

# SKEW-ELLIPTICAL TIME SERIES WITH APPLICATION TO FLOODING RISK

MARC G. GENTON\* AND KEITH R. THOMPSON†

**Abstract.** In this article, skew-elliptical time series are defined in order to account for both skewness and kurtosis, with particular emphasis on the skew-normal and skew- $t$  distributions. The bivariate skew- $t$  distribution is then used to describe a 63 year time series of hourly sea levels measured at Charlottetown, Atlantic Canada. It is shown that the skew- $t$  fits the data better than the normal distribution and it can be used to recover return periods of extreme levels based on a standard analysis of 63 annual maxima. Preliminary results are presented to show how the skew- $t$  distribution may be used to estimate changes in flooding risk resulting from changes in sea level rise, storminess, and other climatic factors.

**Key words.** Extremes, flooding risk, kurtosis, nonstationarity, skewness, storm surge.

**AMS(MOS) subject classifications.** Primary 62M10, 62H05, 62P12.

**1. Introduction.** This article presents time series modeling with a class of continuous multivariate distributions that can simultaneously account for both skewness and heavy tails. To date, most time series modeling has focused on symmetric distributions, with particular emphasis on the normal distribution. Nevertheless, there are many natural phenomena that do not follow the normal law including the example discussed in this paper: the nontidal changes in sea level caused by variations in air pressure and wind acting at the surface of the ocean. Non-normal distributions are needed to model such phenomena and departures from normality can be achieved by varying both the skewness and kurtosis in the distribution. For this purpose, two main approaches are available.

The first approach consists of modifying a random variable, and thus also its quantiles, through an appropriate transformation. It was suggested by John W. Tukey in 1977 and discussed by Hoaglin et al. (1985) in the univariate setting. Basically, a standard normal random variable  $Z$  is transformed to  $Y = \tau_{g,h}(Z)$ , where:

$$(1.1) \quad \tau_{g,h}(Z) = \left( \frac{\exp(gZ) - 1}{g} \right) \exp\left( \frac{h}{2} Z^2 \right).$$

The resulting random variable  $Y$  is said to have a  $g$ -and- $h$  distribution, where  $g$  is a real constant controlling the skewness and  $h$  is a nonnegative real constant controlling the kurtosis, or elongation. The quantiles of

---

\*Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA (genton@stat.ncsu.edu).

†Department of Mathematics and Statistics, and Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1 (keith@phys.ocean.dal.ca).

$Y = \tau_{g,h}(Z)$  can easily be computed since  $\tau_{g,h}$  increases monotonically in  $Z$  and is therefore a bijective transformation. Thus, the  $p$ -th quantile of the distribution of  $Y$  is simply  $\tau_{g,h}(z_p)$  where  $z_p$  is the  $p$ -th quantile of the standard normal distribution. The estimation of the parameters  $g$  and  $h$  from data is carried out by computing and fitting empirical quantiles. However, the generalization to the multivariate setting is not straightforward, mainly because multivariate quantiles need to be defined appropriately. Research is currently underway on this topic.

The second approach consists of modifying the probability density of a random variable instead of the random variable itself. This approach was first developed by Azzalini (1985) for the univariate normal distribution and by Azzalini and Dalla Valle (1996) for the multivariate normal distribution, yielding the so-called skew-normal distribution. Extensions to skew-elliptical distributions were proposed by Azzalini and Capitanio (1999), Branco and Dey (2001), Sahu et al. (2003), and include for instance skew- $t$  and skew-Cauchy distributions. All these distributions are particular types of generalized skew-elliptical distributions recently introduced by Genton and Loperfido (2002), i.e. they are defined as the product of a multivariate elliptical density  $g$  with a skewing function  $\pi$ :

$$(1.2) \quad h_n(\mathbf{x}) = 2|\Omega|^{-1/2} \cdot g[\Omega^{-1/2}(\mathbf{x}-\boldsymbol{\xi})] \cdot \pi[\Omega^{-1/2}(\mathbf{x}-\boldsymbol{\xi})], \quad \mathbf{x} \in \mathbb{R}^n,$$

where  $\boldsymbol{\xi} \in \mathbb{R}^n$  and  $\Omega \in \mathbb{R}^{n \times n}$  are location and scale parameters respectively,  $0 \leq \pi(\mathbf{x}) \leq 1$ , and  $\pi(-\mathbf{x}) = 1 - \pi(\mathbf{x})$ . The skew-elliptical distributions are attractive because their properties are very similar to those of the normal distribution, and include the normal distribution as a particular case.

In this paper we will provide a practical application of the skew-elliptical distribution (1.2) by modeling the distribution of a long time series of sea level and then using the distribution to predict changes in flooding risk associated with rising sea level. It will be shown that the skew- $t$  distribution leads to an effective and parsimonious description of the sea level process and can be used to take into account its strong seasonality and other forms of nonstationarity.

The paper is structured as follows. In Section 2, we define skew-elliptical random processes and describe in detail the skew-normal and skew- $t$  distributions. Section 3 presents the sea level analysis and conclusions are presented in Section 4.

**2. Skew-elliptical distributions and stochastic processes.** We say that a stochastic process  $\{X_t, t \in T\}$  is a skew-elliptical time series if and only if its distribution functions are all multivariate skew-elliptical. In particular, Gaussian times series are skew-elliptical time series. Skew-elliptical time series are very flexible and allow us to model a wide variety of natural phenomena. In this section we focus on two simple examples, namely skew-normal time series and skew- $t$  time series.

**2.1. Skew-normal time series.** A skew-normal time series is a skew-elliptical time series defined by the multivariate pdf (1.2) where  $g$  is a multivariate normal pdf  $\phi_n$  with correlation matrix  $\Omega$  and  $\pi$  is the standard normal univariate cdf  $\Phi$ , i.e.

$$(2.1) \quad h_n(\mathbf{x}) = 2\phi_n(\mathbf{x} - \boldsymbol{\xi}; \Omega) \Phi(\boldsymbol{\alpha}^T(\mathbf{x} - \boldsymbol{\xi})).$$

The parameter  $\boldsymbol{\alpha} \in \mathbb{R}^n$  controls the skewness and  $\boldsymbol{\alpha} = \mathbf{0}$  reduces to the multivariate normal case. A random vector  $\mathbf{x} \in \mathbb{R}^n$  with pdf given by (2.1) is said to be multivariate skew-normal and was first introduced by Azzalini and Dalla Valle (1996). Its first two moments are given by

$$(2.2) \quad E(\mathbf{x}) = \boldsymbol{\xi} + \sqrt{\frac{2}{\pi}}\boldsymbol{\delta},$$

$$(2.3) \quad \text{Var}(\mathbf{x}) = \Omega - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}^T,$$

where

$$(2.4) \quad \boldsymbol{\delta} = \frac{\Omega\boldsymbol{\alpha}}{(1 + \boldsymbol{\alpha}^T\Omega\boldsymbol{\alpha})^{1/2}}.$$

Note that both the expectation and the variance of  $\mathbf{x}$  depend on the skewness parameter  $\boldsymbol{\alpha}$  (or equivalently  $\boldsymbol{\delta}$ ). In order to describe a stationary skew-normal time series with constant mean, we require that  $\boldsymbol{\xi} = \xi\mathbf{1}_n$  and  $\boldsymbol{\delta} = \delta\mathbf{1}_n$ , where  $\mathbf{1}_n = (1 \dots 1)^T \in \mathbb{R}^n$ . Figure 1 depicts the contours of the pdf (2.1) for  $n = 2$ ,  $\boldsymbol{\xi} = (0, 0)^T$ ,  $\Omega$  the correlation matrix with correlation 0.5, with  $\boldsymbol{\alpha} = (0, 0)^T$  (left panel) and  $\boldsymbol{\alpha} = (2, 2)^T$  (right panel). Note that the left panel corresponds to the bivariate standard normal distribution with correlation 0.5.

It is rather straightforward to simulate realizations of skew-normal time series. Indeed, Azzalini and Capitanio (1999) showed that if

$$\begin{pmatrix} Z_0 \\ \mathbf{z} \end{pmatrix} \sim N_{n+1}(\mathbf{0}, \Omega^*), \quad \Omega^* = \begin{pmatrix} 1 & \boldsymbol{\delta}^T \\ \boldsymbol{\delta} & \Omega \end{pmatrix}$$

where  $Z_0$  is a scalar component,  $\Omega^*$  is a correlation matrix, then

$$\mathbf{x} = \begin{cases} \mathbf{z} & \text{if } Z_0 > 0, \\ -\mathbf{z} & \text{otherwise,} \end{cases}$$

has a skew-normal distribution with parameters  $\boldsymbol{\xi} = \mathbf{0}$ ,  $\Omega$ , and  $\boldsymbol{\alpha}$ , where

$$\boldsymbol{\alpha} = \frac{\Omega^{-1}\boldsymbol{\delta}}{(1 - \boldsymbol{\delta}^T\Omega^{-1}\boldsymbol{\delta})^{1/2}}.$$

The multivariate skew-normal distribution enjoys many pleasant properties. For instance, marginal distributions of a multivariate skew-normal

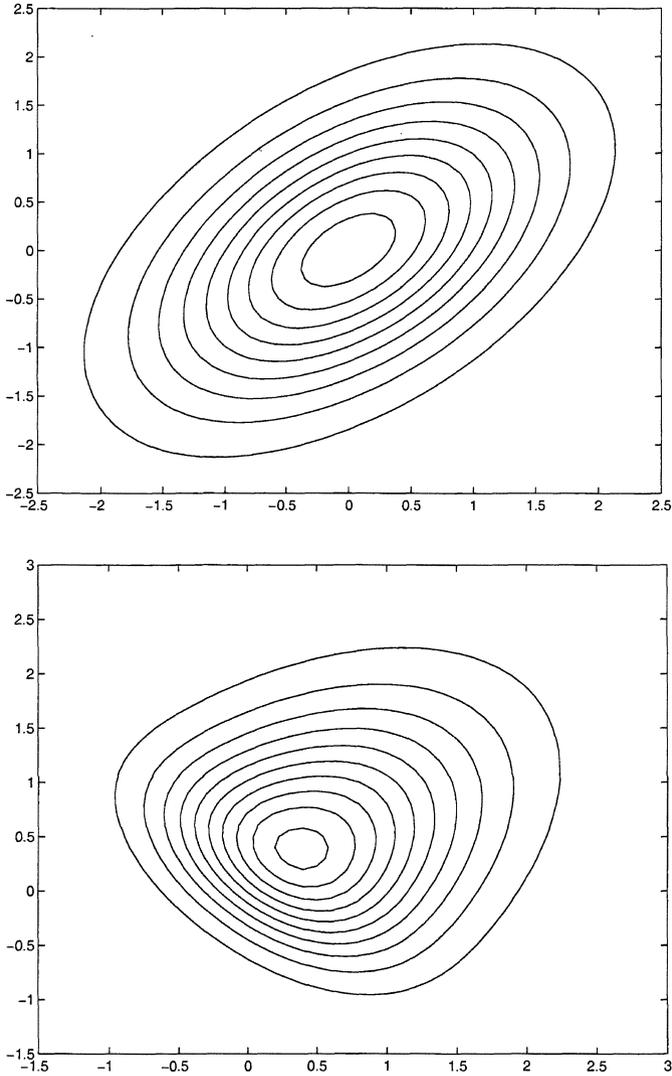


FIG. 1. Contours of the multivariate skew-normal pdf (2.1) for  $n = 2$ ,  $\xi = (0, 0)^T$ ,  $\Omega$  the correlation matrix with correlation 0.5, with  $\alpha = (0, 0)^T$  (top panel) and  $\alpha = (2, 2)^T$  (bottom panel).

distribution are still skew-normal. Further properties and applications of skew-normal distributions can be found in Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), and Genton et al. (2001). One drawback of skew-normal time series is that only skewness is accounted for, and not kurtosis. The next section deals with this issue.

**2.2. Skew- $t$  time series.** A skew- $t$  time series is a skew-elliptical time series defined with the multivariate pdf (1.2) where  $g$  is a multivariate  $t$  pdf  $t_n$  with correlation matrix  $\Omega$  and  $\pi$  is the standard  $t$  univariate cdf  $T$ , i.e.

$$(2.5) \quad h_n(\mathbf{x}) = 2t_n(\mathbf{x} - \boldsymbol{\xi}; \Omega)T(\boldsymbol{\alpha}^T(\mathbf{x} - \boldsymbol{\xi})),$$

where

$$(2.6) \quad t_n(\mathbf{x}; \Omega) = \frac{\Gamma((\nu + n)/2)}{(\pi\nu)^{n/2}\Gamma(\nu/2)} |\Omega|^{-1/2} (1 + \nu^{-1}\mathbf{x}^T\Omega^{-1}\mathbf{x})^{-(\nu+n)/2}.$$

The parameter  $\boldsymbol{\alpha} \in \mathbb{R}^n$  controls the skewness and  $\boldsymbol{\alpha} = 0$  reduces to the multivariate  $t$  case. The parameter  $\nu$  controls the kurtosis and  $\nu = 1$  reduces to the multivariate skew-Cauchy distribution, whereas  $\nu \rightarrow \infty$  reduces to the multivariate skew-normal distribution. The multivariate skew- $t$  distribution was introduced by Branco and Dey (2001) and its first two moments are given by

$$(2.7) \quad E(\mathbf{x}) = \boldsymbol{\xi} + \frac{\Gamma((\nu - 1)/2)}{\Gamma(\nu/2)} \sqrt{\frac{\nu}{\pi}} \boldsymbol{\delta}, \quad \text{if } \nu > 1,$$

$$(2.8) \quad \text{Var}(\mathbf{x}) = \frac{\nu}{\nu - 2} \Omega - \left( \frac{\Gamma((\nu - 1)/2)}{\Gamma(\nu/2)} \right)^2 \frac{\nu}{\pi} \boldsymbol{\delta} \boldsymbol{\delta}^T, \quad \text{if } \nu > 2,$$

where

$$(2.9) \quad \boldsymbol{\delta} = \frac{\Omega \boldsymbol{\alpha}}{(1 + \boldsymbol{\alpha}^T \Omega \boldsymbol{\alpha})^{1/2}}.$$

Note that both the expectation and the variance of  $\mathbf{x}$  depend on the skewness parameter  $\boldsymbol{\alpha}$  (or equivalently  $\boldsymbol{\delta}$ ). In order to describe a stationary skew- $t$  time series with constant mean, we require that  $\boldsymbol{\xi} = \xi \mathbf{1}_n$  and  $\boldsymbol{\delta} = \delta \mathbf{1}_n$ , where  $\mathbf{1}_n = (1 \dots 1)^T \in \mathbb{R}^n$ . Here again, the marginal distributions of a multivariate skew- $t$  distribution are still skew- $t$ , see Branco and Dey (2001).

**3. Modeling the distribution of sea level.** To illustrate the usefulness of multivariate skew- $t$  time series we will now model variations in sea level recorded at Charlottetown, a coastal city on the south shore of Prince Edward Island in the Gulf of St Lawrence (the tide gauge is located at 46.23°N, 63.12° W). Charlottetown was chosen for two reasons. First, it is a low-lying coastal city and so there is considerable interest in the possible effect of climate change on flooding risk. Second, Charlottetown has one of the longest sea level records in Canada and this has allowed us to examine the usefulness of the multivariate skew- $t$  distribution in modeling both seasonality and other forms of nonstationarity.

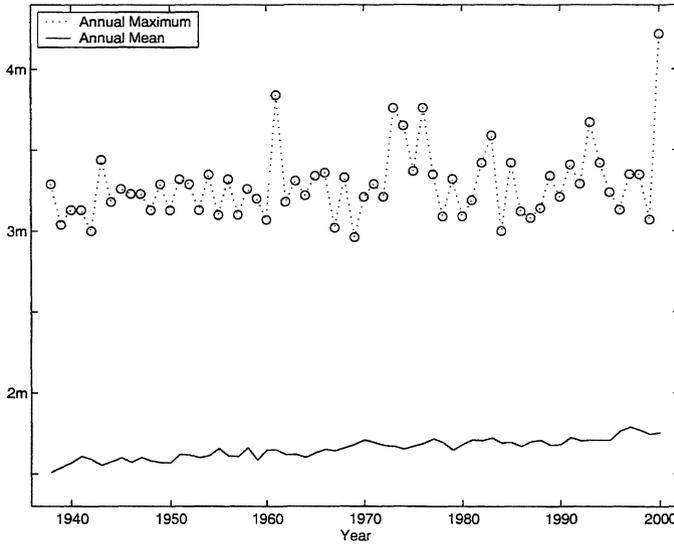


FIG. 2. Annual means and annual maxima of Charlottetown hourly sea levels, 1938 to 2000 inclusive. The levels are specified with respect to local Chart Datum.

**3.1. Annual means, annual maxima, and residuals.** The annual means of sea level at Charlottetown have been increasing almost linearly at a rate of about 3 mm per year from 1938 to the present day (Figure 2, lower trace). This linear trend is due in part to a rise in global sea level of about 1 mm per year; the remainder is believed to be due to subsidence of the Earth's crust in this region.

The annual maxima of the hourly sea levels are, not surprisingly, more variable than the annual means (Figure 2, upper trace). Variations in the annual maxima can exceed 1 m (compare the maxima for 1999 and 2000). From Figure 3 we can conclude that the distribution of the annual maxima for each half of the record are reasonably consistent with a Type I extreme value distribution, i.e. with a cdf of the form  $\exp(-\exp(-(y-\alpha)/\beta))$  where  $\alpha$  and  $\beta$  are location and scale parameters (see for example Leadbetter et al., 1982). From Figure 3 it can also be seen that the probability of an annual maximum not exceeding a specified level is higher for the first half of the record, pointing to an increase in the annual maxima in recent decades.

Note the 1.5 m offset of the two traces in Figure 2. A significant part of this difference between annual means and annual maxima is due to the tide, the dominant component of sea level variability at Charlottetown. The decomposition of sea level into its tidal and aperiodic component is illustrated in Figure 4 for the latter part of January, 2000. During this period an intense storm piled up water in the southern Gulf of St Lawrence and caused extensive flooding of downtown Charlottetown. In fact this

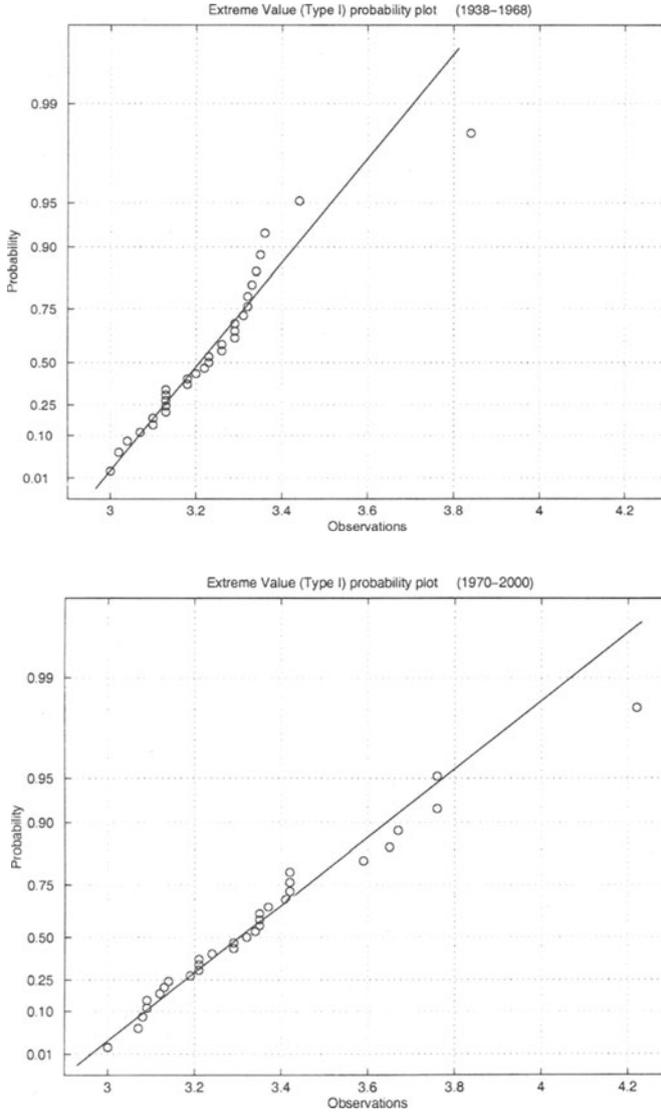


FIG. 3. *Extreme value (Type I) probability plots based on the annual maxima of Charlottetown hourly sea levels. The top and bottom panels are for the first and last halves of the record respectively. The straight lines are the maximum likelihood fits.*

storm caused the highest sea level observed at Charlottetown between 1938 and 2000 (compare Figures 2 and 4).

The tidal component of sea level is forced by the gravitational pull of the sun and moon and, for the purposes of this study, can be treated as deterministic. The difference between the observed sea level and the

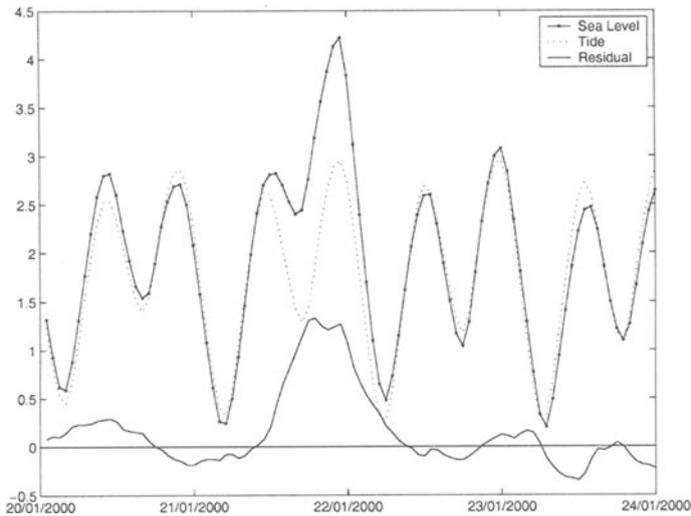


FIG. 4. *Decomposition of sea level into tidal and residual components. The solid line with dots shows the observed hourly sea level variation at Charlottetown from 20 to 23 January, 2000. The dotted line is the predicted tide. The difference between the observed sea level and predicted tide is termed the residual. It is shown by the solid line that fluctuates about zero.*

tide is caused by oceanographic and meteorological factors and is usually called the “residual” (i.e. the part that remains after the predicted tide is subtracted from the observed level). The most important cause of residual variability at Charlottetown is the passage of intense storms and the associated variations in air pressure and wind. It should therefore not be surprising to learn that the residuals exhibit strong seasonal dependence with the largest residuals occurring in winter (Figure 5).

In this study we have treated the residuals as stochastic and modeled their bivariate distribution with a bivariate skew- $t$  distribution. From the fitted bivariate skew- $t$  distribution we were then able to estimate flooding risk as explained in the following section.

Before fitting the bivariate skew- $t$  distribution to the residuals the sea level record was split into two equal halves (1938-1968, 1970-2000 inclusive) and the residuals were stratified by calendar month. This allowed us to model secular and seasonal changes in the residual distribution. Twelve bivariate skew- $t$  distributions were then fit to pairs of adjacent hourly residuals from each half of the record: one for January, one for February and so on through to December.

The method of maximum likelihood was used to fit the bivariate skew- $t$  distribution to pairs of adjacent residuals. Let  $\mathbf{x}_t = (x_t, x_{t+1})^T$  denote a residual pair starting at time  $t$ . To reduce the dependence amongst

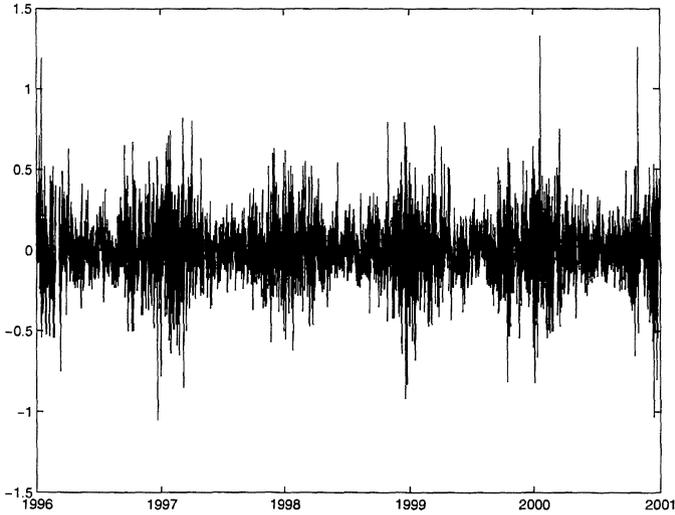


FIG. 5. Variations in the hourly sea level residuals for the last 5 years of the Charlottetown record. Note the strong seasonality in the record with the largest residuals always occurring in winter. The ticks mark the first day of January for each year.

pairs, they were subsampled every  $\tau$  hours and the following likelihood function, which assumes independence, was maximized with respect to  $\xi$ ,  $\Omega_{11}$ ,  $\Omega_{12}/\Omega_{11}$ ,  $\nu$  and  $\alpha$ :

$$h(\mathbf{x}_1, \mathbf{x}_{1+\tau}, \mathbf{x}_{1+2\tau}, \dots) = h_2(\mathbf{x}_1)h_2(\mathbf{x}_{1+\tau})h_2(\mathbf{x}_{1+2\tau}) \cdots$$

where  $h_2(\mathbf{x})$  is the bivariate skew- $t$  distribution (2.5). It is possible to obtain another  $\tau - 1$  estimates by starting the subsampling at times 2 through  $\tau$ . Using the fact that maximum likelihood estimators are asymptotically unbiased we averaged the  $\tau$  estimates to provide a more reliable estimate of the 5 parameters.

Table 1 summarizes the statistics of residual variability and the estimated parameters of the bivariate skew- $t$  distribution for the second half of the record. (We assumed a subsampling rate of  $\tau = 24$  hours which is reasonable given the strength of serial correlation in the residual record.) We see that the sample standard deviation ( $s$ ) exhibits a pronounced seasonal variation with winter values that are more than double the summer values. This is consistent with Figure 5. The autocorrelations at lag 1,  $\hat{\rho}(1)$ , are close to unity but slightly smaller in summer. If the residual process is AR(1) the e-folding time in February is  $-\lceil \log(0.97) \rceil^{-1} = 32$  hours; in August the e-folding time is  $-\lceil \log(0.94) \rceil^{-1} = 16$  hours. This means the residual process has a shorter “memory” in summer and this is another aspect of the seasonality of the process. The kurtosis indicates heavier tails

TABLE 1

Statistics of residual variability and the estimated parameters of the bivariate skew- $t$  distribution for the period 1970 to 2000. The first column gives the month of the year for which the residuals were analyzed. The next four columns give the sample standard deviation, autocorrelation at a lag of 1 hour, kurtosis ( $ku$ ) and skewness ( $sk$ ) calculated directly from the residuals. The remaining 5 columns list the estimated parameters of the bivariate skew- $t$  distribution as defined in the text.

Month	$s$	$\hat{\rho}(1)$	$ku$	$sk$	$\hat{\xi}$	$\hat{\Omega}_{11}^{1/2}$	$\frac{\hat{\Omega}_{12}}{\hat{\Omega}_{11}}$	$\hat{\nu}$	$\hat{\alpha}$
1	0.23	0.96	4.9	0.13	0.005	0.19	0.97	6.2	0.01
2	0.21	0.97	4.1	0.15	-0.036	0.18	0.97	7.5	0.38
3	0.20	0.97	4.4	0.29	-0.015	0.17	0.97	7.2	0.29
4	0.15	0.96	4.3	0.49	-0.046	0.13	0.97	6.6	2.03
5	0.11	0.95	3.9	0.25	-0.056	0.11	0.96	9.9	2.11
6	0.10	0.95	4.0	0.33	-0.035	0.10	0.96	9.6	2.26
7	0.09	0.95	3.6	0.32	-0.039	0.10	0.96	11.8	3.67
8	0.10	0.94	3.6	0.36	-0.056	0.10	0.96	10.4	3.91
9	0.13	0.95	4.4	0.37	-0.032	0.11	0.96	7.5	1.45
10	0.16	0.96	5.4	0.41	-0.033	0.13	0.96	5.8	0.79
11	0.18	0.96	4.2	0.30	-0.024	0.16	0.97	6.9	0.60
12	0.23	0.96	4.8	0.13	-0.010	0.19	0.97	5.4	0.14

than the normal, with lightest tails in summer, and the skewness is positive throughout the year. For the estimated parameters of the bivariate skew- $t$  distribution, we see that estimates of  $\nu$  and  $\alpha$  tend to be highest in summer.

Typical probability plots are shown in Figure 6. It is clear that the skew- $t$  distribution fits the residual histograms better than the normal distribution. Of particular note is the ability of the skew- $t$  distribution to describe the heavy tails of the January residual histogram. Similar fits were found for the other 11 months. Figure 7 shows probability density contours for the bivariate normal and skew- $t$  distributions fit to the residuals from April, 1970 to 2000. This figure clearly shows the ability of the skew- $t$  distribution to model the positive skewness evident in the residuals.

**3.2. Flooding risk.** The extreme value probability plot shown in Figure 3 gives some indication of flooding risk. For example, the probability an annual maximum will not exceed the critical level  $\eta_c = 3.8$  m is estimated to be 0.957 (corresponding to a “return time” of  $1/(1-0.957) = 23.2$

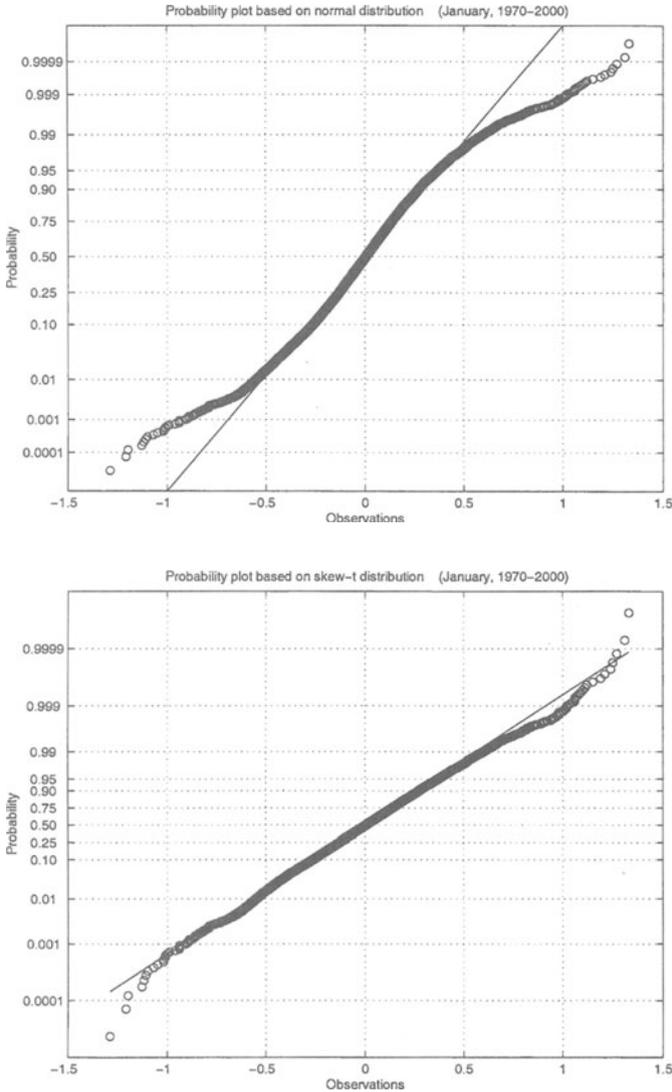


FIG. 6. Probability plots for the January residuals, 1970 to 2000 inclusive. The top-hand panel is based on the normal distribution and the bottom-hand panel is based on the skew- $t$  distribution.

years). Assuming the annual maxima are independent, the probability that no hourly sea level will exceed  $\eta_c$  over a 31 year period (the time span of 1970 to 2000) is  $0.957^{31} = 0.256$ .

To show how the skew- $t$  distributions can be used to calculate flooding probabilities, such as those described in the preceding paragraph, we

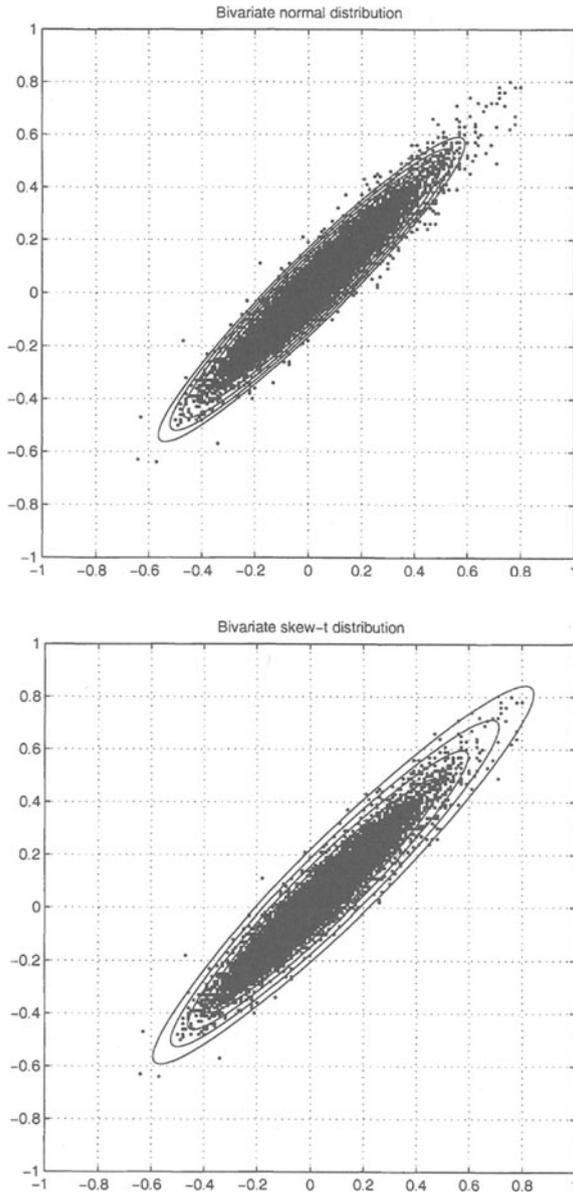


FIG. 7. Contours of the bivariate normal pdf (top panel) and the bivariate skew- $t$  pdf (bottom panel) fitted to all pairs of adjacent April residuals from 1970 to 2000 (using a subsampling rate of  $\tau = 24$  hours). The contour values are logarithmically spaced and equal to  $10^{-2}$ ,  $10^{-1.5}$ ,  $10^{-1}$ ,  $10^{-0.5}$ ,  $10^0$ ,  $10^{0.5}$  and  $10$ . The dots are the observed residual pairs.

expand as follows the probability that all hourly sea levels between hour 1 and  $n$  are less than a specified critical level  $\eta_c$ :

$$(3.1) \quad p_{12\dots n} = p_1 p_{2|1} p_{3|2,1} \cdots p_{n-1|n-2,\dots,1} p_{n|n-1,\dots,1}$$

where  $p_{t|t-1,t-2,\dots,1}$  is the probability the sea level is below  $\eta_c$  at time  $t$  given it was below for the earlier times  $t-1$  through 1. To simplify (3.1) we assume that for large values of  $\eta_c$  it is only necessary to condition on the preceding  $M$  hours. This leads to the approximation

$$(3.2) \quad p_{12\dots n} \approx p_1^{(M)} p_2^{(M)} p_3^{(M)} \cdots p_{n-1}^{(M)} p_n^{(M)}$$

where  $p_t^{(M)} = p_{t|t-1,t-2,\dots,t-M}$ .

To evaluate  $p_t^{(M)}$  we write the sea level at hour  $t$  as follows:

$$\eta_t = \eta_t^T + \eta_t^R$$

where superscript  $T$  and  $R$  denote the tidal and residual components respectively. The event  $\eta_t < \eta_c$  is equivalent to  $\eta_t^R < \eta_c - \eta_t^T$  i.e. the residual at time  $t$  is not large enough to “jump the gap” between the predicted tide and the critical level. Thus  $p_t^{(M)}$  can be expressed in terms of conditional probabilities involving  $\eta_n^R$  which, in turn, can be estimated by fitting an  $M+1$  dimensional skew- $t$  distribution to the observed residuals. This has been done for the case  $M=1$  (i.e. conditioning on the previous sea level) after stratifying the residuals by month to allow for seasonality as explained earlier. It was then a straight forward calculation to compound the conditional probabilities according to (3.2) and estimate the probability the sea level would not exceed  $\eta_c$ .

The result of a typical calculation of  $p_{12\dots n}$  using the skew- $t$  and taking  $\eta_c = 3.8$  m is shown in Figure 8. The small steps in the trace correspond to the seasonal transitions from summer to winter when the residuals tend to be larger and the probability of an exceedance of  $\eta_c$  is greatest. (This is consistent with the seasonal variation in variance shown in Table 1.) The probability  $\eta_c$  will not have been exceeded during this 31 year period is 0.126 according to Figure 8. This is equal to the probability that all annual maximum over this 31 year period are below  $\eta_c = 3.8$  m. If we assume the annual maxima are independent, and denote by  $p$  the probability that an annual maximum is below  $\eta_c$ , it follows that  $p^{31} = 0.126$ . This implies  $p = 0.126^{1/31} = 0.934$ . This calculation allows us to take the probabilities calculated using (3.2) and plot them on the extreme value probability plot of the annual maxima. This has been done in Figure 9 for  $\eta_c$  increasing from 3.40 m to 4.25 m in steps of 0.05 m. The agreement between return periods calculated by the two approaches is encouraging and suggests we may be able to use the skew- $t$  distribution and conditional probability approach to calculate flooding risks in a strongly nonstationary setting (e.g. changing

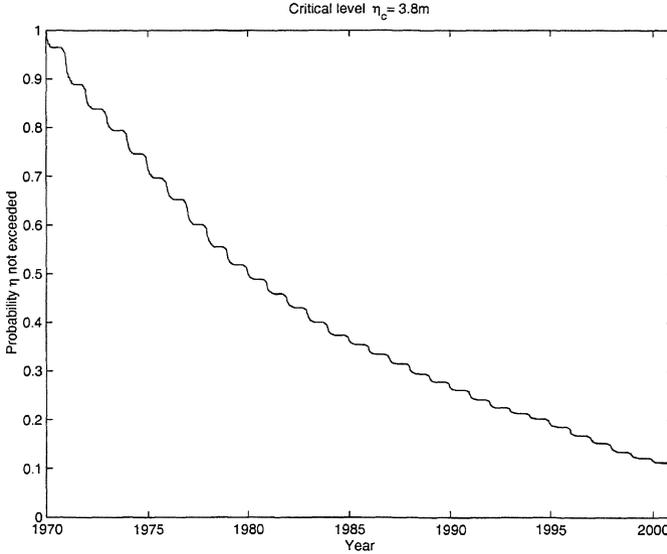


FIG. 8. Probability that sea level will not exceed  $\eta_c = 3.8$  m as a function of time starting January 1, 1970. The trace was obtained by compounding conditional probabilities calculated from the skew- $t$  distributions fit to the residuals from 1970 to 2000. See text for details.

the mean, or perhaps slowly increasing the standard deviation of the skew- $t$  distribution through time to reflect changes in storminess under various climate change scenarios).

To conclude this section we present in Figure 10 a calculation showing the probability of Charlottetown sea level not exceeding  $\eta_c = 4.5$  m over the next 100 years. The predicted tides for this period were calculated in the standard way and the seasonally stratified, bivariate skew- $t$  distributions based on 1970-2000 residuals were used to calculate  $p_t^{(1)}$ . To gauge the effect of rising sea level on flooding risk we added two linear sea level trends to the predicted tide with different slopes: 3 mm per year is the present value and 7 mm per year is the value proposed by the International Panel on Climate Change as a plausible rate for the coming century. Superimposed on the gradual drop and seasonal steps in probability can be seen a small oscillation with a period of about 20 years. This is due to the nodal tide which is forced by moon's gravitational pull; it has an amplitude of several centimeters and a period of 18.6 years. The effect of the increased rate of sea level rise on flooding risk is clearly evident in Figure 10: the probability of at least one exceedance of 4.5 m during the next century is about 0.3 if sea level continues to rise at its present rate, and about 0.8 if it increases to 7 mm per year.

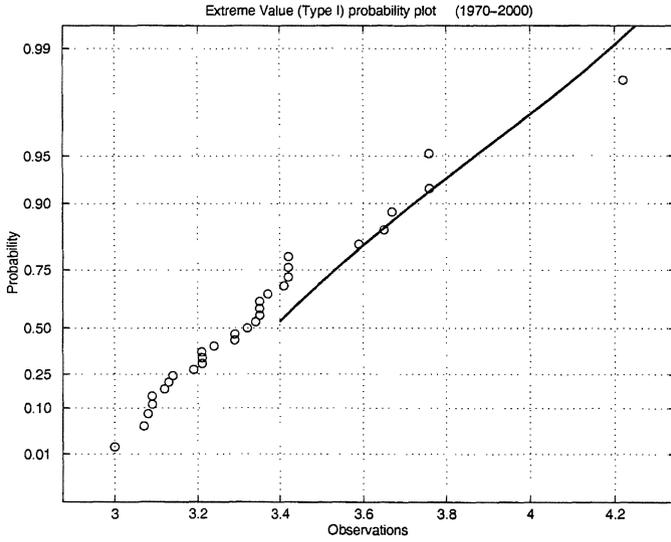


FIG. 9. Extreme value probability plot based on annual maxima of hourly sea levels, 1970-2000 inclusive. The line shows the probabilities calculated using the skew-t distribution and (3.2) as explained in the text.

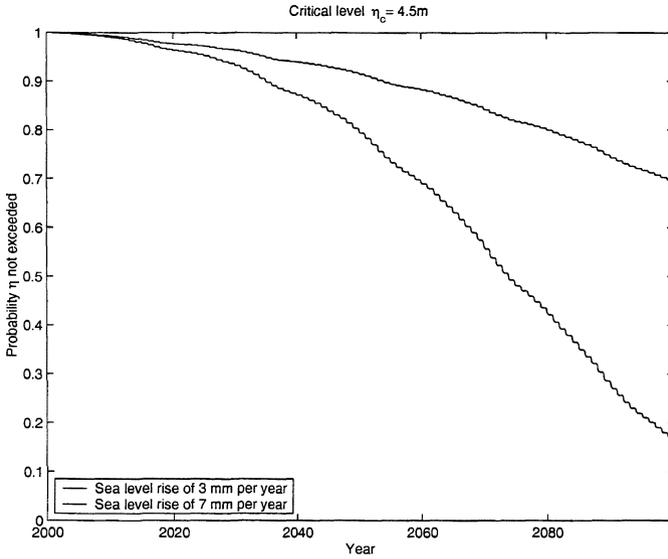


FIG. 10. Probability of not exceeding the critical level  $\eta_c = 4.5\text{ m}$  over the next century assuming a rate of sea level rise of 3 mm per year (upper trace) and 7 mm per year (lower trace). The results are based on compounding conditional probabilities calculated from the skew-t distribution fit to the seasonally stratified residuals, 1970 to 2000 inclusive.

**4. Conclusions.** In this article, the multivariate skew- $t$  distribution has been shown to fit well the distribution of residual sea level at Charlottetown. In particular it captured the heavy tails and skewness in the residuals, features not reproducible by the normal distribution. The bivariate skew- $t$  distribution was used to estimate conditional probabilities of sea level not exceeding a given level, given it was below on the previous time step. By compounding these conditional probabilities it was possible to recover the return periods of extreme sea level calculated in the standard way using extreme value theory and observed annual maxima. This encouraged us to use the conditional probability approach to calculate the risk of flooding a specified critical level as a function of time over the next century. The calculated probabilities reflect the strong seasonality in the sea level process and also long period tidal effects. An increase in the rate of rise of sea level from 3 to 7 mm per year was shown to have a dramatic effect on flooding risk.

The reason simple compounding of conditional probabilities can give reasonable return period of extreme levels is that sea level at Charlottetown is dominated by the tide. This means that even though the highest residuals in the record may not have occurred at the highest tides, which we treat as deterministic, the probability of this coincidence occurring can be estimated without extrapolating into the tails of the residual distribution. This idea has been used by a number of authors following the lead of Pugh and Vassie (1980).

The present approach was designed specifically to quantify flooding risk in a strongly nonstationary setting and the preliminary results presented here are encouraging. There are however a number of issues that need to be addressed before the method is used for practical purposes including, most importantly, the number of residual levels on which to condition i.e. the choice of  $M$  and its dependence on  $\eta_c$  in strongly nonstationary situations.

**Acknowledgments.** Part of this research was carried out while the first author was visiting the second at Dalhousie University, whose hospitality is gratefully acknowledged. The authors thank an anonymous referee for valuable and constructive comments on the article. We would also like to thank Bruce Smith and Chris Field for useful discussions.

#### REFERENCES

- [1] A. AZZALINI, *A class of distributions which includes the normal ones*, Scand. J. Stat., **12** (1985), pp. 171–178.
- [2] A. AZZALINI AND A. DALLA VALLE, *The multivariate skew-normal distribution*, Biometrika, **83** (1996), pp. 715–726.
- [3] A. AZZALINI AND A. CAPITANIO, *Statistical applications of the multivariate skew normal distribution*, J. R. Statist. Soc., B **61** (1999), pp. 579–602.
- [4] M.D. BRANCO AND D.K. DEY, *A general class of multivariate skew-elliptical distributions*, Journal of Multivariate Analysis, **79** (2001), pp. 99–113.

- [5] M.G. GENTON, L. HE, AND X. LIU, *Moments of skew-normal random vectors and their quadratic forms*, *Statistics and Probability Letters*, **51** (2001), pp. 319–325.
- [6] M.G. GENTON AND N.M. LOPERFIDO, *Generalized skew-elliptical distributions and their quadratic forms* (2002), Institute of Statistics Mimeo - Series #2539, under review.
- [7] D.C. HOAGLIN, F. MOSTELLER AND J.W. TUKEY, *Exploring Data Tables, Trends, and Shapes*, Wiley, 1985.
- [8] M.R. LEADBETTER, G. LINDGREN, AND H. ROOTZEN, *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics, Springer-Verlag, 1982.
- [9] D.T. PUGH AND I.M. VASSIE, *Applications of the Joint Probability Method for Extreme Sea Level Computations*, *Proceedings of the Institute of Civil Engineers*, Part 2, 69, 959–975, 1980.
- [10] S.K. SAHU, D.K. DEY, AND M.D. BRANCO, *A new class of multivariate skew distributions with applications to Bayesian regression models*, *The Canadian Journal of Statistics*, **31** (2003).