

## Analyzing Monthly Extreme Sea Levels with a Time-Dependent GEV Model

FERNANDO J. MÉNDEZ AND MELISA MENÉNDEZ

*Ocean and Coastal Research Group, Departamento Ciencias y Técnicas del Agua y del Medio Ambiente, Universidad de Cantabria, Santander, Spain*

ALBERTO LUCEÑO

*Departamento Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander, Spain*

INIGO J. LOSADA

*Ocean and Coastal Research Group, Departamento Ciencias y Técnicas del Agua y del Medio Ambiente, Universidad de Cantabria, Santander, Spain*

(Manuscript received 13 March 2006, in final form 14 September 2006)

### ABSTRACT

A statistical model to analyze different time scales of the variability of extreme high sea levels is presented. This model uses a time-dependent generalized extreme value (GEV) distribution to fit monthly maxima series and is applied to a large historical tidal gauge record (San Francisco, California). The model allows the identification and estimation of the effects of several time scales—such as seasonality, interdecadal variability, and secular trends—in the location, scale, and shape parameters of the probability distribution of extreme sea levels. The inclusion of seasonal effects explains a large amount of data variability, thereby allowing a more efficient estimation of the processes involved. Significant correlation with the Southern Oscillation index and the nodal cycle, as well as an increase of about 20% for the secular variability of the scale parameter have been detected for the particular dataset analyzed. Results show that the model is adequate for a complete analysis of seasonal-to-interannual sea level extremes providing time-dependent quantiles and confidence intervals.

### 1. Introduction

The knowledge of the statistical distribution of extreme sea levels is of practical importance for many purposes, including coastal management, flooding of urban areas and valuable ecosystems, and the design of maritime works. The simplest method is the annual maximum method (AMM), which models annual maximum data using the generalized extreme value distribution. Because of the scarcity of data, some alternatives have been proposed for different geophysical variables, such as the method of the  $r$ -largest maxima (e.g., Smith 1986; Tsimplis and Blackman 1997; Unnikrishnan et al. 2004; Soares and Scotto 2004) or the peak

over threshold approach (e.g., Smith 1989; Davison and Smith 1990; Goda 2000).

The character of the sea level (composed, after averaging out wind surface waves, of three additive components: mean sea level, tidal level, and surge level) and the evaluation of the tide as a deterministic process has led to a number of indirect methods trying to make use of this information [see details of the Revised Joint Probability Method (RJPM) in Tawn and Vassie (1989)]. One important characteristic of all the aforementioned methods is that they often assume the sea level to be a homogeneous random variable, which implies stationarity of the observed process.

However, in the context of environmental processes, such as the extreme sea level, a nonstationary behavior is found at different time scales (seasonal, decadal, and secular). Seasonality is evident in the three components of sea level: monthly mean sea levels (Tsimplis and Woodworth 1994), surge levels (see Pugh 2004, chapter 6), and the highest tide levels (Zetler and Flick 1985;

---

*Corresponding author address:* Fernando J. Méndez, Ocean and Coastal Research Group, Dpto. Ciencias y Técnicas del Agua y del Medio Ambiente, Universidad de Cantabria, Avda. de los Castros s/n, 39005 Santander, Spain.  
E-mail: mendezf@unican.es

Dixon and Tawn 1999). Summing up the three components, the monthly maximum sea level is variable throughout a year and, in addition, the month of occurrence of the highest water level can fluctuate (Smith and Leffler 1980).

Decadal variability or interannual fluctuation is also present in the extreme sea level behavior. This variability is usually expressed in terms of regional climate indices such as the Southern Oscillation index (SOI), the Pacific decadal oscillation (PDO), or the North Atlantic Oscillation (NAO). Recently, Woodworth and Blackman (2004) analyzed a large number of tidal gauges and found a close relation between extreme sea levels and climate indices, supporting the existence of decadal variability in extreme sea levels. Moreover, there exists external forcing acting on the globe, due to the sun and the moon affecting sea level. The 11-yr sunspot cycle can affect extreme sea levels, as pointed out by Smith (1986). Concerning the effect of the moon, there are several cycles (e.g., the 18.61-yr nodal cycle due to the regression of the lunar nodes) that affect the amplitude of the semidiurnal lunar tide as well as the mean sea level (Pugh 2004, chapters 4 and 7).

Concerning secular trends, mean sea level rise is one of the main contributions to extreme sea levels (e.g., Woodworth et al. 1999; Flick et al. 2003; Pugh 2004, chapter 7). This aspect is well known and has been widely studied by the scientific community because of the potential effects on coastal areas (Houghton et al. 2001). Lately, there is some evidence around the globe that suggests that not only sea level rise but also long-term changes in the intensity and frequency of the storminess must be considered (e.g., Meehl et al. 2000; Woodworth and Blackman 2004; Bijl et al. 1999; Bell et al. 2000; Flick et al. 2003; Langenberg et al. 1999; Bromirski et al. 2003; d'Onofrio et al. 1999; Woodworth and Blackman 2002).

Therefore, it would be highly desirable to model the nonstationary behavior of extreme sea levels at different time scales. One possibility for dealing with nonstationarity is the removal of trends, obtaining a net data series suitable for the classical extreme value analysis. This approach is widely used in time series analysis and has been used recently for predicting extreme high water levels in San Francisco (SFO), California (Sobey 2005). Nevertheless, this data precondition is not necessary because the time-dependent generalized extreme value distribution can describe the variability as a function of time or other covariates (see, e.g., Coles 2001, chapter 6; Smith 2001, chapter 8; Katz et al. 2002). Not surprisingly, some time scales have already been analyzed. For example, Smith (1986) and Tawn (1988)

use the  $r$ -largest annual maxima method to study secular variability (linear and quadratic trends) in Venice, Italy; and Lowestoft and Great Yarmouth, United Kingdom; respectively. The first author also models the nodal and the sunspot cycle. The SOI is introduced as a covariate by Coles (2001) to model annual maximum sea levels at Fremantle, Australia, obtaining a significant influence of this climate index. Seasonality is introduced in the RJPM by Dixon and Tawn (1999) and in the peak over threshold approach by several authors. Smith (1989) studied the ground-level ozone, Smith (2001) analyzed wind speeds, and Katz et al. (2002) modeled precipitation maxima, obtaining an annual cycle to describe the behavior of the rainfall throughout the year. However, to the authors' knowledge, no global state-of-the-art seasonal-to-interannual analysis is currently available concerning extreme sea levels. In this work, a global framework designed to cope with all the time scales simultaneously is presented.

The model suggested in this paper is based on the time-dependent generalized extreme value distribution. The nonstationary behavior of extreme sea levels is parameterized using functions of time (linear, quadratic, exponential, and cosine functions) or covariates (e.g., SOI) and degrees of freedom are introduced in the model correspondingly. Independent monthly maxima series are adopted, allowing the efficient modeling of seasonality given that the 12-yearly data are quasi-homogeneously time distributed. As pointed out by Katz et al. (2002), another reason for choosing monthly maxima is that they provide additional information about the upper tail of the distribution. We adopt a "direct method," avoiding the decomposition between the tidal and surge parts (as in the RJPM "indirect method") and the nonlinear interaction. Because tidal levels tend to have a short-tailed distribution whereas surge levels often tend to have longer-tailed distributions, we will always check that the monthly maxima approach using time-dependent parameters allows for an adequate statistical definition of the upper tail of the extreme sea levels, by using several goodness-of-fit and plot measures. The model is applied to a well-known large hourly tidal gauge series (San Francisco) showing the characteristics of the seasonal-to-interannual time scales at this particular site. This approach is also useful when historical information is not hourly recorded, provided that daily or monthly high-water data are known (Bijl et al. 1999).

The paper is organized as follows. Section 2 provides a brief description of the time-dependent generalized extreme value distribution, the method of estimation of parameters, and the criteria for model selection. Section 3 describes the data from the San Francisco tidal

gauge. Section 4 explains the different time scales involved and suggests their possible parameterization. The application of the model is shown in section 5. Finally, some conclusions are given in section 6.

## 2. Methodology

### a. The generalized extreme value distribution in homogeneous populations

Extreme value theory is concerned with the quantification of the stochastic behavior of processes when they approach very large (or very small) values. The annual maximum method is a simple example of an application of that theory. This method uses time series of block maxima for successive years,  $\{Z_t = \max(X_1, \dots, X_N)\}$ , which are called annual maxima series (AMS), where  $t = 1, \dots, n$ , the  $X_t$ s are the reading of the underlying variables throughout the  $N$  sampling intervals in the year, and  $n$  is the number of years. The method assumes that annual maxima of successive years are independent and identically distributed (IID) random variables so that successive years are climatologically and statistically independent. Because only one datum (the annual maximum) per year is used, simplicity in the model is attained at the price of exacerbating scarcity of data and making the quantification of seasonal trends impossible.

A classical theorem in extreme value theory (see Coles 2001, chapter 3) states that, if they exist, the limiting distributions of sequences of appropriately normalized maxima of IID random variables must belong to one of three possible families of probability distribution functions, namely the Weibull, Gumbel, and Fréchet families. The generalized extreme value (GEV) distribution is a reformulation of these limiting families; the corresponding cumulative distribution function (CDF) is given by

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

where  $1 + \xi(z - \mu)/\psi > 0$ ,  $-\infty < \mu < \infty$  is a location parameter,  $\psi > 0$  is a scale parameter, and  $-\infty < \xi < \infty$  is a shape parameter. The Weibull, Gumbel, and Fréchet families correspond to  $\xi < 0$ ,  $\xi = 0$ , and  $\xi > 0$ , respectively. Although this classical theorem applies only to sequences of IID random variables ( $X_1, \dots, X_N$ ), it also holds when the hypotheses are relaxed

moderately, for example, when there exists a relatively weak statistical dependence between these random variables [e.g., the maxima of time series generated by Gaussian stationary autoregressive moving average (ARMA) models converge to the Gumbel CDF; see Galambos 1978, chapter 3]. Consequently, the theorem can be used in practice if every year can be conceived to be composed of many small “sampling” intervals (e.g., hours), such that successive values of the process are approximately identically distributed and show “relatively weak” statistical dependence [e.g., if  $\text{corr}(X_i, X_{i+m}) \log m \rightarrow 0$  as  $m$  increases, where  $m$  is the time lag; see Galambos 1978].

### b. The time-dependent GEV distribution

To mitigate the problem of data scarcity and also to be able to model seasonal trends, it is necessary to consider more than one extreme datum per year. There are three main approaches for choosing the data to be considered within each year: 1) monthly maxima series, 2) exceedances over large thresholds, and 3)  $r$ -largest maxima within the year. For example, Carter and Challenor (1981) and Katz et al. (2002) use the first method; Smith (1989) and Davison and Smith (1990) use the second method; whereas Smith (1986), Tawn (1988), Tsimplis and Blackman (1997), Unnikrishnan et al. (2004), and Soares and Scotto (2004) use the third method.

When using monthly maxima series (MMS), the same reasoning of section 2a can be used to justify that the maximum  $Z_t$  observed in month  $t$  follows a GEV distribution with time-dependent parameters  $-\infty < \mu(t) < \infty$ ,  $\psi(t) > 0$ , and  $-\infty < \xi(t) < \infty$ . Although the number of components  $N$  in vector  $(X_1, \dots, X_N)$  decreases with respect to annual maxima (which deteriorates the asymptotic approximation), the random variables  $X_t$ s can reasonably be expected to be more homogeneous within months than within years (which may be expected to improve the asymptotic approximation and the robustness of the statistical model to inhomogeneities in the series). The functions  $\mu(t)$ ,  $\psi(t)$ , and  $\xi(t)$  can contain sine waves representing seasonal effects (see, e.g., Katz et al. 2002), linear and/or exponential terms representing long-term trends, and covariates representing environmental processes (e.g., El Niño), among others.

We will analyze the performance of the MMS method throughout the paper. In the following, the procedures to estimate the parameters and the quantiles as well as to calculate the confidence intervals are shown. Criteria for model diagnostic and selection are also addressed.

*c. Parameter estimation*

We use the method of maximum likelihood to estimate the model parameters. The location, scale, and shape parameters  $\mu(t)$ ,  $\psi(t)$ , and  $\xi(t)$  are expressed in terms of regression parameters and covariates (Coles 2001). The complete vector of  $p$  regression parameters is denoted by  $\theta$ . The likelihood function of the param-

eters, given  $m$  observations  $\{(t_1, z_1), \dots, (t_m, z_m)\}$  of the period  $t_i$  at which the maximum  $z_i$  is attained, is provided by

$$L(\theta|t_i, z_i) = \prod_{i=1}^m g[z_i; \mu(t_i), \psi(t_i), \xi(t_i)], \tag{2}$$

where

$$g(z; \mu, \psi, \xi) = \frac{1}{\psi} \left[ 1 + \xi \left( \frac{z - \mu}{\psi} \right) \right]^{-(1+1/\xi)} \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\psi} \right) \right]^{-1/\xi} \right\} \tag{3}$$

is the density function of the GEV distribution, which is obtained by differentiating Eq. (1) with respect to  $z$ . The log-likelihood function is

$$l(\theta|t_i, z_i) = - \sum_{i=1}^m \left\{ \log \psi(t_i) + (1 + 1/\xi(t_i)) \log \left[ 1 + \xi(t_i) \left( \frac{z_i - \mu(t_i)}{\psi(t_i)} \right) \right] + \left[ 1 + \xi(t_i) \left( \frac{z_i - \mu(t_i)}{\psi(t_i)} \right) \right]^{-1/\xi(t_i)} \right\}, \tag{4}$$

provided that  $\psi(t_i) > 0$  and  $1 + \xi(t_i)(z_i - \mu(t_i))/\psi(t_i) > 0$ , for  $i = 1, \dots, m$ . For every value of  $\xi(t_i)$  that equals zero, it is necessary to use the appropriate limiting form, replacing the GEV by the Gumbel [Eq. (1) for  $\xi = 0$ ] log-likelihood function,

$$l(\theta|t_j, z_j) = -\log \psi(t_j) - \frac{z_j - \mu(t_j)}{\psi(t_j)} - \exp \left[ - \frac{z_j - \mu(t_j)}{\psi(t_j)} \right]. \tag{5}$$

Maximization of Eq. (4) and/or Eq. (5) yields the maximum likelihood estimate of  $\theta$ , denoted by  $\hat{\theta}$ . As shown below, some of the proposed models have more than 15 regression parameters, making it difficult to maximize the log-likelihood function. Nevertheless, an efficient global optimization procedure has been used successfully in all the cases analyzed, namely, the shuffled complex evolution (SCE) algorithm (Duan et al. 1992), which is widely used in the calibration of highly nonlinear problems. Because the SCE optimization scheme uses random starting points, we have confirmed that the solutions  $\hat{\theta}$  we provide were the unique optimal solutions by running the program several times for every model.

Approximate standard errors for the estimators and confidence intervals for the regression parameters are obtained using standard likelihood theory (see, e.g., Coles 2001). For a large sample size  $n$ , the distribution of  $\hat{\theta}$  is approximately multivariate normal with mean  $\theta$  and covariance matrix given by the inverse of the observed information matrix  $\mathbf{I}_O(\theta)$ ,

$$\mathbf{I}_O(\theta) = - \frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}, i, j = 1, \dots, p, \tag{6}$$

evaluated at  $\theta = \hat{\theta}$ . If an arbitrary term in the inverse of  $\mathbf{I}_O(\theta)$  is  $s_{ij}$ , the square root of the diagonal entry—( $i, i$ )th element,  $s_{ii}$ —is approximately the standard error  $se(\hat{\theta}_i)$  of the maximum likelihood estimator  $\hat{\theta}_i$ . Therefore, confidence intervals for  $\theta_i$  can be obtained in the form  $[\hat{\theta}_i - z_\alpha se(\hat{\theta}_i), \hat{\theta}_i + z_\alpha se(\hat{\theta}_i)]$ , where  $z_\alpha = 1.96$  gives a 95% confidence interval.

*d. Model selection*

The selection of models can be implemented almost automatically when the candidate models are nested with a “maximal” model embracing all the remaining models. In this case, every model can be obtained from the maximal model by setting some of their parameters to constants. If the number of models to be compared is small, model selection can be performed using the maximum likelihood ratio test. Thus, suppose model A can be obtained from model B by setting  $p_B - p_A$  parameters to constants, where  $p_A < p_B$  are the number of parameters in each model. Let  $\hat{l}(p_A) \leq \hat{l}(p_B)$  be the maxima of their respective log-likelihood functions, then model A may be rejected if  $2[\hat{l}(p_B) - \hat{l}(p_A)] > \chi^2(p_B - p_A, 1 - \alpha)$ .

If the number of nested candidate models is large, it is more appropriate to assess the quality of each model by using a penalized function  $-2\hat{l}(p) + pC(n)$ , where  $p$  is the number of parameters and  $\hat{l}(p)$  is the maximum of the log-likelihood resulting from each model. The Akaike information criterion (AIC), the Hannan and Quinn criterion (HQC), and the Bayesian information criterion (BIC; also called Schwarz criterion) corre-

spond to the choices  $C(n) = 2$ ,  $C(n) = \log[\log(n)]$ , and  $C(n) = \log(n)$ , respectively.

Each of these criteria assesses the quality of each model by making a compromise between obtaining a good fit, which is measured by how small the resulting  $-2\hat{l}(p)$  term is, and using a simple model, where simpler models use less parameters than complex models. Therefore, the smaller the criterion, the better the model. Clearly, the HQC and BIC criteria penalize complex models more than the AIC criterion when  $n$  is large.

### e. Model diagnostics

There are a number of procedures for model checking when data are assumed to be identically distributed. However, for a time-dependent random variable it is necessary to standardize the data conditional on the fitted parameter values. Considering that

$$Z_t \sim \text{GEV}[\hat{\mu}(t), \hat{\psi}(t), \hat{\xi}(t)], \quad (7)$$

one can define the standardized variables  $\bar{Z}_t$  as

$$\bar{Z}_t = \frac{1}{\hat{\xi}(t)} \log \left[ 1 + \hat{\xi}(t) \left( \frac{Z_t - \hat{\mu}(t)}{\hat{\psi}(t)} \right) \right], \quad (8)$$

which (conditional on the parameter estimates) follows the standard Gumbel distribution, with the probability distribution function

$$\Pr\{\bar{Z}_t \leq z\} = \exp\{-\exp(-z)\}. \quad (9)$$

Then, probability and quantile plots for the sample of computed values  $\bar{z}_t$  can be obtained using Eq. (8). If  $\bar{z}_{(1)}, \dots, \bar{z}_{(m)}$  are the corresponding sample order statistics, the plotting points (e.g., empirical versus model) for the probability plot are  $\{i/(m+1), \exp[-\exp(-\bar{z}_{(i)})]\}$  while the plotting points for the quantile plot are  $\{-\log(-\log[i/(m+1)]), \bar{z}_{(i)}\}$  for  $i = 1, \dots, m$ . If the estimated model is adequate, both the probability and quantile plots should consist of points close to the unit diagonal. It is important to note that both plots must be checked simultaneously as they contain the same information but on different scales.

### f. Determination of quantiles

The extreme quantiles  $z_q$  for a stationary GEV distribution are obtained by inverting Eq. (1) so that

$$z_q(\theta) = z_q(\mu, \psi, \xi) = \begin{cases} \mu - \frac{\psi}{\xi} [1 - \{-\log(1-q)\}^{-\xi}] & \xi \neq 0 \\ \mu - \psi \log\{-\log(1-q)\} & \xi = 0 \end{cases}. \quad (10)$$

The quantile estimate  $\hat{z}_q = z_q(\hat{\mu}, \hat{\psi}, \hat{\xi})$  is the return level associated with the return period  $1/q$ . Confidence intervals can be obtained, assuming approximate normality for the maximum likelihood estimators, by the delta method (Rice 1994). The standard error of  $z_q$  can be estimated from the formula

$$\text{se}(\hat{z}_q) = \left( \sum_{i=1}^p \sum_{j=1}^p \frac{\partial z_q(\theta)}{\partial \theta_i} \frac{\partial z_q(\theta)}{\partial \theta_j} s_{ij} \right)^{1/2}, \quad (11)$$

where the partial derivatives are evaluated at  $\hat{\theta}$ .

For nonstationary or time-dependent GEV parameters, the calculation of “effective” design values can be carried out using Eq. (10), so that the quantity varies depending on the time of the year (Katz et al. 2002).

## 3. The data

Hourly sea level data from the San Francisco tidal gauge is used in the analysis. SFO data (station 9414290) were obtained from the National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS) Centre for Operational Products and Services (CoOps), with observations beginning in 1901 and ending in 2003 (103-yr data series). As pointed out by Bromirski et al. (2003), San Francisco has the longest continuous tidal record in North America and it is suitable for the investigation of seasonal as well as interdecadal variability.

Annual maxima (circles), monthly maxima (dots), and monthly mean water levels (solid line) for the SFO station are illustrated in Fig. 1. Three monthly gaps appear in the time series (June 1946 and January and February 1978). As seen, the time series reveals an important positive trend as well as possible changes in the variance of the data, although some information is likely hidden due to visual-scale effects. To perform a more detailed examination of all these features, descriptive statistics and exploratory analysis are carried out in the following section.

## 4. Regression model

We shall consider different agents to explain the variability and nonstationarity in the distribution of maxima, such as the annual variability (seasonality), the mean sea level rise, a likely long-term trend for the extreme values, long-term climate variability effects explained by global meteorological indices or interannual astronomical modulations as the 18.61-yr nodal cycle. All these factors can be parameterized providing an easy interpretation of the variability of the data.

It is assumed that the time-dependent location, scale,

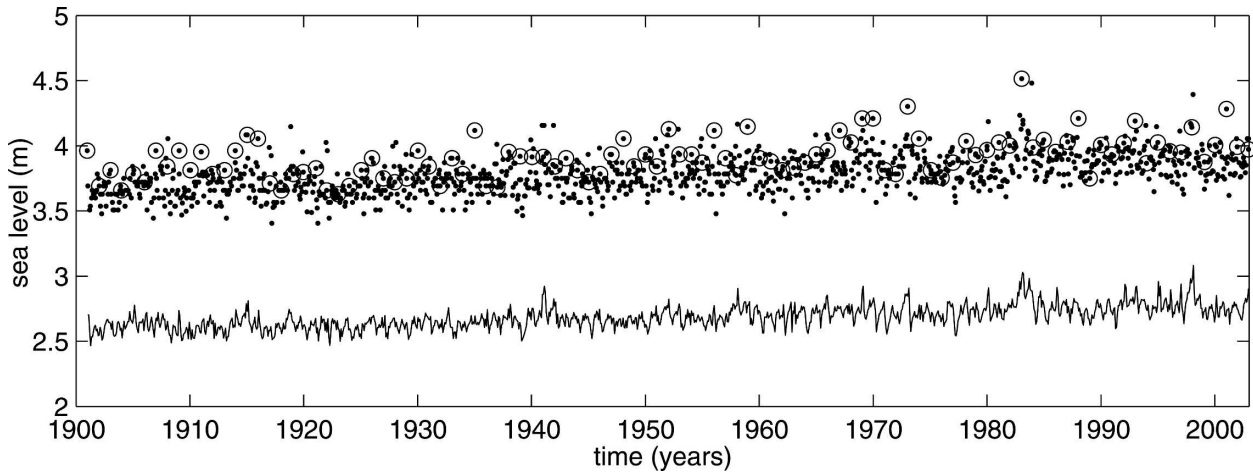


FIG. 1. Monthly mean (solid line), monthly (dots), and annual (circles) maxima series for the SFO tide gauge. (SFO: data from NOAA/NOS CoOps 9414290. Elevations to station datum. MLLW is +1.822 m.)

and shape parameters [ $\mu(t)$ ,  $\psi(t)$ , and  $\xi(t)$ ] of the GEV distribution can be split up in a number of processes with different time scales (we shall use subindex SLR for sea level rise,  $S$  for seasonality, LT for secular trends,  $N$  for astronomical forcing, and CLI for decadal climate variability). The following aggregations of factors are proposed:

$$\mu(t) = \mu_{\text{SLR}}(t) + \mu_S(t) \exp[\mu_{\text{LT}}(t)] + \mu_N(t) + \mu_{\text{CLI}}(t), \quad (12)$$

$$\psi(t) = \psi_S(t) \exp[\psi_{\text{LT}}(t)] + \psi_N(t) + \psi_{\text{CLI}}(t), \quad (13)$$

$$\xi(t) = \xi_S(t) \exp[\xi_{\text{LT}}(t)]. \quad (14)$$

This model introduces some nonlinearity to allow for a long-term variability in the seasonality (e.g., higher increase in winter than in summer). This is mathematically expressed by the coupling between seasonality and the long-term components  $\exp[\mu_{\text{LT}}(t)]$ ,  $\exp[\psi_{\text{LT}}(t)]$ , and  $\exp[\xi_{\text{LT}}(t)]$ . The exponential terms have the property that, for small long-term variations (say  $a \rightarrow 0$ ), the exponential  $\exp(at)$  is approximately equivalent to  $(1 + at)$ , so that  $100a$  can be considered as a yearly percentage of variability, when  $t$  is given in years. Sea level rise is only introduced in the location parameter [Eq. (12)] as it produces only a continuous vertical displacement of the random variable. For the shape parameter [Eq. (14)], only seasonal and secular changes are considered.

Table 1 displays possible regression models for every time-dependent GEV parameter. The corresponding regression parameters are shown in parentheses. The explanation of the parameterization for every component is given below. Note that the regression parameters corresponding to the time-dependent parameters  $\mu(t)$ ,  $\psi(t)$ , and  $\xi(t)$  are named  $\beta_i$ ,  $\alpha_i$ , and  $\gamma_i$ , respectively.

#### a. Seasonality

Monthly maxima are plotted against the day within the year (1–366) for the 103 yr in Fig. 2. The result reveals a clear seasonality along the year that suggests the need to fit a seasonal model to the data. Annual and semiannual cycles are detected with different amplitudes for the maxima throughout the year.

The variability of extreme sea levels within a year can also be addressed by analyzing the different agents involved: mean sea level (MSL), tide level, and surge level (defined as the detrended residual after the harmonic analysis to the 103-yr hourly sea level time series). MSL seasonal oscillations have been widely studied (see, e.g., Tsimplis and Woodworth 1994). This variability is due to local meteorological, oceanographic, and hydrological forcings. For the particular case of SFO, the upper panel in Fig. 3 shows a plot of estimated means plus/minus the standard deviation for MSL for the whole record. Significant vertical variations in MSL throughout a year are found (6 cm). One of the main contributions to the extreme sea level for SFO is the astronomical tide. The upper-central panel in Fig. 3 shows a complete year of hourly tide series. As seen, the highest spring tides are located in December–January and June. Again, a modulation is detected in the highest monthly tide (given by the highest spring tides), as pointed out in the lower-central panel of Fig. 3. Finally, the surge corresponding to the monthly maximum sea level is plotted in the lower panel, showing the highest elevations in winter.

After this preliminary analysis, it seems reasonable to allow for seasonality in the model because most of the processes involved have two maxima and two minima during a year. To support further this prelimi-

TABLE 1. Parameterization of the physical processes.

Physical process	Parameter		
	Location	Scale	Shape
Seasonality	$\mu_S(t) = \beta_0 + \sum_{i=1}^{P_\mu} [\beta_{2i-1} \cos(2i\pi t) + \beta_{2i} \sin(2i\pi t)]$ ( $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \dots$ )	$\psi_S(t) = \alpha_0 + \sum_{i=1}^{P_\psi} [\alpha_{2i-1} \cos(2i\pi t) + \alpha_{2i} \sin(2i\pi t)]$ ( $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \dots$ )	$\xi_S(t) = \gamma_0 + \sum_{i=1}^{P_\xi} [\gamma_{2i-1} \cos(2i\pi t) + \gamma_{2i} \sin(2i\pi t)]$ ( $\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \dots$ )
Relative sea level rise	$\mu_{SLR}(t) = \beta_{SLR} t$ ( $\beta_{SLR}$ )	—	—
Astronomical forcing	$\mu_N(t) = \beta_{N1} \cos(2\pi t/T_N) + \beta_{N2} \sin(2\pi t/T_N)$ ( $\beta_{N1}, \beta_{N2}$ )	$\psi_N(t) = \alpha_{N1} \cos(2\pi t/T_N) + \alpha_{N2} \sin(2\pi t/T_N)$ ( $\alpha_{N1}, \alpha_{N2}$ )	—
Climate variability	$\mu_{CLI}(t) = \beta_{SOI} SOI(t)$ ( $\beta_{SOI}$ )	$\psi_{CLI}(t) = \alpha_{SOI} SOI(t)$ ( $\alpha_{SOI}$ )	—
Secular trends	$\mu_{LT}(t) = \beta_{LT} t + \beta_{LT^2} t^2$ ( $\beta_{LT}, \beta_{LT^2}$ )	$\psi_{LT}(t) = \alpha_{LT} t$ ( $\alpha_{LT}$ )	$\xi_{LT}(t) = \gamma_{LT} t$ ( $\gamma_{LT}$ )

nary evidence, we examine graphically the variability within a year of the location, scale, and shape parameters of the GEV distribution of annual maxima for a given month. The maximum likelihood estimates (MLEs),  $\hat{\mu}$ ,  $\hat{\psi}$ , and  $\hat{\xi}$ , of these parameters versus month (from January to December) are shown in Fig. 4. The location parameter shows two peaks, possibly due to the highest spring tides in December–January and June. The scale parameter is modulated throughout a year with a maximum value in the winter season. Finally, the shape parameter is also modulated throughout a year with minimum values in summer. Figure 4 also shows regression fits to the first two harmonics. The good quality of the fits may suggest that the number of parameters needed in the model can be reduced. Thus, we propose using just 5 location parameters (a constant value and two harmonic terms) instead of the 12 monthly location parameters [used, e.g., by Smith (1989) in an application to ground-level ozone].

In general, a plausible model for the time parameters

can be defined in terms of harmonic functions (the first one for the annual cycle, the second one for the 6-month cycle, and so on). Mathematically, this model can be expressed as (see Table 1)

$$\mu_S(t) = \beta_0 + \sum_{i=1}^{P_\mu} [\beta_{2i-1} \cos(2i\pi t) + \beta_{2i} \sin(2i\pi t)], \tag{15}$$

$$\psi_S(t) = \alpha_0 + \sum_{i=1}^{P_\psi} [\alpha_{2i-1} \cos(2i\pi t) + \alpha_{2i} \sin(2i\pi t)], \tag{16}$$

provided that  $\psi_S(t) > 0$ , and

$$\xi_S(t) = \gamma_0 + \sum_{i=1}^{P_\xi} [\gamma_{2i-1} \cos(2i\pi t) + \gamma_{2i} \sin(2i\pi t)], \tag{17}$$

where  $\beta_0$ ,  $\alpha_0$ , and  $\gamma_0$  are mean values;  $\beta_i$ ,  $\alpha_i$ , and  $\gamma_i$  ( $i > 0$ ) are the amplitudes of the harmonics;  $P_\mu$ ,  $P_\psi$ , and  $P_\xi$  are the number of sinusoidal harmonics in a year; and  $t$  is given in years. In the literature it is very common to find seasonality introduced in the location parameter, but not on the scale and shape parameters. However, it must be noted that not only the location parameter but also the scale and the shape parameters may require harmonic terms. This approach was addressed by Coles and Walshaw (1994) for the modeling of directional extreme wind speeds. The authors analyzed the directional variability of the GEV parameters, obtaining for a particular case three harmonics for the location parameter and two harmonics for the scale and shape parameters, all of them statistically significant.

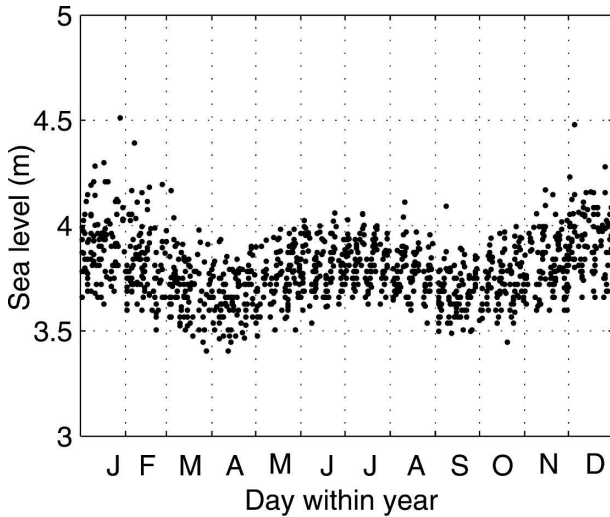


FIG. 2. Monthly maxima sea level against a day within a year.

b. Relative sea level rise

Long-term changes in regional eustatic sea levels (sea level rise) and in vertical land movements together re-

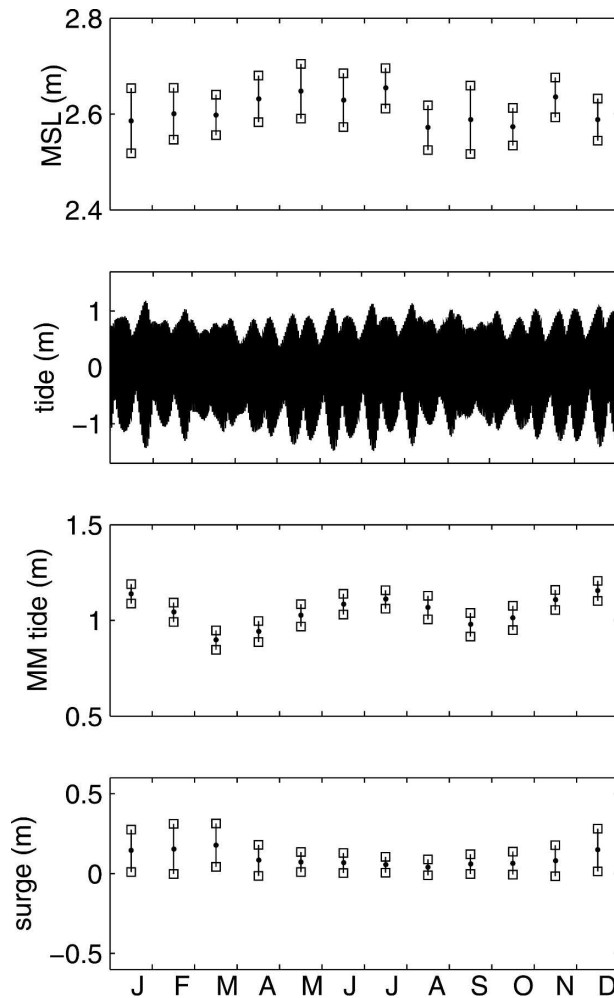


FIG. 3. Estimated means plus/minus standard deviations for (top) monthly mean sea level, (second from top) 1-yr tide series (1952), (third from top) the highest monthly astronomical tide over a 19-yr period, and (bottom) the surge corresponding to the monthly maxima (MM) sea level. Units of MSL, tide, and surge are in meters.

sult in the relative sea level rise of MSL. Tidal records at SFO station show a relative rise in MSL of  $21.7 \text{ cm century}^{-1}$  since 1900 (Flick et al. 2003). In our model, relative sea level rise is assumed to be known and is introduced linearly in the location parameter as  $\mu_{\text{SLR}}(t) = \beta_{\text{SLR}}t$  ( $\beta_{\text{SLR}} = 0.00217 \text{ m yr}^{-1}$ ). Therefore, in the proposed model, the true value of  $\beta_{\text{SLR}}$  is known. This can be done, since the record is well documented and the uncertainty in the estimation of the secular trend is very small. Besides, after visual inspection of the monthly MSL in Fig. 1, a linear trend is considered to be acceptable. Other possibilities for the definition of  $\mu_{\text{SLR}}(t)$  (e.g., quadratic, truncated, cubic splines, etc.) could also be analyzed. However, for the sake of sim-

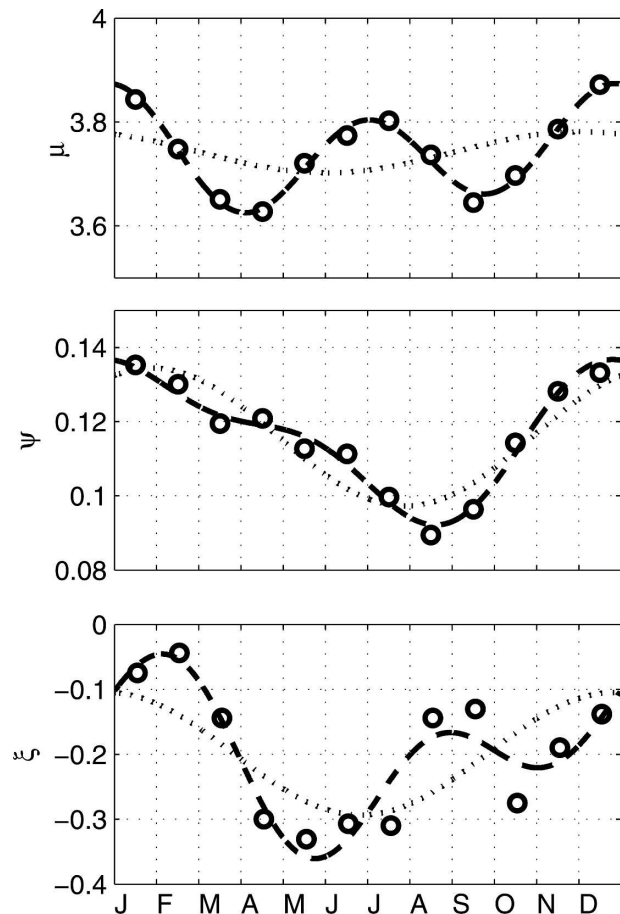


FIG. 4. Scatterplots of annual stationary GEV parameter estimates (for a given month) throughout a year. Regression fit to one (gray line) and two harmonics (black line) is also plotted. Units for location and scale parameters are in meters.

plicity and trying not to mask the interdecadal variability we adopt the linear trend.

### c. Long-term astronomical forcing variability

There are several long-term astronomical forcings such as the 18.61-yr nodal cycle due to the regression of the lunar nodes, the 8.85-yr cycle of the moon's perigee, or the 4.4-yr period between passages of the longitude of perigee past the equinoxes (Cartwright 1974). Their effects result in a slight modulation of mean sea level and a nonnegligible modulation of the semidiurnal component  $M_2$ . In the literature, these cycles have been detected in the series analyzed in the present work. For instance, Zetler and Flick (1985) detected for SFO a modulation of 12-cm amplitude for the prediction of the maximum annual high tide in the period 1983–2000. This modulation is ascribed mainly to the 4.4- and 18.61-yr cycles. Flick et al. (2003) analyzed the yearly



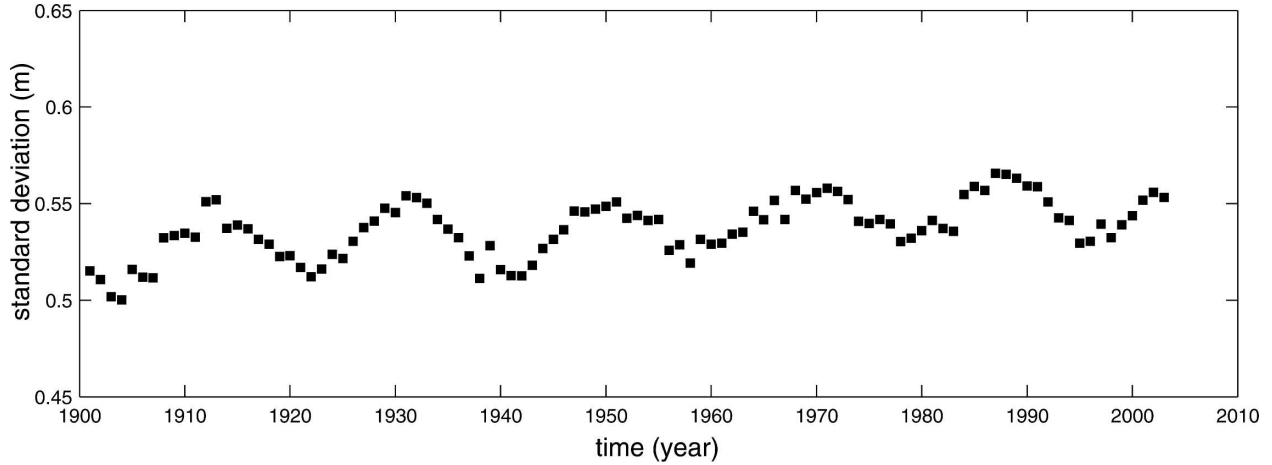


FIG. 5. Annual standard deviation of hourly sea level for SFO.

variability of several tidal datum statistics in numerous U.S. long-term tidal stations obtaining an important modulation in the 18.61-yr cycle. Recently, the 18.61-yr nodal cycle was included by Sobey (2005) as a factor in the definition of extreme high water levels for SFO. The modulation of the annual standard deviation of hourly sea level for SFO is stressed in Fig. 5, suggesting that this factor can affect the magnitude and the dispersion of extreme sea levels.

After this brief analysis of the long-term astronomical factors, we assume that extreme sea levels can be affected in the location and in the scale parameters. Therefore, a tentative relationship for the location and scale parameters can be established as (see Table 1)

$$\mu_N(t) = \beta_{N_1} \cos(2\pi t/T_N) + \beta_{N_2} \sin(2\pi t/T_N), \quad (18)$$

$$\psi_N(t) = \alpha_{N_1} \cos(2\pi t/T_N) + \alpha_{N_2} \sin(2\pi t/T_N), \quad (19)$$

where  $T_N = 18.61$  yr, and  $\beta_{N_1}$ ,  $\beta_{N_2}$  and  $\alpha_{N_1}$ ,  $\alpha_{N_2}$  are the components of the harmonic terms for the location and scale parameters, respectively. Other astronomical factors such as the 4.4-yr cycle can also be considered similarly.

#### d. Climate variability

A critical aspect of understanding the variability in extreme sea levels is the knowledge of the cause and effect of interannual and decadal fluctuations in extreme high water level records. These fluctuations are usually modeled by regional climate indices, based on air pressure data or sea surface temperatures. In this work, we use indices based on air pressure data, such as the SOI.

The variability in the Pacific El Niño–Southern Oscillation (ENSO) is usually characterized by the so-called SOI, which is based on the standard deviation of

the normalized pressure difference between Darwin (Australia) and Tahiti (French Polynesia), varying usually between  $-4$  and  $+4$ , with negative values during El Niño events and positive values during La Niña events. ENSO oscillation is known to affect ocean–atmosphere climate variability not only in the Pacific, but around remote areas of the earth (Kiladis and Diaz 1989; McPhaden 2004; Bell et al. 2000). Recently, a relationship between the occurrence of extreme sea levels and indices of regional climate has been found along the globe for numerous tidal gauges (Woodworth and Blackman 2004).

Therefore, it would be highly desirable to explain the interannual variability of MSL and sea level extremes for SFO station by means of indices such as the SOI. Flick (1998) analyzed the effects on the sea level of historic El Niño events on the Californian coast, and showed that both the mean sea level and the extreme sea level are highly correlated with SOI. In the model (see Table 1) we introduce the location,  $\mu_{\text{CLI}}(t) = \beta_{\text{SOI}} \text{SOI}(t)$ , and the scale parameter,  $\psi_{\text{CLI}}(t) = \alpha_{\text{SOI}} \text{SOI}(t)$ , as global climate covariates, which can be thought of as location and dispersion measures of the distribution of maxima, respectively.  $\text{SOI}(t)$  is the time-varying function of the monthly Southern Oscillation index,  $\beta_{\text{SOI}}$  is the linear coefficient for the location parameter, and  $\alpha_{\text{SOI}}$  stands for the linear coefficient for the scale parameter. As pointed out by Flick (1998), the monthly mean sea level increased during the 1982/83 winter by about 12 cm. This 12 cm is the difference between the 1982/83 and the 1960–78 winter monthly sea levels. This effect is modeled by the location parameter  $\mu_{\text{CLI}}(t)$ . Moreover, during an El Niño event, the storm surge is assumed to increase the variability of the extreme sea levels. Consequently, the scale parameter,  $\psi_{\text{CLI}}(t)$ , attempts to model this climate variability. Note that other

possible regional climate indices could also be considered, such as the Pacific–North America (PNA) pattern (Wallace and Gutzler 1981) or the PDO index (Mantua et al. 1997).

#### e. Secular trends

Secular trends on extreme sea level are associated to changes in surge and tidal components. There is some evidence regarding long-term changes in the intensity and frequency of the surge (meteorologically forced storminess) in different parts around the globe (e.g., Bromirski et al. 2003; Langenberg et al. 1999). With respect to the secular changes in the tide, a number of studies have also been carried out trying to determine whether the tide is changing on a long time scale (Keelig and Whorf 2000), but the time scale (a 1800-yr tidal cycle) is out of the scope of this work. Thus, major changes in the tide are expected to be found due to the tide propagation, which is affected by the geometry. Therefore, natural processes (relative sea level rise and long-term morphology changes) and the anthropogenic permanent impacts (e.g., the construction of dykes; land reclamation in wetland areas) are the main factors that govern long-term variations in the tide (Pugh 2004).

An analysis of the secular trends in the mean tidal range in several tidal gauges of the British Isles, France, Netherlands, and Belgium revealed trends between

–1.8 and 1.3 mm yr<sup>-1</sup> depending on location (Woodworth et al. 1991). For instance, the authors found a negative trend for Brest, France, but a positive one for Newlyn, United Kingdom, confirming that this effect is fully locally dependent.

A number of authors have analyzed the joint effect of tide and surge. For instance, Flick et al. (2003) analyzed the long-term variation of the annual mean higher high water (MHHW; the datum MHHW is defined as the average of only the higher of the daily high waters), and they found statistically significant trends in numerous tidal records in United States. For example, at San Francisco, the authors obtained an increase of 52 mm century<sup>-1</sup> of the MHHW relative to the MSL from 1900 to 1998.

After this analysis, small secular variations are expected. Therefore, we parameterize the location, scale, and shape parameters as  $\mu_{LT}(t) = \beta_{LT}t$ ,  $\psi_{LT}(t) = \alpha_{LT}t$ , and  $\xi_{LT}(t) = \gamma_{LT}t$ . A parabolic parameterization has also been tested for the location parameter,  $\mu_{LT}(t) = \beta_{LT}t + \beta_{LT_2}t^2$  (see Table 1).

#### f. Example of aggregation of the factors

As an example and following the aggregation explained at the beginning of section 4, a possible model to be analyzed can be expressed as

$$\begin{aligned} \mu(t) &= \beta_{SLR}t + [\beta_0 + \beta_1 \cos(2\pi t) + \beta_2 \sin(2\pi t) + \beta_3 \cos(4\pi t) + \beta_4 \sin(4\pi t)]e^{\beta_{LT}t} + \beta_{N_1} \cos(2\pi t/T_N) \\ &\quad + \beta_{N_2} \sin(2\pi t/T_N) + \beta_{SOI} \text{SOI}(t) \\ \psi(t) &= [\alpha_0 + \alpha_1 \cos(2\pi t) + \alpha_2 \sin(2\pi t)]e^{\alpha_{LT}t} \\ \xi(t) &= [\gamma_0 + \gamma_1 \cos(2\pi t) + \gamma_2 \sin(2\pi t)]. \end{aligned} \quad (20)$$

For this particular model  $\beta_{SLR}$  is assumed to be known and the vector parameter to be estimated is

$$\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_{N_1}, \beta_{N_2}, \beta_{LT}, \beta_{SOI}, \alpha_0, \alpha_1, \alpha_2, \alpha_{LT}, \gamma_0, \gamma_1, \gamma_2), \quad (21)$$

where  $\beta_0, \beta_1, \beta_2, \beta_3,$  and  $\beta_4$  model the seasonal behavior of the location parameter;  $\beta_{N_1}$  and  $\beta_{N_2}$  are the amplitudes of the nodal cycle;  $\beta_{LT}$  is the long-term trend of the location parameter;  $\beta_{SOI}$  explains the influence of the climate index in the location parameter;  $\alpha_0, \alpha_1,$  and  $\alpha_2$  model the seasonality of the scale parameter;  $\alpha_{LT}$  represents the long-term trend of the variability of the extreme sea level; and  $\gamma_0, \gamma_1,$  and  $\gamma_2$  model the seasonality in the shape parameter. This model corresponds to the so-called SFO-MM12 model to be analyzed in section 5.

For a given model and a given monthly maxima time series, the maximization of the log-likelihood function

[Eq. (4)] can be done as indicated in section 2. This yields the estimates  $\hat{\theta}$ , their standard errors,  $\text{se}(\hat{\theta})$ , and the maximum likelihood,  $\hat{l}(p)$ . Model performance can be assessed by means of some information criteria, depending on  $\hat{l}(p)$ , the number of parameters  $p$ , and the number of data  $n$  (as explained in section 2d).

## 5. Application of the model

### a. Selection of independent monthly maxima series

One important aspect of the proposed model is the use of monthly maxima series, which must fulfill independency between consecutive months. This aspect is

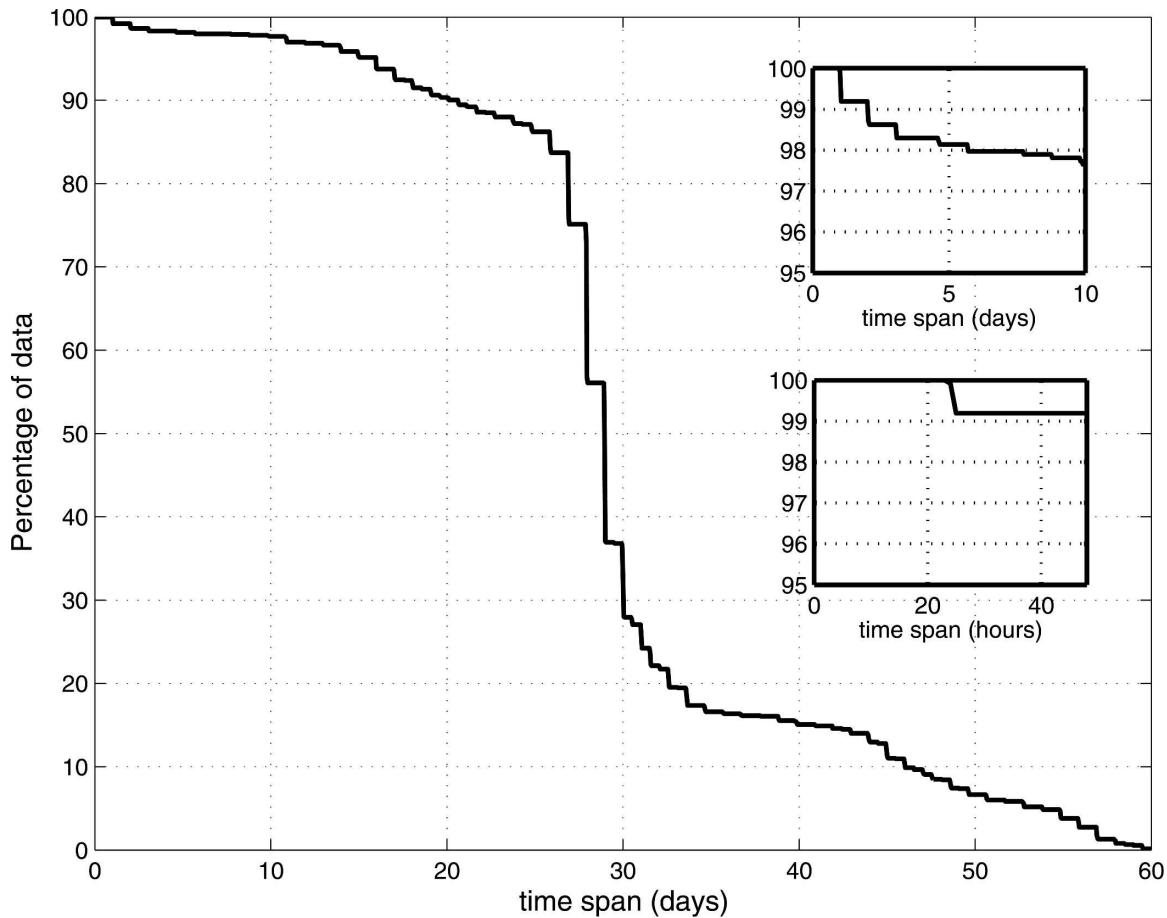


FIG. 6. Percentage of independent monthly extreme data as a function of the time span for San Francisco sea level time series.

usually approached defining a storm length or time span ( $\tau$ ), so that consecutive monthly maxima must be separated at least by a given duration  $\tau$ . A preliminary analysis has detected, from the original extraction of the monthly maxima series  $\{Z_{t_1}, \dots, Z_{t_m}\}$ , the percentage of data that are independent for different time spans (see Fig. 6).

An algorithm to obtain independent monthly maxima series for a given time span  $\tau$ ,  $\{\bar{Z}_{t_1}, \dots, \bar{Z}_{t_m}\}$ , has been implemented. We run the series from  $i = 1, \dots, m - 1$  and, for every pair  $\{Z_{t_i}, Z_{t_{i+1}}\}$  of consecutive monthly maxima in months  $(i, i+1)$ , the duration  $\Delta t_i = t_{i+1} - t_i$  is calculated. If  $\Delta t_i > \tau$ , then  $\bar{Z}_{t_i} = Z_{t_i}$ . If not, the minimum of the consecutive values is disregarded. If, for instance, the maximum  $\max\{Z_{t_i}, Z_{t_{i+1}}\}$  corresponds to month  $i + 1$ , then  $\bar{Z}_{t_{i+1}} = Z_{t_{i+1}}$  and, from the original hourly series in month  $i$ , another monthly maximum  $\bar{Z}_{t_i}$  must be selected in month  $i$ , conditioned to  $t_i \leq t_{i+1} - \tau$ . The same procedure is applied if the maximum corresponds to month  $i$ . In this case,  $\bar{Z}_{t_i} = Z_{t_i}$  and the new  $\bar{Z}_{t_{i+1}}$  must fulfill  $t_{i+1} \geq t_i + \tau$ . In some particular cases, the new time series  $\{\bar{Z}_{t_1}, \dots, \bar{Z}_{t_m}\}$  could still have some

time interval  $\Delta t_i < \tau$ , but this can be overcome in an iterative manner.

There is no general agreement to define the time span, although a number of authors have used durations between 1 and 3 days for the application of the  $r$ -largest annual method (e.g., Tawn 1988; Dixon and Tawn 1994; Tsimplis and Blackman 1997). In this work, we have chosen  $\tau = 48$  h. For this time span the monthly maxima values that must be corrected following the algorithm are about 1%. It is clear that dependency between consecutive months is not a major concern, since just 1% of the original data must be changed so that the results are only very mildly affected by this hypothesis. This conclusion is very useful when long-term historical time series are analyzed, in which only the maximum daily observation or just the monthly maximum value is available (Bijl et al. 1999).

#### b. Analysis of the results

A series of 1233 independent monthly maxima  $\{\bar{Z}_{t_1}, \dots, \bar{Z}_{t_m}\}$  in the period 1901–2003 (three monthly

gaps) is the sample available in San Francisco. It is assumed that time  $t$  refers to the initial year 1901.

Numerous possible models have been tested in order to get the best model that minimizes the BIC criterion. As has been previously stated, this criterion follows the principle of parsimony, establishing a compromise between obtaining a good fit (measured by the maximized log-likelihood function) and penalizing most heavily the number of parameters used. Table 2 shows a summary of the results for SFO for different nested models considered, showing the set of parameters, the MLEs, the standard error (SE) for every parameter in parentheses, the maximum attained by the log-likelihood function for every model ( $l$ ), the number of parameters ( $p$ ), and the BIC and AIC model selection criteria. The reference model with constant GEV parameters is SFO-MM0. As for annual maxima, SFO-MM1 includes sea level rise as a known value [ $\mu_{\text{SLR}}(t) = \beta_{\text{SLR}}t$ ]. The best in terms of minimizing the BIC is model SFO-MM12 [given in Eq. (20)]. The number of parameters of this model is  $p = 16$ , all of them significant at the 0.05 level. The order of inclusion of parameters in Table 2 is the following: astronomical forcing, seasonality in the location, scale and shape parameters, long-term trend in the location and shape parameters, and SOI influence (the exception is model SFO-MM3, where  $\beta_{\text{SOI}}$  is previously included for comparison purposes). Although more intermediate models have been considered by changing the order in which the parameters enter the model, the final model remains unchanged. The intermediate models allow us to assess the importance of the inclusion of the different processes involved.

Figure 7 shows the fit of the models selected by means of the probability and quantile plots. The upper panels show the results for model SFO-MM1, whereas the lower panels show plots for model SFO-MM12. One can see that the residuals in model SFO-MM12 are closer to the straight line, improving the fit with respect to SFO-MM1.

Subsequently, the processes included in model SFO-MM12 are analyzed.

### 1) SEA LEVEL RISE

The sea level rise ( $\beta_{\text{SLR}} = 0.00217 \text{ m yr}^{-1}$ ), is introduced as a fixed value since the San Francisco tidal gauge is highly well documented.

### 2) SEASONALITY

Two harmonics for the location parameter ( $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ ) and one harmonic for both the scale ( $\alpha_0, \alpha_1, \alpha_2$ ) and shape ( $\gamma_0, \gamma_1, \gamma_2$ ) parameters are obtained. The

preliminary analysis made in section 4 revealed this important aspect that is essential when more than one datum per year is considered in an extreme value model. The improvement in the model is conclusive when seasonality is modeled. For instance, the inclusion of the two harmonics in the location parameter (from model SFO-MM2 to SFO-MM4) gives an enhancement in the maximized log-likelihood function of more than 310 units. When considering one harmonic in the scale and shape parameter (from model SFO-MM4 to SFO-MM5 and from SFO-MM5 to SFO-MM7), the improvement is of 69 and 4 units, respectively, which are significant at least at the 0.02 level. The second harmonic in the scale and shape parameters are significant at the 0.11 and 0.17 levels, respectively. Note how, although the results from Fig. 4 for the monthly marginal GEV distribution fits suggested that the second harmonic was adequate for the scale and shape parameters, the complete combination of model SFO-MM12 for the 1233 monthly data is more restrictive. The results obtained for the shape parameter (an annual cycle with nonpositive values) indicate that the tail of the distribution belongs to the Weibull for maxima family and is variable along the year. This is very important for the estimation of large return period quantiles.

### 3) LONG-TERM ASTRONOMICAL FORCING

The nodal cycle in the location parameter is introduced in the model. The amplitude,  $(\hat{\beta}_{N_1}^2 + \hat{\beta}_{N_2}^2)^{1/2} = 0.025 \text{ m}$ , is similar to that obtained by Sobey (2005), fitting the SFO monthly extreme data to a linear trend plus an 18.61-yr sinusoidal component ( $a_5 = 0.0224 \text{ m}$ ). The contribution of the nodal cycle in the scale parameter ( $\alpha_{N_1}, \alpha_{N_2}$ ) is not significant.

### 4) DECADAL VARIABILITY

There is evidence that the effect of SOI is influential on the model in the location parameter. The estimated value of  $\beta_{\text{SOI}}$  in model SFO-MM12 is  $-0.0237$ , with a standard error of 0.003, so that every unit decrease in SOI results in an estimated increase of around 2.4 cm in monthly maximum sea level. Therefore, the 1983 El Niño (SOI  $\approx -3$ ) explained about 7.2 cm. This result is in consonance with the increase detected by Flick (1998) of about 12 cm in the monthly mean sea level in the 1982/83 winter with respect to the 1960–78 period. Another model considering the scale parameter ( $\alpha_{\text{SOI}}$ ) has been tested but it is only significant at the 0.44 level.

### 5) SECULAR TRENDS

The long-term location ( $\beta_{\text{LT}}$ ) and scale ( $\alpha_{\text{LT}}$ ) parameters are included in the best model. The estimated

TABLE 2. Summary of the results for different models for SFO-MM: maximum likelihood estimates, standard errors, maximum likelihood function, number of parameters involved, AIC, and BIC.

MLE (SE)	SFO-MM0	SFO-MM1	SFO-MM2	SFO-MM3	SFO-MM4	SFO-MM5	SFO-MM7	SFO-MM9	SFO-MM11	SFO-MM12
$\beta_0$ (m)	3.728 (0.004)	3.6248 (0.004)	3.626 (0.004)	3.6239 (0.004)	3.6326 (0.003)	3.6353 (0.003)	3.6354 (0.003)	3.6154 (0.005)	3.6146 (0.005)	3.6189 (0.005)
$\beta_1$ (m)	—	—	—	—	0.0214 (0.004)	0.0296 (0.004)	0.0269 (0.004)	0.0265 (0.004)	0.0261 (0.004)	0.0271 (0.004)
$\beta_2$ (m)	—	—	—	—	0.0302 (0.004)	0.0286 (0.004)	0.0262 (0.004)	0.027 (0.004)	0.0276 (0.004)	0.0283 (0.004)
$\beta_3$ (m)	—	—	—	—	0.0823 (0.004)	0.0856 (0.004)	0.0858 (0.004)	0.0842 (0.004)	0.0855 (0.004)	0.0854 (0.003)
$\beta_4$ (m)	—	—	—	—	0.0497 (0.004)	0.048 (0.003)	0.048 (0.003)	0.0489 (0.003)	0.0475 (0.003)	0.0488 (0.003)
$\beta_{L,T}$ (yr <sup>-1</sup> )	—	—	—	—	—	—	—	0.0009 (0.00002)	0.0009 (0.00002)	0.0009 (0.00002)
$\beta_{N_1}$ (m)	—	—	0.0064 (0.005)	—	0.0087 (0.004)	0.0096 (0.003)	0.0095 (0.003)	0.0103 (0.003)	0.0101 (0.003)	0.0071 (0.003)
$\beta_{N_2}$ (m)	—	—	0.0199 (0.005)	—	0.0201 (0.004)	0.0203 (0.003)	0.021 (0.003)	0.0214 (0.003)	0.0198 (0.003)	0.0236 (0.003)
$\beta_{SOI}$ (m)	—	—	—	0.0304 (0.004)	—	—	—	—	—	0.0237 (0.003)
unit <sup>-1</sup> )										
$\alpha_0$ (m)	0.1364 (0.003)	0.1144 (0.003)	0.1133 (0.002)	0.1121 (0.003)	0.0865 (0.002)	0.0858 (0.002)	0.0857 (0.002)	0.085 (0.002)	0.0776 (0.003)	0.0735 (0.003)
$\alpha_1$ (m)	—	—	—	—	—	0.0274 (0.002)	0.0261 (0.003)	0.0253 (0.002)	0.023 (0.002)	0.0186 (0.002)
$\alpha_2$ (m)	—	—	—	—	—	0.0072 (0.002)	0.0084 (0.003)	0.008 (0.003)	0.0083 (0.002)	0.0072 (0.003)
$\alpha_{L,T}$ (yr <sup>-1</sup> )	—	—	—	—	—	—	—	—	0.0018 (0.00076)	0.002 (0.00075)
$\gamma_0$	0.11319 (0.013)	0.1112 (0.012)	0.1077 (0.014)	0.1193 (0.012)	0.0743 (0.015)	0.1177 (0.015)	0.1307 (0.016)	0.1339 (0.016)	0.1294 (0.017)	0.1328 (0.017)
$\gamma_1$	—	—	—	—	—	—	0.0461 (0.02)	0.0613 (0.02)	0.0686 (0.022)	0.0768 (0.021)
$\gamma_2$	—	—	—	—	—	—	0.0519 (0.026)	0.0441 (0.025)	0.0421 (0.026)	0.0325 (0.025)
$\ell$	315.71	818.61	828.17	848.4	1138.51	1207.67	1211.63	1224.14	1227.66	1267.21
p	3	3	5	4	9	11	13	14	15	16
AIC	625.42	1631.22	1646.34	1688.80	2259.02	2393.34	2397.26	2420.28	2425.32	2502.42
BIC	610.06	1615.87	1620.74	1668.32	2212.97	2337.05	2330.74	2348.65	2348.55	2420.53

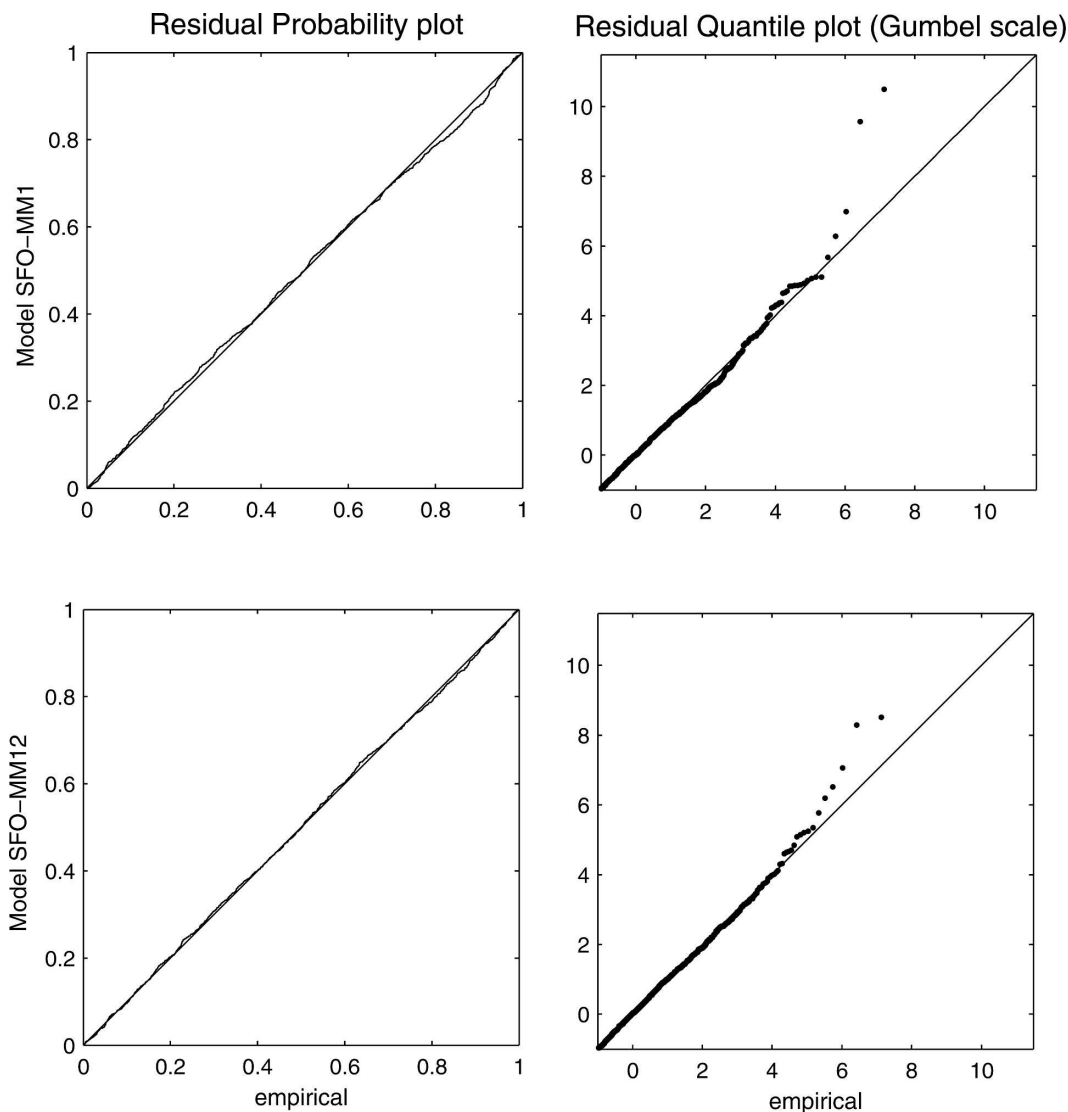


FIG. 7. Quantile and probability plots for models (top) SFO-MM1 and (bottom) SFO-MM12.

value of the location parameter is  $\hat{\beta}_{LT} = 9 \times 10^{-5} \text{ yr}^{-1}$ , with a standard error of  $2 \times 10^{-5}$ . Thus, a yearly average value of the increase of the mode of the distribution can be calculated by  $\hat{\beta}_0 \hat{\beta}_{LT} = 0.00033 \text{ m yr}^{-1}$ , which means an increase in the last 103 yr of 3.4 cm. Similar results were obtained by Flick et al. (2003) analyzing the change of the yearly average MHHW relative to the MSL. The authors obtained for SFO an increase of  $0.00042 \text{ m yr}^{-1}$ . Although the type of analysis and the variable of interest is not the same, it must be noted that the order of magnitude of this result for model SFO-MM12 is in good agreement with respect to Flick et al.'s (2003) work. Moreover, the most impressive result is obtained in the scale parameter. The estimated value is  $\hat{\alpha}_{LT} = 0.002 \text{ yr}^{-1}$ , with a 95% confidence in-

terval of (0.0005, 0.0035). This means an increase of almost 23% [ $\exp(0.002 \times 10^3) = 1.229$ ] in 103 yr in the estimation of the scale parameter, that is, in the storminess or variability of the monthly extreme sea levels. Other models including a parabolic trend in the location parameter ( $p$  value of 0.32) or a linear long-term variation in the shape parameter have been tested but were not significant.

#### 6) CONTRIBUTION OF THE DIFFERENT TIME SCALES

The contribution of every factor in the location and scale parameters in model SFO-MM12 is represented in Fig. 8. The scale parameters ( $\psi$ ,  $\psi_S$ , and  $\psi_{LT}$ ) are vertically distorted to easily visualize its component (note

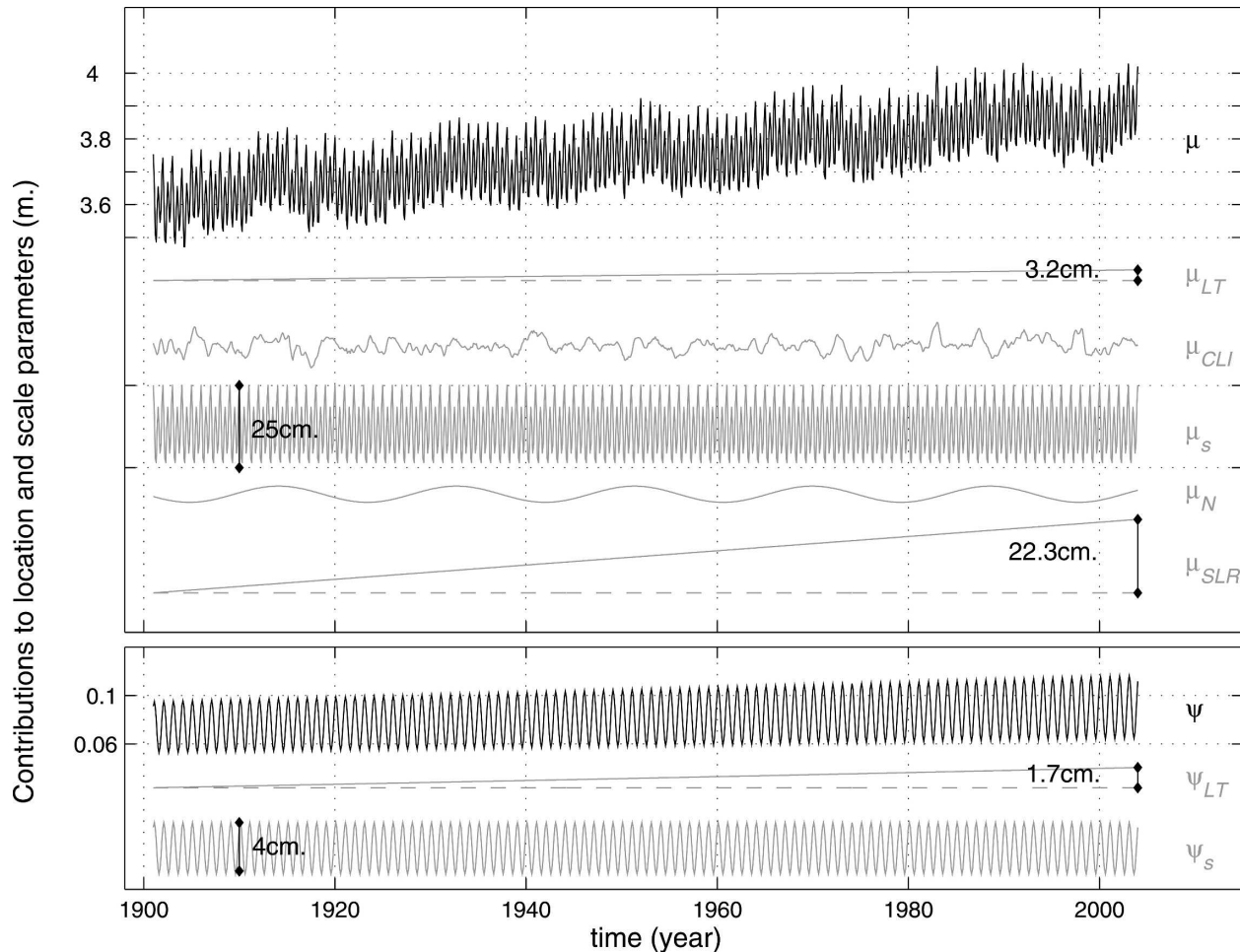


FIG. 8. Contribution of the physical processes in model SFO-MM12 in the location and scale parameters.

the different scales of the location and scale parameters). Secular trends in the location and scale parameters are multiplied by estimated parameters  $\hat{\beta}_0$  and  $\hat{\alpha}_0$ , respectively. One can see that the model is able to take into account the different time scales (sea level rise, seasonality, climate variability, nodal cycle, and secular trend). Besides, it is important to point out that the order of magnitude of every factor is not at all negligible.

#### 7) DETERMINATION OF QUANTILES

The evolution of the location parameter as well as the instantaneous 50-yr return period quantile (bold lines) for years 1900, 1950, and 2000 are shown in Fig. 9. Black lines correspond to 1900 (lower) and 2000 (upper). Results from 1950 are represented by gray lines. The vertical difference is due to the long-term components (sea level rise, secular trend, nodal cycle, and climate variability). It is remarkable how the location

parameter is able to reproduce the two maxima (in December–January and in June) during the year. Besides, note that the behavior of the 50-yr quantile is similar but enhances its magnitude in December and January (due to the seasonal variability of the shape and scale parameters). A 95% upper confidence interval for the 50-yr return period in the year 2000, based on the delta method, is also plotted in the dotted–dashed line following Eqs. (10) and (11). More specifically, these confidence limits have been computed using the formula  $\hat{z}_q(\theta) + 1.96\text{se}[\hat{z}_q(\theta)]$ . We consider the possibility that symmetrical confidence intervals like these may give a misleading representation of the uncertainty involved (Coles 2001). In this case, it is likely that  $\log[\hat{z}_q(\theta)]$  will be better approximated by the normal distribution than  $\hat{z}_q(\theta)$ . We investigate the robustness of our approach by assuming for the moment that the normal approximation applies to  $\log[\hat{z}_q(\theta)]$ , in which case the delta method would justify using the formula  $\hat{z}_q(\theta) \exp\{+1.96\text{se}[\hat{z}_q(\theta)]/\hat{z}_q(\theta)\}$ . For the values of  $\text{se}[\hat{z}_q(\theta)]/\hat{z}_q(\theta)$

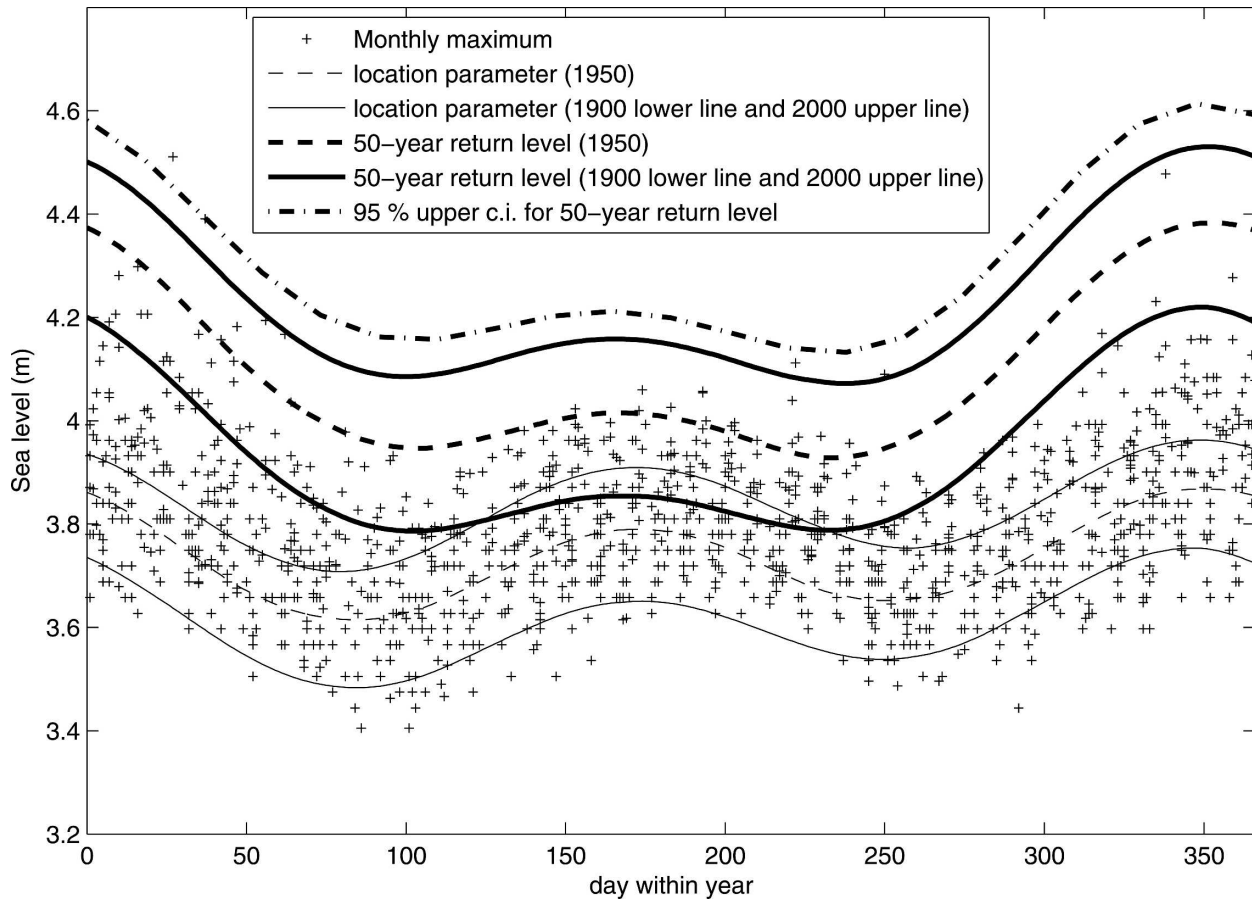


FIG. 9. Location parameter and 50-yr return period quantiles (bold lines) within a year for years 1900 (lower line), 1950 (gray), and 2000 (upper line). The 95% upper confidence interval for the 50-yr return period for the year 2000 is plotted by the gray dashed line. Observed values of monthly maxima sea level are indicated by crosses. Results are for model SFO-MM12.

corresponding to the data used in Fig. 9, the difference between these two formulas is negligible, showing our approach to be robust. Alternatively, one could use the profile likelihood method (Coles 2001), which would require a much heavier computational effort.

## 6. Conclusions

A statistical model to analyze different time scales of the variability of extreme high sea levels is presented. The model is based on the time-dependent generalized extreme value distribution for independent monthly maxima series. Nonstationarity is introduced in the model in terms of different time scales—such as seasonality, interdecadal climate variability and astronomical modulations, sea level rise, and secular trends—that are parameterized as functions of time (linear, quadratic, exponential, and cosine functions) or covariates (e.g., SOI). These processes are included in the location, scale, and shape parameters of the probability distribution of extreme sea levels.

The model is applied to a well-known large tidal hourly time series (San Francisco) showing the characteristics of the seasonal-to-interannual time scales at this particular site. Conclusions for this particular site are as follows: 1) the inclusion of seasonal effects in the location (two harmonics), scale (one), and shape (one) parameters notably improves the understanding of data variability increasing the significance of the parameters considered (e.g., see Table 2); 2) significant correlation with the SOI contribution to the location parameter is detected (each SOI unit explains 2.4 cm); 3) the nodal cycle (in the location parameter) contributes to a better explanation of data variability (3-cm amplitude); 4) a slight positive secular trend in the location parameter is obtained (4 cm in the twentieth century); and 5) an increase of about 20% for the secular variability of the scale parameter has been detected. That means that monthly sea level extremes in SFO are increasing not only in the mean values but also in the variability (storminess).



Results exhibit that the model is valid for a complete analysis of seasonal-to-interannual sea level extremes, providing time-dependent quantiles and confidence intervals. The modeling of the different time scales helps give a better understanding of recent secular trends for the extreme climate events, which are one of the main concerns nowadays (Easterling et al. 2000). Finally, we believe that the proposed model provides very useful insights to building models for other geophysical extreme variables such as extreme rainfalls, temperature, wind speeds, or wave climate, taking into account the peculiarities (location and physical processes) of each particular case.

*Acknowledgments.* The work was partially funded by Projects ENE2004-08172 and CGL2005-05365/CLI from the Spanish Ministry of Educación y Ciencia and by a project called “Wave Climate Duration along the Spanish Littoral” from the Spanish Ministry of Fomento. Fernando J. Méndez and Melisa Menéndez are indebted to the Spanish Ministry of Educación y Ciencia for funding them through the “Ramón y Cajal” program and the FPI program, respectively. Alberto Luceño acknowledges the support of the Spanish Dirección General de Investigación under Grant MTM2005-00287. We thank two anonymous reviewers for helpful comments that notably improved the manuscript.

#### REFERENCES

- Bell, R. G., D. G. Goring, and W. P. de Lange, 2000: Sea level change and storm surges in the context of climate change. *IPENZ Trans.*, **27** (General), 1–10.
- Bijl, W., R. Flather, J. G. de Ronde, and T. Schmith, 1999: Changing storminess? An analysis of long-term sea level data sets. *Climate Res.*, **11**, 161–172.
- Bromirski, P. D., R. E. Flick, and D. R. Cayan, 2003: Storminess variability along the California Coast: 1858–2000. *J. Climate*, **16**, 982–993.
- Carter, D. J. T., and P. G. Challenor, 1981: Estimating return values of environmental variables. *Quart. J. Roy. Meteor. Soc.*, **107**, 259–266.
- Cartwright, D. E., 1974: Years of peak astronomical tide. *Nature*, **248**, 656–657.
- Coles, S. G., 2001: *An Introduction to Statistical Modelling of Extreme Values*. Springer, 208 pp.
- , and D. Walshaw, 1994: Directional modelling of extreme wind speeds. *J. Roy. Stat. Soc.*, **43C**, 139–157.
- Davison, A. C., and R. L. Smith, 1990: Models for exceedances over high thresholds (with discussion). *J. Roy. Stat. Soc.*, **52B**, 393–442.
- Dixon, M. J., and J. A. Tawn, 1994: Extreme sea-levels at the UK A-class site: Site-by-site analysis. Proudman Oceanographic Laboratory Internal Doc. 65, 229 pp.
- , and —, 1999: The effect of non-stationarity on extreme sea-level estimation. *J. Roy. Stat. Soc.*, **48C**, 135–151.
- D’Onofrio, E. E., M. M. E. Fiore, and S. I. Romero, 1999: Return periods of extreme water levels estimated for some vulnerable areas of Buenos Aires. *Cont. Shelf Res.*, **19**, 1681–1693.
- Duan, Q., S. Sorooshian, and V. Gupta, 1992: Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.*, **28**, 1015–1031.
- Easterling, D. R., J. L. Evans, P. Ya. Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje, 2000: Observed variability and trends in extreme climate events: A brief review. *Bull. Amer. Meteor. Soc.*, **81**, 417–425.
- Flick, R. E., 1998: Comparison of California tides, storm surges, and mean sea level during the El Niño winters of 1982–83 and 1997–98. *Shore Beach*, **66**, 7–11.
- , J. F. Murray, and L. C. Ewing, 2003: Trends in United States tidal datum statistics and tide range. *J. Waterw. Port Coastal Ocean Eng.*, **129**, 155–164.
- Galambos, J., 1978: *The Asymptotic Theory of Extreme Order Statistics*. Wiley, 414 pp.
- Goda, Y., 2000: *Random Seas and Design of Maritime Structures*. P. L.-F. Liu, Ed., Advanced Series on Ocean Engineering, Vol. 15, World Scientific, 443 pp.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis*. Cambridge University Press, 881 pp.
- Katz, R. W., M. B. Parlange, and P. Naveau, 2002: Statistics of extremes in hydrology. *Adv. Water Resour.*, **25**, 1287–1304.
- Keelig, C. D., and T. P. Whorf, 2000: The 1,800-year oceanic tidal cycle: A possible cause of rapid climate change. *Proc. Natl. Acad. Sci. USA*, **97**, 3814–3819.
- Kiladis, G. N., and H. F. Diaz, 1989: Global climatic anomalies associated with extremes in the Southern Oscillation. *J. Climate*, **2**, 1069–1090.
- Langenberg, H., A. Pfizenmayer, H. von Storch, and J. Sündermann, 1999: Storm-related sea level variations along the North Sea coast: Natural variability and anthropogenic change. *Cont. Shelf Res.*, **19**, 821–842.
- Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1079.
- McPhaden, M. J., 2004: Evolution of the 2002/03 El Niño. *Bull. Amer. Meteor. Soc.*, **85**, 677–695.
- Meehl, G. A., and Coauthors, 2000: An introduction to trends in extreme weather and climate events: Observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. *Bull. Amer. Meteor. Soc.*, **81**, 413–416.
- Pugh, D. T., 2004: *Changing Sea Levels: Effects of Tides, Weather and Climate*. Cambridge University Press, 265 pp.
- Rice, J., 1994: *Mathematical Statistics and Data Analysis*. 2d ed. Duxbury, 651 pp.
- Smith, R. A., and R. J. Leffler, 1980: Water level variations along California coast. *J. Waterw. Port Coastal Ocean Div.*, **106**, 335–348.
- Smith, R. L., 1986: Extreme value theory based on the  $r$  largest annual events. *J. Hydrol.*, **86**, 27–43.
- , 1989: Extreme value analysis of environmental time series: An application to trend detection in ground level ozone. *Stat. Sci.*, **4**, 367–393.
- , 2001: Environmental statistics. Tech. Rep., Department of Statistics, University of North Carolina, Chapel Hill, NC, 377 pp.
- Soares, C. G., and M. G. Scotto, 2004: Application of the  $r$ -largest

- order statistics for long-term predictions of significant wave height. *Coast. Eng.*, **51**, 387–394.
- Sobey, R. J., 2005: Extreme low and high water levels. *Coast. Eng.*, **52**, 63–77.
- Tawn, J. A., 1988: An extreme value theory model for dependent observations. *J. Hydrol.*, **101**, 227–250.
- , and J. M. Vassie, 1989: Extreme sea levels: The joint probability method revisited and revised. *Proc. Inst. Civ. Eng.*, **87**, 429–442.
- Tsimplis, M. N., and P. L. Woodworth, 1994: The global distribution of the seasonal sea level cycle calculated from coastal tide gauge data. *J. Geophys. Res.*, **99**, 16 031–16 039.
- , and D. Blackman, 1997: Extreme sea-level distribution and return periods in the Aegean and Ionian Seas. *Estuarine Coastal Shelf Sci.*, **44**, 79–89.
- Unnikrishnan, A. S., D. Sundar, and D. L. Blackman, 2004: Analysis of extreme sea level along the east coast of India. *J. Geophys. Res.*, **109**, C06023, doi:10.1029/2003JC002217.
- Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812.
- Woodworth, P. L., and D. L. Blackman, 2002: Changes in extreme high waters at Liverpool since 1768. *Int. J. Climatol.*, **22**, 697–714.
- , and —, 2004: Evidence for systematic changes in extreme high waters since the mid-1970s. *J. Climate*, **17**, 1190–1197.
- , S. M. Shaw, and D. L. Blackman, 1991: Secular trends in mean tidal range around the British Isles and along the adjacent European coastline. *Geophys. J. Int.*, **104**, 593–609.
- , M. N. Tsimplis, R. A. Flather, and I. Shennan, 1999: A review of the trends observed in British Isles mean sea level data measured by tide gauges. *Geophys. J. Int.*, **136**, 651–670.
- Zetler, B. D., and R. E. Flick, 1985: Predicted extreme high tides for California: 1983–2000. *J. Waterw. Port Coastal Ocean Eng.*, **111**, 758–765.