ON SPATIAL EXTREMES: WITH APPLICATION TO A RAINFALL PROBLEM

By T.A. Buishand, L.DE Haan and Zhou

Presented by Zhitao Zhang, Apr28, 2009

1. Introduction and Data description

When a damaging flood has occurred extreme rainfall statistics are frequently used 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 what the amount of rain in a given area is on one day that is exceeded once in 100 years. In this paper, the authors ran simulations based on a stochastic process which generated by modified exponential martingale and gave a prediction of the 100-year quantile.

The authors collected rainfall observations from 32 stations in the province of North Holland for which daily data were available for the fall season in 30-year period from 1971 to 2000.

2. Stochastic process for simulation

Max-stable processes are used to obtain the quantiles of the spatially aggregated rainfall. Here they follow a different approach based on random fields. Apart from the parameters that characterized the upper tail of the marginal distributions, their model has a parameter that determines spatial dependence. This parameter is estimated from the tails of the empirical two-dimensional marginal distributions of daily rainfall in North Holland. The 100-year quantile of the total rainfall over the area is found by simulating synthetic daily rainfall fields using the estimated model.

First, define stochastic processes V_i as below:

$V_i = exp(+W_{1i}(\beta s_1) + W_{2i}(\beta s_2) - \beta(|s_1| + |s_2|)/2)$ where s_1,s_2 are coordinates of a point in the area, and

$$W(s) = \begin{cases} B_1(s), & \text{if } s \geq 0 \\ B_2(-s), & \text{otherwise.} \end{cases}$$

Here, $B_i(s)$ are independent Brownian motions.

Then by Schlather's theorem (see Schlather([2002])

$\eta = \max \frac{V_i}{\sum_{j=1}^{i} E_j}$ is a max-stable process.

where E_i's are i.i.d. standard exponential random variables.

In order to generate extreme values, the above process need to be transformed so that it has a Generalized Pareto Distribution. So they define

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

and

$$X(s_1, s_2) := a(s_1, s_2)\frac{\xi(s_1, s_2)^{\gamma} - 1}{\gamma} + b(s_1, s_2)$$

Then

$$P(\frac{X(s_1, s_2) - b(s_1, s_2)}{a(s_1, s_2)} > x)$$

$$= P(\frac{\xi(s_1, s_2)^{\gamma} - 1}{\gamma} > x)$$

$$= P(\eta(s_1, s_2) > \frac{1}{-log(1 - (1 + \gamma x)^{-\frac{1}{\gamma}})})$$

$$=(1 + \gamma x)^{-\frac{1}{\gamma}}$$

Hence, all marginal distributions are GPD.

3. Simulation

On an arbitrary day, there will be "extreme" rainfall in part of the area and "nonextreme" rainfall (or no rainfall at all) in the rest of the area.

They achieve this in the simulation as follows: on the one hand, they simulate the process for the whole area; on the other hand, they choose at random a day out of the 30(30+31+30) = 2730 days of observed rainfall and they connect the two as follows:

For each station they check whether the observed rainfall on the chosen day is larger than the shift parameter b(s1,s2) for that station. If so, they use the simulated process to get the rainfall at that station. If not, they just use the observed rainfall for the chosen day at that station.

First they connect the monitoring stations with each other, so as to cover the area with Triangles.Any Triangle can be extreme or nonextreme.

1. Nonextreme: this is the case if all Vertices of the Triangle are nonextreme. The rainfall in such a Triangle is just a linear function whose values at the Vertices are the observed values.

2. Extreme: all other cases. In that case the rainfall is mainly determined by the process where the functions a(s1,s2) and b(s1,s2) on the Triangle are chosen as linear functions whose values at the Vertices are the values obtained by local estimation.

More specifically they proceed as follows:

2a. Subdivide each Edge into d intervals of equal length. Connect the separating points on the Edges with each other using lines parallel to the Edges.

2b. Next they determine the rainfall process in each vertex (i.e., vertex of a triangle). For Vertices they already determined the process. For the vertices, there are two cases.

2b.1. On an Edge connecting two nonextreme Vertices in an extreme Triangle, the rainfall is chosen to be the linear function whose values at the Vertices are the observed values. This determines the rainfall for all vertices on such an Edge.

2b.2. The rainfall for every other vertex in an extreme Triangle is determined by the stochastic process.

2c. In order to carry out the numerical integration, they simplify the rainfall process on each triangle in an extreme Triangle. The rainfall in each triangle is given as a linear function whose value at the vertices is the one obtained in part 2b.

The simulation procedure has been repeated 91,000 times. This results in a sample of 91,000 days of rainfall in North Holland. For each day they 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, that is, they determine the 1-1/9100 sample quantile of the integrated rainfall. Dividing by the total area, 2010 km2, they get the average rainfall in the area. We replicate this procedure 60 times.

4. Results

The simulation result shows that the mean 100-year quantiles of area-average rainfall is 58.8mm with a standard deviation 3.16mm, and the minimum value is 54.4mm while the maximum is 66.7.

5. Comments

A setback of the exponential martingale model is the dependence on the coordinate axes. In order to see how important the choice of the axes is, they have repeated the analysis after a 45 degree rotation of the axes and the result does not change a lot (from 10 simulated 100-year quantiles, they get a sample mean of 58.2 mm, and a sample standard deviation of 3.1 mm). After rotation, one of the axes is more or less the prevailing wind direction.

The quantile for the area-average rainfall is smaller than the average of the corresponding quantile for the individual measurement 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 , a similar value of ARF is found for an area of 2010 km^2. The latter refers to annual maximum rainfall rather than seasonal maximum rainfall. The ARF from the GPD fit to the extreme average daily rainfall equals 57.8/66.9 = 0.86.