Characterising the changing behaviour of heatwaves with climate change

Hugo C. Winter[1], Simon J. Brown[2] and Jonathan A. Tawn[3]
Email: hugo.winter@edfenergy.com, simon.brown@metoffice.gov.uk, j.tawn@lancaster.ac.uk

[1]*EDF Energy R&D UK Centre, Interchange, 81-85 Station Road, Croydon, CR0 2AJ, UK.*,
[2]*Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK.*,
[3]*Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, U.K.*

Abstract

Understanding the impact of future heatwaves and the development of effective adaptation strategies requires knowledge of both the changes in heatwave temperatures and their durations. We develop a framework, utilising extreme value theory, which allows for the effect of a covariate on both the marginal quantiles and the temporal dependence structure of daily maximum temperatures enabling the changes in heatwave temperatures (marginal effects) to be identified separately from duration changes (dependence effects). To characterise future heatwave changes we use global mean temperature anomalies as a covariate to provide the metric for climate change. Future daily maximum temperatures and global mean temperature changes are provided by 13 general circulation models (GCMs) from the CMIP5 archive forced with predicted future emissions of radiative forcing agents from the RCP8.5 scenario. For Orleans, central France, we find that for all GCMs temporal dependence is unaffected by greenhouse gas induced climate change indicating that durations of heatwaves that exceed time varying high thresholds (i.e., the 1 year level) will not change in the future. However, all GCMs project significant changes in the temperature margins with events similar to the 2003 European heatwave increasing by 1.3°C to 2.7°C and (8.0°C to 18.7°C) for a 1°C (5°C) increase in global temperature. Collectively our results indicate there could be a significant increase in heatwave risk as the world warms with heatwaves increasing in temperature significantly faster than the global mean and local average temperatures.

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

1 Introduction

Heatwaves are events that are characterised by periods of anomalously hot days and nights which have wide ranging impacts on public health, with increased mortality (Donaldson et al., 2003) and morbidity (Mastrangelo et al., 2007), and on the agricultural sector with impacts on livestock (Dunn et al., 2014) and crop yields. For instance the 2003 heatwave in Europe was estimated to reduce maize yields by up to 36% in Italy (Ciais et al., 2005) while the 2012 heatwave in the USA reduced maize production by 13% in 2012 compared to 2011 (USDA, 2013). The definition of

heatwaves has often been tailored to the impact of interest which has resulted in a wide range of durations and critical levels being used to define a heatwave, such as crop specific levels and durations that vary regionally (Falloon et al. (2014), Wheeler et al. (2000) and Asseng et al. (2013)), which has made comparison across studies difficult. In addition, different sensitivities to the duration of heatwaves and/or their temperature complicates efforts to determine future impacts arising from climate change as information is required for both the changes in duration and temperature.

The role of human activities has been identified in the increases of surface air temperatures that have been observed (e.g., Stott et al. (2004), Christidis et al. (2011)) and potential future changes have been widely reported (e.g., Kharin et al. (2013)). However, these studies have generally been limited by the chosen heatwave metric, be it a temporal mean (such as seasonal mean, Stott et al., 2004) that is significantly longer than typical heatwaves or significantly shorter (such as daily maxima, Kharin et al. (2013), Christidis et al. (2011)). Changes in the behaviour of heatwaves can manifest themselves in either changes to temperature or duration or both and therefore require metrics that can capture both. This paper focuses on how future changes in the behaviour of heatwaves, which might arise from climate change, can be modelled using extreme value theory within a framework that can separately identify changes in critical levels (such as a return level) and the durations and magnitudes of the excursions above these changing (relative) critical levels.

Operational definitions of heatwaves are generally split into three different categories (Koppe et al., 2004) based upon: (i) an air temperature level; (ii) an air temperature level and minimum duration; (iii) indices based upon air temperature and relative humidity; in this paper we focus on cases (i) and (ii). Abaurrea et al. (2007) use values that exceed the 95th percentile of June-August daily maximum temperatures from 1971-2000. A 95th percentile is also used in Stefanon et al. (2012), but they impose the additional specification of a minimum duration of 4 days. Fischer and Schär (2010) use a local 90th percentile with a minimum duration of at least 6 consecutive days. In each approach only the exceedances of a threshold are used during the modelling process. In contrast, Winter and Tawn (2016) model the temperature process above some lower modelling threshold and then derive the properties of the heatwaves above a range of critical levels. Thus, if their model is appropriate for their data, they have a more efficient inference of heatwave properties that can also be derived beyond the range of the observed heatwaves. Relative critical levels are preferred to absolute critical 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, the importance of estimating the duration and severity of events accurately is universally recognised.

Following Winter and Tawn (2016), we analyse the future heatwave characteristics for Orleans, central France, as simulated by 13 state of the art general circulation models (GCMs) from phase 5 of the Coupled Model Intercomparison Project (CMIP5) database (Taylor et al., 2012). Orleans was chosen for it's continental location, absence of orography, long observational record and for being located within the area most affected by the 2003 European heatwave, an event used by Winter and Tawn (2016), and here, as representative of a damaging heatwave. Our ensemble includes GCMs of similar resolution and which also produce plausible representations of climate variability. The GCMs have been forced with predicted future emissions of radiative forcing agents from the RCP8.5 scenario (Riahi et al., 2011) for the period 2006-2090 which together with their differing

climate sensitivities produce increases in global mean temperatures that vary between 2.0 to 4.3 °C in the period 2006 to 2090. Such an ensemble of GCMs brings two sources of uncertainty to this analysis. Each model will have its individual representation of physical processes, some of which are poorly known and even if they are well known are imperfectly modelled within the GCM. In addition, due to the non-linear nature of the climate system each realisation of the future climate will sample the natural internal variability of the climate system, be it with the same or a different GCM. Thus our ensemble here and the resulting range of results will be a combination of these two sources of uncertainty.

We present future changes of heatwaves with respect to changes in global mean temperature rather than for a specific date as this covariate is directly applicable to discussions regarding mitigation targets and can also take advantage of efforts to formally quantify the uncertainty in future warming for specific dates (Harris et al., 2013). We calculate annual global mean temperature anomalies, g_t, with respect to 2006 for each GCM respectively

$$g_t = \tilde{g}_t - \tilde{g}_{2006},$$

where $\tilde{g}_t$ is the global mean temperature. Here $\tilde{g}_t$ and g_t vary over different GCM ensemble members.

Whilst it would be possible to use a more locally defined metric of future change (such as the change in mean European temperatures) this would unhelpfully include more unforced naturally occurring internal variability of the climate system; here we desire to identify the changes that are being forced by greenhouse gas emissions. Initially the nature of such changes are first identified for one member of this ensemble, HadGEM2.ES GCM (Martin, 2011), and then the whole ensemble is used to test whether these changes are consistent across GCMs. Although our ensemble contains some GCMs with large increases in global mean temperature, this does not generate bias and actually aids the efficiency of our study as it increases the signal we are seeking to infer, and through the global temperature anomalies, results can be rescaled to whatever time period or emission scenario as deemed appropriate by the user. We report, therefore, changes in heatwave behaviour for 1°C and 5°C global mean temperature increases to span a wide range of possible combinations of dates, emissions and climate sensitivities.

Let $\{Y_t\}$ denote the time-series of daily maximum temperatures over the June to August summer period. Our statistical modelling approach is based on, and extends, the methods developed in Winter and Tawn (2016) who focused on analysing observed temperature extremes at one location under an assumption of stationarity. If $\{Y_t\}$ was a stationary series, the intensity of a heatwave could be modelled by fitting an extreme value model to exceedances by $\{Y_t\}$ of a high modelling threshold $\tilde{u}$. The most common method is to fit a generalized Pareto distribution (GPD)

$$\mathrm{P}\left(Y_t - \tilde{u} > y \mid Y_t > \tilde{u}\right) = \left(1 + \tilde{\xi} y / \tilde{\sigma}_{\tilde{u}}\right)_+^{-1/\tilde{\xi}} \qquad \text{for} \quad y \geq 0,$$

where $c_+ = \max(c, 0)$ and $\tilde{\sigma}_{\tilde{u}} > 0$ and $\tilde{\xi}$ are the scale and shape parameters of the GPD respectively (Coles, 2001).

Under climate change $\{Y_t\}$ is non-stationary, so approaches that model exceedances above a constant threshold $\tilde{u}$ can be problematic since the sample of exceedances will be dominated by values

from certain points in the time-series (e.g., there are likely to be more exceedances in future years under climate change). Here we account for non-stationarity $\{Y_t\}$ with respect to the covariate global mean temperature anomalies g_t and analyse multiple future potential series at the same location. Our strategy for modelling the probabilistic behaviour of extreme temperatures is two-fold. Firstly, we model the marginal structure using a threshold based approach where the threshold is a function of the covariate. Once the marginal structure has been modelled, we transform the data onto common margins, with the variables now identically distributed over time/covariate. We denote the resulting series $\{X_t\}$ and take this series to have a common Laplace marginal distribution over t. We then model the extremal dependence structure of the $\{X_t\}$ series.

For modelling the dependence between successive extremes values of a stationary series $\{X_t\}$ it is helpful to be aware that there are two different types of dependence structure. At lag τ the type of extremal dependence structure is determined by the value of χ_τ where

$$\chi_\tau = \lim_{x\to\infty} \mathrm{P}(X_{t+\tau} > x \mid X_t > x). \tag{1}$$

When $\chi_\tau = 0$ and $\chi_\tau > 0$ the variables $(X_t, X_{t+\tau})$ are said to be asymptotically independent and asymptotic dependent respectively. So exceedances by the process of a very high threshold can cluster locally if the process is asymptotically dependent for at least one τ, but occur increasingly independently as $x \to \infty$ when the process is asymptotically independent for all $\tau \neq 0$ (Ledford and Tawn, 2003). For asymptotically dependent processes the difference between χ_τ and $\mathrm{P}(X_{t+\tau} > x \mid X_t > x)$ for large x is a second order effect on the characteristics of clusters of extremes events. So any clustering can be viewed to be effectively invariant to the value of x. For statistical modelling of heatwaves this means that dependence features of heatwaves don't change with the critical level (here taken to be x) used to define the heatwave. In contrast for asymptotically independent processes the difference between χ_τ and $\mathrm{P}(X_{t+\tau} > x \mid X_t > x)$ for large x is the first order feature of dependence, as $\chi_\tau = 0$, so dependence features of heatwaves will vary with the critical level and weaken as the level increases until, in the limit, there is no clustering of extremes. In practice, this means that the duration and severity of a heatwave event is permitted to change with critical level, i.e., a heatwave exceeding the 1 year return level will have different dependence characteristics than a heatwave exceeding the 50 year return level. So to model heatwaves it is vital to determine the type of extremal dependence structure that the data exhibit.

The main aim of this paper is to provide a coherent extreme value framework for investigating the effect that climate change will have on the temperature and the duration of heatwave events. Many previous studies have focused on the occurrence of singular hot days and how this might vary with climate change while ignoring changes in the persistence of events. There has been limited past work on marginal and dependence structure modelling for heatwaves. Reich et al. (2014) modelled heatwaves using a GPD to capture marginal characteristics and a max-stable process to model dependence in a Bayesian hierarchical framework. Similarly, Shaby et al. (2016) use a latent switching process to model clusters of large values in heatwaves. Neither of these models are appropriate for our data as they require consecutive values that are asymptotically dependent whereas our analysis suggests that our data exhibit asymptotic independence.

Our approach models the $\{X_t\}$ series as a first order Markov chain, with the tails of the joint distribution of (X_t, X_{t+1}) modelled using the conditional extremal dependence approach of Heffer-

nan and Tawn (2004). This approach offers a more flexible way of estimating extremal dependence properties of Markov chains than previous methods since it captures the full asymptotically justified class of extremal dependence. In comparison, Smith (1992) applies only for a restricted special case and our method holds over a much broader tail region than the approach of Bortot and Tawn (1998).

It is important to stress the difference between marginal non-stationarity and non-stationarity in the extremal dependence structure. Our marginal modelling will remove non-stationarity in the marginal distribution, however it does not account for non-stationarity between consecutive time points. In the framework of Markov chains, incorporating covariates into the conditional distribution of $X_{t+1} \mid X_t$ will allow assessment of how the dependence between values on successive days changes with a covariate. A previous study of the effect of covariates on dependence structure appears in Jonathan et al. (2013) for estimating wave heights in the North Sea as a function of the wave direction. In the heatwave problem, a change in the marginal characteristics leads to a change in the overall strength of a heatwave whereas a change in the dependence characteristics leads to a change in the persistence of events. Both of these factors are important when mitigating for heatwave events.

This paper has the following structure. Section 2 describes the steps to capture non-stationarity in the margins to obtain a marginally stationary series. The conditional extremal dependence approach is outlined and extended to include covariates in Section 3. Methods for simulating clusters of extreme values to derive heatwave properties are briefly mentioned in Section 4. Results for the HadGEM2.ES GCM are given in Section 5 and results over the rest of the GCM ensemble are given in Section 6. Discussion and conclusions are provided in Section 7.

2 Marginal modelling

2.1 Modelling strategy

Daily maximum temperatures at a site are related to a covariate g_t. As we are interested in the behaviour of extreme temperatures we need to be able to model the effect of a covariate on tail behaviour at a site. Davison and Smith (1990) and Northrop and Jonathan (2011) propose different modelling approaches to achieve this by focusing exclusively on the effect of the covariate on the tail. Here we adopt the pre-processing approach of Eastoe and Tawn (2009) where a pre-processing step removes covariate effects from the body of the distribution; residual influences of the covariates on the tails are then accounted for using the methods of Davison and Smith (1990). The pre-processing approach has close parallels with Northrop and Jonathan (2011) since the threshold for the extreme value modelling is derived to be covariate dependent. Without allowing for the threshold to change with the covariate we would have major problems as the non-stationary effect over the series is large, therefore the values that were extreme temperatures early in the series are no longer extreme late in the series. In contrast to the approach of Northrop and Jonathan (2011), the pre-processing approach has major benefits in efficiently estimating covariate effects if the effects of the covariate are somewhat similar in the body and tail of the distribution. Furthermore, here we also need a model for the covariate effect on the body of the distribution, as we are interested in non-extremes values of the process that occur after an extreme value. Northrop and Jonathan (2011) and Davison and Smith (1990) do not model this part of the distribution.

2.2 Pre-processing

Eastoe and Tawn (2009) give a framework for transforming marginally non-stationary data such that constant threshold approaches can be used. Specifically, taking the original non-stationary time-series $\{Y_t\}$ the transformation

$$[Y_t^{\kappa(g_t)} - 1]/\kappa(g_t) = \psi(g_t) + \tau(g_t)Y_t^s,$$

yields the approximately stationary standardised sequence $\{Y_t^s\}$, where $(\psi(g_t), \tau(g_t))$ are location-scale parameters, $\kappa(g_t)$ is the Box-Cox parameter and g_t is the global mean temperature anomaly. In this paper all covariates are included in a linear manner, i.e.,

$$\kappa(g_t) = \kappa_0 + \kappa_1 g_t \qquad \psi(g_t) = \psi_0 + \psi_1 g_t \qquad \log \tau^2(g_t) = \tau_0 + \tau_1 g_t.$$

Higher-order covariate relationships are possible but not investigated here.

2.3 Marginal tail non-stationarity

In practice the Box-Cox location-scale model may not completely capture all of the non-stationarity in the extremes and a GPD model is fitted to the upper tail of the margins of the standardised series $\{Y_t^s\}$ such that

$$F_{Y_t^s}(y) = \begin{cases} 1 - \lambda_{u_s}(g_t)\left[1 + \xi(g_t)\left(y - u_s\right)/\sigma_{u_s}(g_t)\right]_+^{-1/\xi(g_t)} & \text{if} \quad y \geq u_s \\ \tilde{F}(y) & \text{if} \quad y < u_s, \end{cases} \tag{2}$$

where u_s is the modelling threshold for the upper tail of the pre-processed margins, $(\sigma_{u_s}(g_t), \xi(g_t))$ are scale and shape parameters that depend on the covariate such that $\log \sigma_{u_s}(g_t) = \sigma_0 + \sigma_1 g_t$ (where σ_0 depends on the threshold u_s but the subscript is dropped for notational simplicity) and $\xi(g_t) = \xi$, $\lambda_{u_s}(g_t) = 1 - \tilde{F}(u_s)$ and $\tilde{F}(y)$ is the empirical cumulative distribution function of $\{Y_t^s\}_{t=1}^n$. It is assumed that non-stationarity in the body of the distribution is accounted for using the pre-processing and so the stationary empirical distribution function is appropriate for values that fall below or equal to the modelling threshold u_s. Throughout this study the modelling threshold u_s is set at the 90th percentile. To study the extremal dependence structure it is common to transform the marginal to a standard form. We transform Y_t^s, $t = 1, \ldots, n$ onto Laplace margins as follows

$$X_t = \begin{cases} \log\left\{2F_{Y_t^s}(Y_t^s)\right\} & \text{if} \quad F_{Y_t^s}(Y_t^s) < 0.5 \\ -\log\left\{2\left[1 - F_{Y_t^s}(Y_t^s)\right]\right\} & \text{if} \quad F_{Y_t^s}(Y_t^s) \geq 0.5, \end{cases} \tag{3}$$

where $F_{Y_t^s}$ is given in equation (2). Estimates of all the parameters for our data set are given later. We subsequently assume that $\{X_t\}$ is marginally stationary.

3 Modelling temporal dependence

3.1 Markov modelling

To obtain estimates for the duration and severity of heatwave events it is necessary to develop a model for the evolution of the temperature time-series. Here, supported by 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 time-step before. As a consequence, to model the extremes of the transformed stationary time series $X_1, \ldots, X_n$ it is only necessary to model the extremes of pairs (X_t, X_{t+1}) for $t = 1, \ldots, n-1$.

The joint tail approach developed in Smith et al. (1997) uses a bivariate extreme value distribution with a parametric dependence structure to model the extremal dependence of (X_t, X_{t+1}). This distribution leads to the variables being asymptotically dependent, and so $\chi_1 > 0$, where χ_τ is defined by expression (1). Furthermore, the Markov assumption implies that $\chi_\tau > 0$ for all τ. Thus this model is highly restrictive in applications where the variables are asymptotically independent at some lag. The semi-parametric conditional extremal dependence approach outlined in Heffernan and Tawn (2004) allows for a richer class of dependence structures and most importantly allows for both asymptotic dependence and asymptotic independence; see Winter and Tawn (2016) for details of how these two methods differ. The additional flexibility of the conditional extremal dependence approach is useful for our application and is used throughout the rest of this paper.

3.2 Semi-parametric stationary conditional extremal dependence approach

The conditional extremal dependence approach of Heffernan and Tawn (2004) and Heffernan and Resnick (2007) can be used to model the extremes of pairs (X_t, X_{t+1}) for $t = 1, \ldots, n-1$. Heffernan and Tawn (2004) gave their representation for Gumbel margins, but Keef et al. (2013) showed that a more comprehensive approach arises for Laplace margins (equation (3)). The desire is to model (X_t, X_{t+1}) using the distribution of X_{t+1} given that X_t is large (defined as exceeding a high threshold). We want to model the conditional distribution $\mathrm{P}\{X_{t+1} \leq x_{t+1} \mid X_t = x_t\}$ for large x_t. It is therefore natural to consider asymptotic models for this distribution as $x_t \to \infty$. Unless X_{t+1} is suitably normalised the resulting distribution will become degenerate in the limit. Heffernan and Tawn (2004) identify an appropriate way to normalise X_{t+1}. Under the weak simplifying dependence assumptions of Heffernan and Tawn (2004), the specification of Laplace margins ensures that the upper and lower tails are both modelled by a symmetric distribution with exponential tails and permits the definition of a single unified class of normalising functions such that the conditional distribution is

$$\mathrm{P}\left([X_{t+1} - \alpha X_t]/X_t^{\beta} \leq z, X_t - u_X > x \mid X_t > u_X\right) \to G(z)\exp(-x), \tag{4}$$

as $u_X \to \infty$, where u_X is the modelling threshold on Laplace margins, G is a non-degenerate distribution function, $\alpha \in [-1, 1]$ 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 the variables are independent then $\alpha = \beta = 0$ and $G(z)$ is the Laplace distribution function whereas $\alpha = 1$ and $\beta = 0$ corresponds to the situation of asymptotic dependence, $-1 \leq \alpha \leq 0$ to asymptotic dependence with negative extremal dependence and $0 < \alpha < 1$ or $\alpha = 0$ and $\beta > 0$ corresponds to asymptotic independence with positive extremal dependence. This representation extends the asymptotic dependence class of Smith et al. (1997) when $\alpha < 1$.

Modelling using the conditional extremal dependence approach requires the assumption that the limiting form of equation (4) holds exactly for all values of $X_t > u_X$ given that u_X is a sufficiently high threshold. Given this assumption it is possible to write the form of X_{t+1} given that $X_t > u_X$

as

$$X_{t+1} = \alpha X_t + X_t^{\beta} Z_{t+1}, \tag{5}$$

where Z_{t+1} is a random variable with distribution function G and is independent of X_t. As G does not take any simple parametric form, a temporary working assumption of Gaussianity is made, as in Keef et al. (2013), i.e., that $Z_{t+1} \sim \mathrm{N}(\mu, \gamma^2)$, so

$$X_{t+1} \mid \{X_t = x\} \sim N\left(\alpha x + \mu x^{\beta}, \gamma^2 x^{2\beta}\right) \qquad \text{for} \quad x > u_X.$$

The Gaussian assumption permits easy estimation of the set of parameters $(\alpha, \beta, \mu, \gamma)$ by standard likelihood approaches. Although it may appear that using a potential incorrect distribution for G would bias the inference, standard regression results show this is not the case. Furthermore, Lugrin et al. (2016) estimate G using a Bayesian non-parametric estimate, but that gives gives only small improvements relative to our much simpler approach.

Once we have our estimates $(\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\gamma})$, the Gaussian assumption is discarded and a non-parametric estimate of the distribution for Z_{t+1} is formed based simply on model (5). Specifically, let $t_1, \ldots, t_{n_{u_X}}$ be the indices of $t = 1, \ldots, n$ where $x_t > u_X$. Then let

$$\hat{z}_j = (x_{t_j+1} - \hat{\alpha} x_{t_j} - \hat{\mu} x_{t_j}^{\hat{\beta}}) / \hat{\gamma} x_{t_j}^{\hat{\beta}}, \tag{6}$$

for $j = 1, \ldots, n_{u_X}$, where n_{u_X} is the number of x_t exceeding the threshold u_X. Using this sample from Z_{t+1} a non-parametric estimate $\hat{G}$ of the distribution function G is formed using $\hat{z}_j$, $j = 1, \ldots, n_{u_X}$. Heffernan and Tawn (2004) present a simulation study that shows this inference method works well.

3.3 Incorporating covariates

The process of incorporating covariates into the marginal parameters was highlighted in Section 2.2. However, for a more complete analysis of the extremal behaviour it is necessary to ascertain whether the covariate has an effect on the level of extremal dependence by testing if the covariate has a significant effect on the dependence parameters. As in Jonathan et al. (2013) covariates are introduced into the set of parameters $(\alpha, \beta, \mu, \gamma)$ such that

$$\tanh^{-1}[\alpha(g_t)] = \alpha_0 + \alpha_1 g_t \qquad \tanh^{-1}[\beta(g_t)] = \beta_0 + \beta_1 g_t$$
$$\mu(g_t) = \mu_0 + \mu_1 g_t \qquad \log \gamma(g_t) = \gamma_0 + \gamma_1 g_t.$$

An inverse tanh link function is used for $\beta(g_t)$ as well as $\alpha(g_t)$ in this situation, thus constraining each to be in the range $[-1, 1]$. In practice it is very unlikely that $\beta(g_t) < -1$ as this corresponds to $X_{t+1} - \alpha(g_t) X_t$ tending rapidly to zero for large X_t, i.e., X_{t+1} is essentially deterministic given X_t so this restriction of the parameter space is not practically an issue. The impact of the covariate on the dependence structure is assessed using a likelihood ratio test. A non-parametric estimate of the distribution G is formed using equation (6) with the covariate dependent set of parameters $(\alpha(g_t), \beta(g_t), \mu(g_t), \gamma(g_t))$ and the resulting $\{Z_t\}$ are assumed to be independent and identically distributed.

4 Cluster behaviour estimation

With a marginally stationary time-series obtained by pre-processing techniques we wish to estimate whether heatwave events become longer and more severe with climate change. We define a critical level v^Y on original margins, which translates to v^X on Laplace margins, as some level of interest above which events are extreme. Such a level will often be related to the T-year return level and denoted v_T^Y or v_T^X depending on the margin of interest, with v_T^Y time dependent but v_T^X not. A cluster is an extreme event that is defined as a set of points which exceed the critical level v^X (v^Y) which are locally grouped in time but not necessarily consecutive and are preceded and followed by a set amount of non-exceedances (Smith and Weissman, 1994). For a cluster to occur the peak value of the cluster, denoted M, must exceed the critical level. We define the duration D of a cluster as the number of exceedances of v^X (v^Y) in the cluster, i.e.,

$$D = \sum_{t \in C} \mathbb{I}\left(X_t - v^X\right)_+ = \sum_{t \in C} \mathbb{I}\left(Y_t - v^Y\right)_+ ,$$

where $\mathbb{I}(.)$ is the indicator function and C is a set of values comprising a cluster. One common measure of temporal dependence linked to clusters is the threshold dependent extremal index (Leadbetter (1983); Winter and Tawn (2016)), denoted $\theta_X(v^X)$ ($\theta_Y(v^Y)$) depending on the margin of interest. The extremal index is defined as the reciprocal of the average duration of a cluster above v^X (v^Y). Since we assume stationarity within years the value of $\theta_X(v^X)$ and $\theta_Y(v^Y)$ will be approximately equivalent. There are varying definitions of the severity or relative severity of any type of extreme event (e.g., Mishra and Singh (2010)). In this paper we shall refer to the relative severity S of an event as the sum of all excesses of a critical level within an event (cluster) on the original temperature scale, i.e.,

$$S = \sum_{t \in C} \left(Y_t - v^Y\right)_+ . \tag{7}$$

In this study we look to estimate $\mathrm{P}(D > d \mid M > \eta)$ and $\mathrm{P}(S > s \mid M > \eta)$; given a peak value of a cluster is greater than some critical value η, with $\eta \geq v^Y$, these represent the probability of an event that has more than d exceedances of v^Y or has a relative severity greater than s respectively.

Our approach to deriving the properties of clusters of a Markov chain is repeated simulation of the chain in periods with exceedances of a critical level, i.e., when the process exceeds v^Y, with $v^Y \geq u_Y$. We adopt the tail chain approach of Winter and Tawn (2016), an extension of Smith et al. (1997), called peak value chain estimation, by simulating the cluster maximum $M > v^Y$ and then simulating forwards and backwards from this peak value using the conditional model. By its construction, the limiting conditional distribution induces a negative drift in the series so extreme events are well defined as if we ran these tail chains forward sufficiently no future exceedances of v^Y would result. Estimation of the forward chain is implicit in the approach in Section 3.2, estimation of the backward chain requires dependence parameters (α_b, β_b, μ_b, γ_b) for $X_t \mid X_{t+1} > u_X$ to be estimated similarly. The approach behind peak value chain estimation allows full extreme events to be simulated permitting easy estimation of relative severity and duration characteristics as well as any other function of the extreme event. From the peak value tail chain estimation approach, we obtain estimates of $\mathrm{P}(D > d \mid M > \eta)$ and $\mathrm{P}(S > s \mid M > \eta)$ where η is some cluster maximum of interest and d and s are critical values of duration and relative severity respectively. The joint probability of an event exceeding a given duration and relative severity will also be evaluated.

5 HadGEM2.ES Results

The top left plot of Figure 1 shows the HadGEM2.ES GCM temperature data represented as a time series which clearly shows marginal non-stationarity over the length of the record, but within each summer period temperatures are approximately marginally stationary. Figure 1e) shows the time-series of global mean temperature anomaly from HadGEM2.ES. We use the pre-processing method of Section 2.2 and test whether the global mean temperature anomaly covariate has a significant effect on the Box-Cox parameter and the location-scale parameters. The log-link function is used to ensure the non-negativity of the scale parameter $\tau(g_t)$. Incorporating the covariate into the Box-Cox parameter is found to be significant at the 5% level using a likelihood ratio test and this is also the case for the location-scale parameters that vary with the covariate. For this model selection we used backward elimination. Use of AIC would probably have been optimal, but was more computationally demanding as it required all models to be fitted. The estimates of the Box-Cox parameters are $\hat{\kappa}_0 = 1.018$ (0.06) and $\hat{\kappa}_1 = -0.023$ (0.001), where standard errors are given in parentheses. Estimates for the location-scale parameters are given as $\hat{\psi}_0 = 28.304$ (0.1) and $\hat{\psi}_1 = 0.097$ (0.05) and $\exp(\hat{\tau}_0)^{1/2} = 5.536$ (1.008) and $\exp(\hat{\tau}_1)^{1/2} = 0.960$ (1.003) respectively. These estimates show that as global mean temperature anomalies increase the daily maximum temperature data for Orleans increase in mean value, decrease in variability and are becoming more positively skewed.

A GPD$(\sigma_{u_s}(g_t), \xi)$ is fitted to the upper tail of the standardised data (see Section 2.3) to assess whether there is still any residual marginal non-stationarity in the extremes. Throughout this study the modelling threshold u_s for the GPD is set at the marginal 90th percentile. This choice was based on a range of diagnostic methods described by Coles (2001), with the most helpful being parameter stability plots. The covariate does not seem to have an effect on the estimate of the rate parameter $\hat{\lambda}_{u_s} = 0.10$ (0.007) or the shape parameter with $\hat{\xi} = -0.241$ (0.020), the effect having been removed by the pre-processing, but there remains an effect in the margins for the scale parameter. Estimates of the scale parameters are $\exp(\hat{\sigma}_0) = 0.513$ (1.07) and $\exp(\hat{\sigma}_1) = 0.930$ (1.03). These estimates show that as global mean temperature anomalies increase the variability in the upper tail is decreasing at a faster rate than for the body of the distribution of the daily maximum temperature data for Orleans.

These estimates are used to transform the non-stationary series (Figure 1a) into the marginally Laplace distributed series (Figure 1b). This figure shows that the marginal analysis seems to have produced an approximately marginally stationary series. Furthermore the Q-Q plot, in Figure 1c), shows that the transformed series is consistent with the required Laplace marginal distribution. Despite there being some departures from the equality line for large quantiles almost all values lie in the pointwise 95% tolerance intervals.

The marginally non-stationary nature of the time-series means that the value of a T-year return level varies with the value of the covariate. The critical level associated with the 1-year return period (denoted v_1^Y) increases by 1.8°C (1.5, 2.0) or 11.6°C (9.6, 13.1) for an increase in global mean temperature of 1°C or 5°C respectively, relative to global mean temperature in 2006. The figures in parentheses are the 95% confidence intervals, which have been obtained by bootstrapping. Here, and throughout, bootstrap samples account for uncertainty in all components of the analysis, including pre-processing. Return levels are obtained for the 50-year return period that increase by 1.7°C (1.2, 2.1) or 11.1°C (8.7, 13.1) under the same change in covariate. Both these results

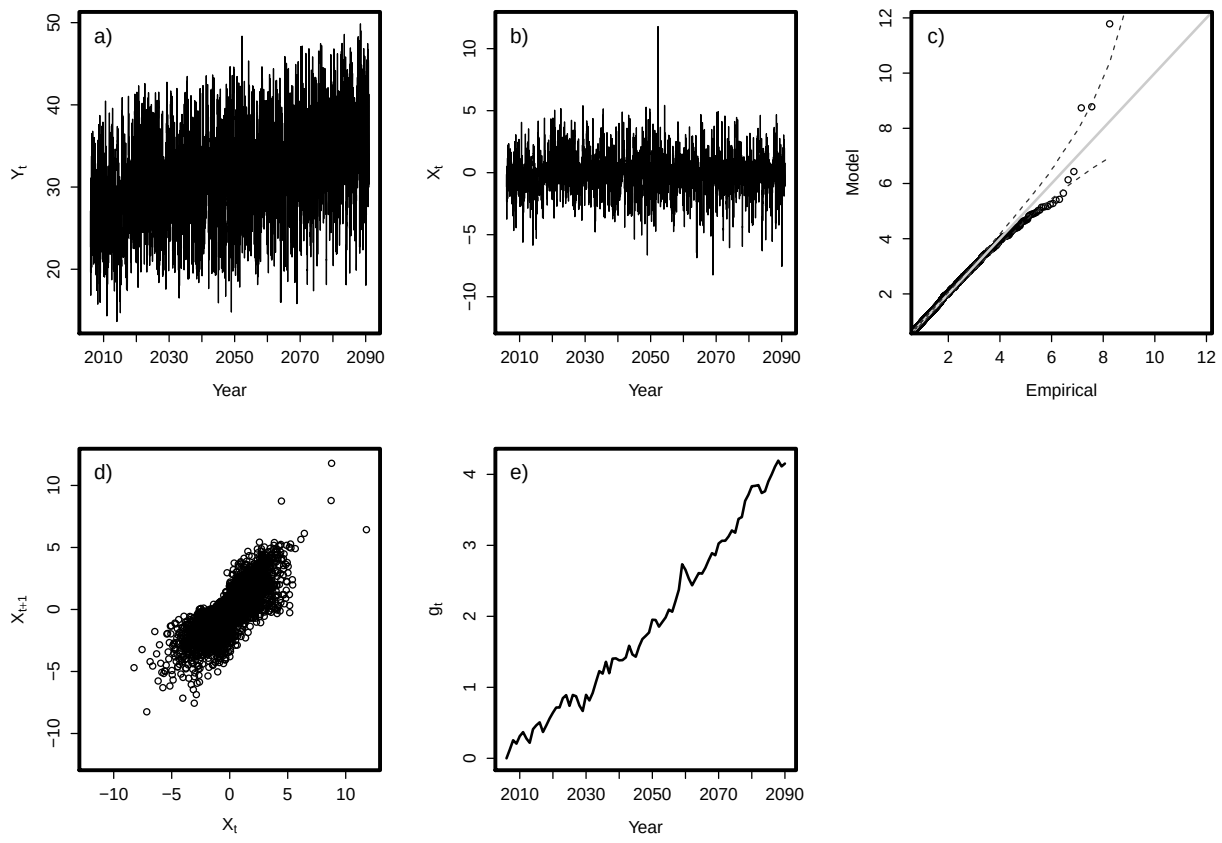

Figure 1: Original June, July and August daily maximum temperature data from HadGEM2.ES GCM represented as a) time-series (^{o}C), b) same data on Laplace margins after pre-processing, c) QQ plot of preprocessed data and theoretical Laplace quantiles, with 95% tolerance intervals, d) pre-processed data as a set of consecutive pairs and e) global mean temperature anomaly taken as covariate. Data from separate years have been concatenated for the time-series plot to show only relevant data. This plot suggests that continuity of data from year to year is induced but separate seasons are modelled as independent.

highlight that extreme temperatures over land change at a significantly greater rate than the global mean.

Estimates are now provided for the extremal dependence between consecutive days using the conditional extremal dependence approach. The approach is evaluated using the modelling threshold u_X, set at the 90th percentile. This choice of threshold was made using diagnostics suggested Heffernan and Tawn (2004), with the most helpful diagnostic being a test of independence of X_t and Z_{t+1} in model (5). Though a likelihood ratio test indicates the covariate has a significant effect on some of the key dependence parameters α and β, bootstrapping does not confirm this. We conclude the data do not exhibit any change in the dependence structure with the covariate and therefore the stationary dependence model from Winter and Tawn (2016) is used to analyse extremal dependence. Estimates of the dependence parameters α and β are given as 0.13 (-0.5, 0.7) and 0.68 (0.5, 0.8) respectively, with bootstrapped 95% confidence intervals in parentheses. Parameter values for the backward chain are given as $\alpha_b = 0.68$ (0.4, 0.9) and $\beta_b = 0.43$ (0.1,

0.5) and since a different pattern is detected in the parameters of the forwards and backwards chains, this suggests non time-reversibility. A likelihood ratio test confirms that the parameter values for both the forward and backward chains are significantly different from $(\alpha = 1, \beta = 0)$ an $(\alpha_b = 1, \beta_b = 0)$ respectively, the situation of asymptotic dependence, and so it is concluded that the data exhibit asymptotic independence. The approaches of Reich et al. (2014) and Shaby et al. (2016), which require asymptotic dependence, are therefore not appropriate for these data. The major consequence of this conclusion that the data exhibit asymptotic independence is that heatwaves above higher critical levels are of shorter duration (see the discussion in Section 1).

Since mean global temperature does not appear to have a significant effect on the extremal dependence of consecutive days we estimate the probability of observing an event with a specific duration using the stationary dependence parameters. As in Winter and Tawn (2016) we estimate three quantities. Firstly, the extremal index to give an estimate of the average length of a cluster. Secondly, the probability of an event whose 3 day average exceeds the time varying 1-year return level of the original temperature scale (v_1^Y). This is suggested to be an important quantity in terms of mortality (Pascal et al., 2013), corresponding to a potential excess mortality of up to 50% and triggering heat health warnings. Finally the probability of observing an event equal to or longer than the 2003 heatwave event, i.e., where there is an event of 11 days above the 1-year return level. These three measures provide information about the average heatwave event expected as well as giving probabilities for very severe and potentially devastating events. To estimate all three quantities we use the methods from Section 4 using our fitted Markov model. The extremal index $\theta_X(v_1^X)$ is given as 0.59 (0.5, 0.7) and suggests an average of just under 2 exceedances in a cluster. The probability of a 3 day event exceeding the 1-year return level in a single year is estimated to be 0.207 (0.1, 0.3), equivalent to an event that happens on average once every 4.8 years. Finally, the probability of a 2003 duration or longer heatwave event is estimated as 0.0005 (1×10^{-5}, 0.004), an event that happens on average once every 2000 years. These compare quite favourably with the values found by Winter and Tawn (2016), using observed temperatures, of 0.49, 0.20 and 0.0006 respectively providing a degree of validation of heatwave characteristics for this GCM though there is some indication that the GCM has a tendency for fewer exceedances per cluster.

One extension of Winter and Tawn (2016) is to consider the relative severity of a heatwave, see equation (7), alongside the duration since the non-linear nature of the marginal transformation means that the relative severity of an event could increase despite there being no difference in the duration of an event. We estimate quantiles of the distribution $S \mid M > v_1^Y$ to see if there is any change in the relative severity with an increase in the global mean temperature. An increase of 1^{o}C in the covariate g_t leads to a change in the median relative severity of -0.04^oC (-0.6, 0.6) and a 5^{o}C increase leads to a change in the median relative severity of -0.17^oC (-0.6, 0.6). The respective values for the change in the 90th quantile of severity are -0.18^oC (-4.4, 4.4) and -0.78^oC (-4.4, 4.4). The results show that S, the measure of the relative severity of a heatwave, may depend on the quantile that is assessed, however the confidence intervals suggest that the effect is not certain as zero is contained within all intervals. It should be noted that since we are using a stationary extremal dependence model all these changes are coming from the effect of the covariate on marginal parameters as previously we observed that v_1^Y increases by 1.8^{o}C for a 1^{o}C increase in the global mean temperature. Thus future changes in heatwave characteristics, as simulated by HadGEM2.ES, are almost entirely driven by changes in the margins and not through changes in

the duration nor increases in the severity once the marginal changes have been taken into account.

6 GCM ensemble results

6.1 Marginal results

We now analyse time-series for the remaining 12 GCMs alongside HadGEM2.ES. Global mean temperatures are available for each of the GCMs and respective anomalies will be used as the covariate g_t and are shown in Figure 2. Daily temperatures from the gridbox containing Orleans are presented in Figure 3. The change in global mean temperature for the period 2006-2090 ranges from 2.0 to 4.3°C across the ensemble as well as exhibiting a spread of inter-annual and longer variability. Likewise there is a considerable range of forced change and variability in the local Orleans temperature (Figure 3). Data from each GCM are taken through the pre-processing and marginal analysis outlined in Sections 2.2 and 2.3. For each GCM the most general form of covariate dependence is assumed for each of the parameters, except for the shape parameter ξ which is assumed to remain constant over covariates but differs between GCMs. This ensures consistency across the models and any changes in marginal parameters are directly comparable. As before we present results for a 1°C and 5°C increase in global mean temperatures. Although these GCMs do not exhibit a 5°C increase in their global mean temperature under this forcing scenario (RCP8.5), this is not an issue as here we are determining the dependence on global temperature not the heatwave characteristics for a specific time within a specific emission scenario, and 4.8°C is the current estimated upper bound of projections for the end of this century (IPCC, 2013).

Table 1 shows estimates for the pre-processing and marginal parameters. The Box-Cox parameter κ shows the clearest trend of all the parameters with the covariate being significant (at the 5% level) and negative for all GCMs, a pattern that was expected due to the upward trend in temperature values. The impact of the covariate becomes more difficult to interpret physically for the remaining parameters as each can be compensating or enhancing the covariate dependence of other parameters. For all GCMs the location parameters ψ_1 are found to be positive whereas $\exp(\hat{\tau}_1)^{1/2} < 1$ for all enseble members introducing a reduction in variance with increased global temperatures. The GPD covariate dependent scale parameter $\exp(\sigma_1)$ has values both greater and less than one. The shape parameter is negative as expected when looking at the upper tail of temperature values.

The separate parameter values do not give a complete picture of the impact of the covariate, as this neglects the interaction between parameters. To address this, Table 2 reports the response of 1-year and 50-year maximum daily temperature return levels for 1°C and 5°C increases in the global mean temperature relative to such temperatures in 2006. Increases are found for all ensemble members for increases in global mean temperature. It is interesting to note that increased radiative forcing can lead to varied responses at different return levels. For some members the temperature response for rarer events is larger, for some smaller, than the more frequent events confirming that an analysis of the extremes is essential alongside any analysis of the average behaviour. The amplification of the global temperature increase is confirmed across the ensemble, ranging between factors of 1.4 to 4. For the larger increase in global temperature this amplification at the extremes can be substantially larger than the amplification in the body of the distribution as measured by the changes in the median of the local summer temperatures (Table 2).

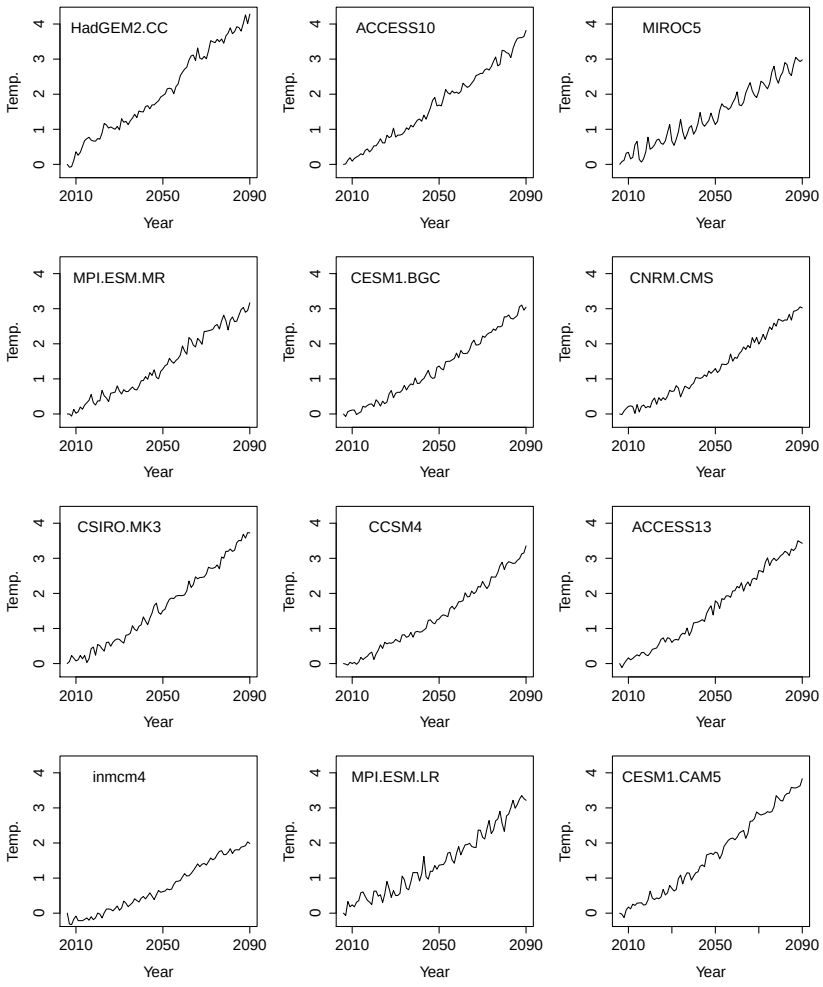

Figure 2: Annual global mean temperature anomalies (°C) for indicated GCMs forced by the RCP 8.5 emission scenario, taken as covariate.

Finally, before analysing the temporal dependence structure for each ensemble member, we need to check that the transformed series after pre-processing are stationary and with Laplace margins. Plots of data (not shown) seem to support stationarity, while the Q-Q plots for each ensemble member, shown in Figure 4, provide good support that the transformed data are well approximated by a Laplace distribution.

6.2 Dependence results

Having noted the significant increase in the magnitude of daily maximum return levels, we explore whether there is any difference in the duration and relative severity characteristics of heatwave events. In Section 5 likelihood ratio and bootstrapping tests suggested that the covariate g_t had no effect on the dependence parameters for the HadGEM2.ES series and this pattern is observed across all GCMs within our ensemble. Although the likelihood ratio tests indicated g_t dependence for a few of the GCMs further bootstrapping tests show such dependence is not robust. We repeat the analysis of the dependence between temperatures on consecutive days from Section 5 for the 13

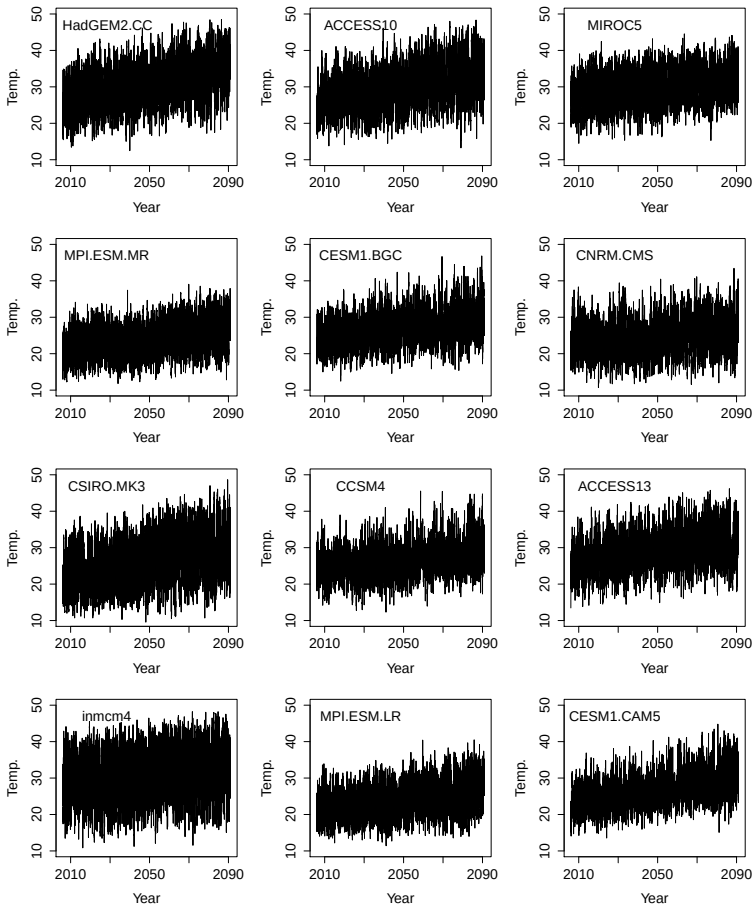

Figure 3: Original June, July and August daily maximum temperature data (^{o}C) for indicated GCMs forced by the RCP 8.5 emission scenario represented as a time-series. Temperatures are taken from the grid box containing Orleans, France.

GCMs after standardising each series to an approximately marginally stationary time-series. The extremal dependence parameters for the ensemble span the range α (0.13, 0.78) and β (0.42, 0.74). The respective values for the backward chain dependence parameters are given as α_b (0.47, 0.73) and β_b (0.13, 0.60). The physical interpretation of α and β is difficult and greater insight is gained through investigating their impact on duration and relative severity as in the following section.

6.3 Duration and relative severity results

We estimate the three measures of the duration outlined in Section 5 for each of the GCMs from our ensemble, the extremal index, the probability of an event whose 3 day average exceeds the one year return level and the probability of an event lasting for 11 days or more above the one year level (Table 3). There is a wide range of extremal indices across the ensemble which is somewhat larger than the range obtained from bootstrapping HadGEM2.ES indicating differences in GCM formulation are having an impact on daily temperature clustering characteristics and the spread is not just due to sampling and natural variability. With regard to validating the GCMs, some models have a high or low extremal index in comparison to the observed value of 0.49. In contrast,

GCM	κ_0	κ_1	ψ_0	ψ_1	$\exp(\hat{\tau}_0)^{1/2}$	$\exp(\hat{\tau}_1)^{1/2}$	$\exp(\sigma_0)$	$\exp(\sigma_1)$	ξ
HadGEM2.ES	1.018	−0.023	28.304	0.097	5.536	0.960	0.513	0.930	−0.241
HadGEM2.CC	0.621	−0.022	10.584	0.019	1.414	0.949	0.627	0.871	−0.287
ACCESS10	0.559	−0.023	9.258	0.005	1.066	0.989	0.652	0.815	−0.254
MIROC5	1.008	−0.020	27.465	0.074	4.551	0.944	0.481	0.999	−0.217
MPI.ESM.MR	0.479	−0.023	6.860	0.014	0.830	0.948	0.501	0.888	−0.221
CESM1.BGC	0.232	−0.014	4.728	0.000	0.325	0.999	0.541	1.021	−0.231
CNRM.CMS	0.430	−0.013	6.574	0.007	0.776	0.979	0.591	0.958	−0.307
CSIRO.MK3	0.674	−0.029	10.483	0.040	2.039	0.959	0.490	0.898	−0.356
CCSM4	0.204	−0.013	4.477	0.002	0.301	0.974	0.520	1.119	−0.214
ACCESS13	0.633	−0.020	10.873	0.014	1.403	0.957	0.660	0.891	−0.322
inmcm4	0.954	−0.023	24.990	0.159	5.963	0.951	0.401	0.901	−0.374
MPI.ESM.LR	0.405	−0.021	6.014	0.006	0.654	0.973	0.462	0.954	−0.239
CESM1.CAM5	−0.063	−0.012	2.853	0.000	0.113	0.977	0.658	0.944	−0.268

Table 1: Pre-processing and marginal parameter estimates for all 13 GCMs.

GCM	summer median		1 year level		50 year level	
	+1°C	+5°C	+1°C	+5°C	+1°C	+5°C
HadGEM2.ES	2.0	9.9	1.8	11.6	1.7	11.1
HadGEM2.CC	2.3	11.5	1.7	12.5	1.3	10.9
ACCESS10	2.4	12.0	2.1	15.9	1.6	14.0
MIROC5	1.6	7.8	1.5	9.1	1.6	9.4
MPI.ESM.MR	1.9	9.5	1.8	12.9	1.5	12.0
CESM1.BGC	1.7	8.5	2.6	17.5	3.2	22.0
CNRM.CMS	1.1	5.7	1.4	8.2	1.4	8.4
CSIRO.MK3	2.4	12.1	2.5	18.6	2.5	18.5
CCSM4	1.6	7.8	2.6	18.1	3.6	26.7
ACCESS13	1.9	9.4	1.6	10.8	1.3	9.8
inmcm4	2.0	10.0	1.9	12.8	1.8	12.1
MPI.ESM.LR	1.8	9.0	2.2	14.7	2.3	15.7
CESM1.CAM5	1.9	9.5	2.2	14.7	2.3	15.8

Table 2: Change in the median local summer temperature, 1-year and 50-year return levels for daily maximum temperature after 1°C and 5°C increase in global mean temperature from 2006 for all GCMs.

validating against the observed 3 day and 11 day event probabilities of 0.20 and 0.0006 we find for both metrics the GCM values lie within the bootstrapped values for HadGEM2.ES of $(0.1, 0.3)$ 3 day, and $(1\times10^{-5}, 0.004)$ 11 day. If the bootstrapped HadGEM2.ES range is taken as an acceptable proxy for the uncertainty in the observations and all GCMs this comparison suggests the GCMs are performing well against these metrics.

As in Section 5 we also estimate whether heatwave events will become more severe with increases in global mean temperature. We generate the distribution of the relative severities, $S \mid M > v_1$,

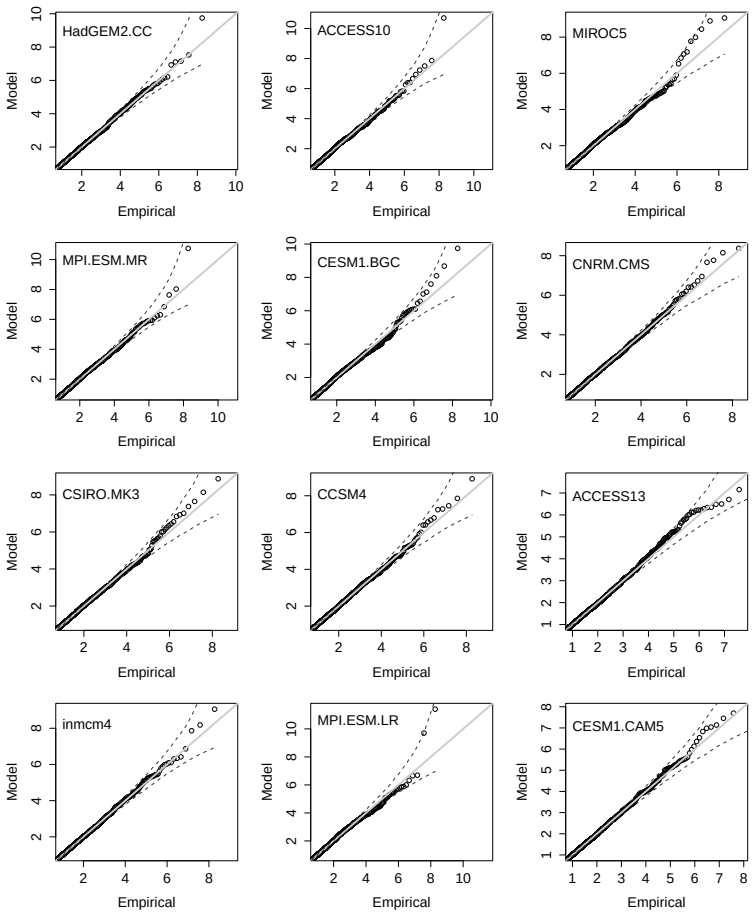

Figure 4: Q-Q plots for transformed GCM temperatures with theoretical Laplace margins. Here 95% tolerance intervals are shown.

for 1°C and 5°C increases in the global mean temperature and estimate 90^{th} quantiles. Severity changes by factors that range by $(0.9, 1.2)$ and $(0.6, 2.6)$ for 1°C and 5°C increases in g_t respectively. All GCMs exhibit severity changes that are, to first order, proportional to g_t changes, however, there is no consensus within this ensemble of the sign of change. This may appear uninformative in providing guidance on future heatwave risk, however, it should be noted that the severity metric is relative to the one year return level. Thus future severity changes are a combination of the changes to the one year level and to changes to temperatures above it. Thus some GCMs may have larger increases in the one year return level which can result in severity reductions and vice versa. Whilst this may be informative for some applications, Table 3 also includes absolute changes (threshold and exceedances changes combined) in the average temperature of 2003 type event, i.e., longer than or equal to 11 days, with average temperatures increasing by (1.3°C, 2.7°C) and (8.0°C, 18.7°C) for 1°C and 5°C increases in g_t respectively. As seen before there is a significant local amplification of the global mean changes that is consistent across the GCM ensemble. Whether this is larger or smaller than increases in the non-extreme temperatures (Table 2) is found to be GCM dependent.

GCM	Extremal index	P(duration)		Severity 90th percentile			Avg temp 11 day event		
		3 day	11 day	2006	+1°C	+5°C	2006	+1°C	+5°C
HadGEM2.ES	0.60	0.21	0.00054	6.2	−0.2	−0.8	37.2	1.9	11.9
HadGEM2.CC	0.51	0.20	0.00060	6.6	−0.6	−2.2	35.3	1.9	13.5
ACCESS10	0.40	0.20	0.00046	6.9	−0.7	−2.5	34.8	2.3	17.1
MIROC5	0.59	0.14	0.00006	4.3	0.1	0.3	35.8	1.4	8.8
MPI.ESM.MR	0.84	0.19	0.00090	5.4	−0.3	−1.1	28.9	1.9	13.3
CESM1.BGC	0.42	0.18	0.00033	5.3	0.7	5.0	31.7	2.3	15.4
CNRM.CMS	0.31	0.26	0.00224	6.8	0.0	0.3	31.4	1.3	8.0
CSIRO.MK3	0.63	0.17	0.00006	4.4	−0.1	−0.1	31.7	2.7	18.7
CCSM4	0.41	0.21	0.00060	6.0	1.2	9.4	31.5	2.2	14.6
ACCESS13	0.47	0.19	0.00042	5.6	−0.4	−1.5	34.5	1.8	11.7
inmcm4	0.33	0.14	0.00013	3.5	−0.2	−1.0	40.1	2.1	13.6
MPI.ESM.LR	0.62	0.21	0.00104	5.0	0.2	1.4	28.7	2.1	14.2
CESM1.CAM5	0.60	0.12	0.00002	4.4	0.1	1.2	29.9	1.8	14.0

Table 3: Estimates of the extremal index, the probability of an event whose 3 day average exceeds the 2006 one year return level, the probability of an event of $\geq$ 11 days above the 2006 one year level, the 90th quantiles of severity and average temperature of heatwave events with 11 day duration for all GCMs. Severity and average temperature are given at their 2006 values together with changes due to global mean temperature increases of 1°C and 5°C .

7 Discussion and Conclusion

We analyse the heatwave characteristics of 13 Global Circulation Models from CMIP5, forced with the RCP 8.5 future emission scenario, and find that increases in global mean temperature is likely to change the behaviour of heatwaves. This change arises primarily through a significant increase in marginal quantities, such as the daily return level, as opposed to increases in the duration or severity (the sum of exceedances above the 1-year return level) that are driven by the dependence structure, that is the day to day dependence of extreme temperatures. Estimated 1-year return levels for daily maximum temperature increase by between (1.4°C, 2.6°C) and (8.2°C, 18.6°C) for global temperature increases of 1°C and 5°C respectively. The dependence of extreme temperatures above this 1-year return level is not found to change with global temperatures indicating that, relative to this (increasing) 1-year return level, heatwave events are not projected to get longer. Though the duration does not change, the average temperature of long duration events, similar to the 2003 European heatwave, are found to increase, amplifying the global temperature changes by a factor of 1.3 to 3.7, more than the corresponding change in median of local summer daily temperatures. Such results confirm that future heatwaves pose a significant challenge to societies wishing to adapt to future climate.

Future changes presented here for heatwaves characteristics confirm previously published results. Kharin et al. (2013) find end of century changes for the 20 year return value of the annual maximum daily maximum temperature, using a larger enemble of CMIP5 GCMs forced with the REP8.5 scenario, of 6°C to 8°C over western Europe. Our comparable metric of daily maximum return levels (Table 2) need to be recalculated to end of century temperatures for each GCM which gives an ensemble mean of 7.0 ± 2.6°C and 7.3 ± 2.8°C for 1 year and 50 year return levels respectively.

Similarly, Sillmann et al. (2013) using a smaller set of CMIP5 models forced with RCP8.5 but with the same index at the 1 year return level find changes of 6.5°C and 7.0°C for Northern Europe and Mediterranean regions respectively, though they see a substantial spread across the ensemble (e.g., ±3°C for Northern Europe). No studies of changes in the temperature of long duration heatwaves are known to the authors with which to compare results presented here. The meteorological conditions required for heatwaves to occur, termed blocking events, are projected to modestly decrease in frequency for the European region (Masato et al., 2013). However, blocking duration is found not to change with future warming (Barnes et al., 2012 and Dunn-Sigouin and Son, 2013) consistent with temperature dependence between consecutive days remaining constant as found here. The generally large spatial nature of blocking phenomena also suggests the results found here will be indictive of changes more widely for this region.

Underlying our approach are a range of assumptions. Although these assumptions have been tested at each stage of the analysis it is worth a further discussion of the two most important. We have assumed that the temperature time-series follows a first-order Markov process to permit the modelling process outlined in the paper. The use of higher order Markov processes has been shown to capture more aspects of the observed extreme temperatures although with higher computational cost and resulting in negligible differences, from a practical perspective, in the subsequent inferences for extreme events Winter and Tawn (2017). Furthermore, we have selected thresholds for extremal marginal and dependence modelling, corresponding to the 90% marginal levels. Although we used diagnostic methods for each choice, it is clear that results will be somewhat sensitive to these choices. With the methods for threshold selection we have used, we would expect higher threshold choices to give similar point estimates to ours but with wider confidence intervals. In contrast, lower threshold choices will give different point estimates and narrower confidence intervals, reflecting bias from taking the threshold to be too low.

In this paper the global mean temperature anomaly has been used as a covariate to provide a metric of climate change; a choice that is often used (e.g. IPCC 2013). Victor and Kennel (2014) suggest that the global mean temperature alone might not be the best way of measuring the level of climate change and put forward a set of measures that include greenhouse gas concentrations and ocean heat content. We note here that the framework developed in the paper could be extended to incorporate any such covariates of interest. However, the use of global temperature allows the advances made to quantify future global temperature changes (Harris et al. 2013) to be utilised and provide a more comprehensive estimate of changes in future extremes as in Brown et al. (2014). This would require a wider assessment of the uncertainty of heatwave changes arising from the formulation of GCMs and their dependence on global temperature. It would also be interesting to investigate data on larger spatial scales to see if there is evidence for changes in the dependence structure over time within GCMs. A similar type of analysis is undertaken by Winter et al. (2016) looking for changes in spatial patterns of heatwaves with ENSO.

Acknowledgements

We gratefully acknowledge the support of the EPSRC funded EP/H023151/1 STOR-i centre for doctoral training. Simon Brown was supported and Hugo Winter partially supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). We also thank Met

Office for data, Pete Falloon for literature and background on how climate change affects food security, and the referees for extremely helpful comments that have improved the text. This work was completed while Hugo Winter was at Lancaster University. CMIP5 data can be accessed at https://pcmdi.llnl.gov/projects/cmip5/

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.

Asseng, S., Ewert, F., and Rosenzweig, C. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, 3:827–832.

Barnes, E. A., Slingo, J., and Woollings, T. (2012). A methodology for the comparison of blocking climatologies across indices, models and climate scenarios. *Climate Dynamics*, 38(11-12):2467–2481.

Bortot, P. and Tawn, J. A. (1998). Models for the extremes of Markov chains. *Biometrika*, 85(4):851–867.

Brown, S. J., Murphy, J. M., Sexton, D. M. H., and Harris, G. R. (2014). Climate projections of future extreme events accounting for modelling uncertainties and historical simulation biases. *Climare Dynamics*, 43(9-10):2681–2705.

Christidis, N., Stott, P. A., and Brown, S. J. (2011). The Role of Human Activity in the Recent Warming of Extremely Warm Daytime Temperatures. *Journal of Climate*, 24(7):1922–1930.

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A. D., Friedlingstein, P., Grunwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T., and Valentini, R. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437(7058):529–533.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Verlag.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B*, 52(3):393–442.

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.

Dunn, R. J. H., Mead, N. E., Willett, K. M., and Parker, D. E. (2014). Analysis of heat stress in uk dairy cattle and impact on milk yields. *Environmental Research Letters*, 9(6):064006.

Dunn-Sigouin, E. and Son, S. W. (2013). Northern Hemisphere blocking frequency and duration

in the CMIP5 models. *J. Geophys. Res.*, 118(3):1179–1188.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C*, 58(1):25–45.

Falloon, P., Challinor, A., Dessai, S., Hoang, L., Johnson, J., and Koehler, A. (2014). Ensembles and uncertainty in climate change impacts. *Frontiers in Environmental Science*, 2:1–7.

Fischer, E. M. and Schär, C. (2010). Consistent geographical patterns of changes in high-impact European heatwaves. *Nature Geoscience*, 3(6):398–403.

Harris, G. R., Sexton, D. M. H., Booth, B. B. B., Collins, M., and Murphy, J. M. (2013). Probabilistic projections of transient climate change. *Climate Dynamics*, 40(11-12):2937–2972.

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.

IPCC (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jonathan, P., Ewans, K., and Randell, D. (2013). Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Engineering*, 79:22–31.

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.

Kharin, V. V., Zwiers, F. W., Zhang, X., and Wehner, M. (2013). Changes in temperature and precipitation extremes in the cmip5 ensemble. *Climatic Change*, 119(2):345–357.

Koppe, C., Kovats, S., Jendritzky, G., and Menne, B. (2004). *Heat-Waves: Risks and Responses*. Number 2. World Health Organization.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(2):291–306.

Ledford, A. W. and Tawn, J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B*, 65(2):521–543.

Lugrin, T., Davison, A., and Tawn, J. A. (2016). Bayesian uncertainty management in temporal dependence of extremes. *Extremes*, 19:491–515.

Martin, G. M. (2011). The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, 4(3):723–757.

Masato, G., Hoskins, B. J., and Woollings, T. (2013). Winter and Summer Northern Hemisphere

Blocking in CMIP5 Models. *JOURNAL OF CLIMATE*, 26(18):7044–7059.

Mastrangelo, G., Fedeli, U., Visentin, C., Milan, G., Fadda, E., and Spolaore, P. (2007). Pattern and determinants of hospitalization during heat waves: an ecologic study. *BMC Public Health*, 7(1):1–8. heatwave health.

Mishra, A. K. and Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, 391(1-2):202–216.

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.

Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially dependent nonstationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22:799–809.

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.

Reich, B. J., Shaby, B. A., and Cooley, D. (2014). A hierarchical model for serially-dependent extremes: A study of heat waves in the western US. *Journal of Agricultural, Biological, and Environmental Statistics*, 19:119–135.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., and Rafaj, P. (2011). Rcp 8.5—a scenario of comparatively high greenhouse gas emissions. *Climatic Change*, 109(1):33–57.

Shaby, B. A., Reich, B. J., Cooley, D., and Kaufman, C. G. (2016). A Markov-switching model for heat waves. *Annals of Applied Statistics*, 10(1):74–93.

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, 118(4):1716–1733.

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.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93:485–498.

Victor, D. G. and Kennel, C. F. (2014). Ditch the 2°C warming goal. *Nature*, 514:30–31.

Wheeler, T. R., Craufurd, P. Q., and Ellis, R. H. (2000). Temperature variability and the yield of annual crops. *Agriculture, Ecosystems and Environment*, 82:159–167.

Winter, H. C. and Tawn, J. A. (2016). Modelling heatwaves in central France: a case study in extremal dependence. *Journal of the Royal Statistical Society: Series C*, 65(3):345–365.

Winter, H. C. and Tawn, J. A. (2017). kth-order Markov extremal models for assessing heatwave risks. *To appear in Extremes*.

Winter, H. C., Tawn, J. A., and Brown, S. J. (2016). Modelling the effect of ENSO on extreme spatial temperature events over Australia. *To appear in Annals of Applied Statistics*.