The estimation of M4 processes with geometric moving patterns

Zhengjun Zhang

Abstract There are many parameters in multivariate maxima of moving maxima processes—or M4 processes. However, the more parameters there are, the more difficult it is to estimate them. It is not just an issue of numerical stability, of course. The statistical precision of the estimates will be poor if the number of parameters is too large. We consider asymmetric geometric structures which correspond to special moving patterns of extreme observations in observed time series. We study the model identifiability and propose parameter estimators. All proposed estimators are shown to be consistent and asymptotically joint normal. Simulation study and real data modeling of North Sea wave height data are illustrated.

Keywords Multivariate nonlinear time series · Max-stable process · Multivariate maxima of moving maxima · Extreme value theory · Empirical distribution · Parameter estimation

1 Introduction

Modeling extreme observations in multivariate time series is a difficult exercise. There have been well-developed approaches in modeling extreme observations in univariate time series. Excellent references addressing probabilistic properties can be found in books by Galambos (1987), Leadbetter et al. (1983), and Resnick (1987), among others. Smith (1990) gives a comprehensive account of statistical aspects, especially the maximum likelihood methods in parameter estimation. Meanwhile, Embrechts et al. (1997) give an excellent viewpoint of

Z. Zhang (✉)
Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA
e-mail: zjz@stat.wisc.edu

modeling extremal events. However, problems concerning the environment, finance and insurance, etc. are multivariate in nature: for example, extremal sea wave heights at different locations (and possibly at different time) may follow similar moving patterns as illustrated in Fig. 3 (details in Sect. 6); floods may occur at different sites along a coastline; the failure of a portfolio investment — i.e., a big loss of the total investment—may be caused by a single extreme price movement or extreme price co-movements among several risk factors. Specification and identification of dependencies cross-sequences and within each sequence in multivariate time series analysis are very important, especially when the occurrences of extremal events are concerned.

There have been many attempts to characterize the dependencies—or the joint probabilities—of extremal events. These attempts—among others— include de Haan and Resnick (1977), de Haan (1985), and Resnick's (1987) point process approach, and Pickands's (1981) representation theorem for multivariate extreme value distribution with unit Fréchet margins. While multivariate extreme value distributions do not have unified parametric form, fortunately, models have been developed that are multivariate extreme value distributions. Many models are reviewed by Coles and Tawn (1991). Coles and Tawn (1994) also demonstrate how statistical methods for multivariate extremes may be applied to a very practical problem of data analysis.

Models for extreme observations in multivariate time series analysis are still limited. Even in the context of univariate time series, only a small number of models have been proposed since 1980s. Deheuvels (1983) defines what he calls the moving minimum (MM) process. The results of Deheuvels (1983) are strong, but the model itself is still not easily tractable for the estimation of parameters. Davis and Resnick (1989) study what they call the max-autoregressive moving average (MARMA(p,q)) process of a stationary process. Some basic properties of the MARMA processes have been shown and the prediction of a max-stable process has been studied relatively completely. However, much less is known about estimation of the MARMA process. For prediction, see also Davis and Resnick (1993). Recently, Hall et al. (2002) discuss moving maximum models which have a direct relation with MM processes. For a finite number of parameters, they propose parameter estimators based on empirical distribution functions. Theoretically, all these models assume Fréchet margins. In most applications, however, data do not follow Fréchet margins. Therefore, a marginal transformation is needed in order to apply these models.

Methods for exceedance modeling over high thresholds are widely used in the applications of extreme value theory. The theory used goes back to Pickands (1975). In statistical applications, one fits a generalized extreme value distribution (GEV) to subsample maxima or a generalized Pareto distribution (GPD) to excesses over a threshold, and the data are transformed to have asymptotically Fréchet margins. The full details of modern statistical treatments can be seen in Embrechts et al. (1997), and Smith (2003) and the references therein. In financial time series studies of extremes, Smith (2003), Zhang (2002, 2005), and Zhang and Smith (2002) applied a GEV data fitting and transformation procedure in their models. Beside these transformation procedures, other alternatives

such as rank transformations can also be adopted. The rank transformation of wave height data will be discussed in Sect. 6.

In the context of multivariate time series analysis, Smith and Weissman (1996) extend Deheuvels' definition to a more general framework—i.e., multivariate maxima of moving maxima (henceforth M4) process. The M4 processes are defined as

$$Y_{id} = \max_l \max_k a_{l,k,d} Z_{l,i-k}, \quad d = 1, \ldots, D, \quad -\infty < i < \infty, \tag{1}$$

where $\{Z_{li}, l \geq 1, -\infty < i < \infty\}$ are an array of independent unit Fréchet random variables. The constants $\{a_{l,k,d}, l \geq 1, -\infty < k < \infty, 1 \leq d \leq D\}$ are nonnegative constants satisfying

$$\sum_{l=1}^{\infty} \sum_{k=-\infty}^{\infty} a_{l,k,d} = 1, \quad \text{for } d = 1, \ldots, D. \tag{2}$$

While MM processes are only specified over one index, however, there are possibilities to easily extend them over two indexes. In the framework of M4 processes, now all random variables Y_{id}s are unit Fréchet random variables, and the logarithms of the bivariate joint distribution functions $Pr(Y_{id} \leq x_{id}, Y_{i'd'} \leq x_{i'd'})$ for any i, i', d, d' are linear functions of parameters. These properties will be emphasized in the next section. It is hoped that the extension of MM processes to M4 processes will make it possible to estimate model parameters easily.

Following de Haan (1984), (1) defines max-stable processes, and we have, for any finite number r and positive constants $\{y_{id}\}$,

$$
\begin{aligned}
&Pr(Y_{id} \leq y_{id}, 1 \leq i \leq r, 1 \leq d \leq D) \\
&= Pr\left(Z_{l,i-k} \leq \frac{y_{id}}{a_{l,k,d}} \text{ for } l \geq 1, -\infty < k < \infty, 1 \leq i \leq r, 1 \leq d \leq D\right) \\
&= Pr\left(Z_{l,m} \leq \min_{1-m \leq k \leq r-m} \min_{1 \leq d \leq D} \frac{y_{m+k,d}}{a_{l,k,d}}, l \geq 1, -\infty < m < \infty\right) \\
&= \exp\left(-\sum_{l=1}^{\infty} \sum_{m=-\infty}^{\infty} \max_{1-m \leq k \leq r-m} \max_{1 \leq d \leq D} \frac{a_{l,k,d}}{y_{m+k,d}}\right).
\end{aligned}
\tag{3}
$$

This is (2.5) of Smith and Weissman (1996) and we have

$$Pr^n(Y_{id} \leq ny_{id}, 1 \leq i \leq r, 1 \leq d \leq D) = Pr(Y_{id} \leq y_{id}, 1 \leq i \leq r, 1 \leq d \leq D)$$

which indicates that $\{\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iD})\}$ are max-stable. Smith and Weissman (1996) have argued that the extreme values of a multivariate stationary process may be characterized in terms of a limiting max-stable process under quite general conditions. They also show that a very large class of max-stable processes may be approximated by M4 processes mainly because those processes have

the same multivariate extreme indexes as M4 processes have (Theorem 2.3 in Smith and Weissman, 1996).

In modeling multivariate time series data by M4 process however, there are an infinite number of parameters in the model. To estimate an infinite number of parameters, even a large number of parameters, would, perhaps, be too ambitious a task unless certain structures of parameters are assumed. Usually, for two finite numbers L and K, we assume the index l varies from 1 to L and the index k varies from $-K_1$ to K_2, where $K_1 \geq 0$, $K_2 \geq 0$, $K = K_1 + K_2 + 1$. The constant K gives the order—or the moving range—of the M4 process used to model a real data set. The order K can be thought as a measure of extremal dependence of a sequence of random variables. This measure tells us that Y_{id} and $Y_{i+K,d}$ are (extremely) dependent when $k \leq K$ but not when $k > K$. Roughly speaking, the extremal dependence between two random variables means that the conditional probability of one variable exceeding a threshold given that another variable has exceeded the threshold is non-zero as the threshold tends to infinity. However, if either K or L is large and the parameters $a_{l,k,d}$ are unconstrained, there will be many free parameters to estimate and the statistical precision of the resulting estimators will be poor. The main purpose of this paper is to propose some models in which all the $a_{l,k,d}$ are expressed in terms of a small number of free parameters. By reducing the size of the parameter space, we hope to improve the statistical properties of the fitting method.

As illustrated in Fig. 3, there may be a characteristic behavior of a process near a local maximum, whereby observations either rise or fall for several time steps before reversing direction. This locally monotone behavior could be modeled by an M4 process with monotone coefficients. Motivated by these facts, we consider the parameters $a_{l,k,d}$ in an M4 model being mixtures of asymmetric, symmetric, and monotone geometric structures on the index k. When the extreme observations in a time period have an increasing (decreasing) trend at the beginning, then a decreasing (increasing) trend after a peak (valley) is reached, these observations may well be approximated by an asymmetric geometric structure or a symmetric geometric structure. When a locally monotone trend is observed, those observations may well be approximated by a locally monotone geometric structure. These structures substantially reduce the number of parameters in the model and make the estimation easier. But they violate the model identifiability assumptions made in Zhang and Smith (2002). The objective of this work is to provide model identifiable conditions, to develop estimating procedures for parameters with asymmetric geometric structures and to model extreme wave heights at different locations on North Sea.

This paper is organized as follows: in Sect. 2, we study some general properties regarding M4 processes. In Sect. 3, some basic results regarding empirical distribution functions are developed. A key lemma regarding the joint limiting distribution is established at the end of the section. In Sect. 4, we introduce the model with asymmetric geometric moving patterns when extremal events occur. For the case of multiple moving patterns in univariate processes, a key lemma regarding the identifiability of all parameters is proved first. Then the

consistency and asymptotic normality of the estimators which are the solutions of a system of nonlinear estimating equations are stated and proved. For the general case of multiple moving patterns in multivariate processes, sufficient conditions under which the model is identifiable are provided and the estimating equations which are used to obtain estimated parameter values are also provided. Consistency and asymptotic normality are illustrated. A simulation example is given in Sect. 5. Real data of wave heights on North Sea is modeled by our proposed models. Conclusions are given in Sect. 7 and the more involved proofs are deferred to Sect. 8.

2 The models

There are infinitely many parameters in (1). But in practice any moving patterns will not last forever. For instance, a heavy rainfall lasts a couple of days and stops, large insurance claims caused by a disaster event last a certain time period, and a sudden extreme price movement in financial markets may cause big jumps or dumps in asset prices for a couple of days and then the market is back to normal until a new extremal event occurs. Under the asymmetric geometric parameter structures, we may be able to model scale transformed observations using a workable M4 model. We content ourselves with concentrating attention on a finite dimensional M4 process model:

$$Y_{id} = \max_{1 \le l \le L} \max_{-K_1 \le k \le K_2} a_{l,k,d} Z_{l,i-k}, \quad d = 1, \ldots, D, \quad -\infty < i < \infty, \qquad (4)$$

where L, K_1 and K_2 are finite non-negative numbers, $\sum_{l=1}^{L} \sum_{k=-K_1}^{K_2} a_{l,k,d} = 1$ for $d = 1, \ldots, D$. Now there are total $(K_1 + K_2 + 1) \times L \times D$ parameters in the model.

Notice that L, K_1 and K_2 are generally unknown in practice. For the purpose of developing an estimation theory, they are assumed known. The references of determining those values include Zhang and Smith (2002), and Zhang (2005), etc. Their methods are based on the concepts of lag-k extreme dependence and a new extreme co-movement measure. In our data (wave heights) modeling, we simply apply clustering method to the transformed unit Fréchet scaled data and find the 'best' values for L, K_1 and K_2.

Under model (4), when an extremal event occurs or when a large Z_{li} occurs, $Y_{id} \propto a_{l,i-k,d}$ for $i \approx k$—i.e., if some Z_{lk} is much larger than all the neighboring Z values, we will have $Y_{id} = a_{l,i-k,d} Z_{lk}$ for i close to k. This indicates a moving pattern existing in observed time series. This moving pattern is called a signature pattern (Smith and Weissman, 1996). The constant L is the maximum number of signature patterns, and K_1 and K_2 characterize the range of sequential dependencies. The constant $K_1 + K_2 + 1$ represents the order of the moving maxima processes.

When we model real data by an M4 process, the original data has to be transformed into unit Fréchet scale. If an exact transformation to an M4 model

exists, but it's not perfectly estimated, than the estimated transformed process will be only approximately M4. Therefore the transformed data may not exactly follow (4). Let us consider the following model and compare it with (4):

$$\tilde{Y}_{id} = \max_{1 \le l \le L} \max_{-K_1 \le k \le K_2} a_{l,k,d} Z_{l,i-k} + N_i, \quad d = 1, \ldots, D, \quad -\infty < i < \infty, \quad (5)$$

where $\{N_i\}$ are iid white noises with mean 0 and standard deviation σ, $\sum_{l=1}^{L} \sum_{k=-K_1}^{K_2} a_{l,k,d} = 1$ for $d = 1, \ldots, D$. A comparison between these two models is shown in Fig. 1. The idea here is to show that an M4 process model may be a good approximation to an observed process which nearly follows (5).

In Fig. 1, Panel a shows simulated data from a symmetric geometric moving pattern M4 process (solid lines) and simulated data from (5) (dashed dotted lines) in a range of 365 days; Panels b–e are plots within local windows. One can see b and d are similar but in a different vertical scales. Similar phenomena can be observed in c and e for solid lines. Panels b and d are thought to follow one signature pattern, meanwhile Panels c and e are thought to follow another signature pattern though the dashed dotted lines in c and e are quite different. We can notice that the dashed dotted lines in Panels b, c, and d are not as significant as the dashed dotted lines in Panel e. These are due to the plot scaling and the facts that large Z_{li}s occur and the first term (the M4 part) in (5) is much larger than the second term (the noise part). From Panel e in Fig. 1, while we see that dashed dotted lines are not showing any symmetric patterns, however, it may still be a better choice to use certain M4 process with geometric moving patterns (solid lines) to approximate those clustered extreme observations (dashed dotted lines).

The theory of Smith and Weissman implies that M4 processes will approximate the extremal behavior of wide classes of multivariate time series models. We will focus on Model (4) and hope this model can be used to approximate most clustered extreme observations, especially when the moving range is large.

Our goal is to estimate all parameters $\{a_{l,k,d}\}$ under the constraints that the parameters are nonnegative and their sum is equal to one for each $d = 1, \ldots, D$. Under model (4), it is easy to obtain the joint distribution of $\{Y_{id}, 1 \le i \le r, 1 \le d \le D\}$ from (3). However, the method of maximum likelihood is not directly applicable in this instance due to the degeneracy of the multivariate joint distribution function of the M4 processes. An alternative way to develop estimators is to use the joint empirical distribution functions which are studied in the next section.

For any two component processes—say the dth and d'th—and choices of non-negative values of y, $y_{id}, y_{i+r,d}, y_{1d}, y_{1d'}$, it follows immediately from (3) and (4) that

$$Pr(Y_{id} \le y) = e^{-1/y}, \tag{6}$$

$$Pr(Y_{id} \le y_{id}, Y_{i+r,d} \le y_{i+r,d}) = \exp\left[-\sum_{l=1}^{L} \sum_{m=1-K_2}^{1+r+K_1} \max\left\{ \frac{a_{l,1-m,d}}{y_{id}}, \frac{a_{l,1+r-m,d}}{y_{i+r,d}} \right\} \right], \tag{7}$$
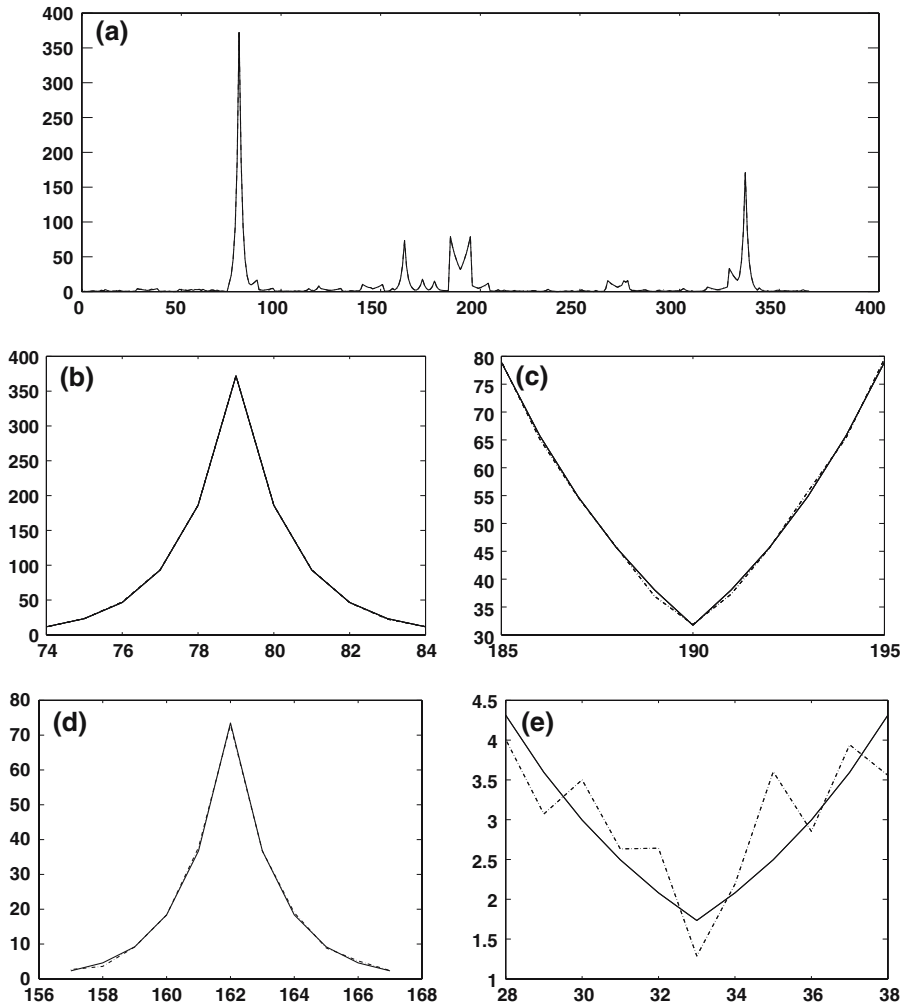
Fig. 1 Plots for an M4 process (*solid lines*), and a sum of an M4 process and a white noise process (*dashed dotted lines*). Panel (a) is a simulated 365 days data for a specific d in (4). b–e are partial pictures drawn from the whole simulated data. They show two different moving patterns— called signature patterns— in certain time periods when extremal events occur

$$Pr(Y_{1d} \le y_{1d}, Y_{1d'} \le y_{1d'}) = \exp\left[-\sum_{l=1}^{L} \sum_{m=1-K_2}^{1+K_1} \max\left\{ \frac{a_{l,1-m,d}}{y_{1d}}, \frac{a_{l,1-m,d'}}{y_{1d'}} \right\} \right], \quad (8)$$

where r is a positive integer and $a_{l,K_2+j,d} = 0$, $a_{l,-K_1-j,d} = 0$, $j = 1,\dots,r$, $d = 1,\dots,D$. When we estimate Model (4), we use a fixed number of r.

Notice that the marginal distributions of the process contain no information about the dependence properties, one must examine at least the bivariate distributions to make any meaningful inferences. In practice, for many processes the

bivariate distributions are sufficient, one does not need trivariate and higher-order joint distributions. Therefore, the theory and methods will be developed for the bivariate distribution functions in this study.

3 Some basic results regarding empirical distribution functions

In this section, we first study the properties of the empirical distribution functions constructed from Model (4). These properties will be used to study asymptotic normality of all estimators constructed in the next section.

Now for each d, let $A_{1d} = (0, x_{1d}) \times (0, y_{1d}), \ldots, A_{md} = (0, x_{md}) \times (0, y_{md})$ be different sets for some $m > 0$ and $I_{(.)}$ be an indicator function. Let $\{Y_{id}, d = 1, 2, \ldots, D; i = 1, 2, \ldots, n+r\}$ be a sample from (4). For a fixed number r, define the empirical probability that $(Y_{id}, Y_{i+r,d})$ falls in A_{jd} as

$$\bar{X}_{A_{jd}} = \frac{1}{n} \sum_{i=1}^{n} I_{A_{jd}} \{Y_{id}, Y_{i+r,d}\}. \tag{9}$$

The asymptotic results related to $\bar{X}_{A_{jd}}$s were first derived in Zhang and Smith (2002) when $r = 1$. In this work, r has to be 2 or larger in order to show that M4 processes with geometric moving patterns are identifiable. This is illustrated in Sect. 4. Now we state some basic results and notations for a general r. Proofs of the following two Lemmas are direct generalizations of Lemmas 3.2 and 3.3 in Zhang and Smith (2002).

The strong law of large numbers (SLLN) implies

$$\bar{X}_{A_{jd}} \xrightarrow{a.s.} Pr(A_{jd}) = Pr(Y_{1d} \le x_{jd}, Y_{1+r,d} \le y_{jd}) \triangleq \mu_{jd}, \quad \text{as } n \to \infty. \tag{10}$$

Lemma 1 *Suppose that $\bar{X}_{A_{jd}}$ and μ_{jd} are defined in (9) and (10), respectively. If $\sigma_{jd} > 0$, then*

$$\sqrt{n}(\bar{X}_{A_{jd}} - \mu_{jd}) \xrightarrow{\mathcal{L}} N(0, \sigma_{jd}^2), \text{ as } n \to \infty,$$

where

$$\sigma_{jd}^2 = \mu_{jd} - \mu_{jd}^2 + 2 \sum_{k=1}^{K_1+K_2+1} \left[Pr\big(Y_{1d} \le x_{jd}, Y_{1+r,d} \le y_{jd}, Y_{1+k,d} \right.$$
$$\left. \le x_{jd}, Y_{1+r+k,d} \le y_{jd}\big) - \mu_{jd}^2 \right].$$

Lemma 2 *Suppose that* $\bar{X}_{A_{jd}}$ *and* μ_{jd} *are defined in (9) and (10), respectively. Then*

$$\sqrt{n}\left(\begin{bmatrix}\bar{X}_{A_{1d}}\\\vdots\\\bar{X}_{A_{md}}\end{bmatrix}-\begin{bmatrix}\mu_{1d}\\\vdots\\\mu_{md}\end{bmatrix}\right)\xrightarrow{\mathcal{L}}N\left(0,\Sigma_d+\sum_{k=1}^{K_1+K_2+1}\{W_{kd}+W_{kd}^T\}\right),\quad as\ n\to\infty,$$

where $\mu_{i,j,d}=Pr\{Y_{1d}\le\min(x_{id},x_{jd}),\ Y_{1+r,d}\le\min(y_{id},y_{jd})\}$, $\mu_{i,i,d}=\mu_{id}$, *the matrix* Σ_d *has the entries* $\sigma_{i,j,d}=\mu_{i,j,d}-\mu_{id}\mu_{jd}$, *the matrix* W_{kd} *has entries* $w_{kd}^{ij}=Pr(Y_{1d}\le x_{id},Y_{1+r,d}\le y_{id},Y_{1+k,d}\le x_{jd},Y_{1+r+k,d}\le y_{jd})-\mu_{id}\mu_{jd}$.

Our goal is to use (7) and (8) and their empirical counterparts—i.e., empirical distribution functions – to construct estimators for all parameters in the model. The probability evaluated at the points $(y_{id},y_{i+r,d})$ in (7) depends on the comparison of $a_{l,1-m,d}/y_{id}$ and $a_{l,1+r-m,d}/y_{i+r,d}$, and similarly in (8). By fixing one of y_{id} and $y_{i+r,d}$, say y_{id}, then $a_{l,1-m,d}/(a_{l,1+r-m,d}y_{id})$ is the change point of $\max(a_{l,1-m,d}/y_{id},a_{l,1+r-m,d}/y_{i+r,d})$ when $y_{i+r,d}$ varies. Without loss of generality, we fix $y_{id}=1$ for simplicity of our calculation. In a real data analysis, we may choose a threshold value u' and fix $y_{id}=u'$ since the data will have to be transformed to have a unit Fréchet marginal distribution according to a generalized extreme value distribution—or a generalized Pareto distribution—fitting of exceedances over a threshold u.
Now define the following two functions which play a key role in constructing parameter estimators.

$$q_d(x)=-x\log\{Pr(Y_{1d}\le 1,Y_{1+r,d}\le x)\},\qquad(11)$$
$$q_{dd'}(x)=-x\log\{Pr(Y_{1d}\le 1,Y_{1d'}\le x)\}.\qquad(12)$$

From (6) and (7), $q_d(x)$ and $q_{dd'}(x)$ can be expressed as

$$q_d(x)=\sum_{l=1}^{L}\left[\sum_{i=0}^{r-1}a_{l,K_2-i,d}x+\max(a_{l,K_2-r,d}x,\ a_{l,K_2,d})+\max(a_{l,K_2-r-1,d}x,\ a_{l,K_2-1,d})\right.$$
$$+\max(a_{l,K_2-r-2,d}x,\ a_{l,K_2-2,d})+\cdots+\max(a_{l,-K_1,d}x,\ a_{l,-K_1+r,d})$$
$$\left.+\sum_{i=0}^{r-1}a_{l,-K_1+i,d}\right],$$
$$d=1,\ldots,D,$$

$$(13)$$

and

$$q_{dd'}(x) = \sum_{l=1}^{L} \sum_{m=1-K_2}^{1+K_1} \max(a_{l,1-m,d}x, \, a_{l,1-m,d'}), \quad d, d' = 1, \ldots, D, \; d < d'.$$
(14)

From (13) and (14), it is easy to see that both $q_d(x)$ and $q_{dd'}(x)$ are piecewise linear functions and their slope changing (jumping) points—$a_{l,i+r,d}/a_{l,i,d}$ and $a_{l,1-m,d'}/a_{l,1-m,d}$—are determined by the parameter values. Meanwhile, when the values of $q_d(x)$ and $q_{dd'}(x)$ are given at a series of points, we can determine the slope jumping points. Under certain conditions, the slope jumping points uniquely determine the values of the parameters. The study of those conditions is postponed to the next section. We now study the properties of the empirical counterparts of $q_d(x)$ and $q_{1d'}(x)$.

The empirical counterparts of $q_d(x)$ and $q_{1d'}(x)$ are defined as:

$$U_d(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{Y_{id} \leq 1, \, Y_{i+r,d} \leq x\}}, \quad \widehat{q}_d(x) = -x \log [U_d(x)], \, d = 1, \ldots, D, \quad (15)$$

$$U_{1d'}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{Y_{i1} \leq 1, \, Y_{id'} \leq x\}}, \quad \widehat{q}_{1d'}(x) = -x \log [U_{1d'}(x)], \, d' = 2, \ldots, D. \quad (16)$$

Let

$$x_{1d}, \, x_{2d}, \, \ldots, \, x_{m_d d}, \quad d = 1, \, \ldots, \, D,$$

and

$$y_{1d'}, \, y_{2d'}, \, \ldots, \, y_{v_{d'} d'}, \quad d' = 2, \, \ldots, \, D,$$

be suitable choices of the points used to evaluate the values of all functions defined. Methods of choosing these points will be addressed in the following sections when we study model identifiability and model estimation. At the moment, they can be regarded as any arbitrary collections of points. The following notations and results will be used to derive the asymptotic properties of all estimators.

$$\mathbf{x} = \left(x_{11}, x_{21}, \ldots, x_{m_1 1}, x_{12}, \ldots, x_{m_D D}, y_{12}, y_{22}, \ldots, y_{v_2 2}, y_{13}, \ldots, y_{v_D D} \right)^{\mathrm{T}},$$

$$\mathbf{U} = \left(U_1(x_{11}), \ldots, U_1(x_{m_1 1}), U_2(x_{12}), \ldots, U_D(x_{m_D D}), U_{12}(y_{12}), \ldots, \right.$$
$$\left. U_{12}(y_{v_2 2}), \ldots, U_{1D}(y_{v_D D}) \right)^{\mathrm{T}},$$

$$\mathbf{q} = \left(q_1(x_{11}), \ldots, q_1(x_{m_1 1}), q_2(x_{12}), \ldots, q_D(x_{m_D D}), \right.$$
$$\left. q_{12}(y_{12}), \ldots, q_{12}(y_{v_2 2}), \ldots, q_{1D}(y_{v_D D}) \right)^{\mathrm{T}},$$

$$\mu_{djd} = E[U_d(x_{jd})] = Pr(Y_{1d} \leq 1, \ Y_{1+r,d} \leq x_{jd}),$$

$$\mu_{1d'jd'} = E[U_{1d'}(y_{j'd'})] = Pr(Y_{11} \leq 1, \ Y_{1d'} \leq y_{j'd'}),$$

$$\mu_{djd, \, d'j'd'} = E\left[\left(I_{\{Y_{1d} \leq 1, Y_{1+r,d} \leq x_{jd}\}} - \mu_{djd}\right)\left(I_{\{Y_{1d'} \leq 1, Y_{1+r,d'} \leq x_{j'd'}\}} - \mu_{d'j'd'}\right)\right]$$

$$\mu_{djd, \, 1d'j'd'} = E\left[\left(I_{\{Y_{1d} \leq 1, Y_{1+r,d} \leq x_{jd}\}} - \mu_{djd}\right)\left(I_{\{Y_{11} \leq 1, Y_{1d'} \leq y_{j'd'}\}} - \mu_{1d'j'd'}\right)\right],$$

$$\mu_{1d'j'd', \, djd} = E\left[\left(I_{\{Y_{11} \leq 1, Y_{1d'} \leq y_{j'd'}\}} - \mu_{1d'j'd'}\right)\left(I_{\{Y_{1d} \leq 1, Y_{1+r,d} \leq x_{jd}\}} - \mu_{djd}\right)\right],$$

$$\mu_{1djd, \, 1d'j'd'} = E\left[\left(I_{\{Y_{11} \leq 1, Y_{1d} \leq y_{jd}\}} - \mu_{1djd}\right)\left(I_{\{Y_{11} \leq 1, Y_{1d'} \leq y_{j'd'}\}} - \mu_{1d'j'd'}\right)\right],$$

$$w^{(k)}_{djd, \, d'j'd'} = E\left[\left(I_{\{Y_{1d} \leq 1, Y_{1+r,d} \leq x_{jd}\}} - \mu_{djd}\right)\left(I_{\{Y_{1+k,d'} \leq 1, Y_{1+r+k,d'} \leq x_{j'd'}\}} - \mu_{d'j'd'}\right)\right],$$

$$w^{(k)}_{djd, \, 1d'j'd'} = E\left[\left(I_{\{Y_{1d} \leq 1, Y_{1+r,d} \leq x_{jd}\}} - \mu_{djd}\right)\left(I_{\{Y_{1+k,1} \leq 1, Y_{1+k,d'} \leq y_{j'd'}\}} - \mu_{1d'j'd'}\right)\right],$$

$$w^{(k)}_{1d'j'd', \, djd} = E\left[\left(I_{\{Y_{1,1} \leq 1, Y_{1,d'} \leq y_{j'd'}\}} - \mu_{1d'j'd'}\right)\left(I_{\{Y_{1+k,d} \leq 1, Y_{1+r+k,d} \leq x_{jd}\}} - \mu_{djd}\right)\right],$$

$$w^{(k)}_{1djd, \, 1d'j'd'} = E\left[\left(I_{\{Y_{11} \leq 1, Y_{1d} \leq y_{jd}\}} - \mu_{1djd}\right)\left(I_{\{Y_{1+k,1} \leq 1, Y_{1+k,d'} \leq y_{j'd'}\}} - \mu_{1d'j'd'}\right)\right].$$

In the above notations, we have triple and quadruple subindexes like djd and $1djd$. For the tripe subindexes, we have $d = 1, 2, \ldots, D$, $j = 1, \ldots, m_d$; for the quadruple subindexes, we have $d = 2, \ldots, D$, $j = 1, \ldots, v'_d$. The ranges in $d'jd'$, and $1d'jd'$ are defined similarly.

It is not convenient to use the above notations to write a multivariate central limit theorem. This can be resolved by defining the following relations and transforming each triple or quadruple sub-index into a single index:

$$\begin{cases} \mu_{djd} \ \rightarrow \ \mu_s, \ q_d(x_{jd}) \rightarrow h_s, & \text{where } s = S_{d-1} + j, \ d = 1, \ldots, D, \\ \mu_{1d'jd'} \ \rightarrow \ \mu_s, \ q_{1d'}(y_{jd'}) \rightarrow h_s, & \text{where } s = S_D + S'_{d'-1} + j, \ d' = 2, \ldots, D, \end{cases}$$

where $S_0 = 0$, $S_d = \sum_{j=1}^{d} m_j$, $d = 1, \ldots, D$, $S'_1 = 0$, and $S'_{d'} = \sum_{j=2}^{d'} v_j$, $d' = 2, \ldots, D$.

Let $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_{S_D + S'_D})'$, $\mathbf{q} = (h_1, h_2, \ldots, h_{S_D + S'_D})'$. We now use the similar relations between the indexes of μ_{djd} and the indexes of μ_s to define the following variables:

$$\sigma_{st} = \begin{cases} \mu_{djd, d'j'd'}, & \text{if } s \leq S_D, \ t \leq S_D, \\ \mu_{djd, 1d'j'd'}, & \text{if } s \leq S_D, \ t > S_D, \\ \mu_{1djd, d'j'd'}, & \text{if } s > S_D, \ t \leq S_D, \\ \mu_{1djd, 1d'j'd'}, & \text{if } s > S_D, \ t > S_D, \end{cases}$$

$$w_k^{st} = \begin{cases} w_{djd,d'j'd'}^{(k)}, & \text{if } s \le S_D, \ t \le S_D, \\ w_{djd,1d'j'd'}^{(k)}, & \text{if } s \le S_D, \ t > S_D, \\ w_{1djd,d'j'd'}^{(k)}, & \text{if } s > S_D, \ t \le S_D, \\ w_{1djd,1d'j'd'}^{(k)}, & \text{if } s > S_D, \ t > S_D, \end{cases}$$

and the matrices:

$$\Sigma = (\sigma_{st}), \quad W_k = (w_k^{st}), \quad \Theta = (\text{diag}\{\mu\})^{-1} \times (\text{diag}\{\mathbf{x}\}).$$

Putting everything above together yields the following lemma. Its proof follows the Mean Value Theorem and the arguments used to establish Lemma 2.

Lemma 3 *For the choices of x_{jd}, $y_{j'd'}$ and the definitions of each variable above, we have*

$$\sqrt{n}(\mathbf{U} - \mu) \xrightarrow{\mathcal{L}} N\left(0, \Sigma + \sum_{k=1}^{K_1+K_2+1} \left\{W_k + W_k^{\mathrm{T}}\right\}\right), \quad as \ n \to \infty,$$

$$\sqrt{n}(\widehat{\mathbf{q}} - \mathbf{q}) \xrightarrow{\mathcal{L}} N\left(0, \Theta\left[\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k^{\mathrm{T}}\}\right]\Theta^{\mathrm{T}}\right), \quad as \ n \to \infty.$$

Having established the asymptotic properties of the statistics $\mathbf{U}$ and $\widehat{\mathbf{q}}$, we now turn our attention to the asymmetric geometric moving models.

4 The models for asymmetric geometric moving patterns

In Fig. 4, we plot transformed wave height data which are obtained by rank transformation and are asymptotically unit Fréchet scaled. We see that the plotted curves show piecewise geometric moving patterns. Motivated by these facts, we consider an asymmetric geometric parameter structure model:

$$a_{l,k+h(l,d),d} = b_{ld}\lambda_{ld}^{(k)+}\phi_{ld}^{(k)-}, \quad d = 1,\ldots,D, \ l = 1,\ldots,L,$$
$$k = -K_1(l,d),\ldots,K_2(l,d), \tag{17}$$

where $b_{ld} > 0$, $\phi_{ld} > 0$, $\lambda_{ld} \ge 0$, $a_{l,h(l,d),d} = b_{ld}$. When $\lambda_{ld} = 0$ or $\lambda_{ld} = 1/\phi_{ld}$, we have a monotone geometric signature pattern. When $\lambda_{ld} = \phi_{ld}$, we get a symmetric geometric signature pattern. The relation $a_{l,h(l,d),d} = b_{ld}$ corresponds to peaks (or valleys) in the lth signature patterns appearing in the dth process. When we have monotone signature patterns, the peaks or the valleys are the left starting points of those signature patterns. The difference $h(l,d) - h(l,d') = m$ means the starting time of the lth signature pattern in the d'th process is m time

units—say days—ahead of the starting time of the lth signature pattern in the dth process.

We assume that when a monotone geometric signature pattern appears the corresponding parameter λ_{ld} is equal to zero. This can be done by re-setting the values of $K_1(l, d)$ and $K_2(l, d)$ accordingly. Examples of monotone geometrically decreasing signature patterns can also be observed in Davis and Resnick's (1993) MARMA(1,q) processes.

In Sect. 2, we argued that the development of parameter estimators is through bivariate joint distributions. In the literature, when $r = 1$ in (13) and (17), the identifiability is shown by a simple case $L = 2$, where $\lambda_{ld} \neq \phi_{ld}$, $K_1(l, d) > 0$, $K_2(l, d) > 0$ in Zhang (2002). When we have pure symmetric or pure monotone geometric moving patterns, we can set $r = 1-$ see Zhang (2002). There is no results for $L > 2$ and asymmetric cases. Fortunately, the identifiability of the model can be shown when r is larger than 1. A natural choice would be to set $r = 2$ which is used in this paper, and the proofs are all based on $r = 2$.

Notice that the number of parameters is now $3 \times D \times L$. The model still has many parameters when L, or D is large. Also notice that moving coefficients $a_{l,k,d}$ associated with the dth univariate process are not associated with any other univariate processes, and hence we can estimate the parameters based on univariate processes. This argument suggests we at least have two ways to estimate parameters $a_{l,k,d}$. One way is to estimate them based on each univariate observed process only, and then identify the co-movement patterns between different processes to build the M4 model and study the joint distributions across sections. Another way is to estimate them based on the multivariate observed process. In fact, the most difficult part of identifying the model is its sequential dependencies (moving patterns), i.e. from each univariate process. We discuss the case $D = 1$ next.

4.1 The case $D = 1$

Since the values of $h(l, d)$ have no effects when we consider the distribution functions (7) for each d, and also notice that different $h(l, d)$ values mean different sample paths in model (4), therefore all $h(l, d)$s are set to zero in this section for simplicity.

The function $q(x)$ now can be written as:

$$q(x) = \sum_{l=1}^{L} b_l \left[\sum_{k=0}^{K_2(l)} \max(\lambda_l^{K_2(l)-k} x, \lambda_l^{K_2(l)-k+2} I_{(k>1)}) + \max(\phi_l x, \lambda_l) \right.$$
$$\left. + \sum_{k=2}^{K_1(l)+2} \max(\phi_l^k x I_{(k \leq K_1(l))}, \phi_l^{k-2}) \right].$$

Since $q(x)$ is a piecewise linear function of x, the jumping points of $q'(x)$ are $1/\phi_1^2, 1/\phi_2^2, \ldots, 1/\phi_L^2, \lambda_1^2, \lambda_2^2, \ldots, \lambda_L^2$ and $\lambda_1/\phi_1, \ldots, \lambda_L/\phi_L$. We assume all non-zero jumping points are distinct except when $\lambda_l/\phi_l = 1$ for some ls which

correspond to symmetric geometric signature patterns. The identifiability is stated in the following lemma.

Lemma 4 *Suppose that $\lambda_i \neq \lambda_j$, $\lambda_i \neq \phi_j$, $\phi_i \neq \phi_j$ for all nonzero λ_i, λ_j, $i \neq j$, $i, j = 1, \ldots, L$. Then there exist $x_1, \ldots, x_m$ such that $q(x_1), \ldots, q(x_m)$ uniquely determine all $b_l, \lambda_l, \phi_l, l = 1, \ldots, L$.*

A proof is postponed to Sect. 8.
Now suppose that $\widehat{q}(x_1), \widehat{q}(x_2), \ldots, \widehat{q}(x_m)$ are estimates of $q(x_1), q(x_2), \ldots, q(x_m)$. Then

$$(\widehat{q}(x_1), \widehat{q}(x_2), \ldots, \widehat{q}(x_m)) \xrightarrow{a.s.} (q(x_1), q(x_2), \ldots, q(x_m)), \quad \text{as } n \to \infty, \qquad (18)$$

where $\widehat{q}(x)$ is defined by (15). The following theorem follows immediately.

Theorem 1 *Suppose $x_1, \ldots, x_m$ are m points such that there are at least two points falling into each interval between any two adjacent slope jumping points of $q(x)$. Let $\widehat{q}(x_i)$ be the estimation of $q(x_i)$, and $\widehat{\mathbf{b}} = (\widehat{b}_1, \ldots, \widehat{b}_L)^{\mathrm{T}}$, $\widehat{\boldsymbol{\lambda}} = (\widehat{\lambda}_1, \ldots, \widehat{\lambda}_L)^{\mathrm{T}}$, $\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1, \ldots, \widehat{\phi}_L)^{\mathrm{T}}$ be the solutions of*

$$\widehat{q}(x_i) = \sum_{l=1}^{L} \widehat{b}_l \left[\sum_{k=0}^{K_2(l)} \max(\widehat{\lambda}_l^{K_2(l)-k} x_i, \widehat{\lambda}_l^{K_2(l)-k+2} I_{(k>1)}) + \max(\widehat{\lambda}_l x, \widehat{\phi}) \right.$$
$$\left. + \sum_{k=2}^{K_1(l)+2} \max(\widehat{\phi}_l^k x_i I_{(k \leq K_1(l))}, \widehat{\phi}_l^{k-2}) \right],$$
$$i = 1, \ldots, m. \qquad (19)$$

Then $(\widehat{\mathbf{b}}^{\mathrm{T}}, \widehat{\boldsymbol{\lambda}}^{\mathrm{T}}, \widehat{\boldsymbol{\phi}}^{\mathrm{T}}) \xrightarrow{a.s.} (\mathbf{b}^{\mathrm{T}}, \boldsymbol{\lambda}^{\mathrm{T}}, \boldsymbol{\phi}^{\mathrm{T}}), \quad \text{as } n \to \infty.$

Proof By (18) and Lemma 4.   □

As long as x is not a jumping point, and $q(x)$ is viewed as a function of all b_l, λ_l and ϕ_l, then $q(x)$ has all continuous first order partial derivatives in a neighborhood of (b_l, λ_l, ϕ_l), $l = 1, \ldots, L$. As a consequence of Theorem 1, Lemma 3 and the Mean Value Theorem, the following limiting distribution of all parameter estimators is obtained.

Theorem 2 *Under the same conditions as in Theorem 1. Let Q be a matrix whose i^{th} row is written as*:

$$Q_i = \left(\frac{\partial q(x_i)}{\partial b_1}, \ldots, \frac{\partial q(x_i)}{\partial b_L}, \frac{\partial q(x_i)}{\partial \lambda_1}, \ldots, \frac{\partial q(x_i)}{\partial \lambda_L}, \frac{\partial q(x_i)}{\partial \phi_1}, \ldots, \frac{\partial q(x_i)}{\partial \phi_L} \right),$$

and

$$J = (Q^{\mathrm{T}} Q)^{-1} Q^{\mathrm{T}}.$$

Then

$$\sqrt{n}\left(\begin{bmatrix}\widehat{\mathbf{b}}\\\widehat{\lambda}\\\widehat{\phi}\end{bmatrix}-\begin{bmatrix}\mathbf{b}\\\lambda\\\phi\end{bmatrix}\right) \xrightarrow{\mathcal{L}} N\left(0, J\Theta\left[\Sigma + \sum_{k=1}^{K_1+K_2+1}\{W_k + W_k^\mathrm{T}\}\right]\Theta J^\mathrm{T}\right), \text{ as } n \to \infty,$$

where $K_1 = \max_{1\le l\le L}\{K_1(l)\}$, $K_2 = \max_{1\le l\le L}\{K_2(l)\}$, W_k, Σ, Θ are defined in Lemma 3.

A proof is shown in Sect. 8.

4.2 The case $D > 1$

Although the solutions of (19) give consistent estimates of the parameters in each component process, the model (4) is not identifiable by $q_d(x)$ only. For example, the following two bivariate processes

$$\begin{cases}Y_{i1} = \max\{a_{1,-1,1}Z_{1,i+1}, a_{1,0,1}Z_{1,i}, a_{1,1,1}Z_{1,i-1}, a_{2,-1,1}Z_{2,i+1}, a_{2,0,1}Z_{2,i}, a_{2,1,1}Z_{2,i-1}\},\\ Y_{i2} = \max\{a_{1,-1,2}Z_{1,i+1}, a_{1,0,2}Z_{1,i}, a_{1,1,2}Z_{1,i-1}, a_{2,-1,2}Z_{2,i+1}, a_{2,0,2}Z_{2,i}, a_{2,1,2}Z_{2,i-1}\};\end{cases} \quad (20)$$

$$\begin{cases}Y'_{i1} = \max\{a_{1,-1,1}Z_{1,i+1}, a_{1,0,1}Z_{1,i}, a_{1,1,1}Z_{1,i-1}, a_{2,-1,1}Z_{2,i+1}, a_{2,0,1}Z_{2,i}, a_{2,1,1}Z_{2,i-1}\},\\ Y'_{i2} = \max\{a_{1,-1,2}Z_{1,i}, a_{1,0,2}Z_{1,i-1}, a_{1,1,2}Z_{1,i-2}, a_{2,-1,2}Z_{2,i}, a_{2,0,2}Z_{2,i-1}, a_{2,1,2}Z_{2,i-2}\}\end{cases} \quad (21)$$

are obviously two different bivariate processes and all parameters can be estimated using (19) for each univariate process under the asymmetric geometric parameter structure assumptions. But the estimates do not tell which one of (20) and (21) is the true model from which the observed data came. This problem can be resolved after considering the joint distributions within each component process and the joint distributions between any two component processes — see the following theorems.

Lemma 5 *Suppose the following conditions are satisfied*

1. *$\lambda_{id} \neq \lambda_{jd}, \lambda_{id} \neq \phi_{jd}, \phi_{id} \neq \phi_{jd}$ for all nonzero $\lambda_{id}, \lambda_{jd}, i \neq j, i, j = 1,\ldots,L$, $d = 1,\ldots,D$.*
2. *all existing non-zero ratios $a_{l,k,1}/a_{l',k'+h(l',d'),d'}$, $l, l' = 1,\ldots,L$ $k, k' = -K_1,\ldots,K_2$, for each $d' > 1$ are distinct, where $h(l,1) = 0$, $l = 1,\ldots,L$, $K_1 = \max_{1\le d\le D,1\le l\le L}\{K_1(l,d) - h(l,d)\}$, $K_2 = \max_{1\le d\le D,1\le l\le L}\{K_2(l,d) + h(l,d)\}$,*
3. *at least one of $a_{l,k,1}/a_{l,k+h(l,d'),d'}$, $k = -K_1,\ldots,K_2$, exists and is not zero when both $\sum_{k=-K_1}^{K_2} a_{l,k,1}$ and $\sum_{k=-K_1}^{K_2} a_{l,k,d'}$ are greater than 0 for each d',*
4. *$q_d(x)$, $q_{1d'}(x)$ are defined in (11) and (12) and their values are given.*

Let

$$x_{1d}, x_{2d}, \ldots, x_{m_d d}, \quad d = 1, \ldots, D,$$
$$y_{1d'}, y_{2d'}, \ldots, y_{v_{d'}d'}, \quad d' = 2, \ldots, D,$$

be suitable choices of points such that at least two x_{id}, x_{jd} values are between any two adjacent slope jumping points of $q_d(x)$, at least two y_{id}, y_{jd} values are between any two adjacent slope jumping points of $q_{1d}(x)$. Then the following system of nonlinear equations

$$
\begin{cases}
q_d(x_{jd}) = \sum_{l=1}^{L} b_{ld} \left[\sum_{k=0}^{K_2(l,d)} \max(\lambda_{ld}^{K_2(l,d)-k} x_{jd}, \lambda_{ld}^{K_2(l,d)-k+2} I_{(k>1)}) + \max(\lambda_{ld} x_{jd}, \phi_{ld}) \right. \\
\qquad\qquad \left. + \sum_{k=2}^{K_1(l,d)+2} \max(\phi_{ld}^{k} x_{jd} I_{(k \le K_1(l,d))}, \phi_{ld}^{k-2}) \right], \quad d=1,\dots,D, \ j=1,\dots,m_d, \\
q_{1d'}(y_{j'd'}) = \sum_{l=1}^{L} \sum_{k=-K_1}^{K_2} \max\{a_{l,k,1} y_{j'd'} I_{-K_1(l,1) \le k \le K_2(l,1)}, a_{l,k+h(l,d'),d'} I_{-K_1(l,d') \le k \le K_2(l,d')}\}, \\
\qquad\qquad d'=2,\dots,D, \ j'=1,\dots,v_{d'},
\end{cases}
\tag{22}
$$

gives the unique solutions which are the true values of the parameters in the model.

A proof is postponed to Sect. 8.

The estimators of all parameters are the solutions of a system of nonlinear estimating equations. The results are stated in the following theorem.

Theorem 3 *Under the same conditions as in Lemma 5. Functions $\widehat{q}_d(x)$, $\widehat{q}_{1d'}(x)$ are defined in (15), (16) respectively. Their values are given. Let Q be a matrix whose rows are written as either*

$$
\begin{aligned}
Q(x_{id}) = \left(\frac{\partial q_d(x_{id})}{\partial b_{11}}, \dots, \frac{\partial q_d(x_{id})}{\partial b_{LD}}, \frac{\partial q_d(x_{id})}{\partial \lambda_{11}}, \dots, \frac{\partial q_d(x_{id})}{\partial \lambda_{LD}}, \right. \\
\left. \frac{\partial q_d(x_{id})}{\partial \phi_{11}}, \dots, \frac{\partial q_d(x_{id})}{\partial \phi_{LD}} \right),
\end{aligned}
$$

or

$$
\begin{aligned}
Q(y_{id}) = \left(\frac{\partial q_{1d'}(y_{id})}{\partial b_{11}}, \dots, \frac{\partial q_{1d'}(y_{id})}{\partial b_{LD}}, \frac{\partial q_{1d'}(y_{id})}{\partial \lambda_{11}}, \dots, \frac{\partial q_{1d'}(y_{id})}{\partial \lambda_{LD}}, \right. \\
\left. \frac{\partial q_{1d'}(y_{id})}{\partial \phi_{11}}, \dots, \frac{\partial q_{1d'}(y_{id})}{\partial \phi_{LD}} \right).
\end{aligned}
$$

Then there exist

$$
x_{1d}, \ x_{2d}, \ \dots, \ x_{m_d d}, \quad d=1, \ \dots, \ D,
$$

and

$$
y_{1d'}, \ y_{2d'}, \ \dots, \ y_{v_{d'} d'}, \quad d'=2, \ \dots, \ D,
$$

such that the solutions of the following system of nonlinear estimating equations

$$
\begin{cases}
\widehat{q}_d(x_{jd}) = \sum_{l=1}^{L} \widehat{b}_{ld} \left[\sum_{k=0}^{K_2(l,d)} \max(\widehat{\lambda}_{ld}^{K_2(l,d)-k} x_{jd}, \widehat{\lambda}_{ld}^{K_2(l,d)-k+2} I_{(k>1)}) + \max(\widehat{\lambda}_{ld}x, \widehat{\phi}_{ld}) \right. \\
\qquad\qquad \left. + \sum_{k=2}^{K_1(l,d)+2} \max(\widehat{\phi}_{ld}^{k} x_{jd} I_{(k\leq K_1(l,d))}, \widehat{\phi}_{ld}^{k-2}) \right], \quad d=1,\ldots,D, \quad j=1,\ldots,m_d, \\
\widehat{q}_{1d'}(y_{j'd'}) = \sum_{l=1}^{L} \sum_{k=-K_1}^{K_2} \max\{\widehat{a}_{l,k,1} y_{j'd'} I_{-K_1(l,1)\leq k\leq K_2(l,1)}, \widehat{a}_{l,k+h(l,d'),d'} I_{-K_1(l,d')\leq k\leq K_2(l,d')}\}, \\
\qquad\qquad d'=2,\ldots,D, \quad j'=1,\ldots,v_{d'},
\end{cases}
$$

satisfy

$$
\sqrt{n}\left(\begin{bmatrix} \widehat{\mathbf{b}} \\ \widehat{\boldsymbol{\lambda}} \\ \widehat{\boldsymbol{\phi}} \end{bmatrix} - \begin{bmatrix} \mathbf{b} \\ \boldsymbol{\lambda} \\ \boldsymbol{\phi} \end{bmatrix} \right) \xrightarrow{\mathcal{L}} N\left(0, J\Theta\Big[\Sigma + \sum_{k=1}^{K_1+K_2+1} \{W_k + W_k^{\mathrm{T}}\}\Big]\Theta J^{\mathrm{T}} \right), \quad \text{as } n \to \infty,
$$

where $\widehat{q}_d(x_{jd})$, $\widehat{q}_{1d'}(y_{j'd'})$ are defined in (15) *and* (16)*, and*

$$
\begin{aligned}
\mathbf{b} &= (b_{11},\ldots,b_{L1},b_{12},\ldots,b_{1D},\ldots,b_{LD})^{\mathrm{T}}, \quad \widehat{\mathbf{b}} = (\widehat{b}_{11},\ldots,\widehat{b}_{L1},\widehat{b}_{12},\ldots,\widehat{b}_{1D},\ldots,\widehat{b}_{LD})^{\mathrm{T}}, \\
\boldsymbol{\lambda} &= (\lambda_{11},\ldots,\lambda_{L1},\lambda_{12},\ldots,\lambda_{1D},\ldots,\lambda_{LD})^{\mathrm{T}}, \quad \widehat{\boldsymbol{\lambda}} = (\widehat{\lambda}_{11},\ldots,\widehat{\lambda}_{L1},\widehat{\lambda}_{12},\ldots,\widehat{\lambda}_{1D},\ldots,\widehat{\lambda}_{LD})^{\mathrm{T}}, \\
\boldsymbol{\phi} &= (\phi_{11},\ldots,\phi_{L1},\phi_{12},\ldots,\phi_{1D},\ldots,\phi_{LD})^{\mathrm{T}}, \quad \widehat{\boldsymbol{\phi}} = (\widehat{\phi}_{11},\ldots,\widehat{\phi}_{L1},\widehat{\phi}_{12},\ldots,\widehat{\phi}_{1D},\ldots,\widehat{\phi}_{LD})^{\mathrm{T}}, \\
Q &= [Q(x_{id})^{\mathrm{T}}, d=1,\ldots,D, i=1,\ldots,m_d, Q(y_{id'})^{\mathrm{T}}, d'=2,\ldots,D, i=1,\ldots,v_{d'}]^{\mathrm{T}},
\end{aligned}
$$

$J = (Q^{\mathrm{T}}Q)^{-1}Q^{\mathrm{T}}$, *and Θ, Σ, W_k are defined the same as in Lemma 3.*

A proof is shown in Sect. 8.

So far, we have not said anything about how to determine the values of the order parameters $L, K_1(l,d), K_2(l,d), d=1,\ldots,D, l=1,\ldots,L$, and all tuning parameters x_{id}, y_{id}. We propose to use nearest neighbor clustering methods to determine these parameter values next.

5 Determining order parameter and tuning parameter values and simulation example

We propose an empirical procedure to determine tuning parameters and the order of moving ranges in this section. We first study some properties of ratios $Y_{i,d}/Y_{i+2,d}$, $i=1,2,\ldots$.

When we have exact Model (4), Zhang and Smith (2004) demonstrate there are infinitely many clusters of extreme observations $Y_{i_j,d}$, $Y_{i_j+1,d}$, $\ldots$, $Y_{i_j+K,d}$, $j=1,2,\ldots$, where i_j is the first location of the jth cluster, such that

$$
\frac{Y_{i_j,d}}{Y_{i_j+2,d}} = \frac{a_{l,K_2,d}Z_{l,i_j-K_2}}{a_{l,K_2-2,d}Z_{l,i_j-K_2}} = \frac{a_{l,K_2,d}}{a_{l,K_2-2,d}}
$$

conditioning on the very large Z_{l,i_j-K_2} values. The ratio $a_{l,K_2,d}/a_{l,K_2-2,d}$ equals one of the three values λ_l^2, $1/\phi_l^2$, or λ_l/ϕ_l.

In Model (5), suppose N_{id}s are bounded random variables, then in the jth cluster $\tilde{Y}_{i_j,d}$, $\tilde{Y}_{i_j+1,d}, \ldots, \tilde{Y}_{i_j+K,d}$, we have

$$\frac{\tilde{Y}_{i_j,d}}{\tilde{Y}_{i_j+2,d}} = \frac{a_{l,K_2,d}Z_{l,i_j-K_2} + N_{id}}{a_{l,K_2-2,d}Z_{l,i_j-K_2} + N_{i_j+2,d}} = \frac{\frac{a_{l,K_2,d}}{a_{l,K_2-2,d}} + \frac{N_{i_j,d}}{a_{l,K_2-2,d}Z_{l,i_j-K_2}}}{1 + \frac{N_{i_j+2,d}}{a_{l,K_2-2,d}Z_{l,i_j-K_2}}} \approx \frac{a_{l,K_2,d}}{a_{l,K_2-2,d}}.$$

This may be seen from scale transformed data. Bearing in mind that we use $\{Y_{id}\}$ process to approximate $\{\tilde{Y}_{id}\}$ process, under geometric moving pattern assumptions, the ratios of $Y_{i_j+s-2,d}/Y_{i_j+s,d}$ in a cluster of extreme observations approximately follow around one or two constant values depending on which model –i.e. monotone or asymmetric – is the true model from which the cluster of extreme observations is observed. This is related to a certain signature pattern. In a different cluster of extreme observations, we may get two different constant values which the computed ratios follow around. The property that the ratios in each cluster of extreme observations follow one or two constant values has been a main advantage of assuming geometric moving patterns since they greatly reduce the complexity of parameter estimations. Of course, if the computed ratios have large variations, it is better to apply general M4 process models and the estimation procedures developed by Zhang and Smith (2002) or other models if any.

The arguments above suggest that the slope jumping points of $q_d(x)$ are at those constants followed by ratios of very large observations. A very naive model may be to use each cluster of extreme observations as one signature pattern, but it may not be the best modeling strategy and it is hard to make statistical inference. In practice, clustering analysis method can be used to cluster those very large observations (above certain thresholds) into groups based on the consecutive ratios of $Y_{i_j+s-2,d}/Y_{i_j+s,d}$. Notice that when the value of L_d increases by 1, the total number of parameters increases by 3. A parsimonious model is to have $L_d = 1$. We propose the following procedure to determine order parameter and tuning parameter values.

1. Start with $d = 1$.
2. Set up $L_d = 1$.
3. For each d and a high threshold u (e.g. use the 90th percentile of the sample), find the maximal length ($\tilde{K}_d$) from all very large clustered extreme observations over the threshold u. Suppose the total number of such clusters is J_d. Within jth cluster, compute all ratios of $Y_{i_j+s-2,d}/Y_{i_j+s,d}$, $s = 2, \ldots, j_l$, where j_l is the length of the jth cluster, and find the point s_0 where the moving pattern changes from increasing (deceasing) trend to decreasing (increasing) trend. For monotone geometric moving patterns, s_0 is 1. For asymmetric geometric moving patterns, let $\mathbf{r}_j = (Y_{i_j+s-2,d}/Y_{i_j+s,d}, s = s_0 - 2, s_0 - 1, s_0, s_0 + 1, s_0 + 2)$. For monotone geometric moving patterns, let $\mathbf{r}_j = (Y_{i_j+s-2,d}/Y_{i_j+s,d}, s = s_0, s_0 + 1, s_0 + 2,, s_0 + 3, s_0 + 4)$.

4. Use nearest neighbor clustering analysis method to cluster the ratio vectors $\mathbf{r}_j, j = 1, \ldots, J_d$ into L_d groups.

5. Within each group, take the averages of the ratios as slope jumping points. Between any two adjacent jumping points, arbitrarily choose two points as x_{jd} values. For example, suppose r_1, r_2 are two adjacent ratios, then a natural choice would be $x_{jd} = r_1 + .25(r_2 - r_1)$, $x_{j+1,d} = r_1 + .75(r_2 - r_1)$. This choice is used in our simulation data and real data analysis. The total number of points x_{id} is denoted as m_d.

6. Within the range of $(1, \tilde{K}_d)$, search best $K_1(l, d)$ and $K_2(l, d)$ such that the fitted squared error $\sum_{i=1}^{m_d} \left\{ \widehat{q}_d(x_{id}) - \tilde{q}_d(x_{id}) \right\}^2$ is the smallest, where $\tilde{q}_d(x)$ is the fitted function to $q_d(x)$, i.e. the true parameter value in $q_d(x)$ are replaced by the estimated values. Denote the estimated b_{ld} values as $\widehat{b}_{ld}$.

7. Let $L_d = L_d + 1$. Go to Step 3 until there is at least one $\widehat{b}_{ld}$ which is 10% smaller than any estimated values of b_{ld}s from the last estimation and less than 0.001 (this number was suggested by simulation results).

8. Let $L_d = L_d - 1$ and use the values determined at Steps 5 and 6 for this L_d value. Let $d = d + 1$. Go to Step 2 until $d > D$.

9. For all clustered groups, assign the same group number to large clustered observations appearing simultaneously in different observed processes.

10. The choices of $y_{j'd}$ can be done from averaging the ratios of Y_{i1}/Y_{id} within the same group numbers obtained in Step 4 between two processes. $y_{j'd}$ can take the middle values of two adjacent ratios or take two values between two adjacent ratios like the previous step. The former method is used in this study.

11. After choosing x_{id} and $y_{j'd}$ values, use them to estimate parameters based on empirical functions $\widehat{q}_d(x)$ and $\widehat{q}_{1d}(x)$ such that the fitted squared error in Step 6 is minimized and signature patterns are matched by minimizing the fitted squared error: $\sum_{j=1}^{m_d} \left\{ \widehat{q}_{1d}(y_{jd}) - \tilde{q}_{1d}(y_{jd}) \right\}^2$, where $\tilde{q}_{1d}(x)$ is the fitted function to $q_{1d}(x)$.

Remark 1 We have suggested several *ad hoc* choices of numbers such as .25, .75, .001. In Step 2, we only keep five ratios which can be used to identify two parameters λ_{ld} and ϕ_{ld}. One can use other numbers, especially when prior information is available.

Remark 2 In some applications or simulations, some of order parameters, L_d, $K_1(l, d)$, $K_2(l, d)$, can easily be determined either by the knowledge of the real problem and time series plots over a high threshold values. As long as this information is given, we may start a model using these parameter values.

Remark 3 In Steps 5 and 10, theoretically, we can choose as many points of x_{jd} and $y_{j'd}$ as possible, but it is not realistic due to the intensive computation and the complexity of inferences. The goal is to choose moderate number of points such that the estimated values of parameters are close to true parameter values.

Remark 4 In Step 6, we adopt a Monte Carlo simulation approach to find best estimates which give best fit to $\widehat{q}_d(x)$ and $\widehat{q}_{1d}(x)$ functions based on the sum of squared errors.

Having established theories and proposed practically applicable methods, we now turn to some simulation example. We have conducted extensive simulation studies. Here we report one example which contains all kinds of geometric moving patterns.

Example 1 This example is a trivariate maxima of moving maxima as defined in (4). Let $\{Z_{l,i} : l = 1, 2; i \geq -11\}$ be two sequences of independent unit-Fréchet random variables. We define three sequences of coefficients:

$$\{a_{l,k,1} = b_{l1}\lambda_{l1}^{|k|}, \; k = 0, \pm 1, \ldots, \pm 5\},$$
$$\{a_{l,k,2} = b_{l2}\lambda_{l2}^{k}, \; k = 0, 1, \ldots, 4\},$$
$$\{a_{l,k,3} = b_{l3}\lambda_{l3}^{(k)+}\phi_{l3}^{(k)-}, \; k = -3, \ldots, 4\},$$

and their values are:

$$(b_{ld}) = \begin{bmatrix} .027065 & .128536 & 1/6 \\ 1/6 & 1/9 & .040737 \end{bmatrix}, \quad (\lambda_{ld}) = \begin{bmatrix} 1.2 & 0 & .5 \\ .5 & 0 & 1.2 \end{bmatrix}, \quad (\phi_{ld}) = \begin{bmatrix} 1.2 & 1.1 & .6 \\ .5 & .5 & 1.2 \end{bmatrix}.$$

The process $\{\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3}) : i \geq 1\}$ is generated by

$$Y_{id} = \max \left(\max_k a_{1,k,d} Z_{1,i-k}, \max_k a_{2,k,d} Z_{2,i-k} \right) + N_{id}, \quad d = 1, 2, 3, \quad (23)$$

where $\{N_{id}\}$ is a white noise process with $N_{id} \sim N(0, .01)$.

Table 1 gives estimated values of all coefficients and standard deviations computed from Theorem 3. All the estimates are very close to the true values and the standard deviations are small. The simulation estimates show the efficiency of the proposed estimating methods.

In the example, we use simulation sample size 10,000. The reason is to obtain more observations over certain threshold value. In some real data analysis, the data set usually contains more than 10,000 observations in each time series—for example, there are huge amount of data in each observed 'tick by tick' stock price time series; the number of points in our wave height data is about 11,000. The wave height data modeling is studied next.

6 Modeling wave height extremes on North Sea

6.1 The problem

Extreme wave heights have severe consequences for coastal structures and adverse effects on shipping and other marine operations. There are numerous studies regarding wave height distributions, regional failure probabilities such as de Haan and de Ronde (1998), Van Gelder et al. (2000), and references therein. However, studies regarding extreme wave height co-movements can hardly be found in the literature. This study aims to model extreme wave height

Table 1 Estimations of coefficients in Example 1

Para.	True value	Estimated value	Para.	True value	Estimated value	Para. Para.	True value	Estimated value
b_{11}	0.0271	0.0305 (0.0253)	λ_{11}	1.2000	1.1996 (0.2818)	ϕ_{11}	1.2000	1.1995 (0.0805)
b_{21}	0.1667	0.1355 (0.0820)	λ_{21}	0.5000	0.5265 (0.2028)	ϕ_{21}	0.5000	0.5265 (0.0786)
b_{12}	0.1285	0.1339 (0.0397)				ϕ_{12}	1.1000	1.1000 (0.0646)
b_{22}	0.1111	0.0942 (0.0424)				ϕ_{22}	0.5000	0.5000 (0.2577)
b_{13}	0.1667	0.1493 (0.0863)	λ_{13}	0.5000	0.5000 (0.0592)	ϕ_{13}	0.6000	0.6000 (0.1208)
b_{23}	0.0407	0.0454 (0.0080)	λ_{23}	1.2000	1.1991 (0.0701)	ϕ_{23}	1.2000	1.1983 (0.3507)

Estimations and standard deviations—in the parentheses—are computed by using Theorem 3. The values of $\widehat{q}_d(x_{jd}), \widehat{q}_{1d}(y_{jd})$ are computed from 10,000 simulated observations. The tuning parameter values used are: {0.2079, 0.3816, 0.5905, 0.7712, 0.9237, 1.000, 1.1097, 1.3292, 1.9811, 3.0653, 4.5093} (in the first process); {0.6198, 1.6198, 3.2066, 5.0000 } (in the second process); {0.1875, 0.3616, 0.5848, 0.7306, 0.7991, 0.8752, 0.9588, 1.1099, 1.3285, 1.7728, 2.4428, 3.4722 } (in the third process); {0.0432, 0.0903, 0.1886, 0.3218, 0.4607, 0.7325, 1.5843, 3.8008, 9.1183 } (between the first process and the second process); {0.3455, 0.3937, 0.4487, 0.5194, 0.5609, 0.5611, 0.5613, 0.6840, 0.8070, 0.8078 0.8086, 1.3123, 1.8636, 1.9622, 2.0660} (between the first process and the third process)

co-movements at two different locations on the North Sea. Data (hourly) was made available by the Dutch Ministry of Transport, Public Works and Water Management (RIKZ). Data was recorded from January 11, 1979 to October 28, 2002. In Fig. 2, time series plots of observations at two different locations, Station Eierlandse gat (Wadden) (ELD) and Station Europlatform (EUR), are in the upper panel and in the middle panel respectively, and a bivariate scatter plot of ELD against EUR is shown in the lower panel. Both time series look stationary. The lower panel suggests that there exist extreme dependencies (in the upper right corner) between these two series.

Figure 3 shows time series plots within certain time periods (in hour). In each plot, the solid line curve is for data observed at ELD, and the dashed dotted curve is for data observed at EUR. In the upper-left panel in Fig. 3, we plot observations within a 100-hour time period from each station. These observations were recorded at different starting time from ELD and EUR. The curve for EUR was observed 10 h later than the curve for ELD. Curves in the upper-right panel were observed at the same time period at both stations. So were curves in the lower-left panel. We can clearly see that wave height moving patterns are very similar between these two stations. It may be appropriate to suggest that there exist extreme height co-movements between waves at different locations on North Sea. Also notice that in the lower-right panel, the curves for lower height waves show different moving patterns between the two locations. These phenomena suggest that an M4 process model may be a good choice to model data like this kind after certain data transformation.
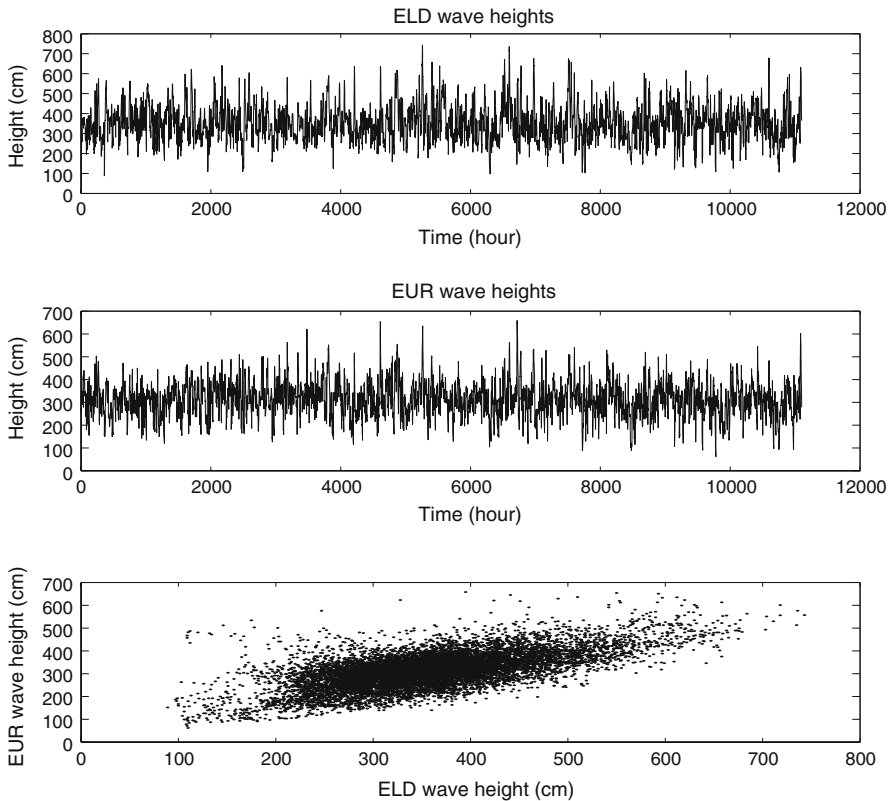
Fig. 2 The upper panel time series plot is for observations at ELD. The middle panel time series plot is for observations at EUR. The lower panel is a scatter plot of ELD against EUR

6.2 Data transformation

As mentioned in the previous sections, in order to model real data by an M4 process, the data must be transformed to have exact unit Fréchet distribution or to asymptotically be unit Fréchet distributed. In the literature, GEV or GPD is often used to fit the original data first, and the data is then transformed based on the fitted distribution functions. Here, we simply apply a rank based transformation procedure. Suppose the rank of U_i is r_i, then the transformed data value is $X_i = -1/\log\left(r_i/(n+1)\right)$. We plot transformed data in Fig. 4. Now the transformed moving patterns and those in Fig. 1 look alike, and hence an M4 model with geometric moving patterns to approximate these transformed data seems acceptable.

6.3 Modeling the transformed data

As mentioned in Sect. 4, parameters a_{lkd}, tuning parameters L_d, $K_1(l,d)$ and $K_2(l,d)$ can be estimated by using the dth observed process only. The idea
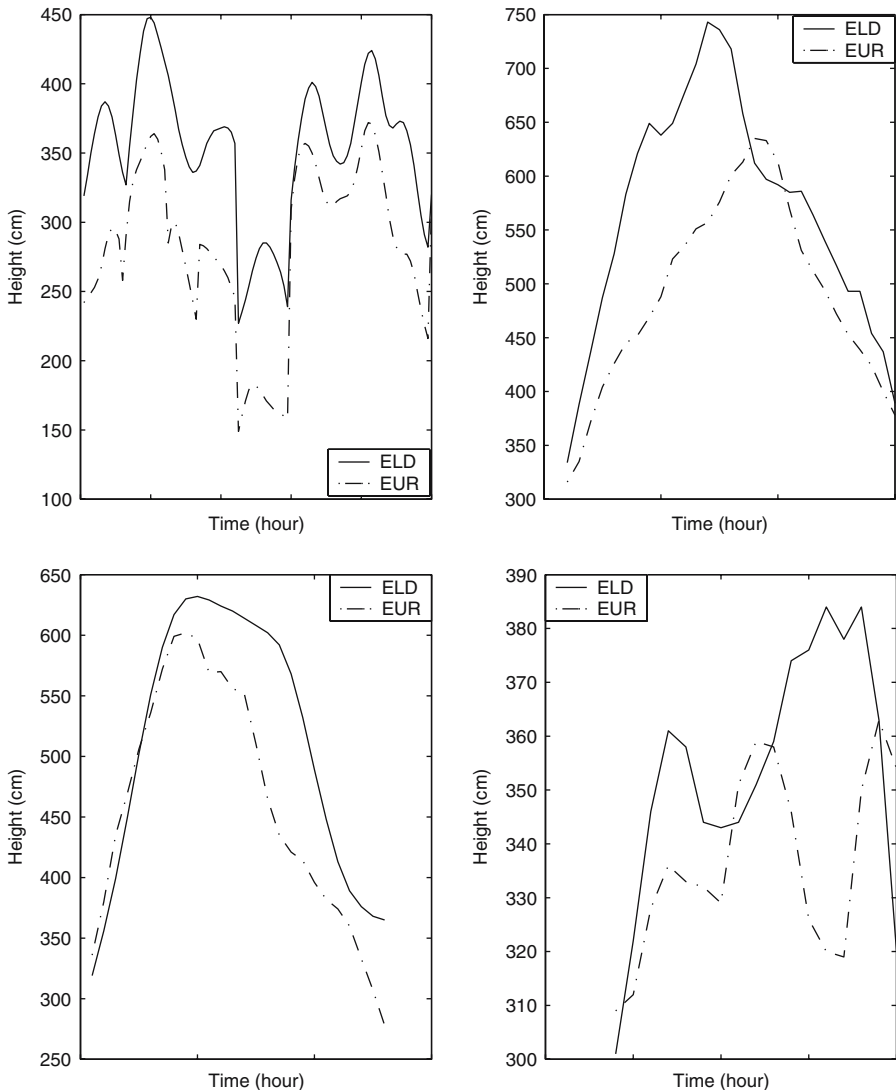
Fig. 3 The upper-left panel plots observations within a 100-hour time period. The dashed dotted curve (EUR) was observed 10 hours later than the solid curve (ELD). The upper-right panel, the lower-left panel, and the lower-right panel, each plots observations within a 35-hour time period. Observations in these three plots are recorded at the same time at both stations

of using functions $q_{1d}(x), d = 2, \ldots, D$, is to match extreme co-movement signature patterns between different processes. Here we first estimate parameter values based on $q_d(x)$ and $\widehat{q}_d(x)$ functions and the procedure proposed in Sect. 5.

For both ELD transformed data and EUR transformed data, after applying the procedure proposed in Sect. 5, we find that both processes are best
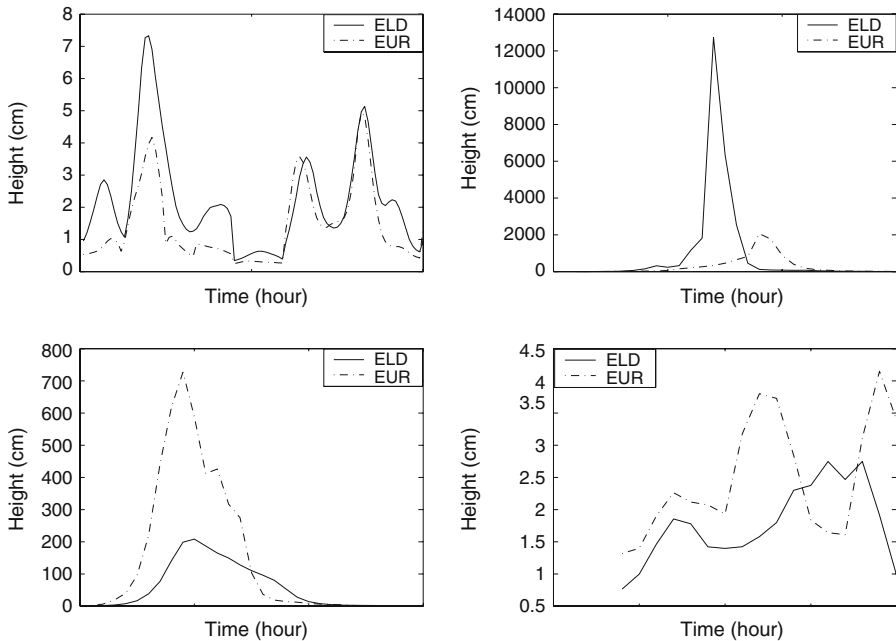
Fig. 4 Plots of transformed data which is asymptotically unit Fréchet distributed. The upper-left panel plots observations within a 100-hour time period. The dashed dotted curve (EUR) was observed 10 hours later than the solid curve (ELD). The upper-right panel, the lower-left panel, and the lower-right panel, each plots observations within a 35-hour time period. Observations in these three plots are recorded at the same time at both stations

Table 2 Estimations of coefficients in ELD transformed data model and EUR transformed data model separately

		Estimated value		Estimated value		Estimated value
ELD	b_{11}	0.1226 (0.0158)	λ_{11}	0.9184 (0.0356)	ϕ_{11}	0.9397 (0.0421)
EUR	b_{12}	0.1416 (0.0201)	λ_{12}	0.8646 (0.0385)	ϕ_{12}	0.8815 (0.0528)

Estimations and standard deviations—in the parentheses—are computed by using Theorem 3

fitted with one signature pattern, i.e. $L_d = 1$, and $K_1 = 3$, $K_2 = 6$, and so $K = K_1 + K_2 + 1 = 10$ (hours) which matches the upper-left panels in Figs. 3 and 4. The tuning parameter values x_{id} used in ELD are 0.6326, 0.8770, 0.9439, 1.0161, 1.0936, 1.4155; and the tuning parameter values x_{id} used in EUR are 0.5606, 0.8058, 0.9225, 1.0573, 1.2104, 1.6087. These values are automatically computed using the procedure in Sect. 5. The estimated parameter value and their standard errors are reported in Table 2.

As illustrated in (20) and (21), the joint distributions between two univariate processes depend on the starting locations of signature patterns in each processes. We use functions $q_{1d}(x)$ (or their empirical counterparts $\widehat{q}_{1d}(x)$) to form

Table 3 Estimations of coefficients in ELD transformed data and EUR transformed data jointly

d	l	Estimated value ($\widehat{b}_{ld}$)	Estimated value ($\widehat{\lambda}_{ld}$)	Estimated value ($\widehat{\phi}_{ld}$)	$(h(l,d))$
1	1	0.0164 (0.0021)	0.9184 (0.0356)	0.9397 (0.0421)	0
1	2	0 0	0.9184 (0.0356)	0.9397 (0.0421)	0
1	3	0.0320 (0.0041)	0.9184 (0.0356)	0.9397 (0.0421)	0
1	4	0.0107 (0.0014)	0.9184 (0.0356)	0.9397 (0.0421)	0
1	5	0.0289 (0.0037)	0.9184 (0.0356)	0.9397 (0.0421)	0
1	6	0.0347 (0.0045)	0.9184 (0.0356)	0.9397 (0.0421)	0
2	1	0 0	0.8646 (0.0385)	0.8815 (0.0528)	0
2	2	0.0638 (0.0091)	0.8646 (0.0385)	0.8815 (0.0528)	0
2	3	0.0216 (0.0031)	0.8646 (0.0385)	0.8815 (0.0528)	0
2	4	0.0264 (0.0037)	0.8646 (0.0385)	0.8815 (0.0528)	0
2	5	0.0128 (0.0018)	0.8646 (0.0385)	0.8815 (0.0528)	2
2	6	0.0170 (0.0024)	0.8646 (0.0385)	0.8815 (0.0528)	2

Estimations and standard deviations—in the parentheses—are computed by using Theorem 3

an M4 process model for both ELD and EUR transformed data. We first give a fact that the following two processes are identical in distribution:

$$Y_i = \max_{k=-K_1,\ldots,K_2} b\lambda^{(k)+}\phi^{(k)-} Z_{i-k},$$

and

$$Y_i' = \max_{l=1,\ldots,L;\, k=-K_1,\ldots,K_2} b_l\lambda^{(k)+}\phi^{(k)-} Z_{l,i-k},$$

where $b = b_1 + \cdots + b_L$, $b_l > 0$, Z_{i-k}, $Z_{l,i-k}$ are independent unit Fréchet random variables. In this analysis, if we let $L_d > 1$, then we have $\lambda_{id} = \lambda_{jd}$, $\phi_{id} = \phi_{jd}$, $i \neq j$ which do not satisfy the conditions in Lemma 5. This is not a problem since the values of $\lambda_{id} = \lambda_{jd}$, $\phi_{id} = \phi_{jd}$, $i \neq j$ can be uniquely determined by functions $q_d(x)$, while the values of b_{ld}s can be uniquely determined by functions $q_{1d}(x)$, $d = 1,\ldots,D$ in this study.

We now let $d = 1$ correspond to ELD, and $d = 2$ correspond to EUR. In order to use Step 11 of Sect. 5, we have tried different values of L_d, $d = 1,2$. The values vary from 1 to 20. Using tuning parameter y_j values 0.4177, 0.5320, 1.7039, 1.9749, 2.2374, the fitted squared error is 2.4158 when $L_1 = L_2 = 1$,

while the fitted squared error is 7.8×10^{-4} when $L_1 = 6$ and $L_2 = 6$. This is the smallest fitted squared error. Summarizing these results, we obtain Table 3.

In Table 3, $\widehat{b}_{21} = 0$ means that the extreme values generated from $\{Z_{2i}\}$ processes only affect wave heights at EUR, not at ELD; similarly, the extreme values generated from $\{Z_{1i}\}$ processes only affect wave heights at ELD, not at EUR. Except these two cases, the extreme values generated from $\{Z_{li}\}$, $l = 3, \ldots, 6$, have effects at both locations. Numbers 2 in the last column show that some peak values of extreme heights at EUR are two hours later than those at ELD. $\widehat{\lambda}_{ld}$ values are repeated in Column 4 due to the estimates are based on a univariate process. So are the $\widehat{\phi}_{ld}$ values.

After a joint model has been established for wave heights at two different locations, we can study various statistical inference problems regarding coastal engineering and marine operations, etc. However these inference problems are beyond the scope of this current study. We put them as our future research directions.

7 Conclusions

M4 processes are built for modeling the extremes of multivariate stationary processes. There are potential applications in finance, insurance and environmental science. Some preliminary financial applications have been conducted in Zhang (2002, 2005), Zhang and Smith (2001, 2002), and Smith (2003). The results of Sect. 3, especially Lemma 3, are very general and can be used in any other case where the asymptotic covariance matrix needs to be obtained. In this paper, we have developed estimating procedures for M4 processes with asymmetric geometric moving patterns of extremes. We have shown the model identifiability (Lemmas 4, 5). The asymptotic properties are proved for all the proposed estimators (Theorems 1, 2, 3).

We have restricted our attention to the asymmetric geometric structures which are also commonly considered in other research areas. We have successfully applied M4 models to wave height data. There is no doubt that other structures are possible. Directions for future research include: M4 models with missing data, and selecting models under the M4 model settings.

8 Proofs

Proof of Theorem 2 For simplicity, we prove the results for $L = 1$. The proof for $L > 1$ is a direct generalization of this proof. Since $\widehat{q}(x) \xrightarrow{a.s.} q(x)$, by continuous mapping theorem,

$$\widehat{b} \xrightarrow{a.s.} b, \quad \widehat{\lambda} \xrightarrow{a.s.} \lambda, \quad \widehat{\phi} \xrightarrow{a.s.} \phi, \quad \text{as } n \to \infty.$$

Notice that for each i, $\widehat{q}(x_i)$ is a function of $(\widehat{b}, \widehat{\lambda}, \widehat{\phi})$. When $(\widehat{b}, \widehat{\lambda}, \widehat{\phi})$ are in a neighborhood of (b, λ, ϕ), $\widehat{q}(x_i)$ can be written as

$$\widehat{q}(x_i) = q(x_i) + \widehat{Q}_i(x_i) \left(\begin{bmatrix} \widehat{b} \\ \widehat{\lambda} \\ \widehat{\phi} \end{bmatrix} - \begin{bmatrix} b \\ \lambda \\ \phi \end{bmatrix} \right),$$

where $\widehat{Q}_i(x_i)$ is a row vector

$$\widehat{Q}_i(x_i) = \left(\frac{\partial \widehat{q}(x_i)}{\partial \widehat{b}}, \frac{\partial \widehat{q}(x_i)}{\partial \widehat{\lambda}}, \frac{\partial \widehat{q}(x_i)}{\partial \widehat{\phi}} \right).$$

The values of the vector are evaluated at the points where each $\widehat{b}$ is replaced by $\widehat{b} + \theta_n(\widehat{b} - b)$, $\widehat{\lambda}$ is replaced by $\widehat{\lambda} + \theta_n(\widehat{\lambda} - \lambda)$, and $\widehat{\phi}$ is replaced by $\widehat{\phi} + \theta_n(\widehat{\phi} - \phi)$, where θ_n is in (0,1). Putting all $\widehat{Q}_i(x_i), i = 1, \ldots, m$, row by row, we then have

$$\widehat{\mathbf{q}} - \mathbf{q} = \widehat{Q} \left(\begin{bmatrix} \widehat{b} \\ \widehat{\lambda} \\ \widehat{\phi} \end{bmatrix} - \begin{bmatrix} b \\ \lambda \\ \phi \end{bmatrix} \right),$$

where the ith row of matrix $\widehat{Q}$ is $\widehat{Q}_i(x_i)$. The proof is completed by applying Lemma 3, the Mean Value Theorem and Slutsky theorem. □

Proof of Lemma 4 Since $q(x)$ is a piecewise linear function, as long as there are at least two points—say x_i and x_{i+1}—between any two adjacent slope changing points, then points $(x_i, q(x_i))$, $i = 1, \ldots, m$ uniquely determine all the values of slope change points. Since $\lambda_i \neq \lambda_j$, $\lambda_i \neq \phi_j$, $\phi_i \neq \phi_j$, $i \neq j$, $i,j = 1, \ldots, L$, any permutations among ϕ_j and λ_l will result in different values of λ_l/ϕ_l which are the true changing points, therefore the values of λ_l and ϕ_l are uniquely determined. To show the uniqueness of b_l, we rewrite $q(x)$ as

$$q(x) = \sum_{l=1}^{L} xe_l \sum_{m=1-K_2(l)}^{2+K_1(l)} \max \left(c_{l,1-m}, \frac{c_{l,2-m}}{x} \right). \tag{24}$$

where $e_l = b_l * \left(\sum_{k=1}^{K_1(l)} \phi_l^k + 1 + \sum_{k=1}^{K_2(l)} \lambda_l^k \right)$, $\sum_j c_{l,j} = 1$ for each l and all $c_{l,j}$ are uniquely determine by the ratios which are the slope change points of $q(x)$.

Suppose now $q(x)$ has a different representation, say

$$q(x) = \sum_{l=1}^{L} xe_l' \sum_{m=1-K_2(l)}^{2+K_1(l)} \max \left(c_{l,1-m}, \frac{c_{l,2-m}}{x} \right) \tag{25}$$

then

$$\sum_{l=1}^{L}(e_l - e'_l) \sum_{m=1-K_2(l)}^{2+K_1(l)} \max\left(c_{l,1-m}, \frac{c_{l,2-m}}{x}\right) = 0 \tag{26}$$

for all $x > 0$.

Suppose we have chosen $x_1, x_2, \ldots, x_{L-1}$ and formed the matrix

$$\Delta = \begin{bmatrix} \sum_{m=1-K_2(l)}^{2+K_1(l)} \max\left(c_{1,1-m}, \frac{c_{1,2-m}}{x_1}\right) & \cdots & \sum_{m=1-K_2(l)}^{2+K_1(l)} \max\left(c_{L,1-m}, \frac{c_{L,2-m}}{x_1}\right) \\ \vdots & \ddots & \vdots \\ \sum_{m=1-K_2(l)}^{2+K_1(l)} \max\left(c_{1,1-m}, \frac{c_{1,2-m}}{x_{L-1}}\right) & \cdots & \sum_{m=1-K_2(l)}^{2+K_1(l)} \max\left(c_{L,1-m}, \frac{c_{L,2-m}}{x_{L-1}}\right) \\ 1 & \cdots & 1 \end{bmatrix}$$

and set

$$\begin{bmatrix} \sum_{m=1-K_2(l)}^{2+K_1(l)} \max\left(c_{1,1-m}, \frac{c_{1,2-m}}{x_1}\right) & \cdots & \sum_{m=1-K_2(l)}^{2+K_1(l)} \max\left(c_{L,1-m}, \frac{c_{L,2-m}}{x_1}\right) \\ \vdots & \ddots & \vdots \\ \sum_{m=1-K_2(l)}^{2+K_1(l)} \max\left(c_{1,1-m}, \frac{c_{1,2-m}}{x_{L-1}}\right) & \cdots & \sum_{m=1-K_2(l)}^{2+K_1(l)} \max\left(c_{L,1-m}, \frac{c_{L,2-m}}{x_{L-1}}\right) \\ 1 & \cdots & 1 \end{bmatrix} \begin{pmatrix} b_1 - b'_1 \\ \vdots \\ \vdots \\ b_L - b'_L \end{pmatrix} = 0.$$

Let $|\Delta|$ be the determinant of the system of linear equations. Assume now the L determinants of the $(L-1) \times (L-1)$ matrices formed from the bottom $L-1$ rows are not all zero. Since $c_{l,k}$ are known and $\sum_{m=1-K_2(l)}^{2+K_1(l)} c_{l,i-m} = 1$, $i = 1, 2$, then there exist $x_{\min}$ and $x_{\max}$ such that when $x_1 < x_{\min}$ or $x_1 > x_{\max}$, all elements of first row in Δ are $\frac{1}{x_1}$ or 1, respectively. This will give two constant rows in $|\Delta|$, so when $x_1 < x_{\min}$ or $x_1 > x_{\max}$, we have $|\Delta| = 0$. When x_1 varies in $[x_{\min}, x_{\max}]$, denoting Δ by $\Delta(x_1)$, then

$$|\Delta(x_1)| = \frac{1}{x_1}\sum c_{i,j}|\Delta|_{1j} + \sum c_{i',j'}|\Delta|_{1j'} \tag{27}$$

where $|\Delta|_{1j} \neq 0$, $|\Delta|_{1j'} \neq 0$ are the $(1,j)$ or $(1,j')$ minors of Δ. Both summations in the right hand side of (27) are over all non-zero minors of the first row of Δ and the corresponding $\frac{c_{i,j}}{x_1}$ or $c_{i',j'}$. If $|\Delta(x_1)| = 0$, by varying x_1 in $[x_{\min}, x_{\max}]$, at some point x, some $\frac{1}{x_1}c_{i,j}|\Delta|_{1j}$ of the summation $\frac{1}{x_1}\sum c_{i,j}|\Delta|_{1j}$ change to $c_{i',j'}|\Delta|_{1j'}$ and add to $\sum c_{i',j'}|\Delta|_{1j'}$, or *vice versa*, and this change results in $|\Delta(x)| \neq 0$. Hence it cannot be true that $|\Delta| = 0$ for all x_1. This argument can be applied to lower dimension matrices. On the other hand, we can start from a 2×2 matrix and extend it to $L \times L$ matrix such that the determinant is not

zero as required. Therefore, there exist constants $x_1, , \ldots, x_{L-1}$ such that each system of linear equations has a unique solution. We then conclude $e_l = e'_l$, for all l. So $q(x)$ uniquely determine all b_l, λ_l and ϕ_l. $\square$

Proof of Lemma 5 By Lemma 4, $q_d(x)$, $d = 1, \ldots, D$ uniquely determine all values of parameters λ_{ld}, b_{ld}, ϕ_{ld}. Since ratios $a_{l,k,1}/a_{l',k'+h(l',d'),d'}$, $l, l' = 1, \ldots, L$, $k, k' = -K_1, \ldots, K_2$, for each d' are distinct, any permutation of index l in $b_{ld'} \lambda_{ld'}^{(k)+} \phi_{ld'}^{(k)-}$ will result in different ratios or jump points on the right hand side of (22), so $q_{1d'}(x)$ uniquely determines $a_{l,k,1}/a_{l,k+h(l,d'),d'}$, $d' = 2, \ldots, D$ for all l and k. So (22) eventually uniquely determines all the true values of all parameters λ_{ld}, b_{ld}, ϕ_{ld} and identifies the model. The reason why $x_{1d}, x_{2d}, \ldots, x_{m_d d}$, $y_{1d'}, y_{2d'}, \ldots, y_{v_{d'} d'}$ uniquely determine all values of λ_{ld}, b_{ld}, ϕ_{ld} is because $q_d(x)$, $q_{1d'}(x)$ are piecewise linear functions which can be uniquely determined by a finite number of points as long as there are at least two points between any two jump points. $\square$

Proof of Theorem 3 Simply notice that as $n \to \infty$,

$$\widehat{q}_d(x_{jd}) \xrightarrow{a.s.} q_d(x_{jd}), \quad j = 1, \ldots, m_d, \ d = 1, \ldots, D,$$
$$\widehat{q}_{1d'}(y_{j'd'}) \xrightarrow{a.s.} q_{1d'}(y_{j'd'}), \quad j' = 1, \ldots, v_{d'}, \ d' = 2, \ldots, D,$$

and apply Lemma 5. Then follow the arguments before Theorem 2. $\square$

References

Billingsley, P. (1995). *Probability and measure* (3rd ed.). New York: Wiley.

Coles, S. G., Tawn, J. A. (1991). Modeling extreme multivariate events. *Journal of Royal Statistical Society, Series B, 53*, 377–392.

Coles, S. G., Tawn, J. A. (1994). Statistical methods for multivariate extrems: an application to structural design (with discussion). *Applied Statistics, 43*(1), 1–48.

Davis, R. A., Resnick, S. I. (1989). Basic properties and prediction of Max-ARMA processes. *Advances in Applied Probability, 21*, 781–803.

Davis, R. A., Resnick, S. I. (1993). Prediction of stationary Max-stable processes. *Annals of Applied Probability, 3*(2), 497–525.

Deheuvels, P. (1983). Point processes and multivariate extreme values. *Journal of Multivariate Analysis, 13*, 257–272.

Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling extremal events for insurance and finance*, Berlin Heidelberg New York: Springer.

Galambos, J. (1987). *Asymptotic theory of extreme order statistics* (2nd ed.). Malabar: Krieger.

de Haan, L. (1984). A Spectral Representation for Max-stable processes. *Annals of Probability 12*(4), 1194–1204.

de Haan, L. (1985). Extremes in higher dimensions: the model and some statistics. In *Proceedings of 45th session international statistics institute*, (paper 26.3). The Hague: International Statistical Institute.

de Haan, L., Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie verw Gebiete*, *40*, 317–337.

de Haan, L., de Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes*, *1*(1), 7–45.

Hall, P., Peng, L., Yao. Q. (2002). Moving-maximum models for extrema of time series. *Journal of Statistical Planning and Inference*, *103*, 51–63.

Leadbetter, M. R., Lindgren, G., Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Berlin Heidelberg New York:Springer.

Pickands, J. III (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131.

Pickands, J. III (1981). Multivariate Extreme Value distributions. *In Proceedings of 43rd session international statistics institute*, pp. (859–878). Buenos Aires.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*. Berlin Heidelberg New York:Springer.

Smith, R. L. (1990). Extreme value theory. In W. Ledermann (Ed.), *Handbook of applicable mathematics (Supplement)*, Chichester:Wiley.

Smith, R. L. (2003). Statistics of extremes, with applications in the environment, insurance and finance. In B. Finkenstadt, H. Rootzén, (Eds.) *Extreme value in finance, telecommunications and the environment*, to be published by Chapman and Hall/CRC.

Smith, R. L., Weissman, I. (1996). Characterization and estimation of the multivariate extremal index. Manuscript, UNC.

Van Gelder, P. H. A. J. M., de Ronde, J. G., Neykov, N. M., Neytchev, P. (2000). Regional frequency analysis of extreme wave heights: Trading space for time. *In Proceedings of the 27th ICCE* (pp. 1099-1112, Vol. 2,) Coastal Engineering 2000, Sydney.

Zhang, Z. (2002). *Multivariate extremes, Max-stable process estimation and dynamic financial modeling*. PhD Dissertation, Department of Statistics, University of North Carolina.

Zhang, Z. (2005). A new class of tail-dependent time series models and its applications in financial time series. *Advances in Econometrics*, *20*(B), 323–358.

Zhang, Z., Smith, R. L. (2001). *Modeling financial time series data as moving maxima processes*. Technical Report, Department of Statistics, University of North Carolina. Submitted to 2001 NBER/NSF time series conference.

Zhang, Z., Smith, R. L. (2002). On the estimation and application of Max-stable processes. Manuscript, Washington University.

Zhang, Z., Smith, R. L. (2004). The behavior of multivariate maxima of moving maxima processes. *Journal of Applied Probability*, *41*, 1113–1123.