Progress in Oceanography 122 (2014) 77–91

Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

Estimating extremes from global ocean and climate models: A Bayesian hierarchical model approach

Eric C.J. Oliver [a,b,*], Simon J. Wotherspoon [a], Neil J. Holbrook [a,b]

[a] Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, TAS, Australia
[b] Australian Research Council Centre of Excellence for Climate System Science, Australia

ARTICLE INFO

Article history:
Received 27 May 2013
Received in revised form 9 December 2013
Accepted 16 December 2013
Available online 24 December 2013

ABSTRACT

Estimating oceanic and atmospheric extremes from global climate models is not trivial as these models often poorly represent extreme events. However, these models do tend to capture the central climate statistics well (e.g., the mean temperature, variances, etc.). Here, we develop a Bayesian hierarchical model (BHM) to improve estimates of extremes from ocean and climate models. This is performed by first modeling observed extremes using an extreme value distribution (EVD). Then, the parameters of the EVD are modeled as a function of climate variables simulated by the ocean or atmosphere model over the same time period as the observations. By assuming stationarity of the model parameters, we can estimate extreme values in a projected future climate given the climate statistics of the projected climate (e.g., a climate model projection under a specified carbon emissions scenario). The model is demonstrated for extreme sea surface temperatures off southeastern Australia using satellite-derived observations and downscaled global climate model output for the 1990s and the 2060s under an A1B emissions scenario. Using this case study we present a suite of statistics that can be used to summarize the probabilistic results of the BHM including posterior means, 95% credible intervals, and probabilities of exceedance. We also present a method for determining the statistical significance of the modeled changes in extreme value statistics. Finally, we demonstrate the utility of the BHM to test the response of extreme values to prescribed changes in climate.

1. Introduction

The recent "marine heat wave" recorded off Western Australia (Pearce and Feng, 2012; Wernberg et al., 2012) has focussed attention on marine extremes, a field that has received relatively less attention than atmospheric extremes. Extreme events can have significant impacts on the physical, chemical, and biological environment and it is not clear how they might change in a changing climate. Understanding the behavior of marine extremes in a changing climate is important for our understanding of the greater marine climate system as well as for predicting potential impacts on ecological habitats (Johnson et al., 2011; Wernberg et al., 2012).

Global climate models (GCMs) are indispensable tools for our understanding of the ocean and atmosphere climate system and how it may be changing under anthropogenic influences. GCMs perform well at capturing the general characteristics of the climate (e.g., the spatial distribution of mean temperatures) but underperform at capturing extreme events. For example, GCMs tend to underpredict the frequency and severity of heavy rainfall events and overpredict the extent of light drizzle (e.g., Perkins et al. (2007)). Despite the fact that GCMs poorly represent the extreme values it is still possible to glean information about the extremes from what the models represent well: the general climate.

Intuitively, the tails of probability distributions are related to the central moments of the distribution – at least for events which are "not too extreme". The shape of a distribution's tails can change significantly due to changes in the central statistics of the distribution, such as the mean or variance (e.g., Mearns et al. (1984) and Wigley (1985)), and one is reminded of the classic Intergovernmental Panel on climate change figure depicting the change in extreme hot and cold events due to changes in the temperature mean, variance, and skewness (Field et al., 2012). While it has been noted that trends in the statistics of extremes may not closely follow the trends of the mean (Katz, 2010) several studies have demonstrated that extreme value statistics can be well represented using the central statistics. For example, the frequency of air temperature extremes in the Asia–Pacific region were shown to be well-predicted by the mean temperature alone (Griffiths et al., 2005). Ballester et al. (2010) have shown that the changes in frequency, length, and intensity of air temperature extremes over Europe, in a climate change scenario, can be closely approximated using changes in the mean, variance and skewness simulated by an ensemble of GCM simulations. Simolo et al. (2011) modeled daily maximum and minimum temperature extremes in Europe using

* Corresponding author at: Institute for Marine and Antarctic Studies, Private Bag 129, University of Tasmania, Hobart, TAS, 7001, Australia. Tel.: +61 03 6226 6663.
E-mail address: eric.oliver@utas.edu.au (E.C.J. Oliver).

0079-6611/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pocean.2013.12.004

the first four L-moments (Hosking, 1990, 1992), showing that the change in the mean provided the best prediction for the changes in the extremes. Further, de Vries et al. (2012) showed that changes in the statistics of cold spells over Europe can be closely linked with changes in the mean and variance of air temperature.

Hierarchical models provide a framework for extreme events to be modeled using the climate statistics (e.g., Casson and Coles (1999), Cooley et al. (2007), and Sillmann et al. (2011)). A hierarchical model has multiple layers (or stages) where the parameters of one model layer are modeled by another layer. For example, a traditional linear regression model can be made hierarchical by allowing the regression coefficients to vary as a function of another set of variables, thereby adding another layer to the model. Here, we model marine extremes using an extreme value distribution (EVD) and in turn the EVD parameters are modeled as a function of some covariates (i.e., the marine climate statistics). This has the advantage that, if we assume the relationship between the covariates and extreme SSTs does not change in time, we can use projections of marine climate statistics under a climate change scenario and the fitted hierarchical model to model the extremes for the projected climate.

Bayesian estimation is particularly well-suited to hierarchical models and such models are called Bayesian hierarchical models (BHMs; Cressie and Wikle (2011)). In the Bayesian framework unknowns are modeled as random variables and so model inputs and outputs are both represented by probability distributions. Therefore, for all variables we can include or obtain estimates of their means and their uncertainty as well as inter-dependence (covariance). This implies that the construction of a model in a Bayesian framework relies on recognizing the inherent uncertainties and the model results reflect these uncertainties. BHMs allow for uncertainties at each level to be specified or modeled explicitly as a parameter (observational uncertainty, process model error, etc.). Several excellent introductions and reviews of Bayesian hierarchical methods in the atmospheric and ocean sciences are provided by Berliner et al. (1998), Cressie and Wikle (2011), and Wikle et al. (2013).

BHMs have been used in the geophysical literature for a number of years. Royle et al. (1999) estimated spatially regular wind fields from sparse scatterometer data using a BHM in which the covariance structure of the wind was conditional upon the atmospheric pressure field (using a hybrid physics–statistics model). This model allowed one to extrapolate the wind estimates where no observations existed, based on the wind–pressure relationship elucidated by the BHM. Berliner et al. (2003) jointly modeled atmospheric and oceanic variables as a function of independent measured data (i.e., scatterometer measurements, altimetry data) with a BHM, allowing for coupling of the atmosphere–ocean variables (air–sea interactions). Milliff et al. (2011) provided a BHM implementation for the generation of surface wind initial conditions for ensemble ocean forecasting, including a detailed explanation of the BHM algorithm. Bayesian methods, and in particular BHMs, have been used to model sea surface temperature variability over a range of time scales including high-frequency variability, the seasonal cycle, multi-decadal trends, and the mean (Higdon, 1998; Lemos and Sansó, 2009; Lemos et al., 2010). Bayesian techniques are also well-established in the ocean ecosystems modeling literature (e.g., Harmon and Challenor (1997)). For example, Fiechter et al. (2013) used a BHM in which the process layer is a Nutrient-Phytoplankton-Zooplankton-Detritus (NPZD) model and others have developed similar models using statistical emulators of NPZD mechanisms (Hooten et al., 2011; Leeds et al., 2013).

Using BHMs for extreme value analysis is a relatively recent development. Casson and Coles (1999) discussed the idea of pooling information on extreme values spatially thereby borrowing information across space to inform the model. The extremes were modeled site-wise in tandem with a latent spatial process model for the variation of parameter values in space. This model used Markov chain Monte Carlo techniques for estimating model parameter values, and it was demonstrated to be skillful by predicting the hurricane climate of the Atlantic and Gulf of Mexico regions. Cooley et al. (2007) developed a hierarchical model where extreme precipitation values observed at weather stations (i.e., point locations) were modeled using a peak-over-threshold extremes model (i.e., the Generalized Pareto Distribution). The parameters of the Generalized Pareto Distribution were then modeled using a latent spatial process. The latent spatial process was expressed using a set of covariates including latitude, longitude, mean precipitation, elevation, and terrain type. The fitted model was then used to interpolate extremes over locations covered by the covariates but for which observed extremes were not available. Similarly, Friederichs and Thorarinsdottir (2012) modeled peak wind using the Generalized Extreme Value distribution, the parameters of which were modeled as a function of covariates such as mean wind speed, wind speed variance, rain rate, atmospheric pressure, and the pressure tendency.

Generally, the parameters of a Bayesian hierarchical model are estimated using Markov chain Monte Carlo algorithms (Casson and Coles, 1999; Cooley et al., 2007; Sang and Gelfand, 2009; Schliep et al., 2010). However, we would like to note that there have also been many hierarchical models with parameter estimation performed by frequentist maximum-likelihood techniques. For example, Sillmann et al. (2011) related extreme air temperature minima over Europe are to atmospheric blocking patterns over the North Atlantic. Other examples include those by Abeysirigunawardena et al. (2009) and Zhang et al. (2010) who used indices of climate variability (e.g., the Southern Oscillation Index, the Pacific-North American teleconnection pattern) as predictors in models of extreme winds in Western Canada and extreme precipitation over North America, respectively.

In this paper we outline the BHM technique and how it can be used to improve estimates of extremes from global ocean and climate models. The basic idea is to fit the BHM to observed extremes using model output climate variables as covariates (mean, variance, etc.). Then by assuming stationarity of the model parameters and given simulated climate variables for a projected future climate we can model future extremes. In doing so we extend the technique of spatial interpolation of extremes (e.g., Cooley et al. (2007)) to include temporal extrapolation. This technique is demonstrated using dynamically downscaled ocean model simulations of global climate projections representing the 1990s and 2060s decades, under the A1B carbon emissions scenario (Nakicenovic et al., 2000). Essentially, this approach is a form of bias correction of the model simulated extremes. The present article focuses on developing a BHM methodology for extremes analysis of ocean temperatures using ocean climate model output data. While we provide a specific case study for the BHM development and its application, our BHM approach presented here is intended to be quite generic so that it can be readily applied in other ocean climate extremes' contexts. For more complete details of the climate change application of our technique developed here please see Oliver et al. (in press) which focusses on understanding changes in sea surface temperature extremes off southeastern Australia in response to future climate change scenarios.

This paper is structured as follows. Extreme value theory and methods of fitting extreme value distributions are presented in Section 2. The Bayesian hierarchical model approach is outlined in Section 3 and demonstrated for extreme sea surface temperatures off southeastern Australia in Section 4. A discussion and conclusions are presented in Section 5.

2. Extreme value theory

Extreme events are those producing climate anomalies which are rare and whose magnitudes deviate significantly from the median of the probability distribution. The statistics of extremes, i.e., those that describe the tails of a probability distribution, are often examined in their own right independently from the central statistic of the distribution, i.e., the mean, variance, etc. The modeling of extreme values has a long history (e.g., Gumbel (1958)). Leadbetter et al. (1983) outlined the theoretical framework for extremal analysis and, more recently, a practical methodological primer has been provided by Coles (2001). There are two main approaches to modeling extreme values: "block maxima" and "peak-over-threshold". In the block maxima approach, the maximum of a time series over a specified block length is modeled using the Generalized Extreme Value distribution. In the peak-over-threshold approach, all values above a specified threshold are modeled using the Generalized Pareto Distribution. Both the "block maxima" and "peak-over-threshold" approaches use only a subset of the data considered to be extreme-valued in order for the description of the tail not to be influenced by non-extreme-valued data. The two approaches are in fact related and the model parameters of one distribution can be derived from the model parameters of the other (Coles, 2001). In both cases, the fitted distribution can be used to estimate the return periods or return levels of extreme events.

2.1. Extreme value models for block maxima

Here we outline the extreme value theory required for the present study. First, consider a stationary sequence of random variables $\{x_t | t = 1, 2, \ldots\}$. Let y denote the maximum of $\{x_t\}$ over a block of length n:

$$y = \max(x_1, x_2, \ldots, x_n). \tag{1}$$

In the limit $n \to \infty$, and if y can be linearly renormalized so that its cumulative distribution function converges, then it must converge to one of three types: Gumbel (Type I), Fréchet (Type II), or Weibull (Type III; e.g., Coles (2001)). These distributions are conveniently summarized by the Generalized Extreme Value (GEV) distribution:

$$F_{\mathrm{GEV}}(y|a,b,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-a}{b}\right)\right]^{-\frac{1}{\xi}}\right\}. \tag{2}$$

Here a is the location parameter, b is the scale parameter (constrained to be positive), and ξ is the shape parameter.

For relatively short records (i.e., the time series x_t contains a small number of blocks of length n) it can be difficult to obtain a stable estimate of the shape parameter. As this is often the case for environmental time series, unless studying sea levels for which centennial records are occasionally available from tide gauges, we will focus on the limit $\xi \to 0$. In this limit the GEV distribution converges to the Type I, or Gumbel, distribution:

$$F_{\mathrm{I}}(y|a,b) = \exp\left[-\exp\left(-\frac{x-a}{b}\right)\right]. \tag{3}$$

Given a vector of block maxima $\boldsymbol{y}$, and estimates of the parameters a and b, a plot of $\boldsymbol{y}$ against the quantiles of the Gumbel distribution (referred to as a Q–Q plot) can be used to demonstrate the fit of the Gumbel distribution. If the Gumbel distribution fits the data well then this relationship is linear; if the relationship deviates significantly from being linear, concave up or concave down, then the data are more appropriately described by the Fréchet (Type II), or Weibull (Type III) distributions, respectively, with significant consequences when extrapolating into the tail of the distribution to make return level estimates. This method will be illustrated using annual SST maxima at six locations off eastern Australia. The annual maxima were derived from daily fields of Advanced Very High Resolution Radiometer (AVHRR) SST observations[1] over a 28-year period (1/1/1982–31/12/2009). Plots of the annual SST maxima against the quantiles of the Gumbel distribution show a linear, or nearly-linear, relationship indicating that the Gumbel distribution provides a good fit to the data (Fig. 1).

Note that we have specified $\{x_t\}$ to be a stationary sequence and sequences of environmental variables are commonly non-stationary, exhibiting low-frequency variability in the mean and variance due to the seasonal cycle and interannual variations. However, if we constrain the block length to be $n \geqslant n_s$, where n_s is the characteristic time scales of these non-stationary variations (e.g., one year for seasonal variations), then the theory outlined above still holds (Coles, 2001).

The return periods and return levels of extreme values can be estimated from the parameters of the Gumbel distribution. The return period for a specified extreme value represents the expected frequency with which, in a probabilistic sense, that extreme value will repeat. For example, if the return period for a 30 °C extreme value, denoted T_{30}, is 50 years then there is a 1 in 50 chance of a 30 °C event occurring in any given year. Conversely, the return level is the extreme value associated with a particular return period. In the previous example, the 50-year return level, denoted z_{50}, is 30 °C. Given a return period T_z, the associated return level z_T can be estimated from the quantiles of the Gumbel distribution using the equation:

$$z_T(y) = a - b\log[-\log F_{\mathrm{I}}(y|a,b)], \tag{4}$$

where the quantiles $F_{\mathrm{I}}(y|a,b)$ are related to the return period T_z by

$$T_z(y) = [1 - F_{\mathrm{I}}(y|a,b)]^{-1}. \tag{5}$$

Similarly, given a return level z_T these equations can be used to calculate the corresponding return period T_z. The return level z_{T} has the same units as does y and the return period has units of the block length n, i.e., if n is one year (i.e., y are annual maxima) then T_z has units of years.

2.2. Parameter estimation

There are two common paradigms within which estimation of the parameters of an extreme value distribution (EVD) can be performed: the frequentist paradigm, in which maximum likelihood is a common approach, and the Bayesian paradigm, in which techniques based on Markov chain Monte Carlo methods are commonly applied (Coles, 2001). Let $\boldsymbol{y}$ be a vector of N annual maxima each of which are assumed to be distributed according to the EVD, i.e., $y_i \sim f(y_i|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ are the EVD parameters (e.g., $\boldsymbol{\theta} = (a, b)$ if the EVD is the Gumbel distribution described above). The likelihood function of $\boldsymbol{\theta}$ is given by

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{N} f(y_i|\boldsymbol{\theta}), \tag{6}$$

where p denotes probability density. Maximum likelihood is an approach that provides an estimator for $\boldsymbol{\theta}$ by finding the value that maximizes the likelihood function. While the maximum likelihood method is simple and efficient for many problems it can become computationally inefficient or even prohibitive for more complex likelihood functions. The Gumbel parameters were estimated by maximum likelihood for the annual SST maxima at the six locations off eastern Australia discussed above (Fig. 1). The linear fit to the data, provided by the Gumbel parameters, indicates that the Gumbel distribution represents the data well (Fig. 1, compare solid lines and black circles).

[1] Obtained from the Jet Propulsion Laboratory, http://podaac.jpl.nasa.gov/.

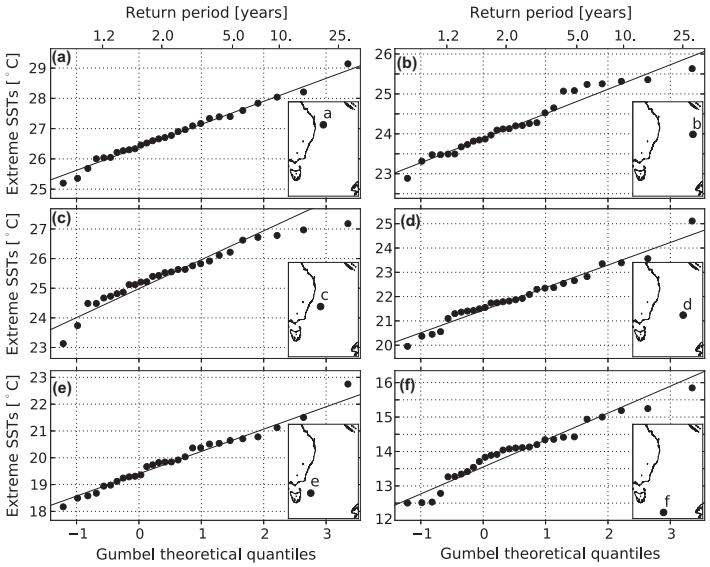

Fig. 1. Annual SST maxima at six locations off eastern Australia against the return period and quantiles assuming a Gumbel distribution. The annual maxima are indicated by the black dots and linear fits based on maximum-likelihood estimation of the Gumbel distribution parameters are shown by solid lines. The *y*-axis represents the return level, in °C, and the upper *x*-axis represents the return period, in years. The location of each SST record is shown in the lower right corner of each panel.

In the Bayesian paradigm, the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ is related to the prior distribution $p(\boldsymbol{\theta})$ and the likelihood function $p(\boldsymbol{y}|\boldsymbol{\theta})$ (given by Eq. 6) by Bayes' rule

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})}, \tag{7}$$

where the denominator is simply a normalizing constant to ensure that $\int_{-\infty}^{\infty} p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} = 1$. A common method for estimating $\boldsymbol{\theta}$ (i.e., fitting the EVD to the data) in the Bayesian paradigm is to draw approximate samples from the posterior distribution of theta using Markov chain Monte Carlo (MCMC) techniques (e.g., Coles (2001)). MCMC techniques do not require knowledge of the normalizing constant and so we use a form of Bayes' rule written as a proportionality where this constant is dropped:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{8}$$

From these MCMC samples one can calculate the posterior mean (an estimator for $\boldsymbol{\theta}$) as well as estimates of other statistics such as the posterior variance or the posterior distribution itself.

3. Estimating extreme value statistics from climate

Consider the annual maxima of an environmental variable of interest (e.g., temperature or rainfall) at J locations. Let this set of annual maxima be denoted $\boldsymbol{Y}$ where $\boldsymbol{Y} = \{\boldsymbol{y}_j | j = 1, 2, \ldots, J\}$ is a list of vectors, with $\boldsymbol{y}_j$ a vector of length N representing the annual maxima at the jth location and N is the number of years in the record. These annual maxima are calculated over a particular time period P_1, most likely the full available extent of the observational record. An extreme value distribution can be fit to the annual maxima to obtain estimates of extreme value statistics at each of the J locations. However, if estimates of extreme value statistics are desired for a time period other than P_1 or for a spatial region that is not covered by the J locations (i.e., for which there are no observed annual maxima) then the direct application of extreme value theory alone is inadequate.

Hierarchical models provide a framework for addressing such problems. Consider the problem outlined above but with the additional information $\boldsymbol{\chi}$ which are covariate variables defined over multiple time periods $\{P_i\}$ at K locations. The original time period P_1 and the set of J locations are each a subset of $\{P_i\}$ and the set of K locations respectively. In other words, the covariates $\boldsymbol{\chi}$ are defined over time periods and locations that include but are not limited to P_1 and J. A hierarchical model can be used to model $\boldsymbol{Y}$ given the covariates $\boldsymbol{\chi}$. If this model provides a satisfactory fit to the annual maxima $\boldsymbol{Y}$ over P_1 and J then it can be used to simulate the annual maxima over the remaining time periods $\{P_i\} - P_1$ and locations $K - J$.

If only a single time period is considered (i.e., $\{P_i\} = P_1$) then this problem involves the interpolation of extreme value statistics in space. For example, given precipitation measurements located sparsely in space (at J locations) and gridded fields of precipitation statistics covering the entire domain (at $K > J$ locations), Cooley et al. (2007) used a hierarchical model, fitted to the observed extremes at the station locations, to interpolate the extreme value statistics over the entire domain. On the other hand, if two time periods are given and $\boldsymbol{Y}$ and $\boldsymbol{\chi}$ are both defined at the same set of locations (i.e., $\{P_i\} = \{P_1, P_2\}$ and $K = J$) then the problem consists of extrapolation of the extreme value statistics in time. For example, given sea surface temperature measurements at all locations within a specified domain and gridded fields of marine climate statistics defined for the 1990s (P_1) and the 2060s under a climate change scenario (P_2), Oliver et al. (in press) used a hierarchical model, fitted to the observed extremes over the 1990s, to estimate extremes for the 2060s. These examples illustrate two possible uses for such a model, e.g., spatial interpolation and temporal extrapolation, but the model can be applied to the general situation which combines these two problems.

3.1. The Bayesian hierarchical model

Hierarchical models lend themselves well to a Bayesian approach (Gelman et al., 2003; Cooley et al., 2007; Oliver et al., in press) and such models are referred to as Bayesian hierarchical models (BHMs). The BHM considered here consists of three layers. The data layer models the observed annual maxima $\boldsymbol{Y}$ using the EVD. The climate process layer models the parameters of the EVD ($\boldsymbol{\theta}_2$) as a function of some covariates ($\boldsymbol{X}$, taken to be the large-scale climate statistics, i.e., the subset of $\boldsymbol{\chi}$ defined for period P_1 and the J locations). The final layer consists of the prior

distributions for the parameters of the climate process layer model ($\boldsymbol{\theta}_1$). This conditional dependence of model layer parameters modifies Relation 8 (Bayes' rule) to yield

$$p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{X}) \propto p(\boldsymbol{Y}|\boldsymbol{\theta}_2)\, p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1,\boldsymbol{X})\, p(\boldsymbol{\theta}_1), \tag{9}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

3.1.1. Data layer

In the first (data) layer we model the annual maxima using the EVD. The technique is general for any choice of block maxima EVD (see Section 2) but we will use the Gumbel distribution in order to illustrate the model. We assume that at each location j, the annual maxima $\boldsymbol{y}_j$ are distributed according to the Gumbel distribution, where $\boldsymbol{y}_j$ is a vector of length N, and N is the number of years in the record. This is expressed as the likelihood function (Eq. 6):

$$p(\boldsymbol{y}_j|a_j,\phi_j) = \prod_{i=1}^{N} f_l(y_{ji}|a_j,\phi_j). \tag{10}$$

Here f_l is the density function for the Gumbel distribution, where the scale parameter b_j has been scaled logarithmically, $\phi_j = \log(b_j)$ to allow for both positive and negative values in the model. When considering the annual maxima at all locations J (i.e., $\boldsymbol{Y} = \{\boldsymbol{y}_j | j = 1, 2, \ldots, J\}$) the likelihood function becomes

$$p(\boldsymbol{Y}|\boldsymbol{\theta}_2) = \prod_{j=1}^{J} p(\boldsymbol{y}_j|a_j,\phi_j) = \prod_{j=1}^{J}\prod_{i=1}^{N} f_l(y_{ji}|a_j,\phi_j), \tag{11}$$

where $\boldsymbol{\theta}_2 = (\boldsymbol{a},\boldsymbol{\phi})$ and $\boldsymbol{a} = \{a_j | j = 1, 2, \ldots, J\}$ and $\boldsymbol{\phi} = \{\phi_j | j = 1, 2, \ldots, J\}$. This equation gives us the first term on the right-hand side of Relation 9.

If modeling the annual maxima with an EVD other than the Gumbel distribution then the parameters of the EVD of choice are contained in $\boldsymbol{\theta}_2$. For example, if using the GEV distribution then $\boldsymbol{\theta}_2 = (\boldsymbol{a},\boldsymbol{\phi},\xi)$ where ξ is the shape parameter at all J locations. It is also possible to formulate the model using a peak-over-threshold EVD (see Cooley et al. (2007) for an example).

3.1.2. Climate process layer

In the second (climate process) layer, we model EVD parameters as a function of a latent spatial process (e.g., Chapter 9, Coles (2001), Cooley et al. (2007)). This latent spatial process is characterized by the covariates in $\boldsymbol{\chi}$ which are typically climate statistics or dynamical variables (the choice of covariates is discussed further in Section 3.3). Let $\boldsymbol{\chi}$ be a $K \times M \times P$ third-order tensor where K is the number of locations, $M-1$ is the number of covariates included in the model ($\boldsymbol{\chi}$ also includes the intercept term used in the linear regression below, which we do not consider a "covariate") and P is the number of time periods over which $\boldsymbol{\chi}$ is defined. So, χ_{kmt} is the value of the mth covariate at location k and time period t. Let $\boldsymbol{X}$ be a $J \times M$ matrix subset of $\boldsymbol{\chi}$ defining the covariates at the J locations and over the single time period P_1 for which $\boldsymbol{Y}$ were observed.

We have assumed that the climate process model for the Gumbel parameters $\boldsymbol{\theta}_2$ can be factored into independent models for $\boldsymbol{a}$ and $\boldsymbol{\phi}$:

$$p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1,\boldsymbol{X}) = p(\boldsymbol{a}|\boldsymbol{\theta}_{1,a},\boldsymbol{X})\, p(\boldsymbol{\phi}|\boldsymbol{\theta}_{1,\phi},\boldsymbol{X}), \tag{12}$$

where $\boldsymbol{\theta}_1 = (\boldsymbol{\theta}_{1,a}, \boldsymbol{\theta}_{1,\phi})$. Each of these independent models characterizes the relationship between the EVD parameters and the latent spatial process. The latent spatial process is modeled as a linear regression of the EVD parameters onto the covariates, i.e.,

$$\boldsymbol{a} = \boldsymbol{X}\boldsymbol{\beta}_a + \boldsymbol{\epsilon}_a, \tag{13}$$

$$\boldsymbol{\phi} = \boldsymbol{X}\boldsymbol{\beta}_\phi + \boldsymbol{\epsilon}_\phi, \tag{14}$$

where $\boldsymbol{\beta}_a$ and $\boldsymbol{\beta}_\phi$ are regression coefficients (vectors of length M) and $\boldsymbol{\epsilon}_a$ and $\boldsymbol{\epsilon}_\phi$ are error terms (vectors of length J).

The error terms are random effects which characterize effects not included in the model. These error terms are distributed normally with zero mean and covariance matrices given by $\tau_a^{-1}\mathbf{I}$ and $\tau_\phi^{-1}\mathbf{I}$ respectively, where $\mathbf{I}$ is the identity matrix and τ_a and τ_ϕ are the precisions.[2] Therefore, the random effects are independent between locations which allows for spatial variation of the model errors. Note that we have not imposed a spatial structure on the random effects (such as through the error covariance structure, e.g., Cooley et al. (2007)) and so our random effects are not spatially auto-correlated (see Cressie and Wikle (2011) for more details on spatial random effects).

The linear regression models can be expressed as J-dimensional multivariate Normal distributions for $\boldsymbol{a}$ and $\boldsymbol{\phi}$:

$$p(\boldsymbol{a}|\boldsymbol{\theta}_{1,a},\boldsymbol{X}) = \mathcal{N}_J(\boldsymbol{X}\boldsymbol{\beta}_a, \tau_a^{-1}\mathbf{I}), \tag{15}$$

$$p(\boldsymbol{\phi}|\boldsymbol{\theta}_{1,\phi},\boldsymbol{X}) = \mathcal{N}_J(\boldsymbol{X}\boldsymbol{\beta}_\phi, \tau_\phi^{-1}\mathbf{I}), \tag{16}$$

where $\mathcal{N}_J(\boldsymbol{X}\boldsymbol{\beta}, \tau^{-1}\mathbf{I})$ denotes a J-dimensional multivariate Normal distribution with mean $\boldsymbol{X}\boldsymbol{\beta}$ and covariance $\tau^{-1}\mathbf{I}$. The parameters of the climate process layer are combined as $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_a, \boldsymbol{\beta}_\phi, \tau_a, \tau_\phi)$, and Eqs. (15) and (16) give us the second term on the right-hand side of Relation 9. We acknowledge that it is not necessary to assume independent distributions for the parameters $\boldsymbol{\theta}_2$, a multivariate distribution could be specified (e.g., Sang and Gelfand (2010)) whereby the relationship between the EVD location and scale parameters is modeled jointly. Furthermore, spatial dependence could be modeled explicitly, such as with a variogram model for the spatial correlation of the EVD parameters (e.g., Cooley et al. (2007)).

Note that in Eq. (11) we have assumed independence between the extremes at different locations, conditional on the latent spatial process. In reality, there will be some spatial dependence of extrema, i.e., we can expect the annual maxima at neighboring grid points to be similar. This assumption will lead to unrealistic spatial structure in the realisations of the annual maxima $\boldsymbol{Y}$ (Davison et al., 2012). However, this is not critical for the estimation of extreme value statistics (i.e. the marginal parameters of the EVD) since spatial dependence is enforced due to spatial correlation of the covariates.

If using an EVD which includes a shape parameter ξ (e.g., the GEV distribution), it is not necessary to model ξ as a function of the covariates, but it can be assumed that it is governed by a physical process fixed in time and space (e.g., Sillmann et al. (2011)). Spatial uniformity of ξ can also be assumed and a single value estimated for the entire domain, thereby borrowing strength across locations in space.

3.1.3. Priors

In the third and final layer we assign priors to the parameters $\boldsymbol{\theta}_1$ of the climate process layer. Each of the parameters in the climate process layer is assumed to be distributed independently, i.e.,

$$p(\boldsymbol{\theta}_1) = p(\boldsymbol{\beta}_a)p(\boldsymbol{\beta}_\phi)p(\tau_a)p(\tau_\phi). \tag{17}$$

Note that it is not necessary to assume independent prior distributions on the parameters $\boldsymbol{\theta}_1$ but we make this choice for computational simplicity.

When there is no prior knowledge regarding how the EVD parameters are related to the covariates, diffuse non-informative priors should be adopted. In general any distribution that is suitably broad will be sufficient. However we have chosen conjugate priors in order to simplify the expressions for the conditional distributions of $\boldsymbol{\beta}_a, \boldsymbol{\beta}_\phi, \tau_a, \tau_\phi$ (see Appendix A). Hence, we have taken the following forms for the priors of each of these parameters

[2] Precision is defined as the inverse of the variance, $\tau = \sigma^{-2}$.

$$p(\boldsymbol{\beta}_a) = \mathcal{N}_M(\boldsymbol{\mu}_a, \tau_{a,0}^{-1}\boldsymbol{I}), \tag{18}$$

$$p(\boldsymbol{\beta}_\phi) = \mathcal{N}_M(\boldsymbol{\mu}_\phi, \tau_{\phi,0}^{-1}\boldsymbol{I}), \tag{19}$$

$$p(\tau_a) = \text{Gamma}(\alpha_a, \theta_a), \tag{20}$$

$$p(\tau_\phi) = \text{Gamma}(\alpha_\phi, \theta_\phi), \tag{21}$$

where the values of $\boldsymbol{\mu}_a, \boldsymbol{\mu}_\phi, \tau_{a,0}, \tau_{\phi,0}, \alpha_a, \alpha_\phi, \theta_a$ and θ_ϕ are chosen to provide suitably broad (diffuse) priors.

3.2. Markov chain Monte Carlo sampling technique

The regression coefficients (e.g., $\boldsymbol{\beta}_a$ and $\boldsymbol{\beta}_\phi$) and error variances (e.g., τ_a and τ_ϕ) can be sampled using the Markov chain Monte Carlo (MCMC) technique (e.g., Coles (2001) and Gelman et al. (2003)). Within each step of the MCMC loop, samples of the EVD parameters (e.g., $\boldsymbol{a}$ and $\boldsymbol{\phi}$) are drawn and accepted or rejected based on the Metropolis rule. Then a Gibbs sampling step (Gelman et al., 2003) is performed to sample for the $\boldsymbol{\beta}$s and the τs, given the EVD parameters. The MCMC simulation should be performed for enough steps to achieve equilibration of the samples and to obtain a large enough sample for subsequent analysis. The initial equilibration period, or "burn-in", is discarded and the remainder is thinned by taking only every dth output step leaving samples from the joint posterior distribution of the $\boldsymbol{\beta}$s and the τs. The thinning interval d is chosen to reduce the serial correlation present in the MCMC output; this correlation is high for consecutive samples but in our experience drops to acceptable levels for $d \simeq O(10)$. The MCMC sampling algorithm is explained in more detail below.

The values of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ must be initialized before running the MCMC loop. The results will not be sensitive to this choice other than to the extent that it influences the length of the burn-in period. It is sensible to initialize the regression coefficients and the EVD parameters to zero and the error precisions to a non-zero and positive value. At each step of the MCMC loop the following procedures are executed:

1. Loop over all locations $j = 1, 2, \ldots, J$ and for each j:
   (a) Draw a candidate sample of θ_{2j} (e.g., a_j and ϕ_j for the Gumbel distribution) denoted θ_{2j}^* using the multivariate linear regression models (Eqs. (15) and (16)) which make use of the covariate information $\boldsymbol{X}$; and
   (b) Accept θ_{2j}^* or reject in favor of θ_{2j} from the previous MCMC step based on the Metropolis rule (e.g., Gelman et al. (2003)).
2. Given $\boldsymbol{\theta}_2$ (e.g., $\boldsymbol{a}$ and $\boldsymbol{\phi}$) at all locations, Gibbs sample (e.g., Gelman et al. (2003)) for the regression coefficients $\boldsymbol{\beta}_a$ and $\boldsymbol{\beta}_\phi$ and precisions τ_a and τ_ϕ by sampling from the conditional distributions $p(\boldsymbol{\beta}_a|\tau_a, \boldsymbol{a}), p(\boldsymbol{\beta}_\phi|\tau_\phi, \boldsymbol{\phi}), p(\tau_a|\boldsymbol{\beta}_a, \boldsymbol{a})$, and $p(\tau_\phi|\boldsymbol{\beta}_\phi, \boldsymbol{\phi})$. (These expressions are given in Appendix A).

The output of the MCMC loop, after discarding the burn-in period and thinning, consists of approximate samples from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X})$.

The samples of $\boldsymbol{\theta}_1$ can now be used to generate a suite of statistics to summarize the extreme value properties of the system. We will assume that the fitted model represents a fundamental physical process linking the covariates to the extreme value statistics and is therefore stationary in time and space. This allows us to use the values of $\boldsymbol{\theta}_1$, which have been fitted to the covariates defined using a subset of $\boldsymbol{\chi}$ (i.e., $\boldsymbol{X}$) to estimate the extreme value statistics over the entire temporal and spatial domain covered by $\boldsymbol{\chi}$. In this way, we are able to spatially interpolate and temporally extrapolate the estimations of the extreme value statistics. This method, and a description of the various statistics that can be obtained, is illustrated in Section 4 for the case of estimating sea surface temperature extremes off southeastern Australia for two time periods over the same spatial domain given observations over only a single time period.

3.3. Predictor selection and model comparison

The first step to formulating the BHM described above is to select all the potential predictors that are to be considered for the definition of the covariate tensor $\boldsymbol{\chi}$. Some general guidelines for selecting potential predictors and a method for comparing models based on subsets of these predictors is described in this subsection.

There are three types of potential predictor variables that should be considered: (i) central moments, (ii) dynamical variables, and (iii) extreme value statistics. Obvious predictors to include would be the first and second central moments (the mean and variance respectively) of the quantity of interest, i.e., mean temperature and temperature variance if considering temperature extremes; standardized moments such as skewness and kurtosis or, more likely, the non-standardized third and fourth central moments are also possible candidates to be included (see Oliver et al. (in press) for a discussion on the relative merit of standardized and central moments as predictors). Dynamical variables of relevance to the problem at hand should also be considered. Examples could be eddy kinetic energy, if considering oceanic extremes, or wind variability, if considering terrestrial extremes. These variables may be able to characterize dynamical mechanisms that are physically linked to the distribution of extreme values. Finally, the extreme value statistics of the ocean/atmosphere/climate model itself can be included as covariates for the parameters of the extreme value model for the observations. For example, if using the Gumbel distribution as the extreme value model then the Gumbel parameters estimated by fitting directly to the annual maxima simulated by the circulation model can be used as predictors. This would be a more explicit form of bias correction; whereas using predictors based on central moments and dynamical variables is an indirect form of bias correction using the relationship between the extreme value statistics and a latent spatio-temporal process. It may also be of value to include different (i.e., nonlinear) scalings of the chosen predictors (e.g., it may be possible that the standard deviation provides a better predictor than the variance).

Given a set of potential predictors it is then necessary to choose which subset provides the best model – this is referred to as model selection. In general, a model based on too few predictors is unable to fully represent the distribution of the extreme value statistics; a model with too many predictors often leads to overfitting. An approach for BHMs to choosing between models is to compare them using the Deviance Information Criteria (DIC, Spiegelhalter et al. (2002) and Gelman et al. (2003)):

$$\text{DIC} = \overline{D(\boldsymbol{\theta}|\boldsymbol{Y})} + p_D, \tag{22}$$

where an overbar represents the posterior mean and

$$D(\boldsymbol{\theta}|\boldsymbol{Y}) = -2\log\{p(\boldsymbol{Y}|\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \boldsymbol{X})\}, \tag{23}$$

$$p_D = \overline{D(\boldsymbol{\theta}|\boldsymbol{Y})} - D(\bar{\boldsymbol{\theta}}|\boldsymbol{Y}). \tag{24}$$

The DIC is the sum of the deviance measure $\overline{D(\boldsymbol{\theta}|\boldsymbol{Y})}$, which decreases as the model fit to the data improves, and the model complexity p_D, which increases with model complexity and provides a penalty for models that fit the data better but do so at the cost of model simplicity, and is meant to control for model overfitting. When comparing two models, the model with the lowest DIC is to be preferred. It has been pointed out that DIC tends to favor overfit models (Ando, 2007) and so it is recommended that model selection be done by using DIC in conjunction with visual assessment of the model results and domain expertise.

If there is an abundance of data, then it is possible to perform a cross-validation procedure. This can be undertaken by fitting the model over a subset of the domain and calculating $D_{\text{rep}} = D(\bar{\boldsymbol{\theta}}|\boldsymbol{Y}_{\text{rep}})$, which provides a measure of model fit over the remaining independent data (or "replicates" $\boldsymbol{Y}_{\text{rep}}$) (Gelman et al., 2003).

4. Example: sea surface temperatures off southeastern Australia

The technique outlined above was used by Oliver et al. (in press) to estimate extreme sea surface temperatures (SSTs) off southeastern Australia from dynamically downscaled climate model output. The annual maxima $\boldsymbol{Y} = \{\boldsymbol{y}_j | j = 1, 2, \ldots, J\}$ were derived from daily fields of AVHRR SST observations over $N = 28$ years (as in Section 2.1). The $J = 2071$ locations are regularly spaced over the domain bounded by 144°E, 169°E, 48°S and 20°S and only where water depth is greater than 200 m. Marine climate statistics were derived from dynamically downscaled ocean climate model simulations representing the 1990s, using realistic forcing fields (ECMWF Reanalysis output; Uppala et al. (2005)), and for the 2060s, under the A1B atmospheric carbon emissions scenario (Sun et al., 2012; Chamberlain et al., 2012; Oliver et al., in press). In Oliver et al. (in press), the best model was chosen, according to the DIC-based selection criteria outlined in Section 3.3, to be that which included the mean SST, the SST variance, the SST third central moment and the eddy kinetic energy as covariates (see also Appendix B). We use that study, which focuses on estimated changes in SST extremes without delving into the details of the model, as a case study to demonstrate the model and illustrate some of the statistics that can be calculated using it.

4.1. Estimating extremes for a projected future climate

Let the four marine climate variables from the chosen model be collected together as the $J \times M \times P$ tensor $\boldsymbol{\chi}$. The number of covariates is four and so $M = 5$ since we must include the intercept term required for the linear regression model; the number of time periods is $P = 2$ (P_1 is the 1990s and P_2 is the 2060s). The observed annual maxima $\boldsymbol{Y}$ were calculated over 1982–2009 and taken to be representative of the 1990s (P_1). We wish to estimate the extreme value statistics for the 2060s (P_2). Following Section 3.1.2, let $\boldsymbol{X}$ be a $J \times M$ subset of $\boldsymbol{\chi}$, being the covariates at the J locations and the single time period P_1. Note that the data source used to define the covariates $\boldsymbol{X}$ (the ocean model) is different from that used to derive the extremes $\boldsymbol{Y}$ (the observed SST record). Following Oliver et al. (in press), the model was fit using the design matrix $\boldsymbol{X}$, and samples from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X})$ were drawn using the MCMC algorithm described in Section 3.2. Given posterior samples of θ_1, samples from the posterior predictive distributions of $\boldsymbol{a}$ and $\boldsymbol{\phi}$ (and $\boldsymbol{b} = \exp \boldsymbol{\phi}$, see Section 3.1.1) can be simulated using Eqs. (15) and (16). Further, given a return period T samples of the return levels $\boldsymbol{z}_T$ can be calculated. Conversely, given a return level z, samples of the return periods $\boldsymbol{T}_z$ can be calculated (see Eqs. (4) and (5)).

We can also extrapolate the extreme value statistics to the time period P_2. Let $\boldsymbol{X}'$ be a $J \times M$ subset of $\boldsymbol{\chi}$, being the covariates at the J locations and the time period P_2. These covariate values are the climate statistics (the SST mean, variance, third moment and eddy kinetic energy) from the 2060s model simulation. By substituting $\boldsymbol{X}'$ for $\boldsymbol{X}$ in Eqs. (15) and (16) we can draw samples from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X}')$ to yield samples of $\boldsymbol{a}$ and $\boldsymbol{\phi}$ (and $\boldsymbol{b}$), as well as return periods and return levels, for the future time period. Note that we assign a prime to all estimations for the future time period: $\boldsymbol{a}', \boldsymbol{\phi}', \boldsymbol{b}', \boldsymbol{T}_z'$ and $\boldsymbol{z}_T'$.

4.2. Posterior means and probability distributions

Draws from the joint posterior distribution of $\boldsymbol{a}$ and $\boldsymbol{\phi}$ can be obtained for both P_1 and P_2 time periods, and by applying Eqs. (4) and (5) we can subsequently obtain draws from the posterior distributions of $\boldsymbol{T}_z$ and $\boldsymbol{z}_T$. We can then calculate a suite of statistics to analyze and summarize the results. The probability distributions of a, b, z_{50} (the 50-year return level) and T_{20} (the return period for an extreme temperature event equaling or exceeding a 20 °C) at the location (152°E, 43°S) are shown in Fig. 2. The distributions of a for P_1 and P_2 are well separated, most likely due to the dependence of this variable on the mean SST which changes significantly between the two periods, and the distributions of b overlap significantly indicating a less dramatic change. The z_{50} are also well separated and there is about a 3 °C shift between the two time periods. Comparing the distributions of T_{20} is more difficult since the mean SST has shifted closer to 20 °C in P_2 and has thus caused the distribution to change from relatively broad and diffuse for P_1 with long return periods (20–50 years) to very focused for P_2 with very short return periods (< 5 years). The simplest summary statistic for these distributions is the posterior mean which is indicated by the filled circles in Fig. 2. However, we can see that other statistics may be of interest, such as the width of the distribution and the degree of overlap between the distributions for the two time periods, and these will be discussed shortly.

The posterior means $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{b}}$ (the posterior mean is denoted by an overbar) are shown in Fig. 3 and $\overline{\boldsymbol{T}_{50}}$ and $\overline{\boldsymbol{z}_{30}}$ are shown in Fig. 4 for the entire spatial domain. Also shown in both Figures are the corresponding estimates from observations obtained by fitting the Gumbel distribution directly to the observed annual maxima using maximum likelihood estimation. We can see that the general spatial pattern and magnitude of the Gumbel parameter estimates from the BHM for P_1 (Fig. 3b and e) are similar to those derived directly from the observations (Fig. 3a and d). Given estimations of $\bar{\boldsymbol{a}}'$ and $\bar{\boldsymbol{b}}'$ (Fig. 3c and f) and $\overline{\boldsymbol{T}_{50}'}$ and $\overline{\boldsymbol{z}_{30}'}$ (Fig. 4c and f) it may also be of interest to compare the change in these quantities between the two time periods. The change in 50-year return levels can be estimated by the difference in posterior means $\Delta \boldsymbol{z}_{50} = \overline{\boldsymbol{z}_{50}'} - \overline{\boldsymbol{z}_{50}}$ (see Oliver et al. (in press) and Fig. 5a). Similarly, one can estimate the change in 30 °C return period by $\Delta \boldsymbol{T}_{30} = \overline{\boldsymbol{T}_{30}} - \overline{\boldsymbol{T}_{30}'}$ (not shown).

The posterior samples provide more information than simply the posterior mean: higher order statistics can be calculated such as the posterior variance or skewness. These can be used to provide error margins on the estimated quantities. For example, we can calculate the lower and upper bounds of the 95% credible intervals (the 2.5th and 97.5th percentiles respectively) from the posterior samples for $\boldsymbol{z}_{50}$ and $\boldsymbol{z}_{50}'$ (Fig. 6a, b, d, and e). The distributions of $\boldsymbol{z}_{50}$ and $\boldsymbol{z}_{50}'$ are generally unimodal and symmetric about the mean (see Fig. 2c) and so the width of the 95% credible interval is a useful summary of the width of the distribution about the posterior mean (i.e., model uncertainty; Fig. 6c and f). We can see that model uncertainty tends to be large over much of the Tasman Sea (south of 30°S) and larger near the Australian coast than offshore. This pattern is very similar to where the Gumbel distribution scale parameter $\boldsymbol{b}$ is also large (Fig. 3e and f). This is expected since the uncertainty on estimates of long return-period extremes, independent of whether or not Bayesian inference has been used to estimate the parameters, increases strongly with increasing $\boldsymbol{b}$ (and weakly with increasing $\boldsymbol{a}$; e.g., Coles (2001)).

4.3. Probabilities of exceedance

We can also calculate quantities that characterize the uncertainty of change between the two time periods, such as "probabilities of exceedance" (e.g., Shaby and Reich (2012)), by examining the posterior samples directly. For example, it is of interest to

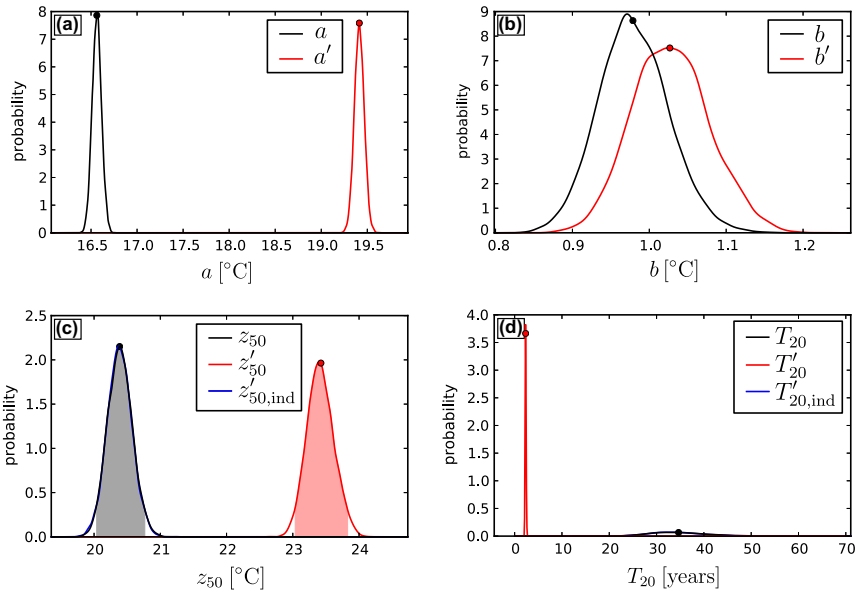

Fig. 2. The posterior distributions of Gumbel parameters a and b, 50-year return levels z_{50} and 20 °C return periods T_{20} at the location (152°E, 43°S) for the 1990s (black lines) and the 2060s (red lines). The filled circles indicate the position of the posterior mean. The shaded area in panel (c) indicates the 95% credible interval for z_{50} and z'_{50}. The blue lines in panels (c) and (d) indicate the independent samples of z_{50} and T_{20} drawn for the 1990s climate (note that these lines are nearly indistinguishable from the black lines since they are approximations of the same probability distribution). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

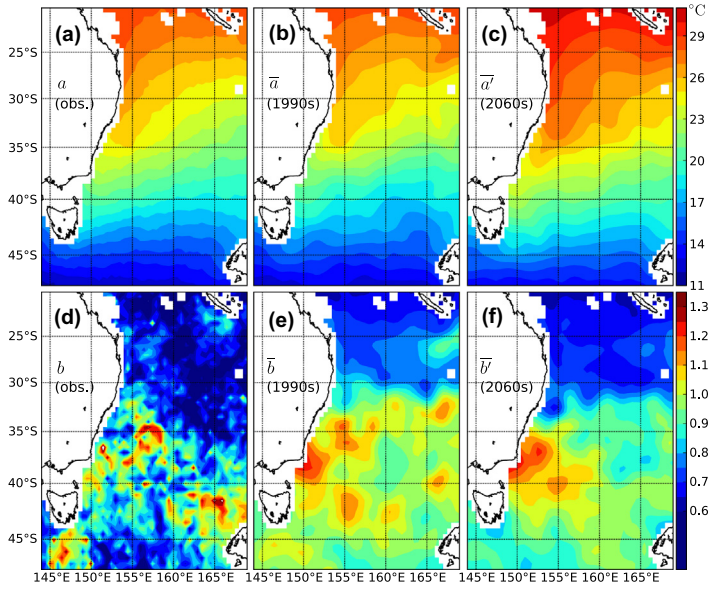

Fig. 3. The Gumbel distribution parameters $\boldsymbol{a}$ and $\boldsymbol{b}$ for the observations (a and d) and posterior means $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{b}}$ from the Bayesian hierarchical model for the 1990s (b and e) and the 2060s (c and f) climates.

know what the probability is that the change of the 50-year return level between P_1 and P_2 will exceed a threshold z^*. This can be calculated by taking the difference of the two posterior samples, $\boldsymbol{z}'_{50} - \boldsymbol{z}_{50}$ and examining the probability that this difference exceeds z^* : prob($\boldsymbol{z}'_{50} - \boldsymbol{z}_{50} > z^*$), which is simply the proportion of the probability distribution of $\boldsymbol{z}'_{50} - \boldsymbol{z}_{50}$ which falls above z^*. The probability distribution of $z'_{50} - z_{50}$ at the location (152°E, 43°S) is shown in Fig. 7a (black curve) and it can be seen that there is ∼100% chance that it will exceed 2 °C, ∼ 50% chance that it will exceed 3 °C, and essentially no chance that it will exceed 4 °C. A similar calculation can be done at the same location for the change in return periods $T'_{20} - T_{20}$ (the probability distribution is shown in Fig. 7b, black curve). The probability that $T'_{20} - T_{20}$ will decrease by at least T^* is the proportion of the probability distribution that lies below T^*. Thus, there is ∼99% chance that T_{20} will reduce by at least 20 years, 61% chance that it will reduce by at least 30 years, and 10% chance that it will reduce by at least 40 years. This information can also be presented in the opposite way: given a probability level, what is the estimated change in a particular variable. This is given simply by the percentiles of $z'_{50} - z_{50}$ or $T'_{20} - T_{20}$. For example, at a probability level of 95%, $z'_{50} - z_{50}$ will be at least 2.57 °C and $T'_{20} - T_{20}$ will change by at least 23.6 years.

We can also move away from the more abstract extreme value statistics and examine the extreme values themselves (i.e., the annual maxima). Given the posterior samples of the Gumbel parameters $\boldsymbol{a}$ and $\boldsymbol{b}$ for both time periods we can simulate annual

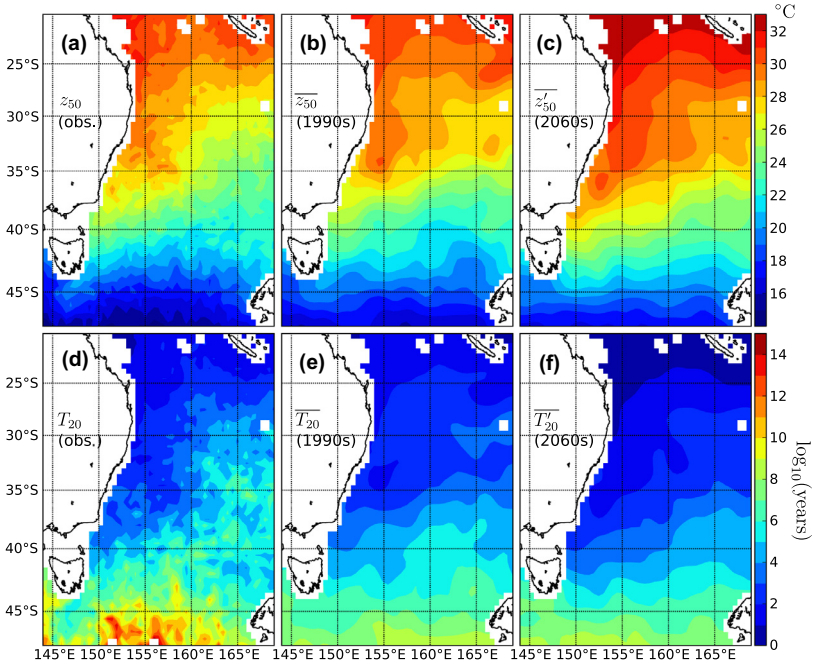

Fig. 4. The 50-year return levels $\boldsymbol{z}_{50}$ and 30 °C return periods $\boldsymbol{T}_{30}$ for the observations (a and d) and posterior means $\overline{\boldsymbol{z}_{50}}$ and $\overline{\boldsymbol{T}_{20}}$ from the Bayesian hierarchical model for the 1990s (b and e) and the 2060s (c and f) climates. The return periods, in years, have been log-transformed.

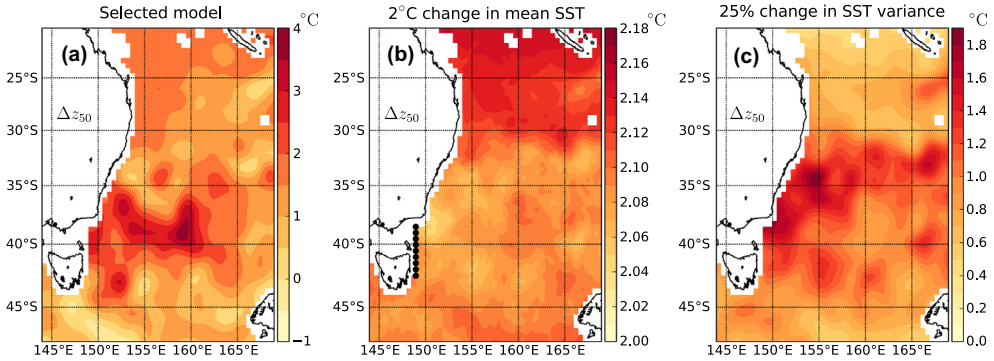

Fig. 5. The change in 50-year return level $\boldsymbol{z}_{50}$ between the 1990s and the 2060s using the model selected by Oliver et al. (in press) (a), for a specified 2 °C change in mean SST from the 1990s climate (b), and for a specified 25% change in SST variance from the 1990s climate (c). The black dots in panel (b) show the locations used for the analysis presented in Section 4.5.

maxima $\boldsymbol{Y}$ by drawing from the Gumbel distribution. Let the simulated annual maxima for P_1 and P_2 be denoted $\hat{\boldsymbol{Y}}$ and $\hat{\boldsymbol{Y}}'$, respectively. Their probability distributions are shown in Fig. 7c for the location (152°E, 43°S). The annual maxima may be a useful measure as they give an indication as to the extreme value that can occur in any particular year, rather than the more abstract (although also useful) measure of the return level at long return periods. We can now examine "probabilities of exceedance" for the annual maxima themselves as we did for the return levels and return periods. Fig. 7d (black curve) shows the probability distribution of $\boldsymbol{Y}' - \boldsymbol{Y}$ and, as above, we can read off the probabilities that the annual maxima will exceed a particular threshold Y^*, i.e., prob$(\hat{\boldsymbol{Y}}' - \hat{\boldsymbol{Y}} > Y^*)$. At this location, there is ~95% chance that the change in annual maxima will be positive, ~70% chance that the change will exceed 2 °C, and ~25% chance that it will exceed 4 °C.

4.4. Testing the significance of the estimated changes

In a climate change scenario such as this, it is important to be able to say how significant, in a statistical sense, the estimated changes in extreme value statistics are. One way of testing this is to draw a secondary sample from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X})$ given the climate in P_1 (i.e., the present day) and compare the changes between it and the original sample drawn from the same distribution. This measures the change that can occur simply due to randomness in the same climate and we define a *significant change* between P_1 and P_2 as one which exceeds that due to randomness at a specified confidence probability. This information allows us to quantify the uncertainty in the model simulations of climate change. This approach is philosophically similar to the method described by Theiler and Prichard (1996) of hypothesis testing by generating "surrogate" data, with the statistical properties of some null hypothesis, and has been used to test for the statistical significance of climate reconstructions (Christiansen et al., 2009) and of observed changes in air temperature extremes (Christiansen, 2013).

Let $\boldsymbol{z}_{50,\text{ind}}, \boldsymbol{T}_{20,\text{ind}}$ and $\hat{\boldsymbol{Y}}_{\text{ind}}$ denote the second (independent) samples drawn from the same distribution that $\boldsymbol{z}_{50}, \boldsymbol{T}_{20}$ and $\hat{\boldsymbol{Y}}$ were drawn from (i.e., the 1990s climate). Their probability distributions, at the location (152°E, 43°S), are shown in Fig. 2(c and d)

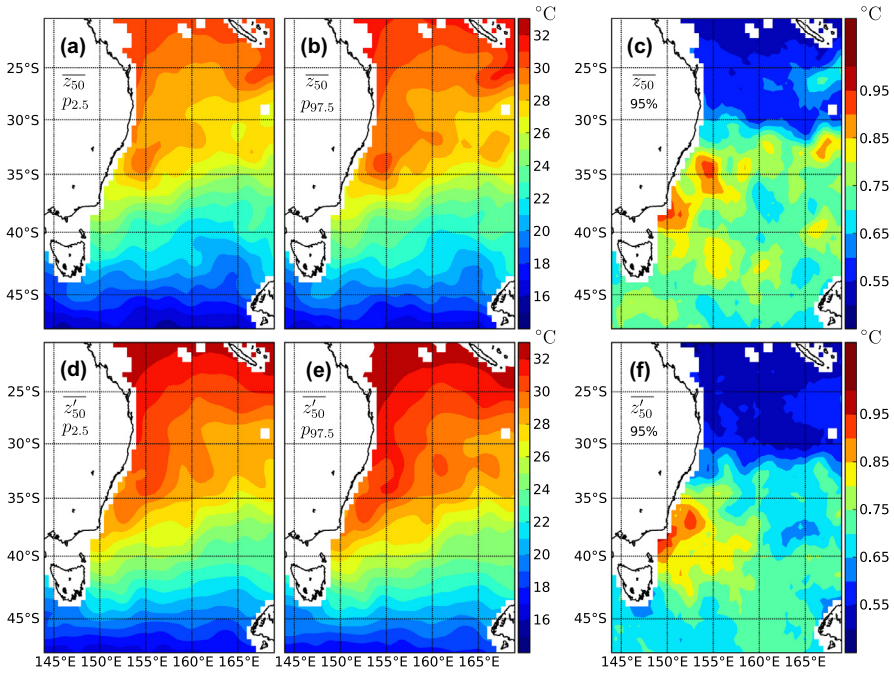

Fig. 6. The 95% credible interval for $\boldsymbol{z}_{50}$ (the 1990s; a–c) and $\boldsymbol{z}'_{50}$ (the 2060s; d–f). The 2.5th (a and d) and 97.5th (b and e) percentiles of the 50-return levels show the lower and upper bounds of the 95% credible interval. The width of the 95% credible interval is shown in (c) and (f) for the 1990s and 2060s respectively. The 95% credible intervals for $\boldsymbol{z}_{50}$ and $\boldsymbol{z}'_{50}$ at the location (152°E, 43°S) are shown in Fig. 2c.

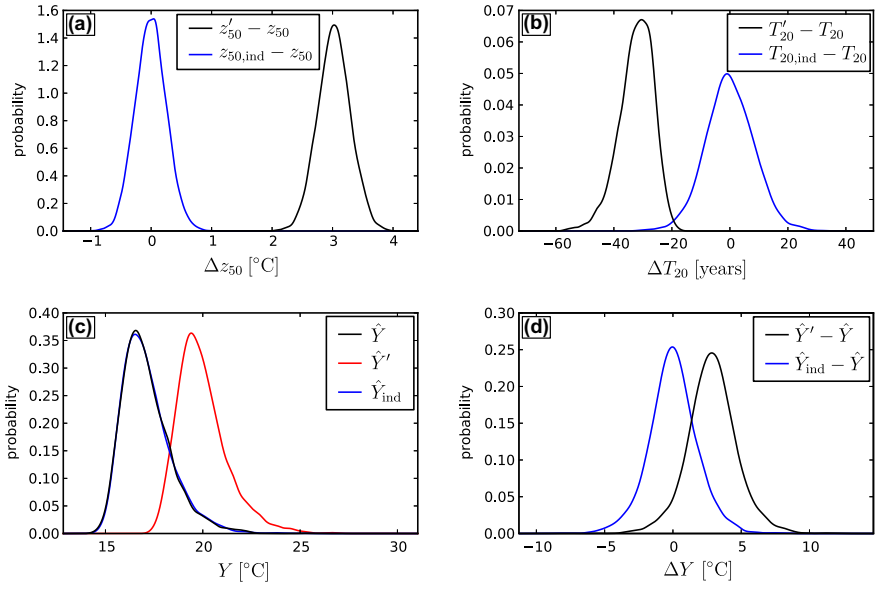

Fig. 7. Posterior distributions of the change in 50-year return level Δz_{50}, the change in 20 °C return period ΔT_{20}, the annual maxima Y, and the change in annual maxima ΔY at the location (152°E, 43°S). In panels (a, b, d), the black lines indicate the distributions of change between the 1990s and the 2060s and the blue lines indicate the distributions of change between two samples given the same 1990s climate. In panel (c), the black and blue lines indicate the distributions of annual maxima from the two samples drawn given the 1990s climate; the red line indicates the distribution of annual maxima drawn given the 2060s climate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and Fig. 7c (blue[3] lines). As described above, the probability that the change in 50-year return levels between P_1 and P_2 will exceed z^* is given by $\mathrm{prob}(\boldsymbol{z}'_{50} - \boldsymbol{z}_{50} > z^*)$; the probabilities that this change can occur by chance in the same climate is given by $\mathrm{prob}(\boldsymbol{z}_{50,\mathrm{ind}} - \boldsymbol{z}_{50} > z^*)$ and $\mathrm{prob}(\boldsymbol{z}_{50} - \boldsymbol{z}_{50,\mathrm{ind}} > z^*)$.[4] If $\mathrm{prob}(\boldsymbol{z}'_{50} - \boldsymbol{z}_{50} > z^*)$ is greater than $\mathrm{prob}(\boldsymbol{z}_{50,\mathrm{ind}} - \boldsymbol{z}_{50} > z^*)$ or greater than $\mathrm{prob}(\boldsymbol{z}_{50} - \boldsymbol{z}_{50,\mathrm{ind}} > z^*)$ then one can say that the probability of the specified change in z_{50} between P_1 and P_2 is significant. Similarly, we can invert this and test if the estimated change in z_{50} is significant with confidence p^* (or, equivalently, at the $1 - p^*$ significance level). This is equivalent to asking the question: does the value of z^* given $\mathrm{prob}(\boldsymbol{z}'_{50} - \boldsymbol{z}_{50} > z^*) = p^*$ exceed the value of z^* given $\mathrm{prob}(\boldsymbol{z}_{50,\mathrm{ind}} - \boldsymbol{z}_{50} > z^*) = p^*$ or $\mathrm{prob}(\boldsymbol{z}_{50} - \boldsymbol{z}_{50,\mathrm{ind}} > z^*) = p^*$?

The probability distributions of the changes due to randomness in an unchanged climate ($z_{50,\mathrm{ind}} - z_{50}$, $T_{20,\mathrm{ind}} - T_{20}$, and $\hat{Y} - \hat{Y}_{\mathrm{ind}}$) at the location (152°E, 43°S) are shown in Fig. 7(a, b, and d, blue

[3] For interpretation of color in Fig. 7, the reader is referred to the web version of this article.

[4] Which is equivalent to $\mathrm{prob}(\boldsymbol{z}_{50,\mathrm{ind}} - \boldsymbol{z}_{50} < -z^*)$.

lines). If there is no overlap between the changes between periods P_1 and P_2 and the changes from the two samples representing period P_1 (compare black and blue lines in Fig. 7a, b, and d) then we can say that, with 100% probability, the estimated changes between periods P_1 and P_2 exceed those due to randomness in an unchanged climate, and are thus significant. This is the case for z_{50} (Fig. 7a) but not for the other statistics as there is overlap in the distributions of ΔT_{20} and in the distributions of ΔY. At a probability level of 95%, the estimated changes in T_{20} between P_1 and P_2 is -23.5 years which is larger than a change of -14 years for the two samples drawn from the P_1 climate. Therefore, the change is significant at this confidence level (95%). However, at a confidence level of 99.5%, the change in return period between P_1 and P_2 is not significant (-20 years, compared against a change due to randomness of -23.8 years). For annual maxima, the situation is more dramatic. At a confidence level of 50% the estimated change in annual maxima between P_1 and P_2 is 2.8 °C, which is significant against a change due to randomness of 0.05 °C. At a confidence level of 75% the change is still significant (1.78 °C versus 1.13 °C due to randomness). However, at a confidence level of 90% the change is no longer significant (0.67 °C versus 2.16 °C due to randomness). We have illustrated these statistics at a point location but a convenient way to summarize the information would be to specify several confidence levels (say 50%, 75%, 90% and 95%) and map the significance of the estimated change in space (see Oliver et al. (in press)).

4.5. Simulating extremes from specified climates

So far we have examined the changes in extreme value statistics between two marine climates simulated by an ocean general circulation model with forcing derived from a global climate model. What if we asked the question: how will the extreme value statistics change from the present (i.e., P_1) if the mean SST increased by a specified amount? Or if the SST variance changed by a certain proportion? These questions can be addressed directly by constructing a new covariate matrix $\boldsymbol{X}_{\text{spec}}$, containing the user specified marine climate, and then drawing samples from the posterior distribution of the model parameters given this climate and then calculating the extreme value statistics. The above examples can be constructed by taking the covariate matrix $\boldsymbol{X}$ for the time period P_1 and modifying the columns containing the mean or the variance accordingly to give $\boldsymbol{X}_{\text{spec}}$ and then sampling from $p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X}_{\text{spec}})$ given the model parameters fitted to $\boldsymbol{X}$.

The change in posterior mean $\boldsymbol{z}_{50}$ for two specified climates can be seen in Fig. 5 along with the corresponding change from the model selected by Oliver et al. (in press). It can be seen that a change in the mean SST of 2 °C will change $\boldsymbol{z}_{50}$ by 2.0–2.2 °C everywhere in a nearly uniform way (Fig. 5b); although the response to a change in mean SST is slightly stronger (by 5–10%) in the tropical Coral Sea than in the midlatitude Tasman Sea. On the other hand, a 25% increase in SST variance leads to increases of $\boldsymbol{z}_{50}$ (1–2 °C, Fig. 5c) over the eddy-rich region of elevated SST variance that is found south of the East Australian Current separation point (~33°S; e.g., Godfrey et al. (1980), Stammer (1997), and Suthers et al. (2011)). These two cases illustrate the utility of the model for examining the response of extreme events to specific changes in the marine climate.

4.6. Application of the extremes model to ecology

The coastal waters of southeastern Australia are experiencing significant ecological habitat changes due to changes in oceanography (warming waters, increasing salinity, etc.; e.g., Holbrook and Bindoff (1997), Ridgway (2007), and Johnson et al. (2011)). Certain species have temperature thresholds, below which they cannot survive, such as the spiny sea urchin (*Centrostephanus rodgersii*)

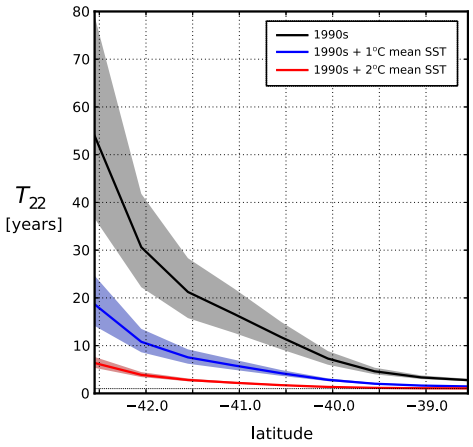

Fig. 8. Return periods for a 22 °C extreme SST event along a North–South section of the shelf break northeast of Tasmania for several specified climates. The return periods are shown for the 1990s climate (black), for the 1990s climate with a specified 1 °C increase in mean SST (blue), and for the 1990s climate with a specified 2 °C increase in mean SST (red). Shaded areas show the 95% credible intervals. The location of this section is shown Fig. 5 (black dots). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

which requires water temperatures $\geqslant 12$ °C for successful larval development (Ling et al., 2009). Even tropical fish species have been moving into the traditionally cold waters around Tasmania to the extent that the local community has been asked to be involved in helping to identify previously unknown fish species that are being found along the coast (www.redmap.org.au). With these changes in mind we have used the extremes model to estimate how warm temperature extremes would change along the Tasmanian shelf break for a specified change in climate. We chose an extreme temperature of 22 °C to represent a suitably warm event that would be rare in Tasmanian waters at present. The return periods T_{22} along a strip of the shelf break northeast of Tasmania (see black dots in Fig. 5) are shown as a function of latitude in Fig. 8 for the 1990s climate (black line), for the 1990s climate with a 1 °C increase in mean SST (blue line), and for the 1990s climate with a 2 °C increase in mean SST (red line). The change in return period is clearly a nonlinear function of the changes in mean SST, i.e., doubling the change in mean SST from 1 °C to 2° more than halves the change in T_{22}. A 22 °C extreme SST event, which is rare in the 1990s climate (T_{22} ranging from 5 to 60 years, depending on latitude), is estimated become merely uncommon ($T_{22} < 20$ years) under a 1 °C change in mean SST and common ($T_{22} < 10$ years) under a 2 °C change in mean SST.

5. Summary and discussion

We have described and demonstrated a technique, based on a Bayesian hierarchical model (BHM), for improving predictions of extreme values from global climate and ocean models. The BHM allows for multi-layer modeling of observed extremes. In the first layer, the observed extreme values are modeled using an extreme value distribution (EVD; e.g., annual maxima are modeled using a Gumbel distribution). In the second layer, the parameters of the EVD are modeled as a function of a set of covariates (e.g., the climate statistics of the environmental variable of interest). There is a third layer which, due to the Bayesian approach taken, expresses prior information about the model parameters. The parameters of the BHM can be estimated by Markov chain Monte Carlo algorithms (MCMC; including the Metropolis rule and the Gibbs sampler) and different model configurations can be

compared using the Deviance Information Criteria (DIC). The model, once fit to observed data, can then be used to interpolate over locations or extrapolate to time periods for which there are no observations but there exists covariate information. This technique, as a form of bias correction, provides a framework for improving predictions of extremes from global climate model output.

The model utility was demonstrated by improving estimates of sea surface temperature extremes off southeast Australia from a downscaled global climate model simulation for the 1990s. The model was then used to estimate extremes for a projected 2060s climate. A more detailed analysis of the model estimates of extreme SSTs, from a climate and oceanography perspective, is provided by Oliver et al. (in press), and here we have described a suite of statistics that can be used to analyze and summarize the model results. Since we have taken a Bayesian approach, and sampled from the posterior distribution by MCMC, we can summarize model estimations by estimating posterior statistics. For example, the posterior mean return periods and return levels, and their differences between the 1990s and 2060s, are perhaps the most useful summaries. However, the width of the posterior distribution also provides information about model uncertainty, e.g., it can provide 95% credible intervals for estimated return periods and return levels.

The posterior samples from the two climates (e.g., the 1990s and the 2060s) can be used to estimate the probability distributions of the change in extreme value statistics. These distributions can be used to provide "probabilities of exceedance", i.e., the probability that the extreme value statistics will exceed a specified threshold. We also outlined a method for testing the statistical significance of these estimations. This was achieved by drawing an independent sample of extreme values statistics for the 1990s climate and testing if, at a specified significance level, the estimated change between the 1990s and 2060s exceeds that which might occur by randomness between two samples for the 1990s climate. This is useful information as it important to provide confidence levels when making climate change predictions since these predictions inherently involve much uncertainty.

Finally, the fitted BHM provides a model for testing the response of the extreme values to prescribed changes in the climate. By specifying changes to the climate for which the model was fit (e.g., changes in the mean or changes in the variability) we can estimate the associated change in extreme value statistics. Such a model can be useful for exploring the relationship between the central climate statistics and the extreme values of the climate system.

In many cases, it has been found that marine ecological stability is defined in terms of temperature thresholds. For example, and as noted above, it has been found that *Centrostephanus rodgersii*, a sea urchin that is invasive to the Tasmanian coastal environment, requires water temperatures in excess of 12 °C for successful larval development (Ling et al., 2009). Thus, it can be useful to map the 12 °C return level isopleth at return periods of interest to the application at hand. That being said, it is also of interest to know how the frequency (return period) of such a critical temperature event might change over time. This is the focus of the subsection on ecological impacts (Section 4.6), and hence why we have focused on both return levels (Figs. 1, 2, and4–7) and return periods (Figs. 1, 2, 4, 7 and 8).

With regards to model stationarity we have assumed that the BHM captures, to first order, a fundamental physical process that links the statistics of the extreme values to the central statistics of the climate system. Essentially, we posited that there exists a relationship between the extremes and climate variables $\boldsymbol{X}$, that can be expressed rather crudely as "extremes" $= f(\boldsymbol{X})$. This relationship expresses fundamental aspects of the climate system that do not change with time. The climate process model, which links the extreme value statistics to the climate variables, essentially takes a linear approximation to this, i.e., $f(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta} + O(\boldsymbol{X}^2)$, where the regression coefficients characterize the fundamental (stationary) physical process and the higher order terms are neglected. Since we have assumed $f(\boldsymbol{X})$ is stationary so too are the parameters of the climate process layer model (e.g., the regression coefficients $\boldsymbol{\beta}$).

We have taken a Bayesian approach in formulating the hierarchical model. The value of the Bayesian approach is threefold. First, Bayesian statistics provides a more comprehensive, ground-up framework for formulating statistical models than a frequentist approach (if one is willing to accept the idea of a prior). This axiomatic approach to statistics is appealing for climate scientists who have been educated in similarly axiomatic fields (e.g., physics). Second, fitting models such as the one outlined in this paper is more easily performed in a Bayesian context, as estimating parameters by maximum likelihood would be unnecessarily difficult due to integrals over bottom-level parameters. We would also need to make assumptions about normality by appealing to asymptotic results for maximum likelihood estimation. A Bayesian approach allows for the use of well-defined existing techniques (MCMC, Metropolis rule, Gibbs sampler) which provide a straightforward and unified approach to fitting this type of model and fits well into the existing literature (e.g., Cooley et al. (2007) and Friederichs and Thorarinsdottir (2012)). Finally, a Bayesian approach allows for uncertainty to carried through the entire calculation in a consistent fashion. The maximum likelihood approach invokes a large sample approximation and assumes the parameter estimates are multivariate and Normally distributed about the true parameters with a covariance determined by the Fisher Information (Chapter 2, Pawitan (2001)). The distribution of model results is in turn dependent upon this Normal approximation. In contrast, the Bayesian approach seeks to simultaneously calculate the true posterior distribution of both the parameters and the model estimations, and the distribution of the model parameters will be correctly reflected in the distribution of the model estimates. This lends itself well to the study of a changing climate as we can make probabilistic estimates and assess the significance of those estimates. The BHM is also easily extended to include uncertainty about observations and covariates (e.g., see Appendix C on observational uncertainty).

Acknowledgements

The authors would like to acknowledge funding from the Australian Research Council Super Science Fellowship (Grant No. FS110200029). This manuscript makes a contribution to objectives of the ARC Centre of Excellence for Climate System Science (ARC CoE CCS) and the CLIVAR-endorsed Southwest Pacific Ocean Circulation and Climate Experiment (SPICE). The authors would like to thank the two anonymous reviewers for their thorough and constructive reviews of this paper.

Appendix A. Conditional distributions for $\boldsymbol{\beta}_a, \boldsymbol{\beta}_\phi, \tau_a$ and τ_ϕ

In order to Gibbs sample for $\boldsymbol{\beta}_a, \boldsymbol{\beta}_\phi, \tau_a$ and τ_ϕ we need expressions for the conditional distributions of those parameters. The posterior distribution for $(\boldsymbol{\beta}_a, \tau_a)$ is given by

$$p(\boldsymbol{\beta}_a, \tau_a|\boldsymbol{a}) = p(\boldsymbol{a}|\boldsymbol{\beta}_a, \tau_a)p(\boldsymbol{\beta}_a, \tau_a), \tag{A.1}$$

$$= p(\boldsymbol{a}|\boldsymbol{\beta}_a, \tau_a)p(\boldsymbol{\beta}_a)p(\tau_a) \tag{A.2}$$

noting that $\boldsymbol{\beta}_a$ and τ_a are distributed independently (Eq. (17)).

The three terms in Eq. (A.2) are given by Eqs. (15), (18) and (20) respectively as:

Table 1
Model selection. The model number and the covariates included in the model are listed in the first and second column respectively. The third, fourth, and fifth columns indicate measures of model fit ($\overline{D}$), model complexity (p_D), and the Deviance Information Criteria (DIC $= \overline{D} + p_D$) scores respectively.

Model	Covariates included in model	$\overline{D}$	p_D	DIC
1	μ	165257	4608	169865
2	μ, σ^2	154443	4655	159098
3	μ, σ^2, m_3	152789	4649	157438
4	μ, σ^2, m_3, K	151965	4641	156606
5	$\mu, \sigma^2, m_3, K, \sigma_\eta^2$	151453	4648	156101
6	$\mu, \sigma^2, m_3, K, \sigma_\eta^2, m_4$	150970	4646	155616

$$p(\boldsymbol{a}|\boldsymbol{\beta}_a, \tau_a) = \left(\frac{\tau_a}{2\pi}\right)^{J/2} \exp\left[-\frac{\tau_a}{2}(\boldsymbol{a} - \boldsymbol{X}\boldsymbol{\beta}_a)^{\mathrm{T}}(\boldsymbol{a} - \boldsymbol{X}\boldsymbol{\beta}_a)\right], \tag{A.3}$$

$$p(\boldsymbol{\beta}_a) = \left(\frac{\tau_{a,0}}{2\pi}\right)^{1} 2\exp\left[\frac{\tau_{a,0}}{2}(\boldsymbol{\beta}_a - \boldsymbol{\mu}_a)^{\mathrm{T}}(\boldsymbol{\beta}_a - \boldsymbol{\mu}_a)\right], \tag{A.4}$$

$$p(\tau_a) = \left(\frac{\tau_a}{\theta_a}\right)^{\alpha_a - 1} \exp(-\tau_a/\theta_a). \tag{A.5}$$

Combining these terms as in Eq. (A.2) and isolating the portions dependent on $\boldsymbol{\beta}_a$ and τ_a respectively, and simplifying, leads to the following conditionals

$$p(\boldsymbol{\beta}_a|\tau_a, \boldsymbol{a}) \propto \mathcal{N}_M\left(\boldsymbol{T}(\tau_a \boldsymbol{X}^{\mathrm{T}}\boldsymbol{a} + \tau_{a,0}\boldsymbol{\beta}_0), \boldsymbol{T}^{-1}\right), \tag{A.6}$$

$$p(\tau_a|\boldsymbol{\beta}_a, \boldsymbol{a}) \propto \mathrm{Gamma}\left(\alpha_a + \frac{J}{2}, (\theta_a^{-1} + SS/2)^{-1}\right), \tag{A.7}$$

where $\boldsymbol{T} = (\tau_a \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \tau_{a,0}\mathbf{I})^{-1}$ and $SS = (\boldsymbol{a} - \boldsymbol{X}\beta_{\boldsymbol{a}})^{\mathrm{T}}(\boldsymbol{a} - \boldsymbol{X}\beta_{\boldsymbol{a}})$. Equivalent expressions exist for the conditionals $p(\boldsymbol{\beta}_\phi|\tau_\phi, \boldsymbol{\phi})$ and $p(\tau_\phi|\boldsymbol{\beta}_\phi, \boldsymbol{\phi})$.

Appendix B. Model selection and validation

In this appendix we outline and summarize the model selection and comparison of observed and modeled extreme SSTs performed by Oliver et al. (in press).

B.1. Model selection

We systematically tested a set of linear regression models, as the climate process layer in the BHM, in order to determine which model provided the best fit. Each model expressed the parameters of the Gumbel distribution (a and ϕ) as a linear combination of covariates and the candidate covariates included mean SST (μ), SST variance (σ^2), SST third moment (m_3), SST fourth moment (m_4), eddy kinetic energy (K), and sea level variance (σ_η^2). Model selection was performed using a stepwise procedure. The DIC scores (see Section 3.3) were used to determine the best model at each step with the lowest value of DIC preferred (analogous to the correlation coefficient in a stepwise regression). Each step of the stepwise procedure is presented as a row in Table 1. The best model was determined to be Model 4 which includes μ, σ^2, m_3, and K. Note that this model did not have the lowest DIC score overall – the models which included m_4 and σ_η^2 had lower DIC scores. However, m_4 and σ_η^2 were highly correlated with σ^2 and K respectively (correlation coefficients of 0.96 and 0.93 respectively; the remaining covariates had correlation coefficients between 0.05 and 0.64) and so they do not provide additional independent predictive information. Furthermore, the model complexity p_D has reached a minimum value for Model 4 (with the exception of Model 1 which includes only μ) and only increases with the inclusion of m_4 and σ_η^2 while the concurrent reduction in DIC scores are negligible. Given that DIC scores have a bias towards overfit models (Ando, 2007), and that later we assume stationarity of the model parameters in order to perform the temporal extrapolation and

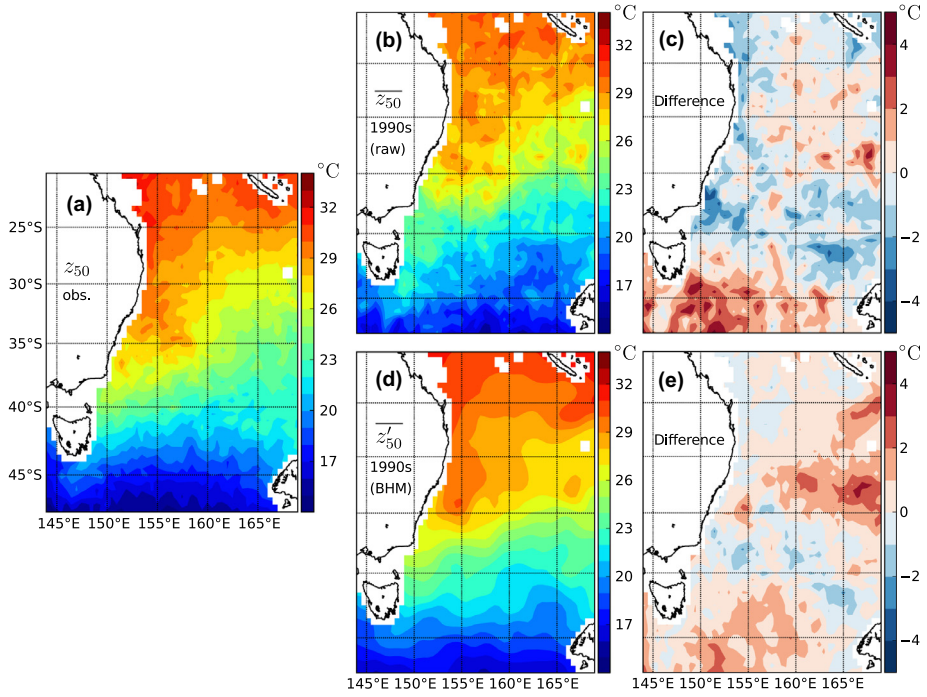

Fig. 9. 50-year return levels (z_{50}) for observed extreme sea surface temperatures (SSTs) and from the ocean model both directly and using the Bayesian hierarchical model (BHM). The z_{50} were estimated using maximum likelihood fits of the Gumbel distribution to annual SST maxima from (a) observations and (b) the ocean model simulation for the 1990s; (d) the z_{50} were also estimated using the BHM on the 1990s ocean model. The difference in the estimates of z_{50} are shown between (c) the observed extremes and the direct ocean model estimates and (e) the observed extremes and the results after using BHM on the the 1990s simulation.

spatial interpolation of extremes, we consider it best to choose a simpler model rather than a possibly overfit model.

B.2. Comparison of observed and modeled extremes

The Gumbel parameters a and ϕ for the observed and ocean model extreme SSTs were determined by fitting the Gumbel distribution directly to the observed annual maxima $\boldsymbol{Y}$ by maximum likelihood estimation and for the BHM by taking the posterior mean of samples drawn from the model selected above (Gumbel parameters from observations and the BHM can be found in Fig. 2). The z_{50} were then estimated using Eq. (4). The z_{50} derived from observations show a strong meridional gradient along with a zonal gradient in the region influenced by the East Australian Current (EAC) separation (25°S and 38°S; Fig. 9a).

The 50-year return levels (z_{50}) derived by fitting the Gumbel distribution to the ocean model SST annual maxima (Fig. 9b) are generally similar to the observations (Fig. 9a). Both the strong meridional gradient over the entire domain and the zonal gradient in the vicinity of the EAC separation are present. However, the zonal gradient in the EAC separation region is less well-defined than in the observations (Fig. 9c). The differences are generally less than ± 3 °C over most of the domain but can be as high as ± 4 °C in some places (e.g., southeast of Tasmania, off the southeast cost of mainland Australia, northwest of New Zealand's South Island, and in the central eastern portion of the domain).

The z_{50} derived from the BHM compare well against those derived from the observations. The modeled z_{50} (Fig. 9d) exhibit a very similar pattern to the observed z_{50} (Fig. 9a) including a strong meridional gradient over the Tasman Sea and Coral Sea as well as a strong zonal gradient in the region influenced by the EAC separation. The differences between the observed and model estimates of z_{50} are generally less than ± 2 °C (Fig. 9e) and are less than ± 1 °C over most of the Tasman Sea and adjacent to the continental shelfs. However, the differences exceed ± 2 °C in some regions including southeast of Tasmania as well in a tongue between about 30°S and 37°S. Overall, the BHM appears to capture the underlying climate process linking the covariates (climate variables) and the extremes over much of the Tasman Sea, Coral Sea, and near the continental shelf, and provides an improvement over the extremes derived directly from the ocean model.

Appendix C. Including uncertainty about observations in the Bayesian hierarchical model

The model described in Section 3 assumes that our data $\boldsymbol{Y}$ are perfectly accurate, i.e., they are "true observations" without measurement error. Let us posit that there exist true observations, denoted $\boldsymbol{Y}_{\text{tr}}$, and that $\boldsymbol{Y}$ are in fact imperfect observations of $\boldsymbol{Y}_{\text{tr}}$. Imperfect observation may be due to measurement (i.e., instrument) error or due to missing values in the record. Let us quantify the observation error as $p(\boldsymbol{Y}|\boldsymbol{Y}_{\text{tr}})$, i.e., the probability of imperfect observations $\boldsymbol{Y}$ given the true observations $\boldsymbol{Y}_{\text{tr}}$. This is often known explicitly, or can be calculated empirically, given the known limitations of the observation equipment, etc. For example, given an error margin of a measurement device one might assume $p(\boldsymbol{Y}|\boldsymbol{Y}_{\text{tr}})$ is a Normal distribution with mean $\boldsymbol{Y}_{\text{tr}}$ and variance informed by the error margin.

The uncertainly about the observations can be easily incorporated into the framework of the Bayesian hierarchical model. This is done by adding another layer to the hierarchy, what we may call the "observational error" layer, as another term in Relation 9

$$p(\boldsymbol{\theta}, \boldsymbol{Y}_{\text{tr}}|\boldsymbol{Y}, \boldsymbol{X}) \propto p(\boldsymbol{Y}|\boldsymbol{Y}_{\text{tr}})\ p(\boldsymbol{Y}_{\text{tr}}|\boldsymbol{\theta}_2)\ p(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \boldsymbol{X})\ p(\boldsymbol{\theta}_1). \tag{C.1}$$

Relation C.1 now includes the observation error probability term $p(\boldsymbol{Y}|\boldsymbol{Y}_{\text{tr}})$, and the true observations $\boldsymbol{Y}_{\text{tr}}$ are simply another parameter to be estimated alongside $\boldsymbol{\theta}$. Note that the "data layer" $p(\boldsymbol{Y}_{\text{tr}}|\boldsymbol{\theta}_2)$ is now expressed as an extreme value model for the (unobserved) true observations rather than the (measured) imperfect observations.

References

Abeysirigunawardena, D.S., Gilleland, E., Bronaugh, D., Wong, P., 2009. Extreme wind regime responses to climate variability and change in the inner south coast of British Columbia, Canada. Atmosphere-Ocean 47 (1), 41–62.

Ando, T., 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. Biometrika 94 (2), 443–458.

Ballester, J., Giorgi, F., Rodó, X., 2010. Changes in European temperature extremes can be predicted from changes in PDF central statistics. Climatic change 98 (1), 277–284.

Berliner, L.M., Royle, J.A., Wikle, C.K., Milliff, R.F., 1998. Bayesian methods in the atmospheric sciences. Bayesian Statistics 6, 83–100.

Berliner, L.M., Milliff, R.F., Wikle, C.K., 2003. Bayesian hierarchical modeling of Air–Sea interaction. Journal of Geophysical Research: Oceans 108 (C4).

Casson, E., Coles, S., 1999. Spatial regression models for extremes. Extremes 1 (4), 449–468.

Chamberlain, M.A., Sun, C., Matear, R.J., Feng, M., Phipps, S.J., 2012. Downscaling the climate change for oceans around Australia. Geoscientific Model Development Discussions 5, 425–458.

Christiansen, B., 2013. Changes in temperature records and extremes: are they statistically significant. Journal of Climate 26, 7863–7875.

Christiansen, B., Schmith, T., Thejll, P., 2009. A surrogate ensemble study of climate reconstruction methods: stochasticity and robustness. Journal of Climate 22 (4), 951–976.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer Verlag.

Cooley, D., Nychka, D., Naveau, P., 2007. Bayesian spatial modeling of extreme precipitation return levels. Journal of the American Statistical Association 102 (479), 824–840.

Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-temporal Data. Wiley, Hoboken, New Jersey.

Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical modeling of spatial extremes. Statistical Science 27 (2), 161–186.

de Vries, H., Haarsma, R.J., Hazeleger, W., 2012. Western European cold spells in current and future climate. Geophysical Research Letters 39 (4), L04706.

Fiechter, J., Herbei, R., Leeds, W., Brown, J., Milliff, R., Wikle, C., Moore, A., Powell, T., 2013. A Bayesian parameter estimation method applied to a marine ecosystem model for the coastal Gulf of Alaska. Ecological Modelling 258, 122–133.

Field, C.B., Barros, V., Stocker, T.F., Qin, D., Dokken, D.J., Ebi, K.L., Mastrandrea, M.D., Mach, K.J., Plattner, G.K., Allen, S.K., et al., 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Cambridge University Press, Cambridge.

Friederichs, P., Thorarinsdottir, T.L., 2012. Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. Environmetrics 23 (7), 579–594.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis. Chapman & Hall/CRC.

Godfrey, J.S., Cresswell, G.R., Golding, T.J., Pearce, A.F., Boyd, R., 1980. The separation of the East Australian Current. Journal of Physical Oceanography 10 (3), 430–440.

Griffiths, G.M., Chambers, L.E., Haylock, M.R., Manton, M.J., Nicholls, N., Baek, H.J., Choi, Y., Della-Marta, P.M., Gosai, A., Iga, N., et al., 2005. Change in mean temperature as a predictor of extreme temperature change in the Asia–Pacific region. International Journal of Climatology 25 (10), 1301–1330.

Gumbel, E.J., 1958. Statistics of Extremes. Columbia University Press, New York.

Harmon, R., Challenor, P.G., 1997. A Markov chain Monte Carlo method for estimation and assimilation into models. Ecological Modelling 101 (1), 41–59.

Higdon, D., 1998. A process-convolution approach to modelling temperatures in the North Atlantic Ocean. Environmental and Ecological Statistics 5 (2), 173–190.

Holbrook, N.J., Bindoff, N.L., 1997. Interannual and decadal temperature variability in the southwest Pacific Ocean between 1955 and 1988. Journal of Climate 10 (5), 1035–1049.

Hooten, M.B., Leeds, W.B., Fiechter, J., Wikle, C.K., 2011. Assessing first-order emulator inference for physical parameters in nonlinear mechanistic models. Journal of Agricultural, Biological, and Environmental Statistics 16 (4), 475–494.

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) 52 (1), 105–124.

Hosking, J.R.M., 1992. Moments or L moments? An example comparing two measures of distributional shape. The American Statistician 46 (3), 186–189.

Johnson, C.R., Banks, S.C., Barrett, N.S., Cazassus, F., Dunstan, P.K., Edgar, G.J., Frusher, S.D., Gardner, C., Haddon, M., Helidoniotis, F., et al., 2011. Climate change cascades: shifts in oceanography, species' ranges and subtidal marine community dynamics in eastern Tasmania. Journal of Experimental Marine Biology and Ecology 400 (1), 17–32.

Katz, R.W., 2010. Statistics of extremes in climate change. Climatic Change 100 (1), 71–76.

Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. Extremes and Related Properties of Random Sequences and Processes, vol. 11. Springer Verlag.

Leeds, W.B., Wilkie, C.K., Fietcher, J., Brown, J., Milliff, R.F., 2013. Modeling 3-D spatio-temporal biogeochemical processes with a forest of 1-D statistical emulators. Environmetrics 24, 1–12.

Lemos, R.T., Sansó, B., 2009. A spatio-temporal model for mean, anomaly, and trend fields of North Atlantic sea surface temperature. Journal of the American Statistical Association 104 (485), 5–18.

Lemos, R.T., Sansó, B., Santos, F.D., 2010. Hierarchical Bayesian modelling of wind and sea surface temperature from the Portuguese coast. International Journal of Climatology 30 (9), 1423–1430.

Ling, S.D., Johnson, C.R., Ridgway, K., Hobday, A.J., Haddon, M., 2009. Climate-driven range extension of a sea urchin: inferring future trends by analysis of recent population dynamics. Global Change Biology 15 (3), 719–731.

Mearns, L.O., Katz, R.W., Schneider, S.H., 1984. Extreme high-temperature events: changes in their probabilities with changes in mean temperature. Journal of Climate and Applied Meteorology 23 (12), 1601–1613.

Milliff, R.F., Bonazzi, A., Wikle, C.K., Pinardi, N., Berliner, L.M., 2011. Ocean ensemble forecasting. Part I: Ensemble Mediterranean winds from a Bayesian hierarchical model. Quarterly Journal of the Royal Meteorological Society 137 (657), 858–878.

Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grubler, A., Jung, T.Y., Kram, T., et al., 2000. Special Report on Emissions Scenarios: A Special Report of Working Group iii of the Intergovernmental Panel on Climate Change. Tech. Rep., Pacific Northwest National Laboratory, Richland, WA (US), Environmental Molecular Sciences Laboratory (US).

Oliver, E.C.J., Wotherspoon, S.J., Holbrook, N.J., in press. Projected Tasman Sea extremes in sea surface temperature through the 21st Century. Journal of Climate, http://dx.doi.org/10.1175/JCLI-D-13-00259.1.

Pawitan, Y., 2001. In All Likelihood: Statistical Modelling and Inference Using Likelihood. OUP, Oxford.

Pearce, A.F., Feng, M., 2012. The rise and fall of the marine heat wave off Western Australia during the summer of 2010/11. Journal of Marine Systems, 139–156.

Perkins, S.E., Pitman, A.J., Holbrook, N.J., McAneney, J., 2007. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. Journal of Climate 20 (17), 4356–4376.

Ridgway, K.R., 2007. Long-term trend and decadal variability of the southward penetration of the East Australian Current. Geophysical Research Letters 34 (13), L13613.

Royle, J.A., Berliner, L.M., Wikle, C.K., Milliff, R., 1999. Case Studies in Bayesian Statistics 4. Springer, p. 367 (Chapter A Hierarchical Spatial Model for Constructing Wind Fields from Scatterometer Data in the Labrador Sea).

Sang, H., Gelfand, A.E., 2009. Hierarchical modeling for extreme values observed over space and time. Environmental and Ecological Statistics 16 (3), 407–426.

Sang, H., Gelfand, A.E., 2010. Continuous spatial process models for spatial extreme values. Journal of Agricultural, Biological, and Environmental Statistics 15 (1), 49–65.

Schliep, E.M., Cooley, D., Sain, S.R., Hoeting, J.A., 2010. A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. Extremes 13 (2), 219–239.

Shaby, B.A., Reich, B.J., 2012. Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. Environmetrics 23 (8), 638–648.

Sillmann, J., Croci-Maspoli, M., Kallache, M., Katz, R.W., 2011. Extreme cold winter temperatures in Europe under the influence of North Atlantic atmospheric blocking. Journal of Climate 24 (22), 5899–5913.

Simolo, C., Brunetti, M., Maugeri, M., Nanni, T., 2011. Evolution of extreme temperatures in a warming climate. Geophysical Research Letters 38 (16), L16701.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (4), 583–639.

Stammer, D., 1997. Global characteristics of ocean variability estimated from regional TOPEX/POSEIDON altimeter measurements. Journal of Physical Oceanography 27 (8), 1743–1769.

Sun, C., Feng, M., Matear, R.J., Chamberlain, M.A., Craig, P., Ridgway, K.R., Schiller, A., 2012. Marine downscaling of a future climate scenario for Australian boundary currents. Journal of Climate 25 (8), 2947–2962.

Suthers, I.M., Young, J.W., Baird, M.E., Roughan, M., Everett, J.D., Brassington, G.B., Byrne, M., Condie, S.A., Hartog, J.R., Hassler, C.S., et al., 2011. The strengthening East Australian Current, its eddies and biological effects – an introduction and overview. Deep Sea Research Part II: Topical Studies in Oceanography 58 (5), 538–546.

Theiler, J., Prichard, D., 1996. Constrained-realization Monte-Carlo method for hypothesis testing. Physica D: Nonlinear Phenomena 94 (4), 221–235.

Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., et al., 2005. The ERA-40 re-analysis. Quarterly Journal of the Royal Meteorological Society 131 (612), 2961–3012.

Wernberg, T., Smale, D.A., Tuya, F., Thomsen, M.S., Langlois, T.J., de Bettignies, T., Bennett, S., Rousseaux, C.S., 2012. An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. Nature Climate Change 3, 78–82.

Wigley, T.M.L., 1985. Climatology: impact of extreme events. Nature 316, 106–107.

Wikle, C.K., Milliff, R.F., Herbei, R., Leeds, W.B., in press. Modern statistical methods in oceanography: a hierarchical perspective. Statistical Science. Invited paper (2013). <http://www.imstat.org/sts/futurepapers.html urlwww.imstat.org/sts/futurepapers.html>.

Zhang, X., Wang, J., Zwiers, F.W., Groisman, P.Y., 2010. The influence of large-scale climate variability on winter maximum daily precipitation over North America. Journal of Climate 23 (11), 2902–2915.