

Modelling heatwaves in central France: a case study in extremal dependence

Hugo C. Winter and Jonathan A. Tawn

Email: h.winter@lancaster.ac.uk, j.tawn@lancaster.ac.uk

Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, U.K.

Abstract

Heatwaves are a phenomena that have large social and economic consequences. Understanding and estimating the frequency of such events is of great importance to climate scientists and decision makers. Heatwaves are a type of extreme event which are by definition rare and as such there exists little data in the historical record to help planners. Extreme value theory is a general framework from which inference can be drawn from extreme events. When modelling heatwaves it is important to take into account the intensity and duration of events above a critical level as well as the interaction between both factors. Most previous methods assume that the duration distribution is independent of the critical level used to define a heatwave, a shortcoming that can lead to incorrect inferences. This paper characterises a novel method for analysing the temporal dependence of heatwaves with reference to observed temperatures from Orleans in central France. This method enables estimation of the probabilities for heatwave events irrespective of whether the duration distribution is independent of the critical level. The methods are demonstrated by estimating the probability of an event more severe than the 2003 European heatwave or an event that causes a specified increase in mortality.

Keywords: conditional extremes, extremal dependence, heatwaves, Markov chain, time-series extremes

1 Introduction

When modelling heatwaves decision makers are most interested in mitigating for disruption and fatalities. The heatwave across Europe in 2003 that caused around 40,000 heat related deaths (Fischer and Schär, 2010) and cost the farming industry around €13.1 billion highlights the potential large scale effects of such an event. High temperatures reduce the capacity of the human body for heat loss and are likely to cause core body temperature to exceed healthy limits (37-39°C). Most casualties in a heatwave are caused by heat exhaustion which leads to heat stroke. Heat exhaustion increases the blood pressure and leads to cardiovascular stress, which if not relieved results in cellular damage and an increased risk of mortality (Donaldson et al., 2003). Young and old people are particularly vulnerable during heatwave events. A day of strong heat could disrupt certain services for a couple of days but is unlikely to cause many fatalities. Conversely, a long sustained period of moderate to high heat may not disrupt services but can lead to many fatalities.

A heatwave is defined as a set of hot days and/or nights that are associated with a marked short-term increase in mortality. To make this definition precise we need to clarify what is meant by a hot day and a set of days. A hot day is when the temperature, or a related variable, exceeds a critical threshold level for health. Koppe et al. (2004) proposed threshold definitions based upon air temperature or indices based upon air temperature and relative humidity. Clearly health impacts increase with both the extent of the temperature excess over the critical threshold and the number of days that such an event lasts for. One way to measure the severity of the heatwave is to count the total number of days that the temperature series exceeds the critical level during a meteorological event, which we refer to as the duration of the heatwave event. During an extreme event, a set of days with temperatures below the critical level could allow respite from heat exhaustion and dramatically change the impact of the event so the duration of the heatwave is an insufficient measure for assessing some health implications. In these cases metrics such as the maximum consecutive sequence of exceedances or aggregated temperatures over the event are more appropriate. Aburrea et al. (2007), Stefanon et al. (2012) and Fischer and Schär (2010) all define a heatwave using a critical temperature threshold corresponding a fixed percentile of daily maximum summer temperatures (in the range 90%-95%) and a specified minimum duration (in the range 1-6 days). Relative critical levels are typically preferred to absolute levels when defining a heatwave since temperature can vary by geographical location and humans are able to adapt to local climate (Nitschke et al., 2011). Although heatwave definitions vary, all correspond to different but well-defined functionals of a meteorological event having temperatures which exceed a critical level.

To estimate the probability of a heatwave we propose a framework based upon extreme value theory. The framework relies on asymptotically justified models for describing the properties of the time series during an extreme temperature event. A broader and more flexible model with stronger asymptotic justification is proposed here than in previous studies. The model is used to simulate replicate extreme events that exceed a critical level permitting the evaluation of the distribution of any functional of the extreme event and hence the probability of a heatwave with specific characteristics. This approach ensures that this methodology applies to any form of heatwave definition of interest to experts from wide-ranging fields, such as heat-health researchers or those studying economic damage linked to heatwaves. Critically it enables the estimation of the probability of heatwaves occurring in a future period that are more extreme in any functional of interest than any of the observed events.

We apply these generic methods to the modelling of observed daily maximum temperatures to estimate the distribution of heatwaves at Orleans in central France, an area that was affected by the 2003 heatwave event. The hottest observed daily maximum temperature in 2003 for Orleans was 39.9°C. The summer daily maximum temperature one year return level, defined as the level exceeded on average once every summer, for Orleans is estimated as 35°C using standard extreme value methods (Coles, 2001). What made the 2003 event so severe for Orleans was that two heatwaves with 2 and 11 consecutive exceedances of the one year level occurred within a four week period. Pascal et al. (2013) quantified the relationship between temperatures and excess mortality over France, finding that if the average of three consecutive daily maximum temperatures exceeds 34°C (34°C, 35°C) excess mortality is 47% (17%, 33%) in Paris (Limoges, Lyon) respectively. Orleans is situated between these three cities and we focus on 35°C as the critical level for defining heatwaves. We note that the excess mortality from observing such a level is high but can vary

between locations. Under the assumption that the summer daily maximum temperature at Orleans, denoted $\{X_t\}$ on day t , is a stationary process we will estimate multiple quantities. These quantities include the joint probability of having an event that lasts at least as long and has peak value at least as severe as the 2003 event and that the average of three consecutive daily maximum temperatures exceeds 35°C . Pascal et al. (2013) also assessed the impact of high daily minimum temperatures coupled with high daily maximum temperatures on excess mortality. Under our framework we are also able to model the joint characteristics of temperature maxima and minima, over any time-scale, during the extreme event. We do not give specific estimates for that case, but in Section 6 we outline the modifications to our approach for modelling daily maxima that are required to address this broader concern.

Using empirical methods to estimate probabilities for the extreme heatwave events of interest to us is not possible so models are required. Here we need models for both the intensity and extremal dependence structure that determine properties of events. The intensity of heatwave events can be modelled by fitting an extreme value model to exceedances of a high modelling threshold u . The most common approach, which applies under weak conditions, is to fit a generalized Pareto distribution (GPD) to threshold excesses, i.e.

$$P(X_t - u > x \mid X_t > u) = \left(1 + \frac{\xi x}{\sigma_u}\right)_+^{-1/\xi} \quad \text{for } x \geq 0, \quad (1)$$

where $c_+ = \max(c, 0)$, $\sigma_u > 0$ and ξ are the scale and shape parameters of the GPD respectively (Coles, 2001).

A time-series of temperature data can be split into independent clusters where within each cluster groups of dependent exceedances occur. In the literature of extreme value theory these clusters are defined using different methods; the most popular technique is the runs method (Smith and Weissman, 1994). Under this method a cluster is ended by a sequence of m consecutive non-exceedances of u and a new cluster is commenced with the next exceedance of u . The run length m can be chosen subjectively; although Ferro and Segers (2003) outline an automated method. Therefore from a time-series it is possible to obtain the number of independent clusters and the values in each cluster. The number of clusters is Poisson distributed (Davison and Smith, 1990) so it remains to model the values within a cluster.

Standard asymptotic measures of cluster features are independent of the critical level. Examples include the distribution of the number of exceedances in a cluster, $\{\pi(i), i \geq 1\}$, associated mean cluster size θ^{-1} , where $\theta \in [0, 1]$ is the extremal index (Leadbetter et al., 1983), and other cluster functionals outlined in Smith et al. (1997) and Segers (2003). The focus on heatwaves highlights the need to not only account for the number of exceedances in a cluster, but also the full profile of the event to enable estimation of features such as the distribution of the number of consecutive exceedances or the average of three consecutive values. The application motivates the study of a new distribution $\pi_C(i)$ of the longest set of consecutive exceedances within a cluster along with the associated consecutive extremal index θ_C .

Under a stationary Markov process assumption, the extremal behaviour of $\{X_t\}$ can be modelled by focusing on the joint distribution of (X_t, X_{t+1}) . Multivariate extreme value theory leads to models for the joint tail through using separate marginal and dependence structures and can

be used to assess dependence between (X_t, X_{t+1}) . Dependence structures can be broadly split into those with asymptotic dependence and those with asymptotic independence (Sibuya (1960), Ledford and Tawn (1996)) determined by the value of χ where

$$\chi = \lim_{x \rightarrow x^*} P(X_{t+1} > x \mid X_t > x), \quad (2)$$

with x^* being the upper limit of the support of the common marginal distribution. In the case when $\chi = 0$ the variables (X_t, X_{t+1}) are said to be asymptotically independent and $\chi > 0$ corresponds to asymptotic dependence. The assumption of a dependence structure that is asymptotically dependent leads to the duration distribution being approximately independent of the critical level. Smith (1992), Coles et al. (1994), Smith et al. (1997), Perfekt (1997) and Yun (1998) use a parametric Markov model for estimating extremal quantities when $\chi > 0$. In contrast if the process is an asymptotically independent Markov chain then clusters in the limit reduce to single exceedances and $\theta = 1$ (Bortot and Tawn, 1998). However, if sub-asymptotic thresholds are considered $P(X_{t+1} > u \mid X_t > u) > 0$, for u as in equation (1), even when $\chi = 0$ and models are required that can capture this dependence as well. In these cases the duration and level of events are not independent.

The semi-parametric conditional extremes approach of Heffernan and Tawn (2004) offers a more flexible way of estimating extremal quantities of Markov chains than existing methods. This is due to a richer class of extremal dependence properties are permitted than those of Smith et al. (1997). These properties also hold over a much broader tail region than the parametric approach of Bortot and Tawn (1998). The approach of Bortot and Tawn (1998) provided models with asymptotic justification for (X_t, \dots, X_{t+m}) only in the region with $X_i > u$ for all $i = t, \dots, t+m$ and u a high threshold whereas we need models that hold for this vector subject only to $X_t > u$. The inclusion of dependence structures that also exhibit asymptotic independence permits the distribution of duration of events to change with critical level. Asymptotic dependence is a special case in the conditional extremes approach that does not require the evaluation of a parametric model. This non-parametric method of estimating extremal quantities of Markov chains can be compared to previous studies that assume a parametric dependence structure with asymptotic dependence.

In Section 2 the definition of a cluster and distributions of exceedances are formalised. Different approaches to model extremal dependence are outlined in Section 3. Section 4 discusses techniques for summarising the behaviour of clusters and compares the values of θ and θ_C . Section 5 presents the temperature data for Orleans, model fit diagnostics and results concerning the probability of observing the events of interest identified above. We focus on applying the conditional extremes approach and demonstrate how results differ from other approaches and show diagnostics that support our approach. Discussion and conclusions are presented in Section 6.

2 Cluster features

To understand clustering of time-series extremes it is necessary to formalise the asymptotic definition of a cluster and to provide a range of summaries. For a series $\{X_t, t = 1, \dots, n\}$ specify a threshold level u_n and a block of length m_n such that as $n \rightarrow \infty$, $u_n \rightarrow x^*$, with x^* as defined for equation (2), such that $nP(X_t > u_n) \rightarrow \tau > 0$ as $n \rightarrow \infty$ and $m_n = o(n)$. Under suitable

long-range mixing conditions the normalised process of times of exceedances of u_n , i.e.

$$\left\{ \frac{t}{n+1}; t = 1, \dots, n, X_t > u_n \right\},$$

converges to a compound Poisson process (Hsing, 1988). A cluster in block $\{1, \dots, m_n\}$ of this process is a set of exceedances of u_n by X_t for $t = 1, \dots, m_n$. The number of such exceedances is

$$N(u_n, m_n) = \#\{X_i > u_n \text{ for } i = 1, \dots, m_n\},$$

and hence a cluster occurs when $N(u_n, m_n) \geq 1$. The cluster size distribution $\pi(i, u_n)$ is defined as

$$\pi(i, u_n) = \text{P}(N(u_n, m_n) = i \mid N(u_n, m_n) \geq 1) \quad \text{for } i = 1, \dots, m_n.$$

Using this definition, it can be seen that $\pi(i, u_n)$ is the probability of obtaining i exceedances of threshold u_n in a block of m_n values given that there is at least one exceedance (i.e. there is a cluster). From this

$$\pi(i) = \lim_{n \rightarrow \infty} \pi(i, u_n) \quad \text{for } i = 1, 2, \dots, \quad (3)$$

is the limiting probability of a cluster of size i . A widely discussed dependence measure is the extremal index. This measure is the reciprocal of the mean of the cluster size distribution of the extremes in a time-series (Leadbetter, 1983). In terms of equation (3) the extremal index θ can be written as

$$\theta^{-1} = \sum_{i=1}^{\infty} i\pi(i).$$

An alternate form for the extremal index is characterised in O'Brien (1987) in the form of $\theta = \lim_{n \rightarrow \infty} \theta(u_n, m_n)$ where

$$\theta(u_n, m_n) = \text{P}(X_2 \leq u_n, \dots, X_{m_n} \leq u_n \mid X_1 > u_n), \quad (4)$$

which links to the runs estimator, discussed in Section 1, with run length m_n . The distribution $\pi(i)$ can be defined (Rootzén, 1988) as

$$\pi(i) = \frac{\theta^{(i)} - \theta^{(i+1)}}{\theta^{(1)}} \quad \text{for } i = 1, 2, \dots, \quad (5)$$

where

$$\theta^{(i)}(u_n, m_n) = \text{P}(N(u_n, m_n) = i \mid X_1 > u_n) \quad \text{for } i = 1, \dots, m_n, \quad (6)$$

defines the probability of viewing i exceedances of a threshold in a block of values given that the first value (X_1) exceeded the threshold and

$$\theta^{(i)} = \lim_{n \rightarrow \infty} \theta^{(i)}(u_n, m_n) \quad \text{for } i = 1, 2, \dots \quad (7)$$

This alternative approach to evaluating $\pi(i)$ is beneficial as it requires the evaluation of the process conditional on $X_1 > u_n$, in contrast to the evaluation of $N(u_n, m_n)$ which starts from an arbitrary X_1 , and hence it is more efficient for computational purposes.

For heatwaves it is also important to model the number of consecutive exceedances. This can be accomplished using the distribution $\pi_C(i)$ stated in Section 1. Specifically, let $C^{(i)}(u_n, m_n)$ be the event

$$C^{(i)}(u_n, m_n) = \{X_1 > u_n, \dots, X_i > u_n, X_{i+1} < u_n \cap \exists t = 2, \dots, m_n : X_t > u_n, \dots, X_{t+i-1} > u_n\},$$

i.e. that at time 1 there is a run of i consecutive exceedances but at no later time in the cluster does a run of this length or more occur. This leads to a measure that is analogous to equations (6) and (7), namely

$$\theta_C^{(i)}(u_n, m_n) = P(C^{(i)}(u_n, m_n) \mid X_1 > u_n) \quad \text{for } i = 1, \dots, m_n,$$

with

$$\theta_C^{(i)} = \lim_{n \rightarrow \infty} \theta_C^{(i)}(u_n, m_n) \quad \text{for } i = 1, 2, \dots$$

Note that both $\theta_C^{(1)}$ and $\theta^{(1)}$ are equal to θ by equation (4) since, in both situations, the event of interest is $\{X_2 < u_n, \dots, X_{m_n} < u_n \mid X_1 > u_n\}$. The distribution of the maximum number of consecutive exceedances within a cluster $\pi_C(i)$ is defined as

$$\pi_C(i) = \frac{\theta_C^{(i)} - \theta_C^{(i+1)}}{\theta_C^{(1)}} \quad \text{for } i = 1, 2, \dots$$

The average length of the longest set of consecutive exceedances in a cluster is given by the reciprocal of the consecutive extremal index θ_C , defined as

$$\theta_C^{-1} = \sum_{i=1}^{\infty} i \pi_C(i).$$

An event that has one exceedance in a cluster directly implies a maximum of one consecutive exceedance whereas the counter implication is not true, and hence $\pi_C(1) \geq \pi(1)$. As a consequence $\pi_C(i)$ experiences a sharper decline than $\pi(i)$ as i is increased.

Smith et al. (1997) investigate the behaviour of the extremal index θ against the parameters in an underlying Markov chain model. Here interest is in θ_C and so using the methodology outlined in Section 4 for evaluating cluster functions, Figure 1 compares the inverted extremal index θ^{-1} and the inverted consecutive extremal index θ_C^{-1} for a Markov chain with bivariate extreme value distribution with logistic dependence (Tawn, 1988) between consecutive values. Here a range of parameter values for the logistic dependence parameter are used, $\gamma \in (0, 1]$. A near perfect linear relationship is observed which shows that $\theta_C^{-1} \approx 0.7\theta^{-1}$. This shows that groups of consecutive exceedances are on average 30% shorter than the average cluster size for this dependence model.

3 Modelling temporal dependence

3.1 Markov modelling

To obtain estimates for $\pi(i)$ and $\pi_C(i)$ and their sub-asymptotic equivalents it is necessary to develop a model for the evolution of the temperature data through time. Many types of parametric

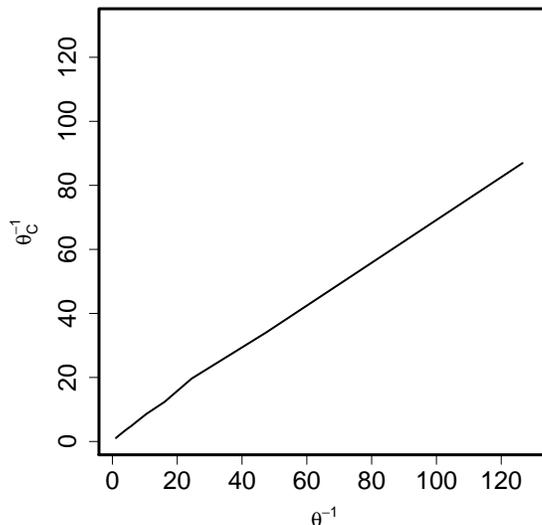


Figure 1: Relationship between inverted extremal index θ^{-1} and inverted consecutive extremal index θ_C^{-1} ; both are functions of logistic dependence parameter γ , for a Markov chain with bivariate extreme value distribution with γ from 0.05 to 1.

and non-parametric models could be constructed. Here, supported by exploratory data analysis, an assumption that the time series follows a first order Markov process is made. By the Markov property the distribution at each time step is only affected by the state of the system at the previous time-step with the resulting joint density of (X_1, \dots, X_n) given by

$$f(x_1) \prod_{t=1}^{n-1} f(x_{t+1}|x_t).$$

The assumption greatly simplifies the modelling process since to model the extremes of a time series X_1, \dots, X_n it is only necessary to model the extremes of pairs (X_t, X_{t+1}) for $t = 1, \dots, n-1$, which have joint distribution function $F(x_t, x_{t+1})$. The model for the marginal exceedances of the threshold u is given in Section 3.2. Our main method for modelling dependence is presented in Section 3.3 and connections with other models are discussed in Section 3.4. Our method is based upon the conditional approach outlined in Heffernan and Tawn (2004). That model allows for a rich class of dependence structures and most importantly allows for asymptotic independence which models the interaction between the duration distribution of an event and a critical level. Asymptotic dependence is a special case within the conditional approach, which motivates a new non-parametric approach and enables comparisons with the methods of Smith et al. (1997).

3.2 Marginal modelling

Following the assumption of stationarity of $\{X_t\}$ the marginal distributions of F are identically distributed. The assumption (1) for a GPD for the marginal excesses of u leads to the model for

the common marginal distribution

$$F(x) = \begin{cases} 1 - \lambda_u \left(1 + \xi \frac{x-u}{\sigma_u}\right)_+^{-1/\xi}, & x \geq u \\ \tilde{F}(x), & x < u, \end{cases} \quad (8)$$

where $\lambda_u = 1 - F(u)$ and $\tilde{F}(x)$ is the empirical cumulative distribution function of $\{X_t\}_{t=1}^n$. Marginal parameters are estimated using a censored likelihood approach. For modelling extremal dependence we need to select an appropriate margin to transform onto. In copula methods (Nelson, 2007) it is common to model dependence with uniform margins, but for extremes simplifications in model form arise when focusing on a different marginal choice. Heffernan and Tawn (2004) model dependence for Gumbel margins. Keef et al. (2013) showed that a more comprehensive approach arises for Laplace margins. Following Keef et al. (2013) we transform X_t , $t = 1, \dots, n$ onto Laplace margins as follows

$$T(X_t) = \begin{cases} \log \{2F(X_t)\} & \text{if } X_t < F^{-1}(0.5) \\ -\log \{2[1 - F(X_t)]\} & \text{if } X_t \geq F^{-1}(0.5). \end{cases}$$

3.3 Semi-parametric conditional extremes approach

The conditional extremes method of Heffernan and Tawn (2004) and Heffernan and Resnick (2007) can be used to motivate a modelling framework for which χ , defined by equation (2), can be either positive or zero. The desire is to model the joint distribution $[T(X_t), T(X_{t+1})]$ using the distribution of $T(X_{t+1})$ given that $T(X_t)$ is large (defined as exceeding a high threshold). A requirement for modelling the conditional distribution $\mathbb{P}\{T(X_{t+1}) \leq T(x_{t+1}) \mid T(X_t) = T(x_t)\}$ is that this distribution should be non-degenerate as $x_t \rightarrow x^*$. As such the Heffernan and Tawn (2004) approach aims to identify normalizing functions $a : \mathbb{R}_+ \rightarrow \mathbb{R}$ and $b : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ that are defined such that for $x > 0$

$$\mathbb{P}\left(\frac{T(X_{t+1}) - a[T(X_t)]}{b[T(X_t)]} \leq z, \frac{X_t - u}{\sigma_u} > x \mid X_t > u\right) \rightarrow G(z) (1 + \xi x)_+^{-1/\xi}, \quad (9)$$

as $u \rightarrow x^*$, where G is a non-degenerate distribution function and σ_u is as in equation (1). The specification of Laplace margins ensures that the upper and lower tails are symmetric and exponential which permits the definition of a single parsimonious class of choices for the normalising functions of

$$a(y) = \alpha y \quad \text{and} \quad b(y) = y^\beta,$$

where $\alpha \in [-1, 1]$ and $\beta \in (-\infty, 1)$. This form of the normalising functions does not affect the limiting dependence model in Heffernan and Tawn (2004) and simplifies the inference for variables which are either negatively or weakly associated. If $\alpha = \beta = 0$ and $G(z)$ is the Laplace distribution function the variables are independent whereas $\alpha = 1$ and $\beta = 0$ corresponds to the situation of asymptotic dependence (given by $\chi > 0$ in equation (2)) and $-1 \leq \alpha \leq 0$ to negative dependence.

Modelling using the conditional extremes approach requires the assumption that the limiting form of equation (9) holds exactly for all values of $X_t > u$ given that u is a sufficiently high threshold, i.e. λ_u is small. Given this assumption it is possible to write the form of X_{t+1} given that $X_t > u$ as

$$T(X_{t+1}) = \alpha T(X_t) + T(X_t)^\beta Z_{t+1}, \quad (10)$$

where Z_{t+1} is a random variable with distribution function G . We also have that Z_{t+1} is independent of X_t and, following the stationary Markov process assumption, the sequence of $\{Z_t\}$ are independent and identically distributed. As G does not take any simple parametric form, to estimate α and β a false working assumption is made, as in Keef et al. (2013), that $Z_{t+1} \sim N(\mu, \sigma^2)$ and as such

$$T(X_{t+1}) | \{T(X_t) = y\} \sim N\left(\alpha y + \mu y^\beta, \sigma^2 y^{2\beta}\right) \quad \text{for } y > T(u).$$

The working assumption permits the estimation of the set of parameters $(\alpha, \beta, \mu, \sigma)$ by standard likelihood approaches. At this stage the estimates for (μ, σ) are discarded and a non-parametric estimate of the distribution for Z is formed by inverting equation (10) to give estimated values of Z_{t+1} . Specifically, let t_1, \dots, t_{n_u} be the indices of $t = 1, \dots, n$ where $x_t > u$ and where n_u is the number of data points exceeding the threshold u . Then let

$$\hat{z}_j = \frac{T(x_{t_j+1}) - \hat{\alpha}T(x_{t_j})}{T(x_{t_j})^{\hat{\beta}}}, \quad (11)$$

for $j = 1, \dots, n_u$. In this way a non-parametric estimate \hat{G} to the distribution function G is formed using $\hat{z}_j, j = 1, \dots, n_u$.

Note that under asymptotic dependence, i.e. $a(y) = y$ and $b(y) = 1$, the transition probability (9), when expressed in terms of the original variable X_t , is given by

$$P(X_{t+1} \leq X_t + z[\sigma_u + \xi(X_t - u)]_+ | X_t > u) \rightarrow G(z). \quad (12)$$

Under an asymptotic dependence assumption in the semi-parametric conditional approach it is known that $\alpha = 1$ and $\beta = 0$. In this situation G is estimated by the empirical distribution of the differences in the original data on the Laplace scale (later referred to as the non-parametric approach), i.e. using the sample $j = 1, \dots, n_u$ of

$$\hat{z}_j = T(x_{t_j+1}) - T(x_{t_j}), \text{ for } x_{t_j} > u. \quad (13)$$

3.4 Connections with alternative approaches

Smith et al. (1997) propose a parametric Markov model for the joint distribution of consecutive values of the time series which limits the dependence structure to asymptotic dependence or exact independence. Here we show the connections between that modelling approach and the conditional approach outlined in Section 3.3. This enables us to show the benefits for modelling the data of relaxing the strong assumptions of asymptotic dependence and a parametric model.

Based on an asymptotic approximation for a high threshold u , Smith et al. (1997) propose a bivariate extreme value distribution copula with GPD marginal tails for the joint distribution function $F(x_1, x_2)$ of (X_t, X_{t+1}) . This joint distribution is given as

$$F(x_1, x_2) = \exp \left\{ - \int_0^1 \max \left(\frac{w}{z_1}, \frac{1-w}{z_2} \right) 2dH(w) \right\} \quad \text{for } x_1 > u, x_2 > u,$$

where

$$z_j = -1/\log \left[1 - \lambda_u \left(1 + \xi \frac{x - u}{\sigma_u} \right)_+^{-1/\xi} \right] \text{ for } j = 1, 2,$$

and H is an arbitrary distribution function on $[0, 1]$ satisfying the moment constraint

$$\int_0^1 w dH(w) = 1/2.$$

The corresponding transition probability for extreme X_t is given by

$$P(X_{t+1} \leq x_t + z[\sigma_u + \xi(x_t - u)]_+ | X_t = x_t) \rightarrow 2 \int_{[1+(1+\xi z)_+]^{1/\xi}]^{-1}}^1 w dH(w), \quad (14)$$

as $x_t \rightarrow x^*$. Following results in Heffernan and Resnick (2007) and Wadsworth et al. (2014), equations (12) and (14) are equivalent expressions despite having slightly different conditioning and limit setups. This gives a formulation for G in this case but also shows that the semi-parametric conditional approach directly extends the approach of Smith et al. (1997).

Smith et al. (1997) make the additional assumption of a parametric model for H , exploring a range of models; see Kotz and Nadarajah (2000) for more models. We follow Smith (1992) and assume the logistic dependence structure with parameter γ . This gives the joint distribution to be

$$F(x_1, x_2) = \exp \left\{ - \left(z_1^{-1/\gamma} + z_2^{-1/\gamma} \right)^\gamma \right\},$$

where $\gamma \in (0, 1]$. Independent variables correspond to $\gamma = 1$ and perfectly dependent variables are given as $\gamma \rightarrow 0$. For intermediate values of γ there is asymptotic dependence with $\chi = 2 - 2^\gamma$. Inference for this parametric family is through the censored likelihood approach of Smith et al. (1997). For this parametric model it follows that $G(z) = [1 + \exp(-z/\gamma)]^{\gamma-1}$. In contrast to this parametric form for G the empirical distribution of the sample given by expression (13) offers greater flexibility for the semi-parametric conditional approach even under an assumption of asymptotic dependence. For more information about non-parametric approaches for multivariate extremes under asymptotic dependence see de Haan and Ferreira (2006).

4 Cluster behaviour estimation

When analysing the behaviour of heatwaves we can look at within cluster and over cluster results. The work in the previous sections has concentrated on within-cluster behaviour since the definition of the distributions in Section 2 are conditional upon a cluster occurring. It is more relevant for applications to have the probability of observing a cluster with specific characteristics in a certain time period. Here we first discuss within cluster behaviour and then discuss how this is extended to over cluster results.

Our approach to deriving the properties of clusters of a Markov chain is the repeated simulation of a segment of the chain in periods with exceedances of a critical level v , i.e. when the process exceeds v , with $v \geq u$, where u is our modelling threshold. There are two different strategies for the generation of the chain in its tail state, known as the tail chain. Smith et al. (1997) suggest simulating the cluster maximum $M > v$ and then simulating forwards and backwards from this. A different method (Rootzén, 1988) involves the simulation of an exceedance of v , i.e. $X_1 > v$, and only requires forward simulation. Cluster properties such as $\theta(v, m)$ and $\pi(v, m)$ can be estimated empirically from repeated simulations of clusters. For example, $\theta(v, m)$ is estimated as either the reciprocal of the average cluster length using the Smith et al. (1997) approach or as the probability

$\theta^{(1)}(v, m)$ in equation (6) using the Rootzén (1988) approach. For general functionals the Smith et al. (1997) approach can always be used. The Rootzén (1988) approach is easiest to implement but in practice requires additional steps to ensure non-negativity of the distributions π and π_C . We use the pool adjacent violators (PAV) algorithm; for more information see Appendix B. For notational simplicity we define N as the number of exceedances above the critical level v in a cluster and N_C as the maximum number of consecutive exceedances in a cluster. For the rest of this section construction of one simulated chain will be discussed; π , π_C and other useful cluster features are evaluated using repeated simulation (2 million tail chains in Section 5). Importance sampling of the initial simulated exceedance is useful to obtain accurate estimates of $\pi(i, v)$ and $\pi_C(i, v)$ for large i in practice but omitted from the discussion below.

The semi-parametric conditional extremes approach is used to generate realisations of a tail chain. We require a starting exceedance of $v \geq u$ to be generated from a $\text{GPD}(\sigma_v, \xi)$, where $\sigma_v = \sigma_u + \xi(v - u)$, and step forward until a chain of length k is simulated. For sufficiently high thresholds the GPD is an appropriate distribution for simulating cluster maxima; at lower levels it may be necessary to simulate cluster maxima using the distribution in Eastoe and Tawn (2012), which for high thresholds converges to the GPD. This length k is chosen large enough to ensure a negligible probability of simulating any more exceedances of v ($k = 40$ is found sufficient in Section 5). Details of the algorithm are given in Appendix A. The asymptotic justification for the algorithm is satisfied only when $X_t > u$. For $X_t < u$ we continue to use the algorithm as it should still provide a reasonable approximation unless $X_t \ll u$. In this case the probability of the chain coming above u again is negligible.

It may be of interest to work out how long a heatwave event might last given that the peak value of the cluster, M , is known to be greater than or equal to η with $\eta \geq v$. Such a question cannot be evaluated efficiently using the forward tail chain methods described above. We use forward and backward tail chains starting from the peak value M . A simulation scheme for the conditional extremes approach to evaluate $\text{P}(N = i \mid M = \eta)$ is outlined in Appendix A. The probability $\text{P}(N_C = i \mid M = \eta)$ for the number of consecutive exceedances N_C above v given the maximum is η can similarly be evaluated. The distribution of the number of exceedances given a maximum greater than η , where $\eta \geq v$, is given by the integral

$$\text{P}(N = i \mid M \geq \eta) = \int_{\eta}^{\infty} \text{P}(N = i \mid M = s) \frac{1}{\sigma_v} \left[1 + \xi \frac{s - v}{\sigma_v} \right]_+^{-1/\xi - 1} ds,$$

which is evaluated in practice using a Monte Carlo approximation. Similar simulation schemes can be produced for the parametric (Smith et al., 1997) and non-parametric approaches for asymptotically dependent tail chains; more details are given in Appendix A.

Extending within cluster results to over cluster results requires the assumption that clusters of the modelling threshold u occur as a Poisson process (Hsing, 1988). The mean number of clusters in period T is given by

$$\tau_u = \theta(u, m) \lambda_u n_T,$$

where $\theta(u, m)$ is the sub-asymptotic extremal index at u for run length m given by expression (4), λ_u is the threshold exceedance probability from equation (8) and n_T is the number of observations within period T . For example, with daily data if the rate of clusters within a summer (92 day

period from June to August) is desired then $n_T = 92$. At a higher level $v > u$ the mean number of clusters in period T is

$$\begin{aligned}\tau_v &= \theta(v, m) \lambda_u n_T \left[1 + \xi \left(\frac{v-u}{\sigma_u} \right) \right]_+^{-1/\xi} \\ &= \tau_u \frac{\theta(v, m)}{\theta(u, m)} \left[1 + \xi \left(\frac{v-u}{\sigma_u} \right) \right]_+^{-1/\xi},\end{aligned}$$

where the change from $\theta(u, m)$ to $\theta(v, m)$ takes the change in mean cluster size at each level into account and the final term adjusts for the marginal rarity.

It is interesting to know the probability $\psi_v(\kappa, \eta)$ of observing at least one cluster in a period T with a desired extremal property. One example is an event that lasts at least κ days above level v and attains a peak value of at least η . For this example

$$\begin{aligned}\psi_v(\kappa, \eta) &= \sum_{j=0}^{\infty} \left\{ 1 - [1 - \bar{\Pi}_v(\kappa, \eta)]^j \right\} \frac{\tau_v^j \exp(-\tau_v)}{j!} \\ &= 1 - \exp[-\tau_v \bar{\Pi}_v(\kappa, \eta)],\end{aligned}\tag{15}$$

where the summation is taken over the number of clusters of the level v and

$$\bar{\Pi}_v(\kappa, \eta) = \text{P}(N \geq \kappa, M \geq \eta \mid M > v).$$

Similar results can be derived with the number of consecutive exceedances N_C as the quantity of interest. Such measures are required when evaluating the probability of observing events at least as severe as the 2003 heatwave.

5 Heatwave application

5.1 Data

Daily temperature observations were taken at Orleans, in central France, for the period 1946-2012. Four missing values exist in the time-series and are omitted; none occur during the 2003 event. Heatwaves are most likely to occur in summer months, here defined as the 92 day period of June-August, so summer season and yearly return levels are equivalent. These three month periods are extracted from each year to form an approximately stationary time-series for the temperature. Sample auto-correlation and partial auto-correlation functions support the assumption of a first-order Markov chain; see Figure 2 for the latter. Since the partial auto-correlation function is affected by heavy tails, we evaluated this function with the data transformed onto Gaussian margins but found no significant change. As such a first-order Markov model is adopted within each summer period and each summer period is treated as independent of others. Figure 2 shows consecutive pairs of the temperature data illustrating strong inter-day dependence.

5.2 Problem and strategy

We want to estimate the probability of observing events such as a heatwave that is more extreme than the 2003 event or that exceeded a specified level of increased mortality. We select a run length

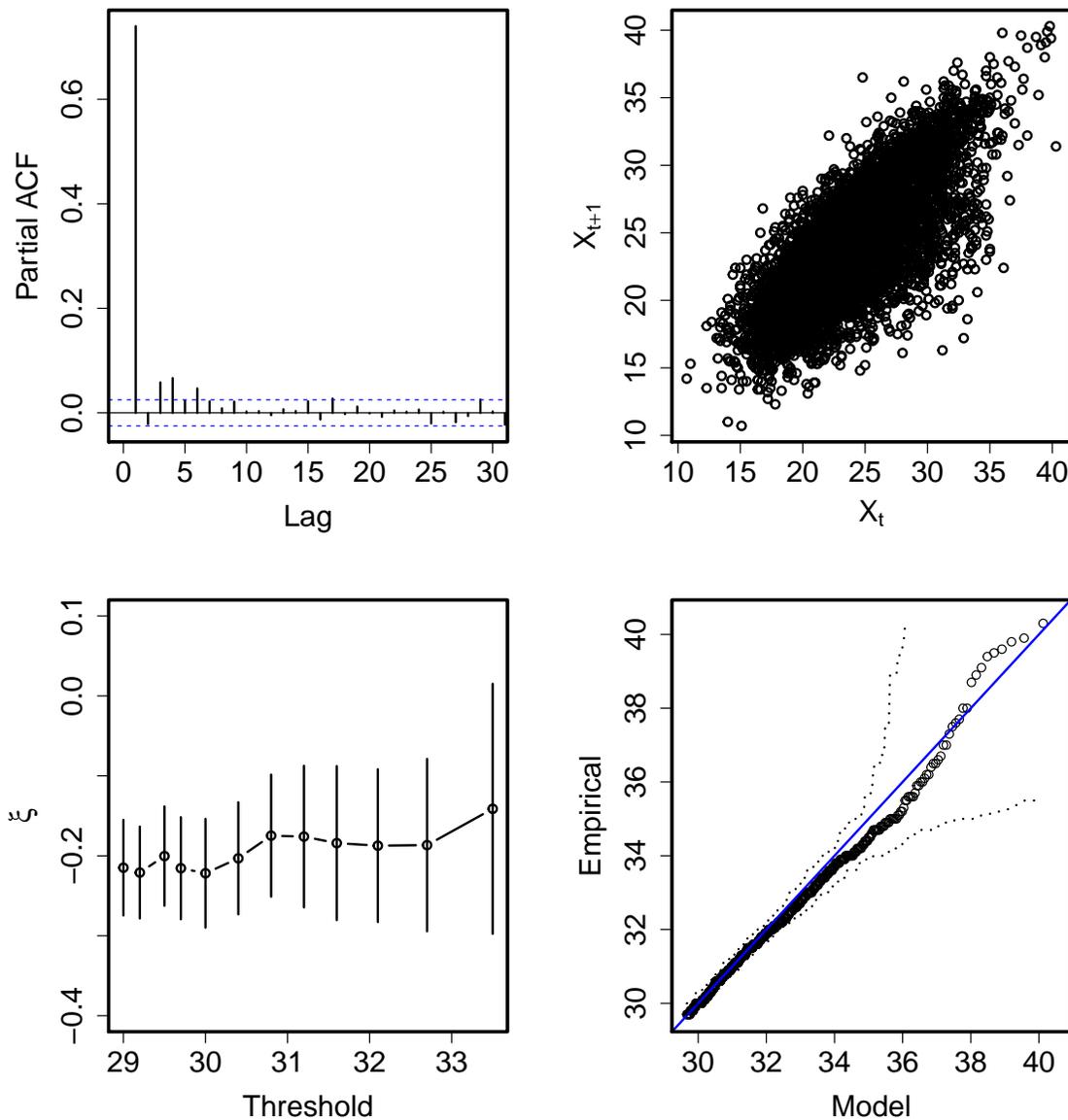


Figure 2: Diagnostic plots for June, July and August temperature data ($^{\circ}\text{C}$) for Orleans: partial autocorrelation (top left), scatter of consecutive pairs (top right), parameter stability plot for GPD shape parameter ξ (bottom left) and QQ-plot of GPD fit with 95% tolerance bounds indicated by the dotted lines (bottom right).

of 3 days to correspond to the typical propagation of weather systems. A larger choice for the run length in practice will make little difference. Using the runs method with a run length of 3 and a critical level equal to the one year return level (denoted v_1 , taking the value 35°C), two independent clusters with 2 and 11 consecutive exceedances respectively are identified within a four week period in 2003. It is expected that the daily maximum temperature series exceeds the 1 year return level on average once a summer. It is highly unlikely to observe 13 exceedances in a year, in particular in Section 5.3 we show that on average we would expect to see only two exceedances for each cluster that exceed such a level.

Empirical estimates of cluster features based upon the runs method are affected by the choice of run length and cannot be used to estimate the required probabilities since they are higher levels than have been observed. Models from Section 3 are used to provide these estimates and do not require a choice of run length due to the assumption of a Markov process. The methods outlined in Section 3.3 are named the conditional and non-parametric approaches respectively and the method outlined in Section 3.4 is named the parametric approach for this study. A comparison between the three models and the empirical estimate obtained via the runs method is used for model diagnostic purposes and is given in Section 5.3. Results regarding the probability of observing heatwaves with the characteristics of interest are given in Section 5.4.

Confidence intervals for all four approaches are generated by bootstrap methods. Runs method declustering defines n_c clusters of varying length and by alternately sampling clusters of exceedances and clusters of non-exceedances from this observed set we generate a bootstrapped sample. This procedure is repeated to generate 1000 replicate data sets to which the models of Section 3 are fitted. Repeated simulation is used to obtain estimates of cluster functionals such as $\pi(i)$ and $\pi_C(i)$ as discussed in Section 4. Bootstrapped 95% confidence intervals for π and π_C are derived by taking the 2.5 and 97.5 percentiles of the estimates obtained from the replicate data sets.

5.3 Diagnostics

First, a GPD is fitted to exceedances of the modelling threshold u , with u chosen using standard diagnostics (Coles, 2001). In particular, we use a parameter stability plot for ξ (Figure 2) and check that estimates of the shape parameter stay consistent above the chosen threshold. Each approach is evaluated using the modelling threshold u , set at the 90th percentile such that 10% of days fall above the threshold (taking the value 29.7°C). Higher levels v for which results are reported will be defined for each different analysis. The rate parameter λ_u is estimated as 0.099 (0.007), where the standard error is given in the parentheses. The GPD scale parameter is estimated as $\hat{\sigma}_u = 3.002$ (0.225) and the shape parameter $\hat{\xi} = -0.215$ (0.033). A QQ-plot evaluated with the modelling threshold at the 90th percentile is provided in Figure 2 and indicates that the GPD is a reasonable fit at this threshold. Deviations from the diagonal are observed at higher thresholds but are contained within 95% tolerance intervals. Parameter stability plots at higher thresholds (not shown) do not indicate any statistically significant change in the parameter estimates.

Fitting the conditional extremes approach leads to an estimate for the dependence parameters of $\hat{\alpha} = 0.713$ (0.072) and $\hat{\beta} = 0.524$ (0.094). Parameter stability plots for the conditional extremes dependence parameters are given in Figure 3 and support that the choice of u is valid. A likelihood ratio test confirms that these parameter values are significantly different from $\alpha = 1$ and $\beta = 0$ and that the data do not exhibit asymptotic dependence. Under the parametric model the logistic dependence parameter is estimated as $\hat{\gamma} = 0.578$ (0.026) with $\hat{\chi} = 0.508$ (0.027). As asymptotic dependence is the only form of dependence allowed in this model, $\hat{\chi} > 0$ despite the evidence from the conditional approach that suggests $\chi = 0$. When using peak value tail chain estimation the dependence parameters for the backward chain are also required and here $\hat{\alpha}_b = 0.816$ (0.061) and $\hat{\beta}_b = 0.512$ (0.096).

Figure 4 shows estimates of $\theta(v, m)$ and $\theta_C(v, m)$ under all approaches for return periods between 0.1 and 1 years, with m set as 3 days. At these levels estimates given by the runs method are

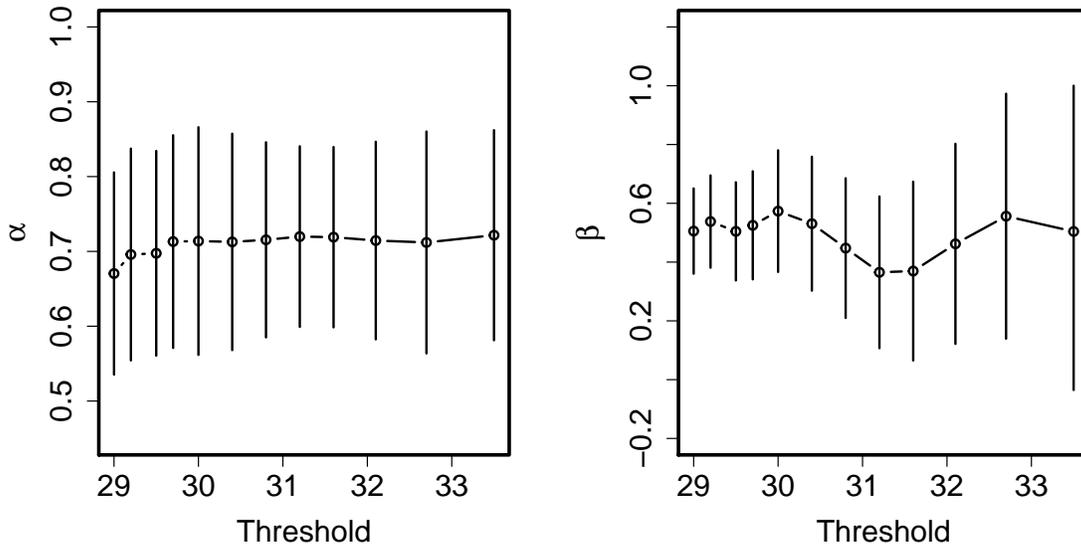


Figure 3: Parameter stability plots for conditional extremes dependence parameters α (left) and β (right). Vertical bars show 95% confidence intervals.

reasonably accurate and are used to assess which approach provides the best fit. The empirical estimate of $\theta(v, m)$ shows a broadly increasing pattern at lower return periods before levelling out and tailing off at higher levels, where less data are available for estimates. The estimate of $\theta(v, m)$ for the conditional extremes approach matches the behaviour of the empirical estimate the best. It is contained within the 95% confidence intervals of the empirical estimate at all levels. This result suggests a reduced sensitivity to the choice of modelling threshold u compared to the asymptotically dependent parametric and non-parametric approaches. The parametric approach is often contained within the confidence intervals, but at lower return periods it overestimates the size of the extremal index. It also fails to pick up the increase of the empirical estimate of the extremal index at lower return levels. The non-parametric approach cannot pick up this behaviour either and usually underestimates the extremal index. Similar patterns are observed for the consecutive extremal index $\theta_C(v, m)$ except that estimates are slightly higher. This pattern is similar for N and N_C and so whilst assessing fit we shall concentrate on estimating functions of N with passing comments only made on N_C .

Estimates of the probability mass function $\pi(i, v)$ for all four approaches are given in Figure 5 (left) for a range of i for which the estimated distributions differ non-negligibly from zero. The critical level is set as $v = v_1$. All distributions are decreasing with i ; there are only slight increases due to sampling noise for longer cluster lengths where results become sparse. The empirical estimate based upon the runs method shows the greatest amount of variability and all other approaches have narrower confidence intervals. The results obtained from the parametric and conditional extremes approaches tend to coincide but show some varying behaviour. At most values of i the confidence intervals for the parametric and conditional extremes approaches are contained within those of the runs estimator. The result suggests that both methods are adequately modelling the data at this critical level. The non-parametric approach gives estimates that seem to generally coincide with

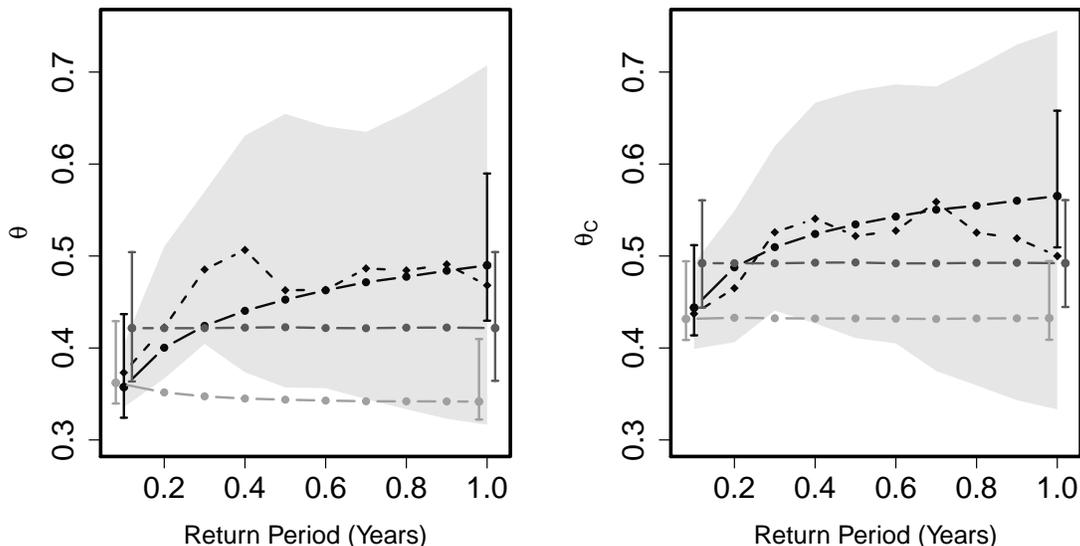


Figure 4: Estimated extremal index $\theta(v, m)$ (left) and consecutive extremal index $\theta_C(v, m)$ (right) obtained with the runs method (dashed, diamonds), parametric approach (grey dots), conditional extremes approach (black dots) and non-parametric approach (light grey dots). Confidence intervals are at the 95% level and are obtained by bootstrapping procedures for all approaches. The interval is given for runs method (light grey shading) at all return periods; for other approaches at 0.1 and 1 year return period (staggered) for visual clarity, where the return level v corresponds to the return period given on the horizontal axis and m is fixed at 3 days.

the empirical runs estimate and there is general agreement with the other approaches, though the estimate of $\pi(1, v_1)$ is lower.

The effect of changing critical level v on $\pi(1, v)$ is presented in Figure 5 (right). The runs estimator has been omitted from the plot since at these high levels the estimates obtained in this way become unreliable with wide confidence bands. Figure 5 (right) confirms that for the parametric and non-parametric approaches the value of $\pi(1, v)$ remains constant at all levels whereas for the conditional extremes approach $\pi(1, v)$ increases as the critical level is increased. The same pattern can be observed for $\pi_C(1, v)$ (not shown). This occurs since the parametric and non-parametric approaches are restricted to asymptotic dependence which does not allow for interaction between duration distribution and critical level. The conditional extremes method can allow for the asymptotically independent behaviour of the series and therefore can have interaction between duration distribution and critical level. The parametric and non-parametric approaches average the dependence over observed levels which leads to the constant behaviour. The confidence intervals of these solely asymptotically dependent methods tend to overlap.

It may also be of interest to know the duration distribution of a cluster given that the peak value was recorded at a specific level. The conditional extremes approach has highlighted that the behaviour of clusters changes with the critical level used to define them. Peak value chain simulation for the conditional extremes approach can be used as outlined in Section 4. Setting $v = v_1$ it is possible to analyse cluster characteristics of events that have a larger peak value. The plot of $P(N = i \mid M = \eta)$ for different peak values η is given in Figure 6. The shape of the distribution

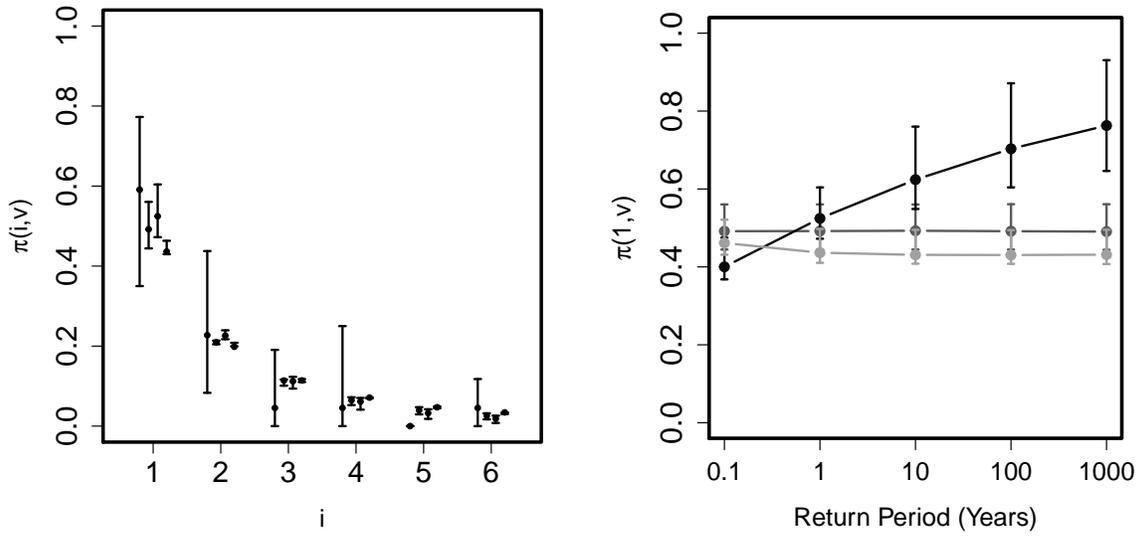


Figure 5: Left: Estimated distribution of $\pi(i, v_1)$ obtained with runs method (far left values), parametric approach (centre left), conditional extremes approach (centre right) and non-parametric approach (far right). Right: The effect of varying the return period (threshold) on $\pi(1, v)$ for parametric approach (grey, middle), conditional extremes approach (black, top) and non-parametric approach (light grey, bottom) plotted against v on a log return period scale. In both plots confidence intervals are at the 95% level and are obtained by bootstrapping procedures for all approaches.

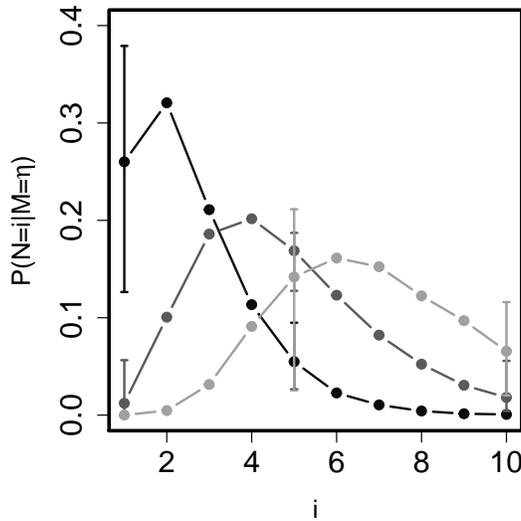


Figure 6: The effect of different cluster maximum sizes on distributions $P(N = i | M = \eta)$ at critical level $v = v_1$ when estimated using the conditional extremes approach. Cluster maxima η at level that would occur on average once in 5 years (black), once in 50 years (grey) and once in 1000 years (light grey). Confidence intervals are given at the 95% level, constructed from 100 replicate data sets with 10000 forward/backward simulated chains and presented at only three durations for visual clarity.

is positively skewed for lower peak values. This result is anticipated since a peak value nearer to the critical level will typically yield a cluster with fewer exceedances of the critical level than for clusters with a larger peak value.

5.4 Results

Empirical analysis in Section 5.2 identified events of length 2 and 11 above the critical level, corresponding to the one year return level. Tools from Section 4 allow us to estimate how likely each event was. In what follows we present estimated probabilities and in parentheses the associated bootstrapped 95% confidence intervals. The probability of observing at least one event in a year that lasts at least 2 days is 0.208 (0.200, 0.216) for the conditional extremes approach, 0.193 (0.183, 0.199) for the parametric approach and 0.175 (0.172, 0.188) for the non-parametric approach. Similarly, the probability of observing at least one event that lasts at least 11 days in a year is 0.001 (1×10^{-4} , 0.004) for the conditional extremes approach, 0.005 (0.002, 0.009) for the parametric approach and 0.012 (0.007, 0.013) for the non-parametric approach. The asymptotically dependent parametric and non-parametric approaches give a much higher probability of observing a long event than the asymptotically independent conditional extremes approach. The same analysis can be completed for the probability of observing at least one event in a year that lasts longer than 11 consecutive days. For the conditional extremes approach the probability is 6×10^{-4} (4×10^{-5} , 0.002), for the parametric approach 0.004 (0.002, 0.006) and for the non-parametric approach this increases to 0.007 (0.004, 0.009).

In Section 1 we noted that periods of 3 days with an average daily maximum temperature above 35°C could lead to an excess mortality which varies over local cities between 17-47%. Using all approaches we can estimate the probability of observing at least one such event in a year. For the conditional extremes approach the probability is given as 0.199 (0.181, 0.226), equivalent to an event that happens on average once every five years. The same probability is given as 0.174 (0.161, 0.180) for the parametric approach and 0.169 (0.157, 0.183) for the non-parametric approach. For the remainder of the analysis we focus on the conditional extremes approach.

The maximum temperature in Orleans in 2003 was recorded at 39.9°C which corresponds to a 1 in 50 year event. The peak value chain estimation method in Section 4 is used to assess the joint probability of an event with a hotter maximum temperature and longer duration than the 2003 heatwave event. The probability of observing a cluster with at least 11 exceedances conditional on a peak value greater than the 2003 temperature is 0.06 (0.008, 0.23). The joint probability for the cluster functionals can be obtained by multiplying the conditional probability by the probability of observing a peak value greater than the 2003 temperature and is estimated as 0.001 (7×10^{-5} , 0.013). Application of equation (15) allows the derivation of over cluster results from the within cluster results given above. As such the probability of observing at least one event in a year that both lasts longer than 11 days and has a peak value greater than 39.9°C is 6×10^{-4} (4×10^{-5} , 6×10^{-3}), approximately equivalent to the 1650 year return period. The equivalent probability for 11 consecutive exceedances is 4×10^{-4} (3×10^{-5} , 5×10^{-3}) for the conditional extremes approach.

6 Discussion and conclusion

The results given in Section 5 show that the interaction between the duration distribution of heatwave events and a critical level is only modelled realistically by methods that account for asymptotic independence. At high critical levels this leads to a reduction in the probability of observing longer events when using the conditional approach over other approaches that can account only for asymptotic dependence. Model selection diagnostics indicate that these lower estimates of the probability of observing longer events at high critical levels reflect the characteristics of the data better. If a user is especially averse to the risk of longer heatwave events then they could be willing to mitigate for such an event using one of the asymptotically dependent approaches. However, our analysis shows that this can considerably inflate the estimated risk. If such a conservative approach is to be taken we have found that two different asymptotically dependent modelling approaches, parametric and non-parametric, give very similar risk estimates.

The assumption that the temperature time-series follows a first order Markov process has been made to permit the modelling process outlined in the paper. Such an assumption was supported by an exploratory data analysis but might be an unrealistic assumption in other applications or for such extreme events as in 2003. Specifically, our approach gives the return period of an event rarer than in 2003 as 1650 years; either this really was an exceptional event or there are subtleties in higher order dependence for the extreme temperature process that are not captured by our Markov model. Making an assumption of higher order Markov processes has not been considered in this paper, but the extension to d -dimensions could be modelled using theory from Heffernan and Tawn (2004). Alternatively no Markov structure assumptions could be made, e.g. as in Eastoe and Tawn (2012), but this comes at the cost of large numbers of parameters and a high dimensional non-parametric distribution G to estimate which is likely to lead to very poor estimates of events more extreme than the event observed in 2003.

Our approach has focused on daily maximum temperatures. As outlined in Section 1, Pascal et al. (2013) point out that extremely hot night time temperatures during a heatwave can also be important in raising mortality. Thus we may be interested in extremes of the series $X_t, Y_t, X_{t+1}, Y_{t+1}, \dots$ where X_t is the daily maximum temperature and Y_t is the daily minimum temperature on day t . A Markov model is likely to be appropriate for the series. Although the series is non-stationary its components X_t and Y_t may be individually stationary with marginal distributions F_X and F_Y . Applying the marginal methods of Section 3 we can transform the X_t and Y_t series to have an identical marginal Laplace distributions. We can then model the dependence structure for the transitions between the series using the conditional extremes approach. These transitions may have parameters that vary between the pairs (X_t, Y_t) and (Y_t, X_{t+1}) but otherwise the methodology developed in the paper can be extended easily to this more general situation. Our approach has also focused on heatwaves at a single site whereas the spatial nature of an event is also critical for the economy and health. Davison and Gholamrezaee (2012) look at the heatwave problem from a spatial perspective, focusing on asymptotically dependent models only and ignoring temporal aspects. Therefore a future open line of research is to draw together our approach with theirs, requiring a fully space-time model for extremes. The first approaches to space-time extremes models are Huser and Davison (2014) and Davis et al. (2013), but these are restricted to asymptotic dependence in both space and time.

The underlying effect of climate change has been ignored during this paper and is an important future extension for each approach. Stott et al. (2004) have investigated the human contribution to the European heatwave of 2003. They suggest that it is very likely that anthropogenic climate change has at least doubled the risk of a heatwave as intense as the event in 2003 in comparison to pre-industrial times. It would be interesting to apply the approaches in this paper to assess the affect of human induced climate change which can affect both marginal and dependence characteristics of the process.

Acknowledgements

We would like to acknowledge funding from EPSRC and Met Office through a CASE studentship. We also thank Met Office for data, specifically Simon Brown for his support and help on heatwaves, and referees for their very helpful comments.

A Tail chain estimation algorithms

Algorithm 1 gives the tail chain generation method for asymptotically independent Markov chains. An exceedance of v is generated from a $\text{GPD}(\sigma_v, \xi)$ and the chain is stepped forward using equation (10) by sampling from the non-parametric estimate of \hat{G} given in equation (11). Particular care must be taken since negative values of the transformed tail chain, i.e. $T(X_t) < 0$, can lead to problems (since $\beta \in (-\infty, 1)$). Since the margins follow a Laplace distribution negative values correspond to values below the median and hence outside the tail region, so following a negative value all further chain values are set to zero and do not effect the cluster properties for the high levels of interest.

Algorithm 1: Simulation scheme to generate tail chain using conditional extremes approach

Input: Dependence parameters (α, β) and non-parametric distribution \hat{G}

- 1 Set threshold v and simulate exceedance using $\text{GPD}(\sigma_v, \xi)$ distribution;
- 2 Set exceedance as X_1^* ;
- 3 **for** i in $1 : k - 1$ **do**
- 4 Make draw Z_i^* with replacement from \hat{G} ;
- 5 Set $X_{i+1}^* = T^{-1} \left(\alpha T(X_i^*) + T(X_i^*)^\beta Z_i^* \right)$;
- 6 **end**

Output: Tail chain X_1^*, \dots, X_k^* with dependence structure given by (α, β)

A non-parametric tail chain simulation scheme based upon the conditional extremes approach provides an alternate method to generate tail chains with asymptotic dependence. As outlined in Section 3.3 this method is a special case of the conditional extremes method with $\alpha = 1$ and $\beta = 0$. It has been shown in equation (13) that the non-parametric estimate of the distribution G is given as the empirical distribution function of the set of differences between the transformed chain at times t and $t + 1$ given $X_t > u$. The chain is stepped forward by sampling a value from the set of differences with replacement and adding the value to the current value of the chain.

The parametric tail chain simulation scheme has a similar form to Algorithm 1. We start with

dependence parameter γ instead of $(\alpha = 1, \beta = 0)$ and instead of simulating from \hat{G} in step 4 we simulate U_i from Uniform(0,1) distribution, set

$$Z_i^* = -\frac{\gamma}{\sigma_u} \log \left(U_i^{1/(\gamma-1)} - 1 \right),$$

and replace step 5 with $X_{i+1}^* = X_i^* + Z_i^* [\sigma_u + \xi (X_i^* - u)]_+$.

For a forward and backward simulation strategy it is necessary to know the peak value of a cluster, here denoted $M = \eta$ where $\eta > v$. By setting $X_0^* = \eta$ a tail chain of length k can be simulated forwards using $(\hat{\alpha}_f, \hat{\beta}_f)$, the estimates of the conditional extremes dependence parameters for the forward chain, and the conditional extremes tail chain simulation scheme outlined above. One difference requires that a chain X_0^*, \dots, X_k^* be discarded if $X_j^* > \eta$ for any $j = 1, \dots, k$. An additional fit of the conditional extremes model must be made for $X_t | X_{t+1} > u$ before the backward simulation step to obtain $(\hat{\alpha}_b, \hat{\beta}_b)$, the estimates of the conditional extremes dependence parameters for the backward chain. If the Markov chain is time-reversible $\alpha_b = \alpha_f$ and $\beta_b = \beta_f$. For the backward simulation a tail chain is constructed using $(\hat{\alpha}_b, \hat{\beta}_b)$ with the same rejection criteria if η is exceeded. By combining the forward and backward simulated tail chains a cluster with peak value η is generated; see Algorithm 2.

Algorithm 2: Simulation scheme to generate realisation of cluster with peak value η using conditional extremes approach

Input: Dependence parameters $(\alpha_f, \beta_f, \alpha_b, \beta_b)$ and non-parametric distributions \hat{G}_f and \hat{G}_b

- 1 Set the peak value η of a cluster as X_0^* ;
- 2 **for** i in $1 : k$ **do**
- 3 Make draw Z_f^* with replacement from \hat{G}_f ;
- 4 Set $X_i^* = T^{-1} \left(\alpha T (X_{i-1}^*) + T (X_{i-1}^*)^\beta Z_f^* \right)$;
- 5 **if** $X_i^* > X_0^*$ **then**
- 6 | Discard current forward chain and return to step 2 ;
- 7 **end**
- 8 **end**
- 9 **for** i in $1 : k$ **do**
- 10 Make draw Z_b^* with replacement from \hat{G}_b ;
- 11 Set $X_{-i}^* = T^{-1} \left(\alpha_b T (X_{-(i-1)}^*) + T (X_{-(i-1)}^*)^{\beta_b} Z_b^* \right)$;
- 12 **if** $X_{-i}^* > X_0^*$ **then**
- 13 | Discard current backward chain and return to step 9;
- 14 **end**
- 15 **end**
- 16 Generate a cluster replicate $X^* = (X_{-k}^*, \dots, X_0^*, \dots, X_k^*)$;

Output: One realisation of a cluster X^* with maximum η and dependence structure given by $(\alpha_f, \beta_f, \alpha_b, \beta_b)$

B Pool adjacent violators algorithm

Using differencing formulas, as in equation (5), with inputs evaluated by Monte Carlo methods from tail chains can result in negative estimates of $\pi(i)$ for certain values of i . Although $\theta^{(i)} \geq \theta^{(i+1)}$ for all i , as only a finite number of chains can be simulated by Monte Carlo, Algorithms from Appendix A provide estimates $\tilde{\theta}^{(i)}$ and $\tilde{\theta}^{(i+1)}$ which do not necessarily satisfy this constraint. Thus $\tilde{\pi}(i) < 0$ if $\tilde{\theta}^{(i)} < \tilde{\theta}^{(i+1)}$. The problem is more prevalent for large values of i since the tail chain simulation scheme generates very few chains with such a large number of exceedances and hence the Monte Carlo variation is large relative to the difference between $\theta^{(i)}$ and $\theta^{(i+1)}$. A solution to this problem, to ensure that estimates $\tilde{\pi}(i)$ of $\pi(i)$ satisfy $\tilde{\pi}(i) \geq 0$ is to use the pool adjacent violators (PAV) algorithm (Robertson et al., 1988). The PAV algorithm generates a monotonically decreasing estimate of $\theta^{(i)}$ and in turn gives non-negative estimates for π and π_C . This is achieved by checking whether $\tilde{\theta}^{(i+1)} \leq \tilde{\theta}^{(i)}$ for all $i = 1, \dots, n$, if not these values are averaged, i.e.

$$\theta_*^{(i)} = \theta_*^{(i+1)} = (\tilde{\theta}^{(i+1)} + \tilde{\theta}^{(i)})/2,$$

and we check whether $\tilde{\theta}^{(i-1)} \geq \theta_*^{(i)}$. If not pooling is continued until decreasing monotonicity is satisfied. The algorithm can lead to ties in consecutive estimated $\theta^{(i)}$ values resulting in $\tilde{\pi}(i) = 0$ but avoids the situation where $\tilde{\pi}(i) < 0$.

References

- Abaurrea, J., Asin, J., Cebrian, A., and Centelles, A. (2007). Modeling and forecasting extreme hot events in the central Ebro valley, a continental-Mediterranean area. *Global and Planetary Change*, 57(1-2):43–58.
- Bortot, P. and Tawn, J. A. (1998). Models for the extremes of Markov chains. *Biometrika*, 85(4):851–867.
- Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Verlag.
- Coles, S. G., Tawn, J. A., and Smith, R. L. (1994). A seasonal Markov model for extremely low temperatures. *Environmetrics*, 5:221–239.
- Davis, R. A., Kluppelberg, C., and Steinkohl, C. (2013). Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society: Series B*, 75(5):791–819.
- Davison, A. C. and Gholamrezaee, M. M. (2012). Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468:581–608.
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B*, 52(3):393–442.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.
- Donaldson, G. C., Keatinge, W. R., and Saunders, R. D. (2003). Cardiovascular responses to heat stress and their adverse consequences in healthy and vulnerable human populations. *International Journal of Hypothermia*, 19(3):225–235.

- Eastoe, E. F. and Tawn, J. A. (2012). Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1):43–55.
- Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B*, 65(2):545–556.
- Fischer, E. M. and Schär, C. (2010). Consistent geographical patterns of changes in high-impact European heatwaves. *Nature Geoscience*, 3(6):398–403.
- Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B*, 66(3):497–546.
- Hsing, T. (1988). On the extreme order statistics for a stationary sequence. *Stochastic Processes and their Application*, 29:155–169.
- Huser, R. and Davison, A. C. (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B*, 76(2):439–461.
- Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.
- Koppe, C., Kovats, S., Jendritzky, G., and Menne, B. (2004). *Heat-Waves: Risks and Responses*. Number 2. World Health Organization.
- Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press.
- Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(2):291–306.
- Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag.
- Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.
- Nelson, R. B. (2007). *An Introduction to Copulas*. Springer, 2nd edition.
- Nitschke, M., Tucker, G. R., Hansen, A. L., Williams, S., Zhang, Y., and Bi, P. (2011). Impact of two recent extreme heat episodes on morbidity and mortality in Adelaide, South Australia: a case-series analysis. *Environmental health*, 10(1):42–51.
- O’Brien, G. L. (1987). Extreme values for stationary and Markov sequences. *The Annals of Probability*, 15(1):281–291.
- Pascal, M., Wagner, V., Le Tertre, A., Laaidi, K., Honoré, C., Bénichou, F., and Beaudeau, P. (2013). Definition of temperature thresholds: the example of the French heat wave warning

- system. *International Journal of Biometeorology*, 57(1):21–29.
- Perfekt, R. (1997). Extreme value theory for a class of Markov chains with values in \mathbb{R}^d . *Advances in Applied Probability*, 29(1):138–164.
- Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.
- Rootzén, H. (1988). Maxima and exceedances of stationary Markov chains. *Advances in Applied Probability*, 20:371–390.
- Segers, J. (2003). Functionals of clusters of extremes. *Advances in Applied Probability*, 35(4):1028–1045.
- Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11:195–210.
- Smith, R. L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability*, 29(1):37–45.
- Smith, R. L., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268.
- Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society. Series B*, 56(3):515–528.
- Stefanon, M., D’Andrea, F., and Drobinski, P. (2012). Heatwave classification over Europe and the Mediterranean region. *Environmental Research Letters*, 7(1):014023.
- Stott, P. A., Stone, D. A., and Allen, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature*, 432:610–614.
- Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.
- Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2014). Modelling across extremal dependence classes. *Submitted to Journal of the Royal Statistical Society: Series B*.
- Yun, S. (1998). The extremal index of a higher-order stationary Markov chain. *Annals of Applied Probability*, 8(2):408–437.