Modelling of Major Flood Arrivals on Chinese Rivers by Switch-time Processes

Pavel Stoynov¹, Plamena Zlateva², Dimiter Velev³ and Xuejun Zong⁴

¹ FEBA, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria
² Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria
³ University of National and World Economy, Sofia, Bulgaria, Email: dgvelev@unwe.bg
⁴ Shenyang University of Chemical Technology, Shenyang, China

Abstract. Nowadays, there is a considerable flood risk in China is. A brief description of major floods on Yangtze River and Huang He River is given. Both big Chinese rivers have long records of floods with severe life and property damages. Quantification of the stochastic behaviour of the largest floods is a key task in the risk assessment and mitigation. An exponential distribution of the time intervals between consecutive floods is assumed in classical study of inter-arrival times of floods. An approach for modelling of flood arrivals on both Chinese rivers by switch-time (ST) processes is proposed. These ST distributions can be considered as distributions of sums of random numbers exponentially distributed random variables. The proposed model specifies explicitly times of occurrence not only of floods but also of higher risk of potential floods. This approach could be useful for making prognoses of floods and for analysing changes in hydrologic behaviour of rivers.

1. Introduction
In recent years, the climate changes lead to an increase in both the number and severity of floods throughout the world. The floods can cause economic and social damage, and lead to serious consequences for the environment and critical infrastructure [1].
A simple definition of the flood is an overflow of a body of water (river, lake) that submerges land or the abnormal accumulation of water on the surface due to excess rainfall and rise of the groundwater level above surface on impermeable or saturated terrains. The major floods have the following characteristics: due to overflowing of rivers and lakes, unexpected and serious breaks in dikes, levees and other protective structures or uncontrolled releases of dam water; coverage of a wide continuous area and rapid spreading to adjacent areas of relatively lower elevation; relatively deep in most parts of the flood-stricken areas [2].
Since the beginning of registration of floods in the world, there are about fifteen major floods [3], [4]. It can be seen that the greatest number of these major floods have occurred in China. The 1931 Central China flood is considered as the deadliest disaster in the world. It started in July 1931, when Yellow river (Huang He), Huai River and Yangtze River water level heights are exceeded. Both big rivers Yangtze and Huang He have long records of floods with significant damages [5].
Large and damaging floods occur every year. The maximum estimated annual damage caused by river floods in one country is recorded in China in 2010, where a total loss of US $51 billion is reported [1]. For this reason, the each approach to the assessment of the flood risk in China is very important and useful.

2. Quantitative statement of the problem
Quantification of the stochastic behavior of the largest floods is a key task in flood risk assessment.
Usually it is assumed that floods occur independently with some average rate of occurrence [7]. For this reason, their occurrence can be modeled by Poisson process and inter-arrival times are independent and exponentially distributed [8]. In this case, they have the distribution of a random variable $\xi \in \text{Exp}(\beta)$ with probability density function.

$$f_\xi(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Therefore, the exponential distribution for intervals between two major floods is assumed [9]. Some observations on inter-arrival times however show that they are heavy-tailed. In probability theory, heavy-tailed distributions are probability distributions whose tails are heavier than the tails of the exponential distribution [8].

A typical representative of heavy-tailed distributions is Pareto distribution. It is applied in some investigations of times between consecutive floods. Heavy tails introduced with Pareto (power law) distribution for these intervals is proposed in [10].

General Pareto distribution $P(\mu, \sigma, \chi, \alpha)$ is a distribution with cumulative distribution function

$$F(x) = 1 - \left[1 + \left(\frac{x - \mu}{\sigma}\right)^{-\frac{1}{\alpha}}\right]^{-\chi}, \ x > \mu.$$  

Here $\mu \in \mathbb{R}$ is a centric parameter, $\sigma > 0$ is a scale parameter, $\chi > 0$ is inequality parameter, and $\alpha > 0$ is a form parameter. Different special cases of this distribution are used [8]

$$P_1(\sigma, \alpha) = P(\sigma, \sigma, 1, \alpha), \quad P_2(\mu, \sigma, \alpha) = P(\mu, \sigma, 1, \alpha), \quad P_3(\mu, \sigma, \chi) = P(\mu, \sigma, \chi, 1).$$

Heavy tailed distributions actually present the possibility for longer inter-arrival times. Another approach to present possibility for longer inter-arrival times is to skip some of the arrivals in the Poisson process. In this case, the inter-arrival time is with Erlang distribution. The Erlang distribution is a two parameter family of continuous probability distribution which may be considered as a sum of independent exponentially distributed random variables.

The approach proposed in this article is to randomize the Erlang distribution of inter-arrival times. This leads to so called switch-time (ST) distributions and switch-time (ST) processes. ST distributions can be considered as distributions of sums of random number exponentially distributed random variables. In this way, without using heavy tailed distributions it can be presented longer and more flexible inter-arrival times [11], [12]. An approach for modelling Inter-arrival time of floods by ST processes is proposed in [13].

The aim of this paper is to model flood arrivals on rivers by switch-time processes. In particular, the major floods of the both Chinese rivers Yangtze and Hunag are considered. These ST processes permit to represent both seasonal and cyclical risks of floods. A benefit of using this approach is that the model specifies explicitly times of occurrence not only of floods but also of higher risk for potential floods. With consistent data sets for flood appearance, the approach could be useful for forecasting of floods and analyzing of changes in hydrologic behaviour of rivers. The model can track also consequences of risk mitigation activities, for example building of dams.

3. Essence of ST distributions and ST processes

It is said that a random variable $\xi$ with probability mass function $f_\xi(x)$ has distribution of $ST(n, \beta)$ family and denote this fact $\xi \in ST(n, \beta)$, if the probability mass function of $\xi$ is given by the formula

$$f_\xi(x) = \begin{cases} \sum_{k=1}^{n} P(D^n = k) f_{\xi}^{(k)}(x) | D^n = k), & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$f_\xi(x) = \begin{cases} \sum_{k=1}^{n} P(D^n = k) f_{G^k}(x), & x \geq 0 \\ 0, & x < 0 \end{cases}$$
Where $G^k$ are random variables with probability mass function $f_{G^k}(x) = f(k, \beta)$ and $D^n$ are positive integer mixing random variables.

Here for $G^k$ it can be adopted different families of distribution.

In this article, the case of ST family of first kind is considered where:

$$G^k \in \Gamma(k, \frac{1}{\beta}) \equiv \text{Erlang}(k, \frac{1}{\beta}), \quad \text{i.e.} \quad \xi \mid D^n \equiv \text{Erlang}(D^n, \frac{1}{\beta}).$$ \hspace{1cm} (5)

Correspondingly, $D^n$ is a random variable, taking values $k = 1, \ldots, (n + 1)$ with probabilities:

$$P(D^n = k) = \frac{C(n, \beta)n!}{\beta^k (n-k+1)!}, \quad k = 1, \ldots, (n + 1)$$ \hspace{1cm} (6)

Where the coefficients $C(n, \beta)$ are given by the formulas:

$$C(n, \beta) = \frac{1}{I(n, \beta)}, \quad I(0, \beta) = \frac{1}{\beta}, \quad I(n, \beta) = \frac{1}{\beta} + \frac{n}{\beta} I(n-1, \beta), \quad n = 1, 2, \ldots$$ \hspace{1cm} (7)

Also, variables $\tilde{D}^n = D^n - 1$ can be introduced taking values $k = 0, \ldots, n$ with probabilities

$$P(\tilde{D}^n = k) = \frac{C(n, \beta)n!}{\beta^{k+1} (n-k)!}, \quad k = 0, \ldots, n.$$ \hspace{1cm} (8)

Then the probability mass function $f_\xi(x)$ of $\xi$ may be presented also by the formula

$$f_\xi(x) = \sum_{k=0}^{n} P(\tilde{D}^n = k) f_\xi(x \mid \tilde{D}^n = k) = \sum_{k=0}^{n} P(\tilde{D}^n = k) f_{G^k}(x)$$ \hspace{1cm} (9)

If a random variable has ST distribution of first kind, we will denote this fact $\xi \in ST1(n, \beta)$. For such a variable, the probability density is density of Erlang distribution with first parameter being another integer-valued random variable with a specific distribution.

If in $ST1(n, \beta)$ it is enforced $D^n = k$, i.e.

$$P(D^n = k) = 1, \quad P(D^n = i) = 0, \quad 1 \leq i \leq k - 1 \leq k + 1 \leq i + 1 \leq n + 1,$$ \hspace{1cm} (10)

Which may be considered as a degenerate ST distribution of first kind, and it is actually obtained Erlang distribution, i.e. $\xi \in \text{Erlang}(k, \frac{1}{\beta})$.

The distribution $ST1(n, \beta)$ can be considered as a special kind of generalized Gamma distribution [12].

It is said that a random variable $A$ has generalized gamma distribution if its probability density function is given by

$$f_A(x) = \frac{u^{s-a}}{\Gamma(\alpha)U(\alpha, \alpha+1-s, u\beta)} e^{-\beta x^a} x^{a-1} (u + x)^{-s}$$ \hspace{1cm} (11)

where $x > 0$, $-\infty < s < +\infty$, $\alpha > 0$, $\beta > 0$, $u > 0$ and

$$U(a, b, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1 + t)^{b-a-1} dt, \quad \text{where} \quad a > 0, \quad z > 0.$$ \hspace{1cm} (12)

The above expression is the integral representation of the hyper-geometric function of second kind [14]. In this case, for $A$ it is given $A \in G^\Gamma(\alpha, \beta, u, s)$. 
It needs to remember that the random variable $\xi$ has a Gamma distribution and it is denotes as $\xi \in \Gamma(\alpha, \beta)$, if its probability mass function is given by

$$f_{\xi}(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$  \hspace{1cm} (13)$$

Here $\Gamma(\alpha)$ is defined by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-\alpha} dx$. It is seen that $\Gamma(\alpha, \beta) = G\Gamma(\alpha, \beta, u, s) = 0$.

Actually, it can written that $STI(n, \beta) = G\Gamma(\alpha = 1, \beta, u = 1, s = -n)$. The Exponential distribution is a special kind of $STI(n, \beta)$ distribution and $STI(0, \beta) \equiv Exp(\beta)$. To recall that the random variable $\xi$ has exponential distribution and it is denotes as $\xi \in Exp(\beta)$, if its probability mass function is given by

$$f_{\xi}(x) = \begin{cases} \beta e^{-\beta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$  \hspace{1cm} (14)$$

The case $STI(1, \beta) \equiv Lindley(\beta)$ can be considered as a weighted version of exponential distribution with weighting function $w(x) = 1 + x$. It can said that the random variable $\xi$ has Lindley distribution and it is denotes as $\xi \in Lin(\beta)$, if its probability mass function is given by

$$f_{\xi}(x) = \begin{cases} \frac{\beta^2}{\beta + 1} (1 + x) e^{-\beta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$  \hspace{1cm} (15)$$

The case $n = 2$ leads to $STI(2, \beta)$ distribution which can be defined as weighted exponential distribution by the weight function $w(x) = (1 + x)^2$.

It is said that the random variable $\xi$ has $STI(2, \beta)$ distribution and it is denotes as $\xi \in STI(2, \beta)$, if its probability mass function is given by

$$f_{\xi}(x) = \begin{cases} \frac{\beta^3}{\beta^2 + 2\beta + 2} e^{-\beta x} (1 + x)^2, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$  \hspace{1cm} (16)$$

To check that the weight function in the definition of the distribution is the right one, it could calculate as $Ew(\xi) = \frac{\beta^3 + 2\beta + 2}{\beta^2}$.

It is said that a process $X(t)$ is a ST process of first kind or $STI(n, \beta)$ process, and it is denoted this fact as $X(t) \in STI(t; n, \beta)$, if the following statements are satisfied:

1) $X(0) = 0$.
2) $X(t)$ is pure jump process with jumps at times $T_i$; $i = 1, 2, \ldots$ and jump sizes $\Delta X(T_i) = 1$.
3) The intervals between two jumps are $\tau_i = T_i - T_{i-1} \in STI(n - 1, \beta)$, $i = 0, 1, \ldots$, $T_0 = 0$.

It is said that $X(t)$ is compound STI($n, \beta$) process if it is replaced condition 3) with condition:

3') $X(t)$ is pure jump process with jumps at times $T_i$; $i = 1, 2, \ldots$ and jump sizes $\Delta X(T_i) = Y_i$, where $Y_i$ are independent and identically distributed random variables.

4. Brief description of major Floods on Yangtze River and Huang He River
4.1. Description of major floods on Yangtze River
The Yangtze River (actually Chang Jiang, the Long River) is the longest river in Asia, the third longest in the world and the longest in the world to flow entirely within one country. It drains one-fifth of the land area of China and its river basin is a home to one-third of the country's population. Flooding along the river has been a major problem. The rainy season in China is May and June in areas south of Yangtze River, and July and August in areas north of it. The huge river system receives water from both southern and northern flanks, which causes its flood season to extend from May to August. Meanwhile, the relatively dense population and rich cities along the river make the floods more deadly and costly. Tens of millions of people live in the floodplain of the Yangtze valley, an area that naturally floods every summer and is habitable only because it is protected by river dikes. The floods large enough to overflow the dikes have caused great distress to those who live and farm there. Floods of note include 1931, 1935, 1954, 1998. The 1931 Central China floods were a series of floods. The floods are generally considered among the deadliest natural disasters ever recorded, and almost certainly the deadliest of the 20th century (when pandemics and famines are discounted). The 1931 floods covered more than 77,700 square km—including the cities of Nanjing and Wuhan—killed more than 300,000 people, and left 4,000,000 more homeless. The Yangtze again flooded in 1935, causing great losses of life. From June to September 1954, the Yangtze River floods were a series of catastrophic floorings that occurred mostly in Hubei Province. Due to unusually high volume of precipitation as well as an extraordinarily long rainy season in the middle stretch of the Yangtze River late in the spring of 1954, the river started to rise above its usual level in around late June. Despite efforts to open three important floodgates to alleviate the rising water by diverting it, the flood level continued to rise until it hit the historic high of 44.67 m in Jingzhou, Hubei and 29.73 m in Wuhan. The number of deaths from this flood is estimated at around 33,000, including those who died of plague in the aftermath of the disaster. The 1998 Yangtze River floods was a major flood that lasted from middle of June to the beginning of September 1998 in China at the Yangtze River. The event is considered the worst Northern China flood in 40 years. In the summer of 1998, China experienced massive flooding of parts of the Yangtze River, resulting in 3704 dead, 15 million homeless and $26 billion in economic loss. Other sources report a total loss of 4150 people, 180 million people are affected, and 13.3 million houses are damaged. The floods caused $26 billion in damages. One of the major objectives of the Three Gorges Dam project was to alleviate flooding on the lower Yangtze. The dam proved effective during the extraordinarily rainy summer of 2010 by holding back much of the resultant floodwaters and thus minimizing the impact of flooding downstream. However, the dam still had to open its floodgates to reduce the high water volume in the reservoir, and flooding and landslides in the Yangtze basin killed several hundred people and caused extensive property damage [15].

4.2. Description of major floods on Huang He River
Huang He (The Yellow River) is the third-longest river in Asia, following the Yangtze River and Yenisei River, and the sixth longest in the world at the estimated length of 5464km. There is a long record of floods caused by Huang He before 19th century. The September–October 1887 flood has been estimated to have killed between 900,000 and 2 million people, and is the second-worst natural disaster in history, excluding famines and epidemics. The Yellow River more or less adopted its present course during the 1897 flood. During the 1931 flood according to various estimates from 1,000,000 to 4,000,000 people are killed, and 88,000 square km are completely inundated, and approximately 21,000 square km more are partially flooded, leaving 80 million people homeless. It is the worst natural disaster recorded (excluding famines and epidemics) On 9 June 1938, during the Second Sino-Japanese War, Nationalist troops under Chiang Kai-Shek broke the levees holding back the river near the village of Huayuankou in Henan. The goal of the operation was to stop the advancing Japanese troops by following a strategy of "using water as a substitute for soldiers" (yishui daibing). The 1938 flood of an area covering 54,000 square kilometer took some 500,000 to 900,000 Chinese lives, along with an unknown number of Japanese soldiers. Huang He floods in 1887, 1931 and 1938 collectively killed millions and are considered to be the three deadliest floods in history and among the most destructive natural disasters ever recorded.
Throughout most of its history, China has attempted to control the Huang He by building overflow channels and increasingly taller dikes, and in 1955 the Chinese embarked on an ambitious 50-year construction plan and flood-control program. This program included extensive dike construction, repair, and reinforcement, reforestation in the loess region, and the construction of a series of dams to control the river’s flow, produce electricity, and supply water for irrigation. Silt-retaining dams have not been completely effective (the accumulation of silt reduces the power-generating capacity of the dams), and they have been criticized by environmentalists. Continued silting in the Huang He has remained a serious problem; however, the river has not burst its banks since 1945, in large part because of the flood-control program [16].

5. Applications of Switch-time processes for modelling of major flood arrivals in China

Applications of Switch-time processes for modelling of major flood arrivals on big Chinese rivers are proposed in the paper. In particular, both Rivers Yangtze and Huang He are considered.

In [16] an example with modelling inter-arrival time between floods by ST processes is presented. Let us suppose that there are two types of events related to floods. The first type is “condition for floods”. The process, which counts number of appearances of this kind of events, is Poisson with parameter $\beta$. The second kind of events we call “real flood”. It can be modelled by the process $STI(n = [\frac{t_{\max}}{t_{\min}}], \beta)$, where $t_{\max}$ is the maximal and $t_{\min}$ - the minimal observed interval between two major floods in a country (area, region) and $[.]$ is the integer part of a number.

For Yangtze river it is estimated that $\frac{1}{\beta} = 35$ and $n = [\frac{t_{\max}}{t_{\min}}] = \frac{61}{12} = 5$. The realizations of the random variables $D^n_i$ are as shown in Table 1.

| N | Year of major flood of | Time interval | $D^n_i$ |
|---|------------------------|--------------|---------|
|   | Yangtze River          | (years)      |         |
| 1 | 1870                   | 61           | 5       |
| 2 | 1931                   | 23           | 2       |
| 3 | 1954                   | 44           | 4       |
| 4 | 1998                   | 12           | 1       |
| 5 | 2010                   | -            | -       |

This means that at every 35 years a possibility of major flood occurs and at every 5 possibilities it can be expected such a flood to really occur.

If we apply Poisson-process modelling, the interval between two consecutive floods should be with exponential distribution. In this case, the mean interval between two floods is 35 years and the model says that a flood should be expected every 35 years. It is seen that this model do not catch the variations of the inter-arrival interval which are better reflected by the model with ST processes.

For Huang He river it is estimated that $\frac{1}{\beta} = 12.75$ and $n = [\frac{t_{\max}}{t_{\min}}] = \frac{26}{7} = 3$. The realizations of the random variables $D^n_i$ are as shown in Table 2.

This means that at every 13 years a possibility of major flood occurs and at every 3 possibilities it can be expected such a flood to really occur.

If we apply Poisson-process modelling, the interval between two consecutive floods should be with exponential distribution. In this case, the mean interval between two floods is 13 years and the model says that a flood should be expected every 13 years. It is seen that this model do not catch the variations of the inter-arrival interval which are better reflected by the model with ST processes.
When we model with ST processes, the parameter $\beta$ is related to the average interval between times of high risk of flood when the conditions for (potential) flood are available. The applied approach gives the possibility to “calibrate” the severity of the floods captured in the model. If we select different thresholds for a flood to be considered as major, we may obtain different degree of refinement of the flood’s scale.

For the considered scales of floods of Yangtze River and Huang He River, we may say that conditions for large-scale flood of Huang He River appear more frequently than conditions for large-scale flood of Yangtze River.

| N | Year of major flood of Huang He River | Time interval (years) | $D^i$ |
|---|--------------------------------------|----------------------|------|
| 1 | 1887                                 | 10                   | 1    |
| 2 | 1897                                 | 8                    | 1    |
| 3 | 1905                                 | 26                   | 3    |
| 4 | 1931                                 | 7                    | 1    |
| 5 | 1938                                 | -                    | -    |

6. Conclusions

An approach for modelling of flood arrivals on rivers by switch-time processes is proposed. In particular the major floods of the both Chinese rivers Yangtze and Hunag are considered. These ST processes permit to represent both seasonal and cyclical risks of floods. ST processes can be used to model floods even for short data records. With longer periods covered and more precise information about floods the models using ST processes could give a better picture of inter-arrival times.

A benefit of using this approach is that the model specifies explicitly times of occurrence not only of floods but also of higher risk for (potential) floods. With consistent data sets for flood appearance, the approach could be useful for making prognoses of floods and analyzing changes in hydrologic behaviour of rivers. The model can track also consequences of risk mitigation activities, for example building of dams.

7. Acknowledgments

The authors express their gratitude to the Bulgarian National Science Fund for the support under the Grant № DFNI-I02/15 from 2014 and the Bulgarian – Chinese scientific cooperation Grant № DNTS/ China 01/6 from 2014.

8. References

[1] Z.W Kundzewicz, et al., 2014 Flood risk and climate change: global and regional perspectives. *Hydrological Sciences Journal*, **59** (1), pp 1–28.
[2] PreventionWeb-Serving the information needs of the disaster reduction community, Hazard – Floods, [Online] Available: http://www.preventionweb.net/english/hazards/flood/
[3] Rankred 2015 *Internet-Technology_Knowledge*, 15Biggest and Worst Floods Ever in History, Nature, Available: http://www.rankred.com/top-10-biggest-and-worst-floods-ever-in-history
[4] O’Connor J and Costa E 2004 *The World's Largest Floods, Past and Present: Their Causes and Magnitudes*. U. S. Geological Survey Circular, **1254**.
[5] List of deadliest floods, Wikipedia, the free encyclopedia, Available: https://en.wikipedia.org/wiki/List_of_deadliest_floods
[6] Asian Development Bank. *Water-related disasters and disaster risk management in the People’s Republic of China*, Mandaluyong City, Philippines: ADB, 2015.
[7] Singh V P, Wang S X, and Zhang L 2005 Frequency analysis of nonidentically distributed hydrologic flood data, *J. Hydrol.*, **307**, pp.175–195.
[8] Stoynov P 2006 Manual of statistics. Sofia University “St. Kliment Ohridski”, Sofia.
[9] Enzel Y, Redmond K, House K and Biondi F 2002 Climate variability and flood frequency at decadal to millenial time scales, Annual Meeting, Geol. Soc. of Am., Denver, Colorado.
[10] Benson D, Schumer R and Meerschaert M 2007 Recurrence of extreme events with power-law interarrival times, Geophys. Res. Lett., 34, L16404, doi: 10.1029/2007GL030767.
[11] Stoynov P 2010 Negative binomial distribution and applications. Proc. of the Third Int. Conf. on Financial and Actuarial Mathematics, Sofia, Bulgaria, pp 2-20.
[12] Stoynov P 2013 Stoynov Process–Definition and Simulations. Proc. of the Sixth Int. Conf. on Financial and Actuarial Mathematics, Sofia, Bulgaria, pp 97-104.
[13] Stoynov P, Zlateva P and Velev D 2015 An Approach for Modelling Inter-arrival Time of Floods Int. Journal of Innovation, Management and Technology, 6 (4), pp 267-271.
[14] Zamani H and Ismail N 2010 Negative binomial-Lindley distribution and its applications Journal of Mathematics and Statistics, 6 (1), pp 4-9.
[15] Encyclopedia Britannica: Yangtze River 2016 http://www.britannica.com/eb/article-9110538/Yangtze-River
[16] Encyclopedia Britannica: Huang He River 2016 http://www.britannica.com/place/Huang-He.