On spatial extremes: with application to a rainfall problem

Chen Zhou
Tinbergen Institute, Erasmus University Rotterdam. P.O. Box 1738, 3000DR Rotterdam, The Netherlands.

Keywords: Spatial extremes, max-stable process, areal reduction factor.
AMS: 60G70;62M30

Abstract

We consider daily rainfall observations at 32 stations in the province of North Holland (The Netherlands) during 30 years. Let Q be the *total* rainfall in this area on one day. An important question is: what is the amount of rainfall Q that is exceeded once in 100 years? This is clearly a problem belonging to extreme value theory. Also it is a genuinely spatial problem.

Recently, a theory of extremes of continuous stochastic processes has been developed. Using the ideas of that theory and much computer power (simulations) we have been able to come up with a reasonable answer to the question above.

1. Introduction

Extreme rainfall statistics are frequently used when a damaging flood has occurred to answer questions about the rarity of the event. Engineers often need extreme rainfall statistics for the design of structures for flood protection. A typical question is e.g. what is the amount of rain in a given area on one day that is exceeded once in 100 years? Or, more mathematically, what is the 100-year quantile of the total rainfall in the area on one day? In this paper this question is investigated for a low-lying flat area in the northwest of the Netherlands. Because it roughly covers the province of North Holland, it will shortly be indicated as North Holland.

There are 32 rainfall stations in the area for which daily data were available for the 30-year period 1971-2000. Only the fall season, i.e. the months September, October and November, is considered. In this season the likelihood of flooding and its impact are relatively large. Because of the restriction to the fall season it is reasonable to assume stationary in time. Stationary in space, except for location and scale, is also assumed.

Since we have to extrapolate from a 30-year to a 100-year period, our problem is an extreme value problem - in the absence of detailed and tractable physical models.

In section 2, we specify the stochastic process used in the simulation. This process is used only to simulate "extreme" rainfall. For non-extreme rainfall we sample from the available data. In section 3 we explain how we combine the two to get a simulated day of rainfall. Section 4 discusses the outcome of the simulation and the answer to our problem.

2. Extremes of Continuous Stochastic Processes

We start with the setup of Extreme Value Theory for continuous stochastic process. Let $\{X(s)\}_{s\in[0,1]}$ be a stochastic process in $C[0,1]$. Consider independent copies $X_1, X_2, \cdots$ of the process X. Compose the continuous stochastic processes

$$\left\{\max_{1\le i\le n} X_i(s)\right\}_{s\in[0,1]}.$$

Suppose that for some positive functions $a_s(n)$ and real functions $b_s(n)$, the sequence of processes

$$\left\{\max_{1\le i\le n} \frac{X_i(s)-b_s(n)}{a_s(n)}\right\}_{s\in[0,1]}$$

converges in $C[0,1]$. If this is the case, we say $X \in \mathcal{D}$. Let us call the limiting process $\{U(s)\}_{s\in[0,1]}$. Then we say $X \in \mathcal{D}(U)$. The following proposition is useful for our purposes (de Haan and Lin [2]).

Proposition 2.1. *$X \in \mathcal{D}$ if and only if the following two statements hold:*
1. (uniform convergence of the marginal distributions) There exists a continuous function $\gamma(s)$ such that, for $x > 0$

$$\lim_{n\to\infty} P\left(\max_{1\le i\le n} \frac{X_i(s)-b_s(n)}{a_s(n)} \le \frac{x^{\gamma(s)}-1}{\gamma(s)}\right) = \exp\left(-\frac{1}{x}\right),$$

uniformly for $s \in [0,1]$.
2. (convergence of the standardized process) With $F_s(x) := P(X(s) \le x)$ for $s \in [0,1]$,

$$\left\{\max_{1\le i\le n} \frac{1}{n(1-F_s(X_i(s)))}\right\} \xrightarrow{d} \{\eta(s)\} \quad \textit{(say)}$$

in $C[0,1]$, where η is a simple max-stable *process. Note that all one-dimensional marginal distributions of the process $1/(1-F_s(X_i(s)))$ are equal to $1-1/x$, $x \ge 1$.*

Remark 2.1.

$$\{U(s)\} \stackrel{d}{=} \left\{\frac{(\eta(s))^{\gamma(s)}-1}{\gamma(s)}\right\}.$$

As a consequence of this proposition, we can study the "simple" process η first and go back to U later, in a straightforward way. To simulate $U(s)$ is the way to produce more "extreme observations" which may even beyond the sample.

Two relatively simple characterizations of simple max-stable processes are known. One of them can serve our purposes. It is given in the following proposition.

Proposition 2.2. *(Schlather [5], de Haan and Ferreira [1] Corollary 9.4.5) Every simple max-stable process* η *in* $C[0,1]$ *can be generated in the following way. Consider a Poisson point process on* $(0,\infty]$ *with mean measure* $r^{-2}dr$*. Let* $\{Z_i\}_{i=1}^{\infty}$ *be a realization of this point process. Further consider i.i.d. positive stochastic processes* $V, V_1, V_2, \cdots$ *in* $C[0,1]$ *with* $EV(s) = 1$ *for all* $s \in [0,1]$ *and* $E\sup_{0\le s\le 1} V(s) < \infty$*. Let the point process and the sequence* $V, V_1, V_2, \cdots$ *be independent. Then*

$$\eta \stackrel{d}{=} \bigvee_{i=1}^{\infty} Z_i V_i.$$

Conversely, each process with this representation is simple max-stable.

Hence we can simulate the simple max-stable process by simulating a Poisson point process.

3. Stochastic Process for Simulating "Extreme" Rainfall

The starting point for the simulation of the rainfall process is the representation of simple max-stable processes in Proposition 2.2. For V we choose the so-called exponential martingale (c.f. Øksendal [4], excercise 4.10). Also we have to extend the process to a process with a two-dimensional index set. We choose the model

$$\eta(s_1, s_2) := \bigvee_{i=1}^{\infty} Z_i \exp\{W_{1i}(\beta s_1) + W_{2i}(\beta s_2) - \beta(|s_1| + |s_2|)/2\} \qquad (1)$$

for $(s_1, s_2) \in \mathbb{R}^2$ (or rather the area under study, North Holland). Here $\{Z_i\}$ is the realization of a Poisson point process on $(0,\infty)$ with mean measure $r^{-2}dr$. The processes $W_{11}, W_{21}, W_{12}, W_{22}, W_{13}, W_{23}, \cdots$ are independent copies of double-sided Brownian motions W defined as follows. Take two independent Brownian motions B_1 and B_2. Then

$$W(s) := \begin{cases} B_1(s), & s \ge 0; \\ B_2(-s), & s < 0. \end{cases} \qquad (2)$$

The positive constant β reflects the amount of spatial dependence at high levels of rainfall, it can be estimated via two dimensional marginal distributions. The process η satisfies the requirements of Proposition 2.2. By Proposition 2.2, the one-dimensional marginal distributions of (1) are all $e^{-1/x}, x > 0$. The process is shift stationary as it should be for our application.

For the simulation of our process we need to simulate a Poisson point process. Simulating a Poisson point process is laborious. We use a simplified way which serves our purpose.

We have now a simple max-stable process that can be simulated rather well. But in fact we need a process that has generalized Pareto marginals, not the standard Fréchet extreme value distribution. Hence we use the process η from (1) but transform the marginal distributions to the generalized Pareto distribution $1 - 1/x$, $x \geq 1$:

$$\xi(s_1, s_2) := \frac{1}{1 - \exp\left\{-\frac{1}{\eta(s_1,s_2)}\right\}} \tag{3}$$

for (s_1, s_2) in the area.

The last step is a further transformation of the marginal distribution that adapts the process to the local shape (γ), scale (a) and shift (b) parameters. These parameters can be estimated from each station separately, using the local sample. For the shape parameter, we use the average of the local estimates of γ. We found the value $\hat{\gamma} = 0.1082$.

The final transformation results into the process

$$X(s_1, s_2) := \hat{a}_{(s_1,s_2)}(n/k)\left(\frac{\xi(s_1, s_2)^{\hat{\gamma}_{n,k}} - 1}{\hat{\gamma}_{n,k}}\right) + \hat{b}_{(s_1,s_2)}(n/k). \tag{4}$$

The process (4) provided the simulated (extreme) rainfall in the area.

4. Simulating a Day of Rainfall

On an arbitrary day, there will be "extreme" rainfall in part of the area and "non-extreme" rainfall (or no rainfall at all) in the rest of the area. We achieve this in the simulation as follows: on the one hand, we simulate the process (4) for the whole area; on the other hand, we choose at random a day out of the 30*(30+31+30)=2730 days of observed rainfall and we connect the two as follows:

For each station we check whether the observed rainfall on the chosen day is larger than the shift parameter $\hat{b}_{(s_1,s_2)}(n/k)$ for that station. If so, we use (4) (i.e., the simulated process) to get the rainfall at that station. If not, we just use the observed rainfall for the chosen day at that station.

We extend this to obtain the rainfall in the entire area. First we connect the monitoring stations with each other, so as to cover the area with Triangles in a specific way. We write Triangles since later on we shall also deal with smaller triangles, also we write Vertex and Edge for a vertex and edge of a Triangle. Any Triangle can be extreme or non-extreme.

1. Non-extreme: this is the case if all Vertices of the Triangle are non-extreme. The rainfall in such a Triangle is just a linear function whose value at the Vertices are the observed values.

2. Extreme: all other cases. In that case the rainfall is mainly determined by the process (4) where the functions $a_{(s_1,s_2)}(n,k)$ and $b_{(s_1,s_2)}(n,k)$ on the Triangle are chosen as linear functions whose values at the Vertices are the values obtained by local estimation. Triangles are divided into small triangles to get a detailed simulation and numerical integration of the areal rainfall.

This is the way we obtained a day of rainfall. Note that the process is continuous and that it is easy to integrate numerically.

5. Result

Our purpose is to study extremes of the total rainfall in North Holland. In particular we want to determine how severe the areal rainfall is that occurs once in 100 years. In other words, we are studying the 1-1/9100 quantile of the daily total rainfall in the area. This quantile will be briefly indicated as the 100-year quantile.

Before presenting the simulation result, we would like to introduce some statistics and results for separate stations. Take Station 251 - West Beemster - as an example (it is located in the middle of the area, and considered as the origin point when simulating the dependence process). By fitting the GPD with shape parameter $\hat{\gamma} = 0.1082$ to the observed extreme daily rainfall amounts at West Beemster, we can estimate the 1-1/9100 quantile for this station. The point estimator is 63.0 mm. It can also be done for the other stations to find the 1-1/9100 quantile in each monitoring station. We can get that the average 1-1/9100 quantile among all the stations is 66.9 mm.

The simulation procedure in Section 4 has been repeated 91,000 times. This results in a sample of 91,000 days rainfall in North Holland. For each day we calculate the total rainfall as the numerical integral of the rainfall process on the area. We take the 10th largest order statistic of this sample, i.e. we determine the 1-1/9100 sample quantile of the integrated rainfall. Dividing by the total area, 2010 km^2, we get the average rainfall in the area. We replicate this procedure 60 times. The sample mean of the simulated quantiles is 58.8 mm, with sample standard deviation 3.16 mm. Hence the standard deviation of the sample mean is 0.41 mm.

The quantile for the area-average rainfall is thus smaller than the average of the corresponding quantile for the individual measuring stations. The areal reduction factor ARF is the ratio of these two quantities, $ARF = 58.8/66.9 = 0.88$. It is remarkable that from the graph in the UK Flood Studies Report (see NERC [3]), a similar value of ARF is found for an area of 2010 km^2.

Acknowledgements: We are grateful for the thorough comments of three referees, the associate editor and the editor. We thank J. Nellestijn for producing figures in the paper

6. Bibliography

[1] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2006.

[2] L. de Haan and T. Lin. On convergence towards an extreme value distribution in C[0,1]. *Ann. Prob.*, 29:467–483, 2001.

[3] Natural Environmental Research Council (NERC). Flood studies report, Vol. II meteorological studies. *Cent. for Ecol. and Hydrol.*, Wallingford , U.K., 1975.

[4] B. Øksendal. *Stochastic differential equations*. Springer, third edition, 1992.

[5] M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5:33–44, 2002.