Stoch Environ Res Risk Assess (2014) 28:717–727
DOI 10.1007/s00477-013-0784-0

ORIGINAL PAPER

Rainfall derivatives pricing with an underlying semi-Markov model for precipitation occurrences

Giovanni Masala

Published online: 30 August 2013

Abstract Weather derivatives represent a new and particular kind of contingent claim which shares a specific underlying weather index. These derivatives are written for different temperature indices, hurricanes, frost, snowfall and rainfall, and they are available for several cities. Our paper focuses on rainfall derivatives. In order to price this kind of derivatives, we have to model daily rainfall sequences at a specific location. For this purpose, we adopt a non-homogeneous parametric semi-Markov model to describe the rainfall occurrences, and a mixture of exponential distributions for rainfall amounts. The underlying Markov process has the obvious two states: dry and wet. In addition, dry and wet sequences are estimated by using best-fitting techniques. The model parameters are determined thanks to classical log-likelihood maximization. We finally price some rainfall contracts issued by the Chicago Mercantile Exchange through Monte Carlo simulation. The numerical applications and the parameter estimations are carried out using real data.

Keywords Semi-Markov process · Waiting time distributions · Rainfall occurrences · Rainfall intensity · Monte Carlo simulation · Rainfall derivatives

1 Introduction

We are aware that weather influences many human activities. For example, it appears to be an important production factor in agriculture. Regrettably, this production factor is difficult to monitor. Consequently, weather risks constitute a major source of uncertainty in many agricultural products (such as wheat, corn, cotton, rice). Adverse weather conditions can lead to huge unexpected losses. Traditionally, producers try to refund these losses by buying insurance instruments. In the mid-1990s, a new class of more flexible contingent claims has emerged, namely weather derivatives (WD). WD are financial instruments that allow the trading of weather-related risks. These WD can be structured as futures, options and swaps. What underlies them are not tradable assets, as in traditional financial derivatives, but are specific weather indices, such as temperature, rainfall, snow, frost, or wind. As an example, the effect of temperature on the demand for energy is clearly understood.

Consequently, these instruments provide firms with the possibility of managing adverse weather events. They also provide financial institutions with a new investment tool that is uncorrelated with other financial instruments, such as equities or bonds. There is also significant potential use for WD in numerous industry sectors, as a notable proportion of the industrialized economy is weather sensitive.

WD first appeared in the US energy industry in 1997. The weather futures markets available at the Chicago Mercantile Exchange (CME) began trading in 1999. The first products traded were two standard temperature contracts (heating degree days and cooling degree days) available for ten U.S. cities. In 2003, the CME Group weather products went global by issuing monthly and seasonal HDD and cumulative average temperature (CAT) contracts for six European locations. In 2005, the first frost contract appeared for Amsterdam. In the same year, CME also launched the first precipitation contract based on snowfall for the U.S. Further enhancements and additions occurred in 2009, 2010, and 2011, including binary

G. Masala (✉)
Faculty of Economics, University of Cagliari, Via S. Ignazio 17, 09123 Cagliari, Italy
e-mail: gb.masala@unica.it

contracts for U.S. snowfall which pays a fixed amount when a given amount of snowfall occurs, and U.S. rainfall contracts that provide precipitation risk management during the spring, summer, and fall months. These products are actively traded on the CME Globex electronic platform.

The main characteristics of rainfall-based WD, which is one of the main topics of this paper, will be revealed in the next sections. We devote this paper to the pricing of such derivatives. The underlying rainfall forecasting model represents the other fundamental goal of our survey. Note that rainfall forecasting is a very common topic in the literature. For example, Haddad et al. (2011) examined design rainfall in Australia, Niu (2013) investigated the temporal patterns of precipitation due to climate anomalies in China, and Shi et al. (2013) link precipitation variation with flood risk.

We will now investigate the pricing features more thoroughly. First, we observe that WD pricing is not straightforward. Indeed, what underlies them are non-tradable and so the market is incomplete. Classical pricing techniques coming from the Black and Scholes environment imposing the existence of a unique pricing measure or perfect replication strategy fail for these kinds of products. In addition, we can assert that there is no direct analogue distributional assumption with respect to weather events in general, as the weather variables considered (temperature, rainfall amounts, etc.) share significantly different properties and are not market-traded assets, so that they can have no market price. Remember that the idea of the distribution of future scenarios corresponds to the estimation of expectations about the future states of the underlying asset. This observation is analogous to producing a weather forecast when considering weather indices as the underlying data. Consequently, managing weather risk can be pursued by devising a reliable model to forecast the evolution of the weather. Indeed, weather-related risk is linked to the high degree of the weather's unpredictable fluctuations.

Another key point in the traditional option valuation literature is the determination of the so-called market prices of risk inherent in the particular instrument. Indeed, the market price of risk is linked to the discount rate to be applied in order to calculate the present value of the option. Regarding WD, the market price of risk assumes a particular interest. As pointed out by Cao et al. (2004), the risk of a weather event cannot be hedged against, while a fundamental achievement of the Black and Scholes model is the possibility, based on arbitrage arguments, to define a replicating portfolio of the underlying stock that can hedge the option payoff. In a survey carried out by Alaton et al. (2002), this problem is applied to the pricing of a temperature derivative, under the assumption of a constant market price of risk defined as the product of a constant λ and the risk-free rate. A similar assumption is stated by Cao et al. (2004) relative to precipitation derivatives. The authors add a constant risk premium to the simulated option prices.

Observe now that, from a statistical point of view, there are two alternatives for the modeling of weather risk. On the one hand, the distribution of the weather event (a weather index such as CDD for temperature derivatives) can be estimated directly. On the other hand, a daily model of the underlying weather variable (for example, mean temperature) can be set up, from which the associated weather index is then deduced. This latter procedure is more complex, but nevertheless it furnishes more accurate results. We will adopt this procedure throughout the paper.

We now summarize the following pricing techniques. The pricing of the options can be carried out by a burn analysis, an index value simulation, and a daily simulation. These procedures are briefly described below.

- Burn analysis (historical simulation technique). The goal of burn analysis is to derive an empirical distribution of the rainfall index by using the historical rainfall data included in our database. Thanks to this empirical distribution, empirical payoffs of the specific WD are determined and discounted through the given risk-free interest rate. Finally, the price of the option is the mean of these discounted payoffs.
- Index value simulation. It requires determining a parametric distribution for the given rainfall index. From these distributions, index values are randomly determined by Monte Carlo simulation. We then deduce the associated payoff of the WD for each scenario. Finally, the price of the option is average discounted payoffs.
- Daily simulation. In this approach, we simulate the daily precipitation instead of the rainfall index. The rainfall index is then derived and the price of the WD is settled as previously described.

In our paper, we will adopt, as already stated, the daily simulation, and compare the results with Burn analysis.

We emphasize that all the methods enumerated should require the quantifying of the market price for weather risk. The market price for weather risk ought to be derived implicitly from price quotations, if a market already existed for these WD. Unfortunately, this is not the case in the situation considered, so that we will ignore this aspect.

When focusing on WD, the underlying precipitation model should be able to capture the following characteristics of daily rainfall:

- the probability of rainfall occurrence follows a seasonal pattern. For example, rainfall in Europe is more likely in winter than in summer;

– the sequence of wet and dry days follows an autoregressive process. This means that the probability of a rainy day is higher if the previous day was wet; in other words, the present state dry/wet is conditioned by the previous state dry/wet;
– the amount of precipitation on a wet day varies with the season. For example, rainfall in Europe has a higher intensity in summer than in winter;
– the volatility of the amount of rainfall also changes seasonally. In Europe, it is higher in summer than in winter.

In this paper, we propose a new approach based on non-homogeneous parametric semi-Markov models for what is involved in the rainfall occurrences, which permit the taking into consideration of a more realistic assumption of events' time dependence. For example, the state of evolution of the system (rain/dry) depends on the previous state as well as the elapsed time from the previous state. This is, to our knowledge, the first application of semi-Markov models in this domain. The main goal is then to price WD issued by CME through a Monte Carlo simulation, after generating rainfall time series. We will focus the numerical application of the model upon a large database representing daily rainfalls in New York in the period 1980–2010.

Homogeneous Semi-Markov processes (HSMP) were independently introduced in the 1950s by Levy (1954) and Smith (1955), with the objective of generalizing the Markov processes. Indeed, in a Markov process environment, the waiting time distribution function in each state must be exponential, whereas in a semi-Markov process environment, these distributions can be of any type.

We must point out that the Markovian approach for modeling rain/dry sequences has been investigated by several authors; for example, see Detzel and Mine (2011), Goncü (2011), Liu et al. (2011), Mhanna and Bauwens (2011), Perera et al. (2002), and Rajagopalan et al. (1996). A detailed theoretical analysis of semi-Markov processes can be found in Howard et al. (1971a, b), and more recently in Janssen and Manca (2006). Since then, these processes have been applied successfully in a wide range of scientific fields. Non-homogeneous semi-Markov processes (NHSMP) were successively introduced by Iosifescu Manu (1972).

Many authors have recently applied semi-Markov models to the field of biomedicine; for example, see Davidov (1999), Di Biase et al. (2007), Satten and Sternberg (1999), Sternberg and Satten (1999), and Foucher et al. (2005). Another area of application concerns geophysics. Some authors have applied semi-Markov models to estimate earthquake occurrences; see Garavaglia and Pavani (2009) and Masala (2012).

The paper is organized as follows. The aim of the paper and the state of art has been illustrated in the current introductory Sect. 1. Section 2 surveys the theoretical aspects of the semi-Markov processes. Section 3 deals with the quantitative models for rainfall forecasting. The rainfall derivatives are described in Sect. 4. The numerical application and the description of the database used are exposed in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Semi-Markov processes

The application of a semi-Markovian model assumes that the present state dry/wet will depend on the previous state and the time elapsed between them. The main purpose of the semi-Markov model then consists of modeling the evolution of dry/wet sequences with respect to the elapsed time from the last state and the last state itself. In a semi-Markov framework, the waiting times in a state before a transition occurs can be of any type; this is an enhancement with respect to the classical pure Markovian approach. In the following, we will adopt the day as time unit because the database at our disposal contains empirical daily rainfall accumulation at the specified location.

Let us now formalize our model from a semi-Markov point of view, after a brief survey of this theory (see Janssen and Manca 2006).

We denote $E = \{E_1, E_2\}$ the state space and $\{\Omega, F, P\}$ represents the usual probability space (where E_1 corresponds to a dry day and E_2 corresponds to a wet day). Let us denote $m = \mathrm{card}E$ (here $m = 2$). We then introduce the following random variables

$$J_n : \Omega \to E \quad S_n : \Omega \to [0, +\infty) \tag{1}$$

where J_n represents the state at the nth transition and S_n represents the chronological time of the nth transition. We denote $N(t)$ as the counting process $\{N(t),\ t \geq 0\}$ which is associated with the point process $(S_n)_{n\in N}$ defined as

$$N(t) = \sup\{n \in N : S_n \leq t\} \tag{2}$$

for every $t \geq 0$.

From this definition, the meaning of the random variable $N(t)$ is clear. It represents the number of transitions occurring in the horizon $[0, t]$.

From the definition of S_n, we can introduce the "duration process" $(X_n)_{n\in N}$ as the following family of random variables:

$$\begin{cases} X_0 = 0 \\ X_{n+1} = S_{n+1} - S_n \end{cases} \tag{3}$$

Henceforth, we deduce that X_{n+1} represents the sojourn time spent in state J_n.

We can now set up the main definition of the theory. The process $(J_n, S_n)_{n\in N}$ is called "non-homogeneous Markov renewal process" if the following relationship holds:

$$P\Big(J_{n+1} = j, S_{n+1} \le t | J_n = i, S_n = s, \vdots J_{n-1}, S_{n-1}, \ldots, J_0, S_0\Big) = P(J_{n+1} = j, S_{n+1} \le t | J_n = i, S_n = s) \quad (4)$$

In addition, we can associate with the renewal process its non-homogeneous semi-Markov kernel Q (for $j \neq i$) defined as

$$\begin{aligned} Q &= \big[Q_{ij}(s,x)\big] \\ &= \big[P\big(J_{N(s)+1} = j, \quad X_{N(s)+1} \le x | J_{N(s)} = i, \quad S_{N(s)} = s\big)\big] \end{aligned} \quad (5)$$

We observe that, in the second member of this definition, x represents a duration time whereas s represents a chronological time.

It has been proved (see Janssen and Manca 2006) that the probability $p_{ij}(s)$ that the system performs its next transition to the state j, given that it entered the state i at time s, can be deduced as:

$$\begin{aligned} p_{ij}(s) &= \lim_{x \to \infty} Q_{ij}(s,x) \\ &= P\big(J_{N(s)+1} = j | J_{N(s)} = i, \quad S_{N(s)} = s\big) \end{aligned} \quad (6)$$

for every $i, j \in E$ and $i \neq j$. Hence, we can introduce the transition probability matrix $P(s) = \big[p_{ij}(s)\big]_{i,j \in E}$ of the embedded non-homogeneous Markov chain $(J_n)_{n \in N}$. In our application, $P(s)$ is a matrix of order m.

We observe that, before entering in state j, the system remains for a certain time x in the state i. At this point, it is opportune to introduce the conditional cumulative distribution function of the waiting time in each state, given the state subsequently occupied. We then set up:

$$F_{ij}(s,x) = P\big(X_{N(s)+1} \le x | J_{N(s)+1} = j, \quad J_{N(s)} = i, \quad S_{N(s)} = s\big) \quad (7)$$

It can be shown that this function is obtained in the following manner:

$$F_{ij}(s,x) = \begin{cases} Q_{ij}(s,x)/p_{ij}(s) & \text{if} \quad p_{ij}(s) \neq 0 \\ 1 & \text{if} \quad p_{ij}(s) = 0 \end{cases} \quad (8)$$

Remark We remind readers that a striking difference between a continuous time Markov process and a semi-Markov process involves the characteristics of the distribution functions $F_{ij}(s,x)$. In a Markov framework, these functions follow a negative exponential distribution function. In a semi-Markov framework, the distribution functions $F_{ij}(s,x)$ can be of any type. Consequently, the transition intensity can be decreasing or increasing. It then follows that the semi-Markov environment is richer than the Markov one.

We assume that the distributions of waiting times admit densities which will be denoted $D(s,x) = \big(f_{ij}(s,x)\big)_{i,j \in E}$.

We now introduce $H_i(s,x)$ the probability that the process leaves state i before or at a time x, given that the state i is entered at chronological time s:

$$H_i(s,x) = P\big(X_{N(s)+1} \le x | J_{N(s)} = i, \quad S_{N(s)} = s\big) \quad (9)$$

These probabilities can be written equivalently as

$$H_i(s,x) = \sum_{j=1}^{m} Q_{ij}(s,x) = \sum_{j=1}^{m} p_{ij}(s) \cdot F_{ij}(s,x) \quad (10)$$

We then observe that the marginal cumulative distribution functions of the waiting time in each state depend on both times.

Let us finally introduce the functions $\bar{S}_i(s,x) = 1 - H_i(s,x)$.

We are now able to define the continuous time non-homogeneous semi-Markov process $Z(t)$. This process represents, for each time t, the state occupied by the process, namely (see Janssen and Manca 2006):

$$Z(t) = J_{N(t)} \quad \text{for} \quad t > 0 \quad (11)$$

We can deduce from these definitions that the semi-Markov process $Z(t)$ is completely determined from the knowledge of the transition matrix $\mathrm{P}(t)$ and the duration matrix $H(s,x)$. We will see hereafter the importance of estimating correctly these two elements.

Remark The non-homogeneity of the process with respect to time is given by the fact that the jump process $p_{ij}(t)$ depends on the chronological time t. In other words, time has influence on the transition probabilities.

We then introduce the so-called "interval transition probabilities" given by the following:

$$\begin{aligned} \phi_{ij}(t,x) =& P(\text{system stays in state } j \text{ at time} \\ & t + x | \text{ entered state } i \text{ at time } t) \\ =& P\big(Z(t+x) = j | J_{N(t)} = i; \quad S_{N(t)} = t\big) \end{aligned} \quad (12)$$

for every $i, j \in E$.

It can be shown that these probabilities satisfy the obvious property $\phi_{ij}(t,x) \neq \phi_{ij}(t+h, x+h)$ for every $h > 0$.

These probabilities give fundamental information for the process's evolution. It is clear that the knowledge of the elements $p_{ij}(t)$ of the transition matrix and the elements $d_{ij}(s,x)$ of the duration matrix will allow us to determine $\phi_{ij}(t,x)$.

The link between these elements is given by the following fundamental relationship, called the evolution equation of the process:

$$\phi_{ij}(t,x) = \delta_{ij} \cdot \bar{S}_{i.}(t,x) + \sum_{\substack{l=1 \\ l \neq i}}^{m} \int_0^x p_{il}(t) \cdot d_{il}(u) \cdot \phi_{lj}(t+u, x-u)\, du \qquad (13)$$

where δ_{ij} is the Kronecker delta.

Finally, we illustrate the Monte Carlo procedure needed to generate sample paths from the semi-Markov process. Each simulation represents a daily sequence of rainfall amounts starting on January, 1 and ending on December, 31. Leap years have been removed so that each scenario is a vector of length 365.

We fix the usual time horizon [0, T] where T represents 1 year. This time, horizon is equally subdivided into 365 parts (daily subdivision).

The algorithm foresees the following steps:

(1) set $k = 0$ and select the initial state $J_0 = i$ (which corresponds to January, 1) with $i \in \{\text{dry; wet}\}$(at this point, we take into consideration the proportion of empirical initial states);
(2) sample the next state from the distribution $p_{J_k,*}$;
(3) sample the sojourn time $X \sim F_{J_k,J_{k+1}}$ (the cycle ends if $X \geq T$);
(4) set the new jump time $S_{k+1} = S_k + X$;
(5) if $S_{k+1} \geq T$ the cycle ends;
(6) otherwise set $k = k + 1$ and repeat from step (II) until $S_{k+1} \geq T$.

This algorithm has been revised from Barbu and Limnios (2008)

3 Rainfall forecasting

As we have already shown, the importance of rainfall forecasting arises from realizing that rainfall is not a market-traded asset and as such the rainfall option contract must be priced directly from states of rainfall. In addition to this, it is not possible to replicate the payoff from a rainfall option using a portfolio of alternative assets as happens in the Black and Scholes environment. These observations, coupled with the well-known fact that the price of an option is directly linked to future states of the underlying asset of value, motivate the fundamental aspect played by rainfall forecasting in our survey.

First, we observe that rainfall sequences are characterized by the alternating of days with no rainfall with intermittent days of rainfall. The rainfall days also experience a large amount of variation in rainfall intensities between them. These features turn out to be a challenging achievement of our model. Indeed, counting distributions which represent rain/no rain sequences must be carefully determined.

In addition, rainfall empirically exhibits much higher volatility than other traditional weather variables and although meteorological analysis acknowledges that rainfall is influenced by the interactions of other weather variables in the dynamic atmospheric system. Consequently, the statistical application of the aforementioned relationships is difficult as the level of rainfall volatility exceeds that of all other variables.

Finally, an additional fundamental modeling consideration is the incorporation of seasonality. Rainfall as with all other weather variables (for example mean daily temperatures at a given location) is seasonal and a selected rainfall model should acknowledge this fact.

3.1 Characteristics of the NHSM process

The previous considerations imply that the rainfall model can be split into two fundamental steps:

- at first, we have to model the sequences of rain/dry days;
- for rainy days, we have to determine the more adequate intensity distribution.

The semi-Markov process clearly enters in the first step. We stated in the previous section that the SM process possesses the two states $E = \{E_1, E_2\}$. In addition, in order to take into account the seasonal behavior, we will operate in a non-homogeneous framework.

An important feature of the SM model is given by the transitions between states. We have here four transitions allowed: $\begin{pmatrix} T_1 & T_2 \\ T_3 & T_4 \end{pmatrix}$.

Transition T_1 represents permanence in a dry state; transition T_2 represents a transition from a dry state to a rain state; transition T_3 represents a transition from a rain state to a dry state and finally transition T_4 represents the permanence in a rain state. Hence we deduce that the only "real" transitions are given by T_2 and T_3.

The waiting times associated to each state play obviously a fundamental role. The ones associated with the states T_2 and T_3 can be banally set up as unitary (i.e. 1 day) while the ones associated with T_1 and T_4 are more striking as they represent respectively dryness and rainy time intervals. Indeed, a sequence of dry/rain data can be interpreted as an alternating sequence of dryness and rainy time intervals with variable length.

In our model we select the more adequate distribution through best fitting techniques amongst a set of typical counting discrete distributions (Poisson, binomial, negative binomial, hypergeometric). The waiting times will be given by a multiple of unitary days.

In order to take into account the seasonality behavior of the rainfalls, we perform a monthly subdivision of our database. This means that for each month we calculate the transition matrix and the waiting time distribution for transitions T_1 and T_4.

Furthermore, the transition matrix will be turned into a "daily" function by interpolating the 12 monthly transitions probabilities previously determined. Results will be revealed in Sect. 5.

3.2 Intensity distribution

The statistical model for the nonzero precipitation amounts is a challenging problem of the model. Most commonly the gamma distribution is chosen for this purpose (see Barkotulla 2010; Castellvi et al. 2004; Leobacher and Ngare 2011; Perera et al. 2002). Hanson and Vogel (2008) proposes the Pearson type-III distribution. Liu et al. (2011) compares several distributions (exponential, mixture of exponentials, gamma, lognormal). Finally, some investigators have found that the mixture of exponential distribution provides a much better fit to daily nonzero precipitation amounts (see Cao et al. 2004; Detzel and Mine 2011; Hardle et al. 2011; Jamaludin and Jemain 2007; Musshoff et al. 2006; Roldan and Woolhiser 1982; Wilks 1999). We adopt the latter choice in our paper.

The density of this distribution is given by

$$f(t) = \frac{\alpha}{\beta} \cdot \exp\left(-\frac{t}{\beta}\right) + \frac{1-\alpha}{\gamma} \cdot \exp\left(-\frac{t}{\gamma}\right) \tag{14}$$

where $0 < \alpha < 1$, $\beta > 0$ and $\gamma > 0$ are real parameters.

The parameters are then estimated through classical maximization of the log-likelihood function

$$\ell(\alpha, \beta, \gamma) = \ln L(\alpha, \beta, \gamma) = \sum_{i=1}^{N} \ln(f(\xi_i)) \tag{15}$$

where $(\xi_i)_{i=1,\ldots,N}$ is a sample of daily precipitation amounts of length N. The maximization problem can be solved numerically thanks to Mathematica 8.0® software.

In order to take into account the seasonality features of the precipitations in a given location, we perform a monthly subdivision of our database. The parameters of the mixture of exponential distribution are then estimated for each month separately.

4 Rainfall derivatives

Rainfall derivatives expand the existing class of the CME Group weather products and they allow traders to manage risk directly linked to precipitation events.

We briefly describe the main features of rainfall derivatives issued by the CME group (information can be retrieved on the web pages http://www.cmegroup.com/trading/weather/). Note that these products are frequently updated and enhanced.

The new rainfall futures and options on futures issued by the CME Group enable the buyer to have access to a wide range of products covering benchmark rainfall indexes and seasonal rainfall totals. These contingents are available at the present time in 10 locations across the United States. The main categories of options issued are the following.

- Rainfall index futures.
  These contracts give the holder price certainty by representing the obligation to buy or sell a specified number of shares of the Rainfall Index on a specified date (they are traded on the CME Globex Market). The Rainfall Index is estimated for a specified month.
- Rainfall Index Options on Futures.
  These contracts are European-style options traded via open outcry; they permit the holder more flexibility as they afford the possibility to buy (in the case of a call) or to sell (in the case of a put) one Rainfall Index Future on a specified date and at a specified price.
- Seasonal Strip Rainfall Futures.
  As regards the strip alternative, the Rainfall Index is estimated for a sequence of specified months.
- Seasonal Strip Rainfall Options on Futures.
  These contracts allow the holder to buy or to sell one respective Seasonal Strip Rainfall future on a specified date and at a specified price.

Other examples include (seasonal Strip) Rainfall Index Binary Options.

Finally, let us enumerate some technical characteristics of these contracts.

For what involves options, the underlying contract is one Seasonal Strip Rainfall Index Futures Contract. The pricing unit is "Dollars per index point". The tick size (minimum fluctuation) is equal to 0.1 index point (namely $50 per contract). Finally, the strike price interval is 0.1 index point in a range of 0–60 index points. Similar characteristics hold for the other rainfall derivatives.

5 Empirical application

The database characteristics will be highlighted in Sect. 5.1. The next Sect. 5.2 is devoted to the estimation of the fundamental parameters of the semi-Markov model, namely the transition probability matrices and the distribution of the waiting times in each state. We further estimate the more adequate rain intensity distributions. Numerical results which represent the main findings of the paper are then given in Sect. 5.3. We finally comment the results in the Sect. 5.4.

5.1 Database description

The database contains rainfall daily accumulation (cm) in New York City (United States) for the period 1980–2010 (11,315 records after removing leap years, corresponding to 31 years). This database comes from Mathematica 8.0® climatic database. The Mathematica command "WeatherData" permits to access the available data. The main source information comprises the "National Oceanic and Atmospheric Administration", the "United States National Climatic Data Center", and the "Citizen Weather Observer Program".

We show in Fig. 1 the monthly precipitation accumulation (cm) for the period 2006–2010.

At first glance, this figure emphasizes a strong seasonal behavior of rainfall accumulation with variable amounts. We will investigate this feature thoroughly in the following subsections.

5.2 Semi-Markov estimation

As already stated, we introduce two observable states (dry/rain). All the transitions are formally allowed.

From the structure of the database, we deduce that the waiting times in each state before a transition occurs is a multiple of days. The theoretical waiting times will then be fitted amongst discrete distributions (using best-fitting techniques).

In this subsection, we enumerate the main steps of our model. Full details will be given in the following subsections. First, we have to estimate the transition matrix P. This matrix is set to be fixed for each month. Second, we determine the waiting times distribution for each transition with a monthly subdivision of our database. Finally, we estimate the rain intensity distributions with the same monthly subdivision.

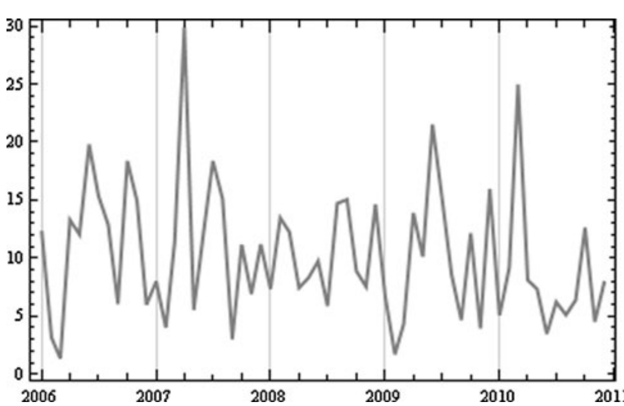

Fig. 1 Monthly data (cm) 2006–2010 (New York City)

5.2.1 The transition matrix P estimation

The data will be aggregated monthly in order to take into consideration the seasonal behavior of the transition probabilities.

The transition matrix P has already been defined in the previous Sect. 2. Whenever real data are available, this matrix can be estimated in the following way. Let us determine all the transitions from one state to another within the given time horizon. At this point, we construct a matrix A whose elements a_{ij} represent the number of transitions from the state i to the state j. We then define:

$$\hat{p}_{ij} = a_{ij} \Big/ \sum_{k=1}^{m} a_{ik} \tag{16}$$

(we set $\hat{p}_{ij} = 0$ whenever the denominator vanishes).

The empirical estimator $\hat{p}_{ij}$ of the transition matrix of the embedded Markov chain p_{ij} shares good asymptotic properties. It can be proved (Barbu and Limnios 2008) that these estimators are maximum likelihood estimators. In other words, they maximize the so-called likelihood function.

This procedure is then applied for each month so that we finally obtain 12 transition matrices. Thanks to an interpolation procedure, we can define the transitions' probabilities on a daily basis. We show (Fig. 2) the results for the transition $p_{22}(t)$.

In the case of the initial state (January, 1 of each year) of the rainfall process in our database, we observe that 64 % starts in state 1 (dry day) and 36 % starts in state 2 (rainy day).

5.2.2 The duration time distribution

After determining the transition matrices, we deduce from our database the waiting times in each state (expressed in days) before a transition occurs. We have already shown that the only significant transitions are given by the

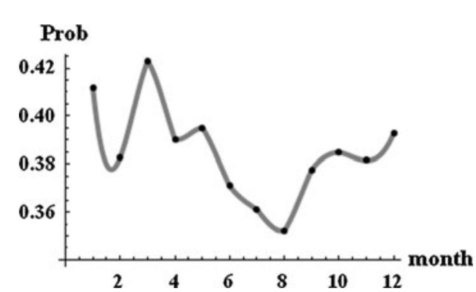

Fig. 2 Transition probabilities $p_{22}(t)$

permanence in the first state (dryness sequences) or in the second state (rainy sequences). Remember that the waiting times for the other transitions are set to unity. The next step consists in determining more adequate distribution functions for the waiting times.

We next analyze the waiting time distributions with a monthly subdivision. The more adequate distributions are then estimated through best-fitting techniques. First, we report in Table 1 the mean empirical waiting times.

These data unveil a significant seasonality behavior. Regarding the theoretical distributions, we adopt best-fitting techniques among the following set of discrete distributions (Best Fit software from Palisade): binomial (B); negative binomial (NB); Poisson (P); and hypergeometric (HG). The results are summarized in Tables 2 and 3.

Another important statement which can be deduced from our database involves the percentage of daily rainfall occurrences. We consider again a monthly subdivision. The results are presented in Table 4 and we can underline again a strong seasonality behavior.

The mean empirical annual precipitation percentage is 37.14 %.

5.2.3 *The rainfall intensity*

The seasonal pattern of rainfall intensity can also be highlighted by examining the mean monthly precipitations. The results are revealed in Table 4. The mean annual precipitation amount is 120.41 cm.

As already stated, the theoretical distributions will be given by a mixture of exponential distributions whose parameters α, β and γ, obtained through classical log-likelihood maximization, are summarized in Table 5.

Table 1 Empirical waiting times (days)

Month	Transition $E_1 \mapsto E_1$	Transition $E_2 \mapsto E_2$
1	3.24	1.57
2	3.13	1.83
3	3.13	1.75
4	3.25	1.96
5	3.49	1.90
6	3.35	1.88
7	3.51	1.67
8	4.24	1.70
9	4.19	1.58
10	4.72	1.62
11	3.36	1.61
12	3.00	1.64

Table 2 Waiting time distributions transition $1 \mapsto 1$

Month	Distrib.	Param. 1	Param. 2
1	P	3.2388	
2	NB	5	0.6153
3	P	3.1345	
4	NB	4	0.5520
5	NB	3	0.4625
6	NB	4	0.5442
7	NB	3	0.4607
8	NB	2	0.3205
9	NB	3	0.4171
10	NB	3	0.3884
11	NB	5	0.5978
12	NB	6	0.6667

Table 3 Waiting time distributions transition $2 \mapsto 2$

Month	Distrib.	Param. 1	Param. 2	Param. 3
1	B	6	0.2617	–
2	HG	58	1,028	32,675
3	HG	55	669	21,074
4	B	17	0.1150	–
5	B	15	0.1268	–
6	B	19	0.0991	–
7	B	6	0.2788	–
8	HG	54	1,029	32,675
9	B	5	0.3159	–
10	B	6	0.2699	–
11	B	8	0.2015	–
12	B	7	0.2338	–

Table 4 Monthly empirical rainfall occurrence and intensity

Month	Rain occurrences (%)	Monthly mean (cm)
1	39.65	8.46
2	38.94	6.85
3	40.37	10.75
4	42.04	10.98
5	40.69	9.56
6	39.46	11.22
7	37.04	12.51
8	32.26	11.41
9	30.54	9.83
10	31.22	9.91
11	35.48	9.72
12	38.09	9.22

Table 5 Rainfall intensity parameters

Month	α	β	γ
1	0.3811	0.1257	1.0346
2	0.3428	0.1114	0.8976
3	0.6338	1.2684	0.1494
4	0.5148	1.4791	0.2255
5	0.5903	1.1661	0.1692
6	0.4467	0.2296	1.5275
7	0.6795	1.5222	0.1718
8	0.6796	1.5933	0.1807
9	0.6236	1.6046	0.1931
10	0.3577	0.1681	1.5003
11	0.7572	1.1737	0.1007
12	0.6945	1.0677	0.1296

5.3 Numerical results

The goal of this section is twofold. First, we want to simulate rainfall sequences via Monte Carlo simulation and compare the results with observed data. Second, the simulated scenarios will be exploited to determine the price of some WD issued by the CME.

Let us consider 10,000 replications. In order to investigate the adequacy of the rainfall model, we compare the monthly mean precipitation accumulation, the percentage occurrences of rainfall days, and the length of wet/dry sequences with the empirical ones.

The first two outcomes are presented respectively in Figs. 3 and 4.

In order to verify the reliability of the model, we can quantify the error through the RMSE indicator and the associated normalized RMSE indicator, which are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}} \qquad (17)$$

$$N - RMSE = \frac{RMSE}{y_{\max} - y_{\min}}$$

where y_t denotes the observed value, $\hat{y}_t$ denotes the simulated value, and n is the length of the samples. These indicators computed over a monthly horizon (for mean precipitation accumulation) gave us $RMSE = 0.2329$ and $N - RMSE = 1.86\ \%$.

We then produce in Table 6 the ordinary 95 % confidence intervals for monthly mean precipitation accumulation with 10,000 Monte Carlo replications.

We exhibit in Table 7 the monthly simulated waiting times. These data are then compared with the empirical ones (given in Table 1).

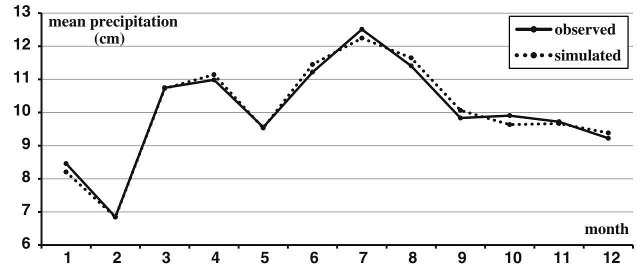

Fig. 3 Monthly mean precipitation accumulation

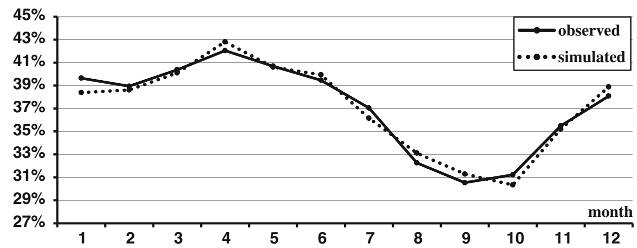

Fig. 4 Percentage occurrences of rainfall days

Table 6 Confidence intervals: monthly mean precipitation accumulation

Month	95 % Confidence interval
1	[8.16, 8.33]
2	[6.73, 6.88]
3	[10.49, 10.72]
4	[11.03, 11.27]
5	[9.49, 9.69]
6	[11.40, 11.65]
7	[12.01, 12.28]
8	[11.45, 11.73]
9	[10.05, 10.31]
10	[9.46, 9.71]
11	[9.39, 9.61]
12	[9.27, 9.46]

We deduce a good accordance between empirical and simulated waiting times.

Finally, we exhibit the pricing results of several rainfall derivatives already described in the previous section.

We will focus on call and put strip options in order to illustrate the pricing method. At this point, remember that the payoff for such options is given, respectively, by the following:

$$C = \text{Max}(I_{month} - K; 0)$$
$$P = \text{Max}(K - I_{month}; 0) \qquad (18)$$

where I_{month} denotes the rainfall index (total precipitation accumulation referred to the covering months) and K denotes the strike price (exercise price) of the contract. Remember that rainfall index and exercise prices are

Table 7 Simulated waiting times (days)

Month	Transition $E_1 \mapsto E_1$	Matching (%)	Transition $E_2 \mapsto E_2$	Matching (%)
1	3.27	0.93	1.59	1.27
2	3.10	−0.96	1.84	0.55
3	3.19	1.92	1.77	1.14
4	3.29	1.23	1.98	1.02
5	3.54	1.43	1.93	1.58
6	3.33	−0.60	1.91	1.60
7	3.47	−1.14	1.69	1.20
8	4.32	1.89	1.72	1.18
9	4.20	0.24	1.59	0.38
10	4.76	0.85	1.61	−0.62
11	3.40	1.19	1.63	1.24
12	3.01	0.33	1.64	0.12

transformed into prices by considering that one rainfall index equals \$500 (with tick size 0.1, namely \$50) where one rainfall index corresponds to 1 inch (2.54 cm) of rainfall accumulation.

As regards Burn analysis, I_{month} is deduced for each year at our disposal in the database (31 years), then we calculate the mean payoff

$$
\begin{aligned}
C &= \frac{1}{31} \cdot \sum_{i=1}^{31} Max\left(I^i_{month} - K; 0\right) \\
P &= \frac{1}{31} \cdot \sum_{i=1}^{31} Max\left(K - I^i_{month}; 0\right)
\end{aligned}
\tag{19}
$$

where I^i_{month} represents the rainfall index for the given month associated to year i.

Concerning the daily simulation, the mean payoff is calculated in the same manner:

$$
\begin{aligned}
C &= \frac{1}{N} \cdot \sum_{j=1}^{N} \mathrm{Max}\left(I^j_{month} - K; 0\right) \\
P &= \frac{1}{N} \cdot \sum_{j=1}^{N} \mathrm{Max}\left(K - I^j_{month}; 0\right)
\end{aligned}
\tag{20}
$$

where N denotes the number of simulations and I^j_{month} represents the rainfall index relative to the given months associated to jth simulation.

Finally, the price of the options is given, in both approaches, by the discounted expected payoffs using the risk-free rate.

The results are shown in Table 8 (Burn approach and daily simulation with 10,000 replications). For the sake of simplicity, we express the payoff with respect to the rainfall index (and not with respect to prices). Assume that the strip period covers the months June to September. These results then represent the sensibility of prices with respect to the strike K.

Table 8 Price options (historical vs. daily simulation)

Strike	Call (Burn)	Call (sim.)	Strike	Put (Burn)	Put (sim.)
6.0	11.70	11.89	30.0	12.30	12.17
6.5	11.20	11.39	30.5	12.80	12.66
7.0	10.70	10.90	31.0	13.30	13.15
7.5	10.20	10.40	31.5	13.80	13.65
8.0	9.70	9.91	32.0	14.30	14.14
8.5	9.31	9.42	32.5	14.80	14.64
9.0	8.73	8.94	33.0	15.30	15.13
9.5	8.25	8.46	33.5	15.80	15.63
10.0	7.78	7.99	34.0	16.30	16.13

5.4 Results and discussion

As regards the rainfall simulation, the mean annual precipitation amount is 120.53 cm while the empirical one is 120.41 cm (difference of 0.10 %). In addition, the mean simulated annual precipitation percentage is 37.11 % while the empirical one is 37.14 % (difference of 0.08 %). These findings emphasize a good agreement between the theoretical model and the empirical data. When considering total precipitation and rainy day percentages on a monthly level (see Figs. 3, 4), we again observe a good agreement with the empirical data. The same conclusion also holds for the monthly waiting times (dry and rain sequences). This means that the model is also able to capture the seasonality behavior of rainfall sequences.

Finally, pricing results are just a consequence of the rainfall simulation procedure. We underline that the simulation procedure and the Burn approach lead to similar results. Actually, the Burn approach represents the average payoff of the derivatives in the past years without incorporating rainfall forecasts in the pricing. For this reason, the model described here is preferable for options weather-based pricing. Moreover, the Burn approach lacks a rigorous mathematical foundation.

We also emphasize that the Burn analysis involves the estimation of a global rainfall index and the rainfall derivatives. It does not involve rainfall forecasts. In addition, the rainfall model proposed is able to estimate the length of dry/rain sequences. This feature is outside the scope of the Burn analysis. We can assert that more complex rainfall derivatives products with these requirements may be correctly priced within the semi-Markov model.

6 conclusions

In this paper, we have determined the daily rainfall sequence at a particular location by applying a parametric semi-Markov non-homogeneous model. The underlying

Markov process presents the two states, rain/dry. The waiting time distributions have been modeled through discrete time distributions, and all the required parameters have been estimated through classical best-fitting techniques. Rainfall intensities have been modeled thanks to a mixture of exponential distributions and the parameters have also been determined through the maximum likelihood procedure.

The main purposes were then to forecast rainfall sequences and to price some rainfall derivatives issued by the Chicago Mercantile Exchange. In order to perform this task, a reliable forecasting model for rainfall sequences is needed. The pricing is determined through a Monte Carlo simulation which consists in generating daily rainfall accumulation for a 1-year time horizon. The payoff of the requested derivative is then deduced. Finally, the option price is simply given by the discounted expected payoff. We note that, for this kind of derivatives, no closed formulas are available. In addition, the classical financial pricing techniques are not applicable as the underlying (rainfall index) is not a tradable asset.

The model has been tested with a large amount of data freely available from Wolfram Mathematica® software for New York City. The simulated data demonstrated a good agreement with empirical data.

References

Alaton P, Djehiche B, Stillberger D (2002) On modelling and pricing weather derivatives. Appl Math Finance 9(1):1–20

Barbu VS, Limnios N (2008) Semi-Markov chains and hidden semi-Markov models toward applications: their use in reliability and DNA analysis. Lecture Notes in Statistics No. 191. Springer, Berlin. doi:10.1007/978-0-387-73173-5

Barkotulla MAB (2010) Stochastic generation of the occurrence and amount of daily rainfall. Pak J Stat Oper Res VI 1:61–73

Cao M, Li A, Wei J (2004) Precipitation modeling and contract valuation: a frontier in weather derivatives. J Altern Invest 7(2):93–99

Castellvi F, Mormeneo I, Perez PJ (2004) Generation of daily amounts of precipitation from standard climatic data: a case study for Argentina. J Hydrol 289:286–302

Davidov O (1999) The steady state probabilities for a regenerative semi-Markov process with application to prevention and screening. Appl Stoch Models Data Anal 15:55–63

Detzel DHM, Mine MRM (2011) Generation of daily synthetic precipitation series: analyses and application in La Plata river Basin. Open Hydrol J 5:69–77

Di Biase G, D'Amico G, Di Girolamo A, Janssen J, Iacobelli S, Tinari N, Manca R (2007) A stochastic model for the HIV/AIDS dynamic evolution. Math Probl Eng 2007:14. doi:10.1155/2007/65636

Foucher Y, Mathieu E, Saint-Pierre P, Durand JF, Daurès JP (2005) A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. Biometrical J 47:825–833

Garavaglia E, Pavani R (2009) About earthquake forecasting by Markov renewal processes. Methodol Comput Appl Probab. doi:10.1007/s11009-009-9137-3

Goncü A (2011) Modeling and pricing precipitation-based weather derivatives. Financial Math Appl 1(1):9–18

Haddad K, Rahman A, Green J (2011) Design rainfall estimation in Australia: a case study using L moments and generalized least squares regression. Stoch Environ Res Risk Assess 25:815–825

Hanson LS, Vogel R (2008) The probability distribution of daily rainfall in the United States. Conference proceeding paper, World Environmental and Water Resources Congress

Hardle W, Osipenko M (2011) Pricing Chinese rain: a multisite multi-period equilibrium pricing model for rainfall derivatives. SFB 649 Discussion Paper 055. Humboldt-Universität zu, Berlin

Howard RA (1971a) Dynamic probabilistic systems, vol I: Markov Models. Wiley, New York

Howard RA (1971b) Dynamic probabilistic systems, vol II: semi-Markov and decision processes. Wiley, New York

Iosifescu Manu A (1972) non-homogeneous semi-Markov processes. Stud Lere Mat 24:529–533

Jamaludin S, Jemain AA (2007) Fitting the statistical distributions to the daily rainfall amount in peninsular Malaysia. J Teknologi 46:33–48

Janssen J, Manca R (2006) Applied semi-Markov processes. Springer, New York

Leobacher G, Ngare P (2011) On modeling and pricing rainfall derivatives with seasonality. Appl Math Finance 18(1):71–91

Levy P (1954) Processus semi-markoviens. In: Erven P, Noordhoff NV. Proceedings of the International Congress of Mathematicians 3, Groningen, pp 416–426

Liu Y, Zhang W, Shao Y, Zhang K (2011) Comparison of stochastic precipitation models based on four distributions. Adv Atmos Sci 28(4):809–820

Masala G (2012) Earthquakes occurrences estimation through a parametric semi-Markov approach. J Appl Stat 39(1):81–96

Mhanna M, Bauwens W (2011) Stochastic single-site generation of daily and monthly rainfall in the Middle East. Meteorol Appl. doi:10.1002/met.256 Wiley OnlineLibrary

Musshoff O, Odening M, Xu W (2006) Modeling and hedging rain risk. American Agricultural Economics Association Annual Meeting, Long Beach

Niu J (2013) Precipitation in the Pearl River basin, South China: scaling, regional patterns, and influence of large-scale climate anomalies. Stoch Environ Res Risk Assess 27:1253–1268

Perera HKWI, Sonnadara DUJ, Jayewardene DR (2002) Forecasting the occurrence of rainfall in selected weather stations in the wet and dry zones of Sri Lanka. Sri Lankan J Phys 3:39–52

Rajagopalan B, Lall U, Tarboton DG (1996) Nonhomogeneous Markov model for daily precipitation. J Hydrol Eng 1(1):33–40

Roldan J, Woolhiser DA (1982) Stochastic daily precipitation models, a comparison of occurrence processes. Water Resour Res 18(5):1451–1459

Satten GA, Sternberg MR (1999) Fitting semi-Markov models to interval-censored data with unknown initiation times. Biometrics 55:507–513

Shi P, Qiao X, Chen X, Zhou M, Qu S, Ma X, Zhang Z (2013) Spatial distribution and temporal trends in daily and monthly precipitation concentration indices in the upper reaches of the Huai River, China. Stoch Environ Res Risk Assess. doi:10.1007/s00477-013-0740-z

Smith WL (1955) Regenerative stochastic processes. Proc R Soc Lond A 232:6–31

Sternberg MR, Satten SA (1999) Discrete-time nonparametric estimation for semi-Markov models of chain-of-events data subject to interval-censoring and truncation. Biometrics 55:514–522

Wilks DS (1999) Multisite downscaling of daily precipitation with a stochastic weather generator. Clim Res 11:125–136