

Rainfall Derivatives

Despite the high exposure of business to rainfall risk, most weather derivatives that have been traded have been based on a temperature index. Nevertheless many end users such as farmers or hydroelectric generators are very sensitive to rainfall magnitude and frequency. The consequences of too much or not enough rain are spread widely and directly or indirectly we all suffer from abnormal rainfall magnitudes.

A typical difficulty with risk analysis based on rainfall is that the magnitude and the frequency of rainfall strongly depend on the site where it is measured whereas the temperature is a more ambient measure. By way of example, if in Heathrow the temperature is 20 C then you can assume that the temperature in Central London is near 20 C (average differences taken off). However, even if it rains heavily in Heathrow you cannot assume that it will rain simultaneously in Central London. There is a high positive probability it does, but it is not worth 100%.

Until now, rainfall process studies have been limited to either cumulative rain over a given period (mainly monthly and yearly) or on the rainfall process over a very small time period (every 5 or 6 minutes). The first topic is quite limited and lacks interest for weather derivatives purposes. The second one is very detailed but irrelevant for market transaction. Instead of statistically analysing the rain, it may be preferable to simulate it.

Only Heathrow data are used from now on.

A. Frequency

The first step in building a stochastic rainfall process is to understand with which probability rainfall occurs. Figure 1 presents the historical frequency of the event "it rains" in Heathrow.

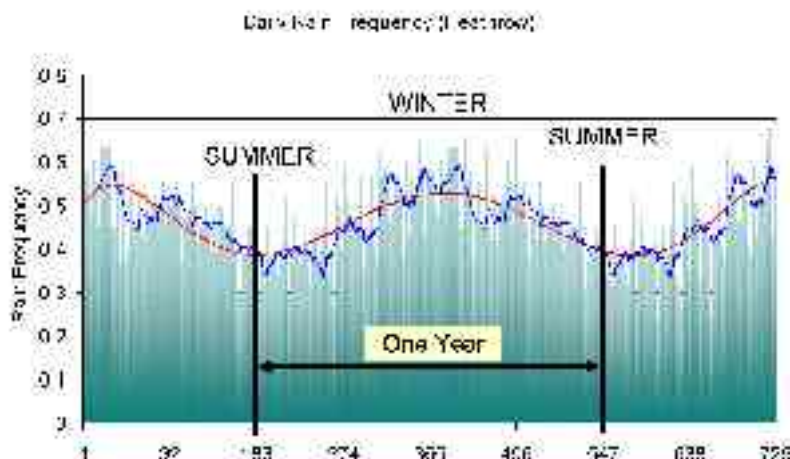


Figure 1

From this study, it appears that it rains more frequently in winter than in summer. The frequency curve is not symmetric and a sinusoidal function might be inadequate to fit it. A smoothing process can be used to correct erratic values (red curve). Subsequent

results are strongly dependent on these probabilities; thus the extraction must be done cautiously.

B. The rainfall persistence

The probability that it rains depends on the day of the year. But does it also depend on the past? Without doubt in the case of Heathrow, it appears that when the weather is dry, the following day is more likely to be dry than rainy, and vice versa. So the probability that it rains is conditional on the past. But the probability that it rains depends on the time of year. As a consequence, it is more likely to rain two consecutively days during wintertime than in summertime and the reverse in summertime. Therefore a natural autocorrelation is created which interferes with a possible true autocorrelation.

The model of the persistence of rain is set hereafter.

We note X_t the event “it rains at day t ”. X_t is Bernoulli distributed:

$$X_t = \begin{cases} 0 & \text{with prob } 1 - p_t \\ 1 & \text{with prob } p_t \end{cases}$$

Where 1 is for the event “it rains”.

We know the historical mean of X_t from our first studies. However, this is insufficient to model the time series X_t . As a matter of fact, assuming the independence of X_t we have $E[X_t | X_{t-1}, X_{t-2}, X_{t-3}] = E[X_t] = p_t$ which is different from $E_H[X_t]$ (the operator E_H is for the historical mean). Therefore the probability p_t has a time dependent expectation and is conditional to $X_{t-1}, X_{t-2}, X_{t-3}, \dots$.

A recurrence is produced to estimate the order of this lag dependence. The method is explained hereafter:

The probability p_t is assumed to be given by: $p_t = \text{Prob}(X_t = 1 | X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_{t-k})$, with $k \in \mathbb{N}^*$.

The aim is to extract the minimum value of k that produces the best fit of the distribution of the length of period of rain. Considering a 365 day year, we assume that $E[p_t] = E[p_{t+365}]$ which means that the climate does not vary over years¹. We first estimate conditional probabilities with $k=1$, then simulate the process and compare the simulated distribution of the length of the period to the smoothed historical one. Then, the method is reproduced with higher value of k until no more information on the probability p_t is added.

Assuming the independence between successive rainy days, we have simulated the rainfall below using the historical probability for 38 years (the same length of period from which figure 2 has been calculated). Supposing $k=1$ a good fit is already obtained:

¹ This condition is not restrictive in the sense that it assumes it will always rain statistically more frequently in winter than in summer and that the reverse should not happen. It also means there is no trend in the rainfall process.

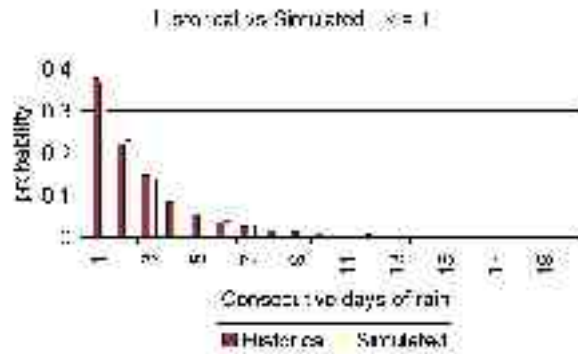


Figure 14

C. The magnitude process

Once the length of the rainy period is known, the intensity of the rainfall must be evaluated for each day. The previous study didn't reveal a dependence on the length of the rainy period. But studies show that there is a dependence on the length of the period.

Since the average conditional to the previous day is different, the distribution is certainly different. Only four events can be enumerated for a rainy day t under the assumption $k=1$: $R_{t-1} / R_t / R_{t+1}$ or $NR_{t-1} / R_t / R_{t+1}$ or $R_{t-1} / R_t / NR_{t+1}$ or $NR_{t-1} / R_t / NR_{t+1}$ with R_t the event it rains at day t and NR_t the event no rain at day t . Four distributions for each day of the year should be estimated. In order to reduce estimation bias errors, a 30-day period, bracketing the day for which the distributions are worked out, is considered. The next figure shows the 4 distributions using Box-Plot for the 1st of January:

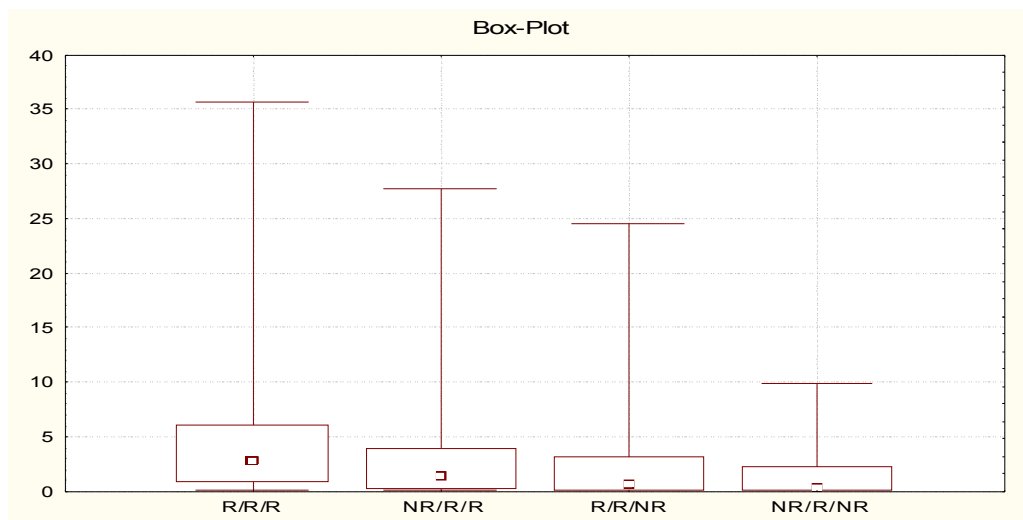


Figure 3

The main characteristics are summarised in this table:

	$R_{t-1} / R_t / R_{t+1}$	$NR_{t-1} / R_t / R_{t+1}$	$R_{t-1} / R_t / NR_{t+1}$	$NR_{t-1} / R_t / NR_{t+1}$
Minimum	0.10	0.10	0.10	0.10
Maximum	35.60	27.80	24.50	9.90

Average	3.96	2.86	2.48	1.53
Stdev	4.27	4.05	4.07	2.34

The four distributions are very different from each other. The average (resp. maximum) between the event $R_{t-1} / R_t / R_{t+1}$ and $NR_{t-1} / R_t / NR_{t+1}$ is approximately divided by 2.5 (resp. 3.5). One can conclude that the distribution of the magnitude of rainfalls depends on the immediate past and future.

Doing the same for each day of the year, all the required information to run the simulations properly is eventually obtained. First, all the rainy days are simulated and then the magnitude of rain is randomly generated using the correct distribution between the four possible ones for each day.

Conclusion

Rainfall magnitude is extremely dependent upon the location where it is measured. This risk can be as high as 100% in a single month for very close locations (20 miles). Because the probability that it rains is non-constant during a whole year, this phenomenon creates a natural autocorrelation in the process. This pitfall has to be avoided and the rainfall process can be decomposed into two steps. The first stage is the frequency process and the second stage is the magnitude given the frequency. In order to know the distribution of the rainfall magnitude, it is just as important to know if it rained the previous day than if it will rain the next day.

Bibliography

- [1] **B. Dischel**, Is precipitation basis risk overstated?, Risk magazine February 2001, or <http://www.risk.net/supplements/weather00/wthr00-rainfallp.htm>
- [2] **Eric Gaume**, Application de l'algorithme METROPOLIS pour l'analyse de sensibilité d'un modèle stochastique de pluie, <http://www.enpc.fr/cereve/HomePages/gaume/gentest/gentest.html>
- [3] **Thauvin V., Gaume E., Roux C.** - A short time step point rainfall stochastic models - Water Sciences and technology, vol. 37, n°11, pp. 34-45