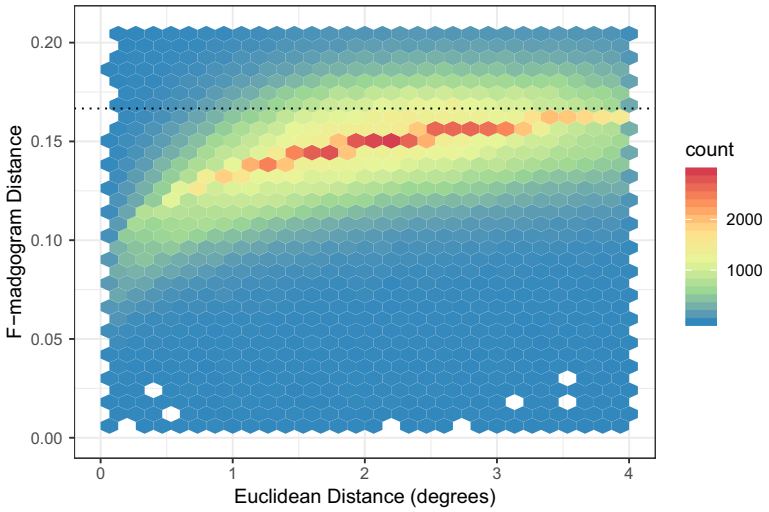

Fig. 9 Plot of the F-madogram distance relative to Euclidean distance. Given the number of pairs we have binned the data instead of showing a scatter plot. Note that the empirical estimator for the F-madogram can take a value above the theoretical range of $\frac{1}{6}$, shown with the dotted line

elliptical curves in different regions indicates similarity of the estimated dependence parameters.

An example of the elliptical level curves is shown in Fig. 10. We observe that the ellipses have optimally partitioned the domain, as no further ellipses could be added or removed. To be confident in this conclusion we have bootstrap sampled the stations and repeated the fitting to visualise the uncertainty in our dependence parameters. We found fitting max-stable models of this type to be useful in deciding the number of clusters and to identify which regions can reasonably be modelled using the same dependence structure. We also develop an intuition for which covariates would be necessary if a non-stationary dependence structure was used and which pair weights would be important in the optimisation.

6.6 Physical Interpretation

The example of Southwest Western Australia has served to highlight different aspects that need to be considered when producing a regionalisation. For this same cut height, we have shown the regionalisation for the whole of Australia in Fig. 11. Note we did not attempt to classify locations that were far from station locations.

We would like to draw attention to specific aspects within this figure where the regionalisation method has performed well. Figure 12 shows examples where the clustering respects that stations are geographically separated by water. Figure 13 shows how the regionalisation performs relative to orography.

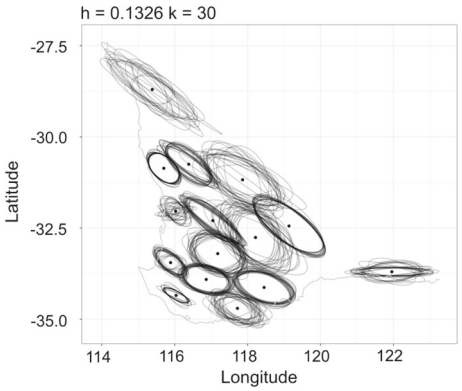

Fig. 10 For a regionalisation generated with a cut height of approximately 0.13, the full dependence is visualised using elliptical level curves. The black points show the median of the stations in that region and elliptical centres

Orographic features are known to strongly influence rainfall. In Australia there is a mountain range that runs up the Eastern Australia coast. We see this orographic feature respected in Fig. 13. There is a clear differentiation between clusters located on the coastal side of the range and those inland. This again reflects differences in the drivers between extremes in coastal areas compared with inland areas (Risbey et al. 2009).

7 Limitations

7.1 Dry regions

The F-madogram distance has interpretation in terms of the partial dependence of extremes provided the extreme value theory assumptions are reasonable. However

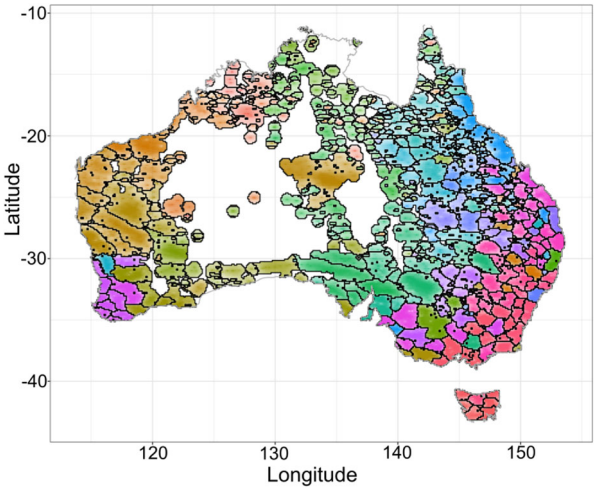

Fig. 11 A regionalisation generated with a cut height of approximately 0.13. Here the colours serve only to distinguish between regions

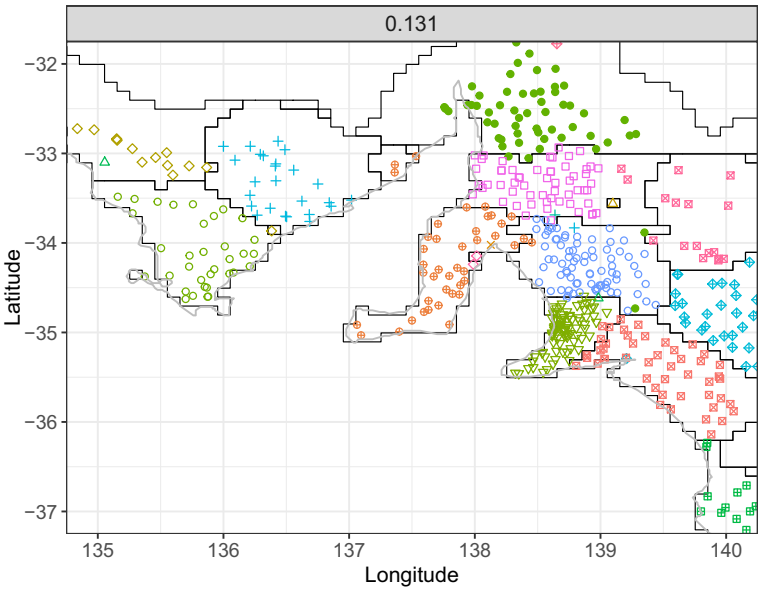

Fig. 12 An example demonstrating that the clustering respects the geographical separation of stations by water. The black lines show the regions, and the shape and colour of the points indicate which stations were clustered similarly

for drier regions, such as parts of inland Australia and Northern Australia, where there is less rainfall, these assumptions are generally invalid (Min et al. 2013). As a consequence the clustering will lack the interpretation in terms of extremal dependence, impacting the related visualisation of the clustering in Euclidean space, and we observe this in Fig. 11. Therefore stations located in dry regions should be considered critically in this kind of analysis.

7.2 Partial dependence

For a given regionalisation, it is tempting to assume that within each region we can assume a fixed dependence structure in our statistical models. However, as acknowledged, the F-madogram is only a measure of partial extremal dependence, not the full extremal dependence. For regions that encompass orography, a single dependence structure is unlikely to be appropriate (eg. Huser and Genton 2016; Oesting et al. 2017).

We observe this to be the case for regions in Tasmania, Fig. 14. At a higher cut height (approximately 0.13), where partial dependence is weaker within clusters, there is no consensus in the size and orientation of the ellipses for regions that encompass orography. At the lower cut height however (approximately 0.11), where the dependence within clusters is stronger, there is consensus in our fitted models. Cut heights therefore need to be chosen with respect to the given application.

We also note that the location of cluster boundaries at the lower cut height better respects the orography, Fig. 15. This regionalisation is consistent with assertions by

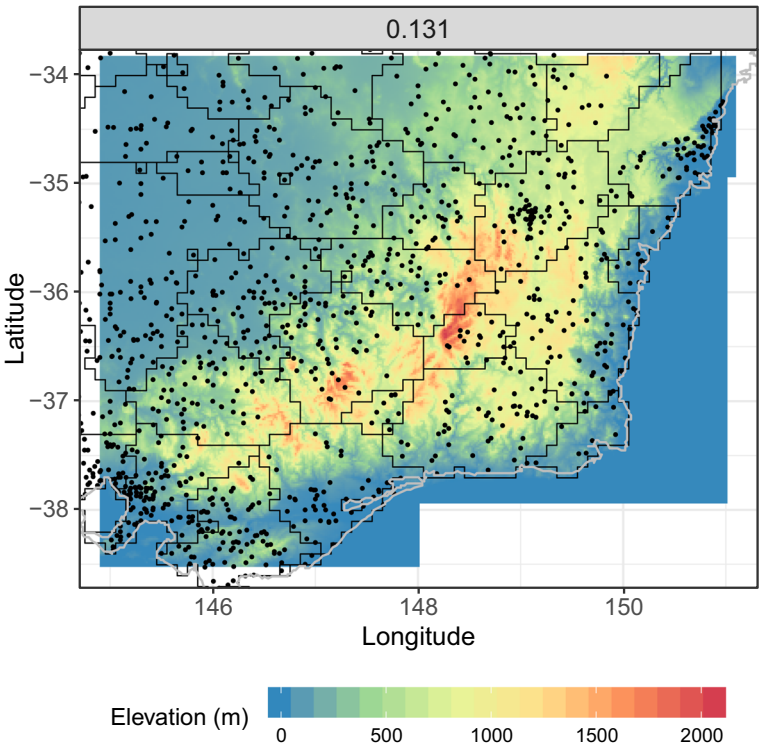

Fig. 13 An example demonstrating that the clustering respects the location of the Great Dividing Range, a mountain range in Australia. Here black lines show the regions and stations are shown as black points for reference

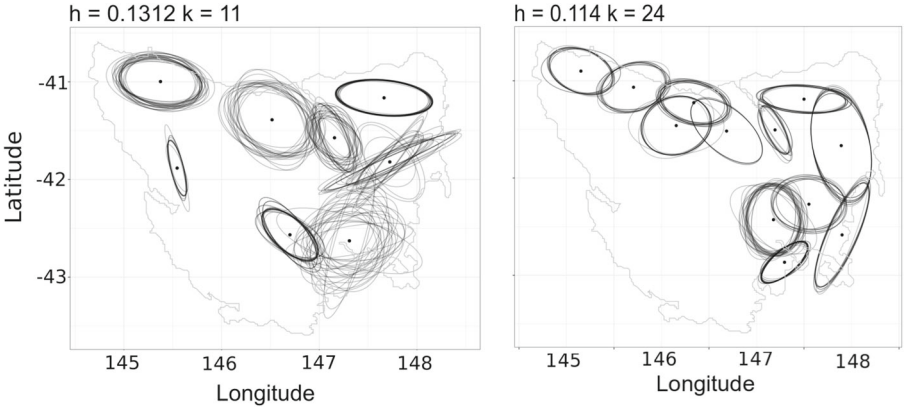

Fig. 14 Visualisation of the full dependence for two different regionalisation. The left figure was generated with a cut height of approximately 0.13 and the right figure with a cut height of approximately 0.11

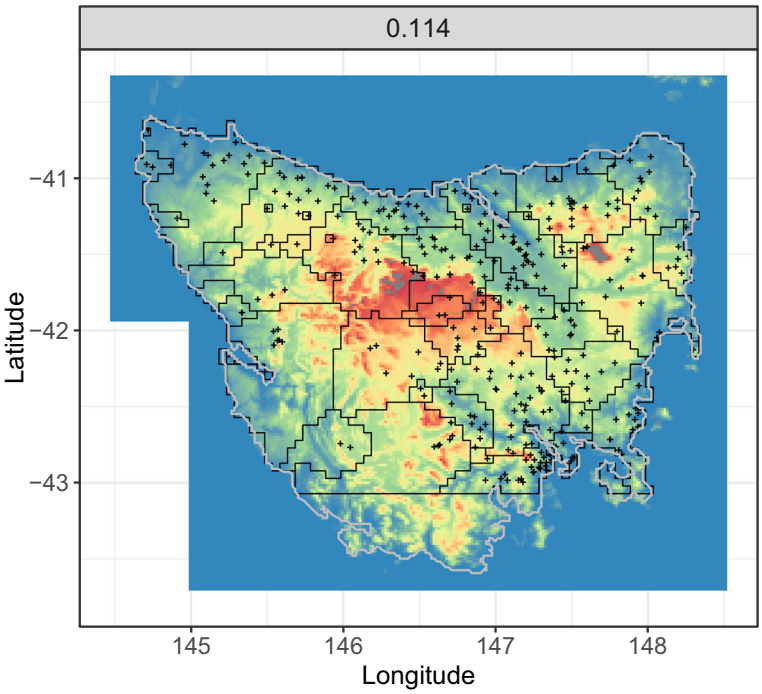

Fig. 15 The regionalisation of Tasmania at a cut height of approximately 0.11 overlaid on an elevation map

Grose et al. (2010) that many small regions are needed for rainfall compared with the East-West split advocated within the National Resource Management (NRM) clusters (CSIRO and Bureau of Meteorology 2015).

8 Conclusions

Using hierarchical clustering with the F-madogram distance, we have created a regionalisation based on the dependence of rainfall extremes. The clustering produced coherent partitions in Euclidean space. This was despite using only the observed, daily annual maxima. Additionally the regions generated from the clusters are broadly consistent with our understanding of climate and topographic features (Stern et al. 2000; Risbey et al. 2009). Given its simplicity, the regionalisation method we have presented is therefore very powerful.

Climate scientists, hydrologists and other researchers can use these regionalisations to improve their understanding about the behaviour of rainfall extremes. The size and shape of the regions provides information about the range of dependence and direction of anisotropy. Also, we can produce different regionalisations for different cut heights, where different cut heights correspond to different levels of regional detail relative to the desired strength of extremal dependence.

In addition to presenting the regionalisation, we highlighted key methodological considerations when using the F-madogram distance for clustering. The F-madogram distance can produce spurious clustering, depending on the underlying station network and the clustering method used. For clustering algorithms that are sensitive to point density this is of particular concern. Therefore for our application, K-medoids was completely unsuitable. This motivated using hierarchical clustering.

In general, we would advocate using hierarchical clustering over K-medoids for two reasons. The agglomerative implementation of hierarchical clustering improves cluster cohesion. Additionally, the ordered partitions have an interpretation in terms of dependence strength.

To understand the partitions relative to the full extremal dependence, we took the additional step of fitting max-stable models. As the dependence structure of our chosen max-stable model was Gaussian, we visualised the range of dependence and direction of anisotropy using elliptical level curves. For our regionalisations, we observed that there are many and varied dependence structures for rainfall extremes in Australia. Even for small regions we found that assuming a single dependence structure was not always suitable, but it depended on topographic features and the cut height chosen.

There are many future directions of this research. Our approach to producing regionalisations can be used to consider different maxima, such as monthly maxima, or different variables, such as temperature. Additionally, here we have also assumed stationarity, but we are curious as to how the dependence of rainfall extremes may vary temporally, such as under different large scale climate drivers (Min et al. 2013; Saunders et al. 2017) or under a changing climate (Westra et al. 2013; Alexander and Arblaster 2017). We would be interested in comparing regionalisation from this method under different time periods (Bador et al. 2015) and comparing regionalisations generated using observations to those from gridded data sets (Jones et al. 2009).

Our future goal for this research is to use the insights to model rainfall extremes on a continental scale, and to understand the impacts across large geographical distances. The regionalisations created can be used to help inform covariate selection and model selection for max-stable processes with non-stationary dependence (Huser and Genton 2016). When we started this research, this goal was aspirational. However, given the knowledge generated about the behaviour and dependence of rainfall extremes in Australia, this is now a very tangible direction for future research.

Acknowledgments Kate Saunders would like to thank Peter Taylor for his support and guidance throughout the course of her Ph.D, during which this research was undertaken. She would also like to thank Phillippe Naveau for his helpful suggestions and guidance during the onset of this work.

References

Alexander, L.V., Arblaster, J.M.: Historical and projected trends in temperature and precipitation extremes in Australia in observations and CMIP5. Weather Clim. Extrem. **15**, 34–56 (2017)

Asadi, P., Engelke, S., Davison, A.C.: Optimal regionalization of extreme value distributions for flood estimation. J. Hydrol. **556**, 182–193 (2018)

Bador, M., Naveau, P., Gilleland, E., Castellà, M., Arivelo, T.: Spatial clustering of summer temperature maxima from the CNRM-CM5 climate model ensembles & E-OBS over Europe. Weather Clim Extrem **9**, 17–24 (2015)

Bernard, E., Naveau, P., Vrac, M., Mestre, O.: Clustering of maxima: Spatial dependencies among heavy rainfall in France. J. Clim. **26**, 7929–7937 (2013)

Carreau, J., Naveau, P., Neppel, L.: Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. Water Resour. Res. **53**, 4407–4426 (2017)

Castro-Camilo, D., de Carvalho, M., Wadsworth, J.L.: Time-varying extreme value dependence with application to leading European stock markets. Ann. Appl. Stat. **12**, 283–309 (2018)

Castro-Camilo, D., Huser, R.: Local likelihood estimation of complex tail dependence structures, applied to US precipitation extremes. Journal of the American Statistical Association, 1–29 (2019)

Castruccio, S., Huser, R., Genton, M.G.: High-order composite likelihood inference for max-stable distributions and processes. J. Comput. Graph. Stat. **25**, 1212–1229 (2016)

Chamberlain, S.: rnoaa: 'NOAA' Weather Data from R. https://CRAN.R-project.org/package=rnoaa, R package version 0.7.0. (2017)

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A.: NbClust: An R package for determining the relevant number of clusters in a data set. J. Stat. Softw. **61**, 1–36. http://www.jstatsoft.org/v61/i06/ (2014)

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, vol. 208. Springer (2001)

Cooley, D., Naveau, P., Poncet, P.: Variograms for spatial max-stable random fields. In: Dependence in Probability and Statistics, pp. 373–390. Springer (2006)

Cressie, N.: Statistics for Spatial Data. Wiley, New York (2015)

CSIRO and Bureau of Meteorology: Climate Change in Australia Information for Australia's Natural Resource Management Regions. CSIRO and Bureau of Meteorology, Australia (2015)

Davison, A.C., Padoan, S.A., Ribatet, M., et al.: Statistical modeling of spatial extremes. Stat. Sci. **27**, 161–186 (2012)

de Haan, L.: A spectral representation for max-stable processes. The Annals of Probability, 1194–1204 (1984)

de Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer Science & Business Media (2006)

Dey, D., Yan, J.: Extreme Value Modeling and Risk Analysis: Methods and Applications. CRC Press (2016)

Dombry, C., Eyi-Minko, F.: Regular conditional distributions of continuous max-infinitely divisible random fields. Electron. J. Probab. **18**, 1–21 (2013)

Dudani, S.A.: The distance-weighted k-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics, 325–327 (1976)

Durre, I., Menne, M.J., Vose, R.S.: Strategies for evaluating quality assurance procedures. J. Appl. Meteorol. Climatol. **47**, 1785–1791 (2008)

Durre, I., Menne, M.J., Gleason, B.E., Houston, T.G., Vose, R.S.: Comprehensive automated quality assurance of daily surface observations. J. Appl. Meteorol. Climatol. **49**, 1615–1633 (2010)

Grose, M.R., Barnes-Keoghan, I., Corney, S.P., White, C.J., Holz, G.K., Bennett, J., Gaynor, S.M., Bindoff, N.L.: Climate Futures for Tasmania: General Climate Impacts Technical Report Cooperative Research Centre. Hobart, Tasmania (2010)

Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer (2009)

Haylock, M., Nicholls, N., et al.: Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910-1998. Int. J. Climatol. **20**, 1533–1541 (2000)

Hosking, J., Wallis, J.: Regional Frequency Analysis. An Approach Based on L-moments. Cambridge University Press Cambridge (1997)

Huser, R., Genton, M.G.: Non-stationary dependence structures for spatial extremes. J. Agric. Biol. Environ. Stat. **21**, 470–491 (2016)

Huser, R., Dombry, C., Ribatet, M., Genton, M.G.: Full likelihood inference for max-stable data. Stat **8**, e218 (2019)

Jones, D.A., Wang, W., Fawcett, R.: High-quality spatial climate data-sets for Australia. Austral. Meteorol. Oceanograph. J. **58**, 233 (2009)
Kaufman, L., Rousseeuw, P.J.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley (1990)
Min, S.K., Cai, W., Whetton, P.: Influence of climate variability on seasonal extremes over Australia. J. Geophys. Res. Atmosph. **118**, 643–654 (2013)
Murtagh, F.: A survey of recent advances in hierarchical clustering algorithms. Comput. J. **26**, 354–359 (1983)
Murtagh, F., Legendre, P.: Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? J. Class. **31**, 274–295 (2014)
Naveau, P., Guillou, A., Cooley, D., Diebolt, J.: Modelling pairwise dependence of maxima in space. Biometrika **96**, 1–17 (2009)
Oesting, M., Schlather, M., Friederichs, P.: Statistical post-processing of forecasts for extremes using bivariate brown-resnick processes with an application to wind gusts. Extremes **20**, 309–332 (2017)
Padoan, S.A., Ribatet, M., Sisson, S.A.: Likelihood-based inference for max-stable processes. J. Am. Stat. Assoc. **105**, 263–277 (2010)
Queensland Floods Commission of Inquiry: Queensland Floods Commission of Inquiry: Final Report. Queensland Floods Commission of Inquiry (2012)
Resnick, S.I.: Extreme Values Point Processes and Regular Variation. Springer, New York (1987)
Ribatet, M.: SpatialExtremes: Modelling Spatial Extremes. https://CRAN.R-project.org/package=SpatialExtremes, R, package version 2.0–2 (2015)
Risbey, J.S., Pook, M.J., McIntosh, P.C., Wheeler, M.C., Hendon, H.H.: On the remote drivers of rainfall variability in Australia. Mon. Weather. Rev. **137**, 3233–3253 (2009)
Rohrbeck, C., Tawn, J.A.: Bayesian spatial clustering of extremal behaviour for hydrological variables. Journal of Computational and Graphical Statistics, pp 1–38 (2020)
Samworth, R.J.: Optimal weighted nearest neighbour classifiers. Ann. Stat. **40**, 2733–2763 (2012)
Saunders, K., Stephenson, A.G., Taylor, P.G., Karoly, D.: The spatial distribution of rainfall extremes and the influence of El Nino Southern Oscillation. Weather and Climate Extremes (2017)
Saunders, K.: An Investigation of Australian Rainfall Using Extreme Value Theory. Ph.D. thesis, University of Melbourne (2018)
Schlather, M.: Models for stationary max-stable random fields. Extremes **5**, 33–44 (2002)
Smith, R.L.: Max-Stable Processes and Spatial Extremes. Unpublished manuscript, Univer (1990)
Stern, H., de Hoedt, G., Ernst, J.: Objective classification of Australian climates. Aust. Meteorol. Mag. **49**, 87–96 (2000)
Tibshirani, R., Walther, G., Hastie, T.: Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **63**, 411–423 (2001)
Westra, S., Alexander, L.V., Zwiers, F.W.: Global increasing trends in annual maximum daily precipitation. J. Clim. **26**, 3904–3918 (2013)