Downscaling of RCM output

Emma Eastoe, Amber Leeson

December 20, 2016

Abstract

Global and regional climate models are deterministic physical models which provide predictions of,amongst other things, future temperatures under a range of climate change scenarios. Predictions are available on grids with cells ranging from XX to YY (GCM) and XX to YY (RCM). The way in which these models are fitted means that they are often good at predicting changes in the mean temperature, but less good at predicting the behaviour of unusual events. Accurate predictions of such events is of particular importance in polar regions, as they can lead to increased ice melt which, in turn, leads to sea level rises. In a recent comparison of observation data and regional climate model output in Greenland, it was observed that regional climate models underestimate the sizes of extreme temperature events, and hence will underestimate future return levels. To resolve this, we fit a statistical downscaling model, based on existing extreme value models, which shows quite marked success in improving estimates of return levels, when compared to the naive approach of using the RCM output directly. Results focus on a single location in Greenland (Summit), but consider four sets of RCM output.

1 Introduction

Many scientists agree that, over the next century, there will be considerable increases in global temperature. This will have a particular impact in polar regions, where increasing temperatures will lead to increased ice melt, leading in turn to rises in global sea levels. Changes in global temperatures have been predicted under a range of climate scenarios using global and regional climate models, with the latter giving predictions at a finer spatial resolution (XX to XX km^2) than the former (XX to XXkm^2). These changes in temperature can then be converted into changes in ice melt using **REF**. Such models are crucial as they provide the only way of being able to understand both the local and the global effects of various climate change scenarios.

Both global (GCM) and regional (RCM) climate models are deterministic physical models, which have both boundary (forcing) conditions and a large number of unknown parameters. In a regional climate model, boundary conditions are usually set by global climate model output, and parameters are estimated by calibrating the model with observational data over a historical time period. Calibration is usually to ensure good fit to some summary statistics, such as the global mean temperature. As a consequence of this, regional climate models to not always predict local behaviour well; where local could refer to both time and space, but also to parts of the temperature space. For example, in an

investigation of large temperature events in both observed time series and RCM output for a number of sites in Greenland by **AL/EE**, it was found that the RCMs do not predict well the size of unusually large temperatures.

In this study, we focus on instrumental (or observational) data taken from the Summit station in Greenland (**COORDS,ALTITUDE**). The matched RCM data was simply taken from the grid cell of the RCM which contains this station. The observational data were obtained from the **STATE SOURCE**. Four different sets of RCM output are considered. The first, which was forced by reanalysis data obtained from the **MORE INFO** ECMWF's ERA-40 (XX-XX) and ERA-Interim (XX-XX) **STATE SOURCE**. We refer to this data as the RCM reanalysis data. The remaining sets of RCM output were each forced by a different global climate model (GCM) and were obtained from **STATE SOURCE**. The forcings used were MIROC5, CanESM1 and NorESM1. These data sets are referred to respectively as MIROC5, CanESM1 and NorESM1 data. The overlapping period for observational data and RCM output at this site is 1997/11/04–2015/06/07; this will be referred to as the past period. There are some gaps in the observational record during this period, but we assume that the missing data were missing at random, and this missingness could therefore be ignored. Predictions from the MIROC5, CanESM1 and NorESM1 data sets are available until the end of the current century.

To see how well the observational and RCM data sets match, consider Figure 1, which shows scatter plots of the observational data against each of the reanalysis and MIROC5 outputs. Initial inspection of the scatter plots suggest quite a strong match between the two data sets. However, if we magnify the upper quadrant of the plot by transforming each data set to have a Laplace distribution (see Section 3 for details on how this is done), as shown in Figure 1b, it becomes clear that the highest observational temperatures do not correspond to the highest RCM outcomes, and vice versa.

To further illustrate this point, consider Figure 2. This shows plots of the return levels estimated using the top 10% of both observational and RCM data, with 95% confidence intervals. Details of return level estimation are given in Section 3.2 and of confidence interval estimation in Section 3.1; the interpretation is that these lines show the temperature that would be exceeded, on average, once every p years, where p is the return period. Figure 2a shows return periods from 5–500 years. It is clear that there is a considerable difference between the two sets of estimates. Of greatest significance if the fact that all return levels estimated using the observational data are positive, whereas done of the return levels based on the RCM output exceed $0^{\circ}C$.

Figures 1 and 2 show why, if interest is in extreme events, it might not be appropriate to simply assume that the RCM output can be used directly as a substitute for observations over a future time period for which observations are, by definition, not yet available. In this paper we propose an algorithm to downscale the extreme temperatures from an RCM. Statistical downscaling works by fitting a stochastic model to the observations and model output in an historical time period. Under the assumption that this model is appropriate in future time periods, data on the observational (or local) scale can then be predicted from future RCM output using the stochastic model.

LITERATURE REVIEW

Our downscaling approach follows Towe et al. (2016) in making much use of existing statistical models motivated by Extreme Value Theory **REF**.

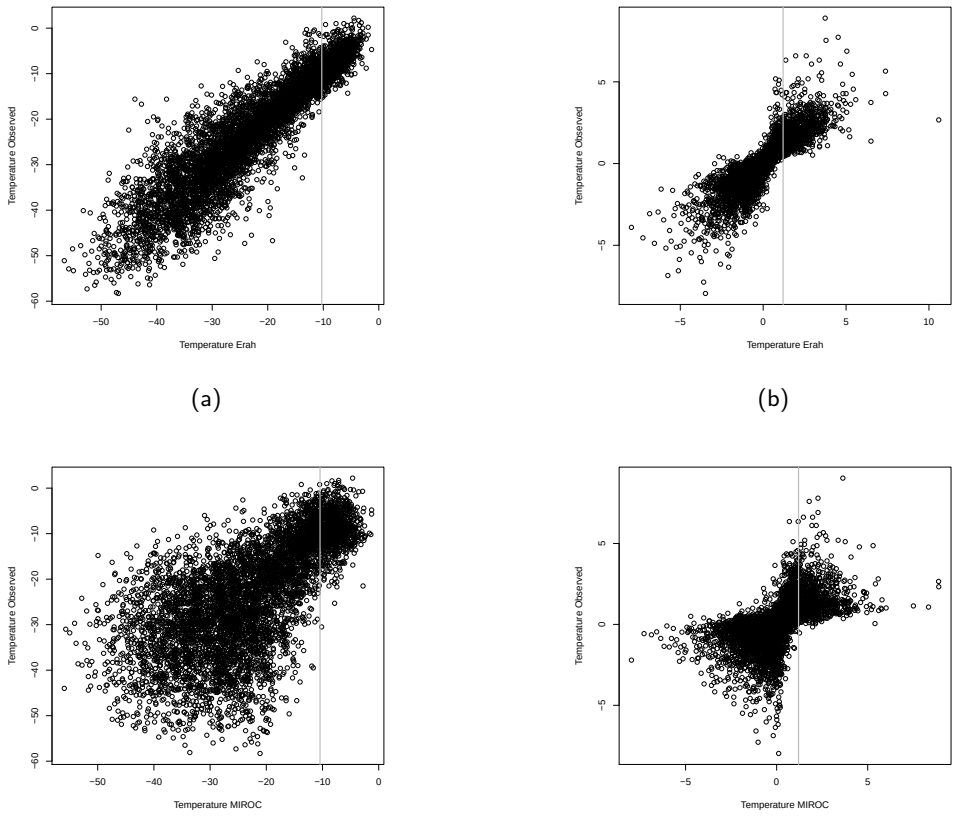
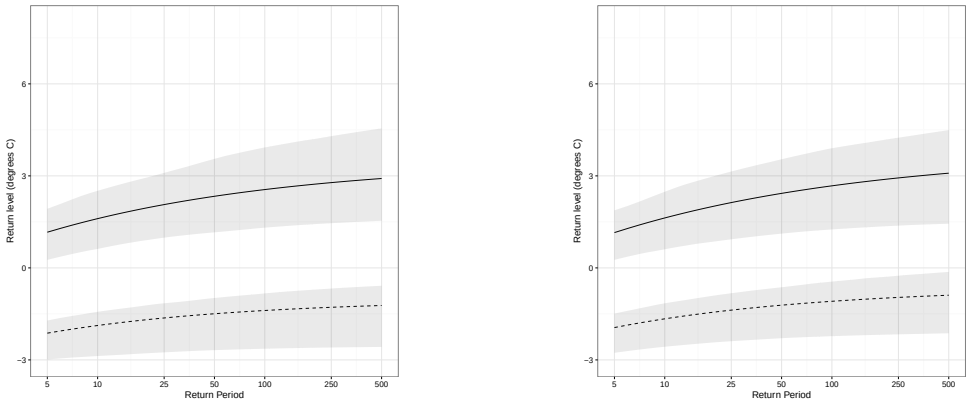


Figure 1: Data for Summit. Top: scatter plots of observed temperatures against RCM output on original (left) and Laplace (right) margins. Vertical line shows the 85% quantile of the RCM output. Bottom: same as top but observations are plotted against MIROC output.

In this paper we focus on downscaling daily maxima temperatures at a single site (Summit) in Greenland. This site is located at XXXX, and is the highest and coldest observational station on the Greenland ice sheet. Daily maxima, rather than measurements on a finer temporal resolution, was made to reduce the within-series dependence with respect, in particular, to the extreme observations.

The outline for the rest of the paper is as follows.In Section 2 we present the statistical models on which our downscaling methods are based. The downscaling algorithm is presented in Section 3. Section 4.1 presents a validation study to confirm that the downscaling model does produce sensible predictions of return levels in a scenario where we have an idea of the truth. In particular, both the RCM reanalysis and the MIROC5 data sets are split into 'historical' (1997–2007) and 'future' (2018–2015) periods. The downscaling model is fitted to the historical period and used to downscale the model output on the future period. The results of teh downscaling are then compared to the observed data in this period. We then show the results of using the downscaling model to downscale the MIROC5, CanESM1 and NorESM1 data sets in two future decades in Section 4.2. Final comments and thoughts are given in Section 5.



(a) (b)

Figure 2: Return levels estimated using a generalised Pareto distribution fitted to the top 10% of the observed (full line) and RCM output (data); RCM estimates are from (a) reanalysis and (b) MIROC output. Shaded regions represent bootstrapped 95% confidence intervals.

2 Extreme Value Methods

Extreme value analysis provides a set of statistical tools, motivated by asymptotic probability theory, to model the largest (or smallest) observations in a sample. For the downscaling algorithm presented in Section 3,we make use of both models for the tails of a univariate random variables (marginal models), and models for the joint tail of bivariate random variables (dependence model).

2.1 Margins

To model univariate marginal behaviour, we take the popular peaks over threshold approach (Davison and Smith, 1990). A high threshold is first identified. The model then consists of two parts: one to model the rate at which threshold exceedances occur and another to model the distribution of the size of these exceedances. More formally, let $X_1, \ldots, X_n$ be a sequence of independent and indetically distributed random variables with unknown marginal distribution function F, and define u to be a sufficiently high threshold. Then we can model the threshold exceedances using the following conditional distribution:

$$\hat{F}(x) = \begin{cases} \tilde{F}(x) & x \leq u; \\ 1 - \phi P(x) & x > u, \end{cases} \tag{1}$$

where $\tilde{F}$ is the empirical distribution function, $\phi = \Pr[X > u]$ and P is the survivor function of the generalised Pareto distribution

$$P(x) = \left[\max\left\{0, \left[1 + \xi\left(\frac{x-u}{\psi}\right)\right]\right\}\right]^{-1/\xi}, \quad x > u.$$

The generalised Pareto distribution has scale $\psi > 0$ and shape ξ parameters.

The key to model fitting is threshold identification; once this is done, parameter estimation is straightforward. Traditionally, diagnostics such as parameter stability plots, and the mean residual life

plot, have expoloited the threshold stability properties of the generalised Pareto distribution in order to identify the threshold (Scarrott and MacDonald, 2012). More recent work has investigated possibilities of a more automated threshold selection procedure (MacDonald et al., 2011; do Nascimento et al., 2012; Wadsworth and Tawn, 2012; Wadsworth, 2016). For the current work we adopt the traditional approach and simply opt for the lowest threshold that gives a reasonable model fit.

Conditional on knowing the threshold, the maximum likelihood estimate of the rate parameter ϕ, based on a Bernoulli model for threshold exceedances, is given by the observed proportion of threshold exceedances. Estimation of the generalised Pareto dparameters can be more tricky, as the parameters are negatively correlated. Possible ways to estimate these parameters include probability weighted moments (Hosking and Wallis, 1987), L-moments (Hosking, 1990), maximum goodness-of-fit estimators (Castillo and Hadi, 1997) and non-linear least squares (Song and Song, 2012), as well as maximum likelihood. We choose to use the latter, as this method can easily be applied to all parts of the full downscaling model.

2.2 Dependence

The conditional multivariate extremes model of Heffernan and Tawn (2004) will be used to model the joint tail behaviour of the observational data and RCM output. Although this model can be specified for a d-dimensional distribution for any $d \geq 2$, we present the bivariate case as this is the only case that we will need. Let $\mathbf{Y} = (Y_1, Y_2)$ denote a bivariate random variable with Laplace margins, as in Keef et al. (2013). Then the conditional bivariate extremes model states that, given a large Y_1, the distribution of the standardised Y_2 can be characterised as

$$\frac{Y_2 - \alpha y_1}{y_1^\beta} \Big| Y_1 = y_1 \sim \mathsf{G}(\mu, \sigma^2), \quad y_1 > u_D. \tag{2}$$

where $-1 \leq \alpha \leq 1$, $\beta \leq 1$ and $\sigma > 0$, and μ and σ^2 are mean and variance parameters for an, as yet unspecified, distribution G. The dependence threshold u_D is chosen so that $\Pr[Y_1 > u_D]$ is small. An equivalent model can be specified for $Y_1|Y_2 > u_D$. In the context of downscaling, where the aim is to predict values on the observational scale (Y_2) given future RCM predictions (Y_1), we in fact only need the conditional model for the variable on the observational scale, given that the RCM scale variable is large. More details are given in Section 3.

Inference for the model for $Y_2|Y_1 > u$ requires estimation of four parameters $\theta = (\alpha, \beta, \mu, \sigma^2)$. The estimates of these parameters are conditional on the threshold u_D, which, as in the marginal model, must be pre-specified. In order to use maximum likelihood inference to estimate the parameter vector θ, a distribution for G must be specified. Heffernan and Tawn (2004) suggest using the Gaussian distribution. Given observations n observations $(y_{1,1}, y_{2,1}), \ldots, (y_{1,n}, y_{2,n})$ with $y_{1,i} > u_D$ for $i = 1, \ldots, n$, this leads to the likelihood function

$$l(\theta) = \prod_{i=1}^{n} \frac{1}{\sigma y_{1,i}^\beta \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{y_{2,i} - \alpha y_{1,i} - \mu y_{1,i}^\beta}{\sigma y_{1,i}^\beta} \right)^2 \right\}.$$

However, for many of the examples in this paper, we found that the Gaussian distribution may be a poor model for the residuals

$$Z_{2,i} = \frac{Y_{2,i} - \alpha y_{1,i}}{\sigma y_{1,i}^\beta}.$$

The empirical distribution of the $\{z_{2,i}\}$ turned out to be both heavier tailed than a Gaussian, and more skewed; see Section 4.1 for more details. We therefore propose that the Gaussian distribution is replaced by a more flexible distribution which still has support on $(-\infty, \infty)$, for example a skewed Normal, logistic, Laplace, skewed t, or Gumbel. Choice of this distribution can be made by considering appropriate quantile-quantile and probability-probability plots for the residuals, by comparing the AIC for the different models, or by calculation of the Anderson-Darling statistic. We see in Section 4.1 and 4.2, that this choice of residual distribution does in fact have a non-negligable impact on the return levels predicted under the downscaling model. In practice, we find that a compromise has to be made between goodness of fit and ease of parameter estimation.

3 Downscaling Algorithm

Statistical downscaling methods use stochastic models to downscale Global or Regional Climate model output to approximate the behaviour of a physical variable, e.g. temperature, at a point location. Typically this is done by fitting a statistical model which links the RCM (GCM) output and observational data taken from the location of interest, over some historical period for which observational data is available. The model is then used to downscale output from the RCM (GCM) in the future period. The major assumption required is that the model relating the RCM (GCM) output and observational data remains unchanged when moving from the historical to the future period.

A number of possible models have been suggested for statistical downscaling. Many of these focusing on downscaling the mean behaviour of the climate variable. Given our interest is specifically in downscaling the extremes, we list only methods which have been developed specifically in this context. These include the CDF- (Michelangeli et al., 2009) and xCDF-transfer (Kallache et al., 2011), regression (Wilby and Wigley, 2000; Mannshardt-Shamseldin et al., 2010) and censored quantile regression (Friederichs and Hense, 2007) methods, and neural networks (Dibike and Coulibaly, 2006). There are also a number of other approaches which were not developed specifically for downscaling extreme events, but have been applied in this area; see Bürger et al. (2012) for details of some of these.

Rather than taking a regression or transfer function type of approach, we propose an algorithm based on a statistical model for mutivariate random variables, in particular on the conditional bivariate extremes model (Heffernan and Tawn, 2004). This work follows on from a similar approach taken by Towe et al. (2016). The algorithm has two parts: model fitting and downscaling. Model fitting uses only the observational data and RCM output for the past period. It comprises two stages: a model for each of the margins (observational and RCM) and a model for their dependence.

We first require some notation. Let Y_i denote the climate variable at the observational scale, with $i = 1$ corresponding to the past and $i = 2$ corresponding to the future periods. Let X_i denote the climate variable at the RCM scale, with $i = 1$ corresponding to the past and $i = 2$ corresponding to the future.

For the marginal model the distribution in equation (1) is fitted separately to both observational data and RCM output. These fitted models are used to transform to Laplace margins via,

$$Y_{L,1} = H^{-1}(\hat{F}_{Y_1}(Y_1)) \quad \text{and} \quad X_{L,1} = H^{-1}(\hat{F}_{X_1}(X_1)) \tag{3}$$

where H is the inverse cdf of the Laplace distribution.

For the dependence part, a dependence threshold is specified for $X_{L,1}$ and the bivariate conditional extremes model specified in equation (2) is fitted to all observations for which $X_{L,1}$ exceeds this threshold. An additional component, to model the sub-extremal dependence, was added following initial results from the validation study (see Section 4.1). To modelthe sub-extremal dependence, a second dependence threshold, which is lower than the first, is specified, and a linear regression model is fitted to all data for which $X_{L,1}$ lies between these two thresholds. If, as we found, the residuals of this model are highly non-Gaussian, a heavy-tailed model such as that given in equation (1) can then be fitted to the residuals of this regression. We found that inclusion of this sub-extremal component helped give more realistic downscaled values.

Because the dependence structure is modelled on Laplace margins, in order to downscale we must first transform the RCM output for the future period to Laplace margins. For physical processes, such as temperature, the support of the generalised Pareto distribution is often found to be bounded above. Therefore using the marginal model fitted to the past RCM output does not make sense; frequently there will be high future values y for which $\Pr[Y_1 \leq y] = 1$ and consequently all such values will be transformed to infinity on the Laplace scale. To resolve this, model (1) is fitted to the future RCM output and this new marginal model is used to transform the future RCM output to Laplace margins, as in equation (3):

$$X_{L,2} = H^{-1}(\hat{F}_{X_2}(X_2)).$$

Using the fits of the bivariate conditional extremes model, and the sub-extremal model if required, values for $Y_{L,2}$ are then simulated conditional on $X_{L,2}$ being large. Finally we need a way to transform the downscaled observational scale data $Y_{L,2}$ back to the original margins. To do this we suggest the transform:

$$Y_2 = L^{-1}(F_{Y_1}^{-1}(H(y_{L,2}))). \tag{4}$$

The $F_{Y_1}^{-1}(H(y_{L,2}))$ component takes $Y_{2,L}$ to the same marginal ddistribution as the past observation scale data. The function L is then used to shift to an appropriate future marginal distribution; the choice of L is motivated by consideration of the CDF-transfer function downscaling methods.

The CDF-transfer function method of Michelangeli et al. (2009) assumes that there exists a transformation T which allows movement between the cdfs of the climate variable observational past and RCM past scales,

$$F_{Y_1}(y) = T(F_{X_1}(y)).$$

Setting $z = F_{Y_1}(y)$ and rearranging, the required transformation function is

$$T(z) = F_{Y_1}(F_{X_1}^{-1}(z)).$$

It is further assumed that the same relationship holds in the future, so that the cdf for Y_2 can be written as a function of the cdfs for Y_1, X_1 and X_2,

$$F_{Y_2}(y) = T(F_{X_2}(y)) = F_{Y_1}(F_{X_1}^{-1}(F_{X_2}(x)). \tag{5}$$

As pointed out in Towe et al. (2016), a major difficulty with this approach is that the support for the cdfs may be bounded. If these bounds differ across the variables X_1, X_2, Y_1 and Y_2, downscaling via

equation (5) is impossible. The colution proposed by Kallache et al. (2011) is to re-scale the data prior to modelling. Instead we follow Towe et al. (2016) who suggest relating not the cdfs, but the quantiles of the variables. Further they propose that this relationship should be between variables at the same scale, i.e. that there exists a transform L such that

$$X_2 \sim L(X_1) \quad \text{and} \quad Y_2 \sim L(Y_1).$$

Then L, which is the same for both the local and global scale variable, ensures that the distribution of the climate variable in the future is the same as the distribution of the transformed climate variable in the future.

A form for L can be obtained as follows. Consider X_1, then

$$F_{X_1}(x) = \Pr[X_1 \leq x] = \Pr[L(X_2) \leq x] = \Pr[X_2 \leq L^{-1}(x)] = F_{X_2}(L^{-1}(x)), \tag{6}$$

which, on rearranging, gives

$$L(z) = F_{X_2}^{-1}(F_{X_1}(z)). \tag{7}$$

Unfortunately, this still suffers from problems when the bounds on the support of the cdfs for X_1, X_2 and Y_1 differ. Towe et al. (2016) suggest that instead, the transformation L, could simply be modelled using *(a)* a linear or exponential relationship for the highest quantiles and *(b)* the empirical relationship elsewhere, with the relationship described in part *(a)* identified by careful examination of a quantile-quantile plot of the highest quantiles of X_1 and X_2. An example will be shown in Section 4.1. The estimated transfer function L can then be applied to equation (4) to take the downscaled observational scale data from the past marginal scale to the future one.

3.1 Inference and uncertainty

As well as point estimates of model parameters we would also like a measure of our uncertainty in them. Our inference will be under the maximum likelihood framework, but because the model has four 'stages', a different likelihood function will be needed for each stage. The first stage is to fit three GP models, one each to the upper tails of the past (observational), past (RCM) and future (RCM) data sets. These models are used to transform the data to have Laplace margins. Given this marginal transformation, the second stage is to fit the conditional bivariate extremes model to the past observational and past RCM data sets. The third (optional) step is to estimate the sub-asymptotic model for this copula. This inself has two sub-steps: firstly a linear model fit and secondly fits of GP tail models to the upper and lower tails of the resulting residuals. The fourth, and final, step is to estimate the relationship L between the upper quantiles of the past (observational) and past (RCM) data sets.

To quantify the uncertainty which accumulates over these model fits, we use a non-parametric block bootstrap algorithm. Blocking is used to preserve the short-scale time dependence. To create a single bootstrap sample, the data were placed in non-overlapping blocks of length X. These blocks were indexed by their position within the year in which they occured. For a given year and within-year index number, the block was resampled, with replacement, across all blocks with that same within-year index. Sampling only across identical within-year index helps to retain the seasonal structure. This was repeated across all blocks in the year, and across all years. The entire bootstrap process was then repeated 200 times, to obtain a reasonable approximation to the sampling distribution.

3.2 Return level estimation

To compare the performance of the various downscaled data sets, both with each other and with the naive option of simply using the RCM output directly, we estimate return levels. These are estimated in the same way for each of the observational data, RCM output and downscaled data sets. For each data set, a generalised Pareto distribution is first fitted to the top 100q% of the data, where $0 \leq q \leq 1$ is relatively small, so that the model is only fitted to the very largest temperatures. The N-year return level is then estimated as

$$y_N = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[q^{\hat{\xi}} \left(1 - \frac{1}{365N} \right)^{-\hat{\xi}} - 1 \right]$$

where $(\hat{\sigma}, \hat{\xi})$ are the maximum likelihood estimates of the generalised Pareto parameters.

4 Results

4.1 Validation Study

We start with a validation study of the reanalysis and MIROC5 data sets. The downscaling model is fitted to the first ten years of each of the data sets and it is then used to downscale the last seven years of model output to the observational scale. The first ten years and last seven years are referred to as the past and future period respectively. Note that downscaling can occur within the modelling, or past, period, and the downscaled data used to assess model fit. In the validation study, model fit was assessed in exactly this way, by comparing return levels estimated from the past downscaled RCM, to return levels estimated from a generalised Pareto distribution fitted directly to the past observational data. To assess prediction, the return levels estimated from the downscaled future RCM output were compared with return levels estimated using the observational data from the downscaled period.

Throughout, a 90% quantile was used to fit the marginal generalised Pareto distributions and an 85% quantile was used for the dependence threshold. Where applicable, a 70% threshold was used to identify the sub-asymptotic model component. Five models for the residuals in the dependence model were considered: normal, skewed normal, Gumbel, skewed t and generalised extreme value (GEV). To obtain return levels, the generalised Pareto distribution was fitted to the top 10% of the data (observational, RCM output or downscaled). To mitigate against simulation uncertainty in the downscaling, 200 downscaled data sets were simulated, return levels were calculated for each data set and the median return level then taken as the final point estimate.

We start by considering the reanalysis data. Quantile-quantile plots (not shown) were used to confirm the fit of the generalised Pareto distribution to each of the margins, prior to transformation to Laplace margins. It is important to note that the shape parameter estimates for each data set (observational past,RCM past and RCM future) were negative, implying a finite upper end-point in the support of each of the distributions. Further, these end-points differed across the different data sets. This has implication when estimating the transformation L as discussed in the previous section.

Turning now to the dependence model, Figure 3 shows standardised probability-probability plots of the residuals (3), for each of the five possible models. The standardised plot is created by plotting

the difference between the empirical and model-based distribution functions against the rank of the residuals. From these plots, it is clear that the Gumbel, skewed t and GEV models perform better than the normal and skew normal distributions. The Anderson-Darling statistics for these three distributions are 18.4 (2 parameters), 0.26 (4 parameters) and 1.43 (3 parameters) respectively, suggesting that the skewed t or GEV is likely to be the best fit.

Finally, we must estimate the transformation L which, as described in Section 3, is used to transform the downscaled data to an appropriate future marginal distribution. Due to the differing ranges of support implied by the marginal generalised Pareto fits, we find a form for this relationship empirically, rather than using the form derived in equation (**??**). Figure 4a shows the quantile-quantile plot of the observational data against the reanalysis output. Figure 4b magnifies the upper end of this plot, by only plotting data points for which the reanalysis component is in the top 10% of the reanalysis output. From this, it is clear that a linear model is appropriate to describe the relationship between the quantiles in the two data sets, but that the coefficients in this relationship depend on whether the model is fitted to all quantiles, or only to the largest ones.

Figures 5a and 5b show estimates of the 5-500 year return levels for the current period as estimated using (i) the observational data, (ii) the reanalysis output and (iii) the downscaled data. For the downscaled output we show estimates for each of the five different residual distributions, both with and without the sub-asymptotic model. All of the downscaling results look much closer to the estimates from the observational data than to the estimates from the reanalysis data. However, the plots do show that the choice of residual model, and also whether or not a sub-asymptotic component is used, makes a difference to the downscaled return levels. The Gumbel, skewed t and GEV residual models all result in downscaled return levels that are closer to the observational ones, than those produced using the Normal and skewed Normal models. Excluding the sub-asymptotic component results in some under-estimation of the return levels, relative to those obtained from the observational data, with the gap between the observational estimates and the downscaled estimates widening as the return period increases. For this reason, Figure 5c shows only downscaled return levels for the future period that were obtained from the model with a sub-asymptotic component. Looking now at these estimates, the Gumbel and skewed t residual models seem to produce the closest return level estimates.

Next, consider the 95% boostrapped confidence intervals shown in Figure 6. We show results for both current and validation periods, but only for Normal, Gumbel and GEV residual models, and only for models with the subasymptotic component. The skewed Normal residual model performed similarly to the Normal model, and the skewed t performed similarly to the Gumbel and GEV. Bearing in mind that there is some uncertainty in the observational estimates, it is clear that all models are performing well, both in the current and validation periods. For this data set, we would propose using the model with Gumbel residuals: out of all the residual models this has performs as well as the more complex GEV and skewed t models in prediction of the return levels.

A similar exercise was carried out for the MIROC5 data, although there is not sufficient space to reproduce all the results here. For this data the Normal and skewed Normal residual models, did not perform as well as the Gumbel, skewed t and GEV residual models, and the model which did not include the sub-asymptotic component did not perform as well as the model with the asymptotic component. Figure 7 shows plots of return levels, with confidence intervals, for the current and validation periods

based on three of the residual models (Normal, skewed t and GEV).

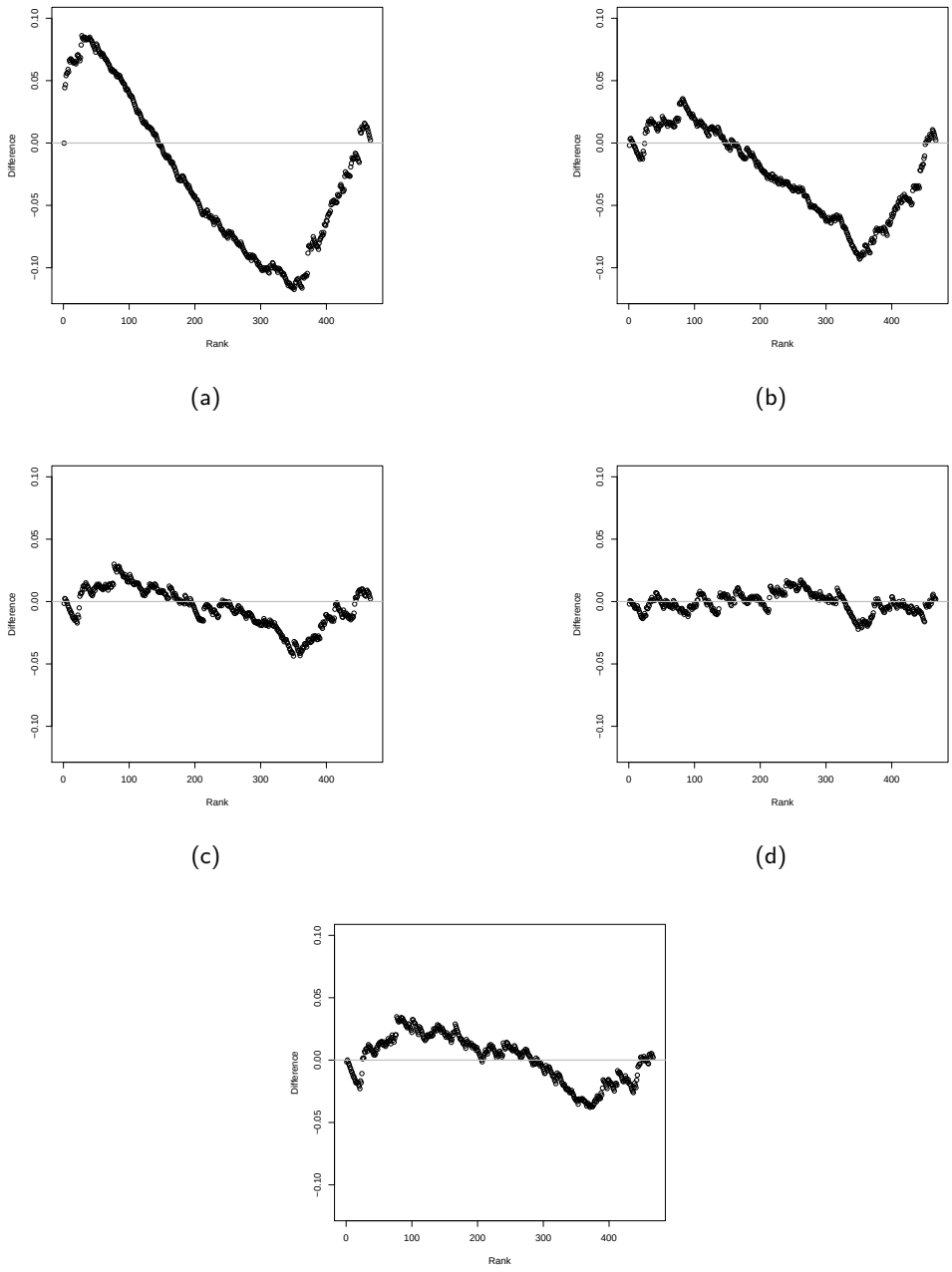
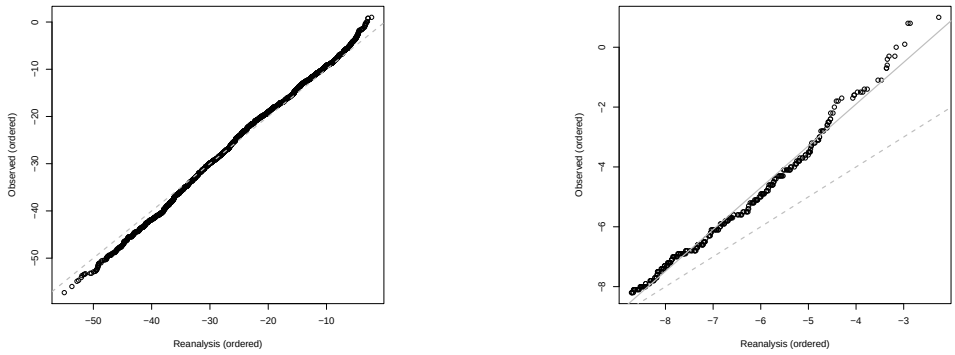
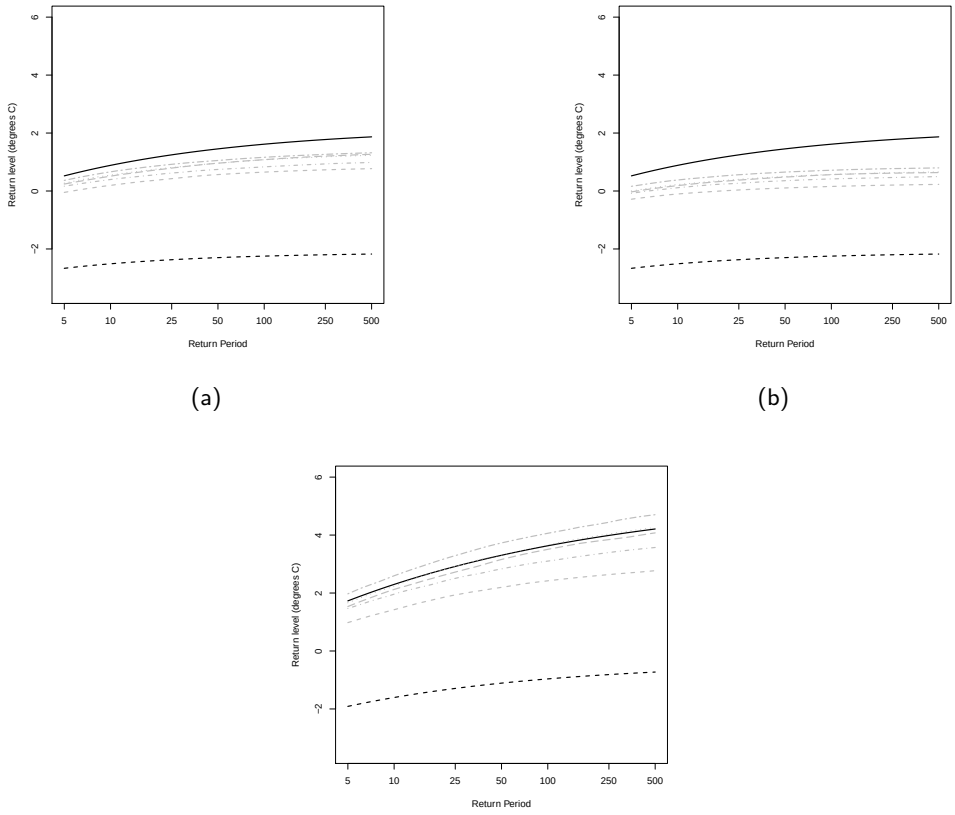


Figure 3: Standardised probability-probability plots for the five residual distributions considered for the conditional bivariate extremal dependence model.The dependence model was fitted to the first ten years of the reanalysis data as part of the validation study.



(a) (b)

Figure 4: Quantile-quantile plots of observational data vs. reanalysis output for the current period in the validation study, for (a) all data and (b) data with reanalysis component in the top 10%. Dashed gray lines show $y = x$; in (b) full gray line shows the linear model fitted to the data shown in this plot.



(a) (b)

(c)

Figure 5: Return levels plotted against return period (in years, on the log scale) for downscaled reanalysis data for the current (a-b) and validation (c) periods. Results from downscaling with (a and c) and without (b) the sub-asymptotic component is also shown. Full (dashed) black lines show estimates made using the observational (reanalysis) data in the relevant period. Gray lines shows downscaling using the Normal (dashed), Gumbel (dotted), skewed Normal (dash dot), skewed t (long dash) and GEV (log dash-dot) residual models.

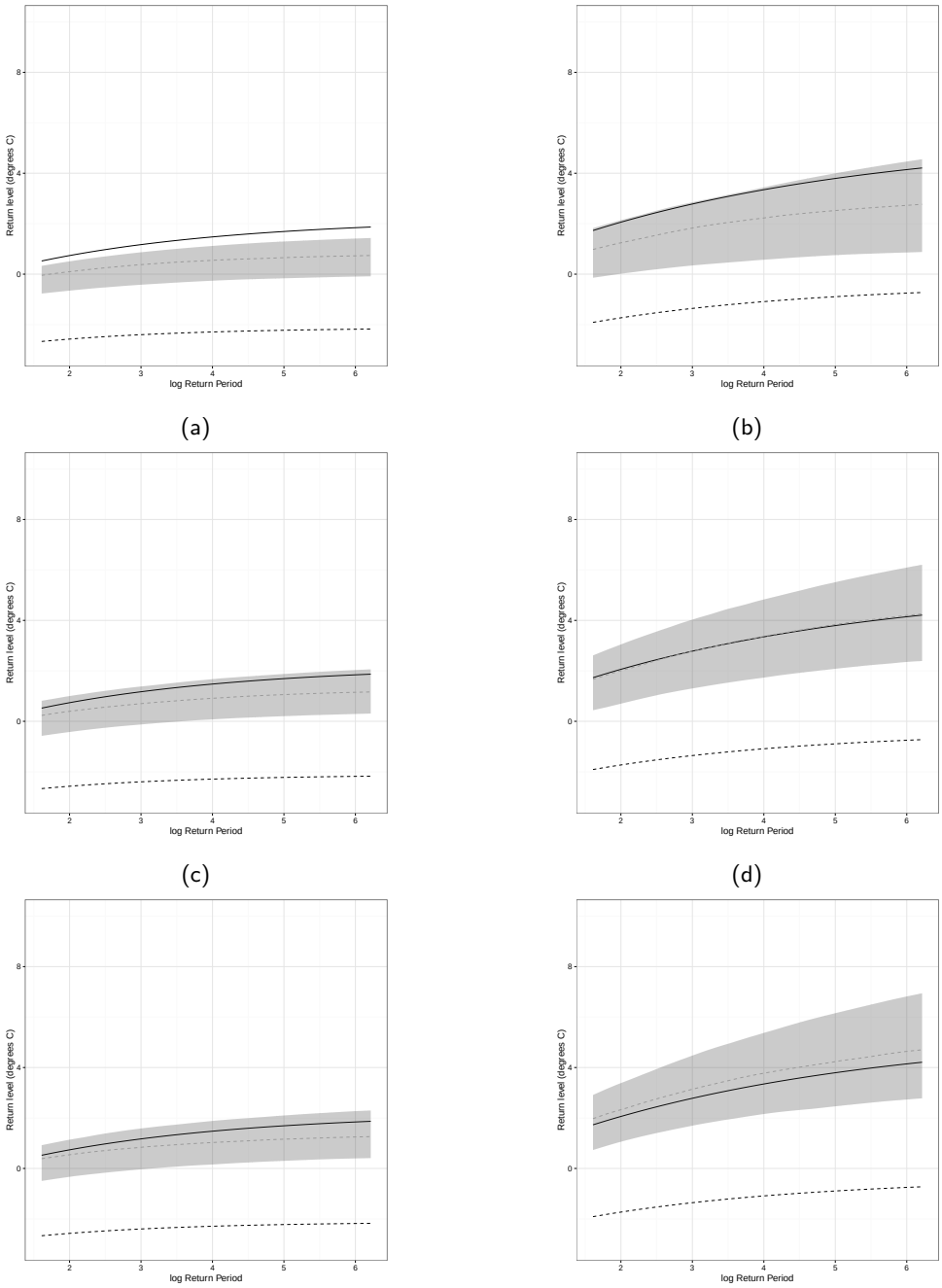


Figure 6: Return levels, as in Figure 5, but with bootstrapped 95% confidence intervals for downscaled reanalysis data for the current (a,c,e) and validation (b,d,f) periods. All downscaling models include the sub-asymptotic component. Residual models used include Normal (a-b), Gumbel (c-d) and GEV (e-f). Full (dashed) black lines show estimates made using the observational (reanalysis) data in the relevant period. Dashed gray lines shows downscaled estimates and shaded regions show confidence intervals.

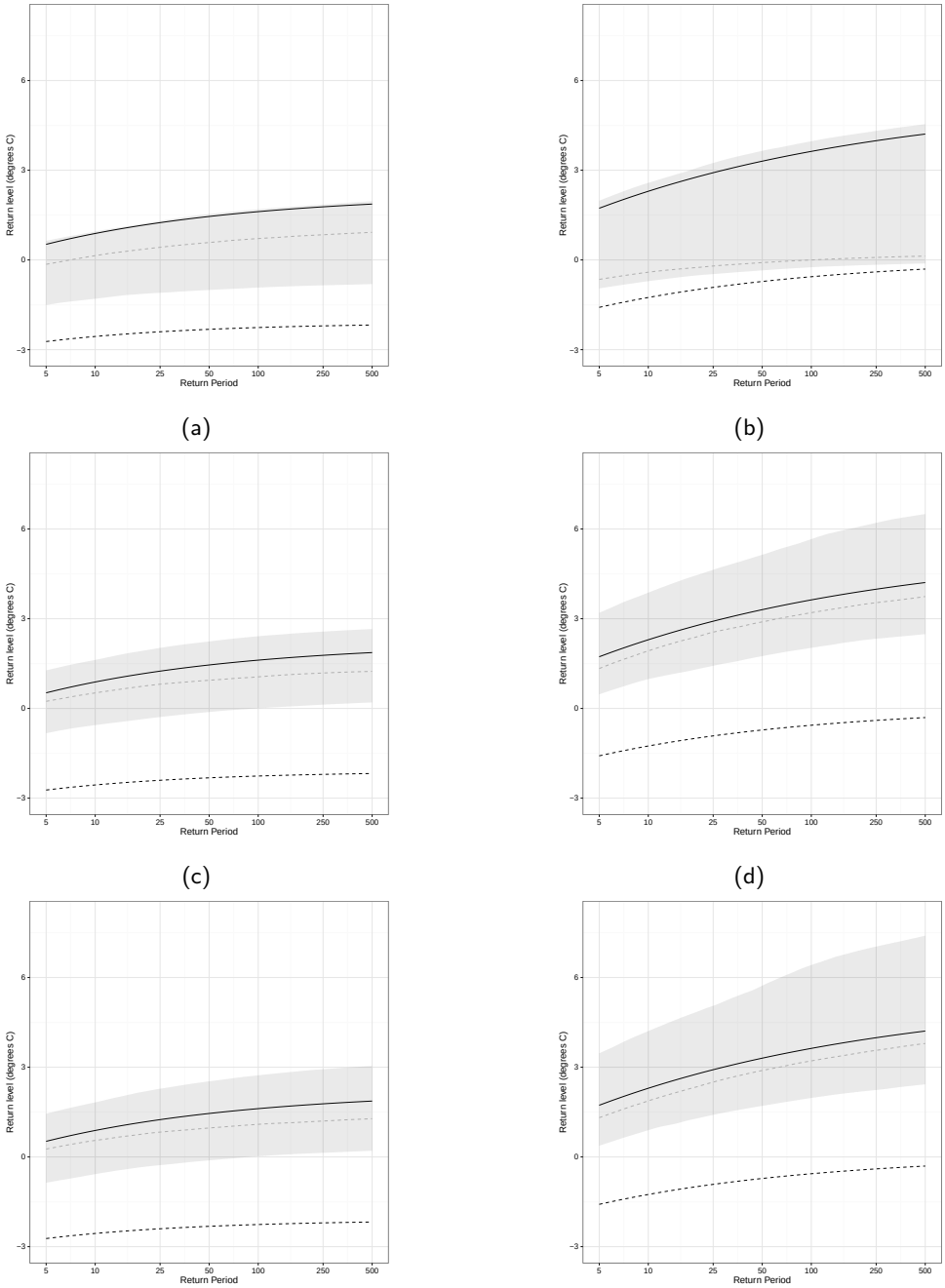


Figure 7: As Figure 5, but for downscaled MIROC5 data: return levels for the current (a,c,e) and validation (b,d,f) periods. All downscaling models include the sub-asymptotic component. Residual models used: Normal (a-b), skewed t (c-d) and GEV (e-f). Full (dashed) black lines show estimates made using the observational (reanalysis) data in the relevant period. Dashed gray lines shows downscaled estimates and shaded regions show confidence intervals.

	MIROC05					
	2020–2029			2050–2059		
Return Period	Normal	GEV	Skewed t	Normal	GEV	Skewed t
10	-0.76	0.19	0.28	2.27	3.65	3.75
	(-1.38,0.15)	(-0.67,1.38)	(-0.68,1.39)	(0.91,3.56)	(2.11,5.37)	(2.08,5.41)
50	-0.50	0.65	0.74	2.59	4.36	4.46
	(-1.14,0.53)	(-0.25,1.94)	(-0.26,1.92)	(1.17,4.03)	(2.66,6.13)	(2.64,6.07)
100	-0.44	0.79	0.88	2.67	4.56	4.67
	(-1.07,0.63)	(-0.14,2.07)	(-0.15,2.06)	(1.23,4.16)	(2.83,6.38)	(2.81,6.28)
500	-0.34	1.00	1.09	2.79	4.86	5.01
	(-0.98,0.77)	(0.01,2.35)	(0.03,2.29)	(1.31,4.34)	(3.10,6.78)	(3.08,6.67)
	CanEsM1					
	2020–2029			2050–2059		
Return Period	Normal	GEV	Skewed t	Normal	GEV	Skewed t
10	0.69	1.50	1.44	2.27	3.65	3.75
	(-0.03,1.48)	(,)	(-0.68,1.39)	(0.91,3.56)	(2.11,5.37)	(2.08,5.41)
50	1.02	1.96	1.92	2.59	4.36	4.46
	(0.26,1.97)	(,)	(-0.26,1.92)	(1.17,4.03)	(2.66,6.13)	(2.64,6.07)
100	1.11	2.08	2.04	2.67	4.56	4.67
	(0.33,2.11)	(,)	(-0.15,2.06)	(1.23,4.16)	(2.83,6.38)	(2.81,6.28)
500	1.24	2.29	2.26	2.79	4.86	5.01
	(0.43,2.34)	(,)	(0.03,2.29)	(1.31,4.34)	(3.10,6.78)	(3.08,6.67)

Table 1: Downscaled return levels ($^{\circ}C$) for MIROC5, CanEsM1 and NorESM1 RCM outputs using the downscaling model with sub-asymptotic component, with comparison between the Normal and GEV (CURRENTLY SKEWED T) models for the residuals. Figures in brackets are 95% bootstrapped confidence intervals.

4.2 Prediction

Given that results from the validation study shown in Section 4.1 suggest that the downscaling model does show good predictive performance, we now predict return levels for two future decades: 2020–2029 and 2059–2059. We do so using RCM output from each of the three different forcings: MIROC5, CanESM1 and NorESM1. Three possible models for the residuals were considered: Normal (the worst fit to the empirical residuals), skewed t and GEV. The latter perform almost identically, and so results are only shown for the Normal and skewed t. As in the validation study, the model without the sub-asymptotic component showed an inferior fit, in terms of reproducing return levels, and so all results shown are for teh model with sb-asymptotic component. Table 1 shows 10-, 50-, 100- and 500-year return levels for each decade, and each

5 Discussion

References

Bürger, G., Murdock, T., Werner, A., Sobie, S., and Cannon, A. (2012). Downscaling extremes-an intercomparison of multiple statistical methods for present climate. *Journal of Climate*, 25(12):4366–4388.

Castillo, E. and Hadi, A. S. (1997). Fitting the generalized pareto distribution to data. *Journal of the American Statistical Association*, 92(440):1609–1620.

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

Dibike, Y. B. and Coulibaly, P. (2006). Temporal neural networks for downscaling climate variability and extremes. *Neural Networks*, 19(2):135–144.

do Nascimento, F. F., Gamerman, D., and Lopes, H. F. (2012). A semiparametric bayesian approach to extreme value estimation. *Statistics and Computing*, 22(2):661–675.

Friederichs, P. and Hense, A. (2007). Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly weather review*, 135(6):2365–2378.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 105–124.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349.

Kallache, M., Vrac, M., Naveau, P., and Michelangeli, P.-A. (2011). Nonstationary probabilistic downscaling of extreme precipitation. *Journal of Geophysical Research: Atmospheres*, 116(D5).

Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the heffernan and tawn model. *Journal of Multivariate Analysis*, 115:396–404.

MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55(6):2137–2157.

Mannshardt-Shamseldin, E. C., Smith, R. L., Sain, S. R., Mearns, L. O., and Cooley, D. (2010). Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. *The Annals of Applied Statistics*, pages 484–502.

Michelangeli, P.-A., Vrac, M., and Loukos, H. (2009). Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophysical Research Letters*, 36(11).

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold es-timation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

Song, J. and Song, S. (2012). A quantile estimation for massive data with generalized pareto distribution. *Computational Statistics & Data Analysis*, 56(1):143–150.

Towe, R., Eastoe, E., Tawn, J., and Jonathan, P. (2016). Statistical downscaling for future extreme wave heights in the north sea.

Wadsworth, J. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126.

Wadsworth, J. and Tawn, J. (2012). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):543–567.

Wilby, R. L. and Wigley, T. (2000). Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology*, 20(6):641–661.