

Projection and Analysis of Extreme Wave Climate

SOFIA CAIRES

Climate Research Branch, Meteorological Service of Canada, Toronto, Ontario, Canada, and Royal Netherlands Meteorological Institute, De Bilt, Netherlands

VAL R. SWAIL AND XIAOLAN L. WANG

Climate Research Branch, Meteorological Service of Canada, Toronto, Ontario, Canada

(Manuscript received 31 May 2005, in final form 30 January 2006)

ABSTRACT

The nonhomogeneous Poisson process is used to model extreme values of the 40-yr ECMWF Re-Analysis (ERA-40) significant wave height. The parameters of the model are expressed as functions of the seasonal mean sea level pressure anomaly and seasonal squared sea level pressure gradient index. Using projections of the sea level pressure under three different forcing scenarios by the Canadian coupled climate model, projections of the parameters of the nonhomogeneous Poisson process are made, trends in these projections are determined, return-value estimates of significant wave height up to the end of the twenty-first century are projected, and their uncertainties are assessed. The uncertainty of estimates associated with the nonhomogeneous Poisson process estimates is studied and compared with the homologous estimates obtained using a nonstationary generalized extreme value model.

1. Introduction

Ship design depends crucially on the knowledge of the most severe significant wave height (SWH) conditions that ships need to withstand during their lifetime, and thus requires reliable estimates of return values. The m -year return value of SWH is defined loosely as the value of SWH that is exceeded on average once every m years; typically of interest is the 20-yr return value, denoted by SWH_{20} , because ships are usually designed to last at least 20 yr. In addition to estimates, it is also important to obtain measures of their accuracy as well as projections as to how they may be affected by future climate changes.

Caires and Sterl (2005) have obtained return-value estimates of SWH data from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) (Uppala et al. 2005), a global reanalysis of meteorological variables from 1958 to 2001. The estimates were obtained by fitting a generalized Pareto distribution (GPD) and a Poisson model to data sampled using the peaks-over-threshold (POT)

method. The annual maxima (AM) method (see, e.g., Coles 2001), another statistically sound approach to return-value estimation in which the generalized extreme value (GEV) distribution is fitted to sample annual/seasonal maxima, was not considered by Caires and Sterl (2005) on the grounds that the rather small sample sizes it deals with yield less precise inferences.

Both of these methods assume the stationarity of extreme values. However, the results of Caires and Sterl (2005) show that return-value estimates based on data from different decades are not compatible with each other, particularly in certain regions of the North Atlantic (NA) and North Pacific (NP) oceans. The incompatibilities in the NP estimates point to a positive trend, and those in the NA estimates to decadal variability. This suggests that the estimation of return values of SWH with data from several decades should be based on a nonstationary model. The realization that the extreme wave climate is nonstationary also implies that current climate alone cannot be used to estimate future return values of SWH.

Another problem involved with the estimates of future SWH return values is that no projections of future SWH fields are available. Although different climate models have been run in various research and meteorological institutes to compute the evolution of the

Corresponding author address: Sofia Caires, Delft Hydraulics, P.O. Box 177, 2600 MH Delft, Netherlands.
E-mail: sofia.caires@wldelft.nl

present climate under different assumptions about the future rates of emission of greenhouse gases (GHGs), that is, under different future *forcing scenarios*, these models do not have coupled wave models and therefore provide no future SWH values. Furthermore, the resolution of the grids on which the computations are done is too low to enable the reliable use of the resulting wind fields to directly force wave models and eventually obtain SWH climate projections offline.

The reasonable reliability of climate projections of sea level pressure (SLP) provided by climate models and the knowledge (e.g., Wang and Swail 2001) that the SWH fields are highly correlated with SLP fields and can be reliably used to regress SWH fields on SLP fields, led Wang and Swail (2005, 2006) and Wang et al. (2004) to project return values of SWH, using a nonstationary generalized extreme value (NS-GEV) model with the parameters being dependent on covariates derived from climate model estimates of future SLP fields under different future forcing scenarios (Wang et al. 2004; Wang and Swail 2005, 2006).

There were three future forcing scenarios considered in the studies by Wang and Swail (2005, 2006) and Wang et al. (2004): a modified version of the Intergovernmental Panel on Climate Change (IPCC) IS92a scenario (Boer et al. 2000), from now on referred to as the IS92a forcing scenario, in which the emissions of GHG correspond to those observed up to 1990 and are then increased at a rate of $1\% \text{ yr}^{-1}$, and the Special Report on Emission Scenarios (SRES; Nakicenovic and Swart 2000), the A2 and B2 forcing scenarios, which also include aerosol forcings. The A2 scenario is similar to the IS92a scenario and the B2 scenario describes a slower future increase of GHG emissions.

Wang et al. (2004) fitted a NS-GEV model with parameters depending on SLP-derived variables to the 1958–97 SLP reanalysis dataset of the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR, Kalnay et al. 1996) and SWH data for the NA from the Swail and Cox (2000) wave dataset. Wang and Swail (2005) did the same analysis for the NP using the Cox and Swail (2001) SWH dataset. Assuming that the fitted model in each case is also valid under different climate scenarios, these authors used projections of SLP fields under the IS92a, A2, and B2 scenarios by the Canadian second Coupled Global Climate Model (CGCM2) from 1990 to 2100 to obtain estimates of SWH_{20} up to 2100 for the NA (Wang et al. 2004) and NP (Wang and Swail 2005).

Using the “observed” (i.e., ERA-40) relationship between SWH_{20} and SLP fields represented by a NS-GEV model, along with projections of SLP fields made with three climate models under three different forcing sce-

narios, Wang and Swail (2006) quantified the uncertainty in projections of SWH_{20} that arise from the differences among the climate models and among the forcing scenarios. They also estimated the multimodel mean projection of changes in SWH_{20} .

The objective of this work is to compare the results obtained using the nonstationary analog of the POT GPD approach with those obtained by Wang and Swail (2004, 2006) and Wang et al. (2004) using a NS-GEV approach. As hinted above, when modeling extremes in the stationary setting the POT GPD approach is often preferable to the AM GEV approach. This seems to be even more so the case in nonstationary settings, because we are dealing with even more parameters. Asymptotically the results should be the same, but because we are dealing with short time series it is important to compare and assess the relative uncertainty of the two approaches.

The datasets we will be using here are the global ERA-40 SWH and SLP descriptions of the present climate and the global CGCM2 SLP projections of future climate scenarios. Because Wang and Swail (2006) analyze only briefly the future global changes in values of SWH_{20} under different forcing scenarios, and only those based on the multiclimate model results, we will analyze here in some detail global changes in SWH_{20} under different forcing scenarios based on the CGCM2 projections.

In section 2 of this article we will describe the datasets used. In sections 3 and 4 we will present the nonstationary models for extremes and other statistical methods used in our analysis. In section 5 we analyze the results, and we finish with conclusions in section 6.

2. Datasets

In this work we shall assume that the ERA-40 dataset correctly describes the present climate, and that the relationships found between its SLP and extremes of SWH are also valid in the future, irrespective of forcing scenario. It should be pointed out that our conclusions are likely to be very sensitive to a departure from these assumptions; also, because the uncertainty resulting from any possible departure cannot be quantified, it is expected that the uncertainty in estimates of future return values of SWH will be underestimated.

From the ERA-40 dataset we will be using SWH and SLP from 1958 to 2001. The SWH data consist of fields of 6-hourly “observations” on a global 1.5° latitude \times 1.5° longitude grid. The SLP data consist of fields of monthly and seasonal means (computed from 6-hourly observations) on a global 2.5° latitude \times 2.5° longitude grid (which is the resolution of the grid on which a large

subset of the ERA-40 data is available online at http://data.ecmwf.int/data/d/era40_daily/). Squared gradients of monthly mean SLP fields were computed and averaged seasonally. Both the seasonal mean SLP and seasonally averaged squared gradients were reexpressed as anomalies relative to their 1961–90 baseline climates, the ERA-40 baseline climates; the resulting fields will be referred to as the seasonal mean SLP anomalies and seasonal SLP gradient indices and denoted by P and G , respectively.

Projections of future monthly and seasonal means of SLP were obtained from CGCM2 (Flato and Boer 2001) simulations for the IS92a scenario from 1961 to 2099, and for the A2 and B2 forcing scenarios from 1990 to 2099. For each scenario, an ensemble with three members is available, that is, for each scenario, three integrations with the same forcing but different initial conditions were run; the differences between the individual members of an ensemble are supposed to be entirely due to natural variability, meaning that each member can be considered as an independent realization of the same stochastic process. The data are on a global 96×48 Gaussian grid ($\approx 3.75^\circ \times 3.75^\circ$). Seasonal mean SLP anomalies relative to the IS92a scenario simulations of the 1961–90 baseline climate, the “simulations baseline climate,” were computed. Squared gradients of monthly mean SLP fields were computed from monthly mean SLP fields adjusted to the ERA-40 baseline climate (replacing the simulated baseline climate by the ERA-40 baseline climate) and averaged seasonally.

The P and G fields were calculated for each ERA-40 SWH 1.5° latitude $\times 1.5^\circ$ global longitude grid point from values at the four nearest SLP grid points using weights proportional to the inverse of the distance.

3. Nonstationary models for extremes

In this section we outline the two approaches to modeling extremes of nonstationary time series as well as the methods used for estimating model parameters. The first two subsections deal with the nonhomogeneous Poisson process (NPP) and NS-GEV models; the other two are about model choice/testing and confidence intervals. This section is more demanding, but the concepts introduced are indispensable to the understanding of our analyses.

a. The nonhomogeneous Poisson process

In the point process approach to modeling extreme values (Smith 1989; Anderson et al. 2001), one looks at the times at which “high values” occur and at their

magnitudes. If t_1, t_2, \dots denote the times at which high values occur and x_1, x_2, \dots denote the corresponding magnitudes of the variable of interest, then the point process consists of a collection of points $\{(t_i, x_i)\}$ in a region of the positive quadrant of the plane. In practice, such a collection of points has first to be extracted from the original time series in such a way that the x_i s can be modeled as independent random variables. The way this is usually done with wave and similar data is by a process of “declustering” in which only the peak (highest) observations in clusters of successive exceedances of a specified threshold or level are retained and, of these, only those that in some sense are sufficiently apart (so that they belong to more or less “independent storms”) will be considered as belonging to the collection of points of the point process. Specifically, in the present application we have treated cluster maxima at a distance of less than 48 h apart as belonging to the same cluster (storm) and hence use only the highest of the cluster excesses.

The process of declustering is thus based on fixing a threshold over which one can consider exceedances and hence define the cluster peaks. This threshold is a number $a = a(t)$, which may depend on time t . For the moment we will assume that $a(t)$ has been chosen, and that the collection of points (obtained by some declustering procedure) representing our point process is given. Although the threshold used for declustering need not be the same as the threshold used for estimating the parameters of the NPP model (it could be lower), in this paper we will always use the same threshold and hence will not distinguish between the two.

Thus, our point process, or rather its “realization,” consists of a collection of points belonging to the plane set $C = \{(t, x): x > a(t), 0 \leq t \leq T\}$, where T is the number of years (in our case) over which observations are available and $a(t)$ denotes the threshold at time t . The NPP model of extremes has the following property. If A is a subset of C , then the number of points occurring in A , which we denote by $N(A)$, is a random variable with a Poisson probability function with mean $\rho(A) = \int_A \lambda(t, x) dt dx$, where

$$\lambda(t, x) = \frac{1}{\sigma(t)} \left[1 + \xi(t) \frac{x - \mu(t)}{\sigma(t)} \right]_{+}^{-(1/\xi(t))-1} \quad \text{for } (t, x) \in C. \quad (1)$$

Respectively, $\mu(t)$, $\sigma(t)$, and $\xi(t)$ are the location, scale, and shape parameters, or rather “parameter functions,” that may depend on time and need to be specified and estimated in practice. In particular, $N(C)$, the total number of points in the whole region C in which we are interested, is Poisson distributed with mean $\rho(C)$.

The function in (1) is called the intensity density of the process. Essentially, $\lambda(t, x)$ regulates the abundance of points around the location (t, x) of the region C in the sense that the larger (smaller) λ is around a neighborhood A of (t, x) , the larger (smaller) the average number of points is in that neighborhood, $\rho(A)$.

In (1), the subscript + is used to mean that $\lambda(t, x)$ is zero unless the expression in parenthesis is >0 . Thus, given $t \in (0, y)$, if $\xi(t) \geq 0$ then the range of x values for which $\lambda(t, x)$ is not zero is $\max[a(t), \mu(t) - \sigma(t)/\xi(t)] < x < \infty$, while if $\xi(t) < 0$ this range is $a(t) < x < \mu(t) - \sigma(t)/\xi(t)$. Consequently, if A is a plane set of points (t, x) that does not satisfy these restrictions, then $\rho(A) = 0$ and hence $N(A) = 0$ with probability one.

The integration of $\lambda(t, x)$ in x can be carried out analytically, and to compute $\rho(C)$ it remains to integrate the result in t ,

$$\begin{aligned} \rho(C) &= \int_C \lambda(t, x) dx dt \\ &= \int_0^T \left[1 + \xi(t) \frac{a(t) - \mu(t)}{\sigma(t)} \right]_{+}^{-(1/\xi(t))} dt. \end{aligned} \quad (2)$$

To incorporate nonstationarity into the process we shall consider the following models for its parameters:

$$\begin{aligned} \mu(t) &= \mu_0 + r_1 P(t) + r_2 G(t), \\ \log[\sigma(t)] &= b_0 + q_1 P(t) + q_2 G(t), \\ \xi(t) &= \xi, \end{aligned} \quad (3)$$

$t = 1, 2, \dots, T$, where r_1, r_2 , etc., are constants and $P(t)$ and $G(t)$ are covariates, that is, observations from a time series that for each time t are to a certain degree related to the peak x occurring at t .

Applying such a nonhomogeneous Poisson process model to data requires estimating the parameters involved in (3). For this purpose we will use the maximum likelihood method (Smith 1989; Anderson et al. 2001). We let $[t_{i,j}, x(i, j)]$, $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, T$, denote the collection of points obtained from the de-clustering procedure, where T is the number of years of data being analyzed and n_i is the number of peaks observed in (a specified season of) year i .

The likelihood function based on this sample is given by $L = \exp[-\int_C \lambda(t, x) dt dx] \prod_{i,j} \lambda[t_{i,j}, x(i, j)]$. The maximum likelihood estimates are the values of the parameters [μ_0, r_1 , etc., in the case of a Poisson model specified by (3)] that maximize this expression, and they usually have to be determined numerically. Instead of maximizing L it is convenient to maximize its logarithm, the log-likelihood function,

$$l = - \int_C \lambda(t, x) dt dx + \sum_{i,j} \log \lambda[t_{i,j}, x(i, j)]. \quad (4)$$

The covariates P and G in terms of which the parameters of the Poisson process will be modeled vary only from year to year, and so [see Eq. (3)] the parameter functions $\mu(t)$, $\sigma(t)$, and $\xi(t)$ will be constant within each year.

The threshold $a(t)$ is taken as a constant, that is, $a(t) = a$; it is defined as the 97th percentile of the empirical distribution of the 6-hourly ERA-40 SWH series for a given season. This specification is based on previous studies with the stationary approach (Caires and Sterl 2005), where it was found that the 97th percentile provided a reasonable threshold value (over which the asymptotic models of extreme value theory were thought to provide a good approximation), and on experimenting with the present model.

The choice of the threshold is usually a problematic aspect of extreme value analyses. If the threshold is too high, very few extreme values will actually contribute to the fitting of the model, which will result in uncertainty in the estimates (e.g., too large confidence intervals). On the other hand, if the threshold is too low there may be too many values in the sample that are not compatible with the asymptotic distribution of extremes, and this may result in biased estimates. To overcome this “variance-bias dilemma,” one has to somehow choose a threshold “in the middle,” and that is the difficult part. In the case of the NPP model the choice of the threshold seems to be even less obvious than in the POT GPD approach, where some experience and empirical rules exist.

Our choice of a constant threshold is based on the assumption that even if the choice of an appropriate threshold would vary in time, such variation should be negligible with respect to the time periods considered and to realistic climatic trends. This assumption has not been checked but seems very plausible to us. Also, we have taken the threshold a little above the threshold used by Caires and Sterl (2005) in the stationary approach, which should compensate for an eventual need for using an increasing threshold.

What usually happens is that the results do not vary much if one varies the threshold in an interval. This is also the case in our application: varying the threshold between the 95th and 98th percentiles yields compatible estimates. On the other hand, considering thresholds above the 93rd percentile seems, according to our and other authors’ experience, high enough a threshold for extremes of SWH to follow approximately a GPD distribution.

The peak exceedances and their times are, as above,

represented by $[t_{i,j}, x(i, j)]$, $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, T$. They correspond to the seasonal peaks of cluster exceedances above the threshold a and the times in which they occur as obtained from the 6-hourly time series of ERA-40 SWH. The declustering method we use in order to arrive at this sample is the one explained earlier.

Because the parameters are constant within each year, it will be convenient to denote the $\mu(t)$, $\sigma(t)$, and $\xi(t)$ by μ_i , σ_i , and ξ_i , respectively, when t falls in year i . In this way, taking into account the piecewise constancy of the parameter functions, using (2) to compute the integral in (4) and (1) to write the expression for $\lambda[t_{i,j}, x(i, j)]$, we can write the log likelihood function as

$$\begin{aligned} l = & -\sum_{i=1}^T \left(1 + \xi_i \frac{a_i - \mu_i}{\sigma_i} \right)_+^{-\frac{1}{\xi_i}} \\ & + \sum_{i=1}^T \sum_{j=1}^{n_i} \log \left\{ \frac{1}{\sigma_i} \left[1 + \xi_i \frac{x(i, j) - \mu_i}{\sigma_i} \right]_+^{-(1/\xi_i)-1} \right\}. \end{aligned} \quad (5)$$

The within-year constancy of the parameters and covariates allows us to write model (3) as

$$\begin{aligned} \mu_i &= \mu_0 + r_1 P_i + r_2 G_i, \\ \log(\sigma_i) &= b_0 + q_1 P_i + q_2 G_i, \\ \xi_i &= \xi, \end{aligned} \quad (6)$$

where again $i = 1, 2, \dots, T$ denotes the year, and P_i and G_i denote the values of the covariates in year i . The models we shall use for the parameters will always be submodels of the “full model” (6), that is, they will be as in (6) except that some of the parameters may be taken as zero from the outset.

If we substitute the right-hand sides of the Eqs. (6) into the log likelihood (5), we obtain an expression depending on several parameters (seven in the case of the full model); it is such an expression that we will have to maximize. The maximization will be carried out numerically using the downhill simplex method (Press et al. 1992).

The m -year return value x_m is obtained from the fact that the mean number of points whose vertical coordinate exceeds a value x_m in m years is $\int_0^m \int_{x_m}^\infty \lambda(t, x) dx dt$. Setting this equal to 1 and using (2) with $T = m$ and $a(t) = x_m$, we see that the m -year return value is the solution x_m to the equation

$$\int_0^m \left[1 + \xi(t) \frac{x_m - \mu(t)}{\sigma(t)} \right]_+^{-(1/\xi(t))} dt = 1. \quad (7)$$

The above expression incorporates the time variability of the parameter estimates in the particular return-value estimate. However, in this study we will be mainly

interested in the yearly variation of return-value estimates because of the time changes in the parameters, that is, in the variation of the return values once the estimates of the parameters for a given time are fixed. Treating the parameters in (7) as constant in t and solving for x_m , we find that the m -year return value based on the NPP parameters at a fixed time t is

$$x_m(t) = \frac{\sigma(t)}{\xi(t)} \left[\left(\frac{1}{m} \right)^{-\xi(t)} - 1 \right] + \mu(t). \quad (8)$$

b. The nonstationary generalized extreme value model

In the NS-GEV approach to extremes (see, e.g., Coles 2001), the maxima of a process over n times of observation within year/season t is assumed to have a NS-GEV distribution function,

$$G_t(y) = \exp \left\langle - \left\{ 1 + \xi(t) \left[\frac{y - \mu(t)}{\sigma(t)} \right] \right\}_+^{-1/\xi(t)} \right\rangle, \quad (9)$$

for $-\infty < y < \infty$, where the parameters satisfy $-\infty < \mu(t), \xi(t) < \infty$, and $\sigma(t) > 0$. In the traditional stationary setting the parameters are constant in time, and (9) is just the GEV distribution function.

In the stationary case the parameters of the GEV distribution and Poisson models are related in the following way. Suppose that the sequence of observations above a certain level satisfies the model specified by (1). Then, the probability that the maximum in a sequence of T observations does not exceed a large value y equals the probability that the number of points in the region $C_y = \{(t, x); x > y, 0 \leq t \leq T\}$ is zero, and thus equals (recall the form of the Poisson probability function) $e^{-\rho(C_y)}$, which can be seen to equal (9) by a computation like that in (2). Thus, in situations where both stationary versions of the methods are applied one expects the parameter estimates to agree, though of course the methods do not use the same amount of data and therefore have somewhat different statistical properties. In the nonstationary case, however, this correspondence between the parameters in the NPP and NS-GEV models is lost to a certain extent, which is just a reflection of the fact that each of the nonstationary models is a rather particular extension of a stationary model (see Anderson et al. 2001, p. 59, for further details on this matter). One of our aims in this work is to study and quantify the differences between the two approaches in terms of the return values rather than the differences between parameters, because these are not really comparable.

We shall apply the NS-GEV distribution to model seasonal annual maxima as in Wang and Swail (2006),

and, just as in the case of the NPP model, model the parameters of the NS-GEV distribution in terms of the covariates P and G using model (6). The parameters of the NS-GEV distribution can also be estimated by the maximum likelihood method (e.g., Coles 2001). We let $[t_i, y(i)]$, $i = 1, 2, \dots, T$, denote the collection of seasonal annual maxima, where again T is the number of years of data. The log-likelihood function based on this sample is then given by

$$\begin{aligned} l = & -\sum_{i=1}^T \left\langle \log \sigma_i + (1 + 1/\xi_i) \log \left\{ 1 + \xi_i \left[\frac{y(i) - \mu_i}{\sigma_i} \right] \right\} \right. \\ & \left. + \left\{ 1 + \xi_i \left[\frac{y(i) - \mu_i}{\sigma_i} \right]^{-1/\xi_i} \right\}_+ \right\rangle. \end{aligned} \quad (10)$$

Again, the maximization will be carried out numerically using the downhill simplex method (Press et al. 1992).

The m -year return value y_m is obtained numerically by solving $\sum_{t=1}^m G_t(y) = 1$ for y . The m -year return value computed with the parameters at a fixed time t , denoted by $y_m(t)$, is defined as the quantile of probability $1 - 1/m$ of the NS-GEV distribution, which is obtained by inverting (9),

$$y_m(t) = \mu(t) - \frac{\sigma(t)}{\xi(t)} \{1 - [-\log(1 - 1/m)]^{-\xi(t)}\}. \quad (11)$$

c. Model choice and testing

We end this section by specifying the models [(3)] describing the temporal variation of the parameters of the NPP and NS-GEV models that we will be interested in estimating, and by stating the method we use to compare two different models.

Ordered in degree of complexity and in the notation of (6), with i indicating the year, the models we shall consider are as follows:

- PM₀: $\mu_i = \mu_0$, $\sigma_i = \sigma$, $\xi_i = \xi$, $i = 1, 2, \dots, T$; all parameters are constant;
- PM₁: $\mu_i = \mu_0 + r_1 P_i$, $\sigma_i = \sigma$, $\xi_i = \xi$, $i = 1, 2, \dots, T$; the location parameter is a function of the time-dependent covariate P , and the other parameters are constant;
- PM₂: $\mu_i = \mu_0 + r_1 P_i + r_2 G_i$, $\sigma_i = \sigma$, $\xi_i = \xi$, $i = 1, 2, \dots, T$; the location parameter is a function of both time dependent covariates, and the other parameters are constant;
- PM₃: $\mu_i = \mu_0 + r_1 P_i + r_2 G_i$, $\log(\sigma_i) = b_0 + q_1 P_i$, $\xi_i = \xi$, $i = 1, 2, \dots, T$; the location parameter is a function of both time-dependent covariates, the scale parameter is a function of the time-dependent covariate P , and the shape parameter is constant; and

- PM₄: $\mu_i = \mu_0 + r_1 P_i + r_2 G_i$, $\log(\sigma_i) = b_0 + q_1 P_i + q_2 G_i$, $\xi_i = \xi$, $i = 1, 2, \dots, T$; both location and scale parameters are functions of the two time-dependent covariates and the shape parameter is constant.

To compare any two of the above models and to decide which is “the best,” we use the likelihood ratio principle (e.g., Coles 2001). The set of models is the same as that considered by Wang and Swail (2001, 2004, 2006) and Wang et al. (2004).

d. Confidence intervals

A common approach to the computation of confidence intervals based on maximum likelihood estimates is that based on the delta method, with either the information matrix or the observed information matrix. A method that is usually more accurate (e.g., Coles 2001), especially when the distribution of estimates is skewed, is the profile likelihood method. In this work we use the latter method, but will not explain it here because of its rather technical nature; see Coles (2001) and references therein.

4. Methods of analysis

a. Trend analysis

Once the dependence of the parameters of the NPP or NS-GEV model on the two covariates (P and G) has been established and estimated, *projected time series* of these parameters, that is, forecasts of their values in the future, can be obtained from available projections or forecasts of the covariates. In this work we obtain such projected series of parameters on the basis of three ensembles of projections of the covariates, each of which was obtained under one of three different climate-forcing scenarios. We will therefore have to consider three time series of projections for each of the parameters under each forcing scenario. The three ensemble members within each climate scenario will be regarded as “replicates,” that is, as three sets of independent time series governed by the same law. As in Wang et al. (2004) and Wang and Swail (2006), the projected time series of parameters will then be analyzed for trends by combining the three members of each ensemble into a single sample, and regression models will be fitted and chosen to represent them.

More specifically, let (π_t^i) ($i = 1, 2, 3$) denote the projected time series of a given parameter of the NPP model obtained in a given scenario and corresponding to the ensemble member i . We consider the following regression models (RM):

$$\begin{aligned} \text{RM}_0: \pi_t^j &= \alpha_0 + \varepsilon_t^j, \\ \text{RM}_1: \pi_t^j &= \alpha_0 + \alpha_1 t + \varepsilon_t^j, \\ \text{RM}_2: \pi_t^j &= \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \varepsilon_t^j, \end{aligned}$$

where the errors ε_t are assumed normal. Using the three ensemble members, the parameters in these models are estimated, and the F test is used to check which of the three RM fits the data better (see, e.g., Jorgensen 1993).

Once a RM is chosen and fitted to the projected time series of parameters, projected return-value estimates, that is, i.e., time series of return values computed from the projected time series of parameters, can be obtained.

In addition to the above three models, we have also considered more complicated models, such as cubic polynomials; however, in no case have we found, on the basis of the F test, the need to fit more than a quadratic model; that is why we restrict ourselves to linear and quadratic models throughout the paper.

The motivation for using regression models instead of the “raw” series of parameters is to obtain a summary of their evolution in terms of a simple parametric form containing only a couple of parameters. This facilitates the interpretation of trends, but naturally has the effect of smoothing the raw series, which might be subject to some criticism. Our justification for this approach is that fitting a regression model to these data should not influence the results as compared to using the raw series. The assumptions behind regression models might also be questionable in our case, but our approach hardly depends on them because our objective is simply to obtain a smoothed representation of the series, rather than a statistical model capable of describing it; also, the use of the F test mentioned above should be seen more as a device for getting a simple and still sufficiently flexible model than as a formal procedure of validating a model.

b. Identifying sources of variability

The projected return-value estimates are determined by several factors: the nonstationary extreme value model (NPP or NS-GEV) used; the climate change signal of the covariates within a certain scenario, which we refer to as the *prescribed forcing* or *forcing-induced variability*; and the forcing scenario. To assess the contribution of these factors to the uncertainty of return-value projections, we use the same technique as Wang and Swail (2006): analysis of variance (ANOVA). We consider both one- and two-way ANOVA for studying the uncertainty resulting from one factor or two simultaneously. For the sake of clarity of the analysis of re-

sults, we outline the assumptions, models, and methods used (for more details see, e.g., Rice 1995; Seber and Lee 2003; Jorgensen 1993; Storch and Zwiers 1999). It should be pointed out, however, that the ANOVA procedures do not take the uncertainty in the projected return-values estimates into account, which restricts their utility.

1) ONE-WAY ANOVA

The application of one-way ANOVA we will consider is exactly the same as that in Wang and Swail (2006) but applied to the SWH₂₀ projection made with the NPP model rather than the NS-GEV model.

Let X_{ij} , $i = 1, \dots, k$, $j = 1, \dots, n$ represent k groups of n random variables such that $X_{ij} = \alpha + \alpha_i + \varepsilon_{ij}$, where the parameter α is called the overall mean, the parameter α_i is called the effect of the i th group or level of the factor under consideration, and ε_{ij} is a zero mean random variable called the error of the j th observation in group i . In checking whether in a particular scenario the forcing-induced variability influences the series of projected return values, the factor “prescribed forcing” has $k = 110$ levels (1990, 1991, ..., 2099) and $n = 3$ random variables corresponding to the three runs in an ensemble.

It is assumed that the errors ε_{ij} are independent and identically distributed (i.i.d.) zero mean normal variables with variance σ^2 .

The main purpose of a one-way ANOVA analysis is to test the hypothesis that all of the α_i parameters are zero, that is, that the factor being considered has no effect. If this null hypothesis is true, then the F ratio

$$F = \frac{\text{SS}_B/(k-1)}{\text{SS}_W/[k(n-1)]},$$

where $\text{SS}_B = n \sum_{i=1}^k (X_{i\cdot} - X_{..})^2$, $\text{SS}_W = \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - X_{i\cdot})^2$ and the dot denotes averaging over the missing subscript, has an F distribution with $(k-1)$ and $[k(n-1)]$ degrees of freedom, denoted by $F[k-1, k(n-1)]$. Respectively, SS_B and SS_W are called the between- and within-groups sum of squares. They reflect the amount of variability resulting from differences between groups and between observations within the same group, and partition the total sum of squares $\text{SS}_T = \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - X_{..})^2$ in the sense that $\text{SS}_T = \text{SS}_B + \text{SS}_W$ with SS_B and SS_W independent.

The proportion of total variance resulting from the factor is defined by

$$\begin{aligned} p &= \frac{(F-1)}{\text{SS}_T} \text{SS}_W \frac{(k-1)}{k(n-1)} \\ &= \left[\text{SS}_B - \frac{(k-1)}{k(n-1)} \text{SS}_W \right] / \text{SS}_T. \end{aligned}$$

TABLE 1. Expressions involved in the two-way ANOVA.

	<i>F</i> ratio	Degrees of freedom	Proportion of total variance
Factor A	$\frac{SS_A/(I-1)}{SS_E/[IJ(T-1)]}$	$I-1, IJ(T-1)$	$\left[SS_A - \frac{(I-1)}{IJ(T-1)} SS_E \right] / SS_T$
Factor B	$\frac{SS_B/(J-1)}{SS_E/[IJ(T-1)]}$	$J-1, IJ(T-1)$	$\left[SS_B - \frac{(J-1)}{IJ(T-1)} SS_E \right] / SS_T$
Interaction	$\frac{SS_{AB}/[(I-1)(J-1)]}{SS_E/[IJ(T-1)]}$	$(I-1)(J-1), IJ(T-1)$	$\left[SS_{AB} - \frac{(I-1)(J-1)}{IJ(T-1)} SS_E \right] / SS_T$
Sums of squares			$SS_A = TJ \sum_{i=1}^I (X_{i..} - X_{..})^2, \quad SS_B = TI \sum_{j=1}^J (X_{.j} - X_{..})^2,$ $SS_{AB} = T \sum_{i=1}^I \sum_{j=1}^J (X_{ij.} - X_{i..} - X_{.j} + X_{..})^2,$ $SS_E = \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^T (X_{ijt} - X_{ij.})^2, \quad SS_T = \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^T (X_{ijt} - X_{..})^2$

Because the number of runs in an ensemble is small, the power of the one-way ANOVA will be small for low values of p ; for example, it can be seen that the power is 0.5 when $p < 0.1$ and close to 1 when $p > 0.2$.

2) TWO-WAY ANOVA

The application of two-way ANOVA that we will consider is exactly the same as in Wang and Swail (2006) but with the three climate models being replaced by the two nonstationary models of extremes.

Let X_{ijt} , $i = 1, \dots, I$, $j = 1, \dots, J$, $t = 1, \dots, T$, be random variables classified into I levels of factor A ($I = 2$ nonstationary models of extremes) and J levels of factor B ($J = 3$ forcing scenarios), with T observations on each ($T = 110$ projected return values from 1990 to 2099), and such that $X_{ijt} = \alpha + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijt}$, where α is an overall mean level, α_i and β_j are parameters accounting for the influence of factors A and B, respectively, δ_{ij} is a parameter expressing the interaction of the two factors, and ε_{ijt} are i.i.d. zero mean normal variables with variance σ^2 .

In our application we are interested in testing the hypothesis that the choice of the nonstationary models of extremes or forcing scenario has no effect, that is, that the α_i s are all zero or that the β_j s are all zero, and that there is no interaction between these factors (i.e., the contribution, if any, of the extreme value model used does not change across climate scenarios). Similarly to what happens in one-way ANOVA, the total sum of squares of the errors is partitioned into independent components reflecting the presence/absence of effects, namely, as $SS_T = SS_A + SS_B + SS_{AB} + SS_E$, where the terms on the right are defined in Table 1, and the tests of the above hypotheses are carried out through *F* ratios with SS_E in the denominator. These are also given in Table 1, along with the degrees of freedom and proportion of the total variance resulting from each factor.

3) CHECKING HOMOGENEITY OF THE VARIANCES

Before applying ANOVA to the data we will be testing the assumption that the errors have equal variances using the Barlett test (e.g., Storch and Zwiers 1999, p. 180); but, keep in mind that the test is very sensitive to nonnormality. Fortunately, the analysis of variance for equal-sized samples (the cases we will be considering here) is robust even under considerable heterogeneity of variances, and the validity of results is affected only slightly even under considerable deviations from normality, especially as the sample sizes increases (Rice 1995; Seber and Lee 2003).

5. Results

The methods described in section 3 were applied to data at each grid point of the ERA-40 SWH grid for each of the following four seasons: January–February–March (JFM), April–May–June (AMJ), July–August–September (JAS), and October–November–December (OND). Extreme value analyses were carried out separately for each of the four seasons. The NPP model was applied to each set of declustered peak excesses of SWH above the 97th percentile of the whole (seasonal) SWH dataset, and the NS-GEV model was applied to each set of seasonal maxima of the 6-hourly ERA-40 SWH series; in each case, the estimates and inferences produced are based on the fitting of the NPP and NS-GEV models to the SWH data (peak excesses or maxima) and associated ERA-40 SLP covariates P and G from 1958 to 2001.

Before we begin the analysis of the results, we try to give a physical interpretation of the models considered. Waves are created by wind and wave heights are roughly proportional to the squared wind speed, which in turn is proportional to the pressure gradient. Pressure systems have marked spatial patterns, and in prin-

ciple a mean SLP anomaly at a certain location could give an indication of the spatial pattern present. Thus, measures of pressure and of its gradient give information about wind velocities and consequently about wave height. By choosing P and G as covariates explaining the time variability in the parameters of the extreme value models we are trying to establish how changes in local and spatial pressure/wind influence, if at all, the extremes of SWH. Note that these covariates incorporate not only possible time trends but also decadal and other scales of variability. Also, their influence may be felt in the form of shifts (models PM_1 or PM_2) and/or changes in spread (models PM_3 or PM_4) in the distribution of extremes, which can be interpreted as increases/decreases in severity and/or variability in extreme wave systems, respectively.

In regards to the choice of the model (6) describing the dependence of the parameters on the covariates, the results of the likelihood ratio tests show that the location parameter is significantly correlated with both P and G , and the scale parameter is not significantly correlated with either P or G . That is, the time variation of the P and G covariates influences the distribution of extremes in the form of shifts but not in the form of changes in spread. This is the case in both the NPP and NS-GEV models. Thus, model PM_2 (section 3c) is the one fitting the data best; this is the model we used to compute the projections of SWH 20-yr return values. This choice of model PM_2 is consistent with what was reported by Wang et al. (2004) and Wang and Swail (2006), and implies that the changes in the m -year return values of SWH resulting from future climate changes are in absolute value independent of m [cf. Eqs. (8) and (11)].

Having adopted model PM_2 to describe the NPP and NS-GEV parameters in terms of the covariates and having estimated its coefficients, we have then computed projections of the location parameter and of the SWH 20-yr return values from 1990 to 2099 based on each ensemble member and under each scenario using the CGCM2 projections of P and G , obtaining a total of nine time series of projections corresponding to three forcing scenarios with three ensemble members each on a global 1.5° latitude \times 1.5° longitude grid, for each nonstationary model of extremes.

Next, we have analyzed the time variability of the projected series of the location parameter from 1990 to 2099 for each climate scenario using the RM presented in section 4a. In each case, the coefficients of the RM were estimated by combining the three members of each ensemble into a single sample. In both the NPP and NS-GEV models the quadratic component of the trend is significantly different from zero at the 5% level

in some regions (typically more than 15% of the locations), especially in JFM and JAS and under the IS92a and A2 scenarios. Because RM_2 is the most general model we consider this is the model we have chosen to estimate the projected location parameter μ_t in order to analyze changes in SWH_{20} under a particular future climate scenario. Wang et al. (2004) and Wang and Swail (2005) have equally analyzed the trends in their projections of the location parameter and have also detected the presence of significant quadratic terms in the trends. This is in contrast to the present climate, where only linear trends were detected (Caires and Swail 2004; Wang and Swail 2005).

To recap, we have established that the best model for describing the dependence of the NPP or NS-GEV parameters on the SLP covariates is model PM_2 , and if such a model is used to obtain projections of the location parameter μ_t from 1990 to 2099, then the model best describing the time trends of these μ_t time series, or equivalently the $SWH_{20}(t)$ time series [see expressions (8) and (11)], is quadratic. In the following subsection we look at the NPP estimates in detail and analyze the estimated future changes of SWH_{20} and the influence of the prescribed forcing on the results. After that we will look at the differences between the NPP and NS-GEV estimates and analyze the uncertainty that results from choosing a particular nonstationary model of extremes.

a. Results of the NPP model

The estimates of the parameters r_1 and r_2 are presented in Fig. 1 for each season; the regions in which the parameters are significantly different from zero (according to the likelihood ratio test comparing PM_1 with PM_0 in the case of r_1 , and PM_2 with PM_1 in the case of r_2) are hatched.¹ The parameter estimates are seasonally dependent. The estimates of r_1 are mostly negative in the high latitudes and positive in parts of the Tropics. Those of r_2 have large magnitudes close to the equator and comparatively low magnitudes at high latitudes, though they are also statistically significant in those regions. These differences between the values of the parameter estimates in the Tropics and at high latitudes express differences in the roles played by P and G in each case, and thus suggest differences in the type of

¹ Statistical significance is evaluated on a grid point-by-grid point basis. It should be pointed out that regional or global interpretations of the results of multiple testing—tests carried out simultaneously at a large number of locations—are not straightforward. We should expect false detections of statistical significance at some grid points for which spatial dependence makes it very difficult to properly correct.

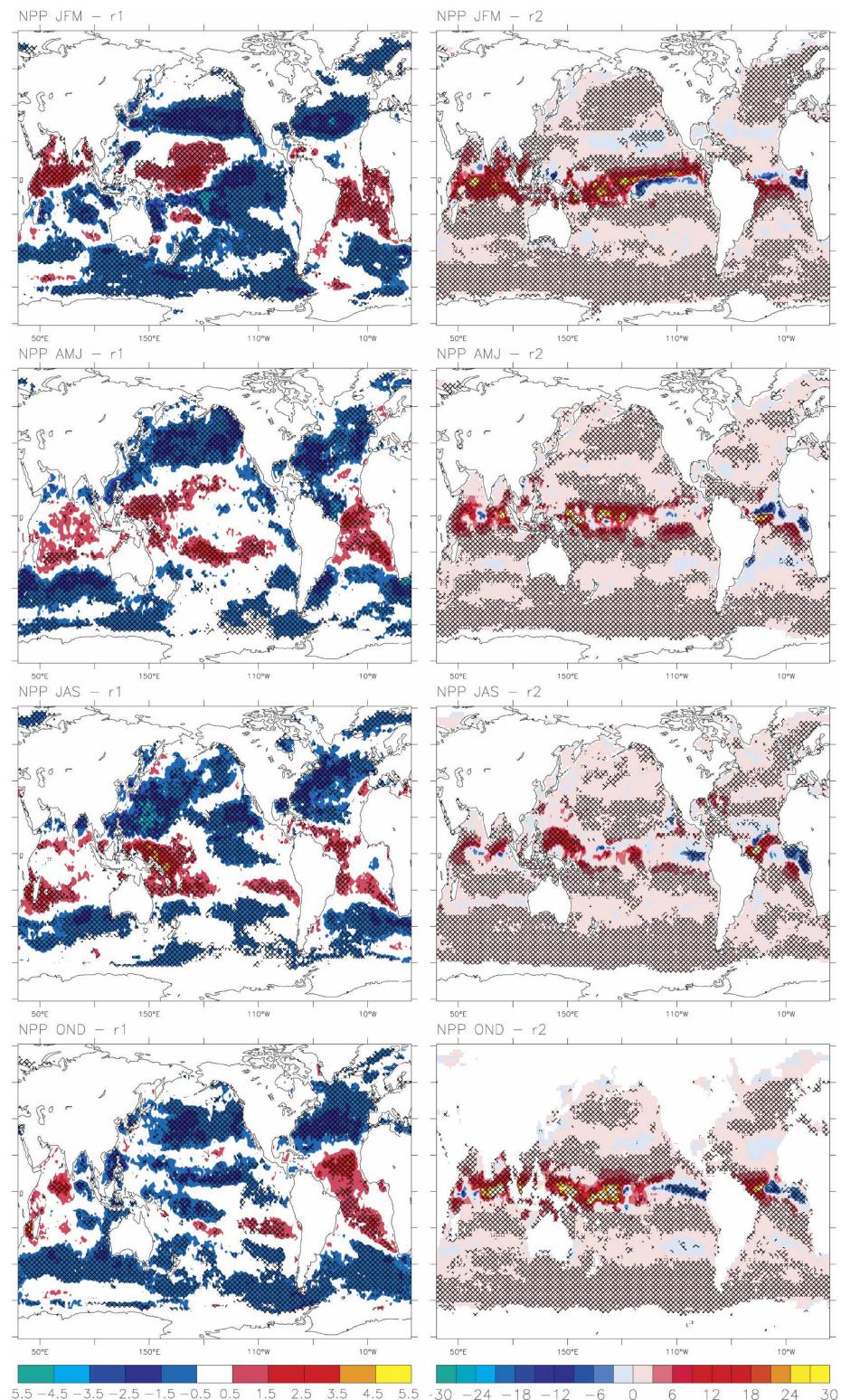


FIG. 1. Estimates of the coefficients r_1 and r_2 linking the location parameter of the NPP model to (left) P and (right) G . Hatching indicates locations where the coefficients are significantly different from zero at a 5% level.

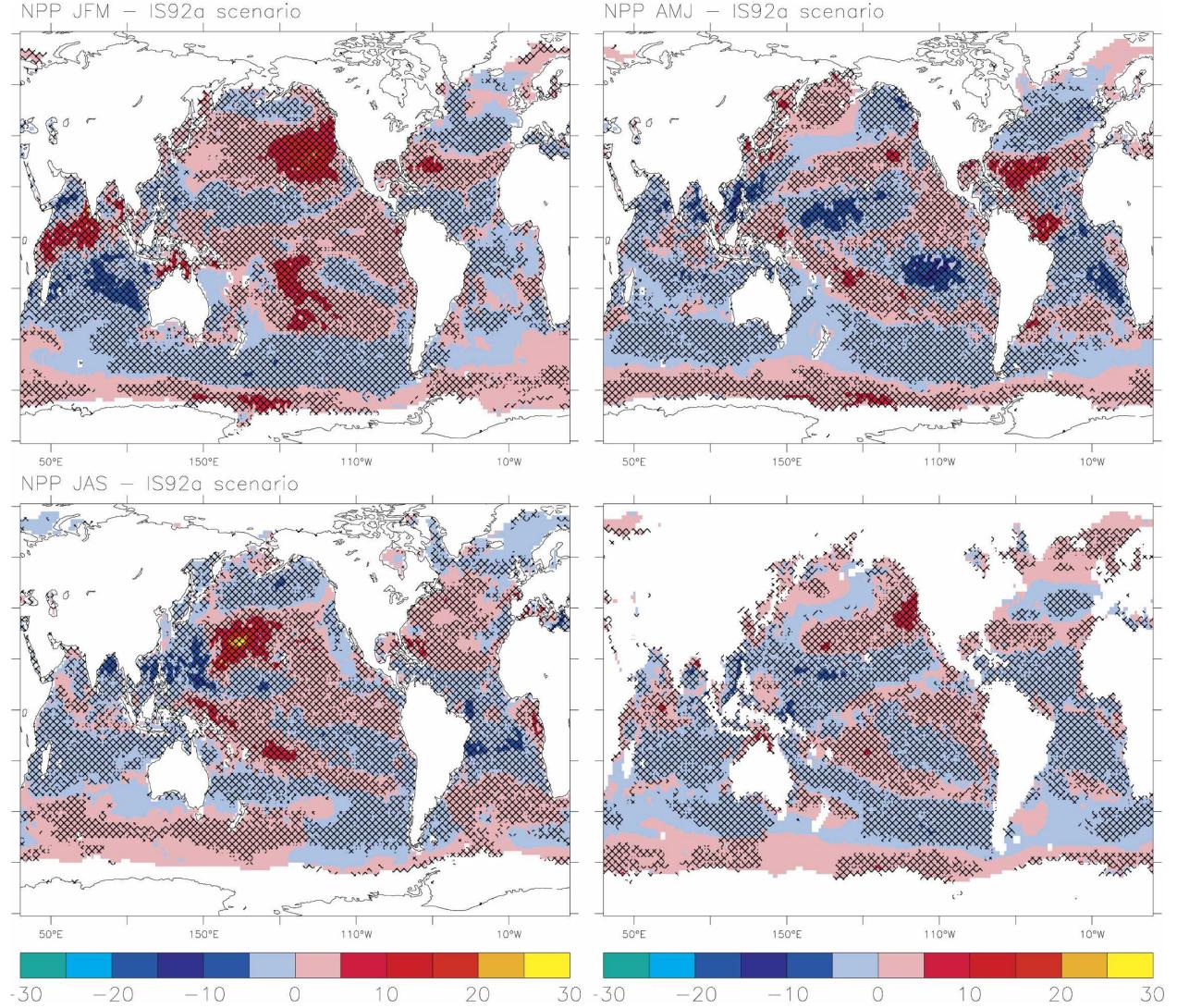


FIG. 2. Percentage changes in the indicated seasonal SWH_{20} from 1990 to 2080 (the estimate of $2080 \text{ SWH}_{20} - 1990 \text{ SWH}_{20}/1990 \text{ SWH}_{20}$), as estimated by combining the three projections under the IS92a forcing scenario. Hatching as in Fig. 1.

processes contributing to the generation of extremes of SWH in the Tropics and at high latitudes. The differences could mean that changes in the spatial pressure modes play the most important role in the changes of extremes at high latitudes, while in the Tropics changes in the intensity of local winds are the most important factor influencing the changes in extremes.

The RM₂ model was chosen to estimate changes in the projected time series of μ_r , which in turn was used to estimate 20-yr return values of SWH for 1990, 2020, 2050, and 2080. The percentage changes between the SWH_{20} estimates for 1990 and 2080 are presented in Figs. 2, 3, and 4 under the IS92a, A2, and B2 scenarios, respectively.

The changes in estimates of SWH_{20} are season and location dependent. The spatial patterns of the estimated changes are very similar under all scenarios. The magnitude of the changes under the IS92a and A2 scenarios are similar, although bigger changes are to be expected in the NP under the IS92a scenario. Differences of magnitude in the changes under the A2 and the IS92a scenarios are probably due to the fact that in the A2 scenario aerosol forcing is also considered. The changes predicted under the B2 scenario are smaller than in the other two, which was already to be expected because of the lower GHG forcing. Thus, the different climate scenarios affect the magnitude of the changes but not their spatial pattern. The region where bigger

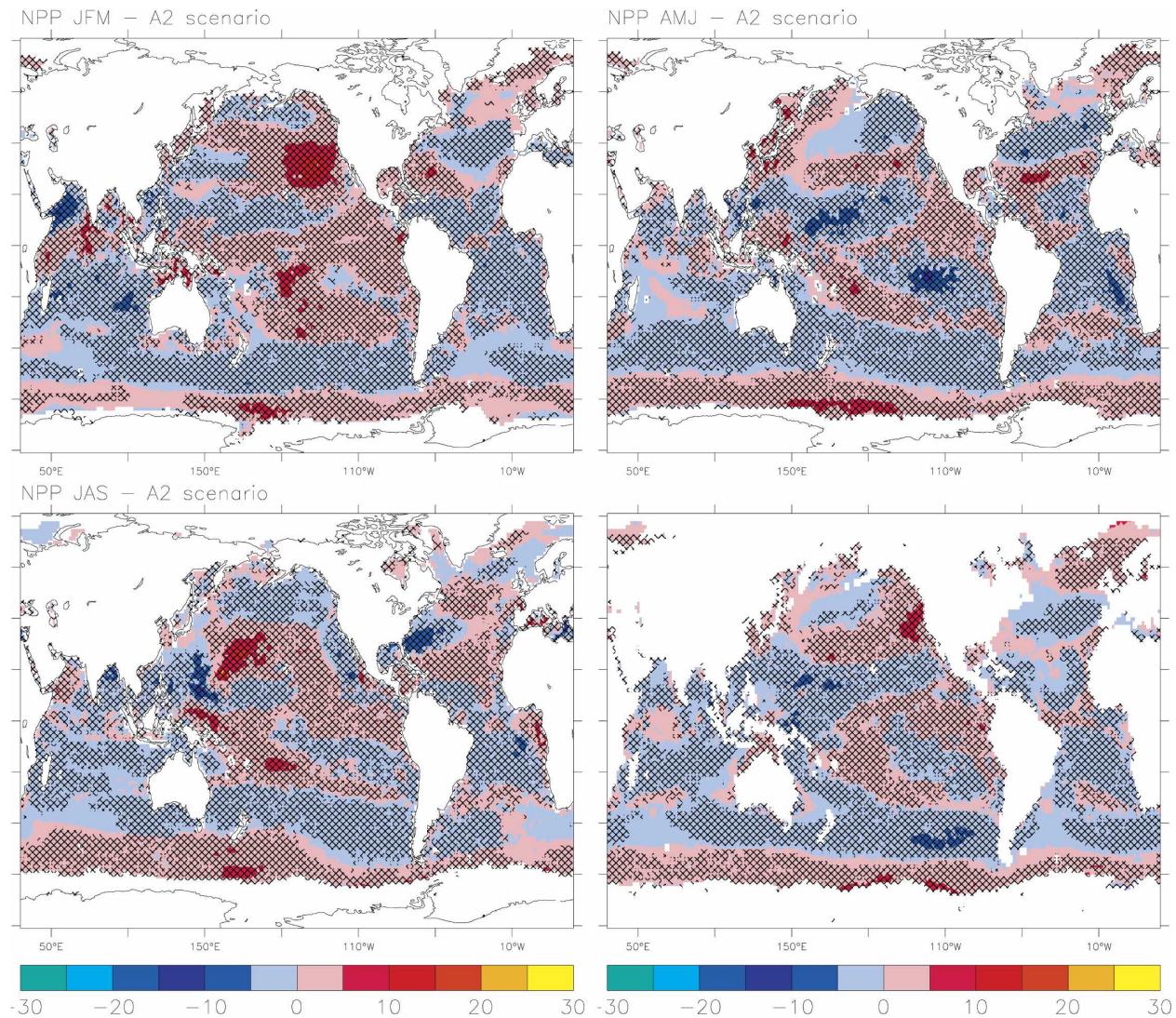


FIG. 3. The same as Fig. 2, but under the A2 scenario.

changes are expected to occur is the North Pacific. Analyzing the results presented in Figs. 2–4 season by season, we can sum up the following conclusions:

- In the JFM season there will be statistically significant decreases in the Southern Ocean (SO), mainly in the band between 40° and about 55° S, and increases at higher latitudes. In the Indian Ocean (IO), increases are expected in the northwest region and decreases in the southeast region, as well as around the Gulfs of Aden and Oman. All forcing scenarios show significant decreases in the central NA and increases in the southwest NA. The changes in the NP are characterized by strong increases (of up to 25%) in its central-east region, and by decreases south of the

Gulf of Alaska and the Bering Sea and westward from the South China Sea. Significant increases are expected in the tropical and central South Pacific (SP) and decreases in the tropical and South Atlantic (SA).

- In AMJ the pattern of changes in the SO and NA is the same as in the previous season. The expected changes in the SA seem to mirror those in the NA; to a lesser extent the changes in the SP mirror those in the NP.
- The most striking feature in JAS is an increase in the central-west NP and a decrease further west and on the Bay of Bengal. This is the season of the austral winter and under both the IS92a and A2 forcing scenarios significant increases are to occur in the SO south and around the southern half of New Zealand.

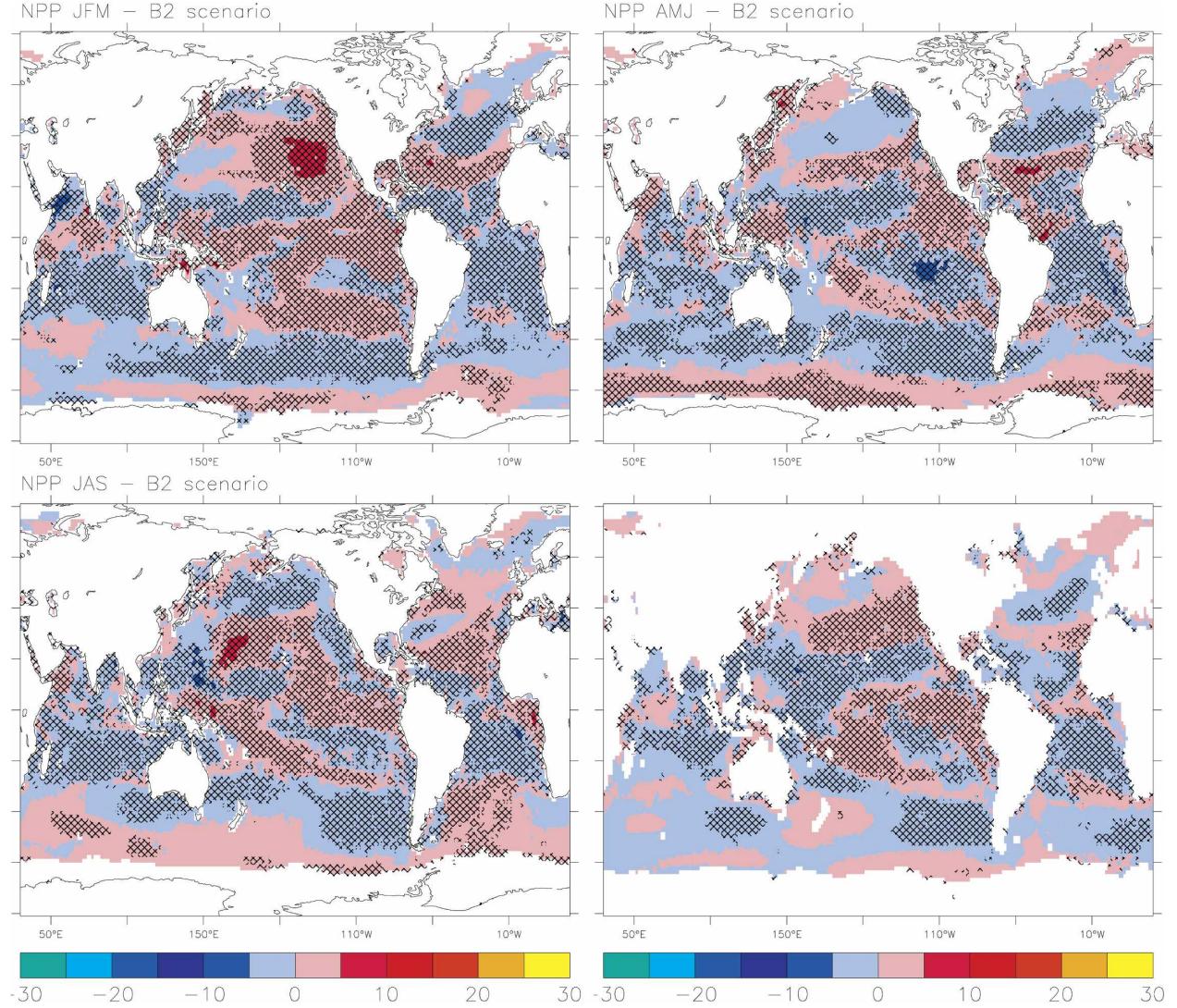


FIG. 4. The same as Fig. 2, but under the B2 scenario.

- In OND the changes are very similar to those in the JFM season except that their magnitude is smaller.

Although the total relative changes from 1990 to 2080 show similar patterns under all scenarios, the rate at which the changes occur within the period, especially in the initial decades, depends on the scenario. Specifically, changes occur more slowly and sometimes in the opposite direction in the initial decades under the SRES climate scenarios. These differences in the time evolution of the projections may be because, contrary to aerosols, GHGs have a cumulative effect in the climate, and once certain concentrations are reached their effects dominate, implying that although the presence of aerosols may initially slow down the changes under

the SRES scenarios, the effect of the GHG forcing would eventually become dominant.

Figures 5 and 6 show, for JFM and JAS, respectively, the projected changes in the SWH_{20} from 1990 to 2020, from 2020 to 2050, and from 2050 to 2080, under the IS92a and A2 forcing scenarios, obtained from the NPP model. Looking at Fig. 5 we see, for example, that in the central-east NP under the IS92a forcing scenario there is an increase in the projections of SWH_{20} from 1990 to 2020, whereas under the A2 forcing scenario there is a decrease. In the same region the increases are similar under both scenarios from 2020 to 2050, and they are bigger under the A2 scenario in the 2050–80 period. These differences are a reflection of the different quadratic trends in μ_t under each scenario. For this

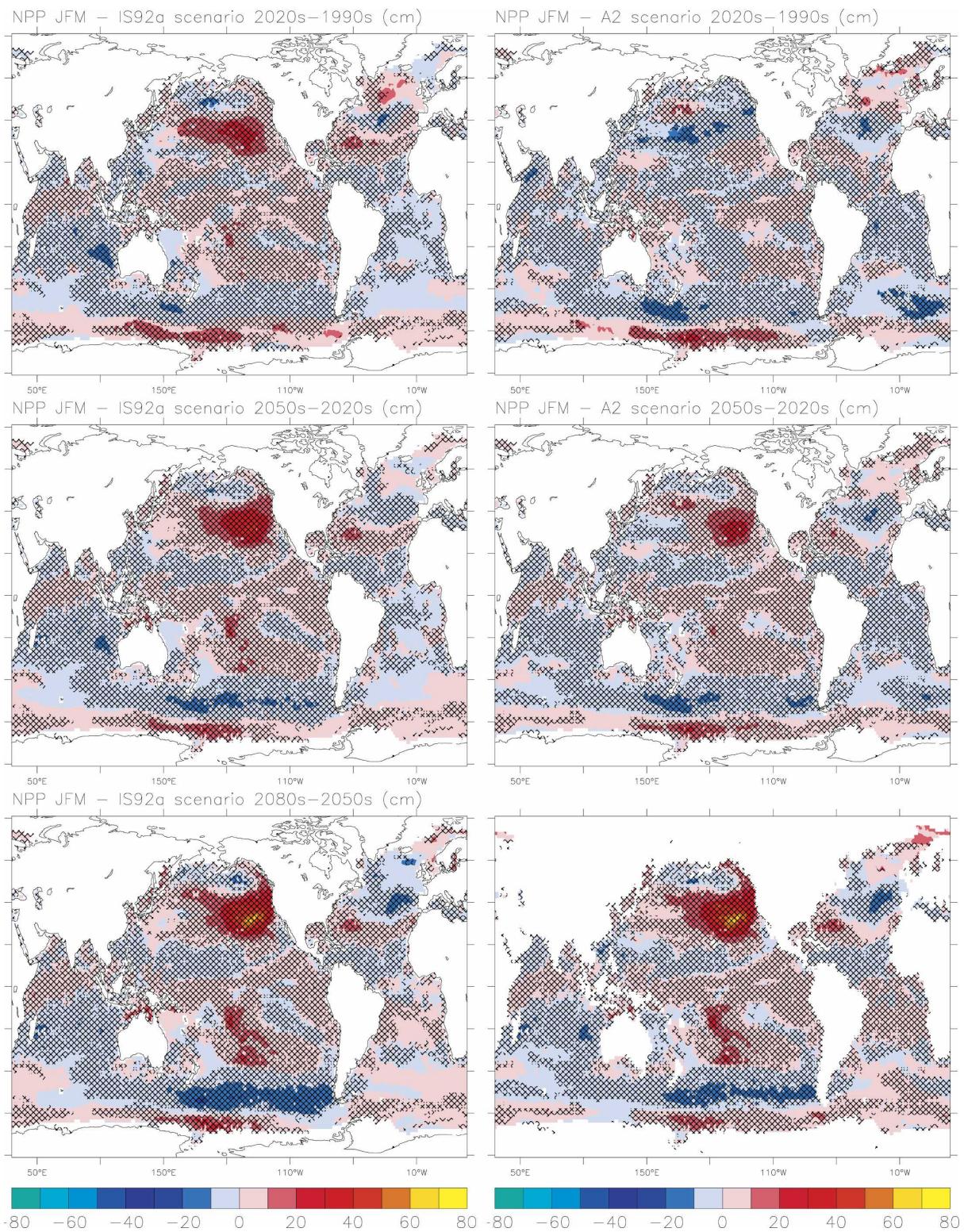


FIG. 5. Changes in the NPP estimates of SWH_{20} (top) from 1990 to 2020 (2020 – 1990), (middle) from 2020 to 2050, and (bottom) from 2050 to 2080 estimated by combining the three projections for the (left) IS92a and (right) A2 forcing scenarios for JFM. Hatching as in Fig. 1.

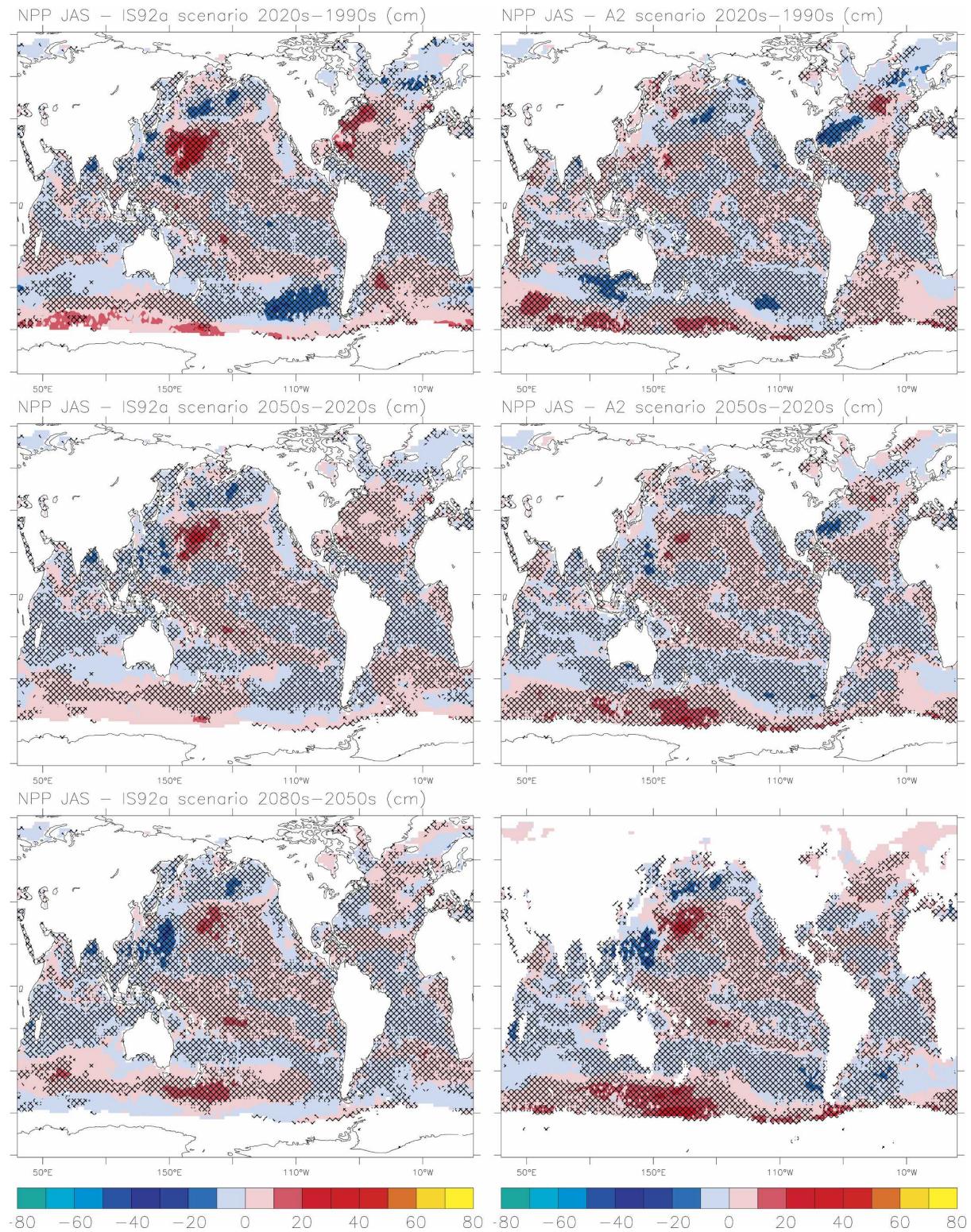


FIG. 6. The same as Fig. 5, but for JAS.

region, the linear term of the trend associated with the A2 time series is negative while that associated with the IS92a time series is positive, but under both scenarios the associated quadratic term is positive (the ones associated with the A2 time series being bigger); that is, under the A2 scenario there is first a decrease and then a quadratic increase with a higher rate as compared to what happens under the IS92a scenario, where there is always an increase. Other examples of different rates of change can be seen in Fig. 6, for instance, in the NA southeast of the American coast and south of New Zealand.

Both the patterns of change and the temporal variations of trends shown here are very close to those observed by Wang and Swail (2005) in projections for the NA and NP based on the NS-GEV model using SLP projections from the CGCM2 climate model and with model parameters inferred from present SWH and SLP fields other than ERA-40. The agreement indicates that the choices of dataset to describe the present climate and nonstationary model for extremes are not very critical.

The patterns of the global changes in projected values of SWH_{20} summarized here are in many aspects similar to those observed by Wang and Swail (2006) in projections obtained with the NS-GEV model on the basis of SLP projections produced by three different climate models combined, and with the model parameters inferred from present climates of ERA-40 SWH and SLP fields. However, the magnitude of the changes predicted by Wang and Swail (2006) is smaller than ours. This is not necessarily due to the choice of the NS-GEV in place of the NPP approach; Wang and Swail (2006) noted that their multimodel projected changes in SWH_{20} are smaller than those based only on CGCM2 projections. Also, in the case of Wang and Swail (2006) there is no big asymmetry between the changes in the NH and SH; their estimated changes in the high latitudes of SH are about as big as those in the NP. These differences between their results and ours indicate that the projections under CGCM2 are more extreme than those under other climate models, and that the former have a NH/SW asymmetry that is not present in the other climate models.

We now proceed by analyzing the uncertainty resulting from the prescribed forcing in the projections of SWH_{20} from 1990 to 2099 obtained with the NPP model using the one-way ANOVA model. We look at the projections of SWH_{20} for 1990–2099 with three independent observations per year [one for each ensemble member, i.e., $k = 110$ and $n = 3$ in the notation of section 4b(i)] under each climate scenario. The Barlett test detects significant differences between the vari-

ances of the errors across the years at a 5% significance level in about 10% of all locations, which suggests no serious violation of the assumptions required by ANOVA; accordingly, we proceed by assuming that the slightly higher proportion of significant results does not have a serious effect on our analysis. Figure 7 shows the seasonal proportion of the total variance resulting from the prescribed forcing under the A2 scenario and the regions where the effect of the prescribed forcing is statistically significant at a 5% level. The effect of the prescribed forcing is statistically significant mostly in the regions where bigger future changes of SWH_{20} are to be expected, and the proportion of the total variance resulting from this effect can be up to 50%. The estimates presented in Fig. 7 are almost identical to those presented by Wang and Swail (2006) in the setting already described, but using A2 SLP projections only from the CGCM2 climate model (see their Fig. 7). The power of ANOVA in this analysis is rather low when the proportion of the total variance resulting from this effect is below 10% and so many regions with statistically significant prescribed forcing may have been missed by this analysis. The prescribed forcing effect on the projections under the IS92a scenario are similar to that observed under the A2 scenario, but under the B2 scenario the proportion of variance resulting from the prescribed forcing is lower and the areas with statistically significant results are smaller (these results are not shown here).

b. Comparisons between the results of the NPP and NS-GEV models

One of our main objectives is to study the uncertainty in projections of SWH_{20} based on the NPP and NS-GEV models. We have already established that both approaches support the same dependence between model parameter estimates and the covariates (viz., that described by the PM₂ model), and that projections based on each of the models yield statistically significant quadratic trends in the time evolution of the projected time series. We now look at the differences between return-value estimates obtained from the two approaches.

We start by looking at how the models fit to the present climate data in terms of error and uncertainty, and at how the estimates of the two compare.

As a way of quantifying the error in the model estimates of return values we have extracted the 25 maxima from the 25 successive 20-yr subsets of the ERA-40 dataset, namely, those of the periods 1958–77, 1959–78, . . . , 1982–2001. We have then computed the corresponding 20-yr return-value estimates for each of these periods from the NPP and NS-GEV models, tak-

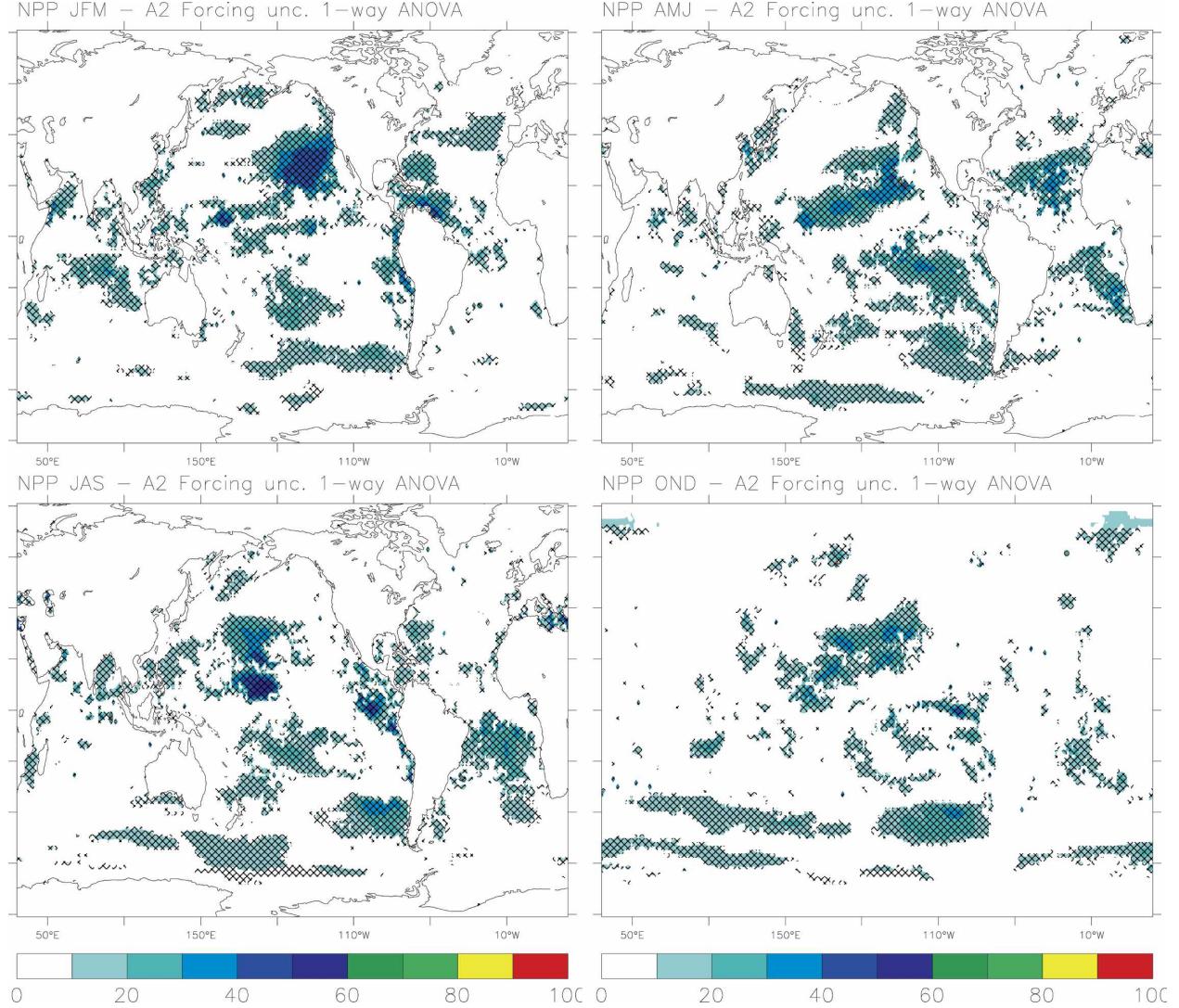


FIG. 7. The proportion of the total variance resulting from the prescribed forcing under the A2 scenario in the indicated seasonal projections of SWH_{20} obtained with the NPP model. Hatching as in Fig. 1.

ing into account the time variability of the location parameter, thus making use of the definitions of return value presented in (7) and just before (11), and in each case computed the relative root-mean-square error (rmse) between the 20-yr maxima and the 20-yr return-value estimates.² Figure 8 presents the relative rmse of the seasonal 20-yr return-value estimates based on the NPP and NS-GEV models. In most of the cases the relative rmse is below 20% and there seems to be no

clear spatial pattern of regions where the fits are poor. Both models seem to fit the data equally well, that is, differences between the relative rmses in the two approaches are negligible. This indicates that in terms of relative rmse the choice of the nonstationary extreme value approach is not important.

We next compare the two approaches in terms of differences between their estimates of SWH_{20} . Figure 9 presents the percentage differences between the NPP and NS-GEV seasonal estimates of SWH_{20} as for 2001 based on inferences from the ERA-40 data. (Any other year could have been chosen in place of 2001; the conclusions that follow do not depend on the year chosen for illustration.) In most of the regions the differences between the estimates is less than 2.5%, but in some

² We define the relative rmse as $\sqrt{\sum_{i=1}^{25} (y_i - x_i)^2 / \bar{x}}$, where the x_i s are the 20-yr maxima extracted from the 20-yr subsets of ERA-40, the y_i s are the corresponding 20-yr return-value estimates obtained from the nonstationary extreme value model (NPP or NS-GEV), and \bar{x} is the mean of the x_i s.

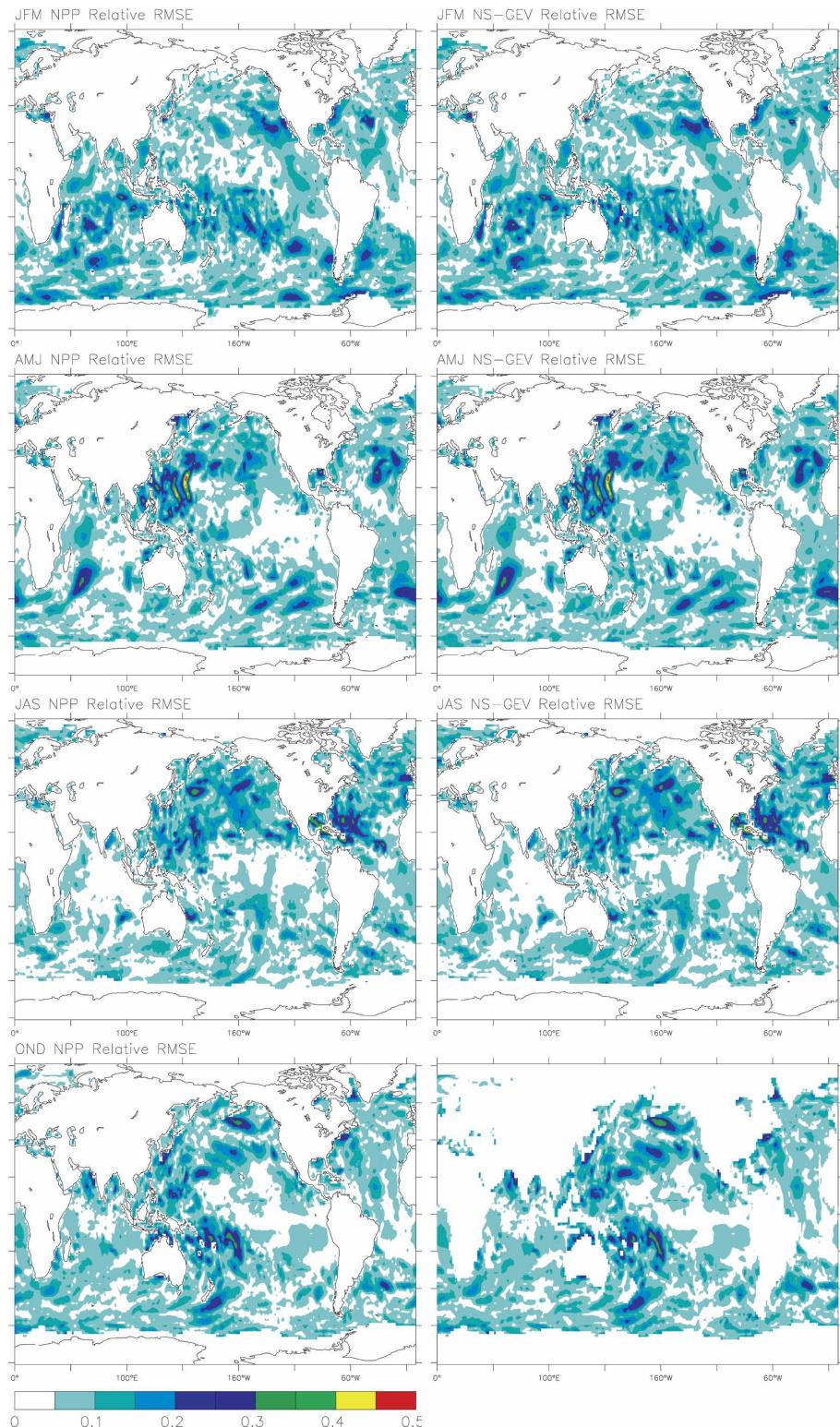


FIG. 8. Relative rmse of the seasonal SWH₂₀ ERA-40 estimates based on the (left) NPP and (right) NS-GEV models.

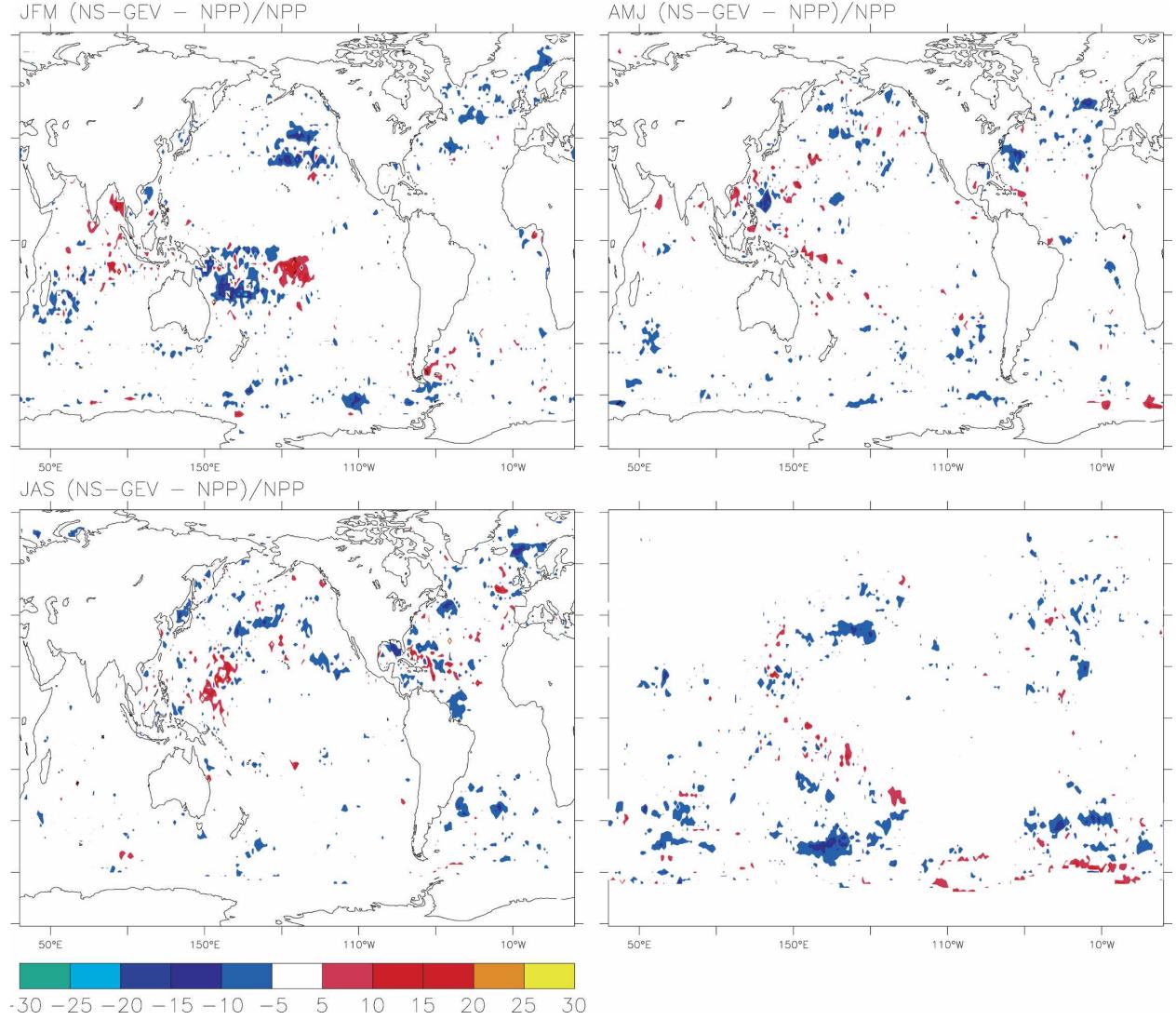


FIG. 9. Percentage differences between the seasonal values of SWH_{20} based on the NPP and NS-GEV model applied to the ERA-40 estimates as for 2001 [$(\text{NS-GEV } \text{SWH}_{20} - \text{NPP } \text{SWH}_{20}) / (\text{NPP } \text{SWH}_{20})$].

regions (mainly in the Tropics) these differences can be as large as 20%. We have looked for regions where the estimates from the models are incompatible in the sense that their 95% confidence intervals do not intersect, but there are no such regions; this suggests that even differences as large as 20% between the NPP and NS-GEV estimates of SWH_{20} are generally not statistically significant. Figure 10 presents the width of the confidence intervals of the seasonal SWH_{20} ERA-40 estimates from the NPP and the NS-GEV as for 2001 in terms of percentage of the corresponding estimate. From the figure we can see that the uncertainty of the estimates is seasonally dependent, that the spatial patterns of uncertainty coincide in the NPP and NS-GEV estimates, and that uncertainty is generally larger in the

NS-GEV estimates (which was already expected because the NS-GEV model uses less data than the NPP model). Comparing Figs. 9 and 10, it is clear that the regions where large differences between the estimates from the two models exist generally coincide with the regions where their uncertainty is large.

We now investigate in a more formal way whether the choice of the extreme value approach affects the projections of SWH_{20} . This is done by means of two-way ANOVA by which we also assess the influence of the choice of the forcing scenario. The time series we consider consists of the changes from 1990 to 2099 in the projections of SWH_{20} (with respect to the corresponding estimate as for 1990). We have created a time series combining all of the ensemble members under

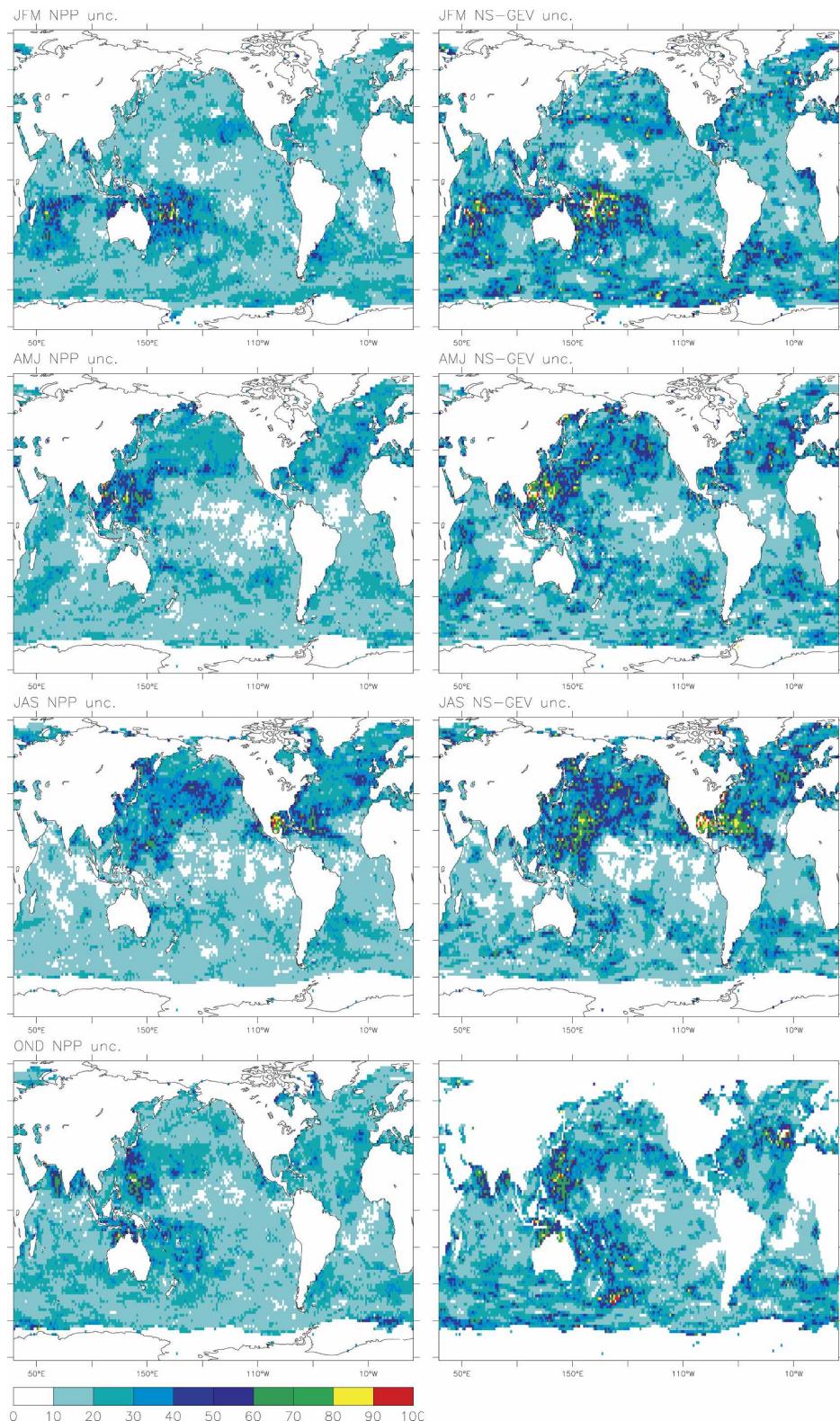


FIG. 10. Width of the 95% confidence intervals of the seasonal SWH_{20} ERA-40 estimates based on the (left) NPP and (right) NS-GEV models as for 2001 in terms of percentage of the corresponding estimate.

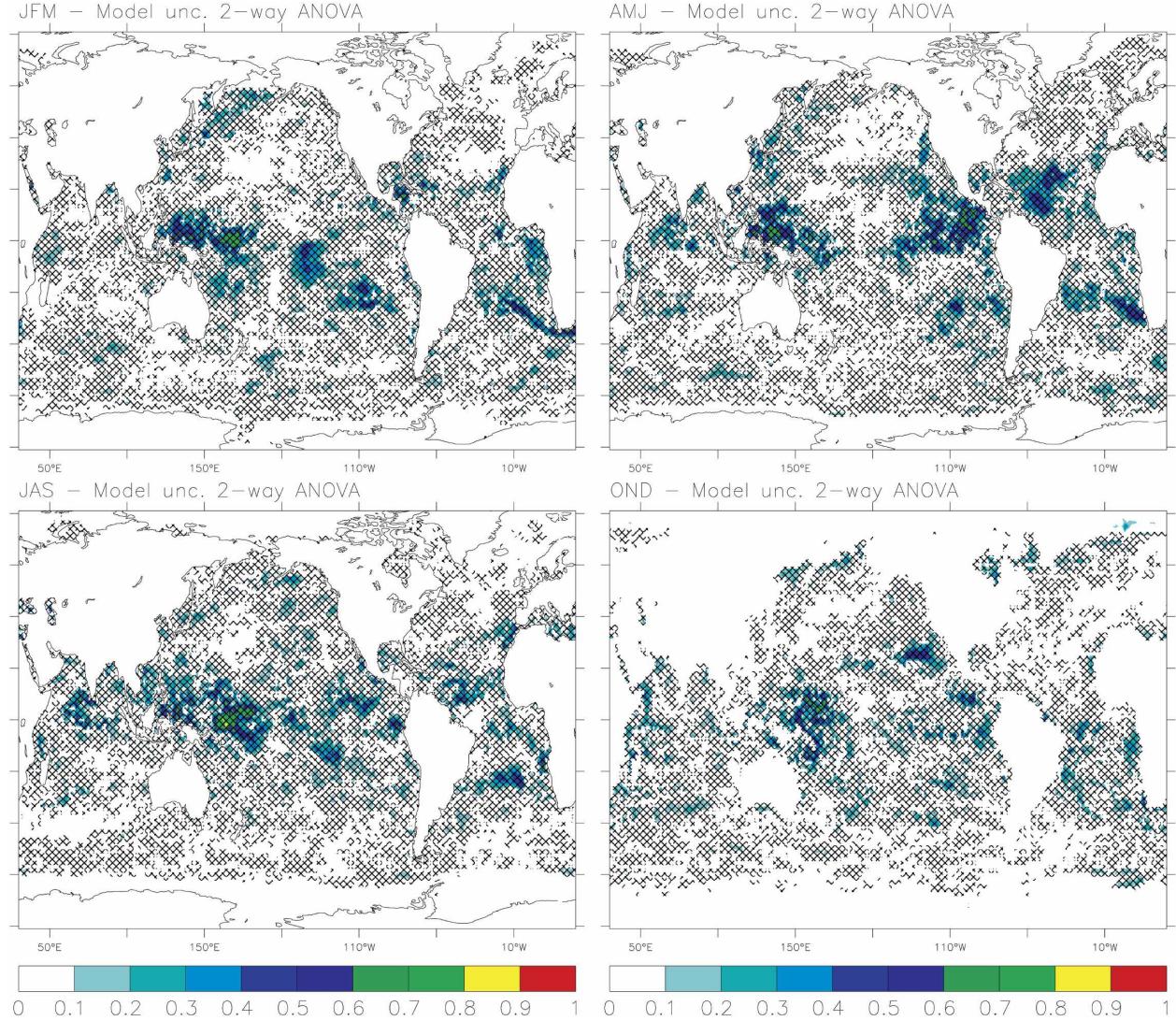


FIG. 11. The proportion of the total variance in the seasonal SWH_{20} that is due to the choice of the nonstationary models of extremes. Hatching as in Fig. 1.

each scenario by computing the ensemble mean at each point in time.

The results of the two-way ANOVA are presented in Figs. 11–13. Regarding differences in variances between the time series of projected SWH_{20} values using the NPP and the NS-GEV models, the Barlett test rejects homogeneity of variances at a 5% significance level in about 75% of the locations. This clearly suggests a systematic departure from the assumptions of the ANOVA model. However, we know that the F tests in the ANOVA model are robust against the inhomogeneity of variances if the sample sizes are large and similar in each group, which is so in our case, so we may still trust the results of the F tests, keeping in mind that the values given in Figs. 11–13 should not be read too lit-

erally. As a double check on the results of the F tests, we have additionally applied the paired Student's t test (see, e.g., Rice 1995) to the time series of SWH_{20} projections based on the NPP and NS-GEV models, and found that differences in mean are indeed significant at the 5% level over roughly the same regions as those in which the two-way ANOVA analysis detects significant differences.

Figure 11 presents the seasonal proportion of the total variance resulting from the choice of the nonstationary model for extremes and the regions where this effect is statistically significant at a 5% level. Statistically significant results occur over large extents of the ocean [i.e., is highly field significant (Livezey and Chen 1983)], and the proportion of variance resulting from it

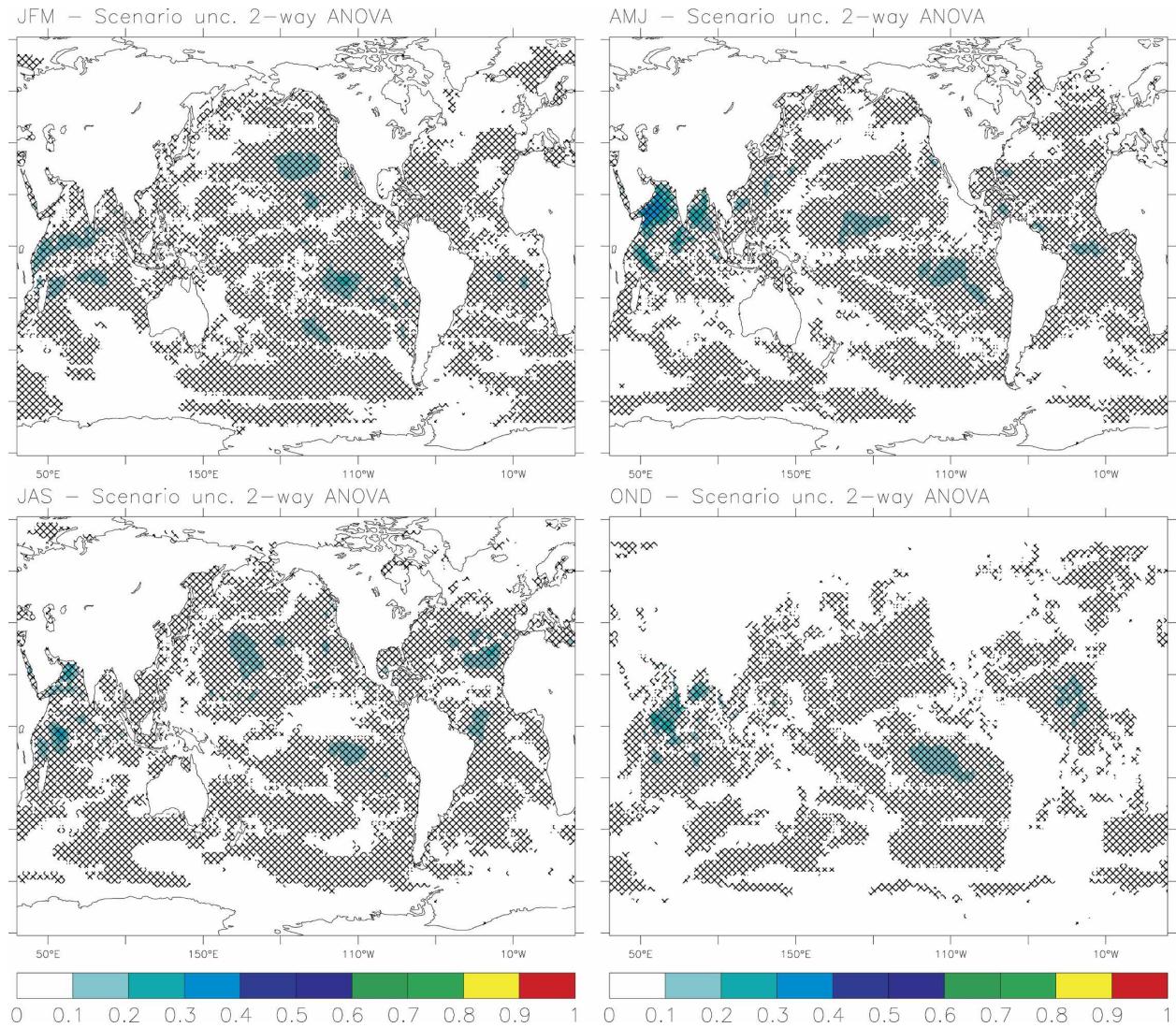


FIG. 12. The proportion of the total variance in the seasonal SWH_{20} that is due to the effect of the forcing scenario. Hatching as in Fig. 1.

can be quite large in the regions where the uncertainty of the NPP and NS-GEV estimates was large. However, in most of the regions where larger changes of SWH_{20} are to be expected, the uncertainty resulting from the choice of the NPP or NS-GEV model is often small and in some cases is even statistically insignificant.

Figure 12 presents the seasonal proportion of the total variance resulting from the choice of the future forcing scenario and the regions where this effect is statistically significant at a 5% level. The effect is statistically significant in large extents of the oceans and the proportion of variances resulting from it can go up to 50% but is in most of the oceans below 10%. The proportion of the total variance resulting from this effect is large, especially in the regions where higher changes in the

projections of SWH_{20} are to be expected. Wang and Swail's (2006) estimates of the total variance resulting from the choice of the future forcing scenario, which were presented only for JFM and JAS, are also mostly 10% but have a spatial pattern not fully coinciding with ours (cf. their Fig. 8).

Figure 13 shows the seasonal proportion of the total variance resulting from the interaction between the choice of the nonstationary model for extremes and the forcing scenario and the regions where this effect is statistically significant at a 5% level. The proportion of the total variance resulting from the interaction is much smaller than that from each of the effects, in most cases below 10%. The regions where the interaction effect is statistically significant are small and mainly in the Trop-

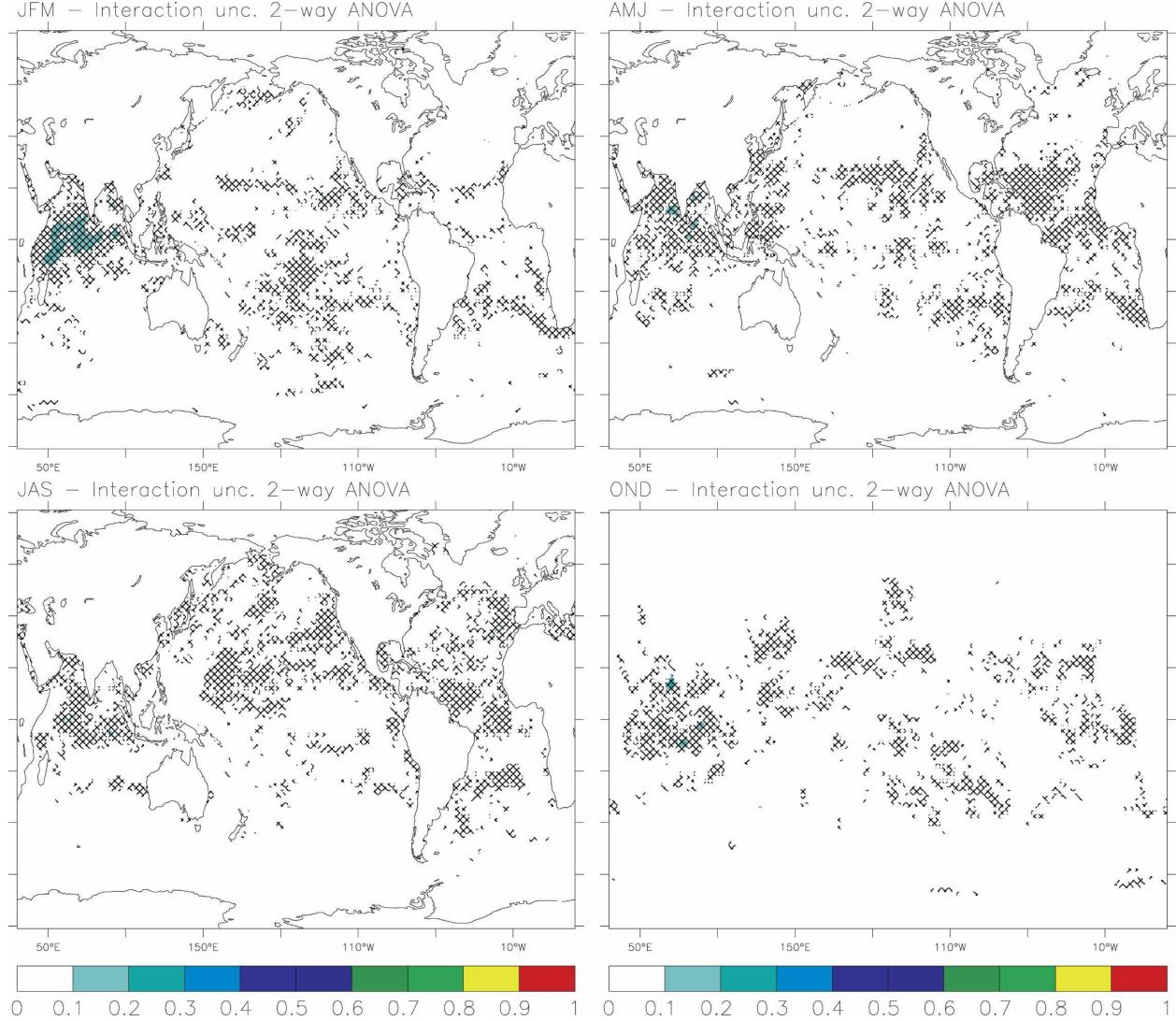


FIG. 13. The proportion of the total variance in the seasonal projections of SWH_{20} resulting from the interaction between choice of nonstationary model of extremes and climate scenario. Hatching as in Fig. 1.

ics, especially in the IO. These regions often coincide with regions where the proportion of the total variance resulting from each of the effects is larger.

Recapping, in this section we have seen that although the results of the extreme value analyses show that the NPP and NS-GEV time series of projected SWH return values are highly correlated and seem statistically compatible, their differences in means and variances are statistically significant when comparing the NPP and NS-GEV projected time series of SWH return values as point estimates (without taking into account their uncertainty) by means of an ANOVA analysis.

6. Conclusions

We have found that a NPP model with a location parameter depending on SLP-derived covariates,

namely, the seasonal mean SLP anomaly and the seasonal SLP gradient index, can be used to describe the present climate of extremes of SWH, as given by the ERA-40 dataset. The spatial pattern of the coefficients linking the location parameter with the SLP-derived covariates seems to suggest that the generation of extreme values of SWH involves different processes in the Tropics and at high latitudes.

The NPP model with parameters estimated from present climate data was then used to compute seasonal projections of SWH_{20} from 1990 to 2099 under the IS92a, A2, and B2 future forcing scenarios using CGCM2 SLP data. Under all forcing scenarios, significant changes in SWH_{20} are to be expected in different regions of the globe with the larger and more significant changes occurring under the more severe GHG emis-

sion scenarios, namely, the A2 and IS92a scenarios. In some cases the rate of future changes depends on the scenario with, for instance, the A2 projections showing a slower rate of change in the initial decades and a higher rate in the later decades when compared with the rates of change in the IS92a projections. It is interesting to note that under all of the future scenarios considered significant positive trend are to be expected in the NP. Positive trends are also present in the present climate of extremes in the NP (Wang and Swail 2005; Caires and Swail 2004), which maybe are already a consequence of increased GHG emissions.

The uncertainty resulting from the choice of nonstationary model for extremes was assessed by obtaining estimates based on the NS-GEV model and comparing them with those based on the NPP model. The estimates from the two models are compatible in the sense that the same dependence on the parameters on the covariates is found, that both projections exhibit the same type of trends, and that estimates at a fixed time point have intersecting 95% confidence intervals. However, the width of the confidence intervals can be quite large—larger than the projected changes—especially in the case of the NS-GEV estimates. The rather large width of the confidence intervals, that is, the rather large uncertainty in the estimates, largely explains the significant differences between changes detected by ANOVA in the time series of SWH_{20} projections obtained from the NPP or the NS-GEV models. These differences occur mainly in the Tropics.

We have measured the effect of the prescribed forcing, forcing scenario, and choice of nonstationary model for extremes in the projected SWH_{20} time series and found that all effects are significant in certain regions. The prescribed forcing and forcing scenario effects are especially large in the regions where larger SWH_{20} changes are expected to occur, and the choice of nonstationary model for extremes is large where the confidence intervals of the NPP and NS-GEV estimates are large, which often do not coincide with the regions where larger SWH_{20} changes are expected to occur.

There are some caveats about the results presented here:

- The projected future changes in SWH_{20} presented here are based only on SLP projections under future forcing scenarios produced by CGCM2. The uncertainty resulting from the choice of climate model was studied by Wang and Swail (2006), whose results show that the uncertainty resulting from the climate model used can be large, especially in the Tropics, where differences between different reanalysis datasets were also found to be large (Caires et al. 2004).
 - Certain climate scenarios predict a retreat of the sea ice cover and therefore the estimates we provide around the sea regions presently covered with ice must be interpreted with care. In our analysis for ERA-40 grid points located in presently partially ice covered regions, observations at times when the grid points are covered were treated as exception/missing values. Eventual waves in grid points at which the sea ice cover may disappear in the future were not considered in this analysis because no present climate wave data in those regions is available. Our opinion is that, because no strong storms originate in those regions, except for the appearance of low waves and associated low extremes in the grid points that eventually may no longer be covered by ice, the effect on the remaining grid points will be negligible.
- Acknowledgments.** We thank ECMWF for the ERA-40 data and the IPCC Data Distribution Centre for the CGCM2 data. We are indebted to Clive Anderson and José Ferreira for statistical advice in the initial stages of this research and to Andreas Sterl, Xuebin Zhang, and the two anonymous referees for comments that led to major improvements on earlier versions of this manuscript.

REFERENCES

- Anderson, C. W., D. J. T. Carter, and P. D. Cotton, 2001: Wave climate variability and impact on offshore design extremes. Report for Shell International and the Organization of Oil and Gas Producers, 99 pp. [Available online at http://info.ogp.org.uk/metocean/JIPweek/WCEReport_2sided.pdf.]
- Boer, G. J., M. C. Flato, M. C. Reader, and D. Ramsden, 2000: A transient climate change simulation with greenhouse gas and aerosol forcing: Experimental design and comparison with the instrumental record for the twentieth century. *Climate Dyn.*, **16**, 405–425.
- Caires, S., and V. Swail, 2004: Climatological assessment of reanalysis ocean data. *Eighth Int. Workshop on Wave Hindcasting and Forecasting*, Oahu, HI, WMO and IOC, WMO Tech. Doc. 1319, JCOMM Tech. Rep. 29, DVD-ROM.
- , and A. Sterl, 2005: 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data. *J. Climate*, **18**, 1032–1048.
- , A. Sterl, J.-R. Bidlot, N. Graham, and V. Swail, 2004: Intercomparison of different wind wave reanalyses. *J. Climate*, **17**, 1893–1913.
- Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, 226 pp.
- Cox, A. T., and V. R. Swail, 2001: A global wave hindcast over the period 1958–1997: Validation and climate assessment. *J. Geophys. Res.*, **106**, 2313–2329.
- Flato, G. M., and G. J. Boer, 2001: Warming asymmetry in climate change simulations. *Geophys. Res. Lett.*, **28**, 195–198.
- Jorgensen, B., 1993: *The Theory of Linear Models*. Chapman and Hall, 240 pp.

- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59.
- Nakicenovic, N., and R. Swart, Eds., 2000: *Emissions Scenarios*. Cambridge University Press, 570 pp.
- Press, W. H., S. A. Flannery, B. P. Teukolsky, and W. T. Vetterling, 1992: *Numerical Recipes in FORTRAN: The Art of Scientific Computing*. Cambridge University Press, 922 pp.
- Rice, J. A., 1995: *Mathematical Statistics and Data Analysis*. 2d ed. Duxbury Press, 650 pp.
- Seber, G. A. F., and A. J. Lee, 2003: *Linear Regression Analysis*. 2d ed. Wiley, 582 pp.
- Smith, R. L., 1989: Extreme value analysis of environmental time series: An application to trend detection in ground level ozone (with discussion). *Stat. Sci.*, **4**, 367–393.
- Storch, H. V., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.
- Swail, V. R., and A. T. Cox, 2000: On the use of NCEP–NCAR reanalysis surface marine wind fields for a long-term North Atlantic wave hindcast. *J. Atmos. Oceanic Technol.*, **17**, 532–545.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- Wang, X. L., and V. R. Swail, 2001: Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes. *J. Climate*, **14**, 2204–2221.
- , and —, 2005: Historical and possible future changes of wave heights in northern hemisphere oceans. *Atmosphere–Ocean Interactions*, W. Perrie, Ed., Advances in Fluid Mechanics Series, Vol. 39, Wessex Institute of Technology Press, 300 pp.
- , and —, 2006: Climate change signal and uncertainty in projections of ocean wave height. *Climate Dyn.*, **26**, 109–126.
- , F. W. Zwiers, and V. R. Swail, 2004: North Atlantic Ocean wave climate change scenarios for the twenty-first century. *J. Climate*, **17**, 2368–2383.