Maxima of moving maxima of continuous functions

Thomas Meinguet

Université catholique de Louvain, Institut de statistique
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
e-mail: thomas.meinguet@uclouvain.be

Abstract: Maxima of moving maxima of continuous functions (CM3) are
max-stable processes aimed at modeling extremes of continuous phenomena
over time. They are defined as Smith and Weissman’s M4 processes with
continuous functions rather than vectors. After standardization of the mar-
gins of the observed process into unit-Fréchet, CM3 processes can model
the remaining spatio-temporal dependence structure.

CM3 processes have the property of joint regular variation. The spectral
processes from this class admit particularly simple expressions. Furt her-
more, depending on the speed with which the parameter functions tend to-
ward zero, CM3 processes fulfill the finite-cluster condition and the strong
mixing condition. For instance, these three properties put together have
implications for the expression of the extremal index.

A method for fitting a CM3 to data is investigated. The first step is to
estimate the length of the temporal dependence. Then, by selecting a suit-
able number of blocks of extremes of this length, clustering algorithms are
used to estimate the total number of different profiles. The number of pa-
rameter functions to retrieve is equal to the product of these two numbers.
They are estimated thanks to the output of the partitioning algorithms in
the previous step. The full procedure only requires one parameter which is
the range of variation allowed among the different profiles. The dissimilarity
between the original CM3 and the estimated version is evaluated by means
of the Hausdorff distance between the graphs of the parameter functions.

AMS 2000 subject classifications: Primary 60G70; secondary 60G60.
Keywords and phrases: CM3, M4, extremes, clusters, spectral process,
extremal index.

1. Introduction

Maxima of moving maxima of continuous functions (CM3) are the analogue
of Smith and Weissman’s M4 processes [15] with continuous functions rather
than vectors. Let \( a_i^{(j)} \) (\( i \in \mathbb{Z}_+ \), \( j \in \mathbb{Z} \)) be strictly positive, real, continuous functions
on a compact domain of \( \mathbb{R}^q \), say \([0,1]^q\). The functions \( a_i^{(j)} \) are the parameter functions. They are assumed to satisfy, for every \( x \in [0,1]^q \), the equality

\footnote{Research supported by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy) and by contract nr. 07/12/002 of the Projet d’Actions de Recherche Concertées of the Communauté française de Belgique, granted by the Académie universitaire Louvain.}
A CM3 process \((X_t)_{t \in \mathbb{Z}}\) is defined by the expression

\[
X_t(x) = \sup_{j \in \mathbb{Z}} \sup_{i \geq 0} a^{(j)}_i(x) Z^{(j)}_{t-i} \quad (x \in [0,1]^q),
\]

where the innovations \(Z^{(j)}_i \ (i \in \mathbb{Z}, \ j \in \mathbb{Z})\) are independent and identically distributed unit-Fréchet random variables, i.e.

\[
P(Z_i \leq z) = \exp\left(-\frac{1}{z}\right) \quad \text{for } z > 0.
\]

The fact that, given real numbers \(\xi_i > 0 \ (i \in \mathbb{N})\) such that \(\xi_1 + \xi_2 + \ldots = 1\), the distribution of \(\max(\xi Z_1, \xi Z_2, \ldots)\) stays unit-Fréchet implies that \(X_t\) has unit-Fréchet margins. However the transformation from \((Z_t)_{t \in \mathbb{Z}}\) to \((X_t)_{t \in \mathbb{Z}}\) induces a dependence structure in time and space. Extremes appear in temporal clusters and, at time \(t\), a large value for \(X_t\) at location \(x\) causes large values at other locations. From this fact, CM3 processes are able to model a wide range of spatio-temporal dependences. The first part of this paper is a study of some properties: spectral process, strong mixing condition, finite-cluster condition and extremal index.

The second objective of this paper is to fit CM3 processes to samples with measurement errors. For that purpose, CM3 will be discretized into M4 of dimension \(D\) selecting \(D\) points \(x_d \ (1 \leq d \leq D)\) in the domain. It will be also assumed that \(0 \leq i < K\) and \(1 \leq j \leq L\) for finite constants \(K\) and \(L\). The practical model studied is thus

\[
X_t(d) = \max_{1 \leq j \leq L} \max_{0 \leq i < K} a^{(j)}_i(x_d) Z^{(j)}_{t-i} + \varepsilon_t(x_d) \quad (1 \leq d \leq D)
\]

where \(\varepsilon_t(x_d)\) are independent \(N(0, \sigma^2)\) random variables. The parameter \(K\) is the length of the temporal dependence and \(L\) is the total number of reproducible patterns that we can observe up to a multiplicative constant in the process.

Figure 1 shows a realization of a CM3 plotted versus a M4.

![Figure 1](image-url)

**Fig 1.** M4 with \(D = 5\) on the left, CM3 on \([0,1]\) on the right (\(K = 3, L = 2\)).

In Section 2, a coherent set of properties for CM3 is established. The motivation is similar as in [13, 14] but now for random continuous functions. Theorem 2.3 is the joint regular variation of those processes, a concept extended to
Banach spaces in [9]. The spectral process of a CM3 has a discrete distribution, given by the theorem. Next, depending on the speed with which the parameter functions $a^{(j)}_i$ tend toward zero, Theorem 2.4 yields the finite-cluster condition and Theorem 2.5 yields the strong mixing condition. These three properties together also have specific implications, for instance the inverse of the extremal index $\theta$ becomes the expected size of clusters of extremes in the sense of [12].

CM3 processes are also examples of max-stable random fields [1, 2]: for every finite space-time subset $A \times T \subset [0, 1]^q \times Z$, the random vector $(X_t(x))_{x \in A, t \in T}$ has a multivariate extreme value distribution. This property of M4 is inherent to CM3 since the law of a continuous random field is characterized by its finite dimensional distributions that are M4 according to Example 2.2.

Section 3 is a preparation for the estimation of the parameter functions. Extremes will play a central role in identifying the recursive patterns and their relative frequencies. So we need to study the probabilistic properties of the blocks of extremes that can be observed in CM3. The harmonic mean makes convenient the expressions of the frequencies of the reproducible patterns that can be observed.

In Section 4, we suggest and compare empirical methods to estimate $K$, $L$ and the parameter functions without assumptions on these objects. It is a complement to [23], where the case $L = 1$ has only been treated, and to [22], where assumptions are made on the parameters.

This study is designed to improve the statistical analyses of extreme events, as done in [17, 18] for instance.

2. Definition and properties

Choose a nonempty compact domain of $\mathbb{R}^q$. To not multiply the notations this compact will be taken to be $[0, 1]^q$. Given an array $Z_i^{(j)}$ ($i \in \mathbb{Z}_+, j \in \mathbb{Z}$) of independent unit-Fréchet random variables, if $a_i^{(j)} : [0, 1]^q \to \mathbb{R}^*_+$ are deterministic strictly positive continuous functions, a \textit{CM3 process} is a stochastic process defined by

$$X_t(x) = \sup_{j \in \mathbb{Z}} \sup_{i \geq 0} (a_i^{(j)}(x)Z_i^{(j)}).$$

If furthermore

$$(\forall x \in [0, 1]^q) : \sum_{j \in \mathbb{Z}} \sum_{i \geq 0} a_i^{(j)}(x) = 1$$

we say that $(X_t)_{t \in \mathbb{Z}}$ is a \textit{standard CM3 process}.

The first result is an imperative condition before any use of CM3 processes. Recall that the sup-norm of a function $f : [0, 1]^q \to \mathbb{R}$ is $\|f\|_\infty = \sup_{x \in [0, 1]^q} |f(x)|$ and that this supremum is achieved.

\textbf{Proposition 2.1.} If

$$\sum_{j \in \mathbb{Z}} \sum_{i \geq 0} \|a_i^{(j)}\|_\infty < \infty,$$

(2.2)
then, for every \( t \in \mathbb{Z} \), \( X_t \) in (2.1) is a random element in \( C([0, 1]^q, \mathbb{R}^+) \) and the process \( (X_t)_{t \in \mathbb{Z}} \) is stationary.

The proofs of the results of this section are relegated to Appendix A.

**Example 2.2.** If \( (X_t)_{t \in \mathbb{Z}} \) is a standard CM3 with \( 0 \leq i < K \) and \( 1 \leq j \leq L \), then for each \( x_1, \ldots, x_D \in \mathbb{R}^q \), the process \( (X_t(x_1), \ldots, X_t(x_D))_{t \in \mathbb{Z}} \) is a standard M4. Under (2.2) \( \|X_t\|_\infty \) is a non-standard M3. In both cases \( 0 \leq i < K \) and \( 1 \leq j \leq L \). This is helpful for the estimation of \( K \) and \( L \) if \( x_1, \ldots, x_D \) are far enough, see 4.2.

A CM3 process is an example of a jointly regularly varying time series. In particular, there exists a process \( (\Theta_t)_{t \in \mathbb{Z}} \) in \( C([0, 1]^q, \mathbb{R}^+) \), called spectral process which is the limit in distribution, as \( x \to \infty \),

\[
\mathcal{L}\left(\frac{X_t}{\|X_0\|_\infty} \mid \|X_0\| \geq x\right) \xrightarrow{d} \mathcal{L}\left(\langle \Theta_t \rangle_{t \in \mathbb{Z}}\right)
\]

in the proper product space. According to [9], this process captures all aspects of extremal dependence, both within space and over time.

**Theorem 2.3.** Setting \( a^{(j)}_i = 0 \) if \( i < 0 \), under condition (2.2), a CM3 process \( (X_t)_{t \in \mathbb{Z}} \) is jointly regularly varying with index \( \alpha = 1 \) and spectral process

\[
\langle \Theta_{-s}, \ldots, \Theta_t \rangle \overset{d}{=} \left(\frac{a^{(j)}_{s+t}}{\|a^{(j)}_i\|_\infty}, \ldots, \frac{a^{(j)}_t}{\|a^{(j)}_i\|_\infty}\right),
\]

where \( (I, J) \) is a random vector on \( \mathbb{Z}_+ \times \mathbb{Z} \) having distribution

\[
P[I, J = (i, j)] = \frac{\|a^{(j)}_i\|_\infty}{\sum_{l \in \mathbb{Z}} \sum_{k \geq 0} \|a^{(j)}_k\|_\infty}, \ i \in \mathbb{Z}_+, \ j \in \mathbb{Z}.
\]

All CM3 processes satisfying (2.2) also satisfy the finite-cluster condition. This property prevents a sequence of extremes occurring in a CM3 from being infinite over time even if \( K = +\infty \) or \( L = +\infty \).

**Theorem 2.4.** Under condition (2.2), a CM3 process \( (X_t)_{t \in \mathbb{Z}} \) satisfies the finite-cluster condition: there exists \( (r_n)_{n \in \mathbb{N}} \) with \( r_n \to \infty \) and \( r_n/n \to 0 \) such that

\[
\lim_{n \to \infty} \limsup_{m \to \infty} P\left( \max_{m \leq |t| \leq r_n} \|X_t\|_\infty > n \mid \|X_0\|_\infty > n \right) = 0. \quad \text{(C)}
\]

Together with the finite-cluster condition, the strong mixing property leads to nice properties. To obtain the strong mixing property a sufficient condition is

\[
\sum_{j \in \mathbb{Z}} \sum_{i \geq 0} i \|a^{(j)}_i\|_\infty < \infty. \quad \text{(2.3)}
\]

Note that (2.2) and (2.3) are trivial whenever \( K = +\infty \) and \( L < +\infty \).
**Theorem 2.5.** Under condition (2.3), a CM3 process \((X_t)_{t \in \mathbb{Z}}\) satisfies the strong mixing condition:

\[
\lim_{m \to \infty} \sup_{A \in \sigma(-\infty,-m)} \left| \frac{P(A \cap B)}{P(A)P(B)} - 1 \right| = 0
\]

where \(\sigma(r,s)\) is the \(\sigma\)-field generated by \(\{X_t \mid r \leq t \leq s\}\).

If \((X_t)_{t \in \mathbb{Z}}\) is a regularly varying time series with index 1, the extremal index \(\theta\) of the univariate time series \((\|X_t\|_{\infty})_{t \in \mathbb{Z}}\) is defined as the quantity between 0 and 1 such that

\[
P(\max_{1 \leq t \leq n} \|X_t\|_{\infty} \leq nx) \to e^{-\theta/x}
\]
as \(n \to \infty\). The extremal index of a CM3 process is the following.

**Proposition 2.6.** Under condition (2.2), if \((X_t)_{t \in \mathbb{Z}}\) is a CM3 process, the extremal index of \((\|X_t\|_{\infty})_{t \in \mathbb{Z}}\) is

\[
\theta = \sum_{j \in \mathbb{Z}} \sum_{i \geq 0} \|a_i^{(j)}\|_{\infty} \sum_{j \in \mathbb{Z}} \max_{i \geq 0} \|a_i^{(j)}\|_{\infty}.
\]

Once conditions (C) and (M) are satisfied, which is the case under (2.3) by Theorem 2.4 and Theorem 2.5, there are further characterizations of the extremal index such that

\[
\theta = \lim_{t \to \infty} \lim_{x \to \infty} P(\max_{i=1,\ldots,t} \|X_i\|_{\infty} \leq x \mid \|X_0\|_{\infty} > x)
\]
and, in this case, \(1/\theta\) is the expected size of clusters of extremes in the sense of [12], which is recalled as follows. Let \(u_n \to \infty\) be a thresholding sequence and \(r_n \to \infty\) be such that the expected number of exceedances in a sample of size \(r_n\) tends toward 0:

\[
E \left[ \sum_{i=1}^{r_n} 1\{\|X_i\|_{\infty} > u_n\} \right] = r_n P(\|X_1\|_{\infty} > u_n) \to 0.
\]

Then, denoting \(M_n := \max\{\|X_1\|_{\infty}, \ldots, \|X_n\|_{\infty}\}\), under (C) and (M), we have

\[
E \left[ \sum_{i=1}^{r_n} 1\{\|X_i\|_{\infty} > u_n\} \mid M_n > u_n \right] = \frac{r_n P(\|X_1\|_{\infty} > u_n)}{P(M_n > u_n)} \to \frac{1}{\theta}
\]
as \(n \to \infty\).

3. Block profiles

In this section we study further probabilistic features of CM3 processes in order to build a method to estimate the parameter functions \(a_i^{(j)}\) in the case \(0 \leq i < K\) and \(1 \leq j \leq L\). The theoretical model for the rest of the paper is thus

\[
X_t(x) = \max_{1 \leq j \leq L} \max_{0 \leq i < K} a_i^{(j)}(x)Z_{t-i}^{(j)} \quad (x \in [0,1]^q)
\]

(3.1)
The estimation method suggested in this paper is based on the fact that a large value of $Z_{i}^{(j)}$ causes large values of $X_{t}$ for $i \leq t < i + K$ and the possibility to have in this case

$$(X_{t}, \ldots, X_{t+K-1}) = Z_{i}^{(j)}(a_{0}^{(j)}, \ldots, a_{K-1}^{(j)}).$$

(3.2)

By “block profile" we mean a sequence $(X_{t}, \ldots, X_{t+K-1})$ satisfying (3.2) for some $1 \leq j \leq L$. The corresponding sequence of functions $(a_{0}^{(j)}, \ldots, a_{K-1}^{(j)})$ will be called “profile" or “pattern”.

In 3.1 we compute the probability of the events (3.2) and their frequencies of occurrence for the different values of $j$. In 3.2 we have a brief look at the correlation between all the possible blocks of length $K$ available in a sample. They are not independent if they overlap. In 3.3 we give the needed sample size to expect that an event of the form (3.2) realizes at least once.

To compute the exact values, the knowledge of the parameter functions is needed, which is particular not the case in the estimation. This is the reason why we also give lower and upper bounds for the true values. These bounds only depend on a unique parameter $C$, which is the maximal variation among the parameter functions $a_{i}^{(j)}$.

3.1. Relative frequencies

To recover the functions $a_{i}^{(j)}$ from a sample of size $T$, the first step is to understand how (3.1) works. Consider as an example a simple situation when $K = 3$ and $L = 2$. A finite number of functions $a_{i}^{(j)}$ is uniformly bounded below by a positive constant since they are strictly positive. Thus, if for instance $Z_{i}^{(2)}$ is large enough, the value of $(X_{1}, \ldots, X_{K})$ at a given position $x \in [0, 1]^{9}$ is

$$X_{1}(x) = \max \left( \begin{array}{ccc} a_{1}^{(1)}(x)Z_{1}^{(1)} & a_{1}^{(1)}(x)Z_{0}^{(1)} & a_{0}^{(1)}(x)Z_{1}^{(1)} \\ a_{2}^{(2)}(x)Z_{2}^{(2)} & a_{1}^{(2)}(x)Z_{0}^{(2)} & a_{0}^{(2)}(x)Z_{1}^{(2)} \end{array} \right) = a_{0}^{(2)}(x)Z_{1}^{(2)}$$

$$X_{2}(x) = \max \left( \begin{array}{ccc} a_{1}^{(1)}(x)Z_{0}^{(1)} & a_{1}^{(1)}(x)Z_{1}^{(1)} & a_{0}^{(1)}(x)Z_{2}^{(1)} \\ a_{2}^{(2)}(x)Z_{0}^{(2)} & a_{1}^{(2)}(x)Z_{2}^{(2)} & a_{0}^{(2)}(x)Z_{2}^{(2)} \end{array} \right) = a_{1}^{(2)}(x)Z_{1}^{(2)}$$

$$X_{3}(x) = \max \left( \begin{array}{ccc} a_{1}^{(1)}(x)Z_{1}^{(1)} & a_{1}^{(1)}(x)Z_{2}^{(1)} & a_{0}^{(1)}(x)Z_{3}^{(1)} \\ a_{2}^{(2)}(x)Z_{1}^{(2)} & a_{1}^{(2)}(x)Z_{2}^{(2)} & a_{0}^{(2)}(x)Z_{3}^{(2)} \end{array} \right) = a_{2}^{(2)}(x)Z_{1}^{(2)}$$

(3.3)

so that the second pattern appears:

$$(X_{1}, X_{2}, X_{3})(x) = (a_{0}^{(2)}, a_{1}^{(2)}, a_{2}^{(2)})(x)Z_{1}^{(2)}.$$
How likely is this kind of events to occur? To compute their probabilities, first remark that (3.3) is equivalent to

\[
\begin{align*}
Z_{-1}^{(1)} &\leq \min_{x} \frac{a_0^{(2)}(x)}{a_2^{(1)}(x)} Z_1^{(2)} && Z_{-1}^{(2)} \leq \min_{x} \frac{a_0^{(2)}(x)}{a_2^{(1)}(x)} Z_1^{(2)} \\
Z_0^{(1)} &\leq \min_{x} \frac{a_0^{(1)}(x)}{a_1^{(1)}(x)} Z_1^{(2)} && Z_0^{(2)} \leq \min_{x} \frac{a_0^{(2)}(x)}{a_1^{(1)}(x)} Z_1^{(2)} \\
Z_1^{(1)} &\leq \min_{x} \frac{a_0^{(1)}(x)}{a_1^{(1)}(x)} \frac{a_2^{(2)}(x)}{a_2^{(2)}(x)} Z_1^{(2)} \\
Z_2^{(1)} &\leq \min_{x} \frac{a_0^{(1)}(x)}{a_1^{(1)}(x)} \frac{a_2^{(2)}(x)}{a_2^{(2)}(x)} Z_1^{(2)} \\
Z_3^{(1)} &\leq \min_{x} \frac{a_0^{(1)}(x)}{a_1^{(1)}(x)} Z_1^{(2)} \\
\end{align*}
\]

For the general case, let \( A(l^*) \) be the event (3.2) with fixed \( j = l^* \):

\[ A(l^*) = \{ \forall x \in [0, 1]^q : (X_t, \ldots, X_{t+K-1})(x) = Z_t^{(l^*)}(a_0^{(l^*)}, \ldots, a_{K-1}^{(l^*)})(x) \}, \]

i.e. \( A(l^*) \) is the event for a \( K \)-block starting a time \( t \) to be a block profile of type \( l^* \). Generalizing (3.4) shows that the event \( A(l^*) \) is the intersection of \((2K-1)L - 1 \) conditions involving the random variables \( Z_t^{(j)} \) for \( t - K + 1 \leq i \leq t + K - 1 \). Remembering that the density of \( Z \) is \( f_Z(z) = z^{-2} \exp(-z^{-1}) \), the probability \( p^{(l^*)} \) of \( A(l^*) \) is

\[
P^{(l^*)} = \int_0^\infty \prod_{l=1}^L \prod_{t=1}^L P(Z \leq \min_x \frac{a_0^{(l^*)}(x)}{a_{K-1}^{(l^*)}(x)} |dz)
\]

\[
\prod_{l=1}^L \prod_{t=1}^L P(Z \leq \min_x \frac{a_0^{(l^*)}(x)}{a_{K-2}^{(l^*)}(x)} \frac{a_1^{(l^*)}(x)}{a_{K-1}^{(l^*)}(x)} |dz)
\]

\[
\ldots
\]

\[
\prod_{l=1}^L \prod_{t=1}^L P(Z \leq \min_x \frac{a_0^{(l^*)}(x)}{a_0^{(l^*)}(x)} \frac{a_{K-2}^{(l^*)}(x)}{a_{K-1}^{(l^*)}(x)} |dz)
\]

\[
\prod_{l=1}^L \prod_{t=1}^L P(Z \leq \min_x \frac{a_0^{(l^*)}(x)}{a_0^{(l^*)}(x)} \frac{a_{K-1}^{(l^*)}(x)}{a_{K-1}^{(l^*)}(x)} |dz)
\]

\[
\prod_{l=1}^L \prod_{t=1}^L P(Z \leq \min_x \frac{a_0^{(l^*)}(x)}{a_0^{(l^*)}(x)} |dz) z^{-2} \exp(-z^{-1}) dz
\]
Denoting $m_k^{(l^*; l^*)}$ the minimum written on line $k$ of (3.5) and by extension $m_k^{(l^*; l^*)} = 1$, the value of $p^{(l^*)}$ is

$$p^{(l^*)} = \frac{1}{1 + \sum_{k=1}^{2K-1} \sum_{l=1}^{L} \frac{1}{m_k^{(l^*; l^*)}}} = \frac{\text{harm}(m_1^{(1; l^*)}, \ldots, m_{2K-1}^{(L; l^*)})}{(2K - 1)L}$$

(3.6)

where harm is the harmonic mean of the $(2K - 1)L$ minima.

It is thus possible to compute $p^{(l^*)}$ exactly given the parameter functions. If the parameter functions $a_i^{(j)}$ satisfy

$$\frac{1}{C^{(l^*)}} \leq \frac{a_k^{(l^*)}(x)}{a_k^{(l^*)}(x)} \leq C^{(l^*)}$$

(3.7)

for all $x, k, k', l, l^*$, then the probability $p^{(l^*)}$ of success to reveal $(a_0^{(l^*)}, \ldots, a_{K-1}^{(l^*)})$ by picking up a random block satisfies

$$p^{(l^*)} := \frac{1}{C^{(l^*)}(2K - 1)L} \leq p^{(l^*)} \leq \frac{C^{(l^*)}}{(2K - 1)L} =: p^{(l^*)}.$$ 

(3.8)

The harmonic mean being more sensitive to small values, the lower bound is actually closer to the exact probability.

Under the knowledge of $K$, the probability that a random $K$-block is a profile differs from pattern to pattern. But under the control condition (3.7), if all $C^{(l^*)}$ are themselves bounded above by a common constant $C$, the probability $p = p^{(1)} + \ldots + p^{(L)}$ for a random block to be any profile can be estimated by

$$p := \frac{1}{C(2K - 1)} \leq p.$$ 

(3.9)

which has the remarkable property not to depend neither on $L$ nor on the dimension of the ambient space.

As an illustration, Table 1 shows the number of found block profiles found versus their expectations, knowing and without knowing the parameter functions for five simulations of (4.1). The different patterns are split in columns.

### 3.2. Correlation

As we have seen in paragraph 3.1, we need $(2K - 1)L/ \text{harm}(m_1^{(1; l^*)}, \ldots, m_{2K-1}^{(L; l^*)})$ independent random blocks from the series $(X_t)_{t \in \mathbb{Z}}$ to expect that at least one is proportioned like the $l^*$th profile. Practically, given a chain $X_1, \ldots, X_T$ with $T$ observations, we have $T - K + 1$ dependent blocks of length $K$. The main pieces of information about the dependence structure between these blocks can be summarized in the following way.
Simulation with parameters $C = 5$, $D = 20$, $K = 5$, $L = 5$ and $T = 5000$

| #1  | #2  | #3  | #4  | #5  | total | #1  | #2  | #3  | #4  | #5  |
|-----|-----|-----|-----|-----|-------|-----|-----|-----|-----|-----|
| 31  | 31  | 30  | 34  | 33  | 159   | 135 | 28  | 20  | 25  | 30  | 44   |
| 31  | 32  | 33  | 29  | 36  | 161   | 135 | 159 | 30  | 43  | 36  | 30   |
| 35  | 35  | 31  | 31  | 33  | 166   | 135 | 166 | 31  | 24  | 30  | 46   |
| 32  | 36  | 33  | 33  | 33  | 165   | 135 | 160 | 33  | 26  | 43  | 27   |
| 30  | 30  | 30  | 31  | 33  | 154   | 135 | 143 | 31  | 28  | 32  | 33   |

Table 1

Estimation of the number of block profiles versus values obtained in simulation.

i) If a $K$-block $(X_t, \ldots, X_{t+K-1})$ is a block profile, it does not overlap with another block profile.

ii) Given that the $K$-block $(X_t, \ldots, X_{t+K-1})$ is not a profile, the probability that one of the $K - 1$ next $K$-blocks $(X_{t+1}, \ldots, X_{t+K})$ is a profile is higher.

iii) If consecutive blocks are not profiles, the probability that the next one is a profile stops increasing after $K$ non-profile blocks.

To see i), for instance, have a look at the third matrix in (3.3). If $(X_1, \ldots, X_3)$ is like the second profile, then in particular

$$a_2^{(2)} Z_1^{(2)} > a_0^{(1)} Z_3^{(1)}.$$ 

But for $(X_3, \ldots, X_5)$ to be like the first profile, we must have

$$a_2^{(2)} Z_1^{(2)} < a_0^{(1)} Z_3^{(1)}.$$ 

Then $(X_3, \ldots, X_5)$ cannot be proportioned like the first profile. Proceed similarly for any two non-disjoint blocks and any two different profiles.

To see ii), if for instance $(X_1, \ldots, X_3)$ in (3.4) is not a profile, that means that at least one of the 15 inequalities is not satisfied, although we do not know precisely how many. Some of the reverse inequalities lie in the conditions for the $K - 1$ next blocks to be profile and some do not. To get the exact incidence, we need to condition on the number of inequalities not satisfied in (3.4) and whether or not they participate in the conditions for the next blocks to be profile. In any case: the probability that the next blocks are profile increases.

To see iii), simply remark that the process is $K$-dependent.

3.3. Sample size

As a consequence of paragraph 3.2, the expected number of profiles of type $l$ is greater than unity in a chain of length at least

$$T = K - 1 + \frac{(2K - 1)L}{\text{harm}(m_1^{(1)}, \ldots, m_{2K-1}^{(1)})},$$

(3.10)
Given an upper bound \( C \) on all the \( C^{(l)} \) in (3.7), the minimum sample size needed to expect at least \( M \) repetitions of a particular profile in the chain is

\[
T = K + 1 + CM(2K - 1) \tag{3.11}
\]

if we do not know the parameter functions but only \( C \).

4. Estimation

The estimation methodology for the parameter functions \( a_i^{(j)} \) starts from a discretization at \( D \) points \( x_d \) \( (1 \leq d \leq D) \) of the domain. CM3 processes from (3.1) are seen as a high-dimensional M4. Furthermore we may want to consider independent and normally distributed errors with variance \( \sigma^2 \) at each measurement point. Thus the “practical model” studied in this section is

\[
X_t(d) = \max_{1 \leq j \leq L} \max_{0 \leq i < K} a_i^{(j)}(x_d) \gamma_{i-1}^{(j)} + \varepsilon_t(x_d) \quad (1 \leq d \leq D) \tag{4.1}
\]

where \((Z_t)_{t \in \mathbb{Z}}\) are independent unit-Fréchet, \( \varepsilon_t(x_d) \) are independent \( N(0, \sigma^2) \) random variables, the \( a_i^{(j)}(x) \) are positive continuous functions defined on \([0, 1]^q\) and for every \( x \in [0, 1]^q \) we have that \( \sum_{j \in \mathbb{Z}} \sum_{i \geq 0} a_i^{(j)}(x) = 1 \).

It is important to note that the profiles \((a_0^{(j)}, \ldots, a_{K-1}^{(j)})\) contain not only the information about the shapes of the profile but also, according to 3.1, their probability of occurrence. The shapes will be denoted \( a_i^{(j)} \) and their frequencies of occurrence \( f^{(j)} \). That is

\[
a_i^{(j)} = \alpha^{(j)} a_i^{(j)}, \tag{4.2}
\]

where the coefficients \( \alpha^{(j)} \) must be chosen so that \( f^{(j)} = p^{(j)} \) in 3.1.

The first step of the procedure is to estimate the length of the tail dependence \( K \). This is done in 4.1 taking the average size of the clusters of exceedances over a threshold. Next blocks of extremes are selected to estimate number of patterns \( L \), the shapes \( a_i^{(j)} \) of the parameter functions and their frequencies \( f^{(j)} \). The algorithm to locate the blocks of extremes explained in 4.2 is based on a multivariate approach. The value of \( L \) is determined in 4.3 as the number of clusters among the chosen blocks of extremes. The functions \( a_i^{(j)} \) are yielded by the natural output (centroids, medoids, ...) of the partitioning algorithm used to determine \( L \) and the values \( f^{(j)} \) through the size of the different clusters. Then the solution \( a_i^{(j)} \) of (4.2) is obtained in 4.4 thanks to an iterative algorithm. To measure the quality of the estimation, the quantification of the dissimilarity between the the original parameter functions and their estimations is done in 4.5 in terms of the Hausdorff distance.

4.1. Estimation of the length of the tail dependence (\( K \))

Eight estimators of \( K \) have been tested for the model (4.1) with \( \sigma = 0 \) when the knowledge of the parameter functions \( a_i^{(j)} \) is replaced by the range \( C \) given in (3.7).
The first step is to select the values considered as extremes. A first approach consists of working on \( \|X_t\|_\infty \) and choosing the values above the threshold \( \max_t(\|X_t\|_\infty)/C \). This will be referred to as the scalar version. A second approach can be to use all the available information by doing the previous operation for the \( D \) components of \( (X_t) \) separately. In this last case the threshold also depends on the location. This method will be referred as the multivariate version.

Once the extremes are selected, a runs declustering generates a sequence \( (s_n) \) of all sizes of clusters of extremes found in the univariate or multivariate scan. More precisely, only contiguous extremes were considered here to make a cluster. This is the runs declustering with \( r = 0 \).

From the sequence \( (s_n) \), we estimate \( K \) through \( \text{mean}(s_n) \), \( \text{median}(s_n) \) or \( \text{mode}(s_n) \) with the nomenclature as follow.

| Average | Time series | Threshold (scalar or vector) |
|---------|-------------|------------------------------|
| \( K_\mu \) | mean \( (\|X_t\|_\infty) \) | \( \max_t(\|X_t\|_\infty)/C \) |
| \( K_\mu^M \) | mean \( (X_t) \) | \( \max_t(X_t(d))/C \) \( 1 \leq d \leq D \) |
| \( K_m \) | median \( (\|X_t\|_\infty) \) | \( \max_t(\|X_t\|_\infty)/C \) |
| \( K_m^M \) | median \( (X_t) \) | \( \max_t(X_t(d))/C \) \( 1 \leq d \leq D \) |
| \( K_o \) | mode \( (\|X_t\|_\infty) \) | \( \max_t(\|X_t\|_\infty)/C \) |
| \( K_o^M \) | mode \( (X_t) \) | \( \max_t(X_t(d))/C \) \( 1 \leq d \leq D \) |

The ceil or floor options to get rid of the decimals are used to build the eight following estimators:

\[
\hat{K}_1 = \text{ceil}(K_\mu), \quad \hat{K}_3 = \text{ceil}(K_\mu^M), \quad \hat{K}_5 = \text{ceil}(K_m), \quad \hat{K}_7 = \hat{K}_o, \\
\hat{K}_2 = \text{round}(K_\mu), \quad \hat{K}_4 = \text{round}(K_\mu^M), \quad \hat{K}_6 = \text{ceil}(K_m^M), \quad \hat{K}_8 = \hat{K}_m^M.
\]

Figure 2 shows the success rate the eight estimators of \( K \) against the length of the simulated chain. The tests were performed with \( N = 50000 \) trials at each step: for \( C \) from 1 to 10, \( D \) from 1 to 20, \( K \) from 1 to 5, \( L \) from 1 to 5, \( \sigma = 0 \) and, for each of these parameters, 10 different sets of coefficients \( a^{(j)}_i \) randomly generated (uniformly, without time or space correlation).

According to these empirical results, the winner for \( T \geq 35 \) is the univariate version and the mode as average cluster size. If \( T \leq 35 \) the best success rate is obtained with the multivariate version and the median.

### 4.2. Extremal clustering

Once we know the length \( K \) of the tail dependence thanks to 4.1, the next step of the procedure studied here to recover the parameter functions \( a^{(j)}_i \) of a theoretical CM3 process (3.1) is to locate the blocks of extremes. Indeed, according to 3.2, the probability that at least one block of length \( K \) in the chain is block profile of type \( l \) in a sample of size \( T \) is greater than

\[1 - (1 - p^{(l)})^{T-K+1}\]

which tends to 1 as \( T \) tends to infinity.
The suggested method locates the positions of blocks of extremes in the practical model (4.1) maximizing the “likelihood” of being a multivariate extreme. This idea comes from the wish not to lose information across the $D$ dimensions. Nevertheless a bad situation can still happen when, for some $0 \leq i^* < K$, all $a_{i^*}^{(j)}(x_d)$ are negligible in comparison with the $a_i^{(j)}(x_d)$, $i \neq i^*$, for instance. If the points $x_1, \ldots, x_D$ of the discretization are far enough to obtain independent-like patterns, it is unlikely that all the $a_i^{(j)}(x_d)$ are negligible in the same time.

We explain the method on the following example with $D = 2$ and $K = 3$:

First step
Using the order statistics, mark the $K$ largest values in the $D$ chains by 1.

\[
\begin{array}{ccccccc}
  d = 1 & 0 & 0 & 0 & 1 & 1 & 0 \\
  d = 2 & 1 & 0 & 1 & 1 & 0 & 0 \\
\end{array}
\]

Second step
Compute the sum of the extremal status for each $t$.

\[
\begin{array}{ccccccc}
  d = 1 & 0 & 0 & 0 & 1 & 1 & 0 \\
  d = 2 & 1 & 0 & 1 & 1 & 0 & 0 \\
  \Rightarrow & \lambda & 1 & 0 & 1 & 2 & 1 & 0 & 1 \\
\end{array}
\]

Third step
Compute the moving sum (MS) of order $K$. This is considered as the likelihood $\lambda$ to have an large value at time $t$ among the $(Z_t^{(j)})_{1 \leq t \leq L}$.

\[
\begin{array}{ccccccc}
  \lambda & 1 & 0 & 1 & 2 & 1 & 0 & 1 \\
  MS & 1 & 1 & 2 & 3 & 4 & 3 & 2 \\
\end{array}
\]
Extract the profile

Find the index $t$ that maximizes the moving sum. Then

$$ (X_{t-K+1}, \ldots, X_t)/\|X_{t-K+1}\|_{\infty} \quad (4.3) $$

is the shape of the first block profile to store in the memory:

$$
\begin{array}{c|cccccccc}
  MS & 1 & 1 & 2 & 3 & 4 & 3 & 2 \\
  X_t(1) & 5 & 3 & 4 & 14 & 19 & 2 & 7 \\
  X_t(2) & 6 & 1 & 10 & 5 & 2 & 1 & 4 \\
\end{array}
$$

Shape$_1(x_1) = (\frac{4}{10}, \frac{14}{10}, \frac{19}{10})$

Shape$_1(x_2) = (\frac{10}{10}, \frac{5}{10}, \frac{2}{10})$

To decide between multiple maxima, for instance if

$$
\begin{array}{c|cccccccc}
  MS & 1 & 1 & 2 & 4 & 2 & 4 & 4 \\
\end{array}
$$

we first choose the single maximum. If there are consecutive maxima, as a second criterion, we take the block that maximizes $\sum_{t \in \text{block}} \|X_t\|_{\infty}$ among those.

Repeat this loop until having gathered the desired number $Q$ of time-disjoint blocks of extremes of length $K$ ($Q = \rho T$ is the suggestion of 3.1 if we only know a uniform bound $C$ on the variation of the parameter functions).

![Fig 3. Blocks of extremes obtained for a CMS with measurement errors ($\sigma = 1$).](image)
4.3. Estimation of the number of patterns (L)

With Q blocks of extremes of length K normalized as in (4.3), the goal is to estimate the number of reproducible patterns L in the observed process. To do this, we create a Q × KD-table inside which each of the Q lines is made of the D temporal vectors of length K placed successively. We estimate L with the number of clusters for the observations of the table.

Partitioning methods

To break the lines of the table up into groups, we tried several algorithms among which five retained our attention: hierarchical clustering with Ward’s aggregation criterion, hierarchical clustering with the Euclidean distance between the centroids [4, 21], k-means with the Euclidean squared distance, k-means with Pearson’s correlation after standardization [11, 16] and finally Partitioning Around Medoids (PAM) with the classical Euclidean distance [19, 20].

Number of clusters

For each of those algorithms we implemented two criteria to determine the number of clusters. The first: one stops when the percentage of the total variance not explained by the clustering is less than 20%, i.e. when

\[
\frac{\text{SSE}}{\text{SS}^{\text{tot}}} = \frac{\sum_{\text{cluster}} \left( \text{nbobs} (\text{cluster}) - 1 \right) \sum_{\text{variable in cluster}} s^2 (\text{variable})}{(Q - 1) \sum_{\text{variable in table}} s^2 (\text{variable})} \leq 0.20.
\]

We refer to this method as the elbow method [6, 8] (see Figure 4).

![Elbow method](image.png)

Fig 4. Elbow: With L = 5, perfect clustering on the left, CM3 with errors on the right.

The second method to find the number of clusters here is the first value that yields a total silhouette \( T_t \text{Sil} \) for the clustering above 85% of Q. Let \( a(q) \) be the average distance between the \( q \text{th} \) observation and the members of its own cluster. Then repeat this operation between the \( q \text{th} \) observation and the members of all
the other clusters, and set \( b(q) \) to the lowest value found. The silhouette \( s(q) \) of the \( q \)th observation is
\[
s(q) = \frac{b(q) - a(q)}{\max\{a(q), b(q)\}}.
\]
Thus \(-1 \leq s(q) \leq 1\) and \( s(q) \) measures how dissimilar the \( q \)th observation is to its own cluster \([5, 7, 10]\). The distance taken into account here is the Euclidean squared distance. We stop the partitioning at the smallest number of clusters satisfying
\[
\frac{TtSil}{Q} = \frac{1}{Q} \sum_{q=1}^{Q} s(q) \geq 0.85
\]
if this occurs. We refer to this method as the silhouette method (see Figure 5).

Both methods are unable to detect that \( L = 1 \). We have thus considered \( \hat{L} = 1 \) when the estimated variance of each variable is less than 0.005.

Estimation
For \( \sigma = 0 \), Figure 6 shows the success rate of the following eleven estimators of \( L \) against the length of the simulated chain.
The tests were performed with $N = 6000$ trials at each step: for $C$ in $\{2, 4, 6, 8, 10\}$, $D$ in $\{1, 5, 10, 15, 20\}$, $K$ from 2 to 5 and $L$ from 1 to 5. The number of blocks of extremes chosen to build the table is $Q = \min(\text{ceil}(pT), 100)$ with $p$ from 3.1. We added the constraint that it has to be possible to see each profile once, i.e. $(K + 1)L \leq T$, to exclude challenges such as finding 5 profiles of length 5 in a chain of length 10.

The combination between PAM and elbow yields the best success rate for small sample sizes $T \leq 50$. For $T \geq 50$ Ward’s algorithm with the silhouette is the best. Here the consensus curve does not provide a better performance.

The gap between $T = 20$ and $T = 500$ is an intermediate area between a situation where the trivial case $Q = L$ frequently occurs and the apparition of asymptotic properties.

These results are not so excellent but, since the implemented algorithm is the “eye” of the analyst, it only sees what we want it to see. For real applications it probably does not matter if not very frequent profiles are missed. Moreover optimizing $C$, $Q$ and the thresholds for a precise sample can improve
4.4. Recovering the parameter functions $a_i^{(j)}$

Once we got the length $K$ of the tail dependence in 4.1 and the number $L$ of different patterns in 4.3, it remains to estimate the parameter functions $a_i^{(j)} (0 \leq i < K, 1 \leq j \leq L)$. Depending on the partitioning method to estimate $L$, we choose for the shapes $a_i^{(j)}$ of the $L$ profiles the natural output of the algorithm: the centroids for the hierarchical clustering and $k$-means and the medoids with PAM. Essentially the difference is that in the first case the estimator of a shape is a mean of observations and in the second case a median observation.

The relationship between the parameter functions $a_i^{(j)}$, the shapes $a_i^{(j)}$ and their frequencies of occurrence is given by (3.6). The theoretical probabilities of occurrence of the different profiles must match the empirical frequencies $(f^{(1)}, \ldots, f^{(L)})$ in the table of 4.3. Thus the last step is to normalize the $a_i^{(j)}$ to obtain $\hat{a}_i^{(j)}$ such that on the one hand

\[
\text{harm} \left( \left( m_k^{(l;1)} \right)_{k,l} \right) / f^{(1)} = \ldots = \text{harm} \left( \left( m_k^{(l;L)} \right)_{k,l} \right) / f^{(L)}
\]

(1 $\leq k \leq 2K - 1, 1 \leq l \leq L$) and on the other hand

\[
\sum_{j \in \mathbb{Z}} \sum_{i \geq 0} \hat{a}_i^{(j)}(x_d) = 1 \quad (1 \leq d \leq D).
\]

We seek a relation of the type $\hat{a}_i^{(l)} = \alpha^{(l)} a_{i,0}^{(l)} (1 \leq l \leq L)$ which is a problem of rank $L$. The solution can be obtained iteratively: if $a_{i,n}^{(j)}$ is the $n^{th}$ update of $a_{i,0}^{(j)}$ and $(p_n^{(1)}, \ldots, p_n^{(L)})$ the probabilities given by (3.6) at step $n$, then define $a_{i,n+1}^{(j)}$ by

\[
a_{i,n+1}^{(j)}(\cdot) = a_{i,n}^{(j)}(\cdot) \frac{f^{(j)}}{p_n^{(j)}}
\]

and stop when the error $|f^{(1)}, \ldots, f^{(L)} - (p_{n+1}^{(1)}, \ldots, p_{n+1}^{(L)})|_\infty$ is small. The iterative algorithm (4.6), if it converges, converges to the solution of (4.4). Indeed, let

\[
M_n^{(1:*)} = \begin{pmatrix}
m^{(1:1*)}_{1,n} & \cdots & m^{(1:1*)}_{2K-1,n} \\
\cdots & \cdots & \cdots \\
m^{(L:1*)}_{1,n} & \cdots & m^{(L:1*)}_{2K-1,n}
\end{pmatrix}
\]

be the $n^{th}$ update of the initial collection of numbers $(m_{k,0}^{(l;*)})_{k,l}$ given the $a_{i,0}^{(j)}$.

Define

\[
T : \mathbb{R}^{(2K-1)L^2} \to \mathbb{R}^{(2K-1)L^2} : (M_n^{(1)}, \ldots, M_n^{(L)}) \mapsto (M_{n+1}^{(1)}, \ldots, M_{n+1}^{(L)})
\]
that the solution of (4.7) shows an output of the full algorithm using PAM, with measurement dependence in (4.7). This completes the procedure to estimate the \( \alpha_{i}^{(j)} \) form observations.

Figure 7 shows an output of the full algorithm using PAM, with measurement errors but given the true values of \( K \) and \( L \).

### 4.5. Distance between two sets of parameter functions

To measure the quality of the estimation, it is necessary to quantify the dissimilarity between the original parameter functions \( \alpha_{i}^{(j)} (0 \leq i < K, 1 \leq j \leq L) \)
and their estimations $\hat{a}_i^{(j)} (0 \leq i < \hat{K}, 1 \leq j \leq \hat{L})$. If $\hat{K} \neq K$ the estimation can certainly be qualified as bad, so that only the case $\hat{K} = K$ requires a discussion.

The order in which the different patterns are retrieved can change and their total numbers can differ. Consequently the Hausdorff distance between the $L$ graphs of $(a_i^{(j)}(x))$, $(1 \leq j \leq L)$ and the $\hat{L}$ graphs of $(\hat{a}_i^{(j)}(x))$, $(1 \leq j \leq \hat{L})$, that are compact in $[0, 1]^q \times \mathbb{R}^K$ (or in $\{1, \ldots, D\} \times \mathbb{R}^K \cong \mathbb{R}^{DK}$ for the discrete version), perfectly suits. Recall that the Hausdorff distance between nonempty compacts $A$ and $B$ is

$$d_H(A, B) = \max\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \},$$

here considered with the Euclidean distance. We have thus to compute $d_H(A, B)$ with

$$A = \bigcup_{j=1}^L \text{graph}[(a_i^{(j)}(x))_{0 \leq i < K}] \quad \text{and} \quad B = \bigcup_{j=1}^\hat{L} \text{graph}[(\hat{a}_i^{(j)}(x))_{0 \leq i < K}]$$

to reach the stated goal.

Fig 7. Recovery of the parameter functions with $D = 20$, $K = 2$, $L = 3$, $\sigma = 1$ and PAM.
To illustrate the procedure developed in 4.4, Figure 8 shows smoothed histograms of the distances from 4.5 for the estimation of the parameter functions with sample sizes $T = 100, 500, 1000, 5000$ given the true values of $K$ and $L$. The test was performed with $N = 5000$ trials for each sample size: $C$ running in $\{2, 4, 6, 8, 10\}$, $D$ in $\{1, 5, 10, 15, 20\}$, $K$ from 2 to 5, $L$ from 1 to 5 with 10 repetitions of each. The other parameters were $Q = \min(\text{ceil}(pT), 100)$ with $p$ from 3.1 and $\sigma = 1$.

**Figure 8.** Smoothed histograms of the Hausdorff distance between the original parameter functions $(a_i^{(j)}(x))_{0 \leq i < K}$ ($1 \leq j \leq L$) and their estimations $(\hat{a}_i^{(j)}(x))_{0 \leq i < K}$ ($1 \leq j \leq \hat{L}$).

### 5. Conclusion

After the linear processes in function spaces with heavy tailed innovations [9], CM3 processes are other examples of jointly regularly varying time series in function spaces. Under (2.2) they enjoy the finite-cluster condition and under (2.3) the strong mixing property.

Further studies could be to determine whether or not the approximation theorem of Deheuvels for M3 [3] and of Smith and Weissman for M4 [15] also hold for CM3, that is to say if max-stable processes in function spaces, excluding the ones containing a deterministic component, can be arbitrarily closely approached by a CM3. From these papers also arises the question of a generalization of the multivariate extremal index. Since such an object becomes hard to figure out in function spaces, it has maybe to be replaced by the spectral process.

About the estimation, in the empirical study on CM3, we saw that the mode correctly estimates more frequently the length of the tail dependence than the median and the mean. For finding the number of patterns, this study revealed
the importance of simulating the behavior of the chosen method before using it on real data.

**Acknowledgements**

The author thanks Johan Segers for helpful discussions throughout the writing of this paper and especially for his written communications about the finite-cluster condition and the strong mixing condition.

The Matlab code written for this paper is available on the Matlab Central File Exchange at [http://www.mathworks.com/matlabcentral/fileexchange/26539](http://www.mathworks.com/matlabcentral/fileexchange/26539).
Appendix A: Proofs of Section 2

Proof of Proposition 2.1. CLAIM 1 - $X_t$ is a random element in $\mathcal{C}([0,1]^q, \mathbb{R}_+)$. Given $t \in \mathbb{Z}$, write $X \overset{d}{=} X_t$ and $X^{(m)} = \sup_{t, |j| \leq m} a_i^{(j)} Z_t^{(j)}$. For $z > 0$, since the $Z_t^{(j)}$ are iid unit-Frèchet,

$$P(||X|| \leq z) = P(\forall i \geq 0, \forall j \in \mathbb{Z} : Z_t^{(j)} \leq z/\|a_i^{(j)}\|_{\infty})$$

$$= \prod_{j \in \mathbb{Z}} \exp(-\|a_i^{(j)}\|_{\infty}/z) = \exp(-\sum_{j \in \mathbb{Z}} \sup_{i \geq 0} \|a_i^{(j)}\|_{\infty}/z).$$

If $\sum_{j \in \mathbb{Z}} \sum_{i \geq 0} \|a_i^{(j)}\|_{\infty} = \infty$, then this probability is zero for all $z > 0$, hence $||X||_{\infty} = \infty$ with probability one.

If $\sum_{j \in \mathbb{Z}} \sum_{i \geq 0} \|a_i^{(j)}\|_{\infty} < \infty$, then $X^{(m)} \leq X \leq X^{(m)} + \sup_{t, |j| > m} \|a_i^{(j)}\|_{\infty} Z_t^{(j)}$ and thus $||X - X^{(m)}||_{\infty} \leq \sup_{t, |j| > m} \|a_i^{(j)}\|_{\infty} Z_t^{(j)}$. Writing

$$Y_m = \sup_{t, |j| > m} \|a_i^{(j)}\|_{\infty} Z_t,$$

we have $Y_m \geq 0$ and $Y_m$ is decreasing in $m$. By a similar computation as above, we find that for $y > 0$,

$$P(Y_m \leq y) = \exp(-\sum_{t, |j| > m} \|a_i^{(j)}\|_{\infty}/y).$$

As a consequence, $Y_m \to 0$ in probability and, by monotonicity, almost surely. Since the uniform limit of a sequence of continuous functions is continuous, by monotonicity, $X$ is continuous with probability one.

Then the fact that, for every $t \in \mathbb{Z}$, the map $\omega \mapsto X_t(\omega, \cdot)$ with values in $\mathcal{C}([0,1]^q, \mathbb{R}_+)$ is measurable follows easily.

CLAIM 2 - $(X_t)_{t \in \mathbb{Z}}$ is a stationary time series. In extension, this property is that, for every $n \geq 0$, every $h \geq 0$ and every measurable set $A \subset \mathcal{C}([0,1]^q, \mathbb{R}_+)$,

$$P((X_0, \ldots, X_n) \in A) = P((X_h, \ldots, X_{n+h}) \in A).$$

The argument is based on two facts:

1) It suffices to pick up $x \in [0,1]^q$, $z_0, \ldots, z_n \in \mathbb{R}$ and verify that

$$P(X_0(x) \leq z_0, \ldots, X_n(x) \leq z_n) = P(X_h(x) \leq z_0, \ldots, X_{n+h}(x) \leq z_n),$$

the case with $k$ points $x_1, \ldots x_k \in [0,1]^q$ being similar.
2) The right hand-side of the expression in 1) admits the expansion

\[
P(X_h(x) \leq z_0, \ldots, X_{n+h}(x) \leq z_n) = P(\sup_{j \in \mathbb{Z}} \sup_{i \geq 0} a_i^{(j)}(x) Z_{h-i}^{(j)} \leq z_0, \ldots, \sup_{i \geq 0} a_i^{(j)}(x) Z_{n+h-i}^{(j)} \leq z_n)
\]

\[
= P(\forall j_0 \in \mathbb{Z}, \forall i_0 \geq 0 : a_{i_0}^{(j_0)}(x) Z_{h-i_0}^{(j_0)} \leq z_0, \ldots,
\]

\[
= P(\forall j \in \mathbb{Z} : Z_{h+n}^{(j)} \leq \frac{z_n}{a_0^{(j)}(x)}).
\]

\[
Z_{h+n-1}^{(j)} \leq \min\left(\frac{z_n}{a_1^{(j)}(x)}, \frac{z_{n-1}}{a_0^{(j)}(x)}\right),
\]

\[
Z_{h+n-2}^{(j)} \leq \min\left(\frac{z_n}{a_2^{(j)}(x)}, \frac{z_{n-1}}{a_1^{(j)}(x)}, \frac{z_{n-2}}{a_0^{(j)}(x)}\right),
\]

\[
\ldots,
\]

\[
Z_{h}^{(j)} \leq \min\left(\frac{z_n}{a_n^{(j)}(x)}, \frac{z_{n-1}}{a_{n-1}^{(j)}(x)}, \ldots, \frac{z_0}{a_0^{(j)}(x)}\right),
\]

\[
Z_{h-1}^{(j)} \leq \min\left(\frac{z_n}{a_{n+1}^{(j)}(x)}, \frac{z_{n-1}}{a_n^{(j)}(x)}, \ldots, \frac{z_0}{a_1^{(j)}(x)}\right),
\]

\[
\ldots
\]

which allows us to dispose of the variable \( h \) by independence.

**Proof of Proposition 2.3.** Given \( s, t \geq 0 \) consider the Dirac masses in \( \mathcal{G}([0,1]^q, \mathbb{R}) \)

\[
\delta(-s, t; i, j) \overset{d}{=} \left(\frac{a_{-s+i}^{(j)}}{\|a_i^{(j)}\|_\infty}, \ldots, \frac{a_{t+i}^{(j)}}{\|a_i^{(j)}\|_\infty}\right)
\]

\[
\overset{d}{=} \left[\begin{array}{c}
\frac{a_{-s+i}^{(j)} Z}{\|a_i^{(j)} Z\|_\infty}, \ldots, \frac{a_{t+i}^{(j)} Z}{\|a_i^{(j)} Z\|_\infty}
\end{array}\right] \|a_i^{(j)}\|_\infty Z > x
\]

with \( i \geq 0 \) and \( j \in \mathbb{Z} \).

A slight variation in the proof of [9, Lemma B.2] using the identity

\[
1\{\max(\|X_1\|, \|X_2\|) > x\} = 1\{\|X_1\| > x\} + 1\{\|X_2\| > x\} - 1\{\|X_1\| > x\}1\{\|X_2\| > x\}
\]

and considering \( T_i^{(j)}(z) = a_i^{(j)} z \) leads to the new conclusion that

\[
E\left[\left\{\sup_{j \in \mathbb{Z}} a_i^{(j)} \mathbb{E}^{(j)}_{X_i < 0} > x\right\} - \sum_{j \in \mathbb{Z}} 1\{a_i^{(j)} Z > x\}\right] = o(P(Z > x)), x \to \infty.
\]

(A.1)
Consequently, for \( f \in \mathcal{F}_b(\mathcal{H}([0, 1]^s, \mathbb{R}^{s+l+1})) \), by virtue of (A.1)

\[
E \left[ f \left( \frac{X_{i+1}}{x}, \ldots, \frac{X_i}{x} \right) \left\| X_0 \right\|_{\infty} > x \right]
= \frac{E \left[ f \left( \frac{X_{i+1}}{x}, \ldots, \frac{X_i}{x} \right) 1 \left\{ \left\| X_0 \right\|_{\infty} > x \right\} \right] P(Z > x)}{P \left( \left\| X_0 \right\|_{\infty} > x \right)}
= \frac{E \left[ f \left( \frac{X_{i+1}}{x}, \ldots, \frac{X_i}{x} \right) 1 \left\{ \sup_{j \in \mathbb{Z}} \sup_{i \geq 0} \left\| a^{(j)}_i \right\|_{\infty} Z^{(j)}_{i-1} > x \right\} \right] P(Z > x)}{P \left( \left\| X_0 \right\|_{\infty} > x \right)}
= \sum_{j \in \mathbb{Z}} E \left[ f \left( \frac{X_{i+1}}{x}, \ldots, \frac{X_i}{x} \right) 1 \left\{ \left\| a^{(j)}_i \right\|_{\infty} Z^{(j)}_{i-1} > x \right\} \right] \frac{P(\left\| a^{(j)}_i \right\|_{\infty} Z^{(j)}_{i-1} > x) P(Z > x)}{P \left( \left\| X_0 \right\|_{\infty} > x \right)} + o(1).
\]

Thanks to the dominated convergence and the continuity of \( f \), the last expression is equal to

\[
\sum_{j \in \mathbb{Z}} E \left[ f \left( \frac{\sup_{l \in \mathbb{Z}} \sup_{k \geq 0} \left\{ a^{(j)}_k Z^{(j)}_{-k-l} \right\}}{x}, \ldots, \frac{\sup_{l \in \mathbb{Z}} \sup_{k \geq 0} \left\{ a^{(j)}_k Z^{(j)}_{-k-l} \right\}}{x} \right) \left\| a^{(j)}_i \right\|_{\infty} Z^{(j)}_{i-1} > x \right] \frac{P(\left\| a^{(j)}_i \right\|_{\infty} Z^{(j)}_{i-1} > x) P(Z > x)}{P \left( \left\| X_0 \right\|_{\infty} > x \right)} + o(1)
= \sum_{j \in \mathbb{Z}} E \left[ f \left( \frac{a^{(j)}_i Z^{(j)}_{-i}}{x}, \ldots, \frac{a^{(j)}_i Z^{(j)}_{-i}}{x} \right) \left\| a^{(j)}_i \right\|_{\infty} Z^{(j)}_{i-1} > x \right] \frac{P(\left\| a^{(j)}_i \right\|_{\infty} Z^{(j)}_{i-1} > x) P(Z > x)}{P \left( \left\| X_0 \right\|_{\infty} > x \right)} + o(1)
= \sum_{j \in \mathbb{Z}} E \left[ f \left( \frac{Z^{(j)}_{i+1}}{x}, \ldots, \frac{Z^{(j)}_{i+1}}{x} \right) \left\| a^{(j)}_i \right\|_{\infty} Z^{(j)}_{i-1} > x \right] \frac{P(\left\| a^{(j)}_i \right\|_{\infty} Z^{(j)}_{i-1} > x) P(Z > x)}{P \left( \left\| X_0 \right\|_{\infty} > x \right)} + o(1).
\]

Using the continuous mapping theorem and the regular variation of \( Z \), the last term converges to

\[
\sum_{j \in \mathbb{Z}} E \left[ f(Y \delta(-s, t, i, j)) \right] \left\| a^{(j)}_i \right\|_{\infty} \frac{1}{\sum_{l \in \mathbb{Z}} \sum_{k \geq 0} \left\| a^{(l)}_k \right\|_{\infty}}
\]

where \( Y \sim \text{Pareto}(1) \). The factors

\[
\pi_{ij} := \frac{\left\| a^{(j)}_i \right\|_{\infty}}{\sum_{l \in \mathbb{Z}} \sum_{k \geq 0} \left\| a^{(l)}_k \right\|_{\infty}}
\]
form a probability distribution on \( \mathbb{Z}_+ \times \mathbb{Z} \), let us say of a random vector \( S \).

According to [9, Theorem 3.1 (iv)], the spectral process of the CM3 process \((X_t)_{t \in \mathbb{Z}}\) is

\[
(\Theta_{-s}, \ldots, \Theta_{t}) \overset{d}{=} \delta(-s, t; S)
\]

which has the announced distribution.

**Proof of Proposition 2.4.** It suffices to check that

\[
\lim_{m \to \infty} \limsup_{n \to \infty} P(\exists t \in \{2m, \ldots, r_n\}: \|X_t\| > n \mid \|X_0\| > n) = 0.
\]

For the convenience of the proof, set \( p = t - i \) so that

\[
X_t = \sup_{j \in \mathbb{Z}} \sup_{p \leq t} a^{(j)}_{t-p} Z_p^{(j)}.
\]

Let \( b^{(j)}_t := \|a^{(j)}_t\|_\infty \) to have

\[
\|X_t\| = \sup_{j \in \mathbb{Z}} \sup_{p \leq t} b^{(j)}_t Z_p^{(j)}.
\]

Given \( m > 0 \), write \( A_n := \{\|X_0\| > n\} \) so that

\[
A_n = \{\exists j \in \mathbb{Z}, \exists p \leq 0 : b^{(j)}_t Z_p^{(j)} > n\}.
\]

We have, as \( n \to \infty \),

\[
nP(A_n)
= n[1 - P(\forall j \in \mathbb{Z}, \forall p \leq 0 : b^{(j)}_t Z_p^{(j)} \leq n)]
= n[1 - \exp(-\frac{1}{n} \sum_{j \in \mathbb{Z}} \sum_{p \leq 0} b^{(j)}_p)]
\to \sum_{j \in \mathbb{Z}} \sum_{p \leq 0} b^{(j)}_p = \sum_{j \in \mathbb{Z}} \sum_{p \geq 0} b^{(j)}_p < +\infty.
\]

For \( m > 0 \), decompose \( \{\exists t \in \{2m, \ldots, r_n\}: \|X_t\| > n\} = C_{m,n} \cup D_{m,n} \) where

\[
C_{m,n} = \{\exists t \in \{2m, \ldots, r_n\}, \exists j \in \mathbb{Z}, \exists p > m : b^{(j)}_{t-p} Z_p^{(j)} > n\},
D_{m,n} = \{\exists t \in \{2m, \ldots, r_n\}, \exists j \in \mathbb{Z}, \exists p \leq m : b^{(j)}_{t-p} Z_p^{(j)} > n\}.
\]

We have

\[
P(C_{m,n})
= 1 - P(\forall t \in \{2m, \ldots, r_n\}, \forall j \in \mathbb{Z}, \forall p > m : b^{(j)}_{t-p} Z_p^{(j)} \leq n)
= 1 - P(\forall j \in \mathbb{Z}, \forall p > m : \max_{2m \leq t \leq r_n} b^{(j)}_{t-p} Z_p^{(j)} \leq n)
= 1 - \exp(-\frac{1}{n} \sum_{j \in \mathbb{Z}} \sum_{p > m} \max_{2m \leq t \leq r_n} b^{(j)}_{t-p})
\leq \frac{1}{n} \sum_{j \in \mathbb{Z}} \sum_{p > m} \max_{2m \leq t \leq r_n} b^{(j)}_{t-p}
\leq \frac{m_n}{n} \sum_{j \in \mathbb{Z}} \sum_{k > 0} b^{(j)}_{t_k} \to 0
\]
as \( n \to \infty \).

Setting \( p = 2m - q \) write

\[
A_n = \{ \exists j \in \mathbb{Z}, \exists q \geq 2m : b^{(j)}_{q-2m}Z_q^{(j)} > n \}
\]

and

\[
D_{m,n} = \{ \exists j \in \mathbb{Z}, \exists q \geq m : \max_{q \leq t \leq q-2m+r_n} b^{(j)}_t Z^{(j)}_t > n \}
\]

hence

\[
nP(A_n \cap D_{m,n}) \leq n \sum_{j \in \mathbb{Z}} \sum_{q \geq 2m} P(\{ b^{(j)}_{q-2m}Z_q^{(j)} > n \} \cap D_{m,n})
\]

\[
\leq n \sum_{j \in \mathbb{Z}} \sum_{q \geq 2m} P(\min\{b^{(j)}_{q-2m}, \max_{t \geq q} b^{(j)}_t Z_t^{(j)} \} > n)
\]

\[
+ n \sum_{j \in \mathbb{Z}} \sum_{q \geq 2m} P(b^{(j)}_{q-2m}Z_q^{(j)} > n)P(D_{m,n}).
\]

The first term of the sum is bounded above by

\[
\sum_{j \in \mathbb{Z}} \sum_{q \geq 0} \min\{b^{(j)}_q, \max_{t \geq 2m} b^{(j)}_t \}
\]

and the second term tends to 0 as \( n \to \infty \) since

\[
n \sum_{j \in \mathbb{Z}} \sum_{q \geq 2m} P(b^{(j)}_{q-2m}Z_q^{(j)} > n) \leq \sum_{j \in \mathbb{Z}} \sum_{q \geq 0} b^{(j)}_q < \infty
\]

and

\[
P(D_{m,n}) = \frac{1}{n} \sum_{j \in \mathbb{Z}} \sum_{q \geq m} \max_{q \leq t \leq q-2m+r_n} b^{(j)}_t \leq \frac{m_n}{n} \sum_{j \in \mathbb{Z}} \sum_{q \geq 0} b^{(j)}_q \to 0.
\]

Consequently

\[
\limsup_{n \to \infty} nP(A_n \cap D_{m,n}) \leq \sum_{j \in \mathbb{Z}} \sum_{p \in \mathbb{Z}} \min\{b^{(j)}_p, \max_{t \geq 2m} b^{(j)}_t \}.
\]

Since \( A_n \) and \( C_{m,n} \) are independent,

\[
P(\exists t \in \{2m, \ldots, r_n\} : \|X_t\| > n \mid \|X_0\| > n)
\]

\[
= P(C_{m,n} \cup D_{m,n} \mid A_n)
\]

\[
= \frac{P([C_{m,n} \cup D_{m,n}] \cap A_m \cap A_n)}{P(A_n)}
\]

\[
\leq \frac{P([C_{m,n} \cap A_n] \cup \{D_{m,n} \cap A_n\})}{P(A_n)}
\]

\[
\leq P(C_{m,n}) + \frac{P(D_{m,n} \cap A_n)}{P(A_n)}
\]
so that
\[ \lim_{n \to \infty} \sup P(\exists t \in \{2m, \ldots, r_n\} : \|X_t\| > n \mid \|X_0\| > n) \]
\[ \leq \sum_{j \in \mathbb{Z}} \sum_{p \geq 0} b_p^{(j)} \left[ \sum_{j \in \mathbb{Z}} \sum_{p \in \mathbb{Z}} \min \{b_p^{(j)}, \max_{t \geq 2m} b_t^{(j)}\} \right] \]
which tends to 0 if \( m \to \infty \).

**Proof of Proposition 2.5.** Again set \( p = t - i \) to have
\[ X_t = \sup_{j \in \mathbb{Z}, p \leq t} a_{t-p}^{(j)} Z_p^{(j)}. \]
For large \( t \geq 0 \), we will prove that we can approach \( X_t \) by
\[ X_t^{(+)} = \sup_{j \in \mathbb{Z}, 1 \leq p \leq t} a_{t-p}^{(j)} Z_p^{(j)} \]
in the sense of
\[ \lim_{m \to +\infty} P(\forall t \geq m : X_t^{(+)} = X_t) = 1. \]
Then the conclusion will follow from the fact that the processes \( (X_t^{(+)})_{t \geq 1} \) and \( (X_t)_{t \leq -1} \) are independent.

As the Borel \( \sigma \)-field on \( \mathcal{C}([0,1]^q) \) is generated by the finite dimensional sets, it is sufficient to check that the limit is 1 on every subset \( \{x_1, \ldots, x_k\} \subset [0,1]^q \).

Passing to the complementary, compute
\[ p_m := P(\exists t \geq m, \exists 1 \leq i \leq k : X_t^{(+)}(x_i) < X_t(x_i)) \]
\[ = P(\exists t \geq m, \exists 1 \leq i \leq k : \sup_{j \in \mathbb{Z}, 1 \leq p \leq t} a_{t-p}^{(j)} Z_p^{(j)} < \sup_{j \in \mathbb{Z}, p \leq t} a_{t-p}^{(j)} Z_p^{(j)}) \]
\[ = P(\exists t \geq m, \exists 1 \leq i \leq k : \sup_{j \in \mathbb{Z}, 1 \leq p \leq t} a_{t-p}^{(j)} Z_p^{(j)} < \sup_{j \in \mathbb{Z}, p \leq t} a_{t-p}^{(j)} Z_p^{(j)}). \]
If \( a, b > 0 \) and \( Z_1, Z_2 \) are standard unit-Fréchet, as in (3.6),
\[ P(aZ_1 < bZ_2) = \frac{1}{1 + \frac{b}{a}} < \frac{b}{a} \]
so that
\[ p_m \leq \sum_{t \geq m} \sum_{i=1}^{k} \sum_{j \in \mathbb{Z}} \sum_{p \leq 0} a_{t-p}^{(j)}(x_i) \sum_{j \in \mathbb{Z}} \sum_{1 \leq p \leq t} a_{t-p}^{(j)}(x_i) =: \sum_{t \geq m} \sum_{i=1}^{k} \sum_{j \in \mathbb{Z}} \frac{N_t(x_i)}{D_t(x_i)}. \]
According to (2.2), \( D_t(x) \) uniformly converges to a continuous function \( D(x) \) as \( t \to \infty \). Without loss of generality, we can assume that \( D(t) \equiv 1 \). Write \( D_t(x) = 1 - \varepsilon_t(x) \) with \( \|\varepsilon_t\|_{\infty} \to 0 \) as \( t \to \infty \). Consequently,
\[ \frac{1}{D_t(x)} = 1 + \frac{\varepsilon_t(x)}{D_t(x)} \]
hence
\[ p_m \leq \sum_{t \geq m} \sum_{i=1}^{k} N_t(x_i) + \sum_{t \geq m} \sum_{i=1}^{k} N_t(x_i) \varepsilon_t(x_i) \frac{\varepsilon_t(x_i)}{D_t(x_i)}. \]

Next, use the fact that for every \( \eta > 0 \), there exists \( m_\eta \) such that for every \( t \geq m_\eta \) and every \( 1 \leq i \leq k \)
\[ 0 \leq \frac{\varepsilon_t(x_i)}{D_t(x_i)} \leq \eta \]
to see that
\[ \sum_{t \geq m} \sum_{i=1}^{k} N_t(x_i) \frac{\varepsilon_t(x_i)}{D_t(x_i)} = o(1) \]
and so obtain
\[ p_m \leq (1 + o(1)) \sum_{t \geq m} \sum_{i=1}^{k} \sum_{j \in \mathbb{Z}} \sum_{p \leq 0} a_{t-p}^{(j)}(x_i). \]

Since
\[ \sum_{t \geq m} \sum_{i=1}^{k} \sum_{j \in \mathbb{Z}} \sum_{p \leq 0} a_{t-p}^{(j)}(x_i) = \sum_{k \geq m} \sum_{i=1}^{k} \sum_{j \in \mathbb{Z}} (k - m + 1) a_k^{(j)}(x_i), \]
thanks to (2.3) we have that \( p_m \) vanishes as \( m \to \infty \), which proves the result.

**Proof of Proposition 2.6.** Since
\[ \| X_t \|_\infty = \| \sup_{j \in \mathbb{Z}} \sup_{t \geq 0} a_{t-i}^{(j)} Z_{t-i}^{(j)} \|_\infty = \sup_{j \in \mathbb{Z}} \| a_{t-i}^{(j)} \|_\infty Z_{t-i}^{(j)} \]
it suffices to transcribe the formula for a M3 from [15].

**References**

[1] Richard A. Davis and Thomas Mikosch. Extreme value theory for space-time processes with heavy-tailed distributions. *Stochastic Process. Appl.*, 118(4):560–584, 2008.

[2] Laurens de Haan and Ana Ferreira. *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. An introduction.

[3] Paul Deheuvels. Point processes and multivariate extreme values. *J. Multivariate Anal.*, 13(2):257–272, 1983.

[4] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning (2nd edition)*. Springer, New York, 2009.
[5] Leonard Kaufman and Peter J. Rousseeuw. *Finding groups in data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1990. An introduction to cluster analysis, A Wiley-Interscience Publication.

[6] David J. Jr Ketchen and Christopher L. Sho. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management J.*, 17(6):441–458, 1996.

[7] R. Jr Lleti, M.C. Ortiz, Sarabia L.A., and M.S Sánchez. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta*, 515:87–100, 2004.

[8] Kantilal Varichand Mardia, John T. Kent, and John M. Bibby. *Multivariate analysis*. Academic Press [Harcourt Brace Jovanovich Publishers], London, 1979. Probability and Mathematical Statistics: A Series of Monographs and Textbooks.

[9] Thomas Meinguet and Johan Segers. Regularly varying time series in Banach spaces. *Submitted, arXiv:1001.3262*, 2010.

[10] Peter J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Comp. and Appl. Math.*, 20:53–65, 1987.

[11] G. A. F. Seber. *Multivariate observations*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1984.

[12] Johan Segers. Approximate distributions of clusters of extremes. *Statist. Probab. Lett.*, 74(4):330–336, 2005.

[13] Johan Segers. Rare events, temporal dependence, and the extremal index. *J. Appl. Probab.*, 43(2):463–485, 2006.

[14] Richard L. Smith. Statistics of extremes, with applications in environment, insurance and finance. Chapter 1 of, Extreme Values in Finance, Telecommunications and the Environment, edited by B. Finkenstadt and H. Rootzen, Chapman and Hall/CRC Press, London. pages 1–78, 2003.

[15] Richard L. Smith and Ishay Weissman. Characterization and estimation of the multivariate extremal index. *Dept. Stat. Oper. Res., Univ. North Carolina, Chapel Hill, NC*, 1996.

[16] H. Spath. *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples*, translated by J. Goldschmidt. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Halsted Press, 1985.

[17] M. Süveges. *Statistical Analysis of Clusters of Extreme Events*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2009.

[18] M. Süveges and A. C. Davison. A dirichlet mixture approach for clusters of extreme events. *Submitted*, 2010.

[19] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition, Third Edition*. Academic Press, Inc., Orlando, FL, USA, 2006.

[20] Mark J. van der Laan, Katherine S. Pollard, and Jennifer Bryan. A new partitioning around medoids algorithm. *J. Stat. Comput. Simul.*, 73(8):575–584, 2003.

[21] Joe H. Ward, Jr. Hierarchical grouping to optimize an objective function.
[22] Zhengjun Zhang. The estimation of M4 processes with geometric moving patterns. *Ann. Inst. Statist. Math.*, 60(1):121–150, 2008.

[23] Zhengjun Zhang and Richard L. Smith. The behavior of multivariate maxima of moving maxima processes. *J. Appl. Probab.*, 41(4):1113–1123, 2004.