UKCP Additional Land Products:
Probabilistic Projections of Climate Extremes
September 2020

Update to Probabilistic Projections

Global Projections

Regional Projections

UKCP18

Marine Projections

Observations

Derived Projections

J.M. Murphy, S.J. Brown and G.R. Harris

Table of Contents

Executive Summary

- This report describes 21ˢᵗ century projections of 20-, 50- and 100-year return levels of daily maximum surface air temperature (TXx), daily precipitation (Rx1day) and 5-day accumulated precipitation (Rx5day) at 25km resolution.

- This new information on climate extremes forms an extension to the UKCP Probabilistic component of the land projections. Results are available for five emissions scenarios (RCP2.6, 4.5, 6.0 and 8.5, and SRES A1B).

- Under RCP8.5 emissions, the projections show future increases in median return levels in all seasons for all variables. By 2070, the increase in median TXx relative to 1990 is 2.5°C in winter, and 3.7-4.3°C in other seasons, based on UK averages of regional values. The smallest median increases occur in summer for Rx1day and Rx5day, with larger increases (5-10mm for Rx1day and 9-13mm for Rx5day) occurring in autumn, winter and spring.

- The results provide broad assessments of known modelling uncertainties, based on 360 climate model simulations combined with observational constraints in the same statistical framework used for other UKCP Probabilistic variables.

- These uncertainties grow during the 21ˢᵗ century. By 2070, for 50-year return levels under RCP8.5, UK averages of the 10th and 90th percentiles are: 32.0°C and 39.9°C respectively for TXx in summer; 40mm and 58mm for Rx1day in winter; 102mm and 142 mm for Rx5day in autumn.

- Data are presented as absolute future values in °C or mm, incorporating a bias correction that ensures consistency with observed return levels for the baseline period of 1981-2000. The results are suitable for analysis of extremes at specific 25km grid squares, but not for analysis of joint risks at spatially distributed locations.

- Users interested in spatially distributed analysis, or other types of extreme event, can obtain suitable climate model data from other UKCP land products. These consist of the sets of 28 global projections (UKCP Global), 12 European regional projections (UKCP Regional) and 12 UK convective-permitting projections (UKCP Local).

- However, these products include more limited representations of uncertainties, in comparison to UKCP Probabilistic. Therefore, the results from this report can provide useful context for studies based directly on model simulations, by revealing potential gaps in sets of potential climate impacts diagnosed from the latter.

1.   Introduction

The latest generation of UK Climate Projections (UKCP) for the 21st century was produced recently (Lowe et al., 2018). The land component (Murphy et al., 2018; Kendon et al., 2019) includes updated probabilistic projections (UKCP Probabilistic, presented at 25km spatial resolution). There is also a new set of global projections (UKCP Global @60km resolution) to support worldwide study of regional climate impacts and their driving physical processes, and two new sets of regional projections facilitating detailed study of impacts over Europe (UKCP Regional @12km, Murphy et al., 2018), and the UK (UKCP Local @2.2km, Kendon et al., 2019).

In the latter case, use of a "convective-permitting" model (with explicit simulation of the dynamics of large convective storms) allowed UKCP, for the first time, to provide projections of extreme precipitation events at the hourly time scale. The other UKCP products are all based on global and regional simulations in which convection is parameterised, hence precluding provision of realistic sub-daily information for precipitation. However, such modelling systems are capable of providing useful advice on future daily surface temperature and precipitation extremes (e.g. Kharin and Zwiers, 2000). Projections of longer-term events, such as multi-day heatwaves or heavy precipitation accumulations (e.g. Clark et al., 2010; Rajczak and Schär, 2017), cold, hot, dry or wet seasonal extremes (e.g. Sexton and Harris, 2015) or multi-season droughts (e.g. Burke and Brown, 2010), can also be obtained.

In UKCP, users can obtain projections of future extremes for the UK or other regions from UK Global (a set of 28 climate model simulations), UKCP Regional (12 simulations) or UKCP Local (12 simulations) These datasets offer flexibility because they consist of raw climate model output (see https://catalogue.ceda.ac.uk/?q=ukcp18&sort_by=), and can therefore be used to derive a wide variety of metrics for analysis of multiple hazards or distributed sectoral risks, as was done with earlier outputs from UKCP09 (e.g. McColl et al., 2012; Palin et al., 2013).

However, the above products provide limited representations of uncertainties compared with the probabilistic projections, which generally show broader ranges of plausible future changes (Murphy et al., 2018; Kendon et al., 2019). This is mainly because the probabilistic projections were derived from a larger set of 360 climate model simulations. These include results from perturbed parameter and multi-model ensembles, and account for uncertainties arising from physical climate system processes, aerosol chemistry and carbon cycle feedbacks at global and regional scales. The probabilistic projections are available for five alternative scenarios of 21st century greenhouse gas emissions (RCP2.6, 4.5, 6.0 and 8.5 (Moss et al., 2010), and SRES A1B (Nakicenovic and Swart, 2000)). The UKCP Global, Regional and Local projections are only available for RCP8.5, although the RCP8.5 scenario results can be used to look at climate change for a wide range of global warming levels.

Currently, the probabilistic projections provide information for monthly, seasonal and annual averages of a set of basic UK climate variables (see https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-guidance-data-availability-access-and-formats.pdf). This report describes extensions to the methodology that facilitate addition of information on selected sub-monthly extremes. These consist of 20-, 50- and 100-year return levels of daily maximum surface air temperature (hereafter TXx), daily precipitation (Rx1day) and 5-day accumulated precipitation (Rx5day).

For example, Brown et al. (2014) estimated a 20-year return level of 35.7°C for summer TXx in London, for 1961-90. Such rare daily events can occur either during short spells of hot weather associated with particular synoptic conditions (e.g. during 2018, McCarthy et al., 2019), or during an extended heatwave such as that of summer 2003 (Pirard et al., 2005). Projections of TXx are therefore useful for assessment of future risks and impacts associated with overheating. This was recognised in the second UK Climate Change Risk Assessment (CCRA2, 2017) as a complex issue affecting many sectors of the economy, including health, transport, energy and the natural and built environments.

Similarly, CCRA2 highlights future risks of surface and river flooding arising from increases in heavy precipitation events, and the importance of providing suitable sustainable drainage and natural flood management measures in response. The Rx1day metric covers contributions to flood risk from daily accumulations that might arise (say) from convective events or the transit of a mobile frontal system, while Rx5day covers sustained events that might be associated with persistent flow conditions or the passage of a large synoptic system. For London, Brown et al. (2014) estimate historical 20-year return periods of 42.1mm for Rx1day in summer, and 78.4mm for Rx5day in autumn.

In common with several previous UK studies (e.g. Brown et al. (2008); Fowler and Ekström, 2009; Burke et al., 2010; Chan et al., 2014) we use extreme value theory (EV), based on theoretical probability distributions that describe extremal properties using a few key parameters. This allows plausible estimates of long-period return levels to be diagnosed from observational or model time series that may not contain specific examples of all the rare events of interest.

The method used in this report is an updated version of Brown et al. (2014), in which EV parameters are estimated by fitting to time series of extreme values from climate model simulations, using an assumed linear dependence on global mean surface temperature (GMST) to represent effects of climate change. When making projections, time-dependent values for EV parameters are created by adding climate change components derived from model simulations to baseline values representative of observations. This approach constitutes a form of bias correction, that allows the final return level projections to be presented as absolute rather than anomalous values. In this respect, the return levels product differs from the monthly, seasonal and annual average data released previously (Murphy et al., 2018). In the latter the UKCP Probabilistic variables were presented as anomalies relative to 1981-2000, leaving consideration of bias correction options to users on an application-specific basis (Fung, 2018).

In section 2 we describe our methodology, explaining how the Bayesian statistical framework underlying the probabilistic projections (Murphy et al., 2018; Harris et al., 2021) is combined with EV theory to support the new products. A selection of illustrative results is shown in section 3, followed by concluding remarks in section 4, which include discussion of the strengths and limitations of the results.

2.   Methodology

The Bayesian method for the probabilistic projections is described in Murphy et al. (2018), while the use of EV theory is based on Brown et al. (2014), with updates. Here, we outline the main elements (Fig. 1) in sections 2.1 and 2.2 respectively, followed by a description in section 2.3 of how the two methods are combined to provide projections of extremes.
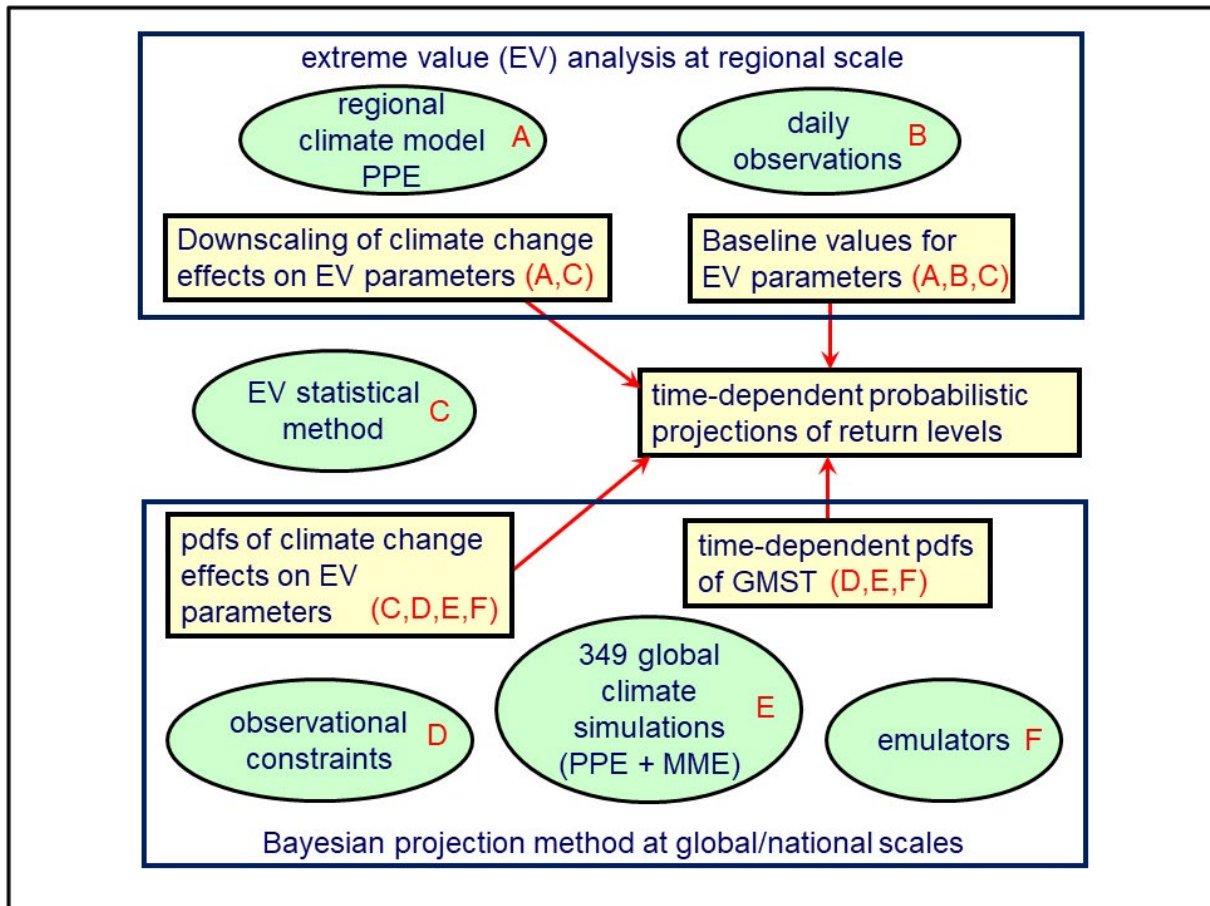

Figure 1. Major elements of the methods for Bayesian projections based on global climate model simulations (lower box), and statistical analysis of regional extreme events (upper box), underpinning the UKCP Probabilistic projections of long-period return levels. Green elements A-F show methodological components, and yellow elements show various outputs (dependent on the bracketed subsets of components), that are combined to produce the final projections.

2.1. Bayesian projections

For a given emissions scenario, the probabilistic projections provide distributions of future climate changes consistent with internal climate variability, and uncertainties in the representation of earth system processes. The results are Bayesian probabilities that represent the relative strength of evidence behind alternative future changes. These are conditional upon the evidence (observations and model simulations) used to derive them, and subjective judgements required to synthesise the evidence in the calculations.

The bullets below provide an overview of major components of the Bayesian method, followed by a summary of the underlying climate model simulations and their main roles in the calculations. Specific details of the implementation for projections of extremes are provided in section 2.3. A fuller outline of the UKCP Probabilistic methodology is available in Murphy et al. (2018). A detailed description will follow in Harris et al. (2021), building on the original implementation of Sexton et al. (2012) and Harris et al. (2013).

- **Definition of a "prior" parameter space.** This space represents uncertainties in earth system processes in a single climate model (HadCM3). Its definition is reliant on judgements from model parameterisation experts.

- **Perturbed parameter ensemble (PPE) simulations.** These explore how historical simulation skill and projected future responses vary across a set of locations in parameter space.

- **Climate model emulation techniques.** These are trained on the PPE results, to provide estimates of past and future climate at any point in the prior parameter space.

- **Estimation of the "structural" component of model error.** This represents the effects of systematic errors that cannot be resolved by varying parameters in HadCM3. Such errors can arise from the effects of processes missing from the model, or from shortcomings in the basic assumptions used in its dynamical integration scheme or physical parameterisations of sub-grid scale phenomena. We quantify this term using results from an independent multi-model ensemble, assuming that differences between outputs of PPE variants and other climate models can be treated as a proxy for structural simulation errors relative to the real world. By construction, this neglects the impact of systematic errors that are common to all climate models. Such errors represent an important but inevitable caveat, that arise from limitations in current modelling capability or incomplete knowledge. For example, explicit representation of atmospheric convective-scale dynamics (Kendon et al., 2019) has recently become possible in some limited area simulations, such as UKCP Local. However, it is not yet feasible in ensembles of long global simulations. In addition, the earth system models used in the probabilistic projections do not yet capture the full complexity of atmosphere-ocean-cryosphere interactions, lacking explicit representations of ice sheets, permafrost or the ocean methane cycle. Once included in future generations of models, there is potential for such interactions to modify the projected changes presented here.

- **Observational constraints.** These are derived from seasonal climatological spatial fields for a set of twelve variables commonly used to evaluate climate models, plus several metrics of historical climate change during the 20th century. The latter include spatial patterns of surface temperature change, and global changes in upper ocean heat content and atmospheric CO_2 concentration.

- **Production of probabilistic projections.** This is done by integrating over parameter space, adding in the estimated structural component of model error in past and future climate, and weighting according to estimated model quality. The weights express a relative likelihood for each point in parameter space.

- **Accounting for uncertainties in both physical and carbon cycle feedbacks on the response to a given emissions scenario.** For each scenario, the projections explore the effects of a range of potential future pathways for CO_2 concentration, as well as a range of climate sensitivities.

Climate model simulations and their roles in the Bayesian method

The Bayesian calculations described by Murphy et al. (2018) involved core components applied to global climate model (GCM) data. This data included GMST and variables for five HadCM3 land boxes covering the UK at a spatial scale of ~300km (see Fig. 3 of Harris et al. (2010), also Fig. 11 in this report). Hereafter, we refer to these grid boxes as UK_GCM. A second component provided downscaling to a 25km national grid, and three sets of aggregated regions (Fung et al., 2018). This was carried out using an 11-member PPE of the regional climate model configuration of HadCM3 (see Murphy et al., 2009), hereafter HadRM3-PPE. In the present application the same ensemble is used for downscaling, but the methodology is adapted for use with extreme value distribution parameters and carried out as part of the EV calculations (see sections 2.2 and 2.3).

Here, therefore, the implementation of the Bayesian framework (Fig. 2) follows that of Murphy et al (2018) at global and national scales, and produces outputs which are then combined with EV outputs (created on the aforementioned 25km grid) to produce the final projections of 21st century return levels (Fig. 1).

The GCM-based calculations use three ensembles of simulations. The first is a 280-member PPE using the coupled atmosphere-mixed-layer-ocean ("slab") configuration of HadCM3 (Box 1 of Fig. 2). This PPE (hereafter SLAB) simulates historical climatology and the equilibrium response to doubled CO_2, exploring the effects of uncertainties in 30 surface and atmospheric parameters controlling a range of physical processes. It also provides a basis for the use of emulation techniques (Box 2) to sample points in this parameter space for which no GCM simulation is available. However, the SLAB model configuration lacks a dynamical ocean component, and does not provide information on transient climate change.

Two further ensembles are used to cover these aspects. The first (Box 3) is a PPE of 57 variants of the earth system configuration of HadCM3, including dynamical ocean and vegetation modules with an interactive carbon cycle (Lambert et al., 2013; Murphy et al., 2014). Members of this ensemble (hereafter the ESPPE), are distinguished by multiple simultaneous perturbations to parameters in the atmosphere, ocean, sulphur cycle and terrestrial ecosystem components. The simulations were driven using historical and future emissions of CO_2 and aerosol precursors, and concentrations of other major greenhouse gases. Natural historical forcing, due to variations in solar irradiance and major volcanic eruptions, was also included. ESPPE simulations are available for the RCP2.6, RCP8.5 and SRES A1B scenarios.

ESPPE results are used to convert projected equilibrium changes from SLAB into estimates of time-dependent global and regional climate changes, and to add sampling of parametric uncertainties in ocean transport, atmospheric sulphur cycle and terrestrial carbon cycle processes. This broadens the range of earth system processes considered in prior distributions of projected changes (Box 5 in Fig. 2), increasing the dimensionality of the sampled parameter space to 54.
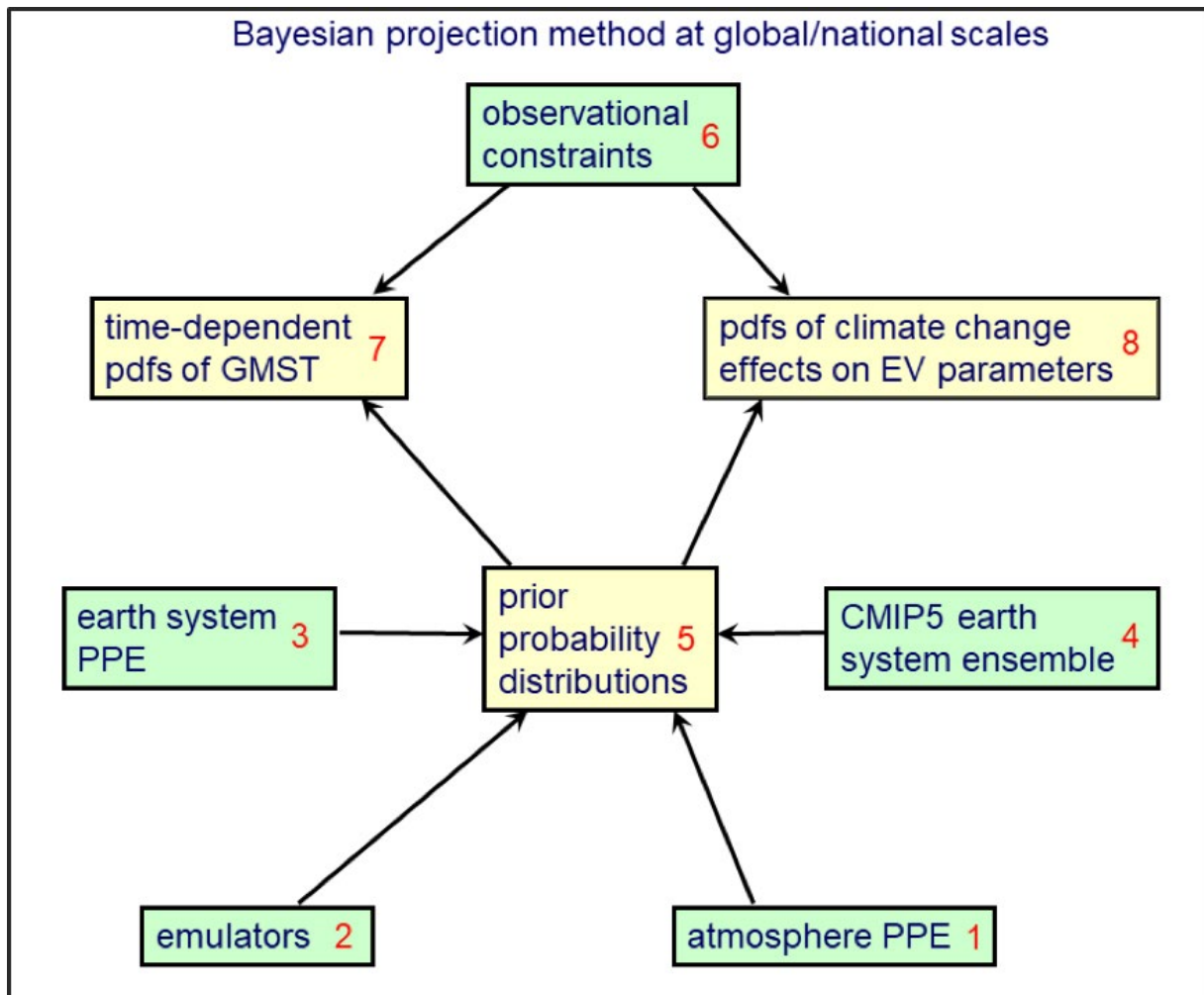
Figure 2. Linkages between components of the Bayesian method underpinning the probabilistic projections based on global climate model output (providing more detail on the lower box of Fig. 1). Green boxes show the main inputs, consisting of GCM simulations, statistical emulators trained on their results, and observational constraints. Yellow boxes show the prior (5) and posterior (7 and 8) probability distributions produced by the calculations.

The final ensemble (Box 4) consists of a set of transient simulations using the RCP8.5 scenario, from twelve CMIP5 earth system models (hereafter CMIP5-ESM - see Table 2.1 in Murphy et al., 2018). Like the ESPPE, these simulations were driven by historical and future CO_2 emissions, providing ranges of future change influenced by both physical and carbon cycle feedbacks. This ensemble is used to account for structural uncertainties (see above), which are quantified in the Bayesian framework using a term called "discrepancy" (Goldstein and Rougier 2004; Sexton et al., 2012).

Discrepancy is specified as a multivariate Gaussian distribution that broadens the spread of projected changes (because it adds to the effects of parametric uncertainties). It can also shift the envelope of projections if the discrepancy distribution shows a non-zero median. The structural uncertainty calculations involve using emulators to search the HadCM3 parameter space for the closest multivariate analogues of the historical and future simulations of each CMIP5-ESM member. Differences between the best analogues and each CMIP5-ESM simulation are used to calibrate the discrepancy distribution.

2.2. Statistical characterisation of extreme events

Extreme value theory (e.g. Coles, 2001) is used in climate research to derive probability distributions that describe the expected occurrence of rare events (e.g. IPCC, 2012). Such events lie in the far upper or lower tails of the full distribution of values for the relevant variable, for example daily temperature maxima that occur less than once or twice per season. Using EV distributions compensates for sampling issues, by supporting estimates of probability for events that may not be present in the relatively short periods of data used to calibrate the distributions. A key assumption is that the chosen EV distribution provides an accurate representation of the complete probability distribution of the relevant events.

Alternative approaches are available to estimate parameters of EV distributions. The block maximum method involves using the most extreme value within each of a set of distinct blocks of data. Here, we provide projections for the meteorological seasons (December to February, March to May, June to August and September to November). For summer, for example, this would involve picking the highest value of TXx, Rx1day or Rx5day from each summer over a set of consecutive years. This method has been used in conjunction with the generalised extreme value (GEV) distribution in several climate change papers, including the Brown et al. (2014) study that forms the basis of our approach here. It has the advantage of avoiding the need to select a threshold to define an extreme event, but restricts the size of the data sample to one event per season.

In the present application, therefore, we use an alternative method: peaks over threshold (PoT). This involves using all events exceeding a specified threshold in a given season, thus considering more of the data, and avoiding the risk of missing multiple extremes that may occur in close proximity. The PoT approach also excludes any seasons which happen not to contain any extreme events. The results depend on the choice of threshold, which is a subjective decision. We select thresholds designed to capture 3% of events per season, thus seeking a balance between increasing sample size and capturing genuine extremes. Following Brown et al. (2008), the number of events is assumed to follow a Poisson process, and the magnitude of exceedances a generalised Pareto distribution (e.g. Katz et al., 2002).

The expected number of exceedances above a level x, conditional on x exceeding the chosen threshold, is parameterised as:

$$1 + \xi \left[\frac{x - \mu}{\sigma} \right]^{-1/\xi} \qquad (1)$$

where μ, σ and ξ are the location, scale and shape parameters (e.g. Coles, 2001). With this choice of formulation, the above EV parameters are equivalent to those used in the GEV distribution for block maxima. We therefore use the same notation as Brown et al (2014), but using threshold exceedances rather than block maxima to calibrate the parameters.

The EV distribution function is:

$$
G(x) = \begin{cases} \exp\left\{ -\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi} \right\}, & \xi \neq 0 \\[2em] \exp\left\{ -\exp\left[-\left(\frac{x - \mu}{\sigma}\right)\right] \right\}, & \xi = 0 \end{cases}
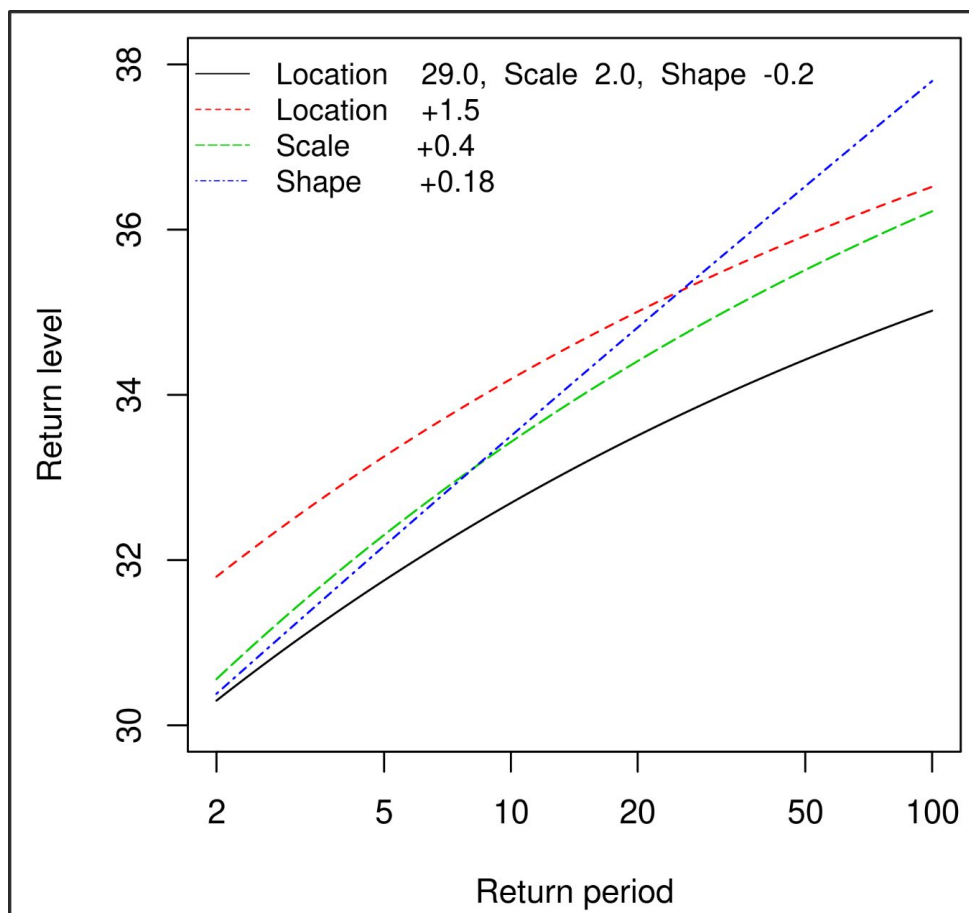$$

(2)

Figure 3. Effects of varying EV parameters on a return level curve. The black curve shows how return level would vary with return period for location, scale and shape values of 29.0, 2.0 and -0.2. The red, green and blue curves show the impacts of increasing location, scale and shape.

Figure 3 demonstrates how changing each parameter influences the variation of return level with return period. The black line is a return level curve assuming μ = 29, σ = 2 and ξ = -0.18. These values are typical for TXx distributions in summer, though the qualitative impacts of changing the parameters apply in general. Increasing μ (red curve) results in a uniform upward shift to higher return levels, analogous to increasing the mean of a Gaussian distribution. Increasing σ (green curve) stretches and flattens the distribution (akin to increasing the standard deviation of a Gaussian), increasing the difference in return level between short and long return periods. Increasing ξ (blue curve) makes the distribution more heavy-tailed, thus increasing the levels of the longer periods.

Calibration of EV parameters for observations and RCM simulations

We use observations of TXx, Rx1day and Rx5day from 1961-2019, pooled with the HadRM3-PPE data of section 2.1, which is available for 1951-2100. The combined model and observed data are used to calibrate a common statistical model for historical and future values of location, scale and shape. The observed data is provided on a 5km grid, derived from high-density networks of station data (Perry et al., 2009). We use versions of the observational datasets regridded to the 25km grid of HadRM3, representing the time-dependent EV parameters at each grid box as:

$$\mu(t) = \mu_0 + \mu_\Delta I + \mu_T GMST(t)$$

$$\sigma(t) = \beta_0 + \beta_\Delta I + \beta_T GMST(t) \tag{3}$$

$$\xi = \xi_0 + \xi_\Delta I$$

Effects of climate change are represented through an assumed linear dependence on GMST, scaled by parameters μ_T and β_T; μ_0, β_0 and ξ_0 represent observed baseline values; μ_Δ, β_Δ and ξ_Δ characterise differences in location, scale and shape between observations (for which the indicator covariate I is set to zero) and model data (I=1). Brown et al. (2014) found that robust climate change signals could not be identified for the shape parameter, which is therefore assumed independent of GMST in Eq (3).

Eq (3) follows Brown et al. (2014), except that a logarithmic transformation is no longer used in the parameterisation of σ. Under this transform, we found that sampling of residual uncertainties in fitted values of β_T, which are assumed Gaussian in our methods for emulation of GCM values (section 2.3.2) and regional downscaling (section 2.3.4), could sometimes lead to unrealistically large future values of σ. This is because application of the inverse (exponential) transform converts Gaussian residuals in β_T into positively skewed residuals in σ. Using a linear relationship in Eq (3) avoids this issue. However, unrealistically large future return levels can still occur, when high values of β_T are combined with large future increases in GMST. A lower bound on σ is also required, in order to avoid negative values, or small positive values implying unrealistic EV distributions in which return level fails to increase with return period (cf Fig. 3). When generating projections, we therefore reject any points in HadCM3 parameter space that lead to future σ values outside the range $0.5\beta_0$ - $2\beta_0$. Section 2.3.5 provides details on the implementation of this condition.

The estimation of parameters in Eq (3) involves a separate fit to each of the 11 members of HadRM3-PPE. Values of GMST(t) for 1951-2100 are obtained from the 11 GCM simulations that drove HadRM3-PPE members. These were transient climate change experiments run from 1900-2100, using the SRES A1B emissions scenario for the future component. Each driving simulation was a member of a PPE of the coupled atmosphere-ocean configuration of HadCM3 (Collins et al., 2011), that used a set of parameter perturbations corresponding to the relevant HadRM3-PPE member. Time series of GMST(t) are smoothed[1] prior to fitting the EV parameters, in order to remove the influence of internal variability and isolate the long-term effects of climate change that μ_T and β_T are intended to represent.

---

[1] A spline is fitted to the data, which removes variability on time scales shorter than approximately 30 years.

For each 25km grid square, the fit[2] is performed by pooling data from the relevant HadRM3-PPE simulation with the observed data, using the HadCRUT4.6.0.0 dataset (Morice et al., 2012, updated) to provide GMST(t) data in the case of the observations. In all cases, time-dependent GMST anomalies are defined relative to the standard UKCP baseline of 1981-2000. The values of μ_0, β_0 and ξ_0 that emerge from the fit are therefore representative of 1981-2000 conditions, even though the full period of observed data is used in their calibration, with account taken of non-stationarity through the observed GMST changes. The same baseline period is used in section 2.3, where Eq (3) is deployed to construct projections.

When I=1 for model data (two exceptions are discussed below), separate observational and bias estimates are derived from each of the eleven pooled fits. The eleven values of μ_0, β_0 and ξ_0 are found to be similar (as in Brown et al., 2014), demonstrating that the calibration successfully identifies RCM biases. We take the average values of μ_0, β_0 and ξ_0 as our best estimates of the baseline observations. When using μ_0, β_0 and ξ_0 to produce the final probabilistic projections (section 2.3.5), we sample uncertainties via bootstrapping[3].

In the case of ξ_0 for Rx1day and Rx5day we set I to zero for the regional model data as well as observations, since tests show that statistically significant differences between regional model and observed shape values cannot be identified from the limited time series available. This is because fitted values of ξ_0 and ξ_Δ are particularly sensitive to individual rare precipitation events (Brown et al., 2014). The regional model and observed data are therefore treated as samples from populations with common shape values.

We also smooth the spatial distributions of ξ_0, by defining each 25km value as the median over a set of 5x5 grid boxes including and surrounding the target location. This is done for TXx, as well as Rx1day and Rx5day. Equivalent smoothing is not performed for μ_0 or β_0.

Figure 4 shows examples of the spatial patterns found for μ_0, β_0 and ξ_0, for TXx and Rx1day in summer and Rx5day in autumn. For TXx, values of μ_0 are lowest in northern Scotland and highest in central, eastern and southern England, exceeding 30°C in places. Not surprisingly, these broad-scale variations resemble those found in the observed pattern of climatological average summer temperature, and other metrics of extremes such as the typical hottest day of summer (e.g. Figs. 4.1b and 4.5a of Murphy et al., 2018). The highest values of β_0 occur mostly over south-east England. Relatively high values of β_0 are also found over parts of East Anglia, central southern England, Wales and the Hebrides.

For Rx1day in summer and Rx5day in autumn, the patterns of μ_0 show maxima over high ground in western regions, as do corresponding climatological precipitation patterns (e.g. Fig. 4.1d of Murphy et al., 2018). Values of β_0 show maxima over the western Highlands for both variables, and the spatial distribution for Rx5day shows lower values over East Anglia and parts of central and southern England, than elsewhere. This may (speculatively) be related to stronger influence, to the north and west, of variability associated with the passage of large synoptic cyclones associated with the North Atlantic storm track.

For summer TXx, ξ_0 is negative everywhere. This indicates bounded EV distributions where the rate of increase in magnitude reduces with rarity. Negative ξ_0 values are also prevalent for Rx5day in autumn, although values are close to zero in some regions. These indicate light-tailed distributions in which the probability of the highest extremes is bounded. For summer Rx1day, most regions show small positive values of ξ_0. These correspond to EV distributions with slightly heavy tails that are unbounded.

---

[2] The EV parameters are estimated by searching for values that maximise the likelihood of the data, using the ISMEV extremes analysis software of Coles (2001). Further details are available in Brown et al. (2014).

[3] A random sample is drawn from the EV distribution defined by the baseline parameters, of an equivalent size to the real observational dataset, and the EV model is refitted to these pseudo-observations (with no RCM data or global temperature terms). This is repeated many times and the spread in parameters is taken to be representative of the uncertainty in the observed data. See Brown et al. (2008) for details.
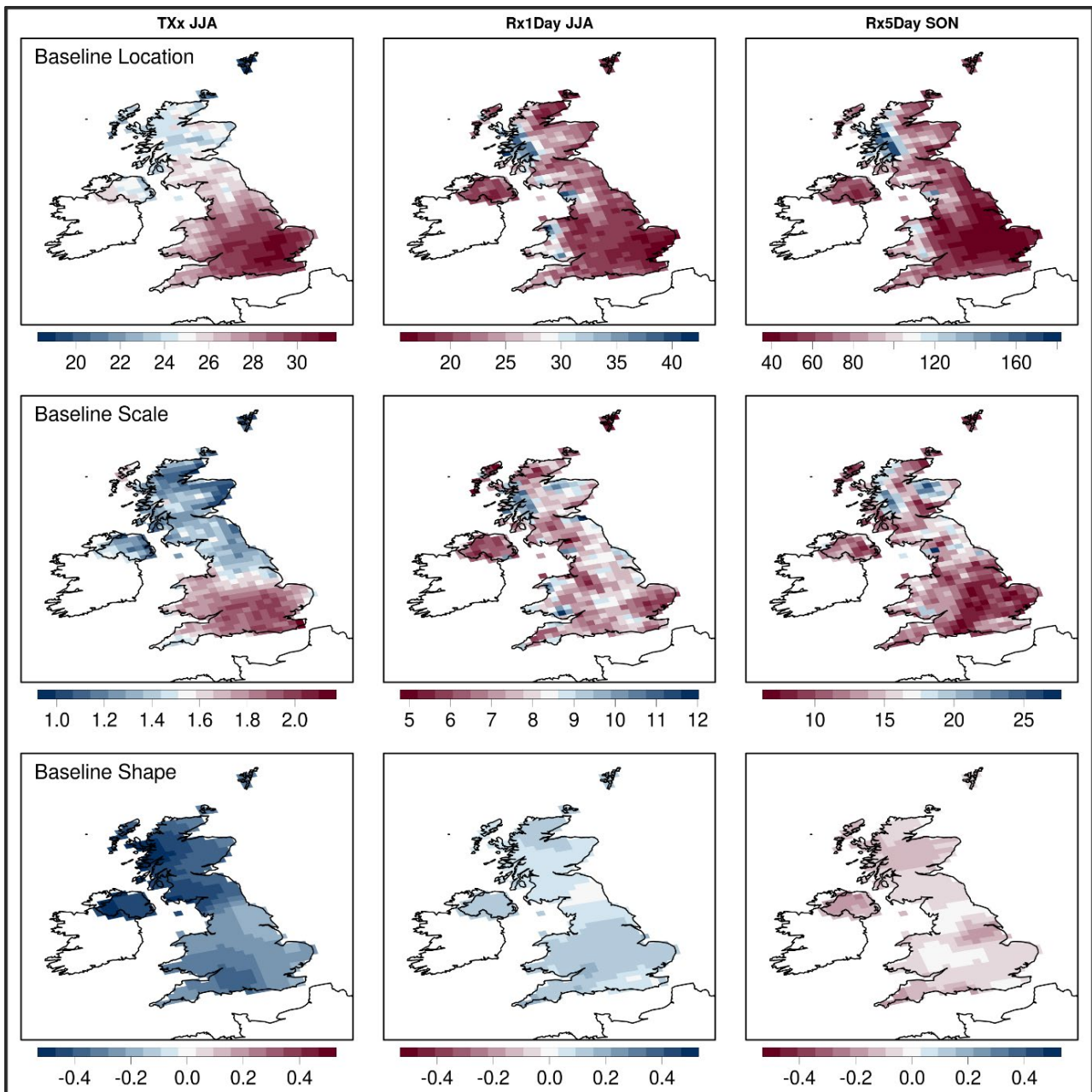
Figure 4. Estimates of μ_0, β_0 and ξ_0, representing observed values for 1981-2000 of location, scale and shape (see Eq (3)), for TXx and Rx1day in summer, and Rx5day in autumn. The values are calibrated from observed data covering 1961-2019 accounting for non-stationarity associated with historical changes in GMST. The results shown are averages of eleven values of μ_0, β_0 and ξ_0 obtained by pooling observed data with output from each member of the HadRM3-PPE of regional climate model simulations. This increases the sample size of events used in calibration, while allowing for model biases by using the indicator variable in Eq (3) to distinguish modelled from observed data. Values are obtained for each 25km grid box of HadRM3. For ξ_0, results are smoothed by taking the median over 25 grid boxes including and surrounding the target location.

Figure 5 shows examples of fitted return level curves for the variables of Fig. 4, for the 25km grid box containing central London. Open circles show the highest 2% of the distribution of values for the relevant season in the 59-year observed record, which amounts to 108 data points. Empirical estimates of corresponding return periods are diagnosed from a simple ranking of the events. This is done by assuming that the mth highest event corresponds to a return period of 59/m years. The EV curves match the empirical estimates quite well, particularly for return periods of 1-10 years. There is more deviation for longer return periods: For Rx5day the empirical values for the top eight events are all higher than the EV estimates, as are the highest events for TXx and Rx1day. This is not surprising, as the empirical estimates are known to be biased high for the rarest events, while giving reasonably accurate estimates for more moderate extremes (Folland and Anderson, 2002).
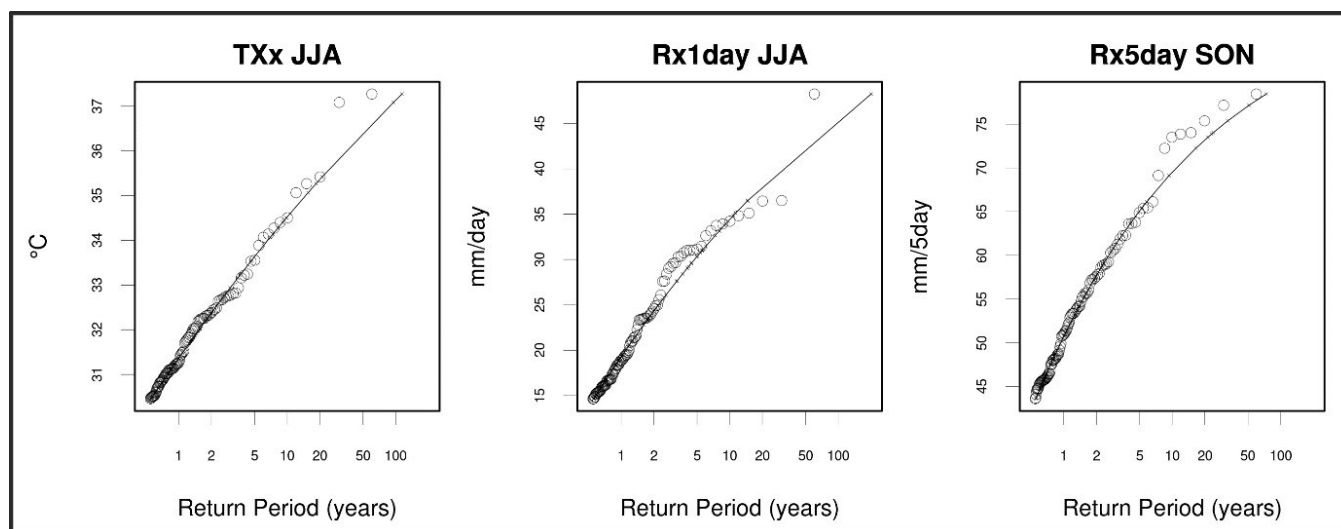

Figure 5. Return level curves for the 25km grid box containing central London, derived from the values of μ_0, β_0 and ξ_0 shown in Fig. 4. Curves for TXx and Rx1day in summer, and Rx5day in autumn, are compared with corresponding observations of the highest 108 events in the 59-year observational record (open circles). These are ranked from high to low (m = 1 →108), assumed to correspond to return periods of 59/m years.

Calibration of EV parameters for GCM simulations

In order to construct the probabilistic projections, values of μ_T and β_T are required for the SLAB, ESPPE and CMIP5 simulations (see section 2.1). Values are obtained for each of the five UK_GCM land points available in HadCM3 (see section 2.1 and Fig. 11). This is done by calibrating Eq (3) separately for each simulation. In this case the model data is not pooled with observations, hence the μ_Δ and β_Δ terms are omitted. In addition to the required values of μ_T and β_T, these GCM fits produce values of μ_0 and β_0 indicative of simulated baseline conditions. These are of interest for GCM evaluation but are not used in the subsequent generation of projections.

In the case of SLAB PPE members, a pair of 20-year simulations of historical and doubled CO_2 climate are available to calibrate μ_T and β_T, based on the 20-year average change simulated in GMST. The relative shortness of these simulations increases the uncertainty in member-specific values, due to limited sampling of internal variability. We therefore winsorize the distribution of 280 SLAB values of μ_T and β_T, to avoid retaining potentially unreliable outliers. This is done by estimating the standard deviation from the interquartile range (assuming a normal distribution) and resetting to the median SLAB value any results that differ from it by more than 3.5 standard deviations.

For the ESPPE and CMIP5 members, longer transient simulations from 1860-2100 are available, so the diagnosed estimates are less noisy and winsorization is not necessary. EV parameters are calibrated[2] for each of these, using smoothed time series of GMST[1]. We illustrate this process in Fig. 6, showing examples for the grid box covering southern England from one member of each ensemble, using the RCP8.5 scenario[4]. Time-dependent values of the 50-year return level (red curves), diagnosed from the relevant EV parameters, are used to illustrate the effects of climate change. Also shown (black curves) are the values of the 3% threshold used to identify extreme events (blue crosses) using the PoT method. The smoothed time variation of the threshold is determined by fitting a spline[1] to empirical values diagnosed annually.

For summer TXx, both simulations show substantial increases in the 50-year return level during the 21st century, leading to projected values approaching or exceeding 50°C by 2100. This is driven by increases in GMST of approximately 4.6°C in the ESPPE member and 4.3°C in HadGEM2-ES (the selected CMIP5 model), by the 2090s relative to 1981-2000. Increases in the 50-year return level substantially exceed those in GMST, amounting to more than 10°C in the ESPPE member, and just under 10°C in HadGEM2-ES. These changes are associated with large future increases in μ_T (2.7 for the ESPPE member and 1.9 for HadGEM2-ES). The latter simulates values close to 39°C during the baseline period, while the baseline values in the ESPPE member are 40°C. In predictive mode (see discussion above), the use of observed baseline values would therefore alter the projected absolute values. For example, applying the observed baseline of ~36°C for London (Fig. 4) would reduce the future values by several degrees. For Rx5day in autumn, both simulations project modest increases in the 50-year return level, but from considerably different baseline values. In this case, applying a common observed baseline would bring the future projections of absolute values closer together.

---

[4] In the case of CMIP5-ESM, values of μ_T and β_T were obtained from RCP8.5 results, as shown in Fig. 6. For ease of comparison, the ESPPE example in Fig.6 is also taken from RCP8.5. In the ESPPE case, however, the values of μ_T and β_T used in section 2.3 were obtained by pooling results from the three available scenarios (RCP2.6, RCP8.5 and SRES A1B – see section 2.1) in order to increase the statistical robustness of the estimates. This was supported by tests showing that scenario-specific values were statistically indistinguishable.
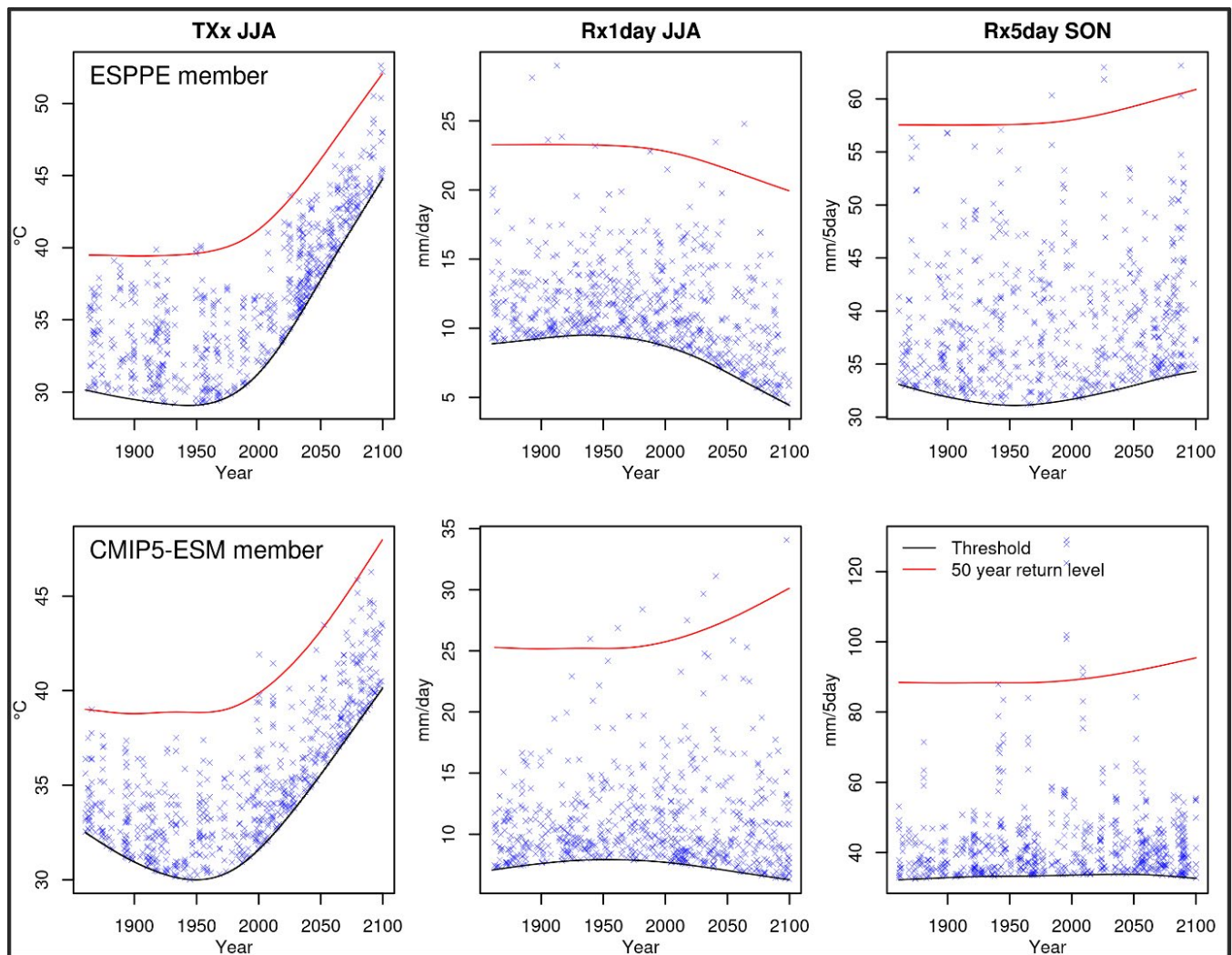
Figure 6. Examples of projected 50-year return levels (red curves) for southern England, derived from a single ESPPE member (top row) and a single CMIP5-ESM member (bottom row). These are obtained for TXx and Rx1day in summer, and Rx5day in autumn, by fitting the EV parameters of Eq (3) to simulated events (blue crosses). These exceed time-varying thresholds (black curves) that are chosen to identify 3% of events per year, on average, as extremes for use in calibration. Further details in text.

For Rx1day in summer, the two GCM simulations give baseline values of ~23mm and ~25mm, respectively. Not surprisingly, these are well below observed baseline estimates at the 25km scale (e.g. ~40mm for London in Figure 4), as the GCM values represent a larger spatial scale of ~300km. The projected changes are different, with the ESPPE member showing a future reduction in the 50-year return level, while HadGEM2-ES shows an increase. These results illustrate the uncertainty in future changes, discussed further in section 3. We would expect the 50-year return level to be exceeded ~5 times between 1860-2100, on average. In practice, the number of exceedances in Fig. 6 ranges from four to seven, indicating that the fitted EV parameters give credible estimates.

2.3.   Method for probabilistic projections of return levels

This section describes the steps used to combine the Bayesian and EV methods of sections 2.1 and 2.2, in order to produce probabilistic projections of return levels for TXx, Rx1day and Rx5day.

2.3.1   Probabilistic projections of GMST

These (see Fig. 1) are required as inputs to Eq (3), when used in predictive mode to produce time-dependent pdfs of the location and scale parameters. We use the pdfs of GMST(t) produced by Murphy et al. (2018), for the A1B, RCP2.6, RCP4.5 and RCP8.5 emissions scenarios. A recap of the method is provided below, with more detail available in Murphy et al. (2018).

The probabilistic projections of GMST were generated using the simple climate model (SCM) of Harris et al. (2013). This predicts global ocean and land surface temperatures using planetary energy balance principles, a one-dimensional advection-diffusion equation for vertical ocean heat transport, and a globally-averaged representation of the earth's carbon cycle. The SCM is used to sample prior pdfs of its driving inputs, which include equilibrium climate sensitivity (ECS[5]), and several parameters controlling ocean heat uptake and carbon cycle feedback.

The prior pdf of ECS is constructed using an emulator calibrated from the 280-member SLAB PPE, using ECS values for CMIP5-ESM members to adjust the distribution to account for structural modelling uncertainty via the discrepancy term (Boxes 1, 2, 4 and 5 in Fig. 2, discussed in section 2.1). Corresponding distributions for ocean heat uptake and carbon cycle parameters are obtained by prescribing priors consistent with the spread of values simulated by the 57-member ESPPE (Fig. 2, Box 3).

Discrepancy estimates are also required for the time-dependent outputs of the SCM, which include upper ocean heat content and atmospheric CO_2 concentration, as well as GMST. These are obtained by finding best analogues from the above input prior distributions, that lead to SCM predictions matching the results of CMIP5-ESM members as closely as possible (Fig. 2, Box 4). The discrepancy terms for heat content, CO_2 concentration and GMST can only be calibrated directly for the RCP8.5 scenario, as CMIP5-ESM results are not available for the others. For these, we assume that the median discrepancy is proportional to the predicted GMST change, using scaling coefficients derived by linear regression to the responses of CMIP5-ESM members for the RCP8.5 scenario (Murphy et al., 2018). The variances of the discrepancy distributions are assumed independent of emissions scenario.

Projections of GMST are produced by running the SCM from 1860-2100, driven by prescribed emissions of CO_2 and prescribed concentrations of other major greenhouse gases, switching from historical to scenario values beyond 2005. The historical forcing also accounts for solar and volcanic influences, whilst aerosol forcing is included by sampling from the pdf specified in the IPCC Fifth Assessment Report (Myhre et al., 2013). This pdf applies to the forcing for 2011 relative to pre-industrial conditions, with values for other historical and future periods estimated by scaling according to prescribed sulphur dioxide emissions.

---

[5] ECS is defined as the equilibrium (steady state) response of globally and annually averaged surface temperature to a doubling of CO_2 concentration in the atmosphere. It is a standard benchmark measuring the long term sensitivity of a climate model, or of the real climate system, to a sustained change in greenhouse gas concentrations.

The SCM projections of GMST are modified, by adding a bias term and sampling from separate distributions representing discrepancy, and residual prediction errors. The bias and residual terms are pre-calibrated[6], by using the SCM to predict the projections of ESPPE members. The bias term accounts for errors in the transient response arising from simplifications used in the SCM, such as its assumption that climate feedbacks are independent of time (e.g. Gregory and Andrews, 2016). The dominant contribution to the residual term arises from internal climate variability, which is omitted by construction from the SCM. Sampling the residual distribution thus ensures that the probabilistic projections account for uncertainty in GMST arising from internal variability, as well as modelling uncertainty in physical and carbon cycle feedbacks, aerosol forcing and ocean heat uptake.
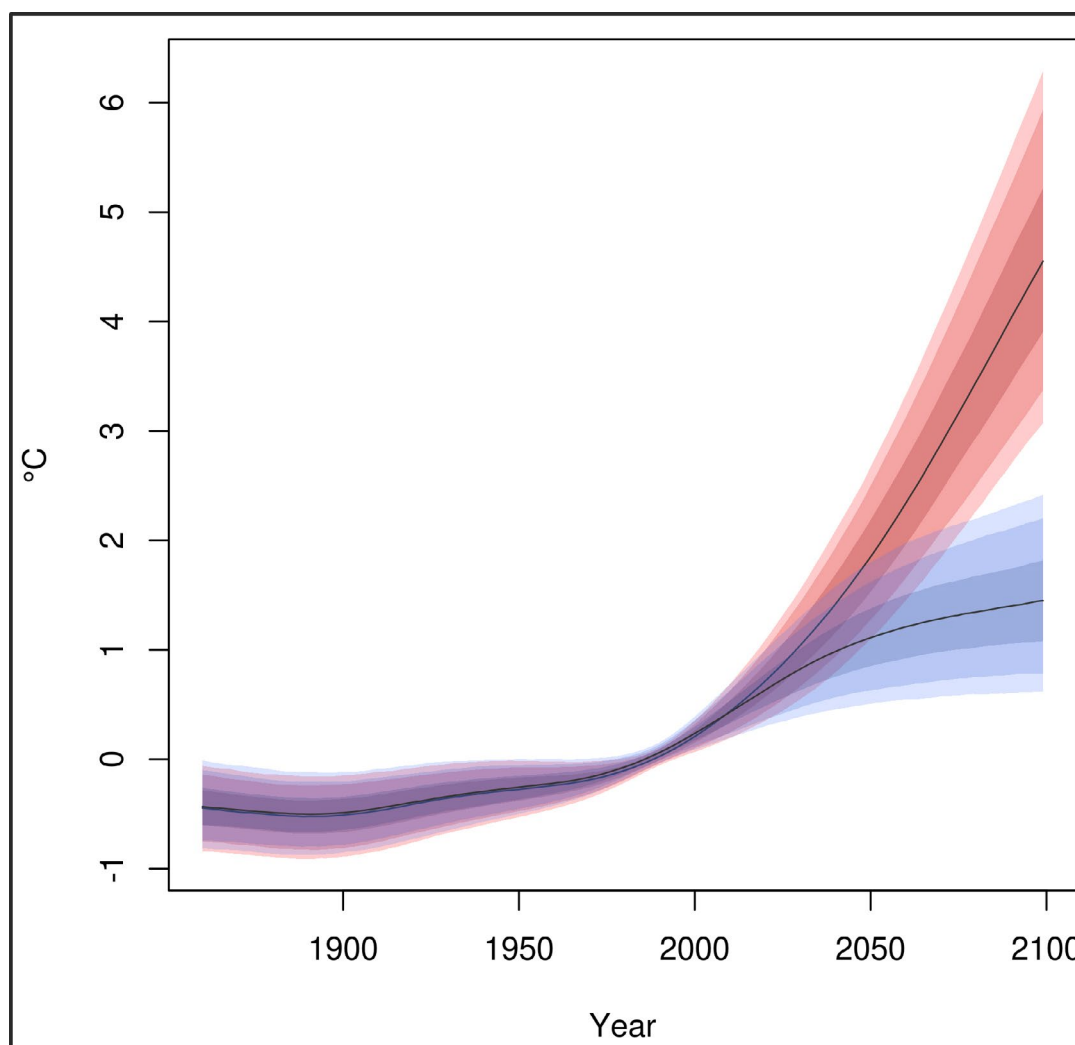

Figure 7. Probabilistic projections of GMST in response to historical changes in radiative forcing to 2005, and to the RCP8.5 scenario from 2006-2100 (red plume). The blue plume shows corresponding results for the RCP2.6 scenario. Anomalies are calculated relative to the 1981-2000 baseline period. Grey curves show the medians of the pdfs, and shading denotes the 5th, 10th, 25th, 75th, 90th and 95th percentiles. The pdfs are calculated from 3000 sampled realisations of annual GMST changes, which are temporally smoothed to isolate time-dependent changes on time scales of 30 years and longer. This allows the GMST projections to be used, via Eq (3), to determine the effects of secular climate change on the evolving characteristics of corresponding EV distributions. Details in text.

[6] Direct calibration of the bias and residual terms is possible for the RCP2.6, RCP8.5 and A1B emissions scenarios, for which ESPPE simulations are available. For RCP4.5 and RCP6.0, we estimate the bias term assuming a linear dependence on the GMST response (as described above for the median discrepancy term), using a scaling coefficient derived from the available ESPPE results (Murphy et al., 2018). The residual variances are assumed independent of emissions scenario.

The posterior probabilistic projections of GMST (Fig. 7) consist of 3000 emulated realisations. These are generated from an initial set of 106 realisations that sample the prior distributions of SCM inputs. These are weighted according to relative likelihood, estimated using the set of observational constraints (Box 6 in Fig. 2) described in section 2.1. The 10^6 realisations are then resampled, with replacement according to weight, to obtain 3000 GMST pathways, convenient for the calculation of time-evolving percentiles[7] of the pdfs for GMST anomalies (black and coloured curves in Fig. 7).

Prior to calculating the pdfs, the 3000 time series of GMST are smoothed using the same spline[1] applied when calibrating μ_T and β_T from individual climate model simulations in section 2.2. This is done in order to remove variability on time scales shorter than ~30 years, hence isolating the effects of long-term climate change. The future plumes of Fig. 7 are therefore slightly narrower than the corresponding GMST pdfs of Murphy et al (2018) (their Fig. 2.8), as the latter include variability on shorter time scales.

2.3.2 Probabilistic projections for climate change effects on EV parameters

In addition to GMST(t), probabilistic projections for μ_T and β_T are also required (Box 8 in Fig. 2) as inputs to the return level projections (Fig. 1). Here, we summarise how these are calculated, for the UK_GCM grid boxes (Fig. 11). As for GMST, 3000 sampled estimates are used to construct the pdfs. Since μ_T and β_T contain no explicit time-dependence, they are analogous to regional variables termed "normalised transient responses (NTR)" by Murphy et al. (2018), which consist of characteristic regional changes per unit warming in GMST. The description below outlines the corresponding implementation for μ_T and β_T.

Initially, prior pdfs of the equilibrium response to doubled CO_2 are produced, using emulators trained on the SLAB PPE results (as for ECS described above). Examples are shown in the top panels of Figs. 8 and 9, using values of μ_T and β_T for the southern England grid box for TXx and Rx1day respectively. The SLAB results are the green histograms, and the green curves the prior pdfs. Following Murphy et al. (2018), these are termed normalised equilibrium responses (NER).

The next step adjusts the prior pdfs to account for drivers of transient climate change, and associated parametric uncertainties, represented in the ESPPE but not in the SLAB simulations. This adjustment (hereafter referred to as the "offset" term) accounts for ocean circulation changes, terrestrial vegetation feedbacks and anthropogenic forcing due to sulphate aerosols. Each of the 57 ESPPE members uses a set of parameter perturbations in its atmospheric components that corresponds to one of 17 members of SLAB (SLAB17, blue dots in Figs. 8 and 9, top panels). The number of ESPPE members corresponding to each SLAB17 member varies between one and four (Lambert et al., 2013).

---

[7] In Murphy et al (2018), we used the term "probability level" to denote cumulative probability thresholds. For example, the 90% probability level defined an outcome with a 10% chance of being exceeded. Here, we use "90th percentile" to denote the equivalent threshold. This is done to avoid confusion with the term "return level", which defines the intensity of the relevant extreme event. Thus, a statement that: "the 90th percentile of the 20-year return level of daily maximum summer temperature is 40°C in 2050" would mean that there is a 10% chance of seeing a 20-year return level higher than 40°C by 2050, according to the UKCP probabilistic projections of daily summer maximum temperature.
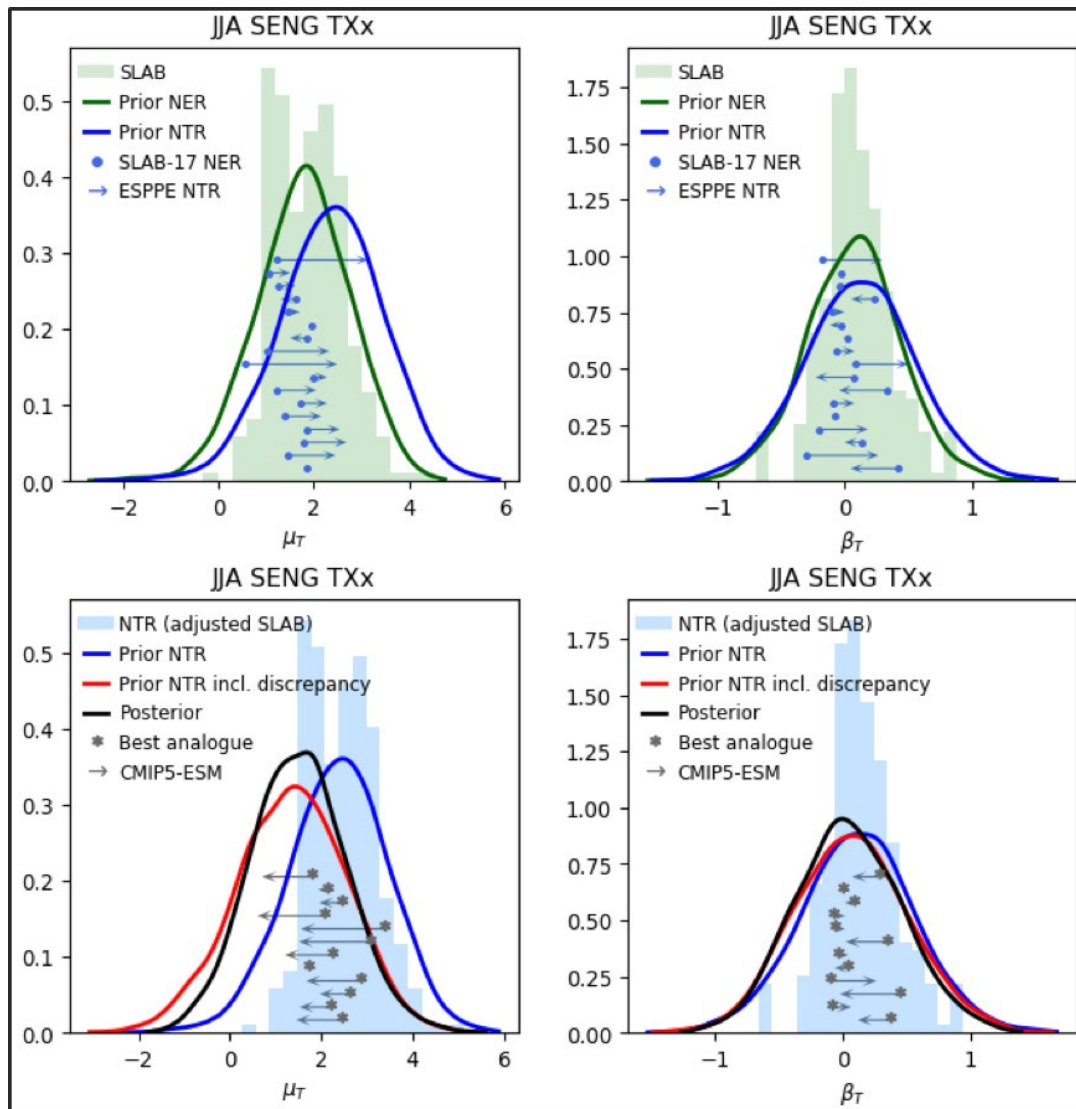
Figure 8. Steps in the calculation of pdfs of μ_T and β_T for GCM grid boxes, using TXx in summer for the southern England box as an example. Upper panels show results from the 280 SLAB simulations of the equilibrium response to doubled CO_2 (green histograms). These are converted into prior pdfs (green curves, labelled NER - normalised equilibrium responses) by using statistical emulation to sample the whole atmospheric parameter space of HadCM3. These are in turn converted into priors for the normalised transient response (NTR, blue curves), that account for ocean circulation changes and additional parametric uncertainties in ocean, sulphur cycle and terrestrial ecosystem processes in the earth system configuration of HadCM3. This is done by applying the average offset between NER and NTR, calculated from the 17 blue lines. Circles show NER values for 17 SLAB model variants, and arrow heads represent responses averaged over corresponding members of the ESPPE of HadCM3 earth system model variants. These are ESPPE variants that use corresponding perturbation sets in their atmosphere component. The lower panels show NTR values for SLAB PPE members (blue histogram) following application of the offset. They illustrate the impact of adding discrepancy to the prior distribution of NTR from the middle panel (cf red and blue curves). In these panels, arrows show twelve individual discrepancy estimates obtained by finding best PPE analogues (asterisks) to each of the CMIP5-ESM simulations (arrow heads), from which the median and variance of the (assumed Gaussian) discrepancy distributions are calculated. The black curves show posterior pdfs of NTR, obtained by weighting points in the HadCM3 parameter space of earth system processes according to relative likelihood, derived from application of the observational constraints listed in section 2.1.

By comparing corresponding members, a 57-member sample of differences is obtained for μ_T and β_T, for each season, variable and GCM grid point. The blue arrows in Figs. 8 and 9 (top panels) show the average difference between each SLAB member, and the 1-4 corresponding ESPPE members. The prior NER pdfs are then converted to prior NTR pdfs, by adjusting the emulated equilibrium responses to account for the additional processes represented in the ESPPE. This is done by adding the 57-member average offset to each emulated outcome, and sampling an associated uncertainty diagnosed from the standard deviation of

the individual offsets. Our approach assumes a Gaussian distribution independent of location in parameter space, since we lack sufficient ESPPE simulations to calibrate potential parametric dependencies of the offset. This results in the blue curves shown in Figs. 8 and 9 (top panels). In general, the blue NTR curves are liable to shift, and also become slightly broadened, compared to the green NER priors. In the examples shown, a significant shift is apparent only for μ_T in the case of TXx (Fig. 8), for which most ESPPE simulations give higher values than corresponding SLAB members.

In the final steps (Figs. 8 and 9, lower panels), the prior NTR distributions are adjusted to account for the discrepancy term (red curves cf blue curves), and then converted into posterior distributions (black curves) using the weights obtained by applying the observational constraints.
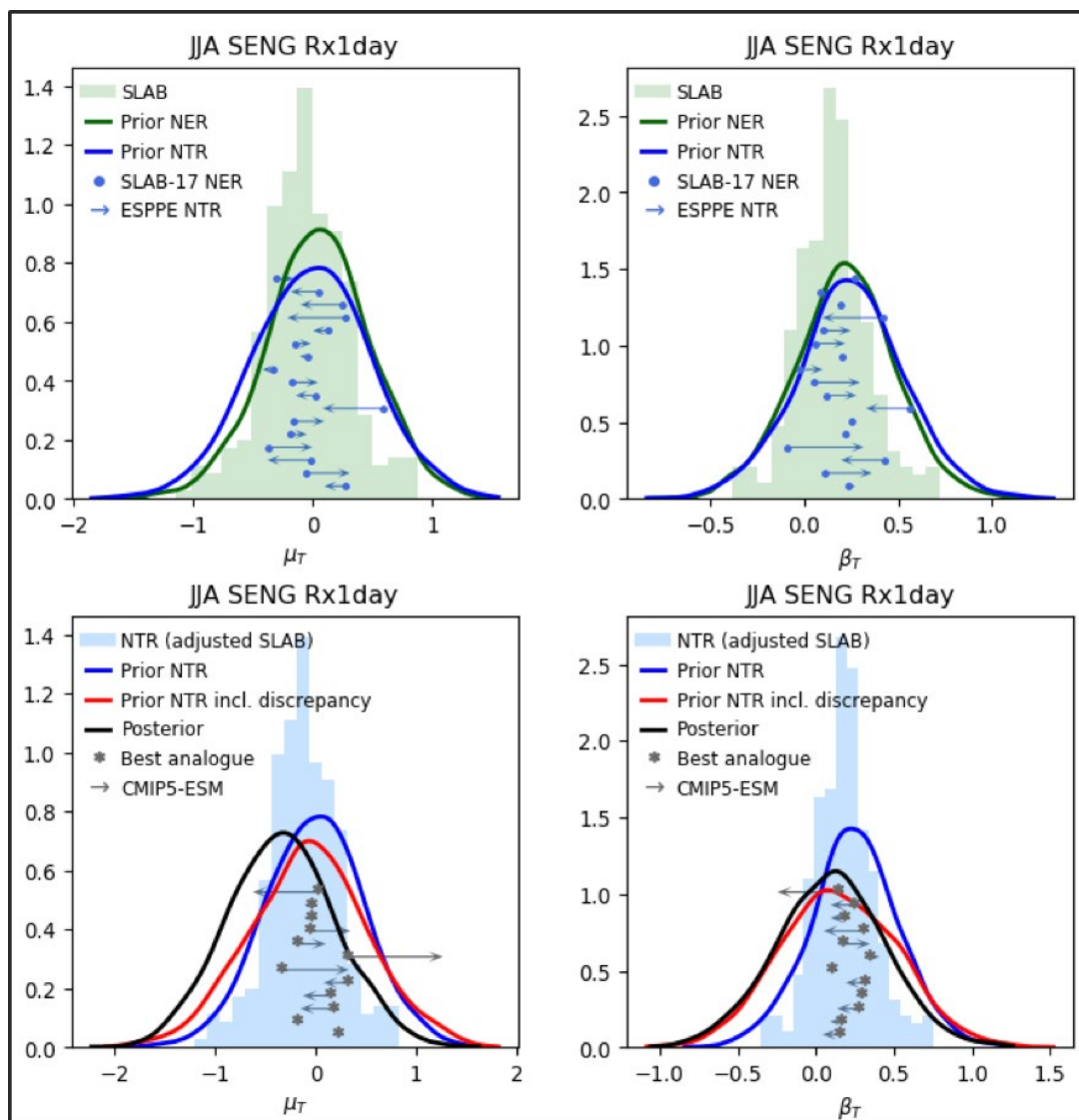

Figure 9. As Fig. 8, for Rx1day in summer in the southern England GCM grid box.

The arrow heads in Figs. 8 and 9 (lower panels) show results from the twelve CMIP5-ESM members used to sample structural model uncertainties, with the arrows representing the differences between these and their best PPE analogues (asterisks). These are found by searching[8] the 54-dimensional space of earth system parameters in HadCM3. The search is conducted by minimising the multivariate distance between a broadly-based set of variables simulated by CMIP5-ESM members, and emulated values from points in the HadCM3 parameter space. This multivariate approach is motivated by our aim to consider a wide range of emergent characteristics in identifying PPE analogues that provide the best overall matches to the complex set of earth system processes represented in each CMIP5-ESM member. This also reduces the risk of identifying an unrealistically close match in any particular metric of interest, that might due to some fortuitous compensation of differences between underlying process-drivers.

The optimisation variables are derived from:

- historical, 20-year mean climatologies of global, seasonal, spatial fields for twelve key climate variables (see Table B.1 of Murphy et al., 2018).

- normalised future transient responses for the same set of variables, also expressed as global, seasonal, spatial fields.

- normalised future transient responses for seasonal, UK-average changes in surface air temperature and precipitation.

- the set of SCM variables controlling the global transient response of GMST (see section 2.3.1).

Following Sexton et al. (2012), the above variables are non-dimensionalised by scaling by the standard deviation of their prior pdf in the PPE, noting that for global spatial variables (which consist of a mix of latitude-longitude and latitude-height fields), this is done by averaging the standard deviations of local or zonal-mean values. Also, the dimensionality of the climatology and normalised transient response datasets is reduced, by using the six leading eigenvectors in each case.

The discrepancy distribution is assumed Gaussian, with a median and spread estimated as the average difference between CMIP5-ESM members and their best analogues, and the standard deviation of the individual differences. In general, adding the discrepancy term will (like the offset term) broaden, and potentially shift, the prior distribution. For TXx, the distribution for μ_T shifts to lower values when discrepancy is added. This opposes the positive shift introduced by the offset term. It occurs because CMIP5-ESM members simulate values smaller than the median of the PPE-based prior distribution (blue curve). For Rx1day, the discrepancy variance is sufficient to drive a noticeable broadening of the prior, for both μ_T and β_T. This reflects the influence of CMIP5-ESM members lying near the extremes of the PPE-based prior. In future work, availability of a larger set of simulations (perhaps through addition of forthcoming CMIP6 models) would help to reduce the sensitivity of the discrepancy calculation to outlying multi-model results.

[8] This search uses the Nelder-Mead algorithm. We find 3 best analogues for each CMIP5-ESM member, distinguished by starting the algorithm from differing, randomly-selected start points in parameter space, and minimising the multivariate mean-square-difference between simulated and emulated values of the variables described in the text. Each arrow in Fig. 8 (lower panels) represents the estimate of structural error for the relevant CMIP5-ESM member, estimated by averaging the distances over the 3 analogues.

In general, applying the observational constraints (black curves in Figs. 8 and 9, lower panels), alters the relative probability of alternative outcomes for of μ_T and β_T, compared with the prior distributions (red curves). For TXx, the effects are relatively minor, featuring a slight narrowing of the pdfs, accompanied by higher relative probabilities for outcomes near the modes. This indicates that in these cases, there is no strong relationship between the historical performance of alternative PPE variants, and the projected values. For Rx1day, the observational constraints downweight positive values of μ_T and upweight negative values, sharpening the distribution and reducing its modal value.

2.3.3   Credibility check on outputs of Bayesian methodology

Following Murphy et al. (2018), we compare the pdfs of μ_T and β_T against results from the underlying GCM simulations, for TXx, Rx1day and Rx5day. This allows us to identify cases where statistical processing in the methodology may have led to pdfs implying significant tail probabilities for outcomes beyond the range covered by the GCM results.

Some extension beyond the GCM range would be expected. This could occur, for example, through the use of emulators (Fig. 2) to estimate PPE outcomes for unsampled parts of parameter space, or by adding the discrepancy term to PPE outcomes to estimate results adjusted to account for structural model error (section 2.1). However, if the tail probabilities are substantial, this may indicate reduced confidence in the pdfs, due to a lack of support from the GCM results for outlying outcomes produced by the statistical calculations.

We measure the fraction F of the 3000 sampled estimates that lie beyond the most extreme model outcome, choosing the larger of the values found at the upper and lower ends of the range (Murphy et al., 2018). In order to include the 280 SLAB simulations of equilibrium climate change (as well as the 69 transient simulations from CMIP5-ESM and the ESPPE) in the GCM range, we adjust the SLAB outcomes to account for additional influences on the characteristic transient response, using the offset term described in section 2.3.2. We use F>15% as a threshold indicating potential credibility issues requiring investigation (Murphy et al., 2018).

Figure 10 shows the credibility tests for TXx and Rx1day in southern England in summer (following the examples of Figs. 8 and 9). For β_T the tails of the pdf extend beyond the smallest and largest simulated values for both TXx and Rx1day. However, all values of F are smaller than 15%, so we interpret the statistical inflation of the pdfs as defensible estimates of GCM results that might have been obtained with larger ensembles. For Rx1day, the high tail arises mainly from emulation of a slightly wider range of outcomes than shown by SLAB (compare blue histogram vs blue pdf in Fig. 9, bottom right panel), whereas the low tail is driven by the addition of spread associated with the discrepancy term (red cf blue pdfs) plus the impact of weighting (black cf red pdfs). This indicates that accounting for structural model uncertainty and observational constraints implies potential for larger reductions in β_T than found in our set of GCM results.
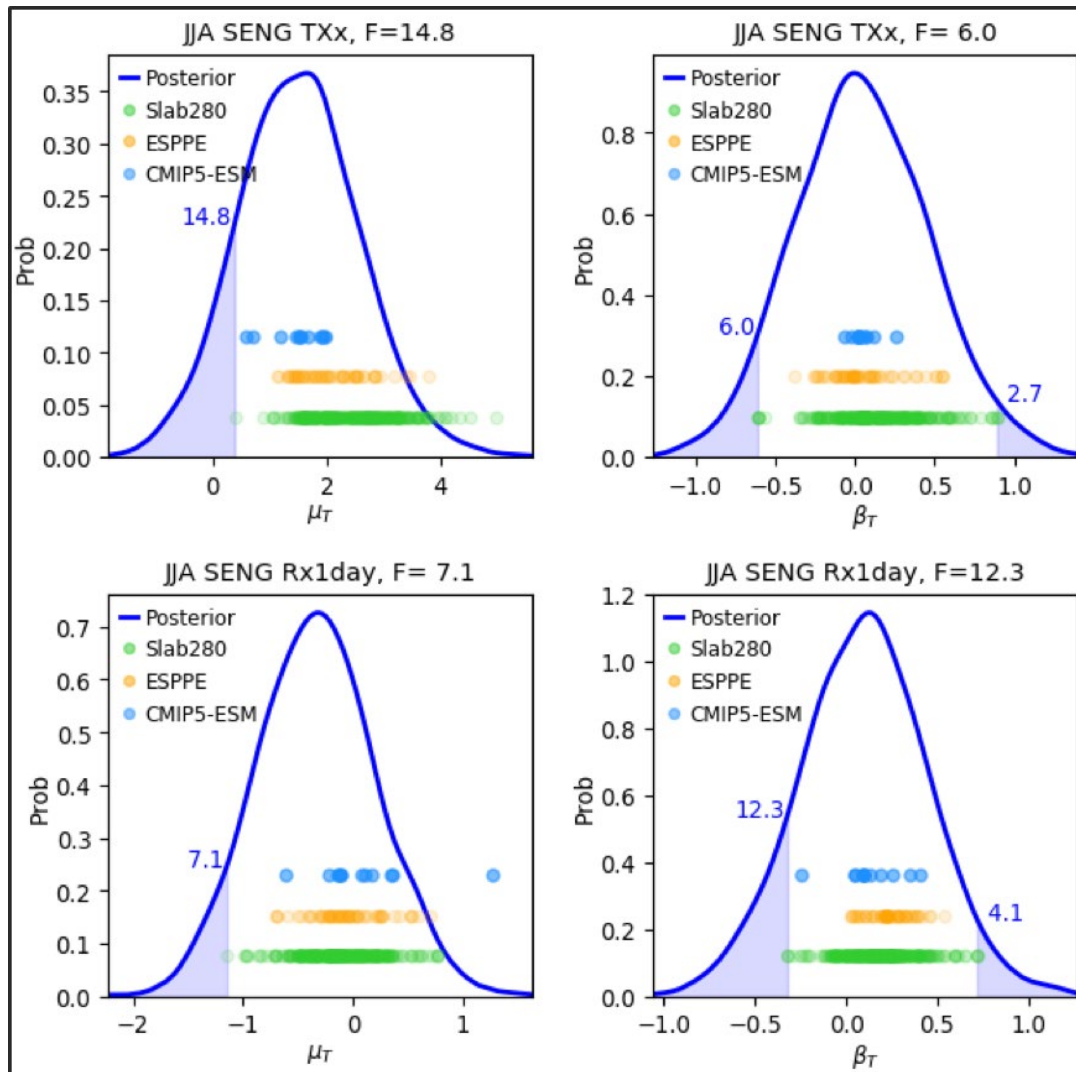
Figure 10. Credibility check on GCM-scale posterior pdfs of μ_T and β_T (blue curves) using TXx and Rx1day in summer for the southern England box as examples. Coloured circles show values of μ_T and β_T from the 349 GCM simulations fed into the Bayesian calculations, and blue numbers show upper and lower tail probabilities for outcomes beyond the range of GCM results. In each case, the larger of these numbers defines F, where F>15% is taken to indicate a level of statistical inflation that flags the relevant tail of the pdf as potentially unrealistic. The GCM values are NTR estimates, derived from the 57 ESPPE (orange), 12 CMIP5-ESM (blue), and 280 SLAB (green) simulations. The SLAB results include the offset term of Fig. 8, applied to convert NER estimates, derived directly from the SLAB simulations, into corresponding estimates of NTR.

For μ_T, Fig. 10 shows extension of the low tails in both cases, large enough to lead to F=14.8% for TXx. This is caused by a significant negative shift in the pdf driven by adding the discrepancy term (Fig. 8, red cf blue curves). The shift arises because the CMIP5-ESM results all lie towards the lower end of the range of simulated PPE outcomes. The Bayesian method is predicated on the assumption that structural biases in the base model used for the PPE are not too large, and that the PPE provides a reasonable first-order estimate of the effects of modelling uncertainty (e.g. Sexton et al., 2012). Instances of a substantial discrepancy term therefore indicate reduced credibility in the relevant pdfs. In this case, we conclude that reduced confidence should be placed in the low tail of μT for summer TXx in Fig. 10. However, despite the negative shift caused by discrepancy, we also conclude that the upper tail of the pdf remains plausible. This is because it essentially encompasses the highest outcomes found in the GCM simulations (recalling that the coupled simulations provide more robust estimates of EV parameters than the SLAB simulations, as they are derived from much longer integrations (section 2.2)).

Table 1. Fraction of seasonal pdfs of μ_T and β_T for which the tail-probability F exceeds 15%.

Variable	F > 15%	
	μ_T	β_T
TXx	0.30	0.10
Rx1day	0.05	0.05
Rx5day	0.00	0.05

For the six EV parameters, Table 1 shows the fraction of occurrences of F>15%, across 20 tests assembled by pooling results from the five UK_GCM grid points for each season. Following Murphy et al. (2018), we assess credibility on this aggregated basis, assuming that a general credibility issue is more likely to exist if a particular EV parameter fails the F>15% criterion in several of the 20 cases. Although all parameters show instances of F>15%, the frequency of occurrence is usually 10% or less, which we assess as acceptable.

The exception is μ_T for TXx, for which F>15% occurs in six out of 20 cases. All of these instances occur at the low end of the pdf. They arise mainly from shifts caused by negative values for the median of the discrepancy distribution (e.g. Fig. 10).

Such shifts will increase the probability of low outcomes for future return levels, noting that the effects of climate change on the location parameter will impart uniform upward or downward shifts to return level curves (Fig. 3), dependent on the sign of μ_T. However, assessments of future risks are likely to focus mainly on the high ends of the return level pdfs of TXx, rather than relatively benign outcomes sampled at the lower ends. This is likely to apply, for example, in impacts studies focused on improving understanding of future risks to public health and infrastructure arising from high temperatures (see Introduction). High future return levels will arise from samples of high-end values for either or both of μ_T and β_T. For TXx, the upper tails of the of the μ_T and β_T pdfs invariably encompass the highest outcomes from the underlying coupled GCM projections (as in Fig. 10, for example), noting that the two instances of F>15% for β_T occur at the low end of the distribution.

Therefore, we assess that the pdfs of future return levels for TXx are suitable for use in impacts studies, with the important caveat that reduced confidence should be placed in low-end outcomes. In particular, outcomes below the 10th percentile are likely to be less strongly supported by available GCM evidence than other parts of the pdf, and the 10% cumulative probability for such outcomes should be interpreted as a conservative estimate. Note also that we clip the return level pdfs for all variables at the 5th and 95th percentiles (see section 2.3.5). This reflects a general caveat that the tails of the return level pdfs may be more sensitive than less extreme outcomes to statistical assumptions in the methodology (potentially including the downscaling aspects discussed below, as well as the GCM-scale calculations covered in this section). For TXx, data is provided for the 5th to 10th percentiles for consistency with other variables. However, users can optionally discard TXx data below the 10th percentile, dependent on their priorities for exploring unlikely low-end outcomes versus focusing on a slightly smaller uncertainty range supported by higher confidence.

2.3.4 Downscaling relationships for climate change effects on EV parameters

Here we describe how the 11-member HadRM3-PPE is used to modify the GCM-scale projections of μ_T and β_T from section 2.3.2. As explained in section 2.1, HadRM3-PPE members were driven by transient climate change simulations using 11 variants of the coupled ocean-atmosphere configuration of HadCM3 (Collins et al., 2011), each GCM-RCM pair using corresponding sets of parameter perturbations. These parameter sets are also shared by 11 of the SLAB17 members.

In Brown et al (2014), linear regressions were used (their Fig. 8) to link each HadRM3-PPE member to the corresponding SLAB simulation. This was done to achieve the link in a single step, with the consequence that the relationships accounted for the effects of ocean thermal inertia and circulation changes on the transient response of UK climate at the national scale, as well as more localised downscaling influences arising from better resolution of mountains, coastlines, mesoscale circulations and storms (e.g. Rummukainen, 2016).

Here, however, the link between the equilibrium and transient responses is quantified by the offset term derived by mapping SLAB17 onto corresponding ESPPE members (section 2.3.2 and Figs. 8 and 9 (upper panels)). In our case, therefore, we require regression relationships between global and regional model changes that isolate the effects of downscaling. This is achieved by regressing HadRM3-PPE changes against their driving coupled simulations. Since there is significant uncertainty in determining the HadRM3-PPE EV parameters a total least squares regression is performed.

Figure 11 shows examples of projected changes in the five UK_GCM grid boxes from one of the ESPPE simulations. These illustrate the spatial information available from our GCM simulations, prior to modification through downscaling. Values of μ_T range from ~1.0-1.5 for TXx in summer, being smallest in Scotland. For Rx1day in summer, northern Scotland shows the largest value of μ_T (~1.2), while a negative value occurs over southern England. Values of μ_T exceed unity everywhere for Rx5day in autumn, the lowest value of ~1.5 occurring over Wales, with values exceeding 2.5 to the north. Values of β_T for Rx1day in summer and Rx5day in autumn are positive at all grid points, the largest values occurring in Scotland in the former case, and over Northern Ireland, northern England and Scotland in the latter. For TXx in summer, there are negative values over Wales and Northern Ireland, with values close to zero elsewhere.

For the downscaling predictands, local HadRM3-PPE values of μ_T and β_T (derived in section 2.2) are first aggregated, by averaging results from 25km grid boxes contained within each UK_GCM grid box. This reflects an assumption that for rare climate extremes, the RCM is more likely to provide added value at the spatial scale of the driving GCM simulations (e.g. Sørland et al., 2018) than at its native grid scale. Examples of the downscaling regressions are shown in Fig. 12. Letters denote results from each of the 11 GCM-RCM pairs. We pool results from the five UK_GCM grid boxes, to give 55 values from which a single, UK-wide regression relationship is calibrated. This is done because the small number of training simulations hinders diagnosis of robust relationships for individual UK_GCM grid boxes (Brown et al., 2014).
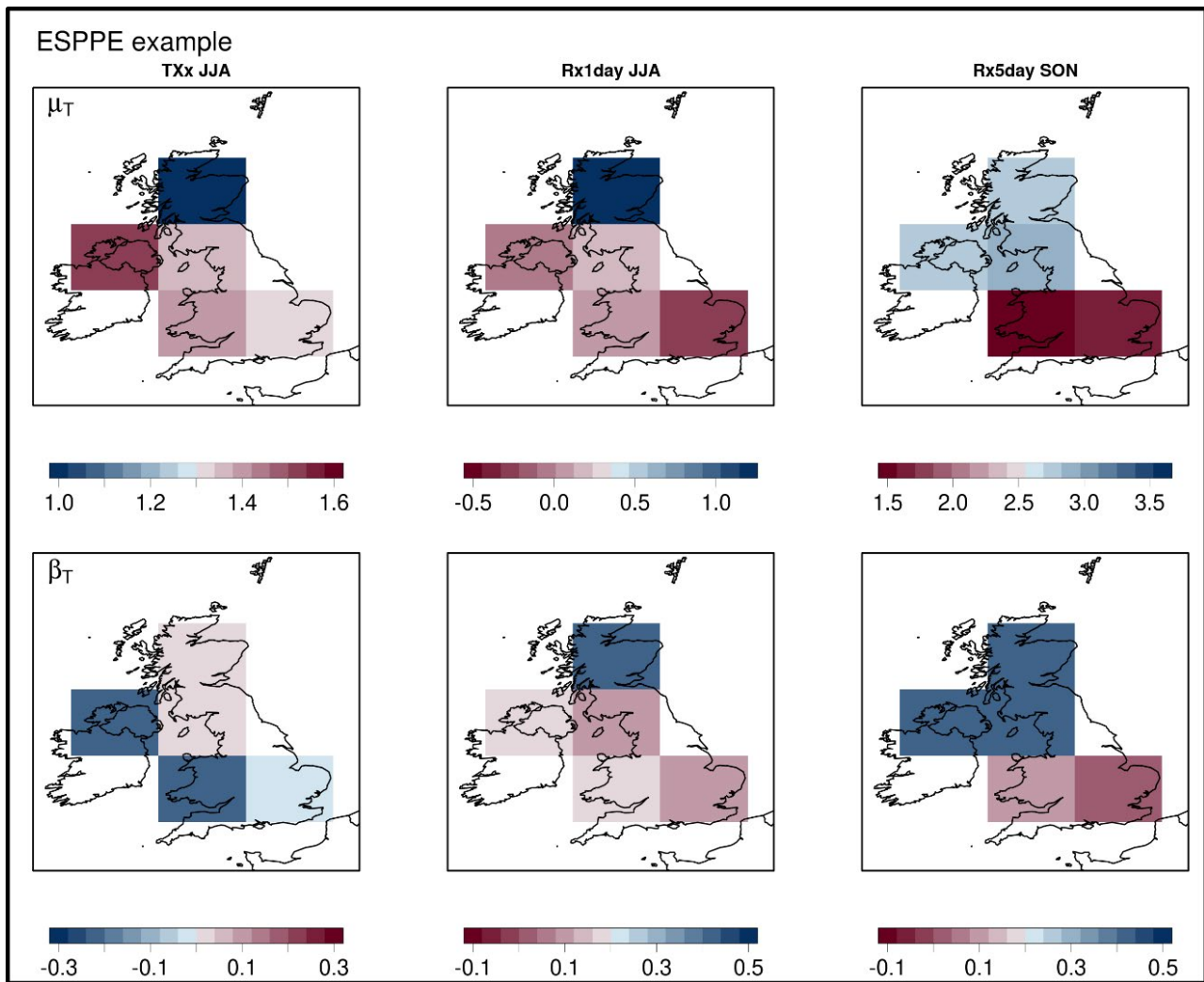
Figure 11. Values of μ_T and β_T from an ESPPE member for the five UK_GCM land points, for TXx and Rx1day in summer, and Rx5day in autumn. The maps show an example of the spatial information available from the GCM simulations, prior to modification using downscaling relationships.

Clear positive relationships (blue lines) are found for all the examples. In comparison with Brown et al. (2014), we find stronger relationships for β_T in all cases shown, and for μ_T in the case of Rx5day in autumn. This is probably due, at least in part, to the use of longer GCM time series (150 years from 1951-2100) to provide the predictor variables, reducing sampling noise compared to the 20-year SLAB simulations used by Brown et al. (2014). Building the relationships directly from the driving GCM simulations (and hence isolating specific downscaling effects as explained above) may also be a factor. Where stronger downscaling relationships occur in comparison to Brown et al. (2014), this implies that climate change effects found in the GCM simulations will exert more influence on the final projections.
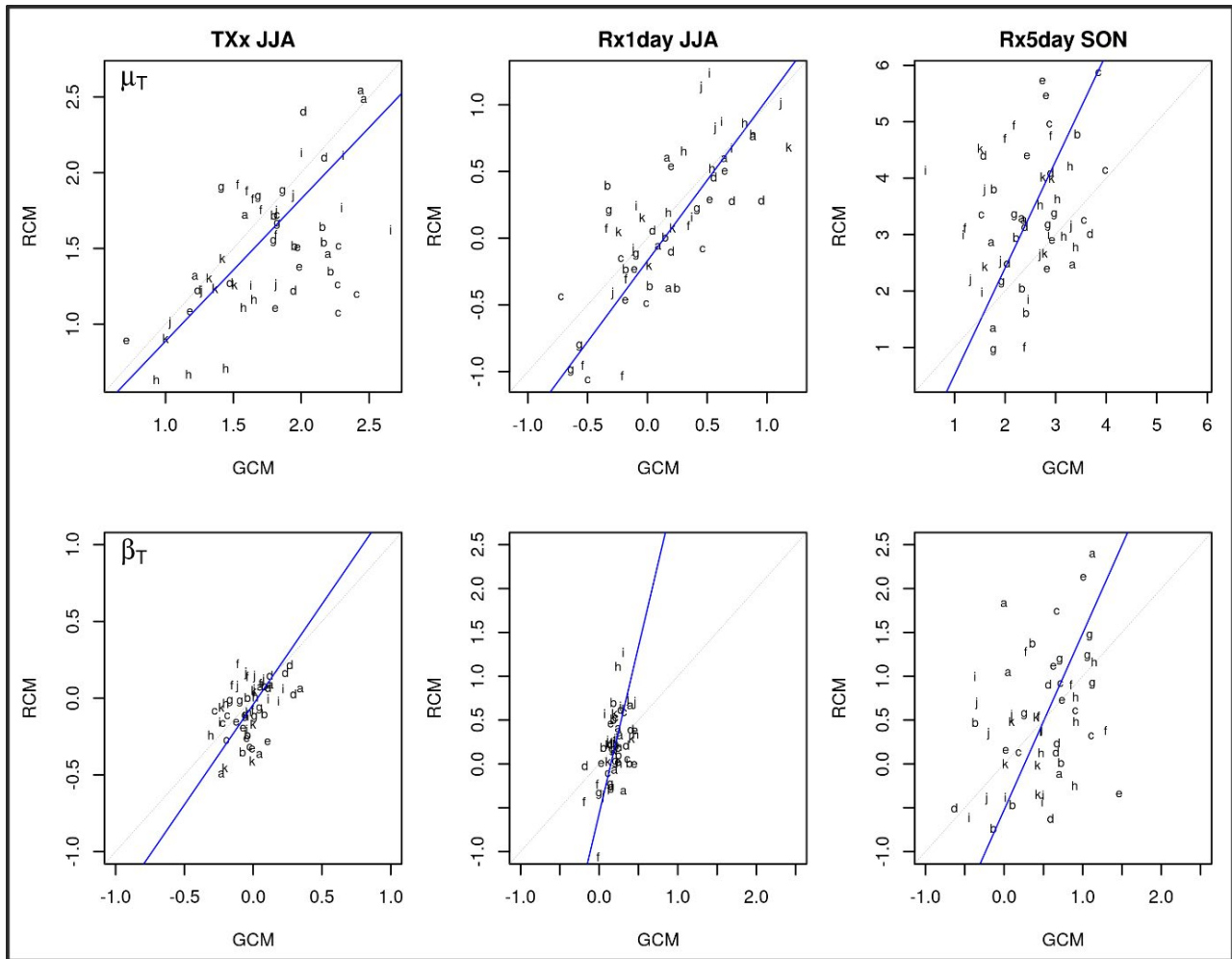
Figure 12. Linear regression relationships (blue lines) used to account for downscaling effects in pdfs of μ_T and β_T, for TXx and Rx1day in summer, and Rx5day in autumn. The relationships are calibrated using results from 11 HadRM3-PPE regional climate model (RCM) projections and their driving global climate model (GCM) projections (see text for details), with specific RCM-GCM pairs denoted as a-k. HadRM3-PPE results are first aggregated to GCM scale, by averaging values of μ_T and β_T over each 25km RCM grid box contained within the relevant GCM grid box (Fig. 11). This provides 55 pairs of GCM and RCM values (five GCM grid boxes x 11 RCM-GCM simulations), which are pooled to derive a single, UK-wide regression for each variable.

The GCM-scale posterior pdfs of μ_T and β_T are then modified using the downscaling relationships. Uncertainty in the latter is represented by sampling a Gaussian residual distribution, calibrated from the scatter of points around the regression lines (Fig. 12). Fig. 13 shows median values from the downscaled pdfs for TXx and Rx1day in summer, and Rx5day in autumn. For TXx in summer, median values are quite similar across the UK, ranging from ~1.25 to ~1.4 for μ_T, accompanied by small negative values of β_T. The full pdfs of μ_T and β_T for TXx explore broad ranges of values in each region (e.g. Fig. 8 (lower panels), which show results prior to downscaling). Projected changes are influenced by the combined effects of these parameters: For example, a positive value of β_T combined with a value of μ_T above 1.0 (e.g. Fig. 8, lower panels) would give rise to a future increase in long-period return levels that outstrips the corresponding change in GMST, while a negative value of β_T coupled with a small or negative value of μ_T could lead to samples showing future reductions in return levels in TXx, despite increases in GMST.
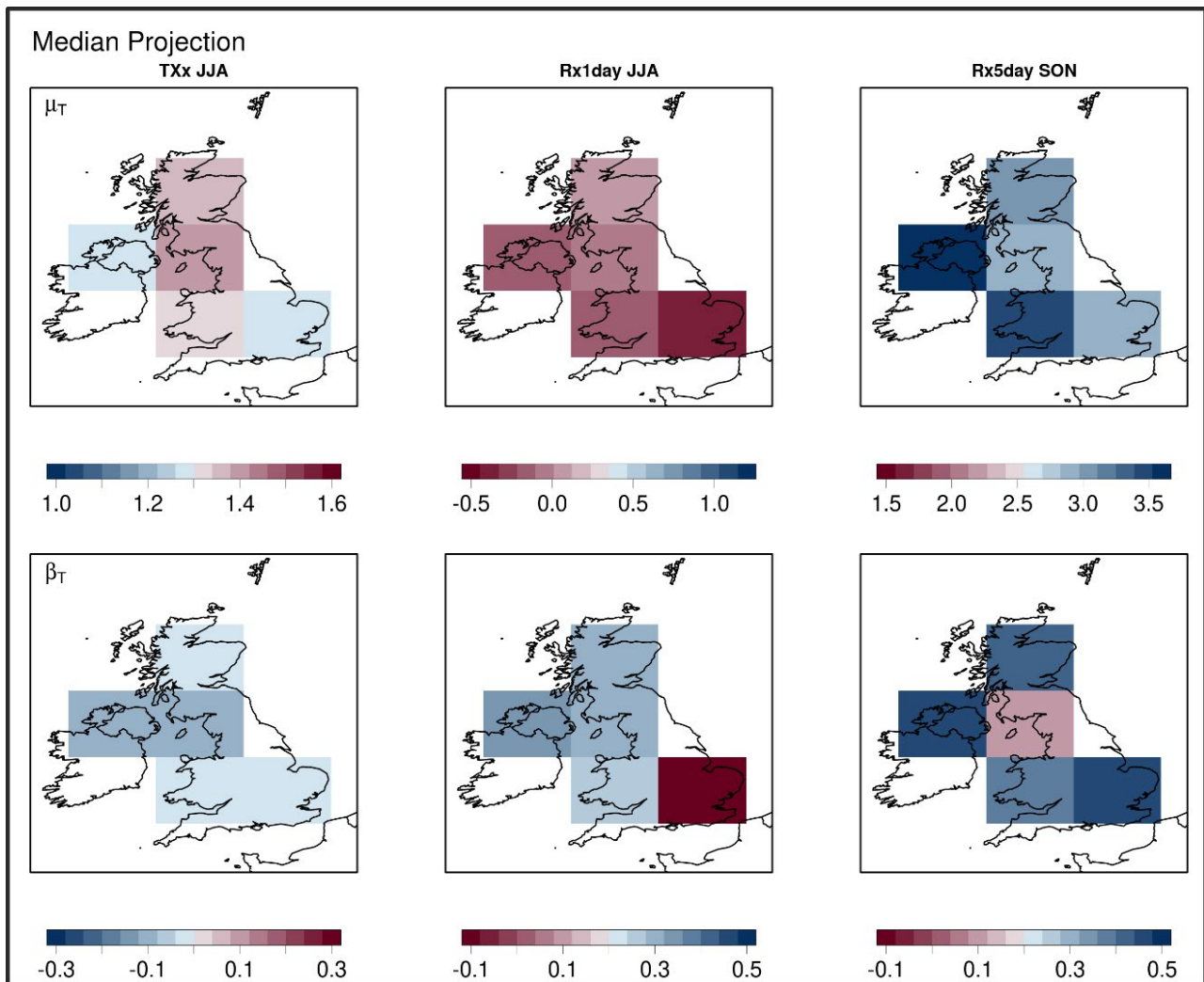
Figure 13. Median values of μ_T and β_T for TXx and Rx1day in summer, and Rx5day in autumn, following modification of their GCM-scale pdfs by application of the downscaling relationships of Fig. 12.

With the exception of Scotland, median values of μ_T for Rx1day in summer are negative. This indicates a median tendency towards a general reduction in return levels. On the other hand, the median values of β_T are positive everywhere apart from southern England. Positive values indicate a greater than even chance of sampling an increase in scale, that would offset the impact of negative μ_T on the rarest events (Fig. 3). The pdfs of μ_T and β_T in Fig. 9 encompass a broad range of both positive and negative values. Since the joint pdf of μ_T and β_T (not shown) does not feature a strong correlation between the two variables, this implies that pdfs of future return levels in southern England in summer will encompass both increases and decreases (see Fig. 15 in section 3).

For Rx5day in autumn μ_T exceeds 2.5 everywhere, and β_T exceeds 0.3 at all GCM boxes apart from northern England. These results indicate a balance of probability in favour of future increases in the return levels for extreme five-day precipitation accumulations.

The spatial patterns of median values differ from those of the individual ESPPE member of Fig. 11, illustrating the importance of sampling a diversity of outcomes from different GCM simulations. We note also that the patterns found in the median (or other percentiles of the pdfs of μ_T and β_T) will not necessarily correspond to those found in any particular GCM simulation.

2.3.5   Projections of return levels: Production and data

For each variable, season, 25km grid box and emissions scenario, the probabilistic projections are derived from 3000 emulated samples of climate change from 1951-2100. Each of these:

- consists of a draw from a specific point in the 54-dimensional parameter space of HadCM3 earth system processes, modified to account for structural modelling uncertainty (section 2.1). Each draw provides a set of values for GMST, μ_T and β_T for the relevant point in parameter space, derived using the statistical procedures described in section 2.3.1 and 2.3.2.

- uses a smoothed[1] GMST pathway from section 2.3.1 (recalling that the smoothing isolates the secular climate change responses assumed to drive changes in EV characteristics via Eq (3)). This is combined with the corresponding values of μ_T and β_T for the UK_GCM grid box that the target 25km grid box lies within (section 2.3.2), modified according to the relevant downscaling relationship (section 2.3.4).

- represents an equally likely outcome from the joint posterior pdf. This is achieved by weighting a larger sample according to relative likelihood, and then resampling with replacement to obtain a core sample of 3000 parameter sets. This is the same sample used by Murphy et al. (2018). Here, however, a further resampling is performed for each variable, season and grid box, to exclude unrealistically small or large future values of the scale parameter σ (see section 2.2). This leads to the exclusion of 10-100 sample members. In each case, the sample is restored to 3000 by random selection from the surviving members.

Equation (3) (with l set to zero) is then used to create time-dependent projections of location and scale for each sample member, by combining GMST(t), μ_T and β_T with the observed baseline values μ_0, and β_0 (Fig. 4). The shape parameter is fixed as ξ_0, independent of time, as explained in section 2.2. However, uncertainties[3] in the baseline parameters are accounted for, when assigning values of μ_0, β_0 and ξ_0 to each of the 3000 samples.

Finally, the EV distribution (Eq 2) is used to infer 20-, 50- and 100-year return levels (in °C or mm) for each sample, from 1951-2100. The 3000 samples are used to create probability distribution functions (pdfs) and cumulative distribution functions (cdfs) for each year, available as data and graphical plots from the UKCP User Interface at https://ukclimateprojections-ui.metoffice.gov.uk/.

The pdfs and cdfs are smoothed using a kernel density estimation technique, following the procedures used for other UKCP Probabilistic variables. Clipping of the pdfs and cdfs is also performed, to remove extreme values that may be less physically plausible than those within the bulk of the probability distribution. Reduced confidence in such tail values may arise, for example, from enhanced sensitivity to the details of the statistical assumptions required in the methodology (section 2.3.3, also Sexton and Murphy, 2012). For the other probabilistic variables, clipping was implemented by resetting values below the 1st percentile of the pdf and above the 99th percentile to those of the 1st and 99th percentiles respectively[9]. This winsorization procedure is also used here, but the clipping is applied at the 5th and 95th percentiles. This is because the use of extreme value theory adds another level of statistical processing to that included in the Bayesian calculations, and therefore increases the overall level of reliance of the high and low tails of the pdfs on the attendant assumptions.

As noted in the Introduction, our presentation of return levels as pdfs and cdfs of absolute values differs from the format of the other UKCP probabilistic projections, in which variables are presented as anomalies (Murphy et al., 2018). The return level projections therefore include a built-in bias correction (through the use of observational values of μ_0, β_0 and ξ_0), whereas users requiring absolute projections of the other variables (e.g. of monthly, seasonal or annual averages of surface air temperature, precipitation, cloud cover, etc) must combine the projected anomalies with a suitable baseline climatology.

There are many potential drivers of bias in the simulated tail distributions of TXx, Rx1day or Rx5day events. For example, these might include errors in the simulated average climatology, as well as various aspects of climate variability such as frequencies of weather types associated with extremes. Our bias correction effectively adjusts "top-down" for the combined effects of such influences on μ, β and ξ. This contrasts with some methods that users might deploy when correcting *specific types of error* in (say) a raw time series of GCM data, such as a bias in the long-term average or in the standard deviation of interannual variability.

In earlier UKCP Probabilistic results, Murphy et al. (2018) provided results from each of the 3000 samples individually, in addition to the pdf and cdf data. These were termed "realisations", each of which provided time series of monthly, seasonal or annual averages containing a temporally coherent representation of internal climate variability (as well as an evolving signal of long-term climate change). At *a specific spatial location*, the set of realisations could therefore be used in similar ways to time series output from an ensemble of climate model simulations. For example, the data could be searched for examples of individual threshold exceedances, or of hazards dependent on multivariate events, such as occurrences of hot, dry summers.

In the present case, however, an individual sample for a particular year consists of an EV distribution that represents *an entire climatological distribution of potential events*, rather than a single plausible outcome. Specifically, each EV distribution effectively accounts for an ensemble of phases of internal climate variability (as determined by the functional form of Eq (2) and the parameters of Eq (3)), combined with a level of long-term climate change represented by GMST(t). Whilst each EV distribution is then converted into individual estimates of return levels, these represent an expression of future climatological properties, rather than outcomes that will necessarily occur in the year in question. Therefore, we choose not to provide the return level projections as a set of individual realisations, as it would not be appropriate to interpret samples of return level in the same way as those of other UKCP Probabilistic variables.

[9] The kernel density estimation is described in a UKCP Technical Note, available at https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-technical-note-clipping-and-baseline-guidance-on-land-strand-1-data-in-ukcp18.pdf. This Note also describes a second type of smoothing, consisting of pooling the 3000 samples over an 11-year sliding window. This is not required here, since temporal smoothing is achieved by using low-pass filtered projections of GMST (see text).

[10] While this applies to analysis of changes through time at a particular location, it does not apply to spatial analysis of outcomes at a particular time. For example, realisations at neighbouring 25km grid squares could not be averaged to represent a spatial average over a bespoke, user-chosen region. This is because the statistical processing required to produce the realisations does not produce outputs with the full spatial coherence of raw climate model output (Murphy et al., 2018).

3.   Examples of the probabilistic projections

In this section we provide examples of the projected return levels, demonstrating the types of information that users can obtain from the data.

For the 25km grid box containing central London, Fig. 14 shows the time evolution of 50-year return levels associated with selected percentiles of their pdfs for the RCP8.5 scenario, for TXx and Rx1day in summer and Rx5day in autumn. For summer TXx the median (a central estimate of the projected range of outcomes) increases modestly during the historical period, and then more rapidly from the 2020s onwards. The growth in spread between the low and high percentiles demonstrates considerable uncertainty in the projections. At the upper end of the pdfs, the return level associated with the 95th percentile eventually reaches ~50°C, consistent with simulated occurrences of such events in some of the underlying climate model simulations (Fig. 6). At the lower end, the value of the 10th percentile increases only a little during the 21st century, while that associated with the 5th percentile reduces slightly.

For Rx1day in summer, future median values of the 50-year return level remain close to the baseline. This contrasts with projected changes for seasonal average precipitation, for which the median shows a significant decline across SE England (Figs. 2.7 and 5.4d of Murphy et al., 2018). Such differences are typical of many climate model projections, in which the heaviest events are more likely than average precipitation to show increases (e.g. Pall et al., 2007). The heaviest events tend to occur when most of the moisture in a volume of air is precipitated out. Therefore, these events are strongly influenced by future increases in the availability of moisture, which increases with temperature at 6-7% per °C via the Clausius-Clapeyron relationship. Changes in seasonal average precipitation may be subject to a different balance of influences. In UKCP Probabilistic, for example, the summer North Atlantic Oscillation appears to play an important role, through future shifts towards its positive phase in most ESPPE members (Murphy et al., 2018). This contributes to a balance of probabilities in favour of drying in the pdfs of future seasonal averages, in contrast to the pdfs for the 50-year daily return level.
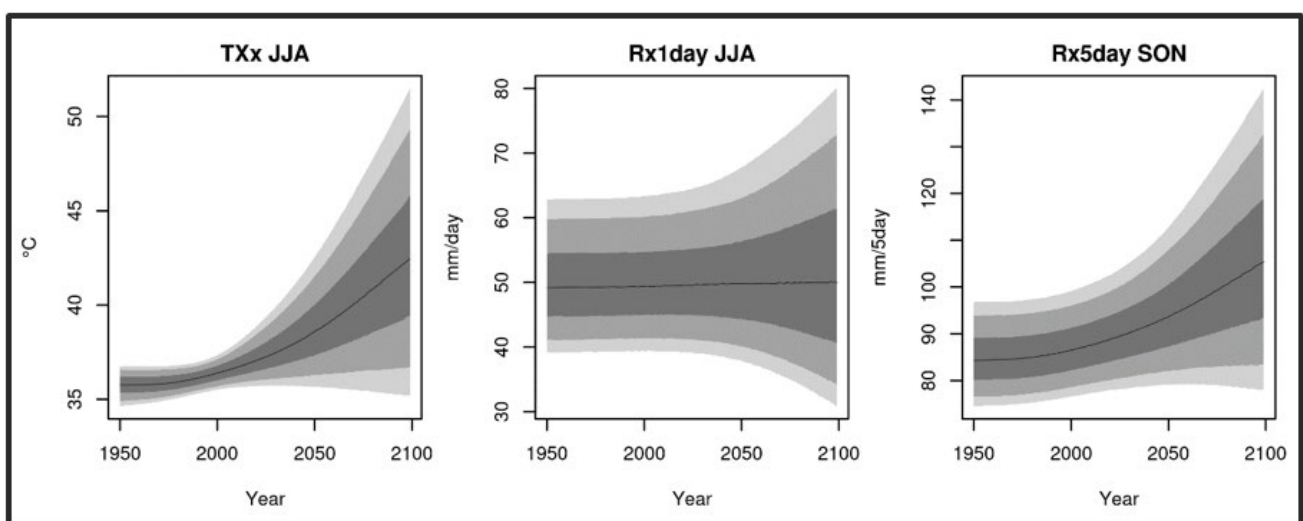

Figure 14. Probabilistic projections of the 50-year return level for the 25km grid box containing central London. Results are shown for TXx (°C) and Rx1day (mm) in summer, and Rx5day (mm) in autumn, in response to historical changes in radiative forcing to 2005, and to forcing from the RCP8.5 scenario for 2006-2100. Black curves show the medians of the pdfs, and the shading shows the 5th, 10th, 25th, 75th, 90th and 95th percentiles.

However, the pdfs of future return level also reveal substantial uncertainties. The 10th percentile shows decreases beyond 2020 (reducing by ~7mm by 2100), whilst the 90th percentile shows substantial increases, increasing by ~10mm by 2100. The Rx1day (and Rx5day) pdfs also show significant uncertainties during the historical period (Fig. 14). The historical uncertainties arise mainly from uncertainty in the baseline values of the EV parameters (section 2.2), with a minor additional contribution arising from the spread in historical changes in GMST (Fig. 7). The historical pdfs of summer TXx also show a spread, though this is smaller in comparison with the future pdfs. The historical spread of TXx is slightly larger prior to 1980 than during 1981-2000. This occurs because the latter is our baseline period for definition of GMST anomalies (section 2.3.1 and Fig. 7). By construction, this reduces the influence of anomalies in GMST during 1981-2000.

Median values of Rx5day in autumn (Fig. 14, right panel) are projected to increase monotonically through the 21st century. During 1981-2000, the median estimate lies close to 85mm. By 2100, the pdfs of absolute return level show a greater than 75% chance of exceeding this value. Whilst the return level associated with the 10th percentile increases only slightly by 2100, that of the 90th percentile increases by more than 30mm.

For London, Fig. 15 shows the impact of alternative emissions scenarios on the 50-year return level for TXx and Rx1day in summer. Beyond 2050, substantial differences develop between projections for the scenarios with the lowest and highest increases in anthropogenic forcing (RCP2.6 and RCP8.5 respectively, where the 2.6 and 8.5 indicate the approximate forcing in Wm^{-2} by 2100 relative to pre-industrial conditions).

For TXx, the strong mitigation measures assumed in RCP2.6 (Moss et al., 2010) restrict future increases in the median return level to about 2°C, and the pronounced upper tail seen beyond 2050 in the RCP8.5 results is absent. This is emphasised by Fig. 15 (top right), that compares the pdfs for 2099. For example, the upper tail of the RCP8.5 distribution (red curve) includes values ~10°C higher than the largest values in RCP2.6 (blue curve, which shows a sharp peak at ~38°C and a relatively narrow interquartile range). The intermediate RCP4.5 (green) and RCP6.0 (black) scenarios show a significantly higher chance than RCP2.6 of seeing return levels exceeding 40°C, however RCP4.5 and RCP6.0 results show very little chance of a value exceeding 50°C, compared with the probability of ~5% found for RCP8.5.

By 2099, the 5th percentile of the TXx distribution in RCP8.5 is slightly lower than at 2000, and similar to the value in RCP2.6, despite the larger GMST increases in RCP8.5 (Fig. 7). This occurs because the pdf of β_T for London, while centred around a value close to zero (Fig. 13), does sample substantial negative values (due to the influence of the GCM-scale results for southern England (Fig. 8)). When coupled with a small positive or negative value sampled from the pdf of μ_T (Fig. 8), this can produce outcomes in which return level reduces with increasing GMST. Note, however, that the low end of the μ_T pdf carries low confidence compared with the bulk of the distribution (section 2.3.3), so the low ends of the TXx return level distributions in Fig. 15 should be regarded as conservative estimates.
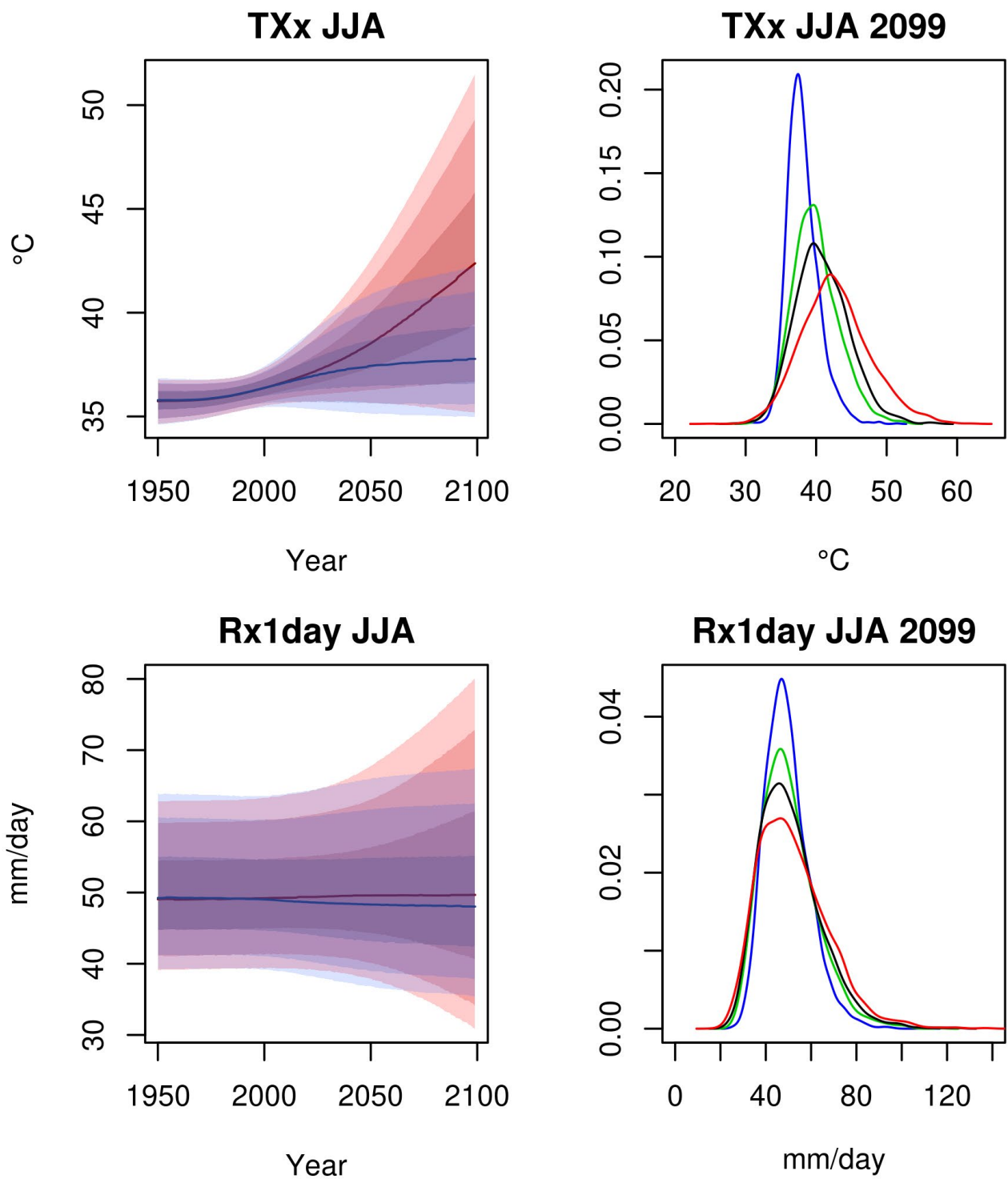
Figure 15. Probabilistic projections of 50-year return levels for TXx (°C) and Rx1day (mm) for the 25km grid box containing central London. Left panels show time-dependent projections, where red and blue shading denote the RCP2.6 and RCP8.5 scenarios respectively (noting that results are identical during the historical forcing period to 2006). Shading denotes the 5th, 10th, 25th, 75th, 90th and 95th percentiles of the pdfs. Right panels show pdfs for 2099, for the RCP scenarios (2.6 blue, 4.5 green, 6.0 black, 8.5 red).

For Rx1day, the 50-year return level pdfs change only modestly during the 21st century, under RCP2.6 emissions (Fig. 15). There are small reductions and increases in the values associated with the 10th and 90th percentiles respectively. This contrasts with the pronounced growth in the range of values under RCP8.5 emissions, beyond ~2050. The posterior pdfs of μ_T and β_T (shown at GCM scale in Fig. 9) both cover a range of positive and negative outcomes. When scaled by the larger increases in GMST found in RCP8.5 (Fig. 7), these drive a larger inflation of uncertainties in future return level. By 2099 (Fig. 15, bottom right panel), 50-year return level values below 30mm, or above 70mm, have very low probabilities for the RCP2.6 scenario, whereas the RCP8.5 results suggest a chance of about 10% for lower or higher values, respectively. The pdfs for RCP4.5 and RCP6.0 give probabilities similar to RCP8.5 for outcomes <30mm, and intermediate between RCP2.6 and RCP8.5 for values >70mm.

Figure 16 compares projections for the 20-, 50- and 100-year return periods, for TXx and Rx1day for London under RCP8.5 emissions. For TXx, the results are qualitatively similar, in the sense that increases in the rate of warming for the median return level, and substantial growth in the range of uncertainty beyond the 2020s, are common to each return period. By construction, longer return periods are associated with higher temperatures. By 2099, for example, the 95th percentile reaches ~52°C for the 100-year return level, compared with ~50°C for the 20-year level. The corresponding medians reach ~41°C and ~42°C, respectively. At the lower ends of the distributions, differences in return level are small.

The probabilistic projections of Rx1day return levels are substantially broader for the 100-year return period, compared to the 20-year values. By 2099, the 90th percentile value exceeds 90mm for the 100-year return period, cf ~60mm for the 20-year period. The 10th percentile values (~38mm and ~30mm respectively) differ less. This contrasts with the TXx results for London, mainly because the values of β_0 and ξ_0 for Rx1day are different from those for TXx, being larger for β_0 and less negative for ξ_0 (see Fig. 4 and discussion in section 2.2). Consequently, the most extreme daily precipitation events (and hence associated uncertainties in the pdfs) will increase more rapidly with rarity than the most extreme temperature events (e.g. Fig. 3).
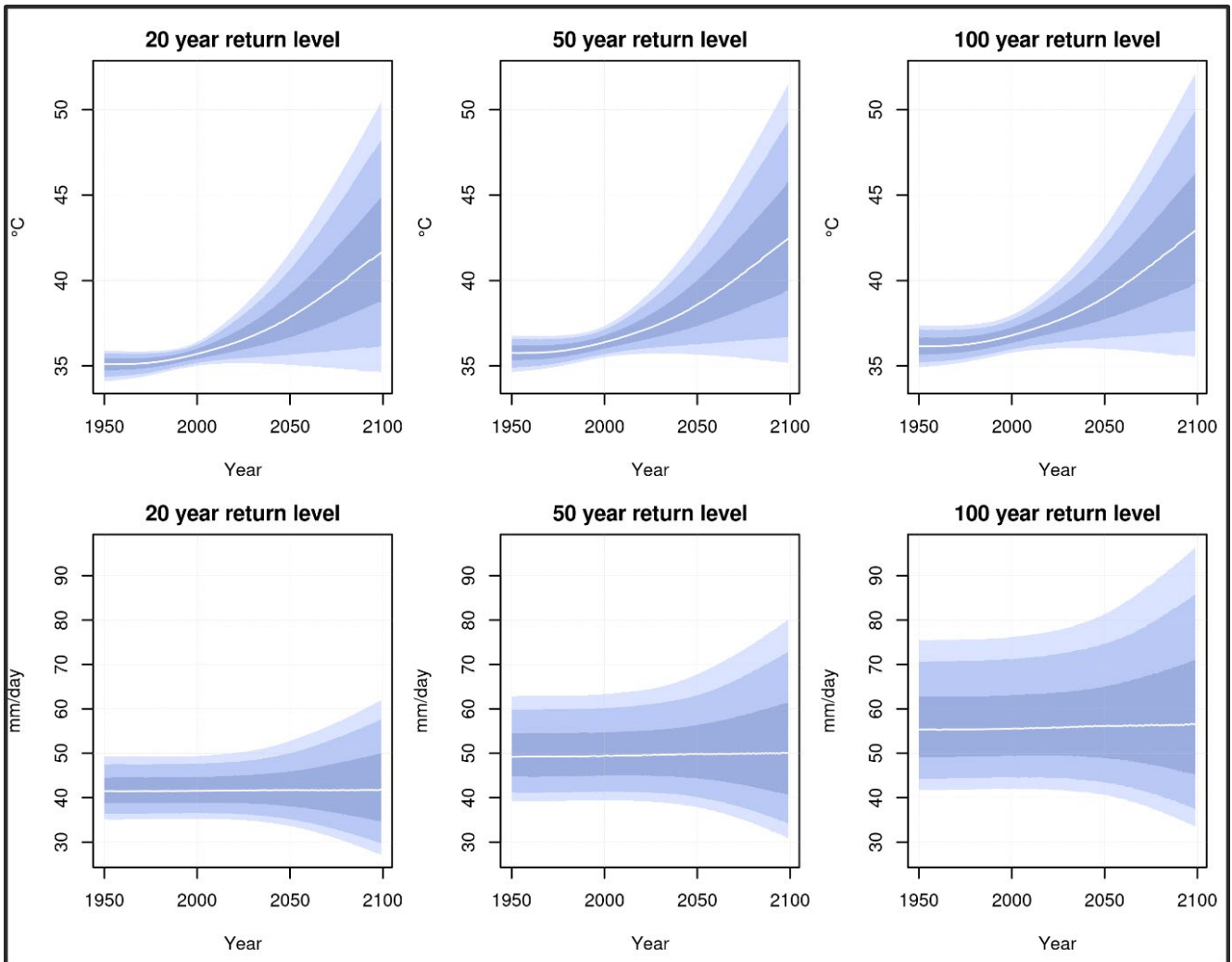
Figure 16. Comparison of probabilistic projections for 20-, 50- and 100-year return levels, for the 25km grid box containing central London. The results show projections for TXx (°C, upper panels) and Rx1day (mm, lower panels) in summer, in response to historical changes in radiative forcing to 2005, and to forcing from the RCP8.5 scenario for 2006-2100. White curves show median changes, and shading denotes 5th, 10th, 25th, 75th, 90th and 95th percentiles of the pdfs.

In future work, we recommend that the physical credibility of these rare temperature or precipitation events, particularly those associated with the upper ends of the probability distributions, should be investigated further by analysing the meteorological conditions that give rise to them in climate simulations (e.g. Fig. 6). Users of UKCP could do this, for example, by selecting appropriate case studies from the sets of Global, Regional or Local projections (e.g. Kendon et al., 2020).

Examples of the projections as spatial patterns are shown in Fig. 17. The maps show 10th, 50th and 90th percentiles of 50-year return levels for 2070, under the RCP8.5 scenario. For TXx in summer (top row), the hottest values are found in SE England, with median values in the range 38-40°C, and 90th percentile values reaching 45°C in places. For a given percentile, lower values are found to the north and west. Over northern Scotland, 90th percentile values reach 33-36°C, whereas 10th percentile values lie between 25 and 30°C. The 10th-90th percentile ranges reach 10°C in SE England and are slightly smaller elsewhere.
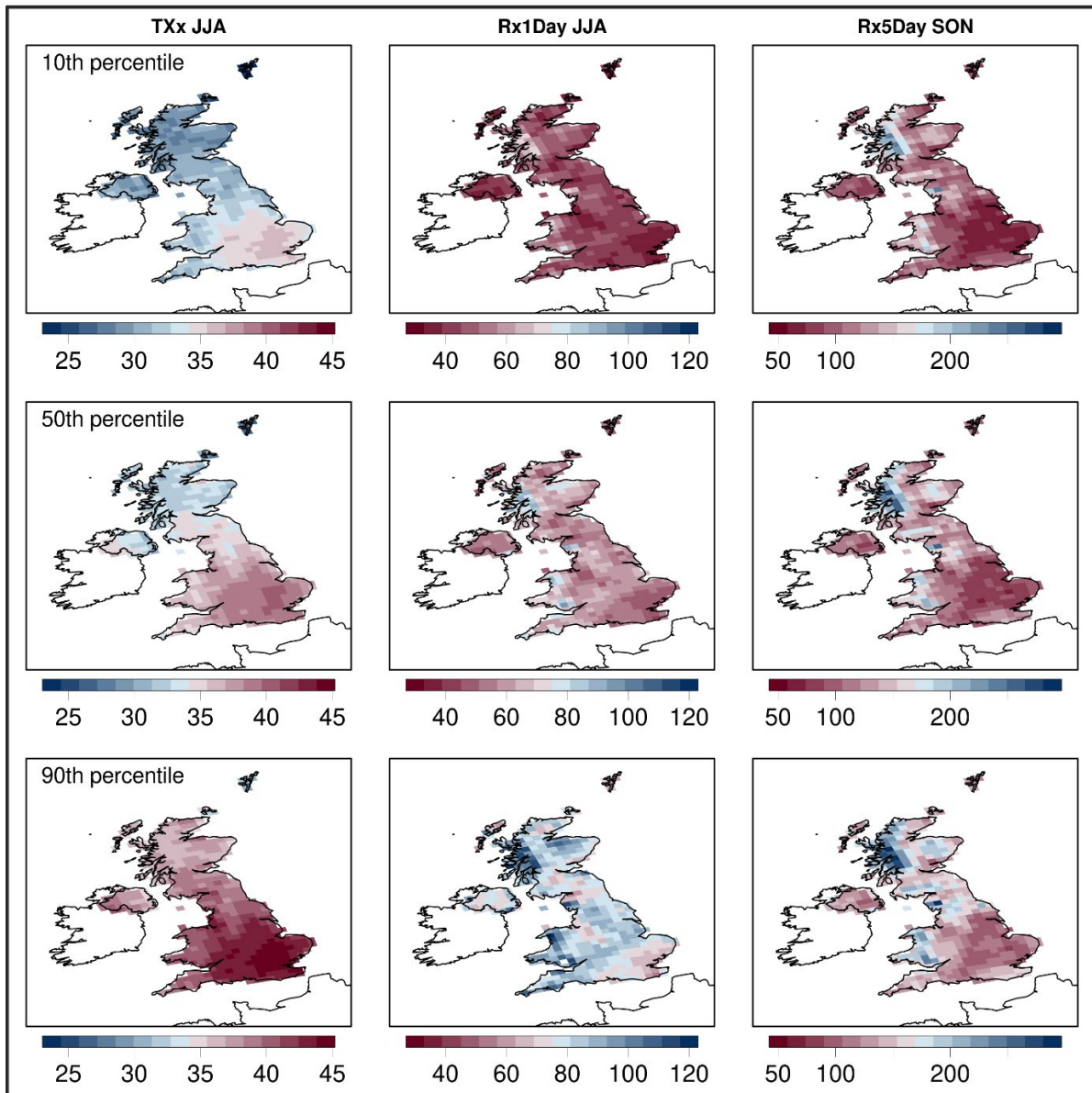
Figure 17. Maps of the 50-year return levels associated with the 10th, 50th and 90th percentiles of the pdfs in 2070, for the RCP8.5 scenario. Upper, middle and lower rows show results for TXx (°C) in summer, Rx1day (mm) in summer, and Rx5day (mm) in autumn.

The maps for Rx1day in summer show their highest values over high ground (notably over the Welsh mountains, the NW Highlands and in the Lake District), and in some coastal regions. These features reflect the influence of spatial detail in the baseline values of μ_0, β_0 and ξ_0 (Fig. 4). At the larger spatial scale of UK_GCM grid boxes, the patterns of future return levels[11] are also influenced by differing values of μ_T and β_T (Fig. 13). At the 90th percentile, future values of the 50-year return level reach 100-120mm in some high-elevation regions, with corresponding median values typically amounting to ~70mm. Many of the lowest return levels occur in Essex and East Anglia, where 10th and 90th percentile shows values of 30-35mm, and 60-75mm, respectively. Uncertainty ranges are substantial everywhere, with 10th-90th percentile ranges typically amounting to ~40mm.

---

[11] Note that the spatial detail in these patterns of return level is not expected, in general, to correspond to the detailed patterns of change in seasonal average temperature or precipitation, shown in Murphy et al. (2018). This is because: (a) the patterns of future change in Murphy et al. (2018) lack, by construction, any influence of baseline climatological values; (b) the downscaling climate change relationships used for seasonal mean variables are locally-specific, and hence contain information at the 25km scale, whereas here we use a single, UK-wide downscaling relationship at all grid points (e.g. Fig. 11). This limits the spatial influence of future climate change information to the GCM (300km) scale (section 2.3.4).

For Rx5day in autumn median values are again highest in elevated regions, locally exceeding 200mm, notably over parts of the NW Highlands. Values are considerably lower (<100mm) over parts of the Midlands and East Anglia, although these values represent increases of >10mm, relative to corresponding baseline values. The 90th percentile return levels show values exceeding 150mm over parts of Wales, northern England and Scotland.

Table 2 provides 50-year return levels for 1990 and 2070 (the latter for the RCP8.5 scenario), for each variable and season. The information consists of UK averages of local values of the 10th, 50th and 90th percentiles at 25km grid squares. This provides a simple national-scale overview of the results, while noting the underlying presence of considerable regional variations, as shown in Fig. 17. Median baseline values are highest in summer for TXx and Rx1day, and in autumn for Rx5day. Future median values show increases, for all variables and seasons. For TXx the increase is largest in summer (4.2°C), and smallest in winter (2.5°C). For Rx1day and Rx5day, the average increases in the median are small in summer (2mm and 3mm respectively), but larger in other seasons (5-10mm for Rx1day, and 9-13mm for Rx5day).

The baseline values show 10th-90th percentile ranges amounting to 1-3-2.1°C for TXx, 13-23mm for Rx1day and 17-25mm for Rx5day. The ranges for the precipitation metrics amount to 32-40% of the median for Rx1day, and 19-25% for Rx5day, due mainly to the influence of significant uncertainties in μ_0, β_0 and ξ_0 (see also Fig. 14). The future 10th-90th ranges amount to 2.8-7.9°C for TXx, 18-34mm for Rx1day and 31-46mm for Rx5day. These are significantly broader than their 1990 counterparts, due to the additional influences of uncertainty in μ_T and β_T, coupled with the growth of spread in future GMST.

Table 2. UK averages of regional 50-year return levels

Season	Year	TXx (°C)			Rx1day (mm)			Rx5day (mm)		
		10	50	90	10	50	90	10	50	90
Winter	1990	15.2	15.8	16.5	36	41	49	82	90	99
	2070	17.0	18.3	19.8	40	47	58	88	100	119
Spring	1990	25.8	26.7	27.8	34	40	48	73	82	91
	2070	27.6	30.7	34.4	36	45	56	76	91	109
Summer	1990	30.4	31.4	32.5	47	57	70	90	102	115
	2070	32.0	35.6	39.9	45	59	79	84	105	130
Autumn	1990	25.3	26.1	27.0	47	54	65	97	108	121
	2070	26.9	29.7	33.1	54	64	79	102	121	142

Average values over all UK 25km grid boxes of 50-year return levels for 1990 and 2070, for each variable and season. The 2070 numbers are projected values under the RCP8.5 scenario. Separate UK averages are provided for low (10th percentile), central (50th percentile) and high (90th percentile) regional outcomes from the relevant pdfs.

Return levels associated with the 10th and 90th percentiles all increase in future, apart from 10th percentile values for Rx1day and Rx5day in summer, which show small reductions. These low-end reductions are likely to arise from outcomes sampling changes in future circulation (such as a shift to the positive phase of the summer North Atlantic Oscillation – see discussion of Fig. 14) that are large enough to offset the effects of a warmer, moister atmosphere in setting the risk of extreme precipitation events.

4. Concluding Remarks

We have described an extension to the probabilistic projections component of UKCP (UKCP Probabilistic, Murphy et al., 2018), in which the underpinning Bayesian framework is combined with extreme value (EV) theory, using an updated version of the method of Brown et al. (2014). This supports projections of 20-, 50- and 100-year return levels of daily maximum surface air temperature (TXx), daily precipitation (Rx1day) and 5-day accumulated precipitation (Rx5day) on a 25-km national grid (Fung et al., 2018), for winter, spring, summer and autumn. These 21st century projections augment the existing probabilistic projections of monthly, seasonally and annually averaged climate variables (Murphy et al., 2018), and are available for five emissions scenarios (RCP2.6, 4.5, 6.0 and 8.5, and SRES A1B).

Compared to historical values for 1990, the projections show increases in median return levels in all seasons for all variables under RCP8.5 emissions. By 2070, the increase in median TXx for the 50year return level is 2.5°C in winter, and 3.7-4.3°C in other seasons, based on UK averages of gridded regional values (Table 2). The smallest median increases occur in summer for Rx1day and Rx5day, with larger values (5-10mm for Rx1day and 9-13mm for Rx5day) occurring in autumn, winter and spring.

In all cases, the future probability density functions (pdfs) reveal ranges that grow with time during the 21st century, reflecting the developing influence of uncertainties in modelling the climate change response. Under the RCP8.5 scenario, the UK average of the 10th percentile for summer TXx in 2070 is 32.0°C for the 50-year return level, whereas the 90th percentile average is 39.9°C. For Rx5day in autumn, corresponding averages are 102mm (10th percentile) and 142mm (90th percentile). These future differences between the 10th and 90th percentile averages, of 7.9°C and 40mm respectively, compare to corresponding differences of 2.1°C and 24mm for 1990.

The pdfs for different emissions scenarios are similar during the next two decades but show significant differences beyond 2050. In particular, return levels associated with the upper tails are larger for RCP8.5 (the scenario with the largest increase in radiative forcing, reaching ~8.5 Wm^{-2} by 2100) compared with RCP2.6 (the scenario with the smallest increase in forcing, that peaks in mid-century and reduces to ~2.6 Wm^{-2} by 2100).

These results are based on 349 global climate model (GCM) simulations and 11 regional climate model (RCM) simulations. Most (348) of these are derived from perturbed parameter ensembles (PPEs) using several configurations of the HadCM3 climate model, while twelve are drawn from a CMIP5 ensemble of earth system models. The latter represent uncertainties due to structural variations in model construction. Use of the Bayesian framework (Sexton et al., 2012; Harris et al., 2013; Murphy et al., 2018) allows the climate model results to be extended to cover unsampled parts of parameter space through the use of statistical emulators, and to be combined with a set of observational constraints derived from metrics of historical model performance. The methodology supports relatively broad assessments of known modelling uncertainties, based on the synthesis of the above lines of evidence. However, the probabilities are also conditional upon the chosen climate modelling information, the choice of constraints, and various subjective choices required to implement the methodology. The latter includes expert prior distributions for uncertain model parameters, and the design and calibration of statistical techniques needed to combine the information.

The return level projections are presented as absolute values (in °C or mm) that incorporate a bias correction. Following Brown et al. (2014), this is done by constructing future extreme value distributions that combine baseline values derived from observations with climate change effects derived from the GCM and RCM simulations. The baseline values are gridded 25km-scale values obtained from UK station data (Perry et al., 2009). This presentation as absolute values differs from earlier UKCP Probabilistic results for monthly, seasonally and annually averaged variables (Murphy et al., 2018), which were issued as anomalies relative to 1981-2000, thus leaving construction of absolute values (and associated bias correction strategies) to be handled by users on an application-specific basis.

As in Murphy et al. (2018), the return level projections are constructed from 3000 samples of potential future climate change. These are drawn from the space of parametric uncertainties in the representation of earth system processes in HadCM3, modified to account for structural modelling uncertainty. Data (available from https://ukclimateprojections-ui.metoffice.gov.uk/) is presented in the form of time-dependent pdfs and cdfs (cumulative distribution functions).

For other UKCP Probabilistic variables, we also provided results from each of the 3000 return level samples individually. Each sample member for the monthly, seasonal and annual variables includes a specific time-varying outcome for internal climate variability in a changing climate. For a given year and region, it can therefore be treated analogously to output from an individual GCM simulation. This means, for example, that users can average individual samples over a number of consecutive years to reduce the influence of internal variability and search the samples for particular examples of threshold exceedances, or multivariate events at a particular spatial location.

However, each individual sample for TXx, Rx1day or Rx5day specifies an EV distribution that, when converted into an estimate of a given return level, accounts for uncertainty arising from different phases of internal climate variability. Whilst an individual sample of (say) average summer precipitation in 2065 can be viewed as one specific realisation of a possible outcome in that year, a corresponding sample of the 20-year return level for (say) Rx1day cannot. Rather, it represents one particular expression of what the climatological distribution of extreme daily precipitation events might look like in 2065. Due to this important conceptual difference, we opted not to provide the return level projections as a set of individual realisations, as it would not be appropriate to interpret samples of return level in the same way as those of other UKCP Probabilistic variables.

Analysis that would require the underlying sampled data, such as study of joint distributions of return level variables, or assessment of ranges of change conditioned on specific global warming levels, will not therefore be possible with the return level projections. However, the full pdfs of return level can be compared with corresponding pdfs of other variables for a given region, period and emissions scenario, as the underlying Bayesian methodology is common to both datasets.

The probabilistic return level projections are suitable for assessment of uncertainties associated with the relevant extremes at specific 25km grid locations. Return levels for spatial averages of TXx, Rx1day and Rx5day over aggregated regions are not provided. However, users requiring a view on typical local values within a broader region could average (say) 10th, 50th or 90th percentile values over all constituent 25km grid boxes, as in Table 2.

The data is not suitable for analysis of joint risks at a network of spatially distributed locations, noting that this would be true even if sampled data was being provided, since the statistical processing required to produce the samples does not produce outputs with the full spatial coherence of raw climate model output (Murphy et al., 2018).

In addition to the new UKCP Probabilistic product described in this report, users can derive projections of future extremes for the UK (or other regions) from alternative UKCP products consisting of data taken directly from climate model simulations: These comprise the sets of 28 global projections (UKCP Global), 12 European RCM model projections (UKCP Regional) and 12 UK convective-permitting projections (UKCP Local). These datasets provide access to a wider range of variables with full spatial coherence, and can be used to study the spatial characteristics of future extreme events, such as relationships between return levels in different parts of the UK, or case studies of specific types of event (for example winter flooding episodes (Huntingford et al., 2014), or summer heatwaves (McCarthy et al., 2019)).

UKCP Local features the highest spatial resolution (2.2km). Its explicit representation of the dynamics of large convective storms offers new capability to predict changes in the characteristics of hourly precipitation events, and it accounts in greater detail for the influences of mountains, coastlines and urban areas (Kendon et al., 2019).

However, UKCP Global, Regional and Local offer more limited strategies for representing uncertainties, in comparison to UKCP Probabilistic (Murphy et al., 2018; Kendon et al., 2019). Therefore, these new UKCP Probabilistic results may provide useful context for studies requiring the flexibility of raw climate model output, by revealing gaps in potential climate impacts diagnosed from the latter. To this end, we encourage future comparisons of the present return level projections with corresponding results derived from UKCP Global, Regional and Local.

References

Brown SJ, Caesar J, Ferro CAT (2008). Global changes in extreme daily temperature since 1950. J. Geophys. Res. 113:D05115.

Brown SJ, Murphy JM, Sexton DMH, Harris GR (2014). Climate projections of future extreme events accounting for modelling uncertainties and historical simulation biases. Clim. Dyn. 43:2861-2705.

Burke EJ, Brown SJ (2010). Regional drought over the UK and changes in the future. J. Hydrol. 394:471-485.

CCRA2 (2017). UK Climate Change Risk Assessment, 2017. Available from https://assets.publishing. service.gov.uk/government/uploads/system/uploads/attachment_data/file/584281/uk-climate-change-risk-assess-2017.pdf

Chan SC, Kendon EJ, Fowler HJ, Blenkinsop S, Roberts NM, Ferro CAT (2014). The value of high resolution Met Office regional climate models in the simulation of multi-hourly precipitation extremes. J. Clim. 27:6155-6174.

Clark RT, Murphy JM, Brown SJ (2010). Do global warming targets limit heatwave risk? Geophys. Res. Lett. 37:L17703.

Coles SG (2001). An introduction to statistical modelling of extreme values. Springer Series in Statistics. Springer, Berlin 224.

Collins M, Booth BBB, Bhaskaran B, Harris GR, Murphy JM, Sexton DMH, Webb MJ (2011). Climate model errors, feedbacks and forcings. A comparison of perturbed physics and multi-model ensembles. Clim. Dyn. 36:1737-1766.

Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, Gao X, Gutowski W, Johns T, Krinner G, Shongwe M, Tebaldi C, Weaver A, Wehner M (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, Qin D, Plattner G-K, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley P, eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1029-1136, doi:10.1017/CBO9781107415324.023.

Folland CK, Anderson C (2002). Estimating changing extremes using empirical ranking methods. J. Clim. 15:2954-2960.

Fowler HJ, Ekström M (2010). Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes. Int. J. Climatol. 29:385-416.

Fung F (2018). UKCP18 Guidance: Bias correction. https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-guidance-data-availability-access-and-formats.pdf

Fung F, Stephens A, Wilson A (2018). UKCP18 Guidance: Data availability, access and formats. https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-guidance-data-availability-access-and-formats.pdf

Goldstein M, Rougier JC (2004). Probabilistic formulations for transferring inferences from mathematical models to physical systems. SIAM J. Sci. Comp. 26:467–487.

Gregory JM, Andrews T (2016). Variation in climate sensitivity and feedback parameters during the historical record. Geophys. Res. Lett. 43:3911-3920.

Harris GR, Collins M, Sexton DMH, Murphy JM, Booth BBB (2010). Probabilistic projections for 21st century European climate. Nat. Hazards Earth Syst. Sci. 10:2009-2020.

Harris GR, Murphy JM, Sexton DMH, Booth BBB (2021). Probabilistic projections for regional climate change accounting for Earth System modelling uncertainty. In preparation.

Humphrey K, Murphy J (2016). UK Climate Change Risk Assessment Evidence Report: Chapter 1, Introduction. Contributing authors: Harris G, Brown S, Lowe J, McCarthy M, Jevrejeva S, Watts G, Johns D, Bell M. Report prepared for the Adaptation Sub-Committee of the Committee on Climate Change, London. Available from https://www.theccc.org.uk/uk-climate-change-risk-assessment-2017/ccra-chapters/introduction/.

Huntingford C, Marsh T, Scaife A, Kendon E, Hannaford J, Kay A, Lockwood M, Prudhomme C, Reynard N, Parry S, Lowe J, Screen J, Ward H, Roberts M, Stott P, Bell V, Bailey M, Jenkins A, Legg T, Otto F, Massey M, Schaller N, Slingo J, Allen M (2014). Potential influences on the United Kingdom's floods of winter 2013/14. Nat. Clim. Change 4:769-777.

Katz RW, Parlange MB, Naveau P (2002). Statistics of extremes in hydrology. Adv. Water Res. 25:1287-1304.

Kendon EJ, Fosser G, Murphy J, Chan S, Clark R, Harris G, Lock A, Lowe J, Martin G, Pirret J, Roberts N, Sanderson M, Tucker S (2019). UKCP Convection-permitting model projections: Science report. Available from https://www.metoffice.gov.uk/research/collaboration/ukcp/guidance-science-reports

Kendon M, Sexton D, McCarthy M (2020). 20°C in the UK winter: a sign of the future ? Weather, accepted for publication.

Kharin VV, Zwiers FW (2000). Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. J. Clim. 13: 3760-3788.

Lambert FH, Collins M, Harris GR, Murphy JM, Sexton DMH, Booth BBB (2013). Interactions between uncertainties in different climate system components simulated by a fully-coupled general circulation model. Clim. Dyn. 41:3055-3072.

Lowe JA, Bernie D, Bett P, Bricheno L, Brown S, Calvert D, Clark R, Eagle K, Edwards T, Fosser G, Fung F, Gohar L, Good P, Gregory J, Harris G, Howard T, Kaye N, Kendon E, Krijnen J, Maisey P, McDonald R, McInnes R, McSweeney C, Mitchell JFB, Murphy J, Palmer M, Roberts C, Rostron J, Sexton D, Thornton H, Tinker J, Tucker S, Yamazaki K, Belcher S (2018). UKCP18 Science Overview Report. Available from https://www.metoffice.gov.uk/research/collaboration/ukcp/guidance-science-reports

McCarthy M, Christidis N, Dunstone N, Fereday D, Kay G, Klein-Tank A, Lowe J, Petch J, Scaife A, Stott P (2019). Drivers of UK summer heatwave of 2018. Weather 74: 390-396.

McColl L, Palin EJ, Thornton HE, Sexton DMH, Betts R, Mylne K (2012). Assessing the potential impact of climate change on the UK's electricity network. Clim. Change 115:821-835.

Morice CP, Kennedy JJ, Rayner NA, Jones PD (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset, J. Geophys. Res. 117:D08101.

Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, Meehl GA, Mitchell JFB, Nakicenovic N, Riahi K, Smith SJ, Stouffer RJ, Thomson AM, Weyant JP, Wilbanks TJ (2010). The next generation of scenarios for climate change research and assessment. Nature 463:747-756. doi:10.1038/nature08823.

Murphy JM, Booth BBB, Boulton CA, Clark RT, Harris GR, Lowe JA, Sexton DMH (2014). Transient climate changes in a perturbed parameter ensemble of emissions-driven earth system model simulations. Clim. Dyn. 43:2855-2885.

Murphy JM, Harris GR, Sexton DMH, Kendon EJ, Bett PE, Clark RT, Eagle KE, Fosser G, Fung F, Lowe JA, McDonald RE, McInnes RN, McSweeney CF, Mitchell JFB, Rostron JW, Thornton HE, Tucker S, Yamazaki K (2018). UKCP18 Land Projections: Science report. Available from https://www.metoffice.gov.uk/research/collaboration/ukcp/guidance-science-reports

Myhre G, Shindell D, Bréon F-M, Collins W, Fuglestvedt J, Huang J, Koch D, Lamarque J-F, Lee D, Mendoza B, Nakajima T, Robock A, Stephens G, Takemura T, Zhang H (2013). Anthropogenic and natural radiative forcing. 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 TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 659–740, doi:10.1017/CBO9781107415324.018.

Nakicenovic N, Swart R (2000). Special Report on Emissions Scenarios. Cambridge University Press.

Palin EJ, Thornton HE, Mathison CT, McCarthy RE, Clark RT, Dora J (2013). Future projections of temperature-related climate change impacts on the railway network of Great Britain. Clim. Change 120:71-93.

Perry MC, Hollis DM (2005). The generation of monthly gridded datasets for a range of climatic variables over the UK. Int. J. Climatol. 25:1041–1054.

Perry M, Hollis D, Elms M (2009). The generation of daily gridded datasets of temperature and rainfall for the UK. National Climate Information Centre, Climate Memorandum No. 24, http://www.metoffice.gov.uk/climatechange/science/monitoring/ukcp09/methods.html.

Pirard P, Vandentorren S, Pascal M, Laaidi K, Le Tetra A, Cassadou S, Ledrans M (2005). Summary of the mortality assessment of the 2003 heat wave in France. Euro-surveillance 10: 153-156.

Rajczak J, Schär C (2017). Projections of future precipitation extremes over Europe: A multimodel assessment of climate simulations. J. Geophys. Res. Atmos. 122:10773-10800.

Rummukainen M (2016). Added value in regional climate modelling. WIREs Clim. Change 7:145-159. doi: 10.1002/wcc.378.

Sexton DMH, Murphy JM, Collins M, Webb MJ (2012). Multivariate prediction using imperfect climate models part I: outline of methodology. Clim Dyn 38:2513-2542.

Sexton DMH, Murphy JM (2012). Multivariate prediction using imperfect climate models part II: robustness of methodological choices and consequences for climate sensitivity. Clim Dyn. 38:2543–2558. doi:10.1007/s00382-011-1209-8

Sexton DMH, Harris GR (2015). The importance of including variability in climate change projections used for adaptation. Nat. Clim. Change 5:931-936.

01274