A step towards efficient inference for trends in UK extreme temperatures through distributional linkage between observations and climate model data

[1]Darmesah Gabda, [2]Jonathan Tawn and [3]Simon Brown

[1]Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Malaysia
[2]Department of Mathematics and Statistics, Lancaster University, UK
[3]Met Office Hadley Centre, UK

Abstract

The aim of this paper is to set out a strategy for improving the inference for statistical models for the distribution of annual maxima observed temperature data, with a particular focus on past and future trend estimation. The observed data are on a 25 km grid over the UK. The method involves developing a distributional linkage with models for annual maxima temperatures from an ensemble of regional and global climate numerical models. This formulation enables additional information to be incorporated through the longer records, stronger climate change signals, replications over the ensemble and spatial pooling of information over sites. We find evidence for a common trend between the observed data and the average trend over the ensemble with very limited spatial variation in the trends over the UK. The proposed model, that accounts for all the sources of uncertainty, requires a very high dimensional parametric fit, so we develop an operational strategy based on simplifying assumptions and discuss what is required to remove these restrictions. With such simplifications we demonstrate more than an order of magnitude reduction in the local response of extreme temperatures to global mean temperature changes.

Keywords: climatological data, distributional linkage, generalised extreme value distribution, spatial extremes, temperature data.

1 Introduction

Extreme events of environmental processes, such as temperature, sea levels and precipitation, are likely to be affected by global climate change. A review of climate extremes encompassing the historical record, the challenges they present to climate models and their possible future impacts is given by Easterling *et al.* (2000). The rate of climate change is not expected to be linear in time in the future, due to the lagged response of the ocean, and so global mean temperature has frequently been used as a metric to represent the time evolution of future climate change (Brown *et al.* 2014). For extreme temperatures, future changes at a location may not follow the same rate as change as the global mean temperature (Clarke et al. 2010), as there can be regional variations in the mean and variance changes, both of which effect extreme temperatures. Therefore there is a need to estimate changes in extreme temperatures at the local scale and to assess how these relate to global mean temperature change. In our analysis we treat the annual global mean temperature as a known covariate and build trend models for extreme temperatures relative to that. A full analysis of extreme temperature trends strictly needs to account for the uncertainty in this covariate, but that is outside the scope of this analysis.

When making inferences of univariate extremes of a stationary process, the starting point of most environmental statisticians is to model the distribution of the annual maxima by a generalised extreme value (GEV) distribution (Coles, 2001). The asymptotic justification for this choice comes from the GEV being the only possible non-degenerate limiting distribution of linearly normalised partial maxima of weakly mixing stationary series (Leadbetter *et al.*, 1983). The GEV has distribution function

$$G(x) = \exp\left[-\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}_{+}^{-1/\xi}\right] \tag{1}$$

with parameters: $\theta = (\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_{+} \times \mathbb{R}$ corresponding to location, scale and shape parameters and the notation $[y]_{+} = \max(y, 0)$ leads to range constraints on the GEV variable. For $\xi = 0$ (taken as the limit as $\xi \to 0$) the upper tail is exponential whereas $\xi > 0$ and $\xi < 0$ corresponds to long and short upper tailed distributions respectively. When there is non-stationarity in the annual maxima then each of the GEV parameters can be adapted to be functions of the covariates to descibe different ways that the distribution changes (Coles, 2001). However, in a wide range of environmental applications we find (based on hypothesis testing) that only the location parameter needs to depend on covariates and it can do this in a linear way. Therefore if there is only one suitable covariate then the location parameter μ of distribution (1) is replaced in year t by

$$\mu_t = \alpha + \beta g_t$$

for some covariate g_t, with the trend parameter being β. This restricted model for extremes of non-stationary data turns out to be sufficient for our analysis. We denote the distribution as being GEV($\theta_t = (\mu_t, \sigma, \xi)$). Here we take g_t as the annual global mean temperature in year t, so β is giving the change in extreme temperature for every 1°C change in annual global mean temperature. Exploratory analysis found that this formulation for the non-stationarity of annual maxima was appropriate for the data studied in this paper, see Gabda (2014). Furthermore, Gabda and Tawn (2017) proposed improving on marginal inference for the GEV distribution by using objectively determined marginal and spatial penalty functions that adapt to the data set being analysed.

There are other well known extreme value modelling approaches, such as threshold exceedances being modelled by the generalised Pareto distribution (GPD) (Davison and Smith, 1990). Threshold methods benefit from using more extreme value data and hence can be more efficient in their inferences than annual maxima methods (Coles, 2001), however, they suffer from potential sensitivity to the threshold choice which is particularly problematic when there are trends (Northrop and Jonathan, 2011). Therefore we restrict our developments to the GEV case, but note that the methods we propose in this paper, and their benefits, are also directly applicable to the GPD.

Trends in extreme values of observed environmental processes are hard to estimate with sufficient precision due to the short duration of the observational data and the relatively small climate change signal over the observation period relative to inter-annual variability. This is not helped by climate processes that can generate decadal scale and longer-term natural variability, a given phase of which can encompass a significant portion, if not all, of an observed record. In contrast, climate models can be used to obtain projections of future, as well as the past, climate changes with independent and uncorrelated realisations of internal variability. In the future the climate change signal will become larger, so climate model data has the advantage of both more data and larger signals. If such climate models represent the required physical processes adequately then they can provide an additional source of information about the current observed changes in extreme temperatures. Specifically, they may then be able to replicate the trend and or other parameters of the GEV distribution during the period of the observational data. This is the underlying assumption adopted here in the use of climate model data to help infer current trends. However, complications with this approach may arise from the observed trend signal potentially being so weak and so providing no real constraint on the climate model trends. In addition, different climate models can produce significantly different trends that arise from their differing representations of the relevant physical processes which complicates their use in inferring the "true" observed changes.

There has been a range of work aiming to jointly characterise observed and climate model data trends. Wuebbles *et al.* (2013) examined the ability of climate model data to capture the observed trends of temperature extremes and heavy precipitation in the United States. Several studies have developed methodologies for modelling observed extreme events with considerations to the uncertainty in the prediction of future climate. Hanel and Buishand (2011) modelled the precipitation from regional climate models (RCM) and gridded observations and found that their estimates from the RCM exhibited a large bias relative to such estimates from observational data. In contrast Kyselý (2002) modelled the annual maximum and the minimum temperatures in observations and RCM and though a multiple regression downscaling method they were able to produce realistic return values of annual maximum and minimum temperatures. Other examples of similar work are given by Katz (2002), Stott and Forest (2007), Coelho *et al.* (2008), Hanel *et al.* (2009) and Nikulin *et al.* (2011).

A key feature with all of these studies is that when the distribution of the observed extreme events is modelled, the parameters have been naively linked, by construction, to the parameters of the distribution of extremes for the climate model, but the uncertainty of these linking parameters or of the climate model extremal parameters have not been accounted for. Since the future level of climate change is uncertain, a wide range of estimates can be made, and therefore it is necessary that such uncertainty is adequately accounted for in any analysis of changes in extremes. Brown *et al.* (2014), for example, model extreme events (temperature and rainfall) using the information from an ensemble of models consisting of both global climate models (GCM) and RCMs. They find a considerable spread in the temperature dependent parameters when fitted to individual ensemble members and that the agreement between values for RCMs and their driving GCMs can be poor and in some cases counter physical (see their Figure 8).

Our objective is to improve inferences for a statistical model of observed temperature maxima by linking the parameters of observed UK temperatures to the equivalent parameters from an ensemble of RCMs and GCMs representing differing, but plausible, future climates. We will explore the linkages for all parameters but pay particular attention to the trends relative to global mean temperature. Unlike previous studies we will jointly account for the uncertainty in the parameters, leading to a statistical model for all UK gridpoints. We derive inference for this large number of parameters via Bayesian methods which enables us to account for the uncertainty of the parameter estimates and which enables us to efficiently pool all the information in our model inferences. However, we have to be aware that we are pooling dependent data. This arises from using data from multiple sites on a spatial grid for the same year and from inter-connected members of the RCM and GCM ensemble. RCMs require boundary conditions which are taken from "parent" GCMs which have the same model formulation as the RCM apart from scale dependent parameters. Therefore consideration of the dependence between these models is required, and we believe we are the first to account for this feature. Additionally, unlike Brown *et al.* (2014), the philosophy here is to consider the GCM ensemble as random sample of possible GCMs with differences assumed to be due to some stochastic process (be it internal sampling variability or GCM formulation) and so aim to find links not just from one individual climate model to the observational data but a common linkage derived from all climate models that are employed.

The outline of this article as follows. Section 2 describes the data used in this study and presents our outline modelling strategy. Our highly ambitious modelling strategy is described in Section 3 identifies key structure in the model parameters. In Section 4 the joint inference of our proposed full model is discussed. The results of applying a simplified version of this model, that ignores the spatial dependence and treats the GCM parameters as known, are presented in Section 5. Section 6 provides a discussion how the simplifications are likely to have affected the results and discusses ways that the inference could be improved.

2 Data and basic model structure

2.1 Data

This study uses observed UK temperature annual maxima at each of 439 sites on a 25 km spatial grid from 1960 to 2009. From the climate model simulations we have temperature annual maxima data from 1950 to 2099 from RCMs with the same spatial grid as the observed data and also from coupled GCMs with a larger grid of 300km which results in 5 grid boxes over the UK domain. We denote the respective time periods with these different data types by T_1 and T_2, with $\mid T_1 \mid= 50$ and $\mid T_2 \mid= 150$. The GCM and RCM models form part of the UK Climate Projections (Murphy et al. 2009) and were specifically designed to sample uncertainty in the future climate response through the perturbation of key but imperfectly understood physical processes. This ensemble provides a range of future climates that are consistent with historical observations and with projections from other climate models (Collins et al. 2011). We focus on an ensemble consisting of 11 GCM members that run from 1950 to 2006 with observed levels of greenhouse gasses and other forcings and thereafter follow the SRES A1B emissions scenario (Nakićenović et al. 2000) to 2099. Each of these GCMs provide boundary conditions to force an additional RCM ensemble with each RCM member having the same parameter perturbations as its "parent" GCM thereby sharing the parameter perturbations and the GCMs' internally generated natural variability. In addition, the annual global mean temperature from 1950 to 2099 for each of the GCMs and the observed global mean temperature for the period of 1960-2009 are available as covariates. For the five GCM grid boxes regions ($r = 1, \dots, 5$) the associated RCM models and observational data have

h_r different sites, $(h_1, \ldots, h_5) = (98, 94, 124, 23, 100)$.

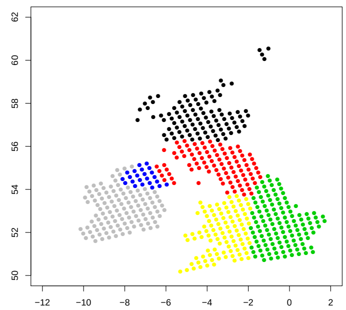


Figure 1: The location of 5 regions with respective of number of points, h_r. Black - Region 1, Red - Region 2, Green - Region 3, Blue - Region 4, Yellow - Region 5.

2.2 Basic model formulation

Let $X_{t,(r,s)}$ denote the observed annual temperature in year t for site s in region r, with $t \in T_1$, $r = 1, \ldots, 5$ and $s = 1, \ldots, h_r$. Focusing on a single site, we assume that $X_{t,(r,s)}$ are independent over t and follow a generalised extreme value distribution,

$$X_{t,(r,s)} \sim \text{GEV}\left(\alpha_{X,(r,s)} + \beta_{X,(r,s)} g_{X,t}, \sigma_{X,(r,s)}, \xi_{X,(r,s)}\right) \tag{2}$$

with a linear trend in a location parameter with covariate $g_{X,t}$ being the observed annual global mean temperature in year t. 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 to which the global mean temperature is better suited. Note that the parameters $(\alpha_{X,(r,s)}, \beta_{X,(r,s)}, \sigma_{X,(r,s)}, \xi_{X,(r,s)})$ do not depend on time, but can vary over region and site. Our choice for these parameters to be independent of time is based on a range of reasons, which include exploratory analysis which shows no evidence of a change in the distribution of residuals around a linear trend (Gabda, 2014) and the pooled assessment of fit over all sites, see Section 3.1.

Now consider the 11 coupled RCM and GCM datasets with maxima in T_2. Let $Y^{(j)}_{t,(r,s)}$ and $Z^{(j)}_{t,r}$ be the RCM and GCM annual maxima respectively in year t, region r, for the j^{th} member of an ensemble, $j = 1, \ldots, 11$ and for site s in region r for the RCM. Then we model

$$Y^{(j)}_{t,(r,s)} \sim \text{GEV}\left(\alpha^{(j)}_{Y,(r,s)} + \beta^{(j)}_{Y,(r,s)} g^{(j)}_{M,t}, \sigma^{(j)}_{Y,(r,s)}, \xi^{(j)}_{Y,(r,s)}\right) \tag{3}$$

and

$$Z^{(j)}_{t,r} \sim \text{GEV}\left(\alpha^{(j)}_{Z,r} + \beta^{(j)}_{Z,r} g^{(j)}_{M,t}, \sigma^{(j)}_{Z,r}, \xi^{(j)}_{Z,r}\right), \tag{4}$$

where $g^{(j)}_{M,t}$ is the GCM numerical model annual global mean temperature for year t in the jth ensemble member.

Here the jth GCM is used to drive the jth RCM, so there is potentially dependence between $Z_{t,r}^{(j)}$ and $Y_{t,(r,s)}^{(j)}$, for each $s = 1, \ldots, h_r$ and for all j, t and r. As $(Y_{t,(r,s)}^{(j)}, Z_{t,r}^{(j)})$, represent dependent componentwise maxima in year t, it is natural to model their joint distribution by a bivariate extreme value distribution (Tawn, 1988). This distribution has GEV marginals and a class of copula that has a restricted formulation, limited to a particular form of non-negative dependence, though it cannot be expressed fully through any finite closed form family. Therefore it is common to take a flexible parametric family in this class of copula, with the most widely used form being the logistic model. Then in year t, the joint distribution of $(Y_{t,(r,s)}^{(j)}, Z_{t,r}^{(j)})$ has the form:

$$G_{t,(r,s)}^{(j)}(y,z) = \exp\left\{-\left(a_y^{-1/\phi} + a_z^{-1/\phi}\right)^{\phi}\right\} \tag{5}$$

where

$$a_y = \left\{1 + \xi_Y\left(\frac{y - \mu_{Y,t}}{\sigma_{Y,t}}\right)\right\}_+^{1/\xi_Y} \qquad a_z = \left\{1 + \xi_Z\left(\frac{z - \mu_{Z,t}}{\sigma_{Z,t}}\right)\right\}_+^{1/\xi_Z},$$

and the dependence parameter $0 < \phi \leq 1$ measures the dependence between the regional model data, $Y_{t,(r,s)}^{(j)}$ and the global model data, $Z_{t,r}^{(j)}$, with dependence increasing from independence ($\phi = 1$) to perfect dependence ($\phi \to 0$) as ϕ decreases. The dependence parameter ϕ is found later to be constant over all sites and regions.

3 Exploratory analysis findings

3.1 Assessing model fit

In Section 2.2 we identified the theoretically motivated GEV distribution as a potential model for each marginal distribution and proposed it would be sufficient for the trends in global annual mean temperature to be modelled through the location parameters only. To assess the validity of this assumption we examined the goodness of the GEV fit to the observations and the climatological model data for each site through Q-Q plots for a set of randomly selected sites. In all cases the fit appeared good, though of course at this level of spatial resolution there are limited data to identify any deviation from the GEV assumption. Therefore, additionally, we constructed pooled P-P plots for each of the observed, RCM and GCM data separately, in each case pooling over sites, regions and years, see Heffernan and Tawn (2001) for a similar example.

These figures are shown in Figure 2. The observational data pooled P-P plot, left panel, is constructed as follows. Let $G_{X_{t,(r,s)}}$ denote the distribution function of $X_{t,(r,s)}$ as given by expression (2). Then for each t, r, s the values $\hat{G}_{X_{t,(r,s)}}(x_{t,(r,s)})$, where $\hat{G}_{X_{t,(r,s)}}$ and $x_{t,(r,s)}$ denote the marginally fitted distribution and the observed data respectively, are sorted and are compared against quantiles of the uniform$(0, 1)$ distribution. The RCM and GCM plots have been constructed similarly with additional replications over ensemble members. Here, and throughout the exploratory analysis, we use likelihood-based inference instead of a full Bayesian analysis for both computational speed and its simplicity of model selection. The results show that the GEV with a trend in the location parameter fits the data well, with a near linear P-P plot for each data type. It should be stressed that here the respective subplots correspond to 21950, 724350 and 8250 data values, thus the near perfect straight line shows the model to be an excellent fit in all three cases, given the immense data volume.

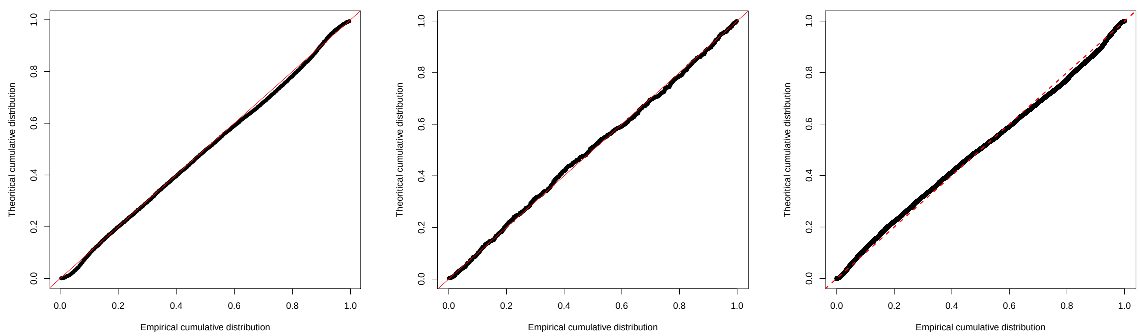


Figure 2: Pooled P-P plots for observed, RCM and GCM annual maximum temperature data (respectively) under GEV marginal models with a trend in the location parameter that is linear in global annual mean temperature.

In these models all the GEV parameters are specified as free, not depending in any way on the parameters of other variables (observed, RCM and GCM) or on the parameters at different sites. Thus the number of parameters is 21292 in total with a break down of 1756 (439×4) for observed data, 19316 ($439 \times 4 \times 11$) for RCM data and 220 ($5 \times 4 \times 11$) for GCM data. In Sections 3.3 and 3.4 respectively we explore, through a detailed exploratory analysis, if we can find any structure between the different parameters. The reason we search for structure between the parameters is that if we can find links, particularly between observation and RCM parameters, then this gives us a greater handle on how climate change will affect the observations, reduce the total number of required parameters and help to improve the efficiency of inference for the observed maxima data. Specifically, as a result of the analysis in Sections 3.3 and 3.4, the number of free parameters is reduced to 1138, a 95% reduction.

3.2 Basic assessment of trends

To help get a first impression on the trends in the different data sets and in the distinct periods of these data sets we fitted the models set out in equations (2), (3) and (4). Specifically, for region r we have h_r estimates of $\beta_{X,(r,s)}$ for each s; $11h_r$ estimates of $\beta^{(j)}_{Y,(r,s)}$ for each s and the 11 ensemble members; and 11 estimates of $\beta^{(j)}_{Z,r}$ for the 11 ensemble members. In Figure 3 we present these estimates, in the form of kernel density estimates for each region and based on 3 different time periods corresponding to the observed data 1960-2009, a future period 2010-2099 covered only by the GCM/RCM models and the full GCM/RCM data 1950-2099. These distributions only show the variation in estimates over sites and ensembles and do not account in any way for the different uncertainties in these estimates.

First consider the results in Figure 3 (left panel). Here we can see that a number of the northern regions (regions 1,2 and 4) have significant proportion of observed trends with values higher than the GCM/RCM models. Some of these are unrealistic e.g., in Northern Ireland with temperatures warming 3 times faster than global annual mean temperatures. Probably this can be explained by local variability in the short observed records and as we will see in Section 3.3 there is no statistically significant difference in observed and RCM trends over sites. Furthermore, by comparison of the GCM/RCM trends over this period with the two other periods, we see no reason to identify separate trends, relative to annual global mean temperature, in Northern Ireland for the different periods. What we can see from comparing Figure 3 left and centre panels is that the RCM/GCM trend estimates seem not to change over the 1960-2099 time period and from comparing the left and right panels that using the longest time period of 1950-2099 gives much less variation in point estimates relative to using just the period of the observed data 1960-2009.

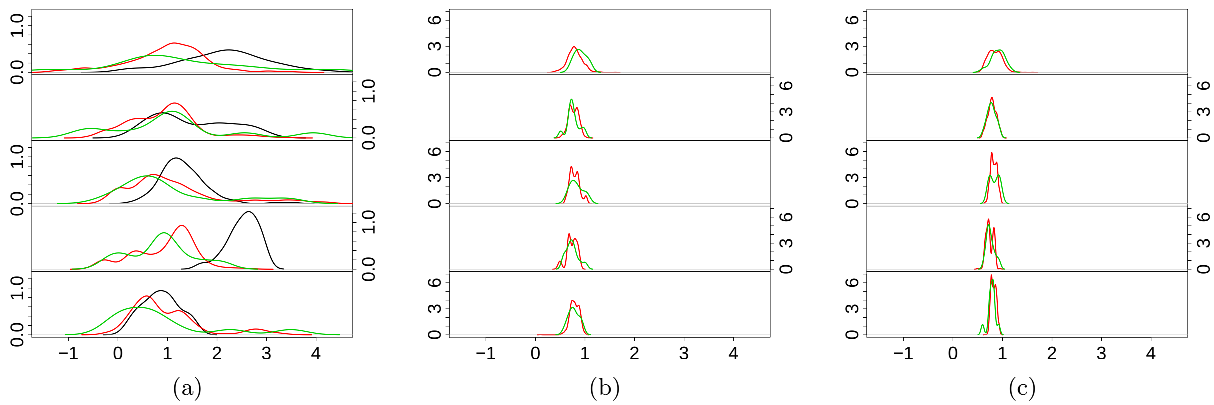


Figure 3: Distribution of the trend parameter estimates for three different periods for each region, Figure top-bottom in a sub-plot: Region 1 to Region 5: observed temperatures (black), RCM (red) and GCM (green). Panels left to right show respectively the estimates based on data for in the intervals corresponding to the observed data 1960-2009, a future period covered only by the GCM/RCM models 2010-2099 and the full GCM/RCM data 1950-2099.

3.3 Observed and RCM parameter linkage

For each site s, in region r, we test for commonality of the GEV parameters for the observed and the RCM data to see which features of the RCM maxima replicate well the features of the observed data maxima. Specifically, we test which components of the parameter vectors

$$(\alpha_{X,(r,s)}, \beta_{X,(r,s)}, \sigma_{X,(r,s)}, \xi_{X,(r,s)}) \text{ and } (\alpha^{(j)}_{Y,(r,s)}, \beta^{(j)}_{Y,(r,s)}, \sigma^{(j)}_{Y,(r,s)}, \xi^{(j)}_{Y,(r,s)})$$

are equal across ensemble members (j) for each (r, s). We present a full discussion of our analysis for the linear gradient parameter and report our findings for the other parameters. Firstly, for the majority of model fits likelihood ratio tests (which exploit the independence of observed and RCM data), with a 5% significance level, are not rejected over the $4829 (= 11 \times \sum_{r=1}^{5} h_r)$ tests. However, the proportion rejected is significantly greater than 5%, and it is not meaningful to consider the observed data trend as being equal to each of the 11 different ensemble trends. Therefore, for each location it is more realistic to think that the RCM ensemble members produce a distribution of possible trends, with the mean of these representing the observed trend. Thus we instead test the hypothesis that

$$\beta_{X,(r,s)} = \frac{1}{11} \sum_{j=1}^{11} \beta^{(j)}_{Y,(r,s)}, \tag{6}$$

for each (r, s). Separately for each site, this test involves a joint fit of the observed data and the 11 RCM ensemble members, exploiting their independence. This test is rejected with a proportion much closer to the size of the test than previously, and therefore we believe that the observed trend is well-captured by the mean of the ensemble of the RCM trends. In addition the mean of the ensemble RCM trends is estimated with a much smaller standard error than $\beta_{X,(r,s)}$ when estimated based on observed data alone. Thus, this identification of a linkage between the parameters gives improved estimation of $\beta_{X,(r,s)}$ through the additional information provided by the RCM data.

In terms of other parameters it is clear that the individual, and average, RCM parameters are statistically significantly different to the parameters of the observed data for both trend intercept (α) and shape parameters (ξ). In contrast, the scale parameters are found to have a similar linkage to the trend gradient, so that for each (r, s)

$$\sigma_{X,(r,s)} = \frac{1}{11}\sum_{j=1}^{11}\sigma^{(j)}_{Y,(r,s)}. \tag{7}$$

3.4 RCM and GCM parameter linkage

For each site s in region r we test for commonality of the GEV parameters for the RCM and the GCM data to see which features of the GCM maxima replicate well the features of the RCM data maxima. Specifically, we test which components of the parameter vectors

$$(\alpha^{(j)}_{Y,(r,s)}, \beta^{(j)}_{Y,(r,s)}, \sigma^{(j)}_{Y,(r,s)}, \xi^{(j)}_{Y,(r,s)}) \text{ and } (\alpha^{(j)}_{Z,r}, \beta^{(j)}_{Z,r}, \sigma^{(j)}_{Z,r}, \xi^{(j)}_{Z,r})$$

are equal over j for each (r, s). When testing such hypotheses we need to account for the dependence between the RCM and GCM for a given (r, s). Using the bivariate extreme value distribution model proposed in Section 2.2, with dependence parameter ϕ, we model the dependence between the RCM $Y^{(j)}$ and the GCM $Z^{(j)}$ for the j^{th} ensemble member. For each (r, s) we get very similar values for the estimated ϕ, with the average value for each of the 5 regions being $(0.56, 0.52, 0.55, 0.52, 0.57)$, with the values not being statistically significantly different at the 5% level. Thus there is no evidence for dependence between RCM and GCM varying over the UK, and a common value of ϕ over ensemble and site can be taken.

For computational simplicity we fixed $\phi = 0.55$ and then tested the required hypotheses on the marginal parameters at the 5% significance level. Again we focus discussion on the trend gradient parameter. Firstly we test for $\beta^{(j)}_{Y,(r,s)} = \beta^{(j)}_{Z,r}$ for all j, r and s, with 83.2% of the tests not rejected, which is substantially in excess of the size of the tests. Next we tested if these trends were linearly related over a region, i.e., if $\beta^{(j)}_{Y,(r,s)} = \kappa^{(r)}_{\beta_0} + \kappa^{(r)}_{\beta_1}\beta^{(j)}_{Z,r}$ with parameters $\kappa^r_{\beta_0}$ and $\kappa^r_{\beta_1}$. This also did not give a convincing fit and additionally led to the estimates of the trends $\beta^{(j)}_{Y,(r,s)}$ having clear jumps at region boundaries. Of course a driving feature for this is the discontinuity in the GCM trends over regions.

As we require observed and RCM trend parameters to change smoothly over a region and across region boundaries we now propose smoothing the GCM region trends, across sites in the region, by constructing weighted means of GCM region trends. Specifically, we define a jth ensemble smoothed GCM trend

$\beta^{(j)}_{Z,(r,s)}$, at site s in region r, as

$$\beta^{(j)}_{Z,(r,s)} = \sum_{\ell=1}^{5} w_{\ell,s}\beta^{(j)}_{Z,\ell} \tag{8}$$

where $\beta^{(j)}_{Z,\ell}$ is the GCM trend parameter for region, ℓ for the jth ensemble member, and its weight, $w_{\ell,s}$, is some monotone decreasing function of the distance $d_{\ell,s}$ of site s to the centre of region ℓ. Here, for simplicity reasons only, we take the function of distance to be the inverse squared distance, so that

$$w_{r,s} = \frac{d_{r,s}^{-2}}{\sum_{\ell=1}^{5} d_{\ell,s}^{-2}}. \tag{9}$$

However, a more flexible alternative, discussed in Section 6, allows for the level of smoothing to adapt to the smoothness of the trends in the RCM. We then find that we reject the hypothesis of

$$\begin{aligned} \beta^{(j)}_{Y,(r,s)} &= \kappa^{(r)}_{\beta_0} + \kappa^{(r)}_{\beta_1}\beta^{(j)}_{Z,(r,s)} \\ &= \kappa^{(r)}_{\beta_0} + \kappa^{(r)}_{\beta_1}\sum_{\ell=1}^{5} w_{\ell,s}\beta^{(j)}_{Z,\ell} \end{aligned} \tag{10}$$

at approximately the size of the test. Thus this linkage between RCM and GCM trends seems reasonable. Therefore, exploiting the linkages (6) and (10), the model we adopt to link the GCM to the observed data trend is via

$$\beta_{X,(r,s)} = \kappa^{(r)}_{\beta_0} + \kappa^{(r)}_{\beta_1}\sum_{j=1}^{11}\sum_{\ell=1}^{5} w_{\ell,s}\beta^{(j)}_{Z,\ell}/11. \tag{11}$$

We repeat the same analysis for the location-intercept, scale and shape parameters to give

$$\alpha^{(j)}_{Y,(r,s)} = \kappa_{\alpha_0} + \kappa_{\alpha_1}\alpha^{(j)}_{Z,(r,s)} = \kappa_{\alpha_0} + \kappa_{\alpha_1}\sum_{\ell=1}^{5} w_{\ell,s}\alpha^{(j)}_{Z,\ell} \tag{12}$$

$$\sigma^{(j)}_{Y,(r,s)} = \kappa^{(r)}_{\sigma_0} + \kappa^{(r)}_{\sigma_1}\sum_{\ell=1}^{5} w_{\ell,s}\sigma^{(j)}_{Z,\ell} \tag{13}$$

$$\sigma_{X,(r,s)} = \kappa^{(r)}_{\sigma_0} + \kappa^{(r)}_{\sigma_1}\sum_{j=1}^{11}\sum_{\ell=1}^{5} w_{\ell,s}\sigma^{(j)}_{Z,\ell}/11 \tag{14}$$

$$\xi^{(j)}_{Y,(r,s)} = \kappa_{\xi_0} + \kappa_{\xi_1}\xi^{(j)}_{Z,(r,s)}, \tag{15}$$

where $\alpha^{(j)}_{Z,(r,s)}$, $\sigma^{(j)}_{Z,(r,s)}$ and $\xi^{(j)}_{Z,(r,s)}$ are smoothed GCM parameters, defined similarly to the GCM smoothed trend (8).

4 Joint Modelling

In Section 3 we identified structure between the parameters of the GEV distributions for observed, RCM and GCM data. If this structure is a reasonable approximation this leaves us with 1138 unknown free marginal parameters instead of the original 21292 free marginal parameters. Furthermore, the exploratory analysis has shown that we only need 1 dependence parameter ϕ. The 1138 parameters comprise: 220 GCM parameters $(\alpha^{(j)}_{Z,r}, \beta^{(j)}_{Z,r}, \sigma^{(j)}_{Z,r}, \xi^{(j)}_{Z,r})$ over $r = 1, \dots, 5$ and $j = 1, \dots, 11$; 878 observed data parameters $(\alpha_{X,(r,s)}, \xi_{X,(r,s)})$ over all 439 sites; and 40 linking parameters $(\kappa^{(r)}_{\alpha_0}, \kappa^{(r)}_{\alpha_1}, \kappa^{(r)}_{\beta_0}, \kappa^{(r)}_{\beta_1}, \kappa^{(r)}_{\sigma_0}, \kappa^{(r)}_{\sigma_1}, \kappa^{(r)}_{\xi_0}, \kappa^{(r)}_{\xi_1})$ for $r = 1, \dots, 5$. The remaining parameters are given as functions of these parameters through expressions (11) and (14) and for observed location intercept and shape parameters and expressions (12), (10), (13) and (15) respectively for RCM location intercept, gradient, scale and shape parameters, We use all the information from the observed extremes data and the RCM and GCM annual maxima to estimate these free parameters, thus a total of 754550 data (50×439 observed values, $150 \times 11 \times 439$ RCM data and $150 \times 11 \times 5$ GCM data).

We could impose some additional structure on the remaining 1138 parameters to reduce the dimensionality of the problem. For example, we would expect that the location and shape parameters $(\alpha_{X,(r,s)}, \xi_{X,(r,s)})$ of the observed data will each individually change smoothly over r and s. In many cases in spatial environmental extreme value modelling no evidence is found for the shape parameter to vary over space. However, those conclusions are often derived from analyses over small spatial regions and limited data. Over larger regions there is evidence for the shape parameter to change, but to change slowly and smoothly. So one approach could be to impose some measure of smoothness over space for the shape parameter, e.g., parametric models (Coles, 2001), smoothing splines (Jonathan, *et al.*, 2014) or generalised additive models (Chavez-Demoulin and Davison, 2005) with latitude and longitude as covariates. However, we anticipate that there are likely to be coastal effects and that they may be lost by immediately fitting such a smooth model over the whole of the UK. We are even less confident about spatial smoothing for the location parameters, at least without much further investigation. This is due to the location parameters being likely to be influenced by distance from the coast, altitude and other topographic features. Therefore at a first level of investigation we prefer not to impose such smooth structure on these parameter, but in Section 6 we return to this issue when we have gathered more information from fitting our unconstrained model.

Given the complex structure of the model, with the very large number of parameters, Bayesian inference is implemented as opposed to our earlier use of likelihood-based methods as the Bayesian approach represents the information in the likelihood surface better, it avoids problems such as getting stuck in local modes, and it fully accounts for all parameter uncertainty in subsequent inferences. As there is no information available about the parameters, other than from the data, we set priors to be non-informative with a large variance, e.g., $N(0, 100^2)$, after the parameters are transformed via a link function onto the space $(-\infty, \infty)$. We apply random walk Metropolis Hastings algorithm to obtain a sample from the posterior distribution of the parameters of our proposed model, where we update each parameter by independently drawing a proposal from the Normal distribution with mean equal to the current value and a value of the variance (tuning parameter) chosen to ensure that the chain mixes suitably, typically set so that the acceptance probability is about 1/4 (Roberts and Rosenthal, 2001). We undertook 10,000 iterations for each region, after a suitable burn-in period.

We need to derive the likelihood for our model, however this is complex due to the various variables, parameter linkages and spatial dependence structure. First consider the likelihood function for a given site located at (r, s). By considering the dependency between the RCM and GCM in region r, and the independence over ensemble members and the independence of the RCM/GCM data from the observed data, then the likelihood function can be written as follows

$$L_{(r,s)} = \left\{ \prod_{t \in T_1} g\left(x_{t,(r,s)}; \boldsymbol{\theta}_{X,(r,s)}\right) \right\} \left\{ \prod_{t \in T_2} \prod_{j=1}^{11} g_2\left(y_{t,(r,s)}^{(j)}, z_{t,r}^{(j)}; \boldsymbol{\theta}_{Y,(r,s)}^{(j)}, \boldsymbol{\theta}_{Z,r}^{(j)}, \phi\right) \right\} \tag{16}$$

where $T_1 = \{1960, 2009\}$ and $T_2 = \{1950, 2099\}$, g and g_2 are the GEV density and the density for the bivariate extreme value distribution (5), and $\boldsymbol{\theta}_{X,(r,s)} = (\alpha_{X,(r,s)}, \beta_{X,(r,s)}, \sigma_{X,(r,s)}, \xi_{X,(r,s)}), \boldsymbol{\theta}_{Y,(r,s)}^{(j)} = (\alpha_{Y,(r,s)}^{(j)}, \beta_{Y,(r,s)}^{(j)}, \sigma_{Y,(r,s)}^{(j)}, \xi_{Y,(r,s)}^{(j)})$ and $\boldsymbol{\theta}_{Z,r}^{(j)} = (\alpha_{Z,r}^{(j)}, \beta_{Z,r}^{(j)}, \sigma_{Z,r}^{(j)}, \xi_{Z,r}^{(j)})$.

Here we use the pseudo likelihood which combines the likelihoods for each individual site under the false working assumption of independence over space, i.e.,

$$L_{\text{false}} = \prod_{r=1}^{5} \prod_{s=1}^{h_r} L_{(r,s)}. \tag{17}$$

To offset this false assumption of spatial independence, we follow the methods developed by Ribatet *et al.* (2012) for handling such a false assumption in a Bayesian context, by making the adjustment to the likelihood of

$$L_{\text{adjusted}} = L_{\text{false}}^{k},$$

where k, $0 < k \leq 1$, is a value to be estimated. If I_{false} and I_{adjusted} denote the observed hessian matrix for L_{false} and L_{adjusted} respectively then

$$I_{\text{adjusted}} = k I_{\text{false}} \tag{18}$$

then variances of the parameters estimated using L_{false} will be k^{-1} times larger when estimated using likelihood L_{adjusted} and consequently the widths of parameter uncertainty intervals from L_{false} will be increased by a factor $k^{-1/2}$. So here k can be interpreted as the reduction factor in the amount of information about the parameters by using L_{adjusted} instead of L_{false}. Then k needs to reflect the loss of information in the data from the presence of spatial dependence in comparison to spatial independence. Thus, careful selection of k is required. Ribatet *et al.* (2012) propose estimating k by exploiting the actual spatial dependence for the data of interest, through setting

$$k = \frac{p}{\sum_{j=1}^{p} \lambda_j}$$

where $(\lambda_1, \ldots, \lambda_p)$ are the eigenvalues of the Godambe information matrix. If the values that contribute to each of the likelihood terms $L_{(r,s)}$ are independent then $k = 1$ and if the sites were perfectly dependent over space then $k = 1/\sum_{r=1}^{5} h_r = 1/439$. For our case though neither such simplification is as straightforward as the data for the GCM in a region r is identical for all sites s in this region. Thus, in practice, we expect $0 < k \ll 1$.

Recall though that we are proposing using Bayesian inference rather than likelihood inference. We therefore have a pseudo-posterior distribution for the parameters of

$$\pi(\boldsymbol{\theta} \mid \text{ data }) \propto L_{\text{adjusted}} \times \pi(\boldsymbol{\theta}) = L_{\text{false}}^{k} \times \pi(\boldsymbol{\theta})$$

where $\pi(\boldsymbol{\theta})$ is the prior. Changing the adjustment factor k leaves the positions modes of the posterior unchanged, but scales the curvature around these modes by k. The impact of this on the inference is that this does not really change in terms of the point estimates but that credibility interval widths are increased by a factor of approximately $k^{-1/2}$.

In summary, in this section we have set out a coherent modelling and inference strategy for getting valid improved efficiency for trend estimates for observations by borrowing information from GCM/RCM data. The problem in implementing this strategy though is its computational complexity. So, in the following section, we illustrate the approach under strongly simplified assumptions which help to overcome the computational burden whilst retaining sufficient features of the strategy that broadly retain its integrity.

5 Illustration of modelling strategy from an over-simplified model

The ideal formulation for the inference, as set out in Sections 3 and 4, is challenging to implement in full. So, to demonstrate the potential benefits of this approach we present results of an analysis which makes strong simplifying assumptions to this ideal formulation. These assumptions will lead to under estimation of the standard deviations for the distribution of trends parameters and hence produce approximate credible intervals that are too narrow to give the nominal coverage. However, in so doing, we illustrate the key steps of the proposed method and show some of its potential benefits. The areas where we make major over-simplifications are:

Spatial penalty adjustment k being fixed Here we take both $k = 1$ and a value of k which depends only on the number of spatial sites, and so we do not evaluate the required adjustment as set out in Section 4. We know in practice k should be much less than 1 and hence using $k = 1$ leads to under estimation of credibility intervals. We also illustrate the analysis with a value of k which we argue is a reasonable approximation, based on intuition, and we explore the differences between the two inferences.

GCM parameters being fixed Here we fix $\boldsymbol{\theta}_{Z,r}^{(j)} = (\alpha_{Z,r}^{(j)}, \beta_{Z,r}^{(j)}, \sigma_{Z,r}^{(j)}, \xi_{Z,r}^{(j)})$ for $r = 1, \ldots, 5$ and $j = 1, \ldots, 11$, thus 220 parameters are treated as fixed in the analysis so their false certainty transmits to under-estimation of uncertainty on the other related parameters. We estimate these 220 parameters using only the GCM data using marginal analysis separately for all r and j. A more complete Bayesian analysis would treat all of these parameters as unknown and the resulting trend estimates would be expected to have wider credible intervals.

Regional instead of UK analysis We undertake the analysis separately for each region, thus instead of using the full pseudo likelihood (17) we use a regional version $L_{\text{false},r} = \prod_{s=1}^{h_r} L_{(r,s)}$.

Thus for the analysis in each region we have 204, 196, 256, 54 and 208 parameters for each of the 5 respective regions.

The trend gradient parameters of the observed data are given in terms of the GCM trends parameters $\{\beta_{Z,r}^{(j)}; r = 1, \ldots, 5, j = 1, \ldots, 11\}$ through expression (11). However, as we have taken these GCM parameters as known, the only source of uncertainty in the estimates of $\beta_{X,(r,s)}$ comes via the unknown linking parameters $(\kappa_{\beta_0}^{(r)}, \kappa_{\beta_1}^{(r)})$ for the region of interest r. Thus only 2 of the 57 parameters that directly determine the observed trend estimates are being appropriately treated as unknown in this illustrative analysis.

Our primary interest is inference for the trend parameter of the observed extreme data and so we focus our discussion on this. We compare three estimates of $\beta_{X,(r,s)}$: the naive maximum likelihood estimator using only observed data from the site itself, and, for two fixed choices of k, our proposed posterior estimator using additional information from the RCM and GCM. We give the results focusing on regions 3 and 5 corresponding to all of Wales and for the part of England south of the north-midlands. We take $k = 1$ corresponding to the false likelihood and $k = h_r^{-1}$ which presumes that there is very strong dependence over the data from the sites in the region and so pooling over sites provides no additional benefit. Thus, this second choice of k is probably too small. For regions 3 and 5 $h_r \approx 100$ and so $k^{-1/2} \approx 10$, and hence when we use the second choice of k we will get credible intervals which are about 10 times wider than if we use the false likelihood ($k = 1$).

Before presenting inference results for $\beta_{X,(r,s)}$, we first examine the estimates of the linking parameters $\kappa_{\beta_1}^{(r)}$ over regions, which gives us information about how the trends of the observed temperature maxima relates to trends in the GCM data (and thus indirectly in the RCM data). Table 1 shows the posterior means and 95% credible intervals of $\kappa_{\beta_1}^{(r)}$. Here we see the benefit for the use of the Bayesian-adjusted analysis over the Bayesian-false method, with the adjustment for spatial dependence giving much wider credible intervals for these parameters. The Bayesian-false inferences give the impression that a different $\kappa_{\beta_1}^{(r)}$ is required for each region, as the credible intervals are non-overlapping under this analysis. Note that the posterior modes for $\kappa_{\beta_1}^{(1)}, \kappa_{\beta_1}^{(2)}$ and $\kappa_{\beta_1}^{(4)}$ are 0.59, 0.88 and 0.88 respectively also appear to support this. However, with the spatial adjustment, it is seen that all credible intervals for $\kappa_{\beta_1}^{(r)}$ will overlap substantially, and thus at least for this linkage parameter we can potentially pool information over regions, though we do not take that approach here on simplicity grounds. Also note that the posterior distributions put the vast majority of their mass in the range $0 < \kappa_{\beta_1}^{(r)} < 1$ for all regions, it shows that the range of trends in the observed data is likely to be less than in the RCM data.

Region	Method	Estimate	95% Uncertainty
3	Bayesian-false	0.49	(0.43, 0.54)
	Bayesian-adjusted	0.49	(-0.17, 1.04)
5	Bayesian-false	0.36	(0.31, 0.41)
	Bayesian-adjusted	0.36	(-0.14, 1.26)

Table 1: The average (and corresponding 95% uncertainty intervals) of estimates for the linkage parameters $\kappa_{\beta_1}^{(r)}$ for GEV trend parameters between the RCM and the GCM (and hence also link observed data with GCM) evaluated using our Bayesian method with the false and adjusted likelihood. In the adjusted likelihood k in region r is taken as $1/h_r$.

Table 2 gives the regional average trend parameter estimate for the observed maxima temperature process from the three inference methods presenting both estimates and associated 95% uncertainty intervals. The naive estimates give a larger average trend estimate in each region than our two Bayesian analyses. This feature suggests that the information from RCM/GCM indicate a lower response rate to changes in annual global average temperature, which is consistent with the exploratory analysis illustrated in Figure 3. However, the key difference is the change in the width of the uncertainty intervals where we can see the potential major benefit from our approach. Firstly, note that the naive estimate gives a 95% confidence interval which shows that the estimates do not significantly differ from 0, and the intervals are very wide. In comparison the false and adjusted likelihoods have credible intervals widths which are reduced by a factor of approximately 400 and 40 respectively relative to the naive interval widths. For both of the values of k that we consider there appears strong evidence of a clear positive trend in extremes with global mean temperatures.

The reason for this level of reduction in uncertainty comes from two factors: our efficient use of the combined information from observed, RCM and GCM data and from our over-simplifying assumptions. Clearly, although we do not expect the reduction in intervals to be as much as 400, as basic knowledge of the data suggests that the false likelihood (when $k = 1$) is failing to account for strong spatial dependence. Taking $k = h_r^{-1}$ over compensates for the spatial dependence and whilst not addressing the other simplifying assumptions that we make it offsets their effects to some degree.

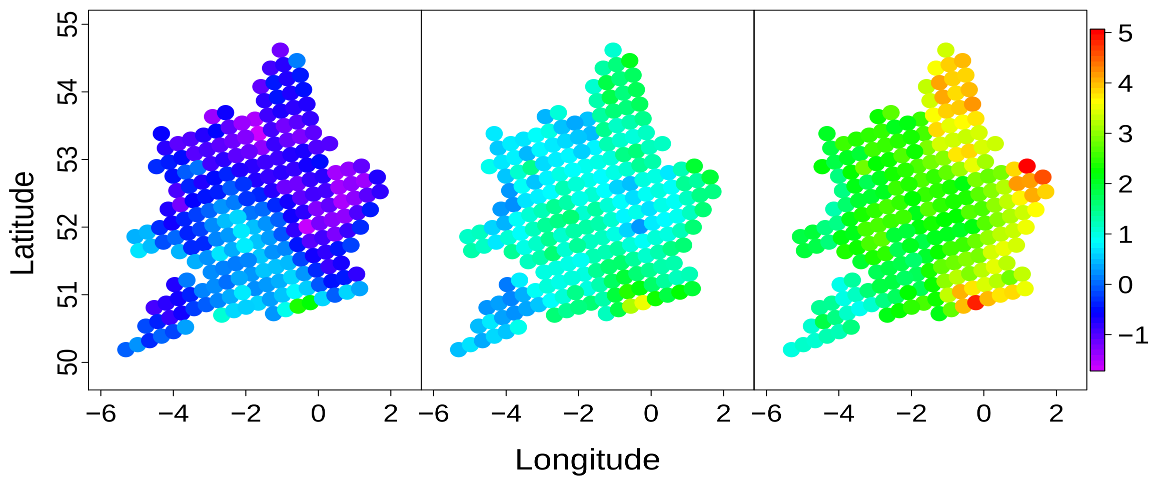
We anticipate that a full analysis without the simplifying assumptions will give estimates and credible regions that are broadly similar to that found here when $k = h_r^{-1}$, i.e., offering a 40 factor reduction in uncertainty relative to the current naive method estimates. To help put this gain of information into context, if we had just used the observed temperature maxima data at a single site then we would have needed a sample of $1,600$ times the current data length (i.e, $80,000$ years) to gain this level of reduction of credible interval width. Of course, to be sure of this, in the future we need to overcome the numerical complexities of the full method and that will enable us to relax these over-simplifying assumptions and rigorously estimate k.

Region	Method	Estimate	95% Uncertainty
3	Naive	1.311	(-0.506, 3.129)
	Bayesian-false	0.802	(0.797, 0.806)
	Bayesian-adjusted	0.802	(0.746, 0.846)
5	Naive	0.868	(-0.237,1.973)
	Bayesian-false	0.816	(0.811, 0.819)
	Bayesian-adjusted	0.816	(0.766, 0.846)

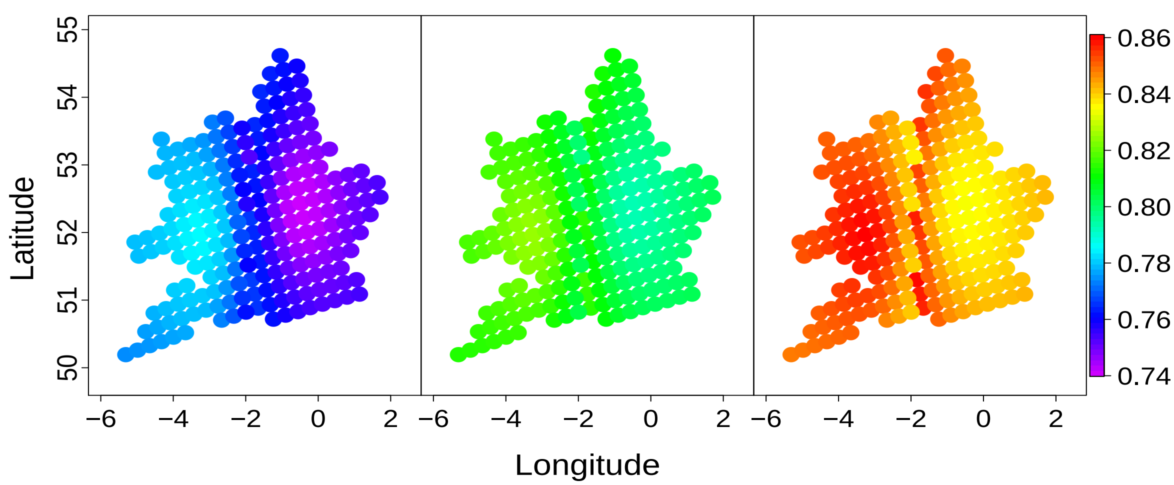
Table 2: The average (and corresponding 95% uncertainty intervals) of estimates of the trend parameter for the observed temperature maxima over each region evaluated using three different methods: naive analysis of observed data only and our Bayesian method with the false and adjusted likelihood. In the adjusted likelihood k in region r is taken as $1/h_r$.

Figure 4 shows the comparisons of these trend parameter estimates and associated uncertainty intervals for the the naive and Bayesian adjusted likelihood methods over these two regions. As already discussed in Section 4, a key feature is the change in uncertainty estimates at each site, whereas here we also see there is a substantial reduction in the spatial variation in the point estimates (a feature not practically affected by our choice of k). From the naive estimates the trends appeared least responsive in the west of the regions (Wales, Cornwall and Devon) and with some spuriously strong positive trends on the south coast, with a 3°C difference in change over these regions for a 1°C change in annual global mean temperature.

Such substantial differences in warming response over relatively small spatial scales are difficult to explain physically. These west-east trend features are reversed in our analysis but with a much smaller variation and a greater spatial coherence to the estimates. There does seem to be a distinctive feature on the Wales-England border in Figure 4 bottom left panel. We believe this feature is an artefact of the grid of the RCM not exactly lining up with the GCM grid, as can be seen in Figure 1. As this artificial feature is seen to be a very small change, once the scale of the plot is accounted for, we note that it does not detract from the main conclusion of our analysis.



(a)



(b)

Figure 4: Maps of the trend estimates for observed temperature maxima over sites in regions 3 and 5: (a) the naive estimator and (b) our Bayesian adjusted method. In both, the middle (left, right) panels correspond to the estimated values and (lower and upper endpoints of 95% uncertainty intervals).

To help with interpretation we focus on these implications for London, corresponding to coordinate (51.5N, 0.3E) in region 3, and for clarity we exclude the uncertainty associated with global mean temperature change. The analysis based on the observed data alone gives that annual maximum daily temperatures in London have increased over 1960-2009 by an estimated 1.22°C, with 95% confidence interval of $(-0.35, 2.79)$°C, whilst global annual mean temperature has increased by 0.88°C. In contrast, our analysis, using all the climate model data as well with the choice of $k = h_r^{-1}$, gives that over this past period the estimated trends has a 95% confidence interval of $(0.68, 0.71)$°C. Furthermore, for a future 2°C increase in global annual mean daily temperature the London annual maximum daily temperature will increase by an estimated 1.59°C with a 95% confidence interval of $(1.54, 1.63)$°C. Thus the inclusion of more evidence has reduced the estimated rate of the response in annual maximum temperatures in London to global mean temperature change and that this estimate now has a level of uncertainty (though subject to caveats due to the residual strong assumptions that we still make) which is of a more helpful magnitude for decision making.

6 Discussion

We have been trying to address the question 'What are the magnitudes and uncertainties of present and future changes in extreme temperatures?' Adaptation pathways, so that society can endure future extreme temperatures, could incur significant cost and therefore it is highly desirable to consider and quantify the uncertainty in projections of future changes in extremes.

This question can be answered through a convolution of the local response to global temperature changes and its uncertainty with the uncertainty in global temperature change at a future date of interest. This paper only deals with the first aspect, looking at the local response sensitivity across climate models. Addressing the question of how the global climate will change is of course the source of extensive independent study, e.g., Knutti et al. (2017), with estimates for the latter part of the century critically depending on different emissions scenarios (Collins et al. 2013).

We have proposed a modelling strategy that utilises the information from climatological model data for the inference of the distribution of observed temperature extremes and their changes through time. The approach here is to take advantage of the additional information from climatological model data with a longer time period to address stochastic uncertainty together with an ensemble of climate model runs to address physical modelling uncertainty. Essentially the analysis is able to efficiently balance the information about the magnitude and uncertainty of the observed trends in the past data with similar information from climate models on past and future changes. Our exploratory analysis has shown which areas of the observed data and climate models can be linked leading to substantial simplification of the statistical modelling. However, implementing such a model remains non-trivial, so to demonstrate the potential advantages of the approach we present an analysis where major assumptions are made. Whilst not being a true representation of reality this analysis shows that considerable reductions in uncertainty can be expected in the estimation of historical and future changes in extreme temperatures relative to using observed data alone. For example, with such simplifications and neglecting any uncertainty in the changes of global temperature, we estimate the annual maximum daily temperatures in London have increased by between 0.68°C and 0.71°C (95% confidence) over the period 1960-2009 in contrast to the naive approach using only observed data which gives a range of -0.35°C to 2.79°C. Furthermore, the high and somewhat unrealistic spatial variability of changes in temperature extremes seen across the UK with the naive approach is greatly reduced resulting in a more physically plausible trend pattern across the UK.

Future work is necessary to overcome the restrictive assumptions we made in Section 5. What is required is to undertake the computationally intensive procedure (simply due to the high dimensionality of the matrix required) described in Section 4 to give a sample based estimate of k, the metric by which likelihoods are adjusted to account for spatial dependence. In addition we have undertaken an analysis with 918 parameters (split over 5 separate regions), so no analysis needed more than 256 parameters to be simultaneously fitted. To address the issues of the GCM parameters being fixed and to expand the analysis to cover the whole UK, we need to extend our fits to having 1138 parameters fitted simultaneously in the Bayesian methods. Conceptually this provides no new problems, but computationally this will be much slower and much more checking is required to ensure that the Markov chain Monte Carlo methods are producing suitably mixing chains to ensure we get convergence of the algorithms. The best way to do this is to trial methods on subsets of the parameters, and this is what we have reported. Additional, complications potentially could arise from strong inter-dependence between the parameters, which may require some blocks of parameters to be jointly updated, rather than to update one by one in turn as our present algorithm does. These issues will only really become apparent when we start to implement the method and monitor convergence.

At the start of Section 4 we decided not to impose smooth spatial structure on the parameters $(\alpha_{X,(r,s)}, \xi_{X,(r,s)})$ in our initial analysis of the data. This resulted in us needing 878 free parameters for this element of the model. Based on the initial analysis it would appear that it is worth exploring now the viability of using smooth estimates of these parameters over space, particularly for the shape parameter. If a simple model form is found to be appropriate for the shape parameter this would substantially reduce the parameter space (reduced by approximately a third). We are less confident in being able to find a sufficiently good smooth model for the location parameters, but once an efficient model is in place for the shape parameters this is worth investigating this aspect further.

We would also like to explore further the simple choice of weighting function (9), to see whether an extension such as

$$w_{r,s} = \frac{d_{r,s}^{-\delta}}{\sum_{\ell=1}^{5} d_{\ell,s}^{-\delta}}.$$

where $\delta > 0$ provides a better fit. We also expect to find that when the GCM trend estimates are not fixed at the marginal estimates then the $\kappa_{\beta_1}^{(r)}$ parameters determining the linkage of RCM to observed data trends will become more spatially coherent, and then it may be possible to see if their regional differences can be removed to produce a more parsimonious model. Both of these extensions though are less important than fully addressing the three areas identified above, that of determining the spatial dependence penalty, fixed GCM parameters and fitting to all UK regions simultaneously.

Acknowledgements

Darmesah Gabda is thankful to the Universiti Malaysia Sabah and Ministry of Higher Education, Malaysia for providing her PhD scholarship. Simon Brown was supported and data provided through the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101).

References

Albert, J. (2007). *Bayesian Computation with R.* Springer.

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

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54, 207-222.

Clark, R. T. and Murphy, J. M. and Brown, S. J. (2010). Do global warming targets limit heatwave risk? *Geophys. Res. Lett.*, 37, L17703, doi:10.1029/2010GL043898.

Coelho, C. A. S., Ferro, C. A. T., Stephenson, D. B. and Steinskog, D. J. (2008). Methods for exploring spatial and temporal variability of extreme events in climate data. *Journal of Climate*, 21, 2072-2092.

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

Collins, M., Booth, B. B. B., Bhaskaran, B., Harris, G., Murphy, J. M., Sexton, D. M. H., and Webb, M. J. (2011). Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles, *Climate Dynamics*, 36, 1737-1766.

Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver and M. Wehner, (2013). Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*

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

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R. and Mearns, L. O. (2000). Climate extremes: Observations, modeling and impacts. *Atmospheric Science*, 289, 2068-2074.

Gabda, D. (2014). *Efficient Inference for Nonstationary and Spatial Extreme Value Problems.* Lancaster University, PhD thesis.

Gabda, D. and Tawn, J. A. (2018). Univariate extreme value inference from small sample sizes in environmental contexts. Submitted to *Extremes.*

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference.* Second edition. Texts in Statistical Science Series. Chapman and Hall/CRC.

Hanel, M. and Buishand, T. A. (2011). Analysis of precipitation extremes in an ensemble of transient regional climate model simulations for Rhine basin. *Climate Dynamics*, 36, 1135-1153.

Hanel, M., Buishand, T. A. and Ferro, C. A. T (2009). A nonstationary index flood model for precipitation extremes in transient regional climate model simulations, *Journal of Geophysical Research*, 114, 1-16.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97-109.

Heffernan, J. E. and Tawn, J. A. (2001). Extreme value analysis of a large designed experiment: a case study in bulk carrier safety. *Extremes*, 4, 359-378.

Hoff, P.D. (2009). *A First Course in Bayesian Statistical Methods.* Springer.

Jonathan, P., Randell, D., Wu, Y. and Ewans, K. (2014). Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *Ocean Engineering*, 88, 520?532.

Katz, R.W. (2002). Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research*, 20, 167-185.

Knutti, Reto and Rugenstein, Maria A. A. and Hegerl, Gabriele C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, 10, 727+.

Kyselý, J. (2002). Comparison of extremes in GCM-simulated, downscaled and observed central-European temperature series. *Climate Research*, 20, 211-222.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and related Properties of Random Sequences and Processes*, Springer, Berlin.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21, 1087-1091.

Murphy, J. M. and Sexton, D. M. H. and Jenkins, G. J. and Booth, B. B. B. and Brown, C. C. and Clark, R. T. and Collins, M. and Harris, G. R. and Kendon, E. J. and Betts, R. A. and Brown, S. J. and Humphrey, K. A. and McCarthy, M. P. and McDonald, R. E. and Stephens, A. and Wallace, C. and Warren, R. and Wilby, R. and Wood, R. A. (2009). UK climate projections science report: Climate change projections. *Met Office Hadley Centre, Exeter, UK*

Nakićenović, N. and Alcamo, J. and Davis, G. and de Vries, B. and Fenhann, J. and Gaffin, S. and Gregory, K. and Grübler, A. and Jung, T. Y. and Kram, T. and La Rovere, E. L. and Michaelis, L. and Mori, S. and Morita, T. and Pepper, W. and Pitcher, H. and Price, L. and Riahi, K. and Roehrl, A. and Rogner, H.-H. and Sankovski, A. and Schlesinger, M. and Shukla, P. and Smith, S. and Swart, R. and van Rooijen, S. and Nadejda, V. and Zhou Dadi. (2000). Emission scenarios. A special report of Working Group III of the Intergovernmental Panel on Climate Change. *Cambridge University Press*

Nikulin, G., Kjellström, E., Hansson, U., Strandberg, G. and Ullerstig, A. (2011). Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations. *Tellus*, 63A, 41-55.

Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights (with discussion), *Environmetrics*, 22(7), 799-816.

Ribatet, M., Cooley, D., and Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22, 813-845.

Roberts, G. O., Rosenthal, J. S., (2001). Optimal scaling for various Metropolis-Hastings algorithms. Statistical Science, 16(4):351-367.

Stott, P. A. and Forest, C. E. (2007). Ensemble climate predictions using climate models and observational constraints, *Philosophical Transactions of The Royal Society A*, 365, 2029-2052.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75, 397–415.

Wuebbles, D., Meehl, G., Hayhoe, K., Karl, T., Kunkel, K., Santer, B., Wehner, M., Colle, B., Fischer, E., Fu, R., Goodman, A., Janssen, E., Kharin, V., Lee, H., Li, W., Long, L., Olsen, S., Pan, Z., Seth, A., Sheffield, J., and Sun, L. (2014). CMIP5 climate model analyses: climate extremes in the United States, *Bulletin of the American Meteorological Society*, 95 (4), 571-583.