Extremes
DOI 10.1007/s10687-013-0172-7

The distribution of the maximum of a first order moving average: the continuous case

Christopher S. Withers · Saralees Nadarajah

Received: 10 June 2011 / Revised: 12 November 2012 / Accepted: 6 March 2013

Abstract We give the cumulative distribution function of M_n, the maximum of a sequence of n observations from a first order moving average. Solutions are first given in terms of repeated integrals and then for the case, where the underlying independent random variables have an absolutely continuous probability density function. When the correlation is positive,

$$P(M_n \le x) \; = \; \sum_{j=1}^{\infty} \beta_{j,x} \; \nu_{j,x}^n,$$

where $\{\nu_{j,x}\}$ are the eigenvalues (singular values) of a Fredholm kernel and $\beta_{j,x}$ are some coefficients determined later. A similar result is given when the correlation is negative. The result is analogous to large deviations expansions for estimates, since the maximum need not be standardized to have a limit. For the continuous case the integral equations for the left and right eigenfunctions are converted to first order linear differential equations. The eigenvalues satisfy an equation of the form

$$\sum_{i=1}^{\infty} w_i (\lambda - \theta_i)^{-1} = \lambda - \theta_0$$

for certain known weights $\{w_i\}$ and eigenvalues $\{\theta_i\}$ of a given matrix. This can be solved by truncating the sum to an increasing number of terms.

C. S. Withers
Applied Mathematics Group, Industrial Research Limited,
Lower Hutt, New Zealand

S. Nadarajah (✉)
School of Mathematics, University of Manchester,
Manchester M13 9PL, UK
e-mail: mbbsssn2@manchester.ac.uk

Published online: 27 March 2013

Keywords Fredholm kernel · Maximum · Moving average · Rainfall

1 Introduction and summary

Many authors have considered extreme value theory for moving average processes, see Rootzén (1978), Leadbetter et al. (1983, page 59), Davis et al. (1985), Davis and Resnick (1985, 1989, 1991), Rootzén (1986), O'Brien (1987), Resnick (1987, page 239, 1988), Deheuvels (1988), Park (1992), Albin (1997), Hall (2001, 2003), Hall and Scotto (2003), Scotto and Ferreira (2003), Hall et al. (2004, 2010), Martins and Ferreira (2004), Fasen (2005, 2006), Klüppelberg and Lindner (2005) and Hall and Moreira (2006). However, the results either give the limiting extreme value distributions or assume that the errors come from a specific class (for example, integer-valued, exponential type, periodic type, stable type, heavy tailed, light tailed, regularly varying tailed, etc). We are aware of no work giving the *exact* distribution of the maximum of moving average processes.

This paper gives a powerful method for deriving the *exact* distribution of extremes of n correlated observations as weighted sums of nth powers of associated eigenvalues. Withers and Nadarajah (2011) illustrated this method for a first order autoregressive process. Here, the method is illustrated for a first order moving average process.

Let $\{e_i\}$ be independent and identically distributed random variables from some cumulative distribution function (cdf) F on $\mathbb{R}$. Consider the first order moving average,

$$X_i = e_i + \rho e_{i-1}, \tag{1.1}$$

where $\rho \neq 0$. In Section 2, we give expressions for the cdf of the maximum $M_n = \max_{i=1}^n X_i$ in terms of repeated integrals. These expressions are obtained via the recurrence relationship

$$G_n(y) = I(\rho < 0) G_{n-1}(\infty) F(y) + \mathcal{K} G_{n-1}(y), \tag{1.2}$$

where

$$G_n(y) = P(M_n \leq x,\ e_n \leq y), \tag{1.3}$$

$$\mathcal{K} r(y) = \mathrm{sign}(\rho) \int^y r((x-w)/\rho) dF(w), \tag{1.4}$$

and $I(A) = 1$ or 0 for A true or false. So, $\mathcal{K}$ is an integral operator depending on x. Here, the dependence of $\mathcal{K}$ on x is suppressed. We shall use this kind of suppression throughout the paper for the sake of simplicity.

For Eq. 1.2 to work at $n = 1$ we define $M_0 = -\infty$ so that $G_0(y) = F(y)$. In Sections 3 and 4, we consider the case when F is absolutely continuous with probability density function (pdf) $f(x)$ with respect to Lebesque measure. In this case, we show that corresponding to $\mathcal{K}$ is a Fredholm kernel $K(y, z)$. We give a solution in terms of its eigenvalues and eigenfunctions. This leads easily to the asymptotic results stated in the abstract. In Section 4, we also discuss a numerical example with F taken to be a standard normal cdf.

Our expansions for $P(M_n \leq x)$ for fixed x are large deviation results. If x is replaced by x_n such that $P(M_n \leq x_n)$ tends to the generalized extreme value cdf, then the expansion still holds, but not the asymptotic expansion in terms of a single eigenvalue, since this may approach one as $n \to \infty$. Various conditions for $P(M_n \leq x_n)$ to tend to the generalized extreme value cdf for moving average processes are gven in Rootzen (1978, 1986), Leadbetter et al. (1983, page 59), Davis et al. (1985), Davis and Resnick (1985, 1989, 1991, 1998), O'Brien (1987), Resnick (1987, page 239, 1988), Deheuvels (1988), Park (1992), Albin (1997), Hall (2001, 2003), Hall and Scotto (2003), Scotto and Ferreira (2003), Hall et al. (2004, 2010), Martins and Ferreira (2004), Fasen (2005, 2006), Klüppelberg and Lindner (2005), and Hall and Moreira (2006).

The conditions for $P(M_n \leq x_n)$ to tend to the generalized extreme value cdf are not always satisfied; that is, $P(M_n \leq x_n)$ may not always tend to a non-degenerate limit as $n \to \infty$. See Nadarajah and Mitov (2002) and references therein for several examples. An attractive feature of results in this paper is that the expansion for $P(M_n \leq x_n)$ holds for every finite n even if $P(M_n \leq x_n)$ does not have a non-degenerate limit as $n \to \infty$.

Another attractive feature is that our results give the exact distribution of M_n for every finite n, especially for small n. Traditionally, the cdf of M_n is set equal to the generalized extreme value cdf for some n considered large even if M_n suitably normalized does not have a non-degenerate limit. There are statistical tests for checking if an extreme value limit exists, but these tests are hardly used in published applications. Furthermore, data sets are often small, so n may not be large enough to justify the use of the generalized extreme value distribution. So, one would expect a model based on the derived expansion to provide a better and a more legitimate fit than the traditional model. In Section 5, we illustrate this fact using two real data sets.

The numerical results in Sections 4 and 5 make use of the main results in Section 2 (Theorem 2.2), Section 3 (Theorem 3.1), and Section 4 (Theorems 4.2, 4.3 and 4.4), so illustrating the main contributions of the paper. The final section (Section 6) discusses usefulness of the method developed for other time series.

Throughout, we set $\int r = \int r(y)dy$. We shall denote the first derivative of a function, say $\omega(\cdot)$, by $\dot{\omega}$.

2 Solutions using repeated integrals

Our goal in this section is to determine

$$u_n = P(M_n \leq x) = G_n(\infty).$$

Theorem 2.1 gives u_n in terms of

$$v_n = [\mathcal{K}^n F(y)]_{y=\infty}.$$

For example,

$$v_1 = -\int F(z)dF(x-\rho z) = -I(\rho < 0) + \int F(x-\rho z)dF(z). \tag{2.1}$$

As noted, the behavior of u_n falls into the two cases: $\rho > 0$ and $\rho < 0$. In the case $\rho < 0$, Theorem 2.2 provides an explicit solution for u_n in terms of Bell polynomials.

Theorem 2.1 *The case $\rho > 0$: For $n \geq 1$,*

$$u_n = v_n. \tag{2.2}$$

The marginal cdf of X_1 is $u_1 = v_1$ given by Eq. 2.1.
The case $\rho < 0$: For $n \geq 0$,

$$u_{n+1} = v_{n+1} + \sum_{i=0}^{n} v_i u_{n-i} \tag{2.3}$$

with the initial value $u_0 = 1$. The marginal cdf of X_1 is $u_1 = 1 + v_1$ of Eq. 2.1.

Proof For $n \geq 1$, G_n of Eq. 1.3 satisfies

$$\begin{aligned} G_n(y) &= P(M_{n-1} \leq x, e_n + \rho e_{n-1} \leq x, e_n \leq y) \\ &= P(M_{n-1} \leq x, e_{n-1} \leq (x-e_n)/\rho, e_n \leq y) \text{ if } \rho > 0 \\ &= \int^y G_{n-1}((x-w)/\rho)dF(w) \\ &= P(M_{n-1} \leq x, e_{n-1} \geq (x-e_n)/\rho, e_n \leq y) \text{ if } \rho < 0 \\ &= \int^y [G_{n-1}(\infty) - G_{n-1}((x-w)/\rho)]dF(w) \\ &= G_{n-1}(\infty)F(y) - \int^y G_{n-1}((x-w)/\rho)dF(w). \end{aligned}$$

That is, for $n \geq 1$, Eq. 1.2 holds. In the case $\rho > 0$, Eq. 2.2 follows since $G_n(y) = \mathcal{K}^n F(y)$. In the case $\rho < 0$, set $a_i(y) = \mathcal{K}^i F(y)$ and $a_i = a_i(\infty)$. By Eq. 1.2, for $n \geq 0$,

$$G_{n+1}(y) = u_n F(y) + \mathcal{K} G_n(y) = \sum_{i=0}^{n} a_i(y) u_{n-i} + a_{n+1}(y).$$

Putting $y = \infty$ gives the recurrence solution (2.3). □

Theorem 2.2 *Set $w_n = v_{n-1}$ and $\mathbf{w} = (w_1, w_2, \ldots)$. In the case $\rho < 0$,*

$$u_n = \widehat{B}_n(\mathbf{w}) \otimes v_n \tag{2.4}$$

for $n \geq 0$, where

$$\widehat{B}_n(\mathbf{w}) = \sum_{j=0}^{n} \widehat{B}_{n,j}(\mathbf{w})$$

is complete ordinary Bell polynomial in $\mathbf{w}$ *and* $\widehat{B}_{n,j}(\mathbf{w})$ *is the partial ordinary Bell polynomial in* $\mathbf{w}$ *tabled on page 309 of Comtet (1974). Furthermore,* $a_n \otimes b_n = \sum_{j=0}^{n} a_j b_{n-j}$ *is the discrete convolution of* a_n *and* b_n*.*

Proof Define the generating functions

$$U(t) = \sum_{n=0}^{\infty} u_n t^n, \; V(t) = \sum_{n=0}^{\infty} v_n t^n.$$

Multiplying Eq. 2.3 by t^n and summing from $n = 0$ gives $(U(t) - V(t))/t = U(t)V(t)$, so that $U(t) = (1 - W(t))^{-1} V(t)$, where $W(t) = tV(t) = \sum_{n=1}^{\infty} w_n t^n$. By definition, for $j = 0, 1, \ldots$

$$W(t)^j = \sum_{n=j}^{\infty} \widehat{B}_{n,j}(\mathbf{w}) t^n.$$

So,

$$(1 - W(t))^{-1} = \sum_{n=0}^{\infty} \widehat{B}_n(\mathbf{w}) t^n.$$

So, Eq. 2.4 follows. □

Some simplification for $\widehat{B}_{n,j}(\mathbf{w})$ follows using $w_1 = 1$: set $\delta(t) = V(t) - 1 = \sum_{n=1}^{\infty} v_n t^n$. Then $\delta(t)^k = \sum_{m=k}^{\infty} \widehat{B}_{m,k}(\mathbf{v}) t^m$, $k \geq 0$. Now expand $W(t)^j = t^j (1 + \delta(t))^j$ and take the coefficient of t^n to obtain

$$\widehat{B}_{n,j}(\mathbf{w}) = \sum_{k=0}^{j} \binom{j}{k} \widehat{B}_{n-j,k}(\mathbf{v}), \; n \geq j \geq 0.$$

For example, either way one obtains the leading u_n as

$$\begin{aligned}
u_0 &= 1, \; u_1 = v_1 + 1, \\
u_2 &= v_2 + 2v_1 + 1, \\
u_3 &= v_3 + 2v_2 + v_1^2 + 3v_1 + 1, \\
u_4 &= v_4 + 2v_3 + v_2(2v_1 + 3) + 2v_1^2 + 4v_1 + 1, \\
u_5 &= v_5 + 2v_4 + v_3(2v_1 + 3) + v_2(v_2 + 5v_1 + 4) + v_1^3 + 6v_1^2 + 5v_1 + 1, \\
u_6 &= v_6 + 2v_5 + v_4(2v_1 + 3) + 2v_3(v_2 + 2 + 2v_1 + 2) + v_2(3v_2 + 3v_1^2 + 6v_1 + 5) \\
&\quad + 4v_1^3 + 10v_1^2 + 6v_1 + 1.
\end{aligned}$$

In-built routines for computing Bell polynomials are widely available. For example, BellY in Mathematica and IncompleteBellPoly in Matlab.

3 The absolutely continuous case

Our solutions (2.2) and (2.3) do not tell us how u_n behaves for large n. Also they require repeated integration. Here, we give solutions that overcome these problems, see Theorem 3.1. This theorem expresses the distribution of the maximum in terms not depending on repeated integration. The solutions use Fredholm integral theory given in Appendix A in Withers and Nadarajah (2011). Write Eq. 1.4 in the form $\mathcal{K}r(y) = \int K(y,z)r(z)dz$, where $K(y,z) = \rho I(x \le y + \rho z) f(x - \rho z)$. Since

$$\begin{aligned} ||\mathcal{K}||_2^2 &= \int\int K(y,z)K(z,y)dydz \\ &= \rho^2 \int\int I(x < y + \rho z) I(x < z + \rho y) f(x - \rho z) f(x - \rho y) dydz \\ &< \rho^2 \int\int f(x - \rho z) f(x - \rho y) dydz = \rho^2, \end{aligned} \tag{3.1}$$

$K(y,z)$ is said to be a Fredholm kernel with respect to Lebesgue measure, allowing the Fredholm theory of Appendix A of Withers and Nadarajah (2011) to be applied, in particular the functional forms of the Jordan form and singular value decomposition. If say, $0 < \rho < 1$, then one can show that $||\mathcal{K}||_2^2 = \int F(x_t)dF(t) \uparrow 1$ as $x \uparrow \infty$, where $x_t = \min(x - \rho t, (x - t)/\rho)$. Let $\{\lambda_j, r_j, l_j \ : \ j \ge 1\}$ be its eigenvalues (singular values) and associated right and left eigenfunctions ordered so that $|\lambda_j| \le |\lambda_{j+1}|$. By Appendix A of Withers and Nadarajah (2011) these satisfy $\lambda_j \mathcal{K} r_j = r_j$, $\lambda_j l_j \mathcal{K} = l_j$ and $\int r_j l_k = \delta_{j,k}$, where $\delta_{j,k}$ is the Kronecker function and we write $\int a(y)b(y)dy = \int ab$. So, $\{r_j(y), l_k(y)\}$ are biorthogonal functions with respect to Lebesgue measure. Set $\nu_j = 1/\lambda_j$. By Eq. 3.1 and Eq. A.3 in Withers and Nadarajah (2011), $1 > ||\mathcal{K}||_2^2 = \sum_{j=1}^{\infty} \nu_j^2$, where ν_j are the singular values, or if the Jordan form is diagonal, the eigenvalues. (We shall use these terms interchangeably.) So, $|\nu_j| < 1$ and $1 + \nu_j > 0$.

Theorem 3.1 *The case $\rho > 0$: For $n \ge 1$,*

$$u_n = \sum_{j=1}^{\infty} \beta_j \nu_j^n, \tag{3.2}$$

where

$$\beta_j = r_j(\infty) \int F l_j \text{ or } l_j(\infty) \int F r_j. \tag{3.3}$$

The case $\rho < 0$: For $n \ge 0$,

$$u_n = v_n \otimes w_n = v_n \otimes N_n \otimes D_n (-1)^n / n!, \tag{3.4}$$

where v_n, w_n and N_n are given by

$$v_n = \sum_{j=1}^{\infty} \beta_j \nu_j^n, \tag{3.5}$$

$$1 - tV(t) = N(t)/D(t) \tag{3.6}$$

and

$$N(t)^{-1} = \sum_{n=0}^{\infty} N_n t^n, \tag{3.7}$$

where

$$D(t) = \prod_{j=1}^{\infty} (1 - \nu_j t), \;\; N(t) = \prod_{j=1}^{\infty} (1 - w_j t). \tag{3.8}$$

Furthermore, D_n is given by

$$D_n/n! = \sum_{1 \le j_1 < \cdots < j_n} \nu_{j_1} \cdots \nu_{j_n} = [1^n], \tag{3.9}$$

the augmented symmetric function, in the notation of Table 10 of Stuart and Ord (1987).

Proof Consider the case, where the Jordan form is diagonal. Suppose that the eigenvalue λ_1 of smallest magnitude has multiplicity M (typically 1). Set

$$B = \sum_{j=1}^{M} \beta_j. \tag{3.10}$$

Then, by Eq. A.4 in Withers and Nadarajah (2011), for $n \ge 1$,

$$v_n = \sum_{j=1}^{\infty} \beta_j \nu_j^n = B \nu_1^n (1 + \epsilon_n),$$

where $\epsilon_n \to 0$ exponentially as $n \to \infty$. So, for $n \ge 1$, by Eq. 2.2, for $\rho > 0$, we have Eq. 3.2.

Now suppose that $\rho < 0$. By Eq. 3.5, for $\max_{j=1}^{\infty} |\nu_j t| < 1$, $V(t) = 1 + \sum_{j=1}^{\infty} \beta_j \nu_j t/(1 - \nu_j t)$. ($\sum_{j=1}^{\infty} \beta_j$ may diverge.) So, $1 - tV(t)$ can be written as Eq. 3.6 with $N(t)$ and $D(t)$ taking the forms given by Eq. 3.8. Note that $D(t)$ is the Fredholm determinant of $K(x, y)$. So, by Eqs. 3.7, 3.8 and the partial fraction expansion, assuming that $\{w_j\}$ are all different,

$$N_n = \sum_{j=1}^{\infty} c_j^{-1} w_j^n,$$

where

$$c_j = \prod_{k \ne j} (1 - w_k/w_j).$$

Also by Fredholm's first theorem—see, for example, Pogorzelski (1966, page 47),

$$D(t) = 1 + \sum_{n=1}^{\infty} D_n (-t)^n / n!,$$

where D_n is given by Eq. 3.9. Table 10 of Stuart and Ord (1987) gives $[1^n]$ in terms of the power sums $(r) = \sum_{j=1}^{\infty} \nu_j^r$. For example, $[1^3] = 2(3) - (2)(1) + (1)^3$. In our case

$$(r) = \int K_r(x, x)dx = \sum_{j=1}^{\infty} \nu_j^r,$$

where, by Eq. A.2 in Withers and Nadarajah (2011),

$$K_r(x, y) = \mathcal{K}^{r-1} K(x, y) = \sum_{j=1}^{\infty} \nu_j^r r_j(x) l_j(y).$$

So, $[1^n]$ has the form

$$[1^n] = \sum_{k=1}^{n} \sum_{n_1+\cdots+n_k=n} A(n_1, \ldots, n_k)(n_1) \cdots (n_k).$$

We have $(1 - tV(t))^{-1} = D(t)/N(t)$ so that $w_n = N_n \otimes D_n(-1)^n/n!$, giving finally Eq. 3.4. □

Note that ν_1 is given by Eq. A.6 in Withers and Nadarajah (2011) with μ Lebesgue measure. When $\rho = 0$ then Eq. 3.2 holds with $\beta_j = \delta_{j,1}$, $\nu_1 = F(x)$. So, we expect that $\nu_1 \to F(x)$ as $\rho \downarrow 0$.

Note that Eq. 3.4 does not give its behavior for large n. However, Eq. 3.4 will be useful for large n if D_n also has an expansion of the form (3.5). To date we have not been able to show this directly. One can show that $D_n = (-1)^n B_n(\mathbf{d})$, where $B_n(\mathbf{d})$ is the complete exponential polynomial, $d_r = -(r-1)!w(n)$, and $w(n) = \sum_{j=1}^{\infty} w_j^n = (n)$ for w. We conjecture that if $d_n = \sum_{j=1}^{\infty} a_j w_j^n$, where $|w_j|$ is strictly decreasing and $|a_j| > 1$, then $B_n(\mathbf{d}) \approx d_n \approx a_1 w_1^n$ as $n \to \infty$.

An alternative approach (for finding an explicit expression for u_n in the case $\rho < 0$) is to try a solution for u_n of the form (3.5), say

$$u_n = \sum_{j=1}^{\infty} \gamma_j \delta_j^n, \tag{3.11}$$

where δ_j decrease in magnitude. Assuming that $\{\delta_j, \nu_j\}$ are all distinct, substitution into the recurrence relation (2.3) gives us the following elegant relations. Note that $\{\delta_j\}$ are the roots of

$$\sum_{k=1}^{\infty} \beta_k/(\delta - \nu_k) = 1$$

and β_k are given by Eq. 3.3. Having found $\{\delta_j\}$, $\{\gamma_j\}$ are the roots of

$$\sum_{j=1}^{\infty} \gamma_j/(\delta_j - \nu_k) \equiv 1.$$

The last equation can be written $\mathbf{A}\gamma = \mathbf{1}$, where $A_{k,j} = 1/(\delta_j - \nu_k)$ and $\mathbf{A} = (A_{k,j} : k, j \geq 1)$. So, a formal solution is $\gamma = \mathbf{A}^{-1}\mathbf{1}$. Numerical solutions can be found by truncating the infinite matrix $\mathbf{A}$ and infinite vectors $\mathbf{1}, \gamma$ to $N \times N$ matrix and N-vectors, then increasing N until the desired precision is reached.

For $\rho > 0$, Eq. 3.2 implies $u_n \approx B\nu_1^n$ and $\nu_1 > 0$, where B is given by Eq. 3.10. Also ν_1 is given by Eq. A.6 in Withers and Nadarajah (2011) with Lebesgue measure μ. Now suppose that $\rho < 0$. By Eq. 3.11, $u_n = \gamma_1\delta_1^n(1 + \epsilon_n')$, where $\epsilon_n' \to 0$ exponentially as $n \to \infty$ and δ_1 has the largest magnitude among $\{\delta_n\}$. The case where multiple δ_n exist of magnitude $|\delta_1|$ requires an obvious adaptation.

4 Expressions for eigenfunctions and eigenvalues

Theorem 4.1 derives differential equations for the left and right eigenfunctions as well as for the resolvent discussed in Section 3. Theorem 4.2 gives a formal solution of the differential equation for the right eigenfunction. Theorem 4.3 provides a similar solution, giving an equation for the jth left eigenfunction in terms of its value at an arbitrary point, taken as x. Theorem 4.4 derives the equation for the eigenvalues mentioned in the abstract. An alternative to this equation is provided by Theorem 4.5.

Theorem 4.1 *The right eigenfunctions satisfy the non-standard linear first order differential equation*

$$\nu_j \dot{r}_j(y) = \text{sign}(\rho) f(y) r_j((x-y)/\rho) \tag{4.1}$$

with the initial value $r_j(-\infty) = 0$. *Similarly, the left eigenfunctions satisfy*

$$\nu_j (d/dz)[l_j(z)/f(x-\rho z)] = \rho^2 l_j(x-\rho z). \tag{4.2}$$

The resolvent satisfies the first order partial differential equations

$$\begin{aligned}(\partial/\partial y)[\{K(y,z,\lambda) - K(y,z)\}/\lambda] &= \text{sign}(\rho) f(y) K((x-y)/\rho, z, \lambda),\\ (\partial/\partial z)[\{K(y,z,\lambda) - K(y,z)\}/\{\lambda f(x-\rho z)\}] &= \rho^2 K(y, x-\rho z, \lambda).\end{aligned} \tag{4.3}$$

Proof The right eigenfunctions satisfy $\nu_j r_j = \mathcal{K} r_j$; that is,

$$\nu_j r_j(y) = \mathcal{K} r_j(y) = \text{sign}(\rho) \int^y r_j((x-w)/\rho) dF(w). \tag{4.4}$$

For example, $\nu_j r_j(\infty) = \rho \int r_j(z) f(x-\rho z) dz$. Differentiating gives Eq. 4.1. Similarly, the left eigenfunctions satisfy $\nu_j l_j = l_j \mathcal{K}$; that is,

$$\nu_j l_j(z) = \int l_j(y) K(y,z) dy = \rho f(x-\rho z) \int_{x-\rho z} l_j(y) dy. \tag{4.5}$$

So,

$$l_j(-\infty) = 0 \text{ if } \rho > 0,\ l_j(\infty) = 0 \text{ if } \rho < 0, \tag{4.6}$$

and by differentiating, we obtain Eq. 4.2. The resolvent satisfies

$$[K(y,z,\lambda) - K(y,z)]/\lambda = \mathcal{K} K(y,z,\lambda) = K(y,z,\lambda)\mathcal{K}.$$

So, $K(-\infty, z, \lambda) = 0$, $K(y, \infty, \lambda) = 0$ if $\rho < 0$, $K(y, -\infty, \lambda) = 0$ if $\rho > 0$ and by differentiation the resolvent satisfies Eq. 4.3. □

Note that Eq. 4.3 may involve the Dirac function $\delta(x)$ since with $x_z = x - \rho z$, we have $(\partial/\partial y)K(y, z) = \rho f(x_z)\delta(y - x_z)$ and $(\partial/\partial z)K(y, z) = \rho^2 f(x_z)\delta(y - x_z) - \rho^2 I(x_z < y)\dot{f}(x_z)$.

For special cases, it is possible to solve Eq. 4.4 or Eq. 4.5 explicitly. Suppose that $F(y) = a\exp(ay)$ on $(-\infty, 0]$, where $a > 0$, and that $\nu_j < 0$, $\rho < -1$, $y \le x$. Taking $r_j(0) = 1$, a solution of Eq. 4.4 is $r_j(y) = \exp(b_j y)$, where $b_j|\nu_j|/a = \exp(b_j x/\rho)$ and $b_j > 0$.

Theorem 4.2 *Set $r(y) = r_j(y)$. Suppose that f and r have Taylor series expansions about zero. Denote the ith derivatives of $f(y)$ by $f_{\cdot,i}(y)$. Set $f_i = f_{\cdot,i}(0)$, $\mathbf{f}' = (f_0, f_1, \ldots)$ and $\mathbf{Y}'_y = \mathbf{Y}' = (y, y^2/2!, \ldots)$. For $l, k \ge 1$, set $Q_{l,k} = q_{l-1,k}$ with $\mathbf{Q} = (Q_{l,k} : l, k \ge 1)$, where*

$$q_{i,k} = \rho^{-k} \sum_{b=0}^{\min(i,k)} \binom{i}{b} f_{i-b}(-1)^b x^{k-b}/(k-b)!.$$

We have

$$r_j(y)/r_j(0) = 1 + \mathbf{Y}'(d_j\mathbf{I} - \mathbf{Q})^{-1}\mathbf{f}, \tag{4.7}$$

where $d_j = \nu_j$ sign(ρ).

Proof Set $c = \lambda_j$ sign(ρ) and $r_i = r_{\cdot,i}(0)$. Expanding $\dot{r}(y) = cf(y)r((x - y)/\rho)$ about zero, for $i \ge 0$, the coefficient of $y^i/i!$ is

$$r_{i+1} = c \sum_{a+b=i} \binom{i}{a} f_a r_{\cdot,b}(x/\rho)(-\rho)^{-b} = c\sum_{k=0}^{\infty} q_{i,k} r_k = cf_i r_0 + c\sum_{k=1}^{\infty} q_{i,k} r_k.$$

Set $\mathbf{R}' = (r_1, r_2, \ldots)$. So, $\mathbf{R} = \mathbf{f}cr_0 + c\mathbf{QR}$, $\mathbf{R} = (\mathbf{I} - c\mathbf{Q})^{-1}\mathbf{f}cr_0$. But $r(y) - r(0) = \mathbf{Y}'\mathbf{R}$. So, we obtain the jth right eigenfunction in terms of its value at zero: $r(y)/r(0) = 1 + \mathbf{Y}'(c^{-1}\mathbf{I} - \mathbf{Q})^{-1}\mathbf{f}$; that is, Eq. 4.7. □

The formal solution of Eq. 4.1 given by Eq. 4.7 is in terms of $r_j(0)$. The value zero is arbitrary: a similar solution can be obtained in terms of $r_j(y_0)$ for any y_0.

For the extreme value cdf $F(x) = \exp\{-\exp(-x)\}$, $\mathbf{f} = (1/e)(1, 0, -1, -1, -7/288, -31/4, \ldots)'$.

Since r_j is unique only up to a constant multiplier, we may take $r_j(0) \equiv 1$. The solution (4.7) can now be implemented by successive approximations. For $N \ge 1$, set $r_{N,j}(y)/r_j(0) = 1 + \mathbf{Y}'_N(d_j\mathbf{I}_N - \mathbf{Q}_N)^{-1}\mathbf{f}_N$, where $\mathbf{Y}_N, \mathbf{f}_N$ are the first N elements of $\mathbf{Y}, \mathbf{f}$ and $\mathbf{Q}_N$ is the upper left $N \times N$ elements of $\mathbf{Q}$. Then one expects that $r_{N,j}(y) \to r_j(y)$ as $N \to \infty$, giving the jth left eigenfunction.

Theorem 4.3 *Set* $c = \rho^2\lambda_j$, $l = l_j$, $e(y) = f(y)^{-1}$, $\boldsymbol{Y}'_y = \boldsymbol{Y}' = (y, y^2/2!, \ldots)$, $\boldsymbol{D}_r = diag(r^i : i \geq 1)$ *and* $r = -\rho$. *Set* $\boldsymbol{W} = (W_{i,j} : i, j \geq 1)$, *where*

$$W_{i,j} = \sum_{a=0}^{\min(j,i+1)} \binom{i+1}{a} e_{\cdot,i+1-a}(x)(-\rho)^{i+1-a}(-x)^{j-a}/(j-a)!.$$

Finally, set $U_i = W_{i,0} = e_{\cdot,i+1}(x)(-\rho)^{i+1}$. *We have*

$$l(z)/l(x) = 1 + \boldsymbol{Y}'_{z-x}(c\boldsymbol{D}_r - \boldsymbol{W})^{-1}\boldsymbol{U}, \tag{4.8}$$

where the multiplier $l_j(x) = l(x)$ *is determined by Eq.* A.1 *in Withers and Nadarajah (2011).*

Proof By Taylor expansions,

$$l(z)e(x-\rho z) = \sum_{i=0}^{\infty}(z^i/i!) \sum_{a+b=i} \binom{i}{a} l_{\cdot,a}(0)e_{\cdot,b}(x)(-\rho)^b.$$

By Eq. 4.2, l satisfies $(d/dz)[l(z)e(x-\rho z)] = cl(x-\rho z)$. Taking the coefficient of $z^i/i!$, for $i \geq 0$,

$$\sum_{a+b=i+1} \binom{i+1}{a} l_{\cdot,a}(0)e_{\cdot,b}(x)(-\rho)^b = cl_{\cdot,i}(x)(-\rho)^i. \tag{4.9}$$

By another Taylor expansion,

$$l_{\cdot,a}(0) = \sum_{k=0}^{\infty} l_{\cdot,k+a}(x)(-x)^k/k!.$$

So, the left hand side of Eq. 4.9 is $\sum_{j=0}^{\infty} W_{i,j}l_{\cdot,j}(x) = V_i l(x) + (\mathbf{WL})_i$, where $L_j = l_{\cdot,j}(x)$, $\mathbf{L}' = (L_1, L_2, \ldots)$ and $V_j = W_{0,j}$. So, Eq. 4.9, for $i \geq 1$, can be written $\mathbf{U}l(x) + \mathbf{W} = c\mathbf{D}_r\mathbf{L}$ so that $\mathbf{L} = (c\mathbf{D}_r - \mathbf{W})^{-1}\mathbf{U}l(x)$, giving

$$l(z) - l(x) = \mathbf{Y}'_{z-x}\mathbf{L} = \mathbf{Y}'_{z-x}(c\mathbf{D}_r - \mathbf{W})^{-1}\mathbf{U}l(x).$$

That is, $l(z)$ is given by Eq. 4.8. □

Theorem 4.4 *In the notation of Theorem* 4.3, *an equation for the eigenvalues is*

$$\sum_{i=1}^{\infty} w_i(c-\theta_i)^{-1} = c - \theta_0 \tag{4.10}$$

for certain weights $\{w_i\}$, *where* $\{\theta_j,\ j \geq 1\}$ *are the eigenvalues of* $\boldsymbol{W}$.

Proof Substituting into Eq. 4.9 at $i = 0$, that is, $W_{0,0}l(x) + \mathbf{V}'\mathbf{L} = cl(x)$, we obtain $\mathbf{V}'(c\mathbf{D}_r - \mathbf{W})^{-1}\mathbf{U} = c - W_{0,0}$. The roots c of this equation are just $\{\rho^2\lambda_j\}$, so Eq. 4.10 follows. □

If $\{c_{N,j},\ j = 1, \ldots, N+1\}$ are the roots of the N dimensional approximation of Eq. 4.10, say $\mathbf{V}'_N\ (c\mathbf{D}_{r,N} - \mathbf{W}_N)^{-1}\ \mathbf{U}_N = c - W_{0,0}$, a polynomial in c of degree $N+1$, then $c_{N,j} \to c_j = \rho^2\lambda_j$ as $N \to \infty$. Having obtained an eigenvalue, one can substitute it into Eqs. 4.7 and 4.8 to obtain the corresponding eigenfunctions up to constants $l(x)$ and $r(0)$. As noted in Appendix A of Withers and Nadarajah (2011), either of these but not both can be arbitrarily chosen. The conditions $r(-\infty) = 0$ and Eq. 4.6 can be verified numerically.

Suppose that $f = \phi$, the pdf of a standard normal $\mathcal{N}(0, 1)$ random variable. Then $e_{\cdot,j}(x) = \phi(x)^{-1} H_j^*(x)$, where

$$H_j^*(x) = \mathbb{E}\left[(x + \mathcal{N}(0,1))^j\right] = \sum_k \binom{j}{2k} x^{j-2k} m_{2k}$$

is the modified Hermite polynomial and $m_{2k} = (2k)!/k!2^k$ is the $(2k)$th moment of $\mathcal{N}(0, 1)$. See Withers and McGavin (2006). Now, using Theorems 2.2, 3.1, 4.2, 4.3 and the numerical methods discussed above, one can calculate $u_n = P(M_n \le x)$ for the case f is a standard normal pdf. Figures 1 and 2 show plots of u_n for $n = 1000$ and $\rho = -0.9, -0.8, \ldots, 0.8, 0.9$. As expected, the distribution of u_n becomes more dominant as ρ goes from 1 to 0 (Fig. 1) and as ρ goes from -1 to 0 (Fig. 2).

Theorem 4.5 *Set* $c = \rho^2\lambda_j$, $e(y) = f(y)^{-1}$, $\boldsymbol{D}_r = diag(r^i\ :\ i \ge 1)$, $r = -\rho$, $x_k = x^k/k!$, $\boldsymbol{U}' = (x_1, x_2, \ldots)$, $A_{i,a} = l_{\cdot,a}(0)\ [e_{\cdot,b}(x)\ (-\rho)^b]_{b=i+1-a}$, $\boldsymbol{A} = (A_{i,a}\ :\ i, a \ge 1)$ *and* $V_i = A_{i,0} = [e_{\cdot,b}(x)\ (-\rho)^b]_{b=i+1}$. *Set* $\boldsymbol{X}$ *to be a matrix with its* i*th row* $(\boldsymbol{0}^{i-1}, x_0, x_1, \ldots)$ *and* $\boldsymbol{B} = c\boldsymbol{D}_r\boldsymbol{X} - \boldsymbol{A}$, *where* $\boldsymbol{0}^i$ *denotes the row* i*-vector of zeros.*

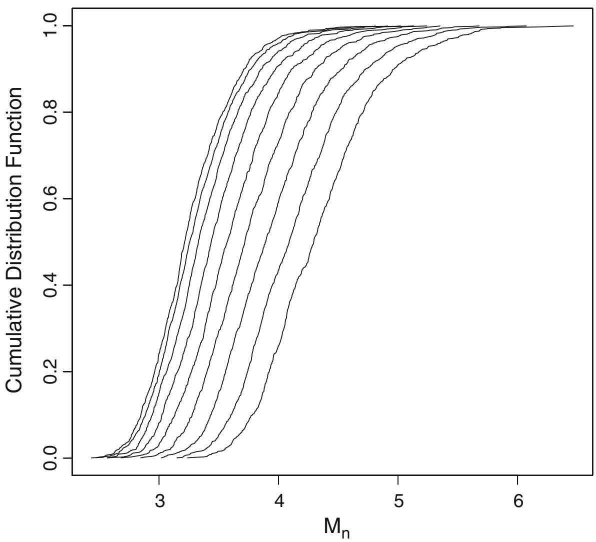

Fig. 1 Plot of $u_n = P(M_n \le x)$ versus x for $n = 1000$ and $\rho = 0.1, 0.2, \ldots, 0.9$ when F is a standard normal cdf. The *curves* from the left to right correspond to increasing values of ρ

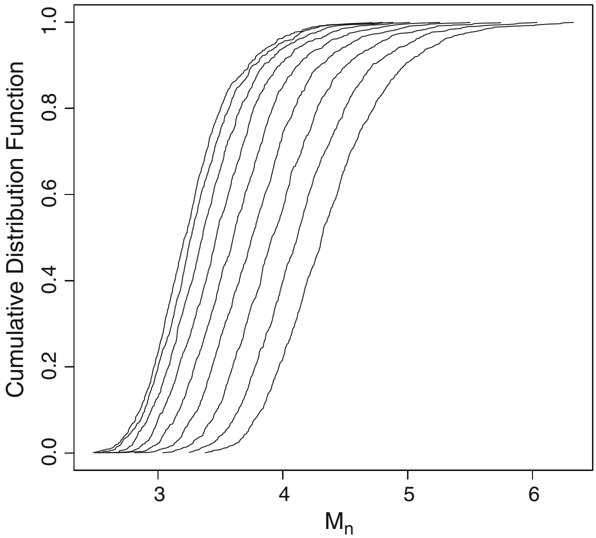

Fig. 2 Plot of $u_n = P(M_n \le x)$ versus x for $n = 1000$ and $\rho = -0.9, -0.8, \ldots, -0.1$ when F is a standard normal cdf. The *curves* from the left to right correspond to increasing values of $-\rho$

We have

$$A_{0,0} + A_{0,1}(\boldsymbol{B}^{-1}\boldsymbol{V})_1 = c + c\boldsymbol{U}'\boldsymbol{B}^{-1}\boldsymbol{V}, \tag{4.11}$$

where $A_{0,0} = -\rho e_{\cdot,1}(x)$ *and* $A_{0,1} = e(x)$.

Proof Expand the right hand side of Eq. 4.9 about $x = 0$, giving $c(-\rho)^i(l(0)+\mathbf{U}'\mathbf{L})$, where $L_j = l_{\cdot,j}(0)$ and $\mathbf{L}' = (L_1, L_2, \ldots)$. For $i \ge 1$, Eq. 4.9 gives

$$V_i l(0) + (A_{i,1}, \ldots, A_{i,i+1}, 0, 0, \ldots)\mathbf{L} = c(-\rho)^i(\mathbf{0}^{i-1}, x_0, x_1, \ldots)\mathbf{L}.$$

That is, $\mathbf{V}l(0) + \mathbf{AL} = c\mathbf{D}_r\mathbf{XL}$. So, $\mathbf{L} = \mathbf{B}^{-1}\mathbf{V}l(0)$, where $l(y)/l(0) = 1 + \mathbf{Y}_y'\mathbf{B}^{-1}\mathbf{V}$ and $\mathbf{Y}_y' = \mathbf{Y}' = (y, y^2/2!, \ldots)$. Note that $\mathbf{X}$ is upper triangular, while $\mathbf{A}$ is lower triangular except for the first super-diagonal. For $i = 0$, Eq. 4.9 gives

$$\sum_{a=0}^{1} A_{0,a} l_a = c \sum_{k=0}^{\infty} l_k x_k = cl(0) + c\mathbf{U}'\mathbf{L}.$$

So, we obtain Eq. 4.11. □

Unfortunately, Appendix A of Withers and Nadarajah (2011) cannot be applied with $\mu = F$ since $\mathcal{K}G(y) = \text{sign }(\rho) \int^y G((x-w)/\rho)\, dF(w)$ is not of the form $\int K(y,z)G(z)dF(z)$. It would be of great interest, and in particular allow a unified approach if Fredholm's theory can be extended to the system $\mathcal{K}\mathcal{O}r = \nu r$, $\mathcal{K}^*\mathcal{O}^*l = \overline{\nu}l$, $l_i^*\mathcal{O}r_j d\mu = \delta_{i,j}$ for $\mathcal{K}$ an $A \times B$ integral operator with kernel $K(y,z) : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ with respect a measure μ, $\mathcal{O}$ a $B \times A$ operator, $*$ the transpose of the complex conjugate, and $\overline{\nu}$ the complex conjugate of ν. For our problem, one could then apply the theory with $\mu = F$, $K(y, z) = \text{sign}(\rho) I(z < y)$, and $\mathcal{O}G(w) = G((x - w)/\rho)$.

5 Application

The aim of this section is to illustrate the usefulness of the derived expansion for $P(M_n \leq x)$. Traditionally, the distribution of M_n is modeled by the generalized extreme value cdf. So, what we seek here is a comparison of models based on the derived expansion and the generalized extreme value cdf.

We use the annual maximum daily rainfall data for the years from 1907 to 2000 for two locations in west central Florida: Clermont and Gainesville. The data were obtained from the Department of Meteorology in Tallahassee, Florida. Withers and Nadarajah (2011) also considered annual maximum daily rainfall data for the years from 1907 to 2000 for west central Florida, but for two different locations.

We would like to emphasize that the aim here is not to provide a complete statistical modeling or inferences for the data sets involved. We refer the readers to Nadarajah (2005) for a comprehensive analysis of the data sets used.

We fitted models based on the derived expansion and the generalized extreme value cdf to the annual maximum rainfall from each of the two locations. The maximum likelihood procedure was used.

Suppose $M_{n,1}, M_{n,2}, \ldots, M_{n,N}$ denote the observations on M_n, where $n = 365$ and N denotes the number of years of data available. Then the likelihood function for the model based on the generalized extreme value cdf is

$$L(\mu, \sigma, \xi) = \prod_{i=1}^{N} \left[\frac{1}{\sigma} \left(1 + \xi \frac{M_{n,i} - \mu}{\sigma} \right)^{-1/\xi - 1} \times \exp \left\{ - \left(1 + \xi \frac{M_{n,i} - \mu}{\sigma} \right)^{-1/\xi} \right\} \right], \tag{5.1}$$

where $-\infty < \mu < \sigma$ is a location parameter, $\sigma > 0$ is a scale parameter, and $-\infty < \xi < \infty$ is a shape parameter.

Let ω denote an intercept parameter for the moving average model in Eq. 1.1. Then the likelihood function for the model based on the derived expansion is

$$L(\rho, \omega) = \prod_{i=1}^{N} \frac{d}{dx} \left[\sum_{j=1}^{\infty} \beta_j \nu_j^n \right] \Bigg|_{x = M_{n,i} - \omega} I\{\rho > 0\} + \prod_{i=1}^{N} \frac{d}{dx} \left[\upsilon_n \otimes N_n \otimes D_n (-1)^n / n! \right] \Big|_{x = M_{n,i} - \omega} I\{\rho < 0\}, \tag{5.2}$$

where β_j, ν_j, υ_n, N_n, D_n are given by Eqs. 3.5, 3.6–3.9, the required right eigenfunctions are given by Eq. 4.7, the required left eigenfunctions are given by Eq. 4.8 and the required eigenvalues are given by Eq. 4.10. The cdf F and its derivatives (required, for example, by Theorems 4.2 and 4.3) were computed empirically. The derivatives of the kind needed in Eq. 5.2 were computed numerically. Numerical methods were used to approximate the various infinite sums, infinite matrices and infinite vectors. The computer code used for implementing the maximum likelihood procedure can be obtained from the corresponding author.

While constructing both Eqs. 5.1 and 5.2, we have assumed $M_{n,1}, M_{n,2}, \ldots, M_{n,N}$ are independent observations. But Durbin and Watson tests (Durbin and Watson 1950, 1951, 1971) show there is a significant evidence of serial dependence in $M_{n,1}, M_{n,2}, \ldots, M_{n,N}$ for both locations. In practice, ignoring serial dependence

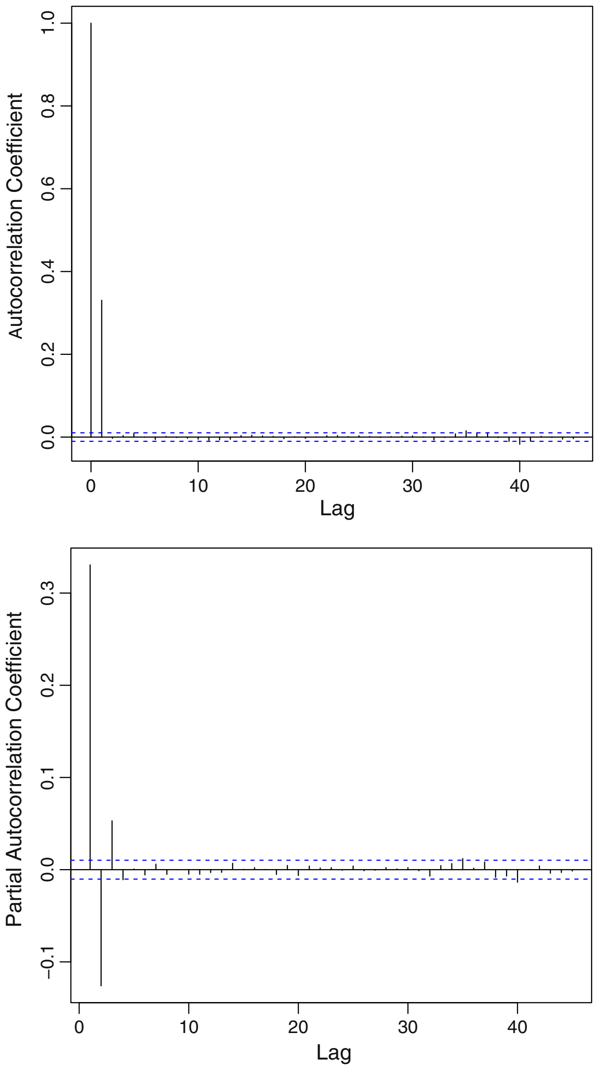

Fig. 3 Autocorrelation coefficient plot (*top*) and partial autocorrelation coefficient plot (*bottom*) for daily rainfall from Clermont

does not affect parameter estimates. Only standard errors are underestimated (Cox and Hinkley 1974). A future work is to see how the two models can be fitted by accounting for serial dependence in $M_{n,1}, M_{n,2}, \ldots, M_{n,N}$.

The optimize function in the R package (R Development Core Team 2012) was used to maximize Eqs. 5.1 and 5.2. The standard errors of the parameter estimates were based on asymptotic normality; that is, they were obtained by inverting the observed information matrices. For each maximization, the optimize function was executed for a wide range of initial values. This sometimes resulted in more than one maximum, but at least one maximum was identified each time. In cases of more than one maximum, we took the maximum likelihood estimates to correspond to the largest of the maxima.

Remarkably, the model based on the derived expansion provided a significantly better fit for Clermont and Gainesville—in spite of the fact the model has only two unknown parameters, ρ and ω, one less than the number of parameters for the generalized extreme value distribution. We now give the details:

- The autocorrelation coefficient plot and the partial autocorrelation coefficient plot for Clermont are shown in Fig. 3. It is clear that the data can be modeled by a first order moving average process. In fact, use of the arma function in the R package produces the residual plot in Fig. 4 which appears reasonable. The model based on the derived expansion yielded $\log L = -320.0$ with the maximum likelihood estimates $\widehat{\rho} = 0.394(0.174)$ and $\widehat{\omega} = 4.014(1.693)$, where the numbers within brackets are the standard errors. The generalized extreme value model yielded $\log L = -341.4$. For the latter model, the maximum likelihood estimates and their standard errors were $\widehat{\mu} = 3.219(0.128)$, $\widehat{\sigma} = 1.186(0.104)$ and $\widehat{\xi} = 0.001(0.073)$.

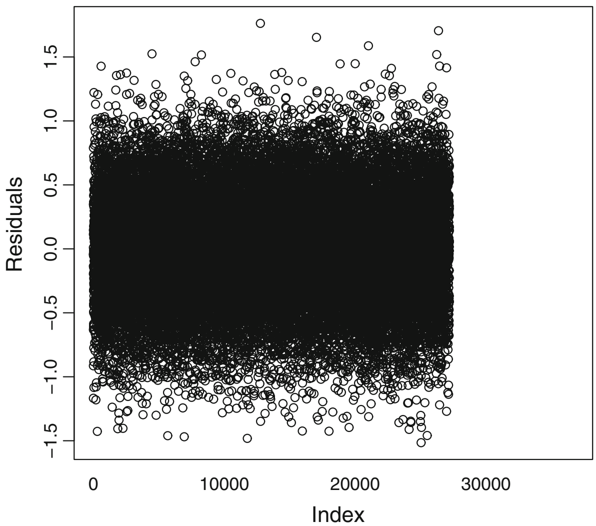

Fig. 4 Residual plot of the fit of a first order moving average to daily rainfall from Clermont

- The autocorrelation coefficient plot and the partial autocorrelation coefficient plot for Gainesville are shown in Fig. 5. It is again clear that the data can be modeled by a first order moving average process, a fact supported by the residual plot in Fig. 6. For this data, the model based on the derived expansion yielded $\log L = -265.8$ with the estimates $\widehat{\rho} = 0.079(0.011)$ and $\widehat{\omega} = 3.635(1.178)$. The generalized extreme value model yielded $\log L = -272.7$. For the latter model, the estimates and their standard errors were $\widehat{\mu} = 3.085(0.092)$, $\widehat{\sigma} = 0.852(0.073)$ and $\widehat{\xi} = 0.002(0.107)$.

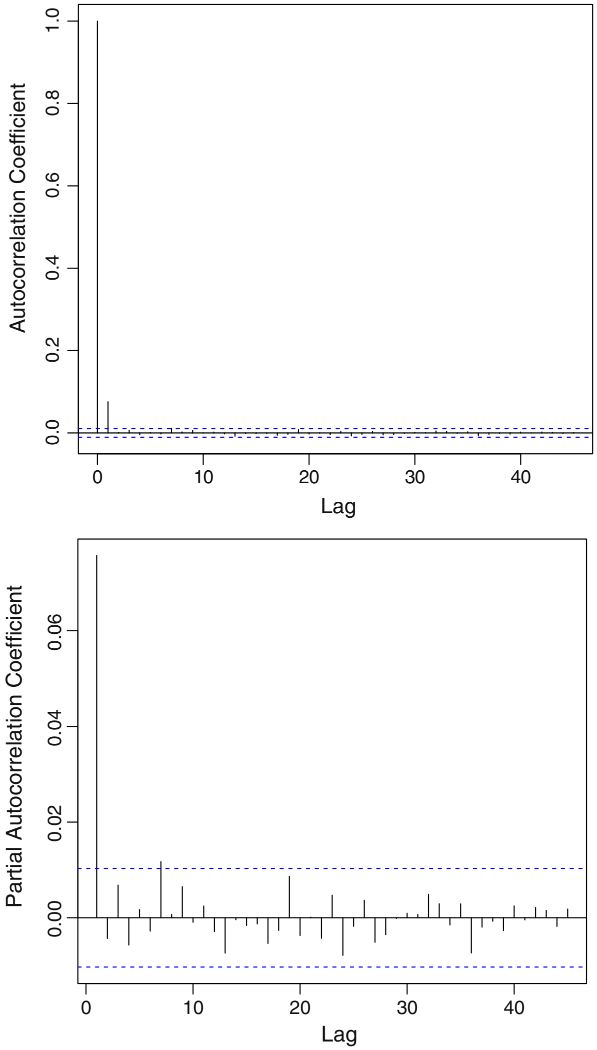

Fig. 5 Autocorrelation coefficient plot (*top*) and partial autocorrelation coefficient plot (*bottom*) for daily rainfall from Gainesville

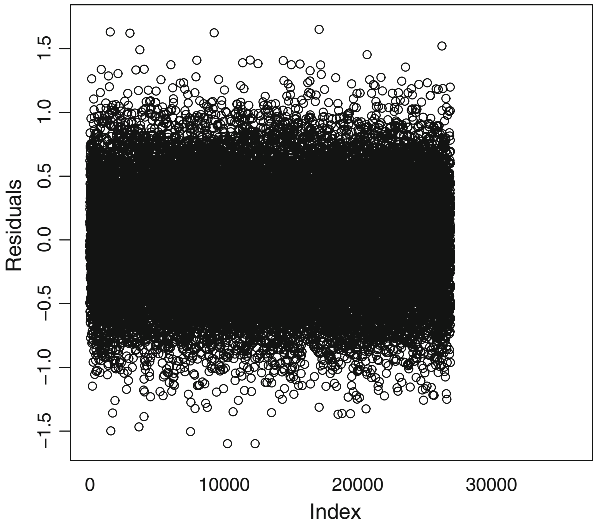

Fig. 6 Residual plot of the fit of a first order moving average to daily rainfall from Gainesville

The conclusion based on the likelihood values can be verified by means of probability-probability plots and density plots. A probability-probability plot plots the observed probabilities against probabilities predicted by the fitted model. The probability-probability plots for the two fitted models and for the two locations are shown in Figs. 7, 8, 9 and 10. We can see that the model based on

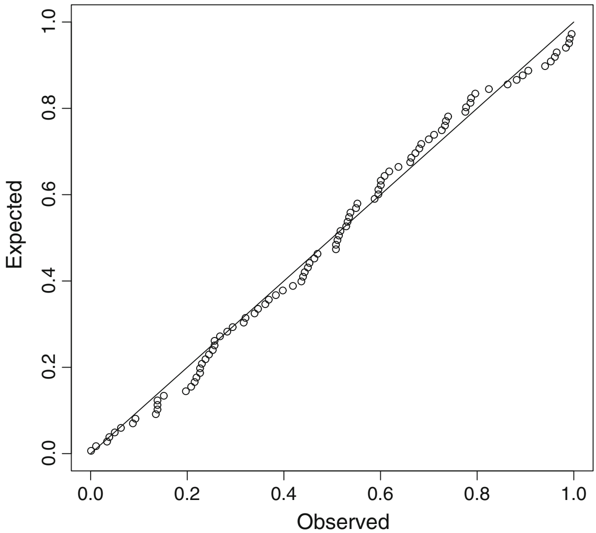

Fig. 7 Probability–probability plot for the fit of the model based on the derived expansion for the annual maximum rainfall data from Clermont

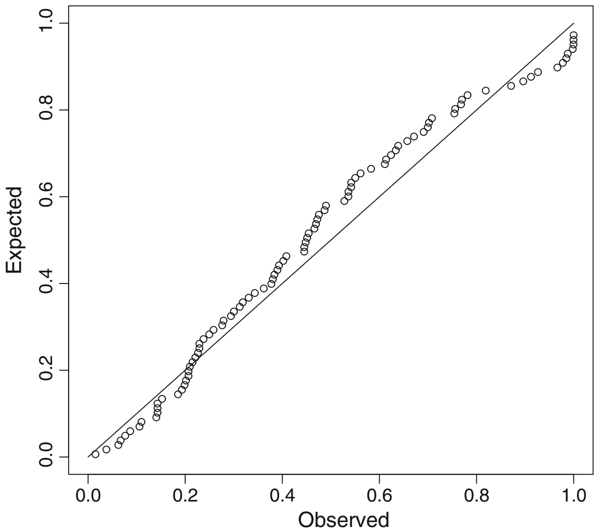

Fig. 8 Probability–probability plot for the fit of the generalized extreme value model for the annual maximum rainfall data from Clermont

the derived expansion has the points much closer to the diagonal line for each location.

A density plot compares the fitted pdfs of the models with the empirical histogram of the observed data. The density plots for the two locations are shown in

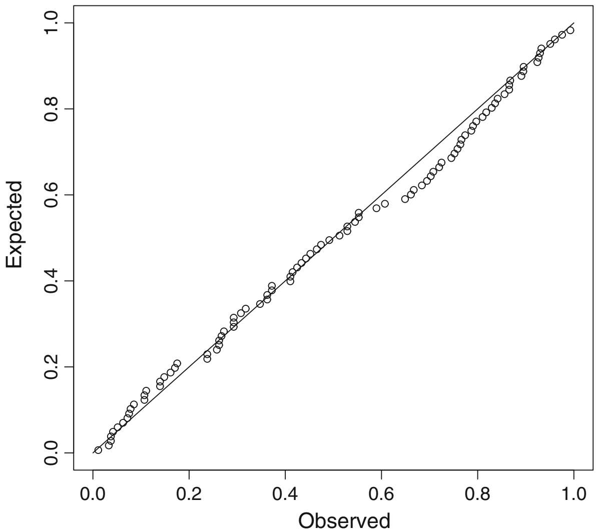

Fig. 9 Probability–probability plot for the fit of the model based on the derived expansion for the annual maximum rainfall data from Gainesville

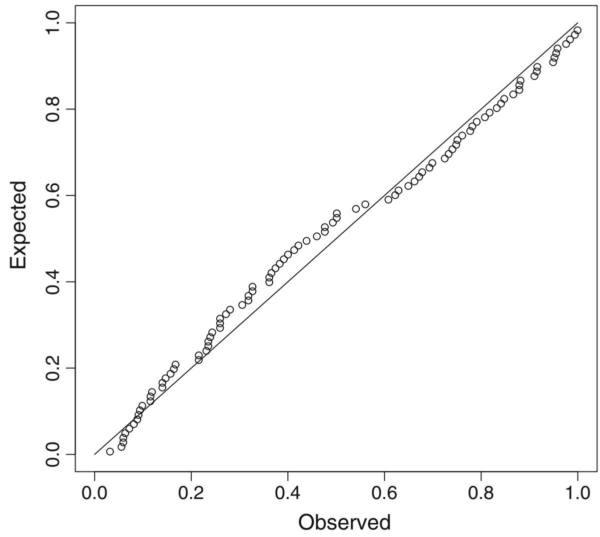

Fig. 10 Probability–probability plot for the fit of the generalized extreme value model for the annual maximum rainfall data from Gainesville

Figs. 11 and 12. Again the fitted pdfs based on the derived expansion appear to capture the general pattern of the empirical histograms much better.

The model based on the derived expansions provides better fits because it is able to capture the features of extreme values for finite n. The generalized extreme value

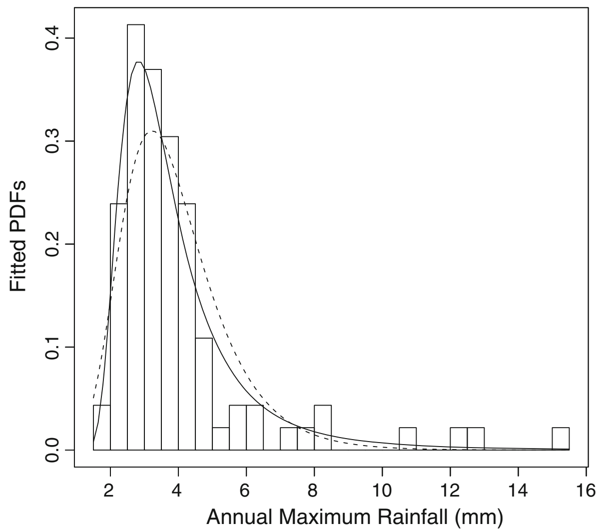

Fig. 11 Fitted pdfs of the two models for the annual maximum rainfall data from Clermont. The *smooth* and *broken curves* correspond to the models based on the derived expansion and the generalized extreme value cdf, respectively

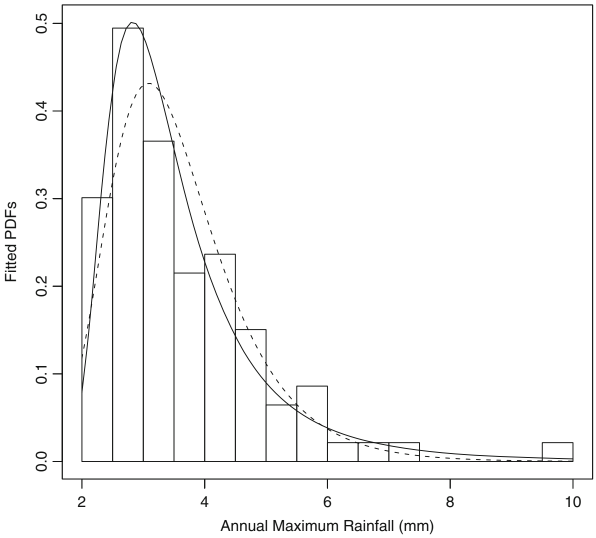

Fig. 12 Fitted pdfs of the two models for the annual maximum rainfall data from Gainesville. The *smooth* and *broken curves* correspond to the models based on the derived expansion and the generalized extreme value cdf, respectively

model assumes n is infinity. It tries to capture the features for finite n by assuming n is infinity. So, the former model can be used to make better predictions of future extreme rainfall.

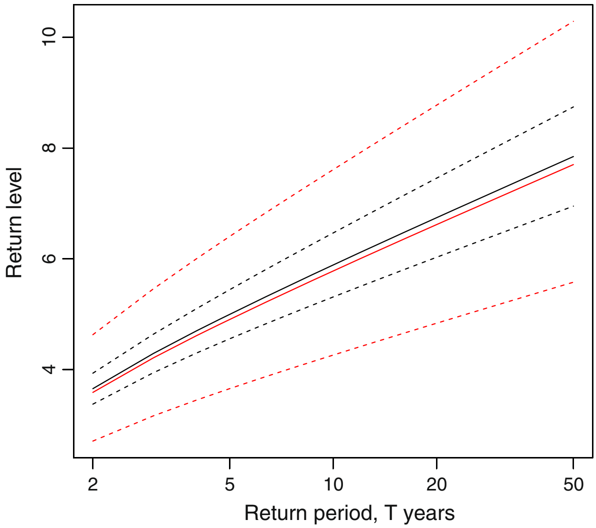

Fig. 13 Return levels and their 95 % confidence intervals for the two models for the annual maximum rainfall data from Clermont. The *curves in black and red* correspond to the models based on the derived expansion and the generalized extreme value cdf, respectively

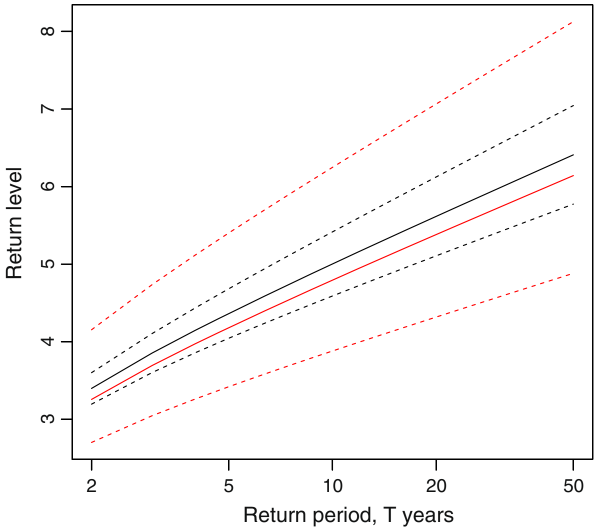

Fig. 14 Return levels and their 95 % confidence intervals for the two models for the annual maximum rainfall data from Gainesville. The *curves in black and red* correspond to the models based on the derived expansion and the generalized extreme value cdf, respectively

Furthermore, the model based on the derived expansions has one less parameter than the generalized extreme value model. So, the predictions can also be expected to be more accurate (for example, narrower confidence intervals, narrower confidence bands, better power functions, and so on) than those based on the generalized extreme value model. The latter model usually leads to distressingly wide confidence intervals for high quantiles. We now illustrate this by computing the return levels. The return level of period T, say x_T, for the generalized extreme value model is

$$x_T = \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1 - 1/T)\}^{-\xi}\right]. \tag{5.3}$$

The return level, x_T, for the derived model is the root of

$$\left.\sum_{j=1}^{\infty} \beta_j \nu_j^n \right|_{x = x_T - \omega} = 1 - 1/T \tag{5.4}$$

if $\rho > 0$ and is the root of

$$v_n \otimes N_n \otimes D_n (-1)^n / n!\big|_{x = x_T - \omega} = 1 - 1/T \tag{5.5}$$

if $\rho < 0$. Plots of Eqs. 5.3, 5.4, and 5.5 versus T are shown in Figs. 13 and 14. The plots also show the 95 % confidence intervals computed by the delta method (Rao 1973, pages 387–389). We see that these intervals are a lot narrower for the derived model. The return level estimates do not appear to differ much between the models.

The proposed model may be computationally more expensive than the generalized extreme value model. But being computationally expensive is not a major problem in this day and age.

Finally, we like to mention that method used to derive the proposed model can be extended for other processes, including processes (for example, long memory processes) for which standard extreme value theory does not apply. This is explained in the next section.

6 Future work

The method of this paper can be applied to derive the exact distribution of M_n for other time series like the: (1) higher order moving average processes; (2) higher order autoregressive processes; (3) ARMA processes; (4) multivariate ARMA processes; (5) long memory processes; (6) discrete versions. The method can also be applied to derive the exact joint distribution of the maximum and minimum for these processes. Another aspect to consider is statistical inference for the derived distributions.

Acknowledgments The authors would like to thank the Editor and the referee for careful reading and for their comments which greatly improved the paper.

References

Albin, J.M.P.: Extremes for non-anticipating moving averages of totally skewed α-stable motion. Stat. Probab. Lett. **36**, 289–297 (1997)

Comtet, L.: Advanced Combinatorics. Reidel, Dordrecht (1974)

Cox, D.R., Hinkley, D.V.: Theoretical Statistics. Chapman and Hall, London (1974)

Davis, R.A., Resnick, S.I.: Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. **13**, 179–195 (1985)

Davis, R.A., Resnick, S.I.: Extremes of moving averages of random variables from the domain of attraction of the double exponential distribution. Stoch. Process. Appl. **30**, 41–68 (1988)

Davis, R.A., Resnick, S.I.: Basic properties and prediction of max-ARMA processes. Adv. Appl. Probab. **21**, 781–803 (1989)

Davis, R.A., Resnick, S.I.: Extremes of moving averages of random variables with finite endpoint. Ann. Probab. **19**, 312–328 (1991)

Davis, R., Marengo, J., Resnick, S.: Extremal properties of a class of multivariate moving averages (with discussion). In: Proceedings of the 45th Session of the International Statistical Institute, vol. 4 (Amsterdam, 1985). Bulletin de l'Institut International de Statistique, **51**, vol. V, pp. 185–192 (1985)

Deheuvels, P.: Extremes of moving averages of iid random-variables. Adv. Appl. Probab. **20**, 10 (1988)

Durbin, J., Watson, G.S.: Testing for serial correlation in least squares regression I. Biometrika **37**, 409–428 (1950)

Durbin, J., Watson, G.S.: Testing for serial correlation in least squares regression II. Biometrika **38**, 159–178 (1951)

Durbin, J., Watson, G.S.: Testing for serial correlation in least squares regression III. Biometrika **58**, 1–19 (1971)

Fasen, V.: Extremes of regularly varying Levy-driven mixed moving average processes. Adv. Appl. Probab. **37**, 993–1014 (2005)

Fasen, V.: Extremes of subexponential Levy driven moving average processes. Stoch. Process. Appl. **116**, 1066–1087 (2006)

Hall, A.: Extremes of integer-valued moving average models with regularly varying tails. Extremes **4**, 219–239 (2001)

Hall, A.: Extremes of integer-valued moving average models with exponential type tails. Extremes **6**, 361–379 (2003)

Hall, A., Moreira, O.: On the extremes of a particular moving average count data model. Stat. Probab. Lett. **76**, 135–141 (2006)
Hall, A., Scotto, M.G.: Extremes of sub-sampled integer-valued moving average models with heavy-tailed innovations. Stat. Probab. Lett. **63**, 97–105 (2003)
Hall, A., Scotto, M.G., Ferreira, H.: On the extremal behaviour of generalised periodic sub-sampled moving average models with regularly varying tails. Extremes **7**, 149–160 (2004)
Hall, A., Scotto, M.G., Cruz, J.: Extremes of integer-valued moving average sequences. Test **19**, 359–374 (2010)
Klüppelberg, C., Lindner, A.: Extreme value theory for moving average processes with light-tailed innovations. Bernoulli **11**, 381–410 (2005)
Leadbetter, M.R., Lindgren, G., Rootzén, H.: Extremes and Related Properties of Random Sequences. Springer, New York (1983)
Martins, A.P., Ferreira, H.: Extremes of periodic moving averages of random variables with regularly varying tail probabilities. SORT **28**, 161–175 (2004)
Nadarajah, S.: Extremes of daily rainfall in west central Florida. Clim. Change **69**, 325–342 (2005)
Nadarajah, S., Mitov, K.: Asymptotics of maxima of discrete random variables. Extremes **5**, 287–294 (2002)
O'Brien, G.L.: Extreme values for stationary and Markov sequences. Ann. Probab. **15**, 281–291 (1987)
Park, Y.S.: Extreme value of moving average processes with negative binomial noise distribution. J. Korean Stat. Soc. **21**, 167–177 (1992)
Pogorzelski, W.: Integral Equations and their Applications, vol. 1. Pergamon, Oxford (1966)
R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (2012)
Rao, C.R.: Linear Statistical Inference and its Applications, 2nd edn. Wiley, New York (1973)
Resnick, S.I.: Extreme Values, Regular Variation, and Point Processes. Springer, New York (1987)
Resnick, S.I.: Moving averages—convolutions of distributions which are both subexponential and in a domain of attraction of an extreme value law. Adv. Appl. Probab. **20**, 5 (1988)
Rootzén, H.: Extremes of moving averages of stable processes. Ann. Probab. **6**, 847–869 (1978)
Rootzén, H.: Extreme value theory for moving average processes. Ann. Probab. **14**, 612–652 (1986)
Scotto, M., Ferreira, H.: Extremes of deterministic sub-sampled moving averages with heavy-tailed innovations. Appl. Stoch. Model. Bus. Ind. **19**, 303–313 (2003)
Stuart, A., Ord, K.: Kendall's Advanced Theory of Statistics, vol. 1, 5th edn. Griffin, London (1987)
Withers, C.S., McGavin, P.: Expressions for the normal distribution and repeated normal integrals. Stat. Probab. Lett. **76**, 479–487 (2006)
Withers, C.S., Nadarajah, S.: The distribution of the maximum of a first order autoregressive process: the continuous case. Metrika **74**, 247–266 (2011)