UNIVERSIDAD NACIONAL DE COLOMBIA

A conceptual stochastic rainfall-runoff model applied to tropical watersheds

Sara María Vallejo Bernal

Thesis presented in fulfilment of the requirements for the degree of:
Master of Science in Applied Mathematics

Advisor: Prof. PhD. Jorge Mario Ramirez Osorio
Co-Advisor: Prof. PhD. Germán Poveda Jaramillo

Research Area: Stochastic Hydrology

Universidad Nacional de Colombia
Faculty of Science, School of Mathematics
Medellín, Colombia
2020

Quien ha visto la esperanza, no la olvida. La busca bajo todos los cielos y entre todos los hombres. Y sueña que un día va a encontrarla de nuevo, no sabe dónde, acaso entre los suyos. En cada hombre late la posibilidad de ser o, más exactamente, de volver a ser, otro hombre.

Octavio Paz

Agradecimientos

A mi esposo Yeisson, por darle a mi corazón, la paz que mi mente necesita. A mi profesor Germán Poveda, por enseñarme que se puede construir la vida, tal cual cómo uno la sueña. A mi profesor Jorge Mario Ramírez, por creer siempre en mí, incluso cuando yo misma dejé de hacerlo. A mi Alma Mater, la Universidad Nacional de Colombia, el lugar donde conocí la esperanza, el lugar donde volví a ser otra mujer.

Abstract

We derive and solve a linear stochastic model for the evolution of discharge and runoff in an order-one watershed. The system is forced by a statistically stationary compound Poisson process of instantaneous rainfall events. The relevant time scales are hourly or larger, and for large times, we show that the discharge approaches a limiting invariant distribution. Hence any of its properties are with regard to a rainfall-runoff system in hydrological equilibrium. We give an explicit formula for the Laplace transform of the invariant density of discharge in terms of the catchment area, the residence times of water in the channel and the hillslopes, and the mean frequency and the probability distribution of rainfall inputs. As a study case, we consider a watershed under a stationary rainfall regime in the tropical Andes and test the probability distribution predicted by the model against the corresponding seasonal statistics. A mathematical analysis of the invariant distribution is performed yielding formulas for the invariant moments of discharge in terms of those of the rainfall. The asymptotic behavior of probabilities of extreme events of discharge is explicitly derived for heavy-tailed and light-tailed families of distributions of rainfall inputs. The scaling structure of discharge is asymptotically characterized in terms of the parameters of the model and under the assumption of wide sense scaling for the precipitation amounts and the inverse of the residence time in the channel. The results give insights into the conversion of uncertainty inherent to the rainfall-runoff dynamics, and the roles played by different geophysical variables. The ratio between the mean frequency of rainfall events to the residence time along the hillslopes is shown to largely determine the qualitative properties of the distribution of discharge. Finally, a purely theoretical approach is proposed to reinterpret the hydrological concept of return period in the context of time-continuous Markov processes.

Keywords: stochastic hydrology, rainfall-runoff modeling, Poisson precipitation.

Un modelo conceptual estocástico de lluvia-escorrentía aplicado a cuencas tropicales

Resumen

En este trabajo derivamos y resolvemos un modelo estocástico lineal para la evolución del caudal y la escorrentía en una cuenca hidrográfica de orden uno. El sistema es forzado por un proceso de Poisson compuesto, estadísticamente estacionario, de eventos de lluvia instantáneos. Las escalas de tiempo relevantes son horarias o mayores, y cuando el tiempo tiende a infinito, mostramos que el caudal se acerca a una distribución invariante límite. Por tanto, cualquiera de sus propiedades está relacionada con un sistema de lluvia-escorrentía en equilibrio hidrológico. Damos una fórmula explícita para la transformada de Laplace de la densidad invariante del caudal en términos del área de la cuenca, los tiempos de residencia del agua en el canal y las laderas, y la frecuencia media y la distribución de probabilidad de los eventos de lluvia. Como caso de estudio, consideramos una cuenca bajo un régimen de lluvias estacionario en los Andes tropicales y evaluamos la distribución de probabilidad predicha por el modelo con las estadísticas estacionales correspondientes. Realizamos un análisis matemático de la distribución invariante obteniendo fórmulas para los momentos invariantes del caudal en términos de los de la precipitación. El comportamiento asintótico de las probabilidades de los eventos extremos del caudal se deriva explícitamente para familias de distribuciones de lluvia de cola pesada y cola ligera. La estructura de escalamiento del caudal se caracteriza asintóticamente en términos de los parámetros del modelo y bajo el supuesto de escalamiento simple para la precipitación y el inverso del tiempo de residencia en el canal. Los resultados dan una idea de la conversión de la incertidumbre inherente a la dinámica lluvia-escorrentía y los roles que juegan las diferentes variables geofísicas. Mostramos que la relación entre la frecuencia media de los eventos de lluvia y el tiempo de residencia en las laderas determina en gran medida las propiedades cualitativas de la distribución del caudal. Finalmente, proponemos un enfoque puramente teórico para reinterpretar el concepto hidrológico de período de retorno en el contexto de procesos de Markov continuos en el tiempo.

Palabras clave: hidrología estocástica, modelo de lluvia-escorrentía, Precipitación de Poisson

Contents

List of Figures

List of Tables

Notation

Symbol	Dimensions	Description
$A(t,x)$	$[L^2]$	Cross-sectional area of the channel.
a	$[L^2]$	Total area of hillslopes.
C	$[L^{1/2}/T]$	Chezy's coefficient.
$f(x)$	$[1/L]$	Probability density function of rainfall amounts.
$\tilde{f}(s)$		Laplace transform of f.
$g(x)$	$[T/L^3]$	Probability density function of discharge.
$\tilde{g}(s)$		Laplace transform of $g(x)$
$h(t,x,y)$	$[L]$	Effective height of the hillslopes runoff.
H	$[1/T]$	Inverse of hillslope mean residence time.
K	$[1/T]$	Inverse of channel mean residence time.
ℓ	$[L]$	Length of the stream.
L	$[L]$	Length of the hillslope.
$N(t)$		Counting process of instantaneous rainfall events.
$p(t)$	$[L/T]$	Proccess of instantaneous rainfall events.
P_n	$[L]$	Net rainfall amount (precipitation minus evapotranspiration) of the event occurring at random time T_n.
$\hat{P}$		Dimensionless net rainfall amount.
$q(t,x)$	$[L^3/T]$	Discharge at a point with coordinate x along the stream.
$Q(t)$	$[L^3/T]$	Discharge at the outlet of the watershed.
$\hat{Q}(t)$		Dimensionless discharge at the outlet of the watershed.
$r(t,x,y)$	$[L^2/T]$	Runoff per unit of cross hillslope length.
$R(t)$	$[L^3/T]$	Total runoff flowing into the channel from its surrounding hillslopes.
S		Stream slope.
t	$[T]$	Time.
T_n	$[T]$	Time of ocurrence of the rainfall event P_n.
v	$[L/T]$	Water velocity along the channel.
w	$[L/T]$	Velocity of the flow down the hillslope.
W	$[L]$	Wet perimeter.
x	$[L]$	Coordinate along the stream.
y	$[L]$	Coordinate along the hillslope.

$\boldsymbol{X}(t)$	$[L^3/T]$	Column vector with the stochastic process of interest. The discharge at the watershed outlet is its first component and the runoff flowing from the hillslopes is the second one.
$\mathbf{M}$	$[1/T]$	Geomorphological square matrix.
$\boldsymbol{Y}(t)$	$[L^3/T]$	Random forcing of the process $\boldsymbol{X}$ due to the precipitation regime $p(t)$.
$\mathbb{CV}$		Coefficient of variation equal to the standard deviation over the mean.
$\mathbb{CV}_g$		Coefficient of variation with respect to the density g, i.e. conditioned on $Q(0)$ distributed as the invariant distribution of $Q(t)$.
$\mathbb{E}$		Expected value.
$\mathbb{E}_g$		Expectation with respect to the density g, i.e. conditioned on $Q(0)$ distributed as the invariant distribution of $Q(t)$.
$\mathbb{P}$		Probability
$\mathbb{P}_g$		Probability with respect to the density g, i.e. conditioned on $Q(0)$ distributed as the invariant distribution of $Q(t)$.
$\mathbb{R}$		Set of real numbers.
$\mathbb{V}\mathrm{ar}$		Variance.
$\mathbb{V}\mathrm{ar}_g$		Variance with respect to the density g, i.e. conditioned on $Q(0)$ distributed as the invariant distribution of $Q(t)$.
α		Dimensionless number equal to μ/σ.
β		Dimensionless number equal to H/K
$\gamma(s)$		Dimensionless probability density function of discharge.
$\tilde{\gamma}(s)$		Laplace transform of $\gamma(s)$.
$\delta(t)$	$[1/T]$	Delta function
η		Dimensionless number equal to H/λ
θ		Pareto Type I distribution shape parameter.
κ	$[L]$	Pareto Type I distribution minimum value parameter.
λ	$[1/T]$	Precipitation mean frequency.
μ	$[L]$	Inverse Gaussian distribution mean.
o		Little-o notation.
π		Pi number.
ρ	$[L]$	Gamma distribution scale parameter.
σ	$[L]$	Inverse Gaussian distribution scale parameter.
$\phi(s)$		Dimensionless probability density function of rainfall amounts.
$\tilde{\phi}(s)$		Laplace transform of $\phi(s)$.
ω		Gamma distribution shape parameter.

1. Introduction

Understanding and modeling the process by which rainfall is converted into runoff in watersheds is a key problem in hydrology. The uncertainty related to occurrence and intensity of rainfall induces uncertainty in the discharge in non-trivial ways that are mediated, in turn, by intertwined physical processes constraining the flow of water through the landscape. See for example Te Chow et al. (1962); Gupta et al. (2007).

Traditionally, an empirical approach has been proposed to address this challenge. The mathematical models consist of a system of deterministic equations that can be implemented by calibrating some parameters based on the response to a predetermined and well-known precipitation event so that the simulated discharge matches the *in situ* measurements (Snyder, 1938; Nash, 1957, 1959; James et al., 1987; Bhunya et al., 2011). Once calibration is achieved, responses to different precipitation events are studied under the assumption that the estimated parameters will remain constant regardless of the new rainfall regime in the watershed. In spite that diverse models have enriched the hydrological literature (Nguyen et al., 2015; Quintero et al., 2016; Hrachowitz and Clark, 2017), multiple attempts to understand and predict rainfall-runoff processes have proven to be a significant source of uncertainty in hydrological design (Moradkhani and Sorooshian, 2009; Beven, 2011).

However, a distinct body of literature has been developed within the framework of conceptual stochastic models pioneered by Eagleson (1972). This new approach uses a probabilistic description of rainfall as input to a dynamical model of runoff, routing and discharge. In such models, the watershed is considered as an ensemble of interconnected conceptual storages – hillslopes and stream links – that dynamically respond to a randomly evolving precipitation field (Reggiani et al., 1998). Unlike the traditional approach, conceptual stochastic models do not intend to focus on the uncertainty associated with the response of the watershed to individual rainfall events. The purpose is to understand the emergent properties of the dynamic system and its relation to the probabilistic properties of the uncertainty.

This thesis aims at deriving and studying robust probabilistic properties of the rainfall-runoff process, namely the equilibrium probability distribution of discharge and its dependence on both, the geophysical properties of the watershed, and the probabilistic structure of rainfall. To achieve this goal we consider a solvable, conceptual, mathematical model recently introduced by Ramirez and Constantinescu (2020). There, the authors solve and analyze the linearized mass and momentum

equations of Gupta and Waymire (1998) for the surface runoff and discharge at all streams and hillslopes within the river network of a spatially heterogeneous watershed forced by a random rainfall field. Here, we review and expand on some of the results of Ramirez and Constantinescu (2020) within the context of the hydrological processes involved in runoff generation in an order-one watershed, and the associated transfer of uncertainty.

In essence, the model considered here reduces the hillslope-channel system to a couple of connected linear reservoirs under a random instantaneous forcing. The reservoirs track the evolution of hillslope runoff $R(t)$ and river discharge $Q(t)$ via coupled stochastic linear differential equations driven by a random precipitation process $p(t)$. Evapotranspiration is ignored, and p is assumed to be net, runoff producing, rainfall. The time scales for t are of the order of hours or longer, and we are in particular interested in the behavior as $t \to \infty$.

Net precipitation is considered in the model as a compound Poisson Process. What this means is that the rainfall process $p(t)$ is random, with rainfall events that are assumed instantaneous, happening at an average rate of λ events per unit time. The precipitation amount falling over the hillslopes during each event is independent of anything else, and significantly, of arbitrary probability distribution density f.

Under all these simplifications, one can solve the stochastic differential equations for $R(t)$ and $Q(t)$ and readily simulate them. The model itself is fully derived from first principles and solved in chapter 2, where we also highlight the specific hypothesis that allowed us to simplify the rainfall-runoff equations into a linear system. A comparative review of previously proposed solvable mathematical rainfall-runoff models can be found there as well.

But the most important result is that the model yields an exact characterization for the invariant probability density g of the discharge Q which we present in chapter 3. In the case of a watershed with a single river, g is written in terms of the rate of rainfall events λ, the probability density of the rainfall amount f, the area of the watershed a, and two inverse time scales: H and K, determining the structure of travel times for sub-surface and river flow, respectively. The resulting expression for g gives the uncertainty structure of discharge whenever subject to a climate in statistical equilibrium.

The overarching objective of this work is to study the dependence of important features of g on all other parameters. Our main concern is the role played by each component of the system in determining emerging qualitative properties of the distribution of discharge; therefore, our model is not intended for quantitative predictions of discharge on any particular basin. Notwithstanding this, we do include a case study in chapter 4, where some of the predictions of the model are tested against real data with the following explicit purpose: illustrate that whenever a small catchment is under a rainfall regime that approximately satisfies the hypotheses of the model, the probability

distribution of discharge is in fact adequately approximated by the invariant distribution computed via the model. This validation exercise also yielded the following interesting observation: the constant H determining the distribution of hillslope travel times is very difficult to statistically identify, especially when compared to K. This is of course of relevance to our results listed above.

In chapter 5 we analyze the properties of the invariant distribution of discharge in relation to the precipitation and the geomorphological features of the watershed. In section 5.1 we derive the structure of the invariant moments of discharge, while in section 5.2 we study the asymptotic properties of the probabilities of extreme events, namely, the behavior of the right tail of the invariant distribution. We also provide a rigorous mathematical framework to study and analyze some traditional concepts of the hydrology in the context of our model, such as the scaling properties of the spatially averaged rainfall rates, residence times and river flows, which we present in section 5.3, and the return period of extreme hydrological events that is reinterpreted for time-continuous stochastic process in section 5.4.

Finally, the conclusions are drawn in chapter 6. Some relevant technical details about the mathematical techniques used in this work are included in the Appendix.

2. A conceptual stochastic rainfall-runoff model

In this chapter we formulate and solve linearized versions of the mass and momentum conservation equations for an order-one watershed forced by a stochastic precipitation process. We start by performing a first-principle derivation where we highlight how the storage-discharge relationships can be simplified in order to arrive at a linear model. Then we pose a stochastic differential equation for the rainfall-runoff system and its solution. Finally, we include a comparative review of previously proposed solvable mathematical rainfall-runoff models.

Consider a watershed with a single river of length ℓ [L] surrounded by hillslopes of total area a [L^2]. See Figure 1. We will derive linear versions of the mass and momentum conservation equations for the discharge Q [L^3/T] at the end of the stream, and the total runoff R [L^3/T] flowing into the channel from its surrounding hillslopes. Time is denoted by t and refers to the evolution of hydrological variables at hourly to daily time scales.

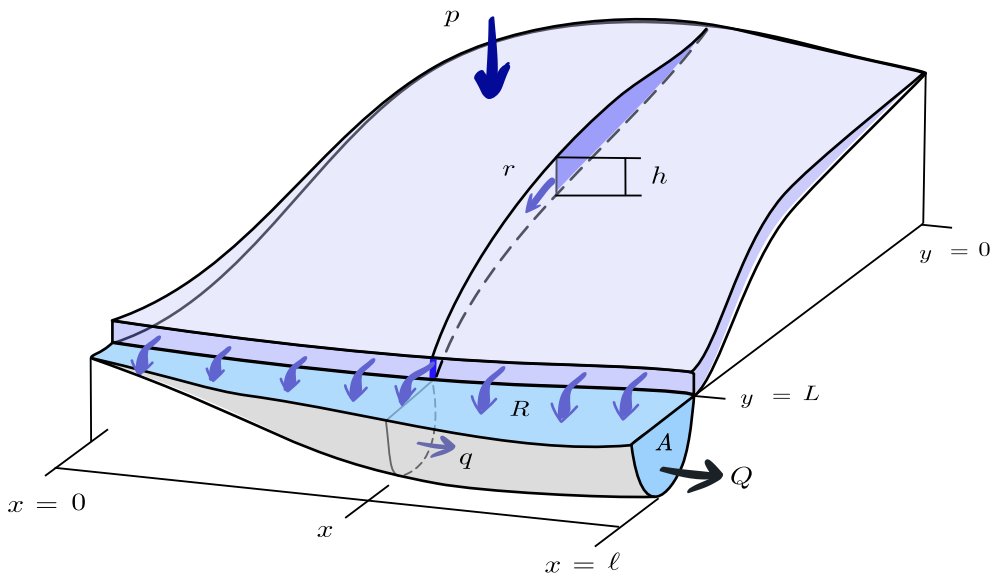

Figure 1. Schematic representation of the channel and the hillslopes, showing all the variables involved in the formulation of the stochastic rainfall-runoff model.

Rain on the hillslope is assumed to be net, runoff-producing rainfall. Namely, we assume that the watershed is subject to a precipitation intensity process $p(t)$ [L/T] composed of a sequence of random rainfall events, such that during the n-th event, an amount P_n [L] of water falls uniformly over the hillslope. Each P_n is assumed to be the total height of water (rainfall minus evapotranspiration) available for infiltration and surface runoff.

The generated runoff per unit of cross-hillslope length is $r(t, x, y)$ [L^2/T]. For simplicity we assume that r includes both Hortonian and Dunnian overland and sub-surface flow. Lateral flows are neglected, possibly preventing the extension of our framework to the case of steep mountain basins, where lateral flows play a decisive role in soil saturation dynamics during intense rainfall events (Botter et al., 2007b). At a point with coordinate x [L] along the stream, the discharge is $q(t, x)$ [L^3/T]. Here we will arrive at a linear model for the total runoff $R(t) := \int_0^\ell r(t, x, L)\,\mathrm{d}x$ and the discharge at the end of the river $Q(t) := q(t, \ell)$.

2.1. Conservation of mass and momentum

Conservation of mass for the hillslope-channel system can be written as equations for the time derivative of the cross-sectional area of the channel $A(t, x)$ [L^2] and the effective height of the hillslopes runoff $h(t, x, y)$ [L], in terms of the flows r and q:

$$\frac{\partial h}{\partial t} = -\frac{\partial r}{\partial y} + p(t), \quad \frac{\partial A}{\partial t} = -\frac{\partial q}{\partial x} + r(t, x, L), \tag{2-1}$$

along with the boundary conditions

$$r(t, x, 0) = 0, \quad q(t, 0) = 0. \tag{2-2}$$

A complete description of the model needs the appropriate equations of conservation of momentum on the hillslopes and channel, or equivalently, the specification of a relationship between h, A and the flows r, q respectively. The conservation of momentum equation is taken to be the simplest possible and the hillslope is modeled as a linear reservoir under the assumption that the runoff at each point is proportional to the total upstream storage, $r(t, x, y) = H \int_0^y h(t, x, \psi)\,\mathrm{d}\psi$. Namely, there exists $H > 0$ [1/T] such that $\frac{\partial r}{\partial y} = Hh$. Upon integration of the conservation of mass equation on the hillslope one gets

$$\frac{1}{H}\frac{\partial}{\partial t}\int_0^L \frac{\partial r}{\partial y}\,\mathrm{d}y = \int_0^L \left(-\frac{\partial r}{\partial y} + p(t)\right)\,\mathrm{d}y,$$

which by (2-2) and after further integration with respect to x, yields

$$\frac{\mathrm{d}R}{\mathrm{d}t} = H\left(-R + ap\right). \tag{2-3}$$

The equation for conservation of mass in the channel can be integrated along the longitudinal variable to obtain

$$\frac{\partial}{\partial t}\int_0^\ell A(t,x)\,\mathrm{d}x = -Q(t) + R(t). \tag{2-4}$$

The momentum conservation equation for the channel can be cast as a relationship between the total storage $\int_0^\ell A(t,x)\,\mathrm{d}x$ and the discharge $Q(t)$. For example, Menabde and Sivapalan (2001) use Chezy's resistance law $q = C(S/W)^{1/2}A^{3/2}$ (where W [L] denotes the wet perimeter, S denotes the stream slope and C [$\mathrm{L}^{1/2}$/T] is the Chezy coefficient) along with the simplification $\frac{\partial A}{\partial x} = \frac{\partial W}{\partial x} = \frac{\partial W}{\partial t} = 0$ to obtain

$$\frac{\partial}{\partial t}\int_0^\ell A(t,x)\,\mathrm{d}x = \ell\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{2}{3}\ell\left(\frac{1}{C}\sqrt{\frac{W}{S}}\right)^{\frac{2}{3}} Q^{-\frac{1}{3}}\,\frac{\mathrm{d}Q}{\mathrm{d}t} = \frac{1}{K(Q)}\frac{\mathrm{d}Q}{\mathrm{d}t}. \tag{2-5}$$

Note that the assumption of uniform flow implies the absence of pronounced topographic effects and prevent from the use of more detailed models exploiting topographic properties (Botter et al., 2007b). Equations (2-4) and (2-5) yields to

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = K(Q)\left(-Q + R\right). \tag{2-6}$$

Here we take one step further and suppose $K(Q) \equiv K$ is constant. This is equivalent to assuming $\frac{\partial A}{\partial x} = \frac{\partial v}{\partial x} = \frac{\partial v}{\partial t} = 0$, where v [L/T] is the water velocity along the channel.

2.2. Dynamical interpretation of H and K

Since the quantities H and K will play an important role in what follows, we now provide a dynamical interpretation of them. We define the residence time as the time spent by a water particle within the hillslope (or the channel), seen as the control volume for the underlying transport process (Botter et al., 2007a).

We denote the position of a water particle flowing down the hillslope as $y(t)$ [L] and its velocity as $w = \frac{\mathrm{d}y}{\mathrm{d}t}$ [L/T]. By continuity, $r = wh$. The assumption $\frac{\partial r}{\partial y} = Hh$ is therefore equivalent to

$$\frac{\partial w}{\partial y} = H, \quad \frac{\partial h}{\partial y} = 0,$$

which along with the boundary conditions (2-2) yields to $w = Hy$. Thus, $y(t)$ satisfies $\frac{\mathrm{d}y}{\mathrm{d}t} = Hy(t)$ and $y(t) = y_0 e^{Ht}$ with y_0 denoting the initial position of the water particle. The time it takes for a particle with initial position y_0 to reach the channel at $y(t) = L$ is

$$t = \frac{1}{H}\log\left(\frac{L}{y_0}\right),$$

and its mean can be computed by integrating over all the possible values of the initial position y_0

$$\bar{t} = \frac{1}{L}\int_0^L \frac{1}{H}\log\left(\frac{L}{y_0}\right)\,\mathrm{d}y_0 = \frac{1}{H}.$$

The constant H can therefore be interpreted as the inverse of *hillslope mean residence time*.

Now, denote the velocity of the flow down the channel as v [L/T]. By continuity and the assumption that $\frac{\partial v}{\partial x} = 0$, we have that $Q = vA$. Therefore, the assumption $\frac{\partial v}{\partial t} = 0$ is equivalent to

$$\frac{\partial Q}{\partial t} = v\frac{\partial A}{\partial t}.$$

Recall from (2-5) that

$$\ell\frac{\partial A}{\partial t} = \frac{1}{K}\frac{\partial Q}{\partial t},$$

then, the constant K [1/T] is then given by the inverse of the mean time it takes to transverse the channel, $K = \frac{v}{\ell}$.

This results are equivalent to assuming that the catchment-scale runoff and discharge are characterized by an exponential residence time distribution, and that all uncertainties related to the transport processes occurring within the hillslopes and the channel are parametrized by H and K respectively. See McGuire et al. (2005) and references therein.

2.3. Poissonian precipitation

The uncertainty in the model comes solely from the precipitation process p in (2-3) which, at time scales of interest here, can be approximated as a random sequence of instantaneous events. As noted in Botter et al. (2008), this implicitly postulates that the size of the considered basin is smaller than the correlation scale of rainfall events and that the timescales of the process of interest are greater than the characteristic duration of single rainfall events. We therefore write p as a random sums of impulses occurring at random times $T_1, T_2, \ldots$, namely

$$p(t) := \sum_{n=0}^{\infty} P_n \delta(t - T_n), \tag{2-7}$$

where $\{P_n : n = 1, 2, \ldots\}$ is the sequence of independent and identically distributed (i.i.d) rainfall amounts, having common but arbitrary probability density function

$$f(y)\,\mathrm{d}y = \mathbb{P}(P_n \in \mathrm{d}y), \quad n = 1, 2, \ldots$$

The inter-arrival times $\{T_0, T_1 - T_0, T_1 - T_2, \ldots\}$ are assumed to be independent and identically distributed with fixed exponential distribution with mean $1/\lambda > 0$ [1/T]. The counting process

$$N(t) := \max\{n \geq 0 : T_n \leq t\},$$

is a Poisson process, while $\int_0^t p(s)\,\mathrm{d}s = \sum_{n=1}^{N(t)} P_n$ is a compound Poisson process of intensity λ and jump distribution f.

The formulation above concerns the case in which an order-one watershed is subject to a random, yet *statistically stationary precipitation regime* $p(t)$ of instantaneous Markovian events. Namely, we are interested in the hourly to daily evolution of the precipitation, runoff and discharge within a months to year-long season where the statistical properties of precipitation and the geomorphological properties of the catchment can be assumed constant.

We are well aware that no precipitation event is instantaneous, however this assumption is reasonable when the time scale of the storm event is much smaller than the timescale of other continuous time processes. We thus regard a precipitation event as 'instantaneous' if its duration is much smaller than the the time scale of the hydrological response, defined as the mean water retention time in the catchment (Botter et al., 2013).

Note also that the physical parameters H and K of the model are also assumed deterministic, although in reality are highly uncertain, as illustrated in chapter 4. Botter (2010) consider a similar model where the residence times are assumed random spatially. See also (Ramirez and Constantinescu, 2020)

2.4. Solution

We now explicitly solve the 2-dimensional system of equations given by (2-3) and (2-6) for the two processes of interest: $Q(t)$ denoting the total stream flow at the most downstream point of the river, and $R(t)$ denoting the total runoff from the hillslopes into the river. The model is succinctly written in terms of the following two-dimensional stochastic differential equation

$$\mathrm{d}\boldsymbol{X}(t) = \mathbf{M}\boldsymbol{X}(t)\,\mathrm{d}t + \mathrm{d}\boldsymbol{Y}(t), \tag{2-8}$$

where

$$\boldsymbol{X}(t) = \begin{bmatrix} Q(t) \\ R(t) \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} -K & K \\ 0 & -H \end{bmatrix}, \quad \boldsymbol{Y}(t) = \sum_{n=1}^{N(t)} \begin{bmatrix} 0 \\ HaP_n \end{bmatrix}.$$

If $-\mathbf{M}$ is a matrix with non-negative eigenvalues, then the only solution to the stochastic differential equation (2-8) is the stochastic process

$$X(t) = e^{\mathbf{M}t}\boldsymbol{X}(0) + \int_0^t e^{\mathbf{M}(t-s)}\,\mathrm{d}Y(s), \quad t \geq 0.$$

The eigenvalues of $-\mathbf{M}$ are $\lambda_1 = K$ and $\lambda_2 = H$, then by integration with respect to $\boldsymbol{Y}$ as in

(A-14), we obtain the explicit solution

$$\boldsymbol{X}(t) = e^{\mathbf{M}t}\boldsymbol{X}(0) + \sum_{n=1}^{N(t)} e^{\mathbf{M}(t-T_n)} \begin{bmatrix} 0 \\ HaP_n \end{bmatrix}, \tag{2-9}$$

whit

$$e^{\mathbf{M}t} = \sum_{k=0}^{\infty} \frac{\mathbf{M}^k t^k}{k!}.$$

The stochastic process $\boldsymbol{X}$ belongs to the family of *piecewise deterministic Markov process* (Davis, 1984), which owes to the fact that the sample paths of $\boldsymbol{X}$ evolve deterministically between precipitation events and are only perturbed at the storm times $\{T_n : n \geq 1\}$: $R(t)$ jumps by a random amount at each T_n, while $Q(t)$ suffers a discontinuity in the derivative as shown in Figure 6. $\boldsymbol{X}$ is also a process of the *Ornstein-Uhlenbeck type* since it satisfies the stochastic differential equation (2-8) and is driven by the process of pure jumps $\{\boldsymbol{Y}(t),\, t \geq 0\}$.

Note that the assumption (2-7) for the rainfall field is a mathematical necessity: the exponential distribution for the inter-arrival times $T_{n+1} - T_n$ ensures that $\boldsymbol{X}$ is a Markov process and thus a solution to a stochastic differential equation of the form (2-8).

2.5. Related approaches in the literature

The analysis of conceptual stochastic models in hydrology has a long history going back to the seminal work of Eagleson (1972). Here we review some of the most significant approaches that have used solvable stochastic models for rainfall-runoff processes, and how they relate to the model and results presented here.

One distinct approach, pioneered by Koch (1985), has used the classical concept of the unit hydrograph (Dooge, 1959) for the probabilistic modeling of a catchment's response to random precipitation fields via 'shot noise processes'. There, scaled versions of a prescribed deterministic unit hydrograph are randomly aggregated according to a precipitation field modelled typically as a compound Poisson process of instantaneous events (see Claps et al. (2005) for a review). Although used mostly for simulation purposes as in Morlando et al. (2016), shot noise processes are simple enough so that explicit mathematical results can be derived. For example, and with clear relation to the work presented here, in Konecny (1992) the authors consider a shot noise process with a unit hydrograph of fixed exponential decay and derive the invariant distribution of discharge peaks. Shot noise models with instantaneous rainfall events and exponentially decaying hydrographs are the one-dimensional version of model developed by Ramirez and Constantinescu (2020).

The Geomorphological Instantaneous Unit Hydrograph (GIUH) has provided hydrologists with a framework to describe the response of a catchment in terms of the geomorphology of its river

network (Rodríguez-Iturbe and Valdes, 1979; Gupta et al., 1980). The expression for the hydrograph in a GIUH model is interpreted as the sum over all possible paths leading to the basin's outlet, of the probability densities of travel times associated with each path, times the probability of the path. The travel time density of each path is, in turn, given by the convolution of densities of travel times of individual hillslopes and streams found along the path. See Gupta et al. (1980, equations 15-16). As for the particular form of the travel time distribution, several explicit mathematical models have been formulated (see Van der Tak et al., 1989; Rinaldo et al., 1991; Saco and Kumar, 2002), including the original proposal by Rodríguez-Iturbe and Valdes (1979) of exponentially distributed travel times. Our formulation is equivalent to the random aggregation of GIUH specialized on the case of an order one catchment with exponentially distributed travel times on the hillslope and the channel.

In the seminal work by Rodriguez-Iturbe et al. (1999), the authors derive equations for at-a-point soil moisture dynamics driven by stochastic rainfall in the form of Poissonian instantaneous events with exponentially distributed amounts. Following this line of research, in Botter et al. (2007a) and Botter et al. (2007b), the authors derive and validate analytic expressions for the dynamics of soil moisture and runoff at a point. As in here, the ultimate goal is studying the stationary distribution of runoff and its relationship to the statistical properties of rainfall and as mediated by the geophysical properties of the soil column. This particular model has been applied, extended and generalized in Botter et al. (2008, 2009); Botter (2010); Suweis et al. (2010); Basso et al. (2015). Under some scenarios, which we detail in chapter 5, some of the results found along this impressive line of research, are validated by our approach. There are also important differences and complementary aspects, which we now summarize:

i. The conceptual derivation of our model is based on the mass and momentum conservation equations for the distributed hillslope-channel system. We thus arrive at a linear two-tiered system, where the linearization can be traced back to hypotheses on the transfer of momentum for hillslope and channel flow.

ii. All of our mathematical derivations consider an arbitrary probability density f of the rainfall amount P, and we give necessary and sufficient conditions on f, for the existence of an invariant distribution of discharge. We explicitly include results in what follows for three very different families of distributions.

iii. In Botter et al. (2007a), precipitation is considered at daily time scales, for which variations of river discharge only require the slow component of the hydrological response to be resolved. Here, we explicitly keep track of fast and slow components, as encoded by the inverse time-scales H and K respectively, thus allowing for a detailed mathematical analysis of the effect that the time scale separation $H \ll K$ has on different aspects of the model. We show that, in fact, our formulation reduces to that of Botter et al. (2007a) in the limit

as $H/K \to 0$. Furthermore, for a large range of H/K, we show that the single-reservoir model approximates very well the results of the two-reservoir system.

iv. Two key findings of Botter et al. (2007a) are mathematically demonstrated via our approach, for arbitrary f and ratio H/K, in fact. First the existence of the threshold $\lambda/H = 1$ controlling the shape of the invariant distribution of discharge, and second the analytical relationship between the coefficients of variation of discharge and rainfall.

v. The analysis of the extreme events and the scaling structure of discharge in sections 5.2 and 5.3 are, to the best of our knowledge, a novel contribution. We do it here applying the mathematical results of Ramirez and Constantinescu (2020) to the case of an order one watershed.

vi. The return period is one of the fundamental concepts for the applications of hydrology in engineering, as it is used for hydrological design and risk analysis. In section 5.4 we provide a mathematical interpretation of this concept by extending its definition for a sequence of independent realizations of a random variable, to the context of a time continuous stochastic process. We also propose a new expression to estimate it based on the results of chapter 3.

vii. Lastly, our model extends naturally to watersheds with arbitrary and heterogeneous river networks via the framework of Gupta and Waymire (1998). The mathematical treatment of this general case is the subject of Ramirez and Constantinescu (2020).

3. The invariant density

In this chapter we determine the invariant probability density of discharge at the watershed outlet. Conditions for the existence and the characterization of invariant distributions for Ornstein-Uhlenbeck type processes are given in Sato and Yamazato (1984, Theorems 4.1-4.2). We now apply their result to X.

Theorem 3.1 (Sato and Yamazato (1984))**.** A necessary and sufficient condition for the existence of a unique invariant density for $\boldsymbol{X}$ is

$$\int_1^\infty \log(y) f(y)\,\mathrm{d}y < \infty. \tag{3-1}$$

If (3-1) holds, then the distribution of $\boldsymbol{X}(t)$ converges weakly to a distribution with Laplace transform given by

$$\tilde{\boldsymbol{g}}_{\boldsymbol{X}}(s_1, s_2) = \exp\left\{-\frac{\lambda}{H}\int_0^1 \frac{1-\tilde{f}(aH(s_2 u + s_1 m(u))}{u}\,\mathrm{d}u\right\}, \tag{3-2}$$

where f is the common density of rainfall amounts, $\tilde{f}$ its Laplace transform and the function $m(u)$ is given by

$$m(u) := \frac{1}{1-\beta}\left(u - u^{1/\beta}\right), \quad \beta := \frac{H}{K}. \tag{3-3}$$

Moreover, the process $\boldsymbol{X}$ is ergodic, and $\boldsymbol{g}_{\boldsymbol{X}}$ is its unique invariant density. See also Konecny (1992) and Ramirez and Constantinescu (2020).

Theorem 3.1 states that under suitable conditions on f, and for any initial condition $\boldsymbol{X}(0) \in \mathbb{R}_+^2$, the distribution of $\boldsymbol{X}(t)$ converges to a limiting distribution as $t \to \infty$. Moreover, the limiting distribution is invariant (also called stationary) in the sense that $\boldsymbol{X}(0)$ is distributed as such, the distribution of $\boldsymbol{X}(t)$ will remain unchanged for all $t \geq 0$. Here and in what follows, the limit as $t \to \infty$ refers to the passage of enough days for the system to forget its initial condition and achieve convergence in distribution. By invariant distribution, we thus mean seasonal invariance: the statistical characterization of the hydrological response of the catchment to the seasonal conditions of precipitation.

The random variable of most interest in $\boldsymbol{X}(t)$ is its first entry $Q(t)$, the discharge at the watershed's outlet. While the distribution of $Q(t)$ evolves with time, and depends on the initial

condition $Q(0)$, here we are concerned with the limiting invariant density g of the process $Q(t)$:

$$g(x)\,\mathrm{d}x = \lim_{t\to\infty} \mathbb{P}(Q(t) \in \mathrm{d}x), \quad x > 0, \tag{3-4}$$

and its Laplace transform $\tilde{g}$ with respect to discharge, namely

$$\tilde{g}(s) = \int_0^\infty e^{-sx} g(x)\,\mathrm{d}x, \quad s > 0. \tag{3-5}$$

Note that the Laplace transform in (3-2) is bi-dimensional, $\tilde{\boldsymbol{g}}_{\boldsymbol{X}}(s_1, s_2) = \mathbb{E}_{\boldsymbol{g}} e^{s_1 Q + s_2 R}$. The invariant distribution of Q is characterized by the following formula for $\tilde{g}$, which is obtained as $\tilde{g}(s) = \tilde{\boldsymbol{g}}_{\boldsymbol{X}}(0, s)$:

$$\tilde{g}(s) = \exp\left\{-\frac{\lambda}{H}\int_0^1 \frac{1 - \tilde{f}(Has\,m(u))}{u}\,\mathrm{d}u\right\}. \tag{3-6}$$

Some remarks are in order. Note first that $\tilde{g}(-s)$ is the moment generating function of the invariant distribution of Q and thus characterizes it. Once $\tilde{g}$ has been computed, the actual density g may be recovered by a numerical inversion algorithm. For all computations reported here, we use the classical algorithm by Zakian (1969). The process $\boldsymbol{X}$ can also shown to be ergodic, and thus g is the ergodic limit of the distribution of $Q(t)$ as $t \to \infty$ (see Kallenberg, 2002). The significance of this result is the implication that if the statistical properties of the precipitation field are stationary for a sufficiently long period of time, the watershed will eventually attain a statistical invariant regime determined by g. Furthermore, time averages of any functional of Q can be approximated by corresponding ensemble averages with respect to g.

3.1. Dimensionless analysis of the invariant density

In order to perform further analyses on the invariant distribution g of Q, let us calculate the invariant densities of the dimensionless precipitation amounts and discharge. Namely, consider the common mean of rainfall amounts $\mathbb{E}[P_n] = \int_0^\infty x f(x)\,\mathrm{d}x$. From now on we just write $\mathbb{E}[P]$. Define the normalized rainfall amounts $\hat{P} = P/\mathbb{E}[P_n]$, having common probability density ϕ. Also, define the normalized discharge process $\hat{Q}(t) = Q(t)/\lambda a \mathbb{E}[P]$, and denote γ its invariant probability density.

$$\hat{P} = \frac{P}{\mathbb{E}[P_n]}, \quad \hat{Q}(t) = \frac{Q(t)}{\lambda a\,\mathbb{E}[P]}. \tag{3-7}$$

Equation (A-9) in the Appendix relates the probability densities of the original and the normalized random variables. Equation (A-10) do so for the Laplace transforms of the probability densities. According to those formulas, for the precipitation we have

$$\phi(x) = \mathbb{E}[P] f\big(x\,\mathbb{E}[P]\big), \quad \tilde{\phi}(s) = \tilde{f}\left(\frac{s}{\mathbb{E}[P]}\right).$$

For the Laplace transform of the discharge we have

$$\tilde{\gamma}(s) = \tilde{g}\left(\frac{s}{\lambda a\, \mathbb{E}[P]}\right),$$

then equation (3-6) for $\tilde{g}$ can be written in terms of the Laplace transforms of ϕ and γ as follows

$$\tilde{\gamma}(s) = \exp\left\{-\frac{1}{\eta}\int_0^1 \frac{1-\tilde{\phi}\left(s\eta\, m(u)\right)}{u}\,\mathrm{d}u\right\}, \quad \text{where } \eta = \frac{H}{\lambda}. \tag{3-8}$$

Table 1. Families of probability distributions commonly used as models for the precipitation in the tropics.

	Pareto (κ, θ) Type I	Gamma (ω, ρ)	Inverse Gaussian (μ, σ)
$f(x)$	$\theta\kappa^{\theta}x^{-\theta-1},\ x \geq \kappa$	$\frac{1}{\rho^{\omega}\Gamma(\omega)}e^{-x/\rho}x^{\omega-1}$	$\sqrt{\frac{\sigma}{2\pi x^3}}\exp\left(-\frac{\sigma(x-\mu)^2}{2x\mu^2}\right)$
$\tilde{f}(s)$	$\theta E_{1+\theta}(\kappa s)$	$(1+\rho s)^{-\omega}$	$\exp\left(\frac{\sigma}{\mu}\left(1-\sqrt{1+\frac{2\mu^2 s}{\sigma}}\right)\right)$
$\mathbb{E}[P]$	$\frac{\kappa\theta}{\theta-1}$	$\rho\omega$	μ
$\phi(x)$	$\theta\left(\frac{\theta-1}{\theta}\right)^{\theta}x^{-\theta-1},\ x \geq \frac{\theta-1}{\theta}$	$\frac{\omega^{\omega}}{\Gamma(\omega)}e^{-x/\omega}x^{\omega-1}$	$\sqrt{\frac{\sigma/\mu}{2\pi x^3}}\exp\left(-\frac{\sigma/\mu(x-1)^2}{2x}\right)$
$\tilde{\phi}(s)$	$\theta E_{1+\theta}\left(\frac{\theta-1}{\theta}s\right)$	$\left(1+\frac{s}{\omega}\right)^{-\omega}$	$\exp\left(\frac{1-\sqrt{1+2s\alpha}}{\alpha}\right),\ \alpha=\frac{\mu}{\sigma}$
Parameters	$2.3 < \theta < 11$ $0.028 < \kappa < 2.27$	$0.7 < \omega < 100$ $0.5\times10^{-3} < \rho < 3.6$	$0.05 < \mu < 2.5$ $0.035 < \sigma < 250$

Note. The Parameters column displays approximate ranges for the parameters of each distribution that would fit the mean and the coefficient of variation reported in Álvarez-Villa et al. (2011). On the expression for $\tilde{f}$ of the Gamma distribution, $E_n(z)$ denotes the generalized exponential integral function.

3.2. Qualitative analysis of the invariant density

In our subsequent analysis we will consider the invariant distribution g of the discharge Q under three different parametrizations of the probability distribution f of the rainfall amount. These are, the Pareto of type I, Gamma and inverse Gaussian. The Pareto distribution has heavy tails and only finitely many moments. The Gamma distribution is a generalization of the exponential

distribution and features exponentially decaying tails. These two families, along with the lognormal distribution, are commonly used as models for the precipitation in the tropics (Cho et al., 2004; Salisu et al., 2010). Since equation (3-2) requires an explicit expression for the Laplace transform of f, we use the inverse Gaussian distribution instead of the lognormal. Chapter 4 shows an application of the former to a set of rainfall data in the Colombian Andes. The families of distributions, along with typical parameter ranges are summarized in Table 1. In all cases, condition (3-1) holds for the probability density functions f and ϕ.

Figure 2 shows plots of the invariant distribution γ for the non-dimensional discharge Q^*. Each plot was obtained by numerically inverting its Laplace transform given by (3-8) with ϕ equal to the non-dimensional forms of the probability densities listed in Table 1. Two features are most notable. First, changes of over at least three orders of magnitude for the value of β play an insignificant role on the qualitative behavior of γ. Secondly and more interesting, is that $\eta = \eta_c = 1$ represents a threshold where γ (and thus g) changes from unimodal to a monotone decreasing function. This important transition was observed by Botter et al. (2007a) in the context of soil moisture and discharge dynamics for the case when P is exponentially distributed. We show here that this critical value holds also for our model regardless of the value of β, and extends to other distributions.

Theorem 3.2. Suppose the distribution of P satisfies (3-1) and let g be the probability density of Q. Then, if $0 < \eta < 1$, $\lim_{x\to 0^+} g(x) = 0$.

Proof. It suffices to prove that $\gamma(x) \to 0$ as $x \to 0$, which by the initial value theorem is equivalent to proving $s\tilde{\gamma}(s) \to 0$ as $s \to \infty$ (see for example Beerends et al., 2003). Since $\tilde{\phi}(s) = \mathbb{E}e^{-sP}$, Jensen's inequality ensures that $\tilde{\phi}(s) \geq e^{-s}$ for all s. Moreover $m(u) \leq m_0(u) := \frac{u}{1-\beta}$ for $0 \leq u \leq 1$. We thus get the following bound

$$s\tilde{\gamma}(s) \leq s \exp\left\{-\int_0^1 \frac{1 - e^{-s\eta m_0(u)}}{\eta u}\,\mathrm{d}u\right\} \tag{3-9}$$

$$= s \exp\left\{-\frac{1}{\eta}\left(\gamma_e - \mathrm{Chi}\left(\frac{s\eta}{\beta - 1}\right) + \mathrm{Shi}\left(\frac{s\eta}{\beta - 1}\right) + \log\left(\frac{s\eta}{-\beta}\right)\right)\right\}, \tag{3-10}$$

where γ_e denotes the Euler gamma constant, and Shi and Chi are the hyperbolic sine and hyperbolic cosine integral functions, respectively. The function on the right hand side of (3-10) converges to zero if $0 < \eta < 1$ and diverges to infinity for $\eta > 1$. □

In terms of hydrological processes, η can be understood as the ratio between the mean time between precipitation events and the mean residence time on the hillslopes; it represents the link between the precipitation regime and the geomorphology of the basin, and it is the most important dimensionless number of the model. We now propose an explanation for the critical value $\eta_c = 1$. When $\eta < 1$, the mean time between rainfall events is less than the mean residence time on the hillslopes. This implies that the rate of water deposition is greater than that of evacuation,

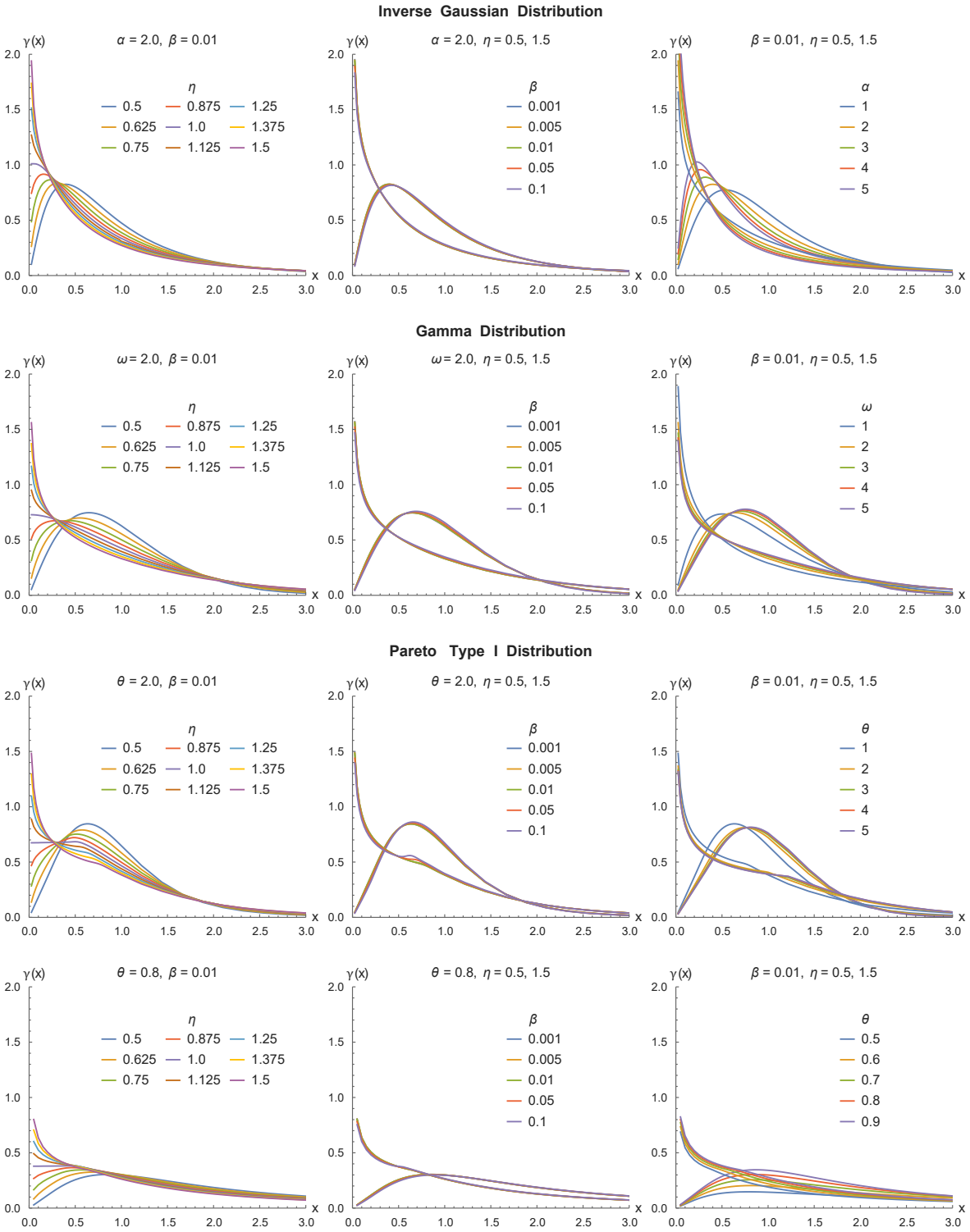

Figure 2. Plots of $\gamma(x)$ for different combinations of the parameters η, β, and each probability density considered for $\phi(s)$. The leftmost panels shows the transition of γ from a unimodal to a monotone decreasing distribution at the critical $\eta_c = 1$. In the center and right panels, two curves are plotted for each value of the varying parameter, the unimodal curves correspond to $\eta = 0.5$ while the monotone curves have $\eta = 1.5$.

therefore runoff can accumulate on the slopes before reaching the channel and the invariant density of the discharge becomes more massive at higher values. Conversely, when $\eta > 1$, water on the hillslopes reaches the channel faster than the flow due to rain, therefore there is less accumulation of runoff and the invariant density becomes more massive at the smaller values of the flow.

4. A case study of the discharge distribution

We now focus on presenting and discussing the results of the application of the conceptual model described in chapter 2 to a specific case study of an order-one tropical watershed: La Gruta catchment, located at the headwaters of the Pesca river, at Boyacá, Colombia, on the eastern mountain range of the Andes. See Figure 3. Our aim is to illustrate that whenever a small catchment is under a rainfall regime that approximately satisfies the hypotheses of the model, the probability distribution of discharge is in fact adequately approximated by the invariant distribution given by (3-6).

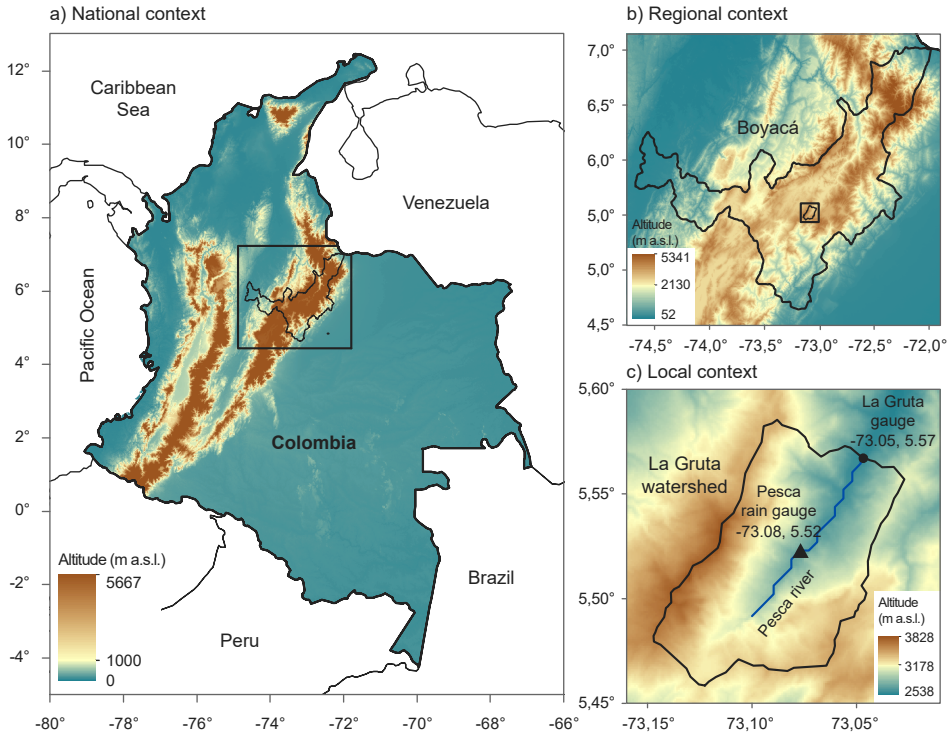

Figure 3. Location of the study watershed over Colombia. (a) National context, (b) regional context and (c) local context delimiting La Gruta basin. The geographical location of the discharge and rain gauges and the recorded time series were provided by IDEAM. The DEM and the digital river network used to produce this figure were downloaded from HydroSHEDS.

La Gruta catchment has an estimated drainage area of $a = 103.79$ km^2, an average elevation of 3160 m a.s.l., the length of the channel is $\ell = 19.56$ km, and the average slope of the Pesca River is 3.72%. The predominant economic activities of the area are agriculture and livestock farming, so the hillslopes are mainly used as pastures and croplands. The annual cycle of precipitation is bimodal with wet seasons during March to May (MAM) and September to October (SON) (Urrea et al., 2019), and the average precipitation is between 800 and 1200 mm/year (Álvarez-Villa et al., 2011). The rain gauge Pesca is located right in the middle of the watershed (5.52 latitudinal decimal degrees and -73.08 longitudinal decimal degrees, 2678 m.a.s.l). The available precipitation records comprise the period from January 1968 to December 2001 with hourly resolution. The outlet of the watershed was chosen as the location of the stream flow gauge La Gruta, (5.57 latitudinal decimal degrees, -73.05 longitudinal decimal degrees, 2500 m a.s.l.) which recorded discharge at daily resolution from February 1968 to December 2010. Both gauges are operated by the hydro-meteorological service of Colombia: Instituto de Hidrologia, Meteorologia y Estudios Ambientales (IDEAM).

The model presented in chapter 2 is based on two assumptions to be verified: the representation of rain as a compound Poisson process, and the stationarity of rainfall and discharge processes. For the precipitation process $\{P(t) : t \geq 0\}$ to be represented as a compound Poisson process, the duration of each rainfall event should be sufficiently short as to be regarded as instantaneous with respect to the time scale of the hydrological response, and the times between events $\{T_0, T_1 - T_0, T_1 - T_2, \dots\}$ must be i.i.d. exponentially distributed. Additionally, the rainfall amounts must come from a stationary random process and the probability density function f of each $\{P_n : n = 1, 2, ...\}$ must be parametrized. We validate the assumptions with hypothesis testing techniques with a significance level equal to 0.05. For the first assumption, we perform a Kolmogorov-Smirnov goodness-of-fit test for the null hypothesis that two distribution functions are the same. The desired result is therefore not to reject the null hypothesis with a very high p-value, bigger or equal than 0.8, which have the weakest possible evidence against our statistical fitting. On the other hand, in order to perform a stationarity test without contradicting the implicit Markovian nature of the discharge and the precipitation processes of the model, we assume that the corresponding time series are samples of an order-one autoregressive processes and we perform a Dickey–Fuller F unit root test. Rejecting the null hypothesis allows concluding that the data comes from a stationary time series. In this case, very small p-values are preferred.

Considering the precipitation seasonality and to ensure large enough data samples and relaxation into a stationary regime, we performed the hypothesis testing for time intervals of at least 80 days. To that aim, we transformed the time series of hourly precipitation into a time series of instantaneous precipitation events, by assuming that the n-th event occurred during the n-th block of consecutive hours with rainfall records. We took the centroid of the consecutive hours as the arrival time T_n and the sum of the corresponding rainfall records as the column P_n of water uniformly dropped over the hillslopes around the stream. We did not reconstruct missing records

of rainfall or discharge, but only considered time intervals with 15% or less of missing data. The following two time intervals, which we henceforth refer to as *validation windows*, were selected:

- June 11, 1968 - August 29, 1968 (JJA 1968). With a record length of 80 days and no missing data.
- September 26, 1989 - December 23, 1989 (OND 1989). With a record length of 89 days and 13.62% of missing precipitation data.

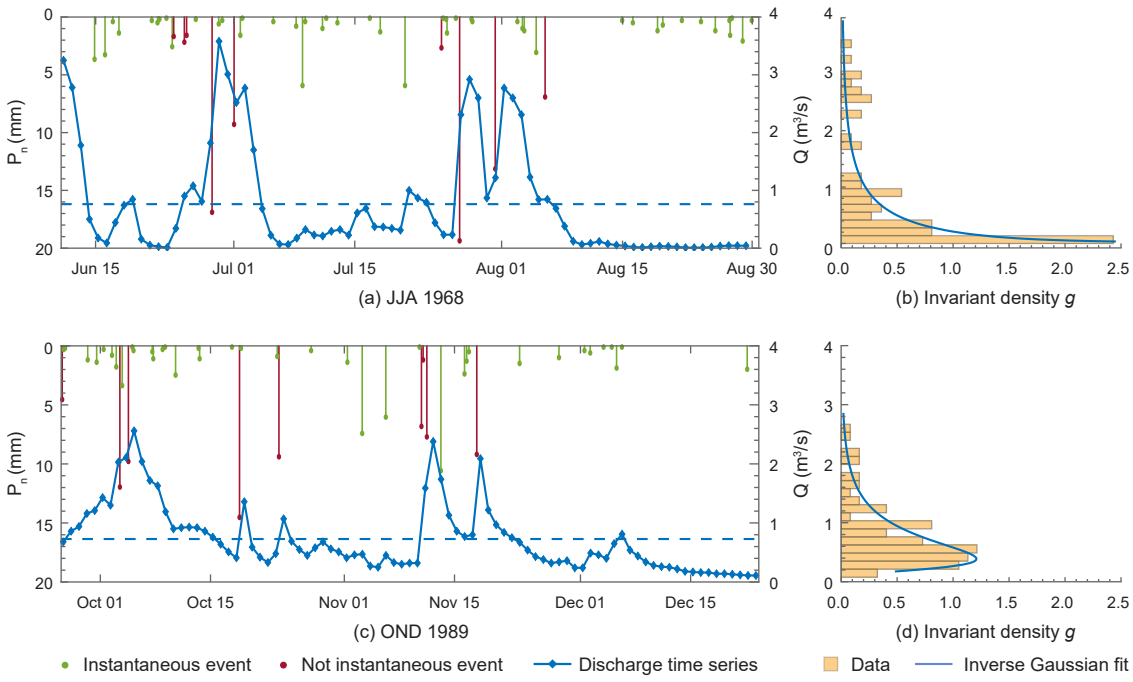

Figure 4. Precipitation (mm) and discharge (m^3/s) time series for the considered validation windows (a) JJA 1968 and (c) OND 1989. Green points denote instantaneous precipitation events and red points denote the non-instantaneous ones. The right panels show the comparison between the discharge histogram and the computed invariant density g which has the best p-value obtained in the validation windows (b) JJA 1968 and (d) OND 1989. For JJA 1968 the best p-value is 0.9345 and was obtained with $K = 0.92$ hr^{-1} and $H = 0.046$ hr^{-1}. For OND 1989 the best p-value is 1 and was obtained with $K = 0.92 hr^{-1}$ and $H = 0.0058$ hr^{-1}.

For the chosen validation windows, more than 80% of the events recorded by the rain gauge lasted for 3 consecutive hours or less, and about 50% of the events lasted for less than one hour. Since the mean recession time of La Gruta watershed is of the order of few days, and in order to maximize the rainfall sample, for this study case we regard a precipitation event as *instantaneous*, with respect to the daily time scale of the hydrological response, if its duration is less or equal to 3 hours. The corresponding time series of precipitation events and daily discharge are shown in Figure 4(a,c). Green points denote the instantaneous precipitation events and red points denote the

Table 2. Results for the the statistical tests and parameter estimation

	JJA 1968		OND 1989	
Test	p-value	Parameters	p-value	Parameters
$\{T_{n+1} - T_n\} \sim \exp(\lambda)$	0.9987	$\lambda = 0.025$ hr^{-1}	0.8574	$\lambda = 0.018$ hr^{-1}
$\{P_n\}$ is stationary	7.8600×10^{-7}		6.4240×10^{-5}	
$\{P_n\} \sim \text{IG}(\mu, b)$	0.7812	$\mu = 1.07 \times 10^{-3}$ m $b = 4.50 \times 10^{-4}$ m	0.5792	$\mu = 1.45 \times 10^{-3}$ m $b = 4.05 \times 10^{-4}$ m
$\{Q(t)\}$ is stationary	0.0224		0.0321	
$Q(t) \sim g(H, K)$	0.9345	$H = 0.046$ hr^{-1} $K = 0.92$ hr^{-1}	1.0000	$H = 0.0058$ hr^{-1} $K = 0.92$ hr^{-1}

non-instantaneous ones, which were discarded in what follows. Given that large discharge records caused by those events were also related to some preceding instantaneous events, and that there are instantaneous events with magnitudes comparable to those of the non-instantaneous ones, we did not remove any discharge record.

The sequence of rainfall amounts was found to best fit inverse Gaussian (IG) distributions for both validation windows. The results of the hypothesis testing are reported in Table 2. Note that the hypothesis of a stationary precipitation regime implies via (5-1) that the watershed should attain hydrological balance. In this regard we obtained that during JJA 1968, $\lambda a \mathbb{E}P = 0.757$ m^3/s which approximates quite well the average discharge of 0.762 m^3/s reported during that period. Similarly, for the OND 1989 period, we obtained $\lambda a \mathbb{E}P = 0.741$ m^3/s and an average discharge of 0.728 m^3/s.

The parametrization for f and the estimated values for the catchment area a and the precipitation frequency λ, allowed us to compute $\tilde{g}$ in (3-6) as a function of the yet unknown inverse mean residence times H and K. The corresponding invariant distribution density g of Q can be then estimated by numerically inverting its Laplace transform $\tilde{g}$ via the classic algorithm in Zakian (1969). The computations of g were performed on a non-regular grid of the parameter space (K, H) constructed as follows: We consider an initial estimate of the form $K = v/\ell$ where v is the average water speed in the channel assumed here to take values ranging from 1 to 5 m/s in steps of 0.25 m/s. We then made $H = \beta K$ and considered equally spaced values of $0 < \beta < 1$. Figure 5 shows, as a function of (K, H), the p-values for goodness-of-fit tests between the invariant distribution g calculated from (3-6) and the empirical distribution of Q from the discharge data. For the JJA 1968 period, we were able to identify acceptable fits for $\beta \in [0.025, 0.2]$; the period OND 1989 required values one order of magnitude less than that, $\beta \in [0.0025, 0.02]$. The estimated parameters H, K for each validation period were taken as those from the analyzed grid that produced the highest p-value. See last row of Table 2.

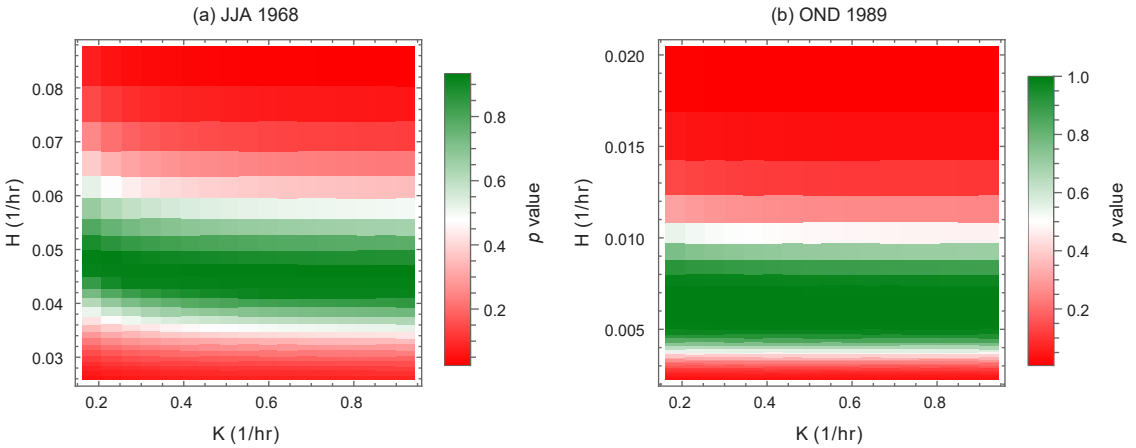

Figure 5. p-values for goodness-of-fit tests between the invariant distribution g and the empirical distribution of the discharge data, as a function of (K, H), for the validation windows (a) JJA 1968 and (b) OND 1989. The details about the non-regular grid for the parameter space (K, H) can be found in chapter 4.

Figure 4(b,d) shows the comparison between the empirical and analytic probability density functions corresponding to the chosen H, K parameters for each validation window. During the period JJA 1968 it rained on average 1.39 times more often and with an average amount of 0.74 times that of OND 1989, however the mean residence time on the hillslopes was 8 times higher during JJA 1968 than in OND 1989. This variations of H led to changes of $\eta = H/\lambda$ large enough as to completely change the behavior of the probability distribution of discharge. According to Botter et al. (2013), when flow-producing rainfall events are relatively frequent, such that their mean inter-arrival is smaller than the mean residence time of the contributing hillslopes $\lambda > H$, $(\eta < 1)$, the range of the discharge recorded by a streamflow gauge between two subsequent events is reduced, and a persistent supply is guaranteed to the stream from the hillslopes. Therefore, the discharge is weakly variable around the mean and quite predictable (See Figure 4(c,d)). When the mean interarrival between flow-producing rainfall events is larger than the mean residence time of the hillslopes $\lambda < H$, $(\eta > 1)$, a wider range of stream flows is observed between events because the stream is allowed to dry significantly before the arrival of new runoff. The temporal pattern of discharge is thus more unpredictable and the preferential state of the system is typically lower than the mean (See Figure 4(a,b)). The plots 4(b) and 4(d) suggest that our model was able to reproduce the changes, both in the mean and the shape of the density function of discharge, and to accurately capture both hydrological regimes.

Lastly, Figure 6 shows the comparison between the empirical and analytic recession curves of discharge for a period of 30 days of the OND 1989 validation window. The analytical model for $\boldsymbol{X}$ in (2-9) was forced by the precipitation data of that period and by an initial value equal to the first measured discharge. When comparing the process Q to the actual value of the discharge

time series, we observe that the model reasonably captured both the magnitude of the increment and the recession rate of the discharge after the precipitation events.

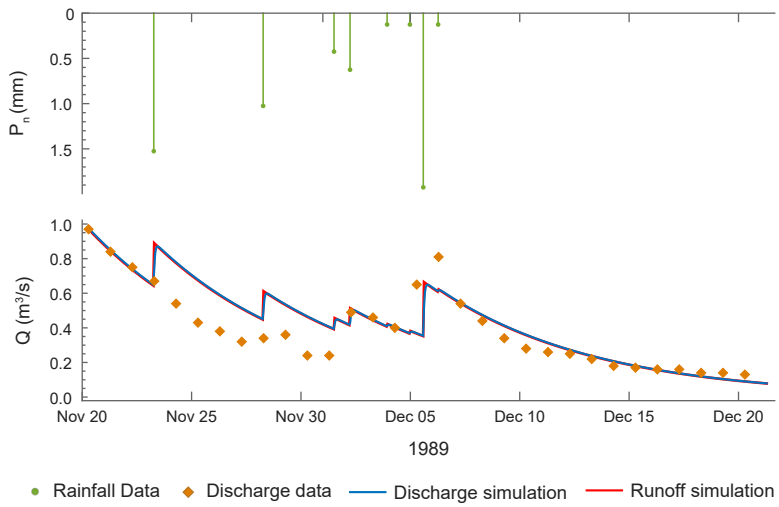

Figure 6. Simulated $Q(t)$ and $R(t)$ (m^3/s) for the order one watershed in Figure 3, during the time period comprised from November 20, 1989 to December 20, 1989. The hillslopes area is $a = 103.79$ km^2 and the geomorphological parameters are $H = 0.0058$ hr^{-1} and $K = 0.92$ hr^{-1}. Rainfall amounts P_n (mm) and occurrence times T_n (hr) were taken from the time series of instantaneous precipitation events in Figure 4, denoted by green points. The initial condition for $\boldsymbol{X}(t)$ was taken to be the first measured discharge of the considered time period. The orange diamonds denote the discharge data form the recorded time series at daily resolution.

5. Analysis of the invariant density

In this chapter we exploit the properties of the Laplace transform of the density g to derive important characteristics of the invariant distribution of Q. We also provide a rigorous mathematical framework to study and analyze some traditional concepts of the hydrology in the context of our model.

5.1. Moments

The first characteristic we analyze is the relation between the moments of the discharge, the moments of the precipitation and the geomorphological features of the watershed. In what follows, we will write $\mathbb{P}_g$ or $\mathbb{E}_g$ to denote probabilities or expectations with respect to the density g, namely conditioned on $Q(0)$ distributed as the invariant distribution of $Q(t)$, which means that the discharge is under an invariant regime and guarantees that the stochastic process $Q(t)$ is stationary.

We start with taking expectations throughout the stochastic differential equation (2-8) and using invariance in the form $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{\boldsymbol{g}}[\boldsymbol{X}(t)] = 0$, to compute the invariant mean of the discharge

$$\mathbb{E}_g[Q] = \lambda a\, \mathbb{E}[P_n], \tag{5-1}$$

where $\mathbb{E}[P_n] = \int_0^\infty x f(x)\,\mathrm{d}x$ denotes the common mean of rainfall amounts. From now on we just write $\mathbb{E}[P]$. Equation (5-1) for $\mathbb{E}_g Q$ makes explicit the fact that under a statistical invariant regime for P, the discharge will remain in a state of hydrological balance with the precipitation.

For arbitrary moments, we consider the normalized discharge $\hat{Q}$. Recall that the Laplace transform of γ is closely related to the moment generating function of the random variable $\hat{Q}$ (see equation (A-8)), and that we can compute moments of any order by differentiating

$$\mathbb{E}_\gamma[\hat{Q}^n] = \left.\frac{\mathrm{d}^n \tilde{\gamma}(-s)}{\mathrm{d}s^n}\right|_{s=0},$$

which yields to the following theorem.

Theorem 5.1. If $\hat{P}$ has n finite moments, then $\hat{Q}$ also has n finite moments given by

$$\mathbb{E}_\gamma[\hat{Q}^n] = B_n\left(\left\{\eta^{i-1}\, c_i(\beta)\, \mathbb{E}[\hat{P}^i]\right\}_{i=1}^n\right), \quad n = 1, 2, 3, \ldots, \tag{5-2}$$

where B_n denotes the n-th complete exponential Bell polynomial, and the constants $c_i(\beta)$ are given by

$$c_i(\beta) = \frac{1}{i} \prod_{k=1}^{i-1} \frac{k}{k + \beta(i-k)}, \quad i = 1, 2, 3, \dots \tag{5-3}$$

We use the convention $\prod_{k=1}^{i-1} x_k = 1$ is $i = 1$, so that $c_1(\beta) = 1$.

See section A.4 in the appendix for the definitions and properties of the Bell polynomials.

Proof. Denote the function inside the exponential in (3-8) by

$$h(s) := -\frac{1}{\eta} \int_0^1 \frac{1 - \tilde{\phi}(s\eta\, m(u))}{u} \, \mathrm{d}u,$$

then, the n-th invariant moment of $\hat{Q}$ is

$$\mathbb{E}_\gamma[\hat{Q}^n] = (-1)^n \left. \frac{\mathrm{d}^n e^{h(-s)}}{\mathrm{d}s^n} \right|_{s=0}, \quad n = 1, 2, 3, \dots \tag{5-4}$$

Faa di Bruno's formula for the n-th derivative of $e^{h(-s)}$, along with the property (A-18) of the Bell polynomials gives

$$\frac{\mathrm{d}^n e^{h(-s)}}{\mathrm{d}s^n} = (-1)^n \tilde{\gamma}(-s) \sum_{k=1}^{n} B_{n,k} \left(\left\{ \frac{\mathrm{d}^i h(-s)}{\mathrm{d}s^i} \right\}_{i=1}^{n-k+1} \right), \tag{5-5}$$

where $B_{n,k}$ is the incomplete exponential Bell polynomial.

Calculating the i-th derivative of $h(-s)$ and replacing $s = 0$, yields to

$$\begin{aligned}
\left. \frac{\mathrm{d}^i h(-s)}{\mathrm{d}s^i} \right|_{s=0} &= -\frac{1}{\eta} \int_0^1 -\frac{1}{u} \frac{\mathrm{d}^i}{\mathrm{d}s^i} \tilde{\phi}(-\eta s\, m(u)) \, \mathrm{d}u \Big|_{s=0} \\
&= -\frac{1}{\eta} \int_0^1 -\frac{1}{u} \mathbb{E}[(\eta\, m(u) \hat{P})^i] \, \mathrm{d}u \\
&= \eta^{i-1} \mathbb{E}[\hat{P}^i] \frac{\beta \Gamma(i) \Gamma\left(\frac{i\beta}{1-\beta}\right)}{(1-\beta)^i \Gamma\left(\frac{i}{1-\beta}\right)} \\
&= \eta^{i-1} \mathbb{E}[\hat{P}^i]\, c_i(\beta).
\end{aligned} \tag{5-6}$$

Note that $\frac{i}{1-\beta} = \frac{i\beta}{1-\beta} + i$. Since $i \in \mathbb{N}$, the recurrence property of the gamma function (A-21) gives

$$\Gamma\left(\frac{i}{1-\beta}\right) = \Gamma\left(\frac{i\beta}{1-\beta}\right) \prod_{k=0}^{i-1} \frac{k + \beta(i-k)}{1-\beta}.$$

Also, for natural numbers the gamma function is related to the factorial as follows

$$\Gamma(i) = (i-1)! = \prod_{k=1}^{i-1} k.$$

These two expressions yield to

$$\begin{aligned} c_i(\beta) =& \frac{\beta\Gamma(i)\Gamma\left(\frac{i\beta}{1-\beta}\right)}{(1-\beta)^i\Gamma\left(\frac{i}{1-\beta}\right)} \\ =& \frac{\beta}{(1-\beta)^i}(i-1)!\prod_{k=0}^{i-1}\frac{1-\beta}{k+\beta(i-k)} \\ =& \frac{1}{i}\prod_{k=0}^{i-1}\frac{k}{k+\beta(i-k)}, \quad i=1,2,3,\ldots \end{aligned}$$

Finally, replacing (5-6) and $s=0$ in (5-5), and then replacing the result in the equation (5-4), one arrives at the expression for $\mathbb{E}_g[\hat{Q}^n]$

$$\mathbb{E}_\gamma[\hat{Q}^n] = B_n\left(\left\{\eta^{i-1}\, c_i(\beta)\, \mathbb{E}[\hat{P}^i]\right\}_{i=1}^{n}\right), \quad n=1,2,3,\ldots$$

□

For instance, the first four moments of the normalized discharge $\hat{Q}$ are:

$$\begin{aligned} \mathbb{E}_\gamma[\hat{Q}] &= 1, \\ \mathbb{E}_\gamma[\hat{Q}^2] &= 1 + \frac{\eta\mathbb{E}[\hat{P}^2]}{2(1+\beta)}, \\ \mathbb{E}_\gamma[\hat{Q}^3] &= 1 + \frac{3\eta\mathbb{E}[\hat{P}^2]}{2(1+\beta)} + \frac{2\eta^2\mathbb{E}[\hat{P}^3]}{3(2+\beta)(1+2\beta)}, \\ \mathbb{E}_\gamma[\hat{Q}^4] &= 1 + \frac{3\eta\mathbb{E}[\hat{P}^2]}{(1+\beta)} + \frac{3\eta^2\mathbb{E}[\hat{P}^2]^2}{4(1+\beta)^2} + \frac{8\eta^2\mathbb{E}[\hat{P}^3]}{3(2+\beta)(1+2\beta)} + \frac{3\eta^3\mathbb{E}[\hat{P}^4]}{2(3+\beta)(2+2\beta)(1+3\beta)}. \end{aligned}$$

The arbitrary moments of Q can be easily found by multiplying the equation (5-2) by the discharge mean

$$\mathbb{E}_g[Q^n] = \left(a\lambda\,\mathbb{E}[P]\right)^n \mathbb{E}_\gamma[\hat{Q}^n], \quad n=1,2,3,\ldots \tag{5-7}$$

An implicit result of equations (5-2) and (5-7) is that, under the limiting distribution, the discharge will have exactly as many moments as the precipitation. This is reflected in a direct relationship between the weight of the tails of g and those of f as we deduce in section 5.2.

If the common distribution of P_n has finite second moment, then we obtain the invariant variance of Q as

$$\mathbb{V}\mathrm{ar}_g[Q] = \lambda\frac{a^2HK}{2(H+K)}\left(\mathbb{V}\mathrm{ar}[P] + \mathbb{E}[P]^2\right). \tag{5-8}$$

The implication of (5-8) together with the corresponding equation for $\mathbb{E}_g Q$ in (5-1), is that the mean and variance of Q are proportional to the corresponding statistics of the time averaged rainfall $\frac{1}{t}\int_0^t p(s)\,\mathrm{d}s$ over any period of time t. Although not a very surprising conclusion from a linear model, the fact that the proportionality coefficients are explicitly given in terms of the physical properties of the river network, might prove a useful first order approximation.

Finally, denote by $\mathbb{CV}_g[Q]$ the coefficient of variation of Q with respect to the invariant distribution. The following expression relates the coefficients of variation of P and Q,

$$\mathbb{CV}_g[Q] = \sqrt{\frac{\eta}{2(\beta+1)}(1+\mathbb{CV}[P]^2)}, \tag{5-9}$$

and generalizes the result of Botter (2010) to the case of non-zero β (except that their equation (10) seems to have a misplaced factor of two). Note that the equation for the second moment of $\hat{Q}$ is equivalent to $\mathbb{CV}_g[Q]^2 = \mathbb{V}\mathrm{ar}_\gamma[\hat{Q}]$.

Table 3 summarizes the results for the distributions of interest. For exponentially distributed precipitation amounts, one has $\mathbb{CV}[P] = 1$ and $\mathbb{CV}_g[Q] = \sqrt{\frac{\eta}{\beta+1}}$ which equals $\sqrt{\eta}$ in the limit $\beta \downarrow 0$, as reported in Botter et al. (2013).

Table 3. Expressions for the the coefficient of variation of Q with respect to the invariant distribution.

Distribution of P	$\mathbb{CV}_g[Q]^2$
Pareto (k,θ), $\theta > 2$	$\dfrac{\eta k(\theta-1)^2}{2\theta(\theta-2)(1+\beta)}$
Gamma (ω,ρ)	$\dfrac{\eta\rho(1+\omega)}{2\omega(1+\beta)}$
Inverse Gaussian (μ,σ)	$\dfrac{\eta(\mu+\sigma)}{2\sigma(1+\beta)}$

In Figure 7 we show qualitative differences in the invariant distribution of discharge across different cases that share the same η, $\mathbb{CV}[P]$ and $\mathbb{CV}_g[Q]$. Note from panel (a) on the right, that for small values of η and $\mathbb{CV}[P]$, and hence of $\mathbb{CV}_g[Q]$, there is essentially no qualitative difference in γ for the three distributions of P considered here. This similarity disappears specially as $\mathbb{CV}[P]$ increases and even if $\mathbb{CV}_g[Q]$ is held constant with the Pareto distribution exhibiting the most distinctive behavior (see panel d). Namely, the parametrization of the distribution of P reflects significantly on the distribution of Q. Note for each plot on the right, the three curves for γ share the same values for $\eta, \beta, \mathbb{CV}[P]$ and $\mathbb{CV}_g[Q]$.

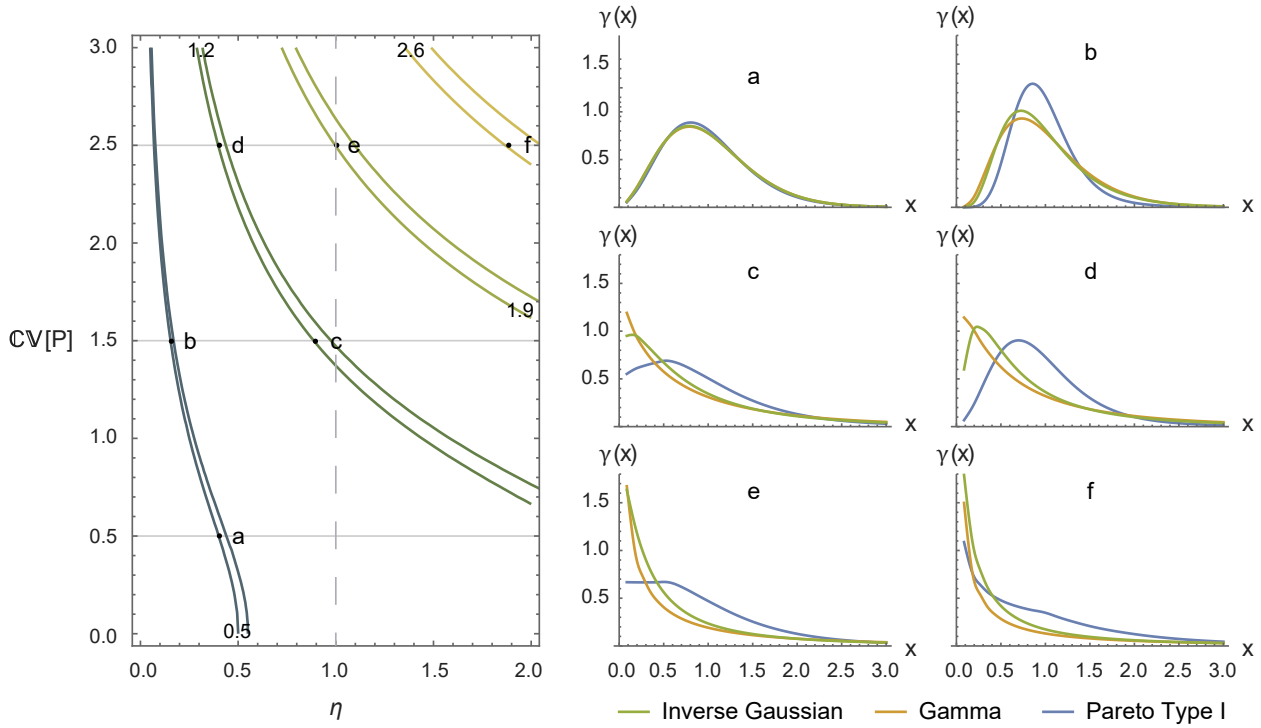

Figure 7. Left panel: contour plot of $\mathbb{CV}_g[Q]$ as a function of η and $\mathbb{CV}[P]$ given by equation (5-9). Each contour plot comes in pairs: one is drawn with $\beta = 0.001$ and the other with $\beta = 0.1$ as to showcase the limited effect of this variable. For each pair $(\eta, \mathbb{CV}[P])$ labeled with a letter on the left figure, the plots on the right show the corresponding invariant distribution γ of the normalized discharge. Every set of three curves is computed with the same values of η, $\beta = 0.01$ and ϕ chosen from the members of each distribution family such that they share the same value of $\mathbb{CV}[P]$. The resulting triplets of γ also have the same coefficient of variation.

5.2. Asymptotics of extreme events

We now characterize the asymptotic behavior of the probabilities of extreme events of discharge, namely, the behaviour of the right tail of the invariant distribution. We are interested in the asymptotic behavior of $\int_x^\infty g(y)\,\mathrm{d}y = \mathbb{P}_g(Q > x)$ as $x \to +\infty$ and how its decay depends on the geomorphological properties of the watershed and the probability density f of rainfall. Our analysis focuses on the Pareto and Gamma families of distributions as they exhibit contrasting power-like and exponentially decaying tails. See Table 1.

First we introduce some notation and language. Suppose $f(x)$ and $g(x)$ are positive functions that decay to zero as $x \to \infty$. We write $f(x) \sim g(x)$ if $\lim_{x\to+\infty} \frac{f(x)}{g(x)} = 1$. If the limit equals zero, we write $f(x) = o(g(x))$, which essentially means that $f(x) \ll g(x)$ as $x \to \infty$. We say that the probability density of a random variable X has a *heavy right tail* if $\mathbb{P}(X > x)$ decays to

zero slower than any exponential, namely

$$\lim_{x\to\infty} e^{\xi x}\mathbb{P}(X > x) = \lim_{x\to\infty} \frac{1 - F(x)}{e^{-\xi x}} = +\infty \quad \text{for all } \xi > 0,$$

where $F(x)$ is the distribution function of X. On the other hand, if

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(X > x) = -\xi,$$

for some $\xi > 0$, we say that the distribution of X has *light tails* of rate ξ and write $\mathbb{P}(X > x) \simeq e^{-\xi x}$.

For example, the Pareto distribution is always heavy-tailed since $1 - F(x) \sim \kappa^\theta x^{-\theta}$. See Table 1. Conversely, the Inverse Gaussian and the Gamma distributions cannot be classified as heavy tailed irrespective of their scale parameter. We have, in fact,

$$\mathbb{P}(X > x) \simeq e^{-x/\rho}, \qquad X \stackrel{d}{=} \text{Gamma}(\omega, \rho), \tag{5-10}$$

$$\mathbb{P}(X > x) \simeq e^{-2x\mu^2/\sigma}, \quad X \stackrel{d}{=} \text{Inverse Gaussian}(\mu, \sigma). \tag{5-11}$$

Note that the asymptotic relations $\simeq$ and $\sim$ are not equivalent, with the Gamma(ω, ρ) distribution providing an important example: (5-10) holds but the precise asymptotic equality is $\mathbb{P}(X > x) \sim (x/\rho)^{\omega-1} e^{-x/\rho}$.

The decay properties of the tail of a distribution can be inferred from studying the properties of the Laplace transform of its density as $s \to 0$. We now apply this theory to the distribution of discharge.

If $f : \mathbb{R} \to \mathbb{R}$ is locally integrable and vanishes on $(-\infty, 0)$, its Laplace-Stieltjes transform is

$$\hat{f}(s) = s \int_0^{+\infty} e^{-sx} f(s)\, \mathrm{d}x,$$

for all s for which the integral converges absolutely. If f has a derivative f', then the Laplace-Stieltjes transform of f is the Laplace transform of f'

$$\hat{f}(s) = \tilde{f}'(s).$$

Theorem 5.2 (Bingham et al. (1989, Theorem 1.7.1.)). **Karamata Tauberian Theorem**. Let U be a non-decreasing right-continuous function on $\mathbb{R}$ with $U(x) = 0$ for all $x < 0$. If $c \geq 0$, $\rho \geq 0$, the following are equivalent

$$U(x) \sim \frac{cx^\rho}{\Gamma(1+\rho)} \quad \text{as } x \to +\infty,$$
$$\hat{U}(s) \sim cs^{-\rho} \quad \text{as } s \to 0^+.$$

When $c = 0$, the results are to be interpreted as $U(x) = o(x^\rho)$ and $\hat{U}(s) = o(s^{-\rho})$, respectively.

Theorem 5.3 (Bingham et al. (1989, Theorem 1.7.2.)). **Monotone Density Theorem**. Let $U(x) = \int_0^x u(y)\,\mathrm{d}y$. If $U(x) \sim cx^\rho$ as $x \to +\infty$, where $c, \rho \in \mathbb{R}$, and if u is ultimately monotone, then

$$u(x) \sim c\rho x^{\rho-1} \quad \text{as } x \to +\infty.$$

When $\rho = 0$, the result is to be interpreted as $u(x) = o(x^{-1})$.

For the Pareto distributed rainfall, the result follows from an application of the Karamata tauberian theorem and the monotone density theorem. Namely the distribution of Q is heavy tailed for all values of the parameter θ. For $0 < \theta < 1$, equation (5-12) gives the precise asymptotic tail behavior. If, $\theta > 1$, P will have exactly $\lfloor\theta\rfloor$ moments, and by the analysis in section 5.1, so will Q. Equation (5-13) then follows from general theory (see for example Chung, 2001, Excercise 3.2.5).

Theorem 5.4. Suppose $P \sim \text{Pareto}(\kappa, \theta)$, for some $\kappa > 0, \theta > 0$. Then

$$\text{If } 0 < \theta < 1 : \mathbb{P}_g(Q > x) \sim \frac{\beta\,\Gamma(\theta)\Gamma\left(\frac{\theta\beta}{1-\beta}\right)}{(1-\beta)^\theta\,\Gamma\left(\frac{\theta}{1-\beta}\right)}\lambda(\kappa a)^\theta H^{\theta-1}x^{-\theta} =: cx^{-\theta} \quad \text{as } x \to +\infty, \tag{5-12}$$

$$\text{If } \theta \geq 1 : \qquad \mathbb{P}_g(Q > x) = o(x^{-1}) \qquad \text{as } x \to +\infty. \tag{5-13}$$

Proof. Denote $\Phi(x) := \mathbb{P}(Q > x) = 1 - \int_0^x g(y)\,\mathrm{d}y$ and $\Psi(x) := \int_0^x \Phi(y)\,\mathrm{d}y$. Then

$$\hat{\Psi}(s) = \tilde{\Psi}'(s) = \tilde{\Phi}(s) = \frac{1}{s}(1 - \tilde{g}(s)). \tag{5-14}$$

In this case, $f(x) \sim x^{-1-\theta}$ as $x \to \infty$ and $\tilde{f}(s) = \theta E_{1+\theta}(\kappa s)$. Power series expansion of $\tilde{f}(s)$ around $s = 0$, yields

$$\tilde{f}(s) = -\kappa^\theta s^\theta \Gamma(1-\theta) + 1 - \frac{\kappa\theta s}{\theta - 1} + \mathcal{O}(s^2).$$

If $0 < \theta < 1$, $\tilde{f}(s) = -\kappa^\theta s^\theta \Gamma(1-\theta) + 1 + \mathcal{O}(s)$, and equation (3-6) becomes

$$\tilde{g}(s) = \exp\left\{\frac{-\lambda}{H}(Ha\kappa)^\theta \Gamma(1-\theta)s^\theta \int_0^1 \frac{(u - u^{1/\beta})^\theta}{u(1-\beta)^\theta}\,\mathrm{d}u + \mathcal{O}(s)\right\}.$$

By direct integration, $\tilde{g}(s) \sim \exp(-cs^\theta)$ as $s \to 0^+$, with

$$c = \frac{\lambda(Ha\kappa)^\theta}{K(1-\beta)^\theta}\frac{\Gamma(1-\theta)\Gamma(\theta)\Gamma\left(\frac{\theta\beta}{1-\beta}\right)}{\Gamma\left(\frac{\theta}{1-\beta}\right)}. \tag{5-15}$$

Substitution of this results in equation (5-14), along with the series expansion of the exponential function yields $\hat{\Psi}(s) \sim cs^{\theta-1}$ as $s \to 0^+$. By theorems 5.2 and 5.3, one gets

$$\Psi(x) \sim \frac{c}{\Gamma(2-\theta)}x^{1-\theta} \quad \text{as } x \to +\infty,$$

$$\Phi(s) \sim \frac{(1-\theta)c}{\Gamma(2-\theta)}x^{-\theta} \quad \text{as } x \to +\infty.$$

Therefore

$$\mathbb{P}_g(Q > x) \sim \frac{\beta\,\Gamma(\theta)\Gamma\left(\frac{\theta\beta}{1-\beta}\right)}{(1-\beta)^\theta\,\Gamma\left(\frac{\theta}{1-\beta}\right)}\lambda(\kappa a)^\theta H^{\theta-1}x^{-\theta} \quad \text{as } x \to +\infty.$$

If $\theta > 2$, $\tilde{f}(s) = 1 - \frac{\kappa\theta s}{\theta-1} + \mathcal{O}(s^2)$, and equation (3-6) becomes

$$\tilde{g}(s) = \exp\left\{\frac{-\lambda a\kappa\theta}{\theta-1}s\int_0^1 \frac{(u-u^{1/\beta})^\theta}{u(1-\beta)^\theta}\,\mathrm{d}u + \mathcal{O}(s^2)\right\}.$$

By direct integration, $\tilde{g}(s) \sim \exp(-cs)$ as $s \to 0^+$, with

$$c = \frac{\lambda a\kappa\theta\beta}{(\theta-1)(1-\beta)^\theta}\frac{\Gamma(\theta)\Gamma\left(\frac{\theta\beta}{1-\beta}\right)}{\Gamma\left(\frac{\theta}{1-\beta}\right)}. \tag{5-16}$$

Substitution of this results in equation (5-14), along with the series expansion of the exponential function yields $\hat{\Psi}(s) \sim c$ as $s \to 0^+$. By theorems 5.2 and 5.3, one gets

$$\begin{aligned}\Psi(x) &\sim c && \text{as } x \to +\infty,\\ \Phi(s) &= o(x^{-1}) && \text{as } x \to +\infty.\end{aligned}$$

Therefore

$$\mathbb{P}_g(Q > x) = o(x^{-1}) \quad \text{as } x \to +\infty$$

□

The result (5-18) for the Gamma distributed rainfall requires a subtler approach for which we follow Nakagawa (2005). Let f be a function with Laplace transform $\tilde{f}(s) = \int_0^\infty f(x)e^{-sx}\,\mathrm{d}x$ with $s = \sigma + i\tau$. We say that σ_0 is the abscissa of convergence of $\tilde{f}$ if the integral converges for $\sigma > \sigma_0$ and diverges for $\sigma < \sigma_0$. For example, the abscissa of convergence for the probability density function of a Gamma(ω, ρ) distribution is $\sigma_0 = -1/\rho$.

Theorem 5.5 (Nakagawa (2005, Theorem 1))**.** For a non-negative random variable X with density function $f(x)$, let $-\xi$ be the abscissa of convergence of $\tilde{f}(s)$. If $\xi > 0$ and $s = -\xi$ is a pole for $\tilde{f}$, then:

$$\lim_{x\to+\infty}\frac{1}{x}\log\mathbb{P}(X > x) = -\xi. \tag{5-17}$$

In the notation introduced in this chapter, equation (5-17) is written $\mathbb{P}(X > x) \simeq e^{-\xi}$. Note that applying L'Hospital rule twice on the limit (5-17) yields $f'(x)/f(x) = -\xi + o(1)$, which implies that $f(x) = c_0 e^{-\xi x + o(1/x)}$ as $x \to \infty$ for some positive constant c_0.

The asymptotic behavior of $P_g(Q > x)$, in the case P is distributed as Gamma(ω, ρ), follows from computing the abscissa of convergence $s = -\xi$, proving it is a pole for $\tilde{g}$ and performing the integral in (3-6). Then, the theorem by Nakagawa (2005) yields the following exponential decay for the tails of Q.

Theorem 5.6. Suppose $P \stackrel{d}{=} \text{Gamma}(\omega, \rho)$ with $\omega > 0$, $\rho > 0$, then

$$\mathbb{P}_g(Q > x) \simeq \exp\left(\frac{-(1-\beta)}{Ha\rho\left(\beta^{\frac{\beta}{1-\beta}} - \beta^{\frac{1}{1-\beta}}\right)}x\right) =: \exp(-\xi x). \tag{5-18}$$

Proof. First, we find the pole of $\tilde{f}$ and show it is its abscissa of convergence.

$$\tilde{f}(s) = \int_0^{+\infty} f(x)e^{-sx}\,\mathrm{d}x = \frac{1}{\rho^\omega\Gamma(\omega)}\int_0^{+\infty} x^{\omega-1}e^{-x(s+1/\rho)}\,\mathrm{d}x = \frac{1}{(1+\rho s)^\omega}.$$

Consider the function $M(u) = Ham(u) = \frac{Ha}{1-\beta}(u - u^{1/\beta})$ and recall that $\beta = \frac{H}{K} < 1$. For $0 < u < 1$, $u > u^{1/\beta}$ and $M(u)$ is a non-negative function with a local maximum at $u^* = \beta^{\beta/(1-\beta)}$, $M(u^*) = \frac{Ha}{1-\beta}(\beta^{\beta/(1-\beta)} - \beta^{1/(1-\beta)})$.

Now, we proceed to find a pole for $\tilde{g}$. The convergence of $\int_0^1 \frac{1-\tilde{f}(sM(u))}{u}\,\mathrm{d}u$ for $s > 0$ is guaranteed by Zato and Yamazato's theorem. For $s \leq 0$, we have:

$$\begin{aligned}
\int_0^1 \frac{1-\tilde{f}(sM(u))}{u}\,\mathrm{d}u &= \int_0^1 \frac{(1+sM(u)\rho)^\omega - 1}{u(1+sM(u)\rho)^\omega}\,\mathrm{d}u \\
&\geq \frac{1}{(1+sM(u^*)\rho)^\omega}\int_0^1 \frac{\left(1+\frac{Ha\rho s}{1-\beta}(u-u^{1/\beta})\right)^\omega - 1}{u}\,\mathrm{d}u \\
&\geq \frac{(Ha\rho s)^\omega}{(1-\beta)^\omega(1+sM(u^*)\rho)^\omega}\int_0^1 \frac{(u-u^{1/\beta})^\omega - 1}{u}\,\mathrm{d}u \\
&\geq \frac{(Ha\rho s)^\omega}{(1-\beta)^\omega(1+sM(u^*)\rho)^\omega}\int_0^1 \left((u-u^{1/\beta})^\omega - 1\right)\,\mathrm{d}u \\
&= \frac{(Ha\rho s)^\omega}{(1-\beta)^\omega(1+sM(u^*)\rho)^\omega}\left[-1 + \frac{\beta\Gamma(1+\omega)\Gamma\left(\frac{\beta(1+\omega)}{1-\beta}\right)}{(1-\beta)\Gamma\left(\frac{1+\omega}{1-\beta}\right)}\right].
\end{aligned}$$

Then $\tilde{g}(s) \to 0^+$ when $s \downarrow -1/\rho M(u^*)$. But, by Jensen's inequality we know that

$$\tilde{g}(s) = \mathbb{E}e^{-sQ} \geq e^{-s\mathbb{E}Q},$$

therefore, $\tilde{g}(s)$ cannot tend to zero, the integral can not converge and $s = -\frac{1}{\rho M(u^*)}$ is a pole of $\tilde{g}$. Theorem 5.5 yields:

$$\begin{aligned}
\lim_{x\to+\infty}\frac{1}{x}\log\mathbb{P}_g(Q > x) &= -\frac{1-\beta}{Ha\rho(\beta^{\beta/(1-\beta)} - \beta^{1/(1-\beta)})} \\
\mathbb{P}_g(Q > x) &\sim \exp\left[-\frac{(1-\beta)x}{Ha\rho(\beta^{\beta/(1-\beta)} - \beta^{1/(1-\beta)})}\right] \quad \text{as } x \to +\infty.
\end{aligned}$$

□

The asymptotic results (5-12) and (5-18) can be illustrated using, respectively, log-log and log plots of g. Differentiating (5-12) with respect to x gives $g(x) \sim c\theta x^{-\theta-1}$, hence for large x on a log-log plot, the graph of $g(x)$ is asymptotically equal to that of $c\theta x^{-\theta-1}$. The asymptotic result (5-18) implies that $g(x) = c_0 e^{-\xi x + o(1/x)}$ for some constant c_0 (see Appendix) and therefore, for large x, the graph of $g(x)$ in a log plot must have slope $-\xi$. These comparisons are shown in Figure 8, Note that according to table 1, the parameter θ for Pareto distributed tropical rainfall is usually greater than 1, however we plot the case of $\theta < 1$ to illustrate a more interesting mathematical result.

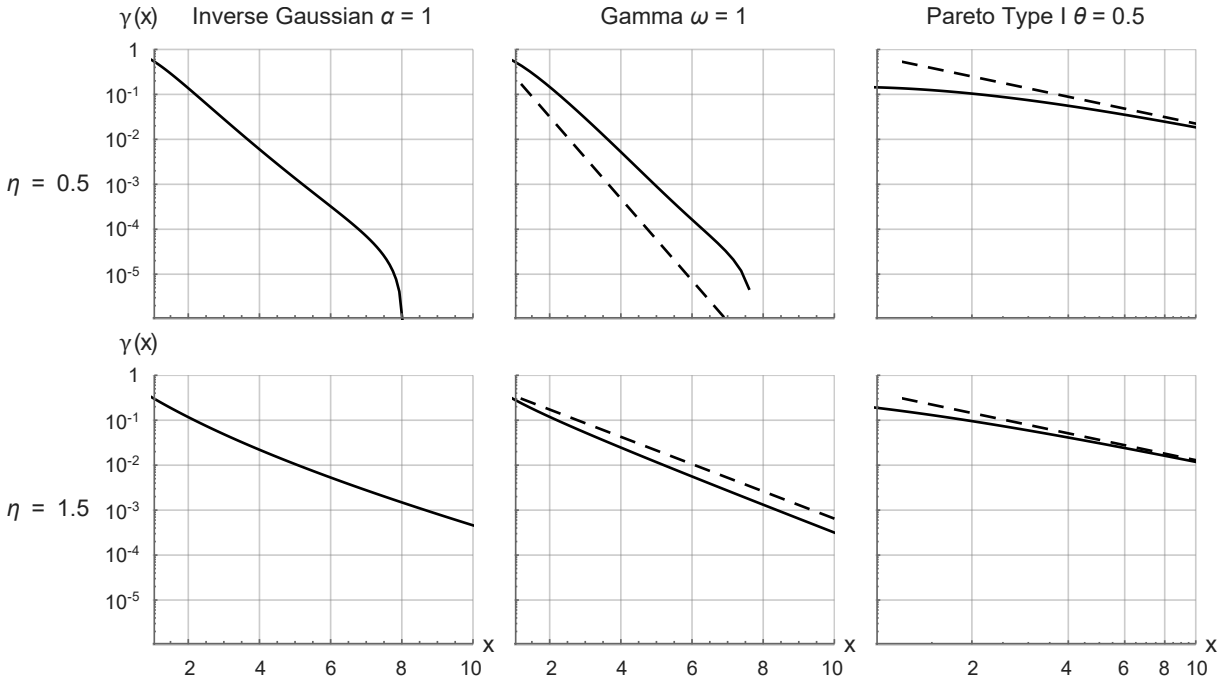

Figure 8. Plots of the non-dimensional invariant distribution $\gamma(x)$ for large values of x and different scenarios. For the case of $\hat{P}$ distributed as an Inverse Gaussian or Gamma distribution, we show a logarithmic plot; for Pareto distributed $\hat{P}$, we show a double logarithmic plot. We used $\beta = 0.01$ for all plots, subcritical $\eta = 0.5$ for plots in the upper row, and supercritical $\eta = 1.5$ for the lower row. The straight dashed line in the middle column has a slope $-\xi$ given in equation (5-18). For the Pareto case, the straight dashed lines have the form $cx^{-\theta}$ in equation (5-12). For the case $\eta = 0.5$ and $\hat{P} \stackrel{d}{=}$ Gamma(1), the numerical Laplace inversion algorithm fails to adequately describe γ before the asymptotic behavior can be observed.

5.3. Scaling of discharge

The traditional framework for studying the spatial variability of the dynamical processes that take place in a watershed is based on the concept of scaling. The formal definition of scaling in hy-

drology, which we now present, was stated by Gupta and Waymire (1990).

Let $Y(\boldsymbol{x})$ be an arbitrary random field indexed by $\boldsymbol{x} \in D \subseteq \mathbb{R}^d$. The parameter $\boldsymbol{x}$ can represent time, or multidimensional space, or space-time, and D is an arbitrary region of the d-dimensional Euclidean space $\mathbb{R}^d$. For $\lambda > 0$, let $Y_\lambda(\boldsymbol{x})$ denote the rescaled random process defined by

$$Y_\lambda(\boldsymbol{x}) = Y(\lambda \boldsymbol{x}) \quad \boldsymbol{x} \in \mathbb{R}^d.$$

We will call $Y(\boldsymbol{x})$ simple scaling if for each λ there is a scale function C_λ, of the form λ^θ, such that

$$Y_\lambda(\boldsymbol{x}) \stackrel{d}{=} \lambda^\theta Y(\boldsymbol{x}). \tag{5-19}$$

The parameter θ can be either positive or negative and is referred to as the scaling exponent. From now on, we write

$$Y_\lambda \stackrel{d}{=} \lambda^\theta Y_1.$$

There is an important consequence of simple scaling of $Y(\boldsymbol{x})$ given in (5-19). If Y_λ has finite moments $\mathbb{E}[Y_\lambda^h]$ of order h, then the strict sense simple scaling in (5-19) implies equality of moments

$$\mathbb{E}[Y_\lambda^h] = \lambda^{h\theta}\mathbb{E}[Y_1^h], \quad h = 1, 2, 3, \tag{5-20}$$

or equivalently,

$$\log \mathbb{E}[Y_\lambda^h] = h\theta \log \lambda + \log \mathbb{E}[Y_1^h]. \tag{5-21}$$

The property (5-21), being a moment rather than a distributional property, is referred to as *wide sense simple scaling* respect to λ and for positive processes, is reflected in two basic forms:

$$\text{log-log linearity in } \log \mathbb{E}[Y_\lambda^h] \text{ versus } \log \lambda \text{ for each } h, \tag{5-22}$$

$$\text{linearity of slope change: } h \to h\theta. \tag{5-23}$$

Wide sense simple scaling is much weaker than strict sense simple scaling. For example, if $\mathbb{E}[Y_\lambda^h] = \infty$ for some h, then wide sense simple scaling is meaningless, but strict sense simple scaling can still hold.

If a random variable exhibits wide sense simple scaling respect to λ and we normalize or rescale it by its mean, which effectively nondimensionalizes it, the moments of the resulting random variable are independent of λ (Peckham and Gupta, 1999).

$$\mathbb{E}[\hat{Y}_\lambda^h] = \mathbb{E}\left[\left(\frac{Y_\lambda}{\mathbb{E}[Y_\lambda^1]}\right)^h\right] = \frac{\mathbb{E}[Y_\lambda^h]}{(\mathbb{E}[Y_\lambda^1])^h} = \frac{\lambda^{h\theta}\mathbb{E}[Y_1^h]}{(\lambda^\theta \mathbb{E}[Y_1^1])^h} = \mathbb{E}\left[\left(\frac{Y_1}{\mathbb{E}[Y_1^1]}\right)^h\right] = \mathbb{E}[\hat{Y}_1^h]. \tag{5-24}$$

On the other hand, we define Y_λ to be *wide sense multiscaling* respect to λ if its statistical moments obey (5-22) but the slopes $s(h)$ as a function of h do not obey (5-23), i.e., are nonlinear in h. This can be expressed as

$$\log \mathbb{E}[Y_\lambda^h] = s(h) \log \lambda + \log \mathbb{E}[Y_1^h]. \tag{5-25}$$

We now consider spatially averaged rainfall rates, residence times and river flows to relate the wide sense scaling properties of these variables under the assumptions of our model. Our result is stated in the following theorem and illustrated in figure 9.

Theorem 5.7. Suppose that P is wide sense simple scaling with a scale parameter $\hat{a} = a/a_1$ and that K follows a power law with $\hat{a}$, i.e.

$$\mathbb{E}[P_a^h] = \hat{a}^{h\theta} \mathbb{E}[P_1^h], \tag{5-26}$$

$$K = k\hat{a}^\omega. \tag{5-27}$$

Let $\hat{a} \to \infty$, then the discharge is wide sense simple scaling with scale parameter $\hat{a}$ and scaling exponent

$$h(\theta + 1) \quad \text{if } \omega < 0, \tag{5-28}$$

$$h(\theta + \omega + 1) \quad \text{if } \omega > 0. \tag{5-29}$$

Proof. If K is related to $\hat{a}$ through a power law, so is H,

$$H = \beta k \hat{a}^\omega,$$

replacing this assumptions in (5-2), the moments of the discharge Q are given by

$$\mathbb{E}[Q_a^h] = (a_1 \lambda \hat{a}^{(\theta+1)} \mu)^h B_h \left(\left\{ \left(\frac{\beta k \hat{a}^\omega}{\lambda} \right)^{i-1} c_i(\beta)\, \mathbb{E}[\hat{P}_1^i] \right\}_{i=1}^{h} \right), \quad h = 1, 2, 3, \ldots \tag{5-30}$$

When $\hat{a} \to \infty$, the asymptotic behavior of $\mathbb{E}[Q_a^h]$ depends on the powers of $\hat{a}$

$$h(\theta + 1) + \omega(i - 1), \quad i = 1, 2, 3, \ldots, h.$$

and the dominating power $m(h)$ in (5-30) is

$$m(h) = \begin{cases} h(\theta + 1) & \text{if } \omega < 0 \\ h(\theta + \omega + 1) - \omega & \text{if } \omega > 0 \end{cases}.$$

Therefore, when $\hat{a} \to \infty$ the discharge is wide sense simple scaling

$$\mathbb{E}[Q_a^h] \sim \begin{cases} \hat{a}^{h(\theta+1)} (a_1 \lambda \mu)^h & \text{if } \omega < 0 \\ \hat{a}^{h(\theta+\omega+1)} \dfrac{(a_1 \lambda \mu)^h}{\hat{a}^\omega} \left(\dfrac{\beta b}{\lambda} \right)^{h-1} c_h(\beta) \mathbb{E}[\hat{P}_1^h] & \text{if } \omega > 0 \end{cases} \quad \text{as } \hat{a} \to \infty. \tag{5-31}$$

□

Gupta and Waymire (1990) observed a multiscaling structure in the instantaneous rainfall data measured during the Global Atmosphere Research Program, Atlantic Tropical Experiment (GATE). Here, we relaxed it to the wide sense simple scaling hypothesis of P for academic purposes. On the other hand, Leopold and Maddock (1953) proposed power laws of the hydraulic geometry, one of which represents the mean velocity in the channel as a function of the mean discharge at the outlet of the watershed. This, together with the well know power law between the mean discharge and the area of a river basin, supports the viability of the hypothesis for K. We proceed to validate the conclusion of theorem 5.7 against our data.

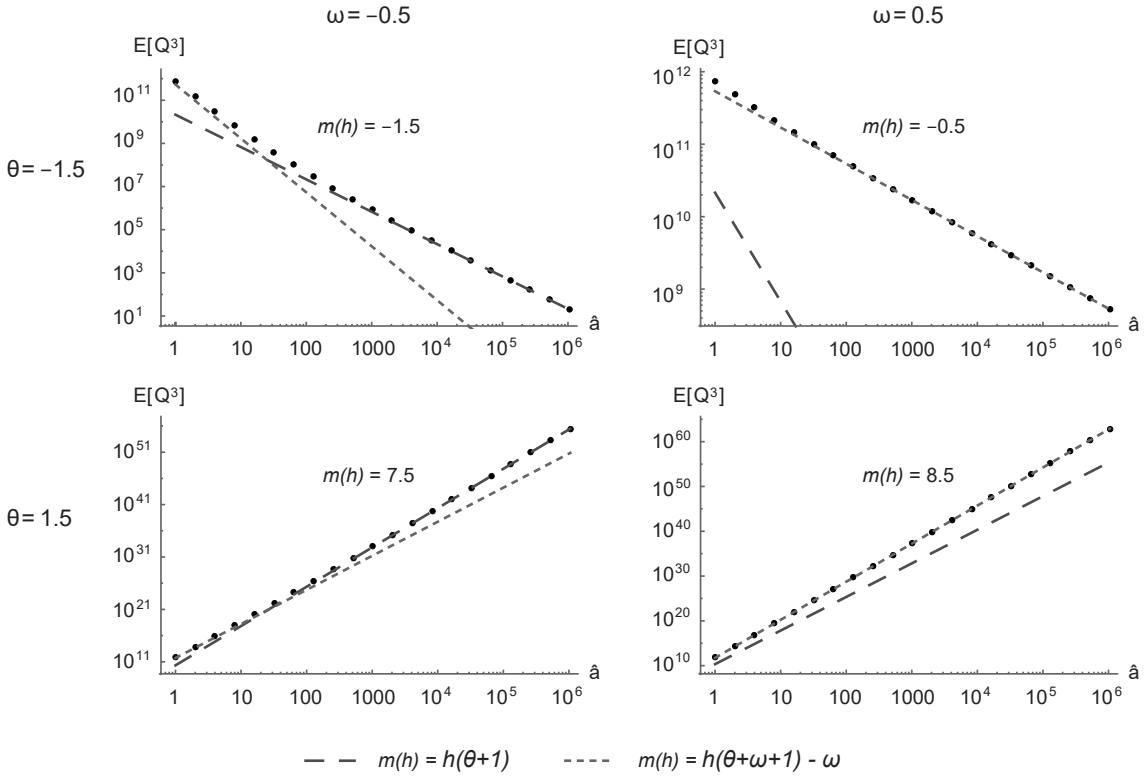

Figure 9. Log-Log plots of $\mathbb{E}[Q_a^3]$ as a function of $\hat{a}$. We used an inverse Gaussian distribution for $\hat{P}$ and the parameters of the validation window JJA 1968 in table 2. The dots were calculated with equation (5-30), the dashed and doted lines were calculated with equation (5-31). The scaling of the discharge is an asymptotic result. In the left panels, the dots approach the dashed line while in the right panels, the dots approach the doted line as $\hat{a} \to \infty$.

From figure 9 it is evident that the dots approach the same line (dashed or dotted) regardless the value of θ, which makes evident the relative importance of the hypothesis for the mean residence time in the channel. The exponent ω in the power law for K not only affects the magnitude of the scaling exponent, but also determines the scaling structure of discharge. Also, it is important to note that the scaling of the discharge is an asymptotic result and the data approaches the

expression in (5-31) as $\hat{a} \to \infty$. This is a remarkable difference with the hydrological literature, where the theoretical wide sense scaling is defined with an equality.

5.4. Return period

The return period is one of the fundamental concepts for the applications of hydrology in engineering, as it is used for hydrological design and risk analysis. In this section We propose a novel interpretation of the return period in the context of the time continuous stochastic process and we present a new expression to estimate it under the assumptions of our model.

5.4.1. Return period in hydrology

In the classical hydrology literature an extreme event occurs when a random variable X is greater or equal than a certain threshold u. The recurrence interval τ is the time between occurrences of the event $[X \geq u]$ and it is also a random variable. The return period $\mathcal{T}$ of the extreme event is the expected value of τ (Te Chow et al., 1962)

$$\mathcal{T} = \mathbb{E}[\tau]. \tag{5-32}$$

If the random variable X is measured every fixed time interval Δt, then we have a time series of observations with a uniform temporal resolution, i.e. daily, monthly, annually, etc. By observations we mean a sequence $\{X_1, X_2, X_3, \dots\}$, of realizations of a random variable X. If we also suppose that the observations are independent and identically distributed, the following heuristic allows to relate the return period to the probability $p = \mathbb{P}(X \geq u)$ of occurrence of the extreme event. For each observation, there are two possibilities: a "success" if $X \geq u$ with probability p, and a "failure" if $X \leq u$ with probability $(1-p)$. If we have $\tau - 1$ failures followed by a success, we have a recurrence interval of discrete duration τ, with probability $(1-p)^{\tau-1}p$. The expected value for τ is given by

$$\mathbb{E}[\tau] = \sum_{\tau=1}^{\infty} \tau(1-p)^{\tau-1}p = p\sum_{\tau=1}^{\infty} \tau(1-p)^{\tau-1} = p\frac{1}{p^2} = \frac{1}{p}.$$

Therefore, the probability of occurrence of an extreme event in any observation is equal to the inverse of its return period

$$\mathbb{P}(X \geq u) = \frac{1}{\mathcal{T}}. \tag{5-33}$$

For example, a 50-year flood has a $1/50 = 0.02$ or 2% chance of happening or being exceeded in any one year and a 100-year flood has a 0.01 or 1% chance of happening or being exceeded in any one year.

In engineering applications, the estimated return period is a statistic. For example, one cannot determine the magnitude of the extreme event with a return period of 100 years with a time series

of 30 years. Even if the return period is a lot less than the record length of the time series, a good estimation of an expected value requires a good sample of recurrence intervals. Typically, the hydrological records are not long enough. Instead, one traditionally fits a statistical model for the tail of the distribution of X and predicts the magnitude of such an unobserved extreme event under the assumptions of equation (5-33).

5.4.2. Marked Poisson Point Process

We now want to extend the concept of return period, which was defined for a sequence of independent realizations of a random variable, to the context of a time continuous stochastic process $\{X(t), t \geq 0\}$. The Marked Poisson Process (MPP) in the real line with intensity λ, describing the occurrence over time of positive random events, is a natural example to motivate the theoretical analysis we are about to present.

Consider a MPP which is a double sequence $(\{t_i\}_{i=1}^{\infty}, \{Y_i\}_{i=1}^{\infty})$ of $\mathbb{R}^+$-valued random variables t_i and Y_i, such that the inter-arrival times $\tau_i = t_{i+1} - t_i$, $i = 1, 2, 3, \ldots$, are exponential variables with mean $1/\lambda$ and the marks $\{Y_i\}$ are i.i.d. variables. See Appendix. One can interpret t_i, if finite, as the timepoint a which the i-th recording of an event with magnitude Y_i takes place.

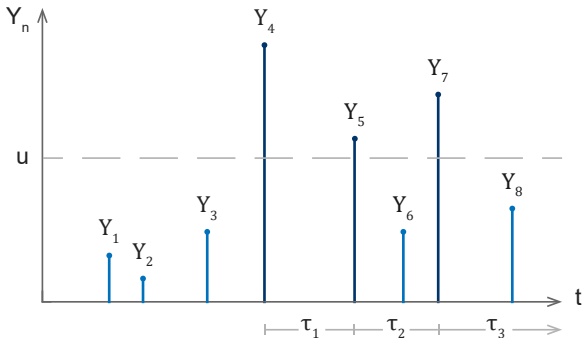

Figure 10. Schematic representation of a Marked Poisson Process. The dark marks are the ones whose magnitude exceeds the level u and therefore form the process $Y(u)$ and define the inter-arrival times τ_i.

Let us consider a MPP with intensity $\lambda = 1$, illustrated in figure 10. If we set an arbitrary level $u > 0$, it can be shown that the process $Y(u) = \{(t_i, Y_i) : Y_i \geq u\}$ is another MPP with intensity $\rho_u = \mathbb{P}(Y_i \geq u)$ (Jacobsen, 2006). Therefore, $N_u(t) = \#\{Y_i : Y_i \geq u\}$ is a Poisson process with intensity ρ_u and the interrarivals $\{\tau_1, \tau_2, \tau_3, \ldots\}$ are exponential variables with mean $1/\rho_u$, i.e.

$$\mathbb{E}[N_u(t)] = t\,\mathbb{P}(Y_i \geq u), \tag{5-34}$$

$$\mathbb{E}[\tau_i] = \frac{1}{\mathbb{P}(Y_i \geq u)}, \quad i = 1, 2, 3, \ldots \tag{5-35}$$

Note that the formulation (5-35) conveys the same meaning as (5-33), thus giving an interpretation of the concept of return period of events separated by continuous inter-arrival times, namely $\mathcal{T} = \mathbb{E}[\tau_i]$.

Equation (5-33), which we refer to as the *Binomial Heuristic*, is the traditional interpretation of the return period as the inverse of the probability of a Bernoulli trial in the binomial distribution. This approach is not adequate for hydrological modeling because each observation of a hydrological variable does not necessarily represent an independent Bernoulli trial. However, the results obtained for engineering applications will be similar to those obtained under the Poisson heuristic approximation, because when the number of observations of the random variable goes to infinity and the probability of occurrence of the extreme event goes to zero, namely $i \to \infty$ and $u \to \infty$ in (5-33), the binomial distribution is a good approximation of the Poisson distribution.

5.4.3. Poisson Clumping Heuristic

Distributional questions concerning rare events associated with random processes may be reformulated as questions about sparse random sets. The sparse random sets, in our case the observations of $[X \geq u]$, often resemble i.i.d random clumps thrown down randomly, i.e. centered at points of a Poisson process. The Poisson clumping heuristic, which we now present as stated by Aldous (1989), consists of approximating these sparse random sets by a mosaic process $\mathcal{S}$. A mosaic process $\mathcal{S}$ in $\mathbb{R}^+$ is the union of i.i.d. random intervals $\mathcal{I}_i = (t_i - \epsilon_i/2, t_i + \epsilon_i/2)$ with Poisson random centers t_i happening at a rate λ per unit length and i.i.d. $\epsilon_i > 0$. We call λ the clump rate.

For $x \in \mathbb{R}^+$, let N_x be the number of intervals of $\mathcal{S}$ which contain x. Write $I_i = |\mathcal{I}_i|$. Aldous (1989, Lemma A2.1) shows that N_x has a Poisson distribution with mean $\lambda\mathbb{E}[I_i]$. Then, for a large interval $[a, b] \in \mathbb{R}^+$, we can approximate $\mathcal{S} \cap [a, b]$ as the union of those intervals of $\mathcal{S}$ whose centers lie in $[a, b]$, and this approximation gives

$$\mathbb{P}(\mathcal{S} \cap [a, b] \text{ empty}) \approx \mathbb{P}(N_x = 0) = \exp(-\lambda\,(b - a)). \tag{5-36}$$

Now, let $\{X(t),\, t \geq 0\}$ be a continuous-time stationary random process, and suppose that we are interested in the distribution of $M_t = \sup_{s\in[0,t]} X(s)$ for large t. We can define the random set $\mathcal{S}_u = \{t : X(t) \geq u\}$, and we have

$$\mathbb{P}(M_t < u) = \mathbb{P}(\mathcal{S}_u \cap [0, t] \text{ empty}).$$

For u large, $\mathcal{S}_u$ is a sparse stationary random set (See figure 11). Suppose $\mathcal{S}_u$ can be approximated by a sparse mosaic with some clump rate λ_u. Then, by (5-36)

$$\mathbb{P}(\mathcal{S}_u \cap \text{ empty}) \approx \exp(-\lambda_u t).$$

Finally, define the first hitting time to u as $T_u = \min\{t : X(t) \geq u\}$. Studying approximations to the distribution of M_t for large t is equivalent to studying hitting times since $\mathbb{P}(M_t < u) =$

$\mathbb{P}(T_u > t)$. Therefore

$$\mathbb{P}(T_u > t) = \mathbb{P}(M_t < u) \approx \exp(-\lambda_u t), \tag{5-37}$$

and the Poisson clumping heuristic states that the first hitting time T_u can be approximated as an exponential random variable.

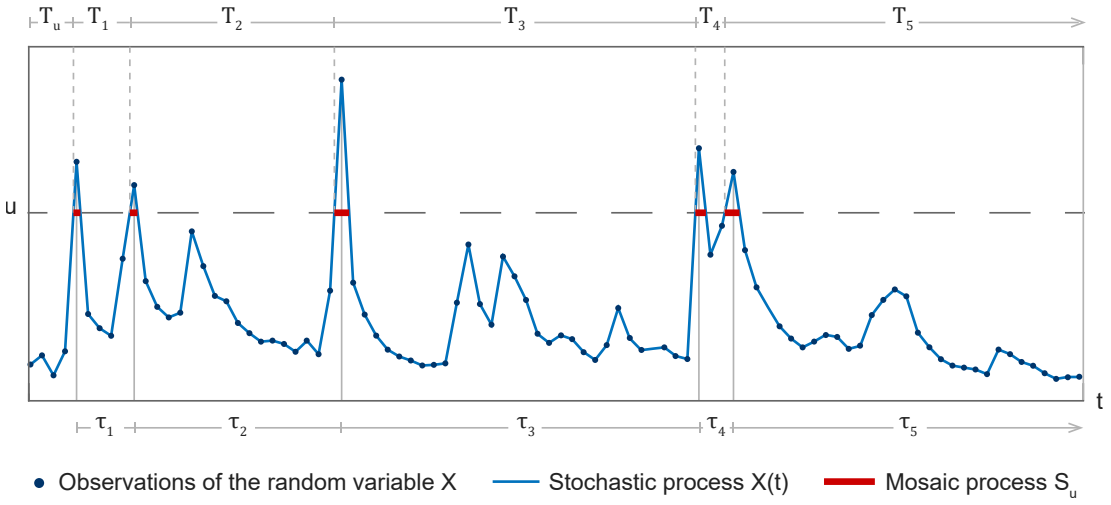

Figure 11. Illustration of the interpretation of the return period concept for a continuous time stochastic process $\{X(t), t \geq 0\}$, from which observations $X_i = X(i\Delta t)$, $i = 0, 1, 2, \ldots$ are made with a uniform time step Δt. The random variables τ_i are the inter-arrivals of the observations X_i exceeding u, while the random variables T_i are the inter-arrivals of the upcrossings of the stochastic process $\{X(t), t \geq 0\}$. Note that for u large, $\{t : X(t) \geq u\}$ is a sparse random set which we approximate by a mosaic process $\mathcal{S}_u$.

In practice, level crossing counting is often used as a means to describe the variability and extremal behavior of a continuous stochastic process. For example, the maximum of the process in an interval is equal to the lowest level above which there exists no genuine level crossing, provided, of course, that the process starts below that level. Since it is often easier to find the statistical properties of the number of level crossings than to find the maximum distribution, crossing methods are of practical importance (Lindgren, 2006).

For sample functions of a continuous process $\{X(t),\, t \in \mathbb{R}^+\}$ we say that $X(t)$ has an upcrossing of the level u at t_0 if, for some $\epsilon > 0$, $X(t) \leq u$ for all $t \in (t_0 - \epsilon, t_0]$ and $X(t) \geq u$ for all $t \in [t_0, t_0 + \epsilon)$. For an interval $[0, t]$, write $N_u^+(X(t))$ for the number of u-upcrossings by $X(s)$, with $s \in [0, t]$. By the *intensity of upcrossings* we mean any function $\rho_u^+(t)$ such that

$$\int_{[0,t]} \rho_u^+(s)\, \mathrm{d}s = \mathbb{E}[N_u^+(X(t))].$$

For a stationary process, $\rho_u^+(t) = \rho_u^+$ is independent of t and ρ_u^+ is the mean number of upcrossings

per time unit at level u.

$$\mathbb{E}[N_u^+(X(t))] = t\rho_u^+. \tag{5-38}$$

Note that (5-38) is the continuous version of (5-34), replacing the probability of exceeding u by the rate ρ_u^+.

If $\{X(t), t \geq 0\}$ has absolutely continuous sample paths, the heuristic (5-37) takes the form: for u large,

$$\mathbb{P}(T_u > t) = \mathbb{P}(M_t < u) \approx \exp(-t\rho_u^+), \tag{5-39}$$

which meas that the hitting time T_u is approximately exponential with mean $1/\rho_u^+$.

Now, define $\{T_i\}_{i=1}^{\infty}$ as the sequence of inter-arrival times between upcrossings. For a continuous-parameter Markov process, by the strong Markov property, the inter-arrivals have the same distribution of the first hitting time T_u, therefore

$$\mathbb{E}[T_i] \approx \frac{1}{\rho_u^+}.$$

Finally, if we interpret the random variable τ in (5-32) as a discrete approximation to the random sequence $\{T_i\}$ of inter-arrival times, under the framework of the Poisson clumping heuristic, we can estimate the return period of a continuous-time Markov process as

$$\mathcal{T} \approx \frac{1}{\rho_u^+}. \tag{5-40}$$

Given that for u large, equation (5-40) provides a new expression to estimate the return period of a hydrological extreme event, we now attempt an exact calculation of ρ_u^+ based on the invariant density of discharge (3-6).

There exists a very famous formula, found by Marc Kac and Steve O. Rice (Rice, 1944), for ρ_u^+, the expected number of upcrossings of a level u. The simplest version is valid for stationary processes $\{X(t), t \in \mathbb{R}\}$ with absolutely continuous sample paths and absolutely continuous distribution with density $f_{X(t)}(u) = f_{X(0)}(u)$, independent of t. For such a process, the derivative $\dot{X}(t)$ exists almost everywhere, and the conditional expectation $\mathbb{E}\big(\dot{X}(0)^+ \,\big|\, X(0) = u\big)$ exist, with $\dot{X}(0)^+ = \max\big(0, \dot{X}(0)\big)$. The upcrossings intensity is therefore given by

$$\begin{aligned} \rho_u^+ &= \mathbb{E}[N_u^+(X(t))] = \int_0^{\infty} z f_{(X,\dot{X})}(u,z)\,\mathrm{d}z \\ &= f_{X(0)}(u)\,\mathbb{E}\big(|\dot{X}(0)^+| \,\big|\, X(0) = u\big), \end{aligned} \tag{5-41}$$

where $f_{(X,\dot{X})}(u)$ is the join density of $X(t)$ and $\dot{X}(t)$. This expression holds for almost any u, whenever the involved densities exist.

This formula can be replaced in the context of the discharge to find an expression to estimate the return period $\mathcal{T}$ of the extreme event $[Q \geq u]$.

Theorem 5.8. If $\{Q(t),\, t \geq 0\}$ is a time continuous stationary process with density function characterized by (3-2), then the Laplace transform of the upcrossings intensity ρ_u^+ at level u is given by

$$\tilde{\rho}^+(s) = K \int_0^\infty \int_0^\infty \boldsymbol{F}(u,r) e^{-su}\, \mathrm{d}r\, \mathrm{d}u = K \tilde{\boldsymbol{F}}(s,0), \tag{5-42}$$

with

$$\tilde{\boldsymbol{F}}(s,s_2) = -\frac{\partial}{\partial s_2} \tilde{\boldsymbol{g}}(s, s+s_2).$$

Proof. Lets consider the stochastic process $Q(t)$, which under the assumptions of our model, is stationary, continuous, differentiable almost everywhere, and has a density g. According to (5-41), the upcrossings intensity can be estimated as

$$\begin{aligned}
\rho_u^+ &= g(u)\, \mathbb{E}\big[\dot{Q}^+(t) \,\big|\, Q(t) = u\big] && (5\text{-}43)\\
&= g(u) \int_0^\infty v\, \mathbb{P}\big(\dot{Q}^+(t) \in \mathrm{d}v \,\big|\, Q(t) = u\big) && (5\text{-}44)\\
&= g(u) \int_0^\infty v\, \mathbb{P}\big(K(-Q(t) + R(t)) \in \mathrm{d}v \,\big|\, Q(t) = u\big) && (5\text{-}45)\\
&= g(u) \int_0^\infty \frac{v\, \mathbb{P}(Q(t) \in \mathrm{d}u, K(-u + R(t)) \in \mathrm{d}v)}{\mathbb{P}(Q(t) \in \mathrm{d}u)} && (5\text{-}46)\\
&= g(u) \int_0^\infty v\, \boldsymbol{g}\left(u, \frac{v}{K} + u\right) \frac{\mathrm{d}v}{K}, && (5\text{-}47)
\end{aligned}$$

where $\boldsymbol{g}$ is the invariant distribution of the process $\boldsymbol{X}(t)$, i.e. the join invariant distribution of $R(t)$ and $Q(t)$.

Let $x = \frac{v}{K} + u$, then

$$\rho_u^+ = \mathbb{E}[N_u^+(Q(t))] = K \int_u^\infty (x-u)\, \boldsymbol{g}(u,x)\, \mathrm{d}x. \tag{5-48}$$

The Laplace transform of $\rho^+(u)$ is

$$\tilde{\rho}^+(s) = K \int_0^\infty \int_u^\infty (x-u)\, \boldsymbol{g}(u,x) e^{-su}\, \mathrm{d}x\, \mathrm{d}u.$$

We would like to relate it to the Laplace transform of g, for which there exists the analytical expression (3-6). Let $r = x - u$, then $\mathrm{d}r = \mathrm{d}x$ and

$$\tilde{\rho}^+(s) = K \int_0^\infty \int_0^\infty r\, \boldsymbol{g}(u, r+u) e^{-su}\, \mathrm{d}r\, \mathrm{d}u. \tag{5-49}$$

Let $\boldsymbol{h}(u,r) = \boldsymbol{g}(u, r+u)$, then

$$\tilde{\rho}^+(s) = K \int_0^\infty \int_0^\infty r\, \boldsymbol{h}(u,r) e^{-su}\, \mathrm{d}r\, \mathrm{d}u.$$

Finally, let $\boldsymbol{F}(u,r) = r\,\boldsymbol{h}(u,r)$, then

$$\tilde{\rho}^{+}(s) = K \int_0^\infty \int_0^\infty \boldsymbol{F}(u,r) e^{-su}\,\mathrm{d}r\,\mathrm{d}u = K\tilde{\boldsymbol{F}}(s,0),$$

with

$$\tilde{\boldsymbol{F}}(s,s_2) = -\frac{\partial}{\partial s_2}\tilde{\boldsymbol{h}}(s,s_2).$$

□

Even though expression (5-42) characterizes the upcrossings intensity, it cannot be explicitly written in terms of $\tilde{g}$, nor computationally implemented to estimate the return period, since we are interested in a rare event which is not possible to sample numerically.

6. Conclusions

We have derived and solved a mathematical model for the rainfall-runoff process in a catchment of order one. The model is linear, implying that its predictions should be understood as a first order approximation of the actual dynamics. Our main result is the characterization of the limiting invariant distribution of discharge in relation to the basin area, residence times on hillslopes and channel, and the statistical properties of the rainfall process. At a very fundamental level, our results give a mathematically minded glimpse of how the uncertainty in the rainfall process interacts with the geophysical variables within a catchment to produce uncertainty in the discharge.

The invariant distribution of discharge is characterized in equation (3-8), and depends on two a-dimensional parameters. The first one is $\eta = H/\lambda$, relating the time scale associated with the occurrence of rainfall events and that of hillslope runoff. The second one, $\beta = H/K$, is essentially the ratio between the residence times for channel and hillslope flow. We have confirmed and expanded the results obtained by Botter et al. (2007a) on the important role played by η on the invariant density g of discharge. The threshold $\eta = H/\lambda = 1$ largely determines the shape of the probability density of the invariant distribution of discharge: g is unimodal for $\eta < 1$, while monotone decreasing for $\eta > 1$. On the other hand, we show that the parameter β plays a very marginal role on determining g when considered over a large range of physically sensible values, $10^{-4} < \beta < 10^{-1}$ (see Figure 2). Even at sub-daily time scales, the slow component of the system dominates the dynamics and the transfer of uncertainty from precipitation to discharge. Namely, most of the properties of the system can be recovered in the limit $H/K \to 0$. We conclude that, at least in the case of an order-one catchment, the at-a-point dynamics modeled by a single reservoir as in Botter et al. (2007a) are a good approximation to the double-reservoir model presented here.

In the case study presented in chapter 4 we were careful enough to find a catchment subject to a statistically invariant rainfall dynamics of approximately instantaneous Poissonian events, where the hydrological budget evidences that evapotranspiration plays a minor role. From our results there, we can draw two conclusions. First, that whenever its hypothesis are sufficiently satisfied, our model can be parametrized for H and K so that the predicted invariant distribution matches the seasonal ensemble of discharge. Moreover, that the qualitative changes in the distribution of discharge predicted at the threshold $\eta = 1$ do occur in natural catchments. Secondly, and this is quite troubling, that while H is the most important parameter of the model, it seem to depend itself on the precipitation regime. While this is probably a manifestation of the true non-linear nature of the underlying phenomena, one should be careful on assigning H a value based simply

on geomorphological considerations.

The transfer of uncertainty between the precipitation process and the discharge is summarized by the mathematical results in chapter 5. First, we relate all moments of P and Q via equation (5-2) which yields the relation (5-9) between the corresponding coefficients of variations $\mathbb{CV}(P)$ and $\mathbb{CV}_g(Q)$. Figure 7 shows, however, that this flow of uncertainty is far from trivial: the distribution of discharge can have quite different behaviors even the values of $\mathbb{CV}(P)$, $\mathbb{CV}_g(Q)$ or η are held constant. Another important implication from equation (5-2) is that under the invariant distribution, Q has exactly as many moments as does the rainfall amounts P_n. This observation is magnified in section 5.2 where we establish that the invariant distribution of Q inherits the tail behavior of the distribution of P. For the cases of P distributed as Pareto and Gamma distributions, which represent heavy-tailed and light-tailed examples respectively, equations (5-12) and (5-18) characterize the weight and index of the tails of g in terms of all other parameters of the model. From both results we can conclude that an increase in variability of seasonal rainfall statistics, will likely produce an increase of uncertainty in discharge, and that this increase will depend heavily on the residence time H.

In section 5.3 we supposed a wide sense scaling structure of the spatially averaged rainfall rates and residence times, and analyze the scaling structure of discharge. It turned out to be another asymptotic result with the discharge being wide sense scaling and approaching the expression in (5-31) as the scale parameter goes to infinity. The scaling of K seems to be a fundamental hypotheses, even more important than the scaling of P, given the threshold $\omega = 0$ which largely determines the scaling structure of discharge. Finally, our approach to the hydrological concept of return period is purely mathematical as we wanted to provide a rigorous theoretical framework to extend this concept to the time-continuous stochastic process context. We were able to reinterpret the return period of a sequence of independent realizations of a random variable, as the inverse of the intensity of a Marked Poisson Process, but also as the inverse of the intensity of upcrossings of a stationary Markov process. Moreover, the comparison between the Poisson clumping heuristic and the binomial heuristic let us justify why the classical definition of return period in hydrology works well for engineering applications, even when based on questionable hypothesis of independence.

A. Appendix

The mathematical framework of this work is based on fundamental concepts of the probability theory, which we now present based on the classical literature by Bhattacharya and Waymire (2009), Durrett (1999) and Walsh (2012).

A.1. Random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **random variable** is a function $X : \Omega \to \mathbb{R}$ such that for all $B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{F}$. We call $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ an event and we denote its probability as $\mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$.

The **distribution function** of a random variable X is the function $F : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = \mathbb{P}(X \leq x).$$

We say that the random variables X and Y are **independent** if and only if for all Borel sets A and B,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \tag{A-1}$$

If $\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq y)$ for all $x, y \in \mathbb{R}$, we write $X \overset{d}{=} Y$.

If there exists a function $f_X(x)$ on $\mathbb{R}$ such that the distribution function $F_X(x)$ of X is

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,\mathrm{d}y,$$

then the function f_X is called the **density** of X and we write

$$\mathbb{P}(X \in \mathrm{d}x) = f_X(x)\,\mathrm{d}x.$$

Let $h(x)$ be a smooth, one-to-one function, i.e., strictly monotone and continuously differentiable. If $Y = h(X)$, then, the density of Y is

$$f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|}. \tag{A-2}$$

Let $g \geq 0$ be a real valued Borel measurable function and suppose X has density f_X. We define the **expected value** of the random variable $g(X)$ as

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x)\,\mathrm{d}x.$$

In particular, $\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)\,\mathrm{d}x$ and we say that the random variable X is integrable if $\mathbb{E}[|X|] < \infty$

The **variance** of X is

$$\mathbb{V}\mathrm{ar}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \tag{A-3}$$

and the **coefficient of variation** is

$$\mathbb{C}\mathbb{V}[X] = \frac{\sqrt{\mathbb{V}\mathrm{ar}[X]}}{\mathbb{E}[X]}. \tag{A-4}$$

The **Laplace transform** of the density f_X is defined as

$$\tilde{f}_X(s) = \mathbb{E}[e^{-sX}] = \int_{\mathbb{R}} f_X(x) e^{-sx}\,\mathrm{d}x, \tag{A-5}$$

and the **moment generating function** of X as

$$M_X(\alpha) = \mathbb{E}[e^{\alpha X}] = \int_{\mathbb{R}} f_X(x) e^{\alpha x}\,\mathrm{d}x, \tag{A-6}$$

If $M_X(\alpha)$ exists for α in a neighborhood of the origin, one can compute moments of any order by differentiating

$$\mathbb{E}[X^n] = \left.\frac{\mathrm{d}^n M_X(\alpha)}{\mathrm{d}\alpha^n}\right|_{\alpha=0}. \tag{A-7}$$

Note that the Laplace transform of f_X is closely related to the moment generating function of X since

$$\tilde{f}_X(-s) = M_X(s). \tag{A-8}$$

Let X be a random variable with probability density f and denote $\mu = \mathbb{E}[X]$. Define the normalized random variable $\hat{X} = X/\mu$ and denote by ϕ its probability density, then

$$\phi(x) = \mu f(\mu x) \tag{A-9}$$

$$\tilde{\phi}(s) = \tilde{f}\left(\frac{s}{\mu}\right). \tag{A-10}$$

A continuous random variable T is said to have an **exponential distribution** with mean $1/\lambda$ if

$$P(T \leq t) = 1 - e^{-\lambda t}, \quad t \geq 0$$

A discrete random variable X is said to have a **Poisson distribution** with intensity λ if

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

A.2. Continuous-parameter Markov processes

A **stochastic process** $X = \{X(i) : i \in I\}$ is a collection of random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, indexed by a variable i in a set I and taking values in a set E of the d-dimensional Euclidean space $\mathbb{R}^d$, $E \subseteq \mathbb{R}^d$. Typically, the indexing variable i is the time, $t \in \mathbb{R}^+ = [0, \infty)$, and we write $X = \{X(t), t \geq 0\}$.

A stochastic process $\{X(t), t \geq 0\}$ has independent increments if for each n and each $0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \cdots \leq s_n < t_n$, the random variables $X(t_i) - X(s_i)\, i = 1, 2, \ldots, n$ are independent. The increments are stationary if for each t, the distribution of $X(s+t) - X(s)$ is independent of s.

Let $\{X(t), t \geq 0\}$ be a time-continuous stochastic process. We will write $\mathbb{P}_x$ to denote the probabilities of $X(t)$ conditioned on $X(0) = x$, i.e. for an event B in the σ-field of Borel subsets of E, $B \in \mathcal{B}(E)$,

$$\mathbb{P}_x(X(t) \in B) = \mathbb{P}\big(X(t) \in B \,\big|\, X(0) = x\big).$$

We say that $X(t)$ is a continuous-parameter **Markov process** if it has the Markov property, i.e. if the conditional distribution of the future given the present and the past only depends on the presents. This is mathematically stated as: for $s \geq 0$, $t > s$ and $B \in \mathcal{B}(E)$

$$\mathbb{P}\big(X(t) \in B \,\big|\, \mathcal{F}(s)\big) = \mathbb{P}_{X(s)}(X(t) \in B)$$

where $\mathcal{F}(s)$ is the σ-algebra generated by $\{X(u) : 0 \leq u \leq s\}$, i.e. the σ-algebra with all the information of the stochastic process until time s.

A stopping time τ is a positive real-valued random variable with the property that for every fixed time s, $[\tau \leq s] \in \mathcal{F}(s)$. This means that the occurrence of the event $[\tau \leq s]$ depends only on the information of the past and the present $\{X(u) : 0 \leq u \leq s\}$, not on any future information.

A Markov process has the **strong Markov property** if for $s \geq 0$, $\tau < \infty$ and $B \in \mathcal{B}(E)$

$$\mathbb{P}\big(X(\tau + s) \in B \,\big|\, \mathcal{F}(\tau)\big) = \mathbb{P}_{X(\tau)}(X(s) \in B).$$

If $\{X(t), t \geq 0\}$ is a stochastic process with independent increments, then $\{X(t), t \geq 0\}$ is a Markov process.

For a Markov process, we define the **transition kernel** $p(t; x, y)$ as the conditional probability

$$p(t; x, y)\, \mathrm{d}y = \mathbb{P}_x(X(t) \in \mathrm{d}y).$$

If the density of $X(0)$ is g_0, we write $\mathbb{P}_{g_0}$ to denote probabilities with respect to the density g_0, namely conditioned on $X(0)$ distributed as g_0, which are related to the transition kernel as

$$\mathbb{P}_{g_0}(X(t) \in B) = \int_E \int_{y \in B} p(t; x, y) g_0(x)\, \mathrm{d}y\, \mathrm{d}x, \quad B \in \mathcal{B}(E)$$

We say that g is an **invariant density** of the Markov process $X(t)$ if, for every $t \geq 0$ and for every $x \in \mathbb{R}$ we have that $g(x)\,\mathrm{d}x = \mathbb{P}_g(X(t) \in \mathrm{d}x)$. This means that the distribution of $X(t)$ does not depend on time, namely if $X(0)$ has a density g, then for every t, the random variable $X(t)$ has the same density g.

A.3. Poisson process

Let $\tau_1, \tau_2, \tau_3, \ldots$ be an i.i.d. sequence of exponentially distributed random variables with mean $1/\lambda > 0$. Let $T_n = \tau_1 + \tau_2 + \cdots + \tau_n$ for $n \geq 1$, $T_0 = 0$, and define a counting process $N(t) = \max\{n : T_n \leq t\}$, $t \geq 0$. The random variables τ_n are referred to as **inter-arrival times**, T_n as the arrival times and $N(t)$ is the number of arrivals by time t. We say that $\{N(t), t \geq 0\}$ is a **Poisson process** with intensity λ.

The Poisson process $N(t)$ with intensity λ has the following three properties:

i. $N(0) = 0$,

ii. $N(t)$ has stationary independent increments,

iii. For all $0 \leq s \leq t$, $N(t) - N(s)$ has a Poisson distribution with intensity $\lambda(t-s)$. This implies that $\mathbb{E}[N(t)] = \lambda t$.

$N(t)$ is a continuous parameter Markov process.

A **Marked Poisson Process** (MPP) with intensity λ is a double sequence $(\{T_n\}_{n=1}^{\infty}, \{Y_n\}_{n=1}^{\infty})$, such that $\{T_n\}$ are i.i.d. exponential variables with mean $1/\lambda$ and $\{Y_n\}_{n=1}^{\infty}$ are i.i.d. non-negative random variables having a common density f_Y and independent of the associated Poisson process $\{N(t), t \geq 0\}$. The random variables $\{Y_n\}_{n=1}^{\infty}$ are referred to as the marks. This means that if we associate a random value Y_n to each arrival time T_n, we form a MPP.

If we sum all the random values of $\{Y_n\}$ until time t, we form a Compound Poisson Process. Let $(\{T_n\}_{n=1}^{\infty}, \{Y_n\}_{n=1}^{\infty})$ be a MPP with intensity λ and let $N(t) = \max\{n : T_n \leq t\}$, $t \geq 0$ be the associated Poisson process. The process defined by

$$Z(t) = \sum_{n=1}^{N(t)} Y_n \tag{A-11}$$

is called a **compound Poisson process (CPP)**. The process (A-11) also has independent increments and is therefore Markovian.

The expected value of the CPP can be computed as the product of the mean number of arrivals by time t, $\mathbb{E}[N(t)]$, and the common mean of the marks, $\mathbb{E}[Y_n]$, i.e.

$$\mathbb{E}[Z(t)] = \lambda t\, \mathbb{E}[Y_n] \tag{A-12}$$

The variance of the CPP is

$$\mathbb{V}\mathrm{ar}[Z(t)] = \lambda t\, \mathbb{E}[Y_n^2] \tag{A-13}$$

The stochastic integral of a stochastic process $\{\phi(t), t \geq 0\}$ with respect to compound Poisson process $\{Z(t), t \geq 0\}$ is

$$\int_0^t \phi(s)\, \mathrm{d}Y(s) = \sum_{n=1}^{N(t)} \phi(t_n) Y_n. \tag{A-14}$$

A.4. Bell Polynomials and gamma function

Definition A.4.1. The partial or incomplete exponential Bell polynomial $B_{n,k}$ is defined as

$$B_{n,k}(\{x_i\}_{i=1}^{n-k+1}) = B_{n,k}(x_1, x_2, \ldots, x_{n-k+1}) = n! \sum_{\boldsymbol{j} \in I(n,k)} \prod_{i=1}^{n-k+1} \frac{1}{j_i!} \left(\frac{x_i}{i!}\right)^{j_i}. \tag{A-15}$$

Here, $\boldsymbol{j} \in I(n,k)$ denotes that the sum is taken over all vectors $\boldsymbol{j} = \{j_1, j_2, \ldots, j_{n-k+1}\}$ of non-negative integers such that

$$\sum_{i=1}^{n-k+1} j_i = k, \quad \sum_{i=1}^{n-k+1} i j_i = n. \tag{A-16}$$

The sum

$$B_n(\{x_i\}_{i=1}^{n}) = B_n(x_1, x_2, \ldots, x_n) = \sum_{k=1}^{n} B_{n,k}(\{x_i\}_{i=1}^{n-k+1}) \tag{A-17}$$

is called the n-th complete exponential Bell polynomial.

For instance, the first four complete exponential Bell polynomials are

$$\begin{aligned} B_1(x_1) =& x_1, \\ B_2(x_1, x_2) =& x_1^2 + x_2, \\ B_3((x_1, x_2, x_3) =& x_1^3 + 3x_1x_2 + x_3, \\ B_4(x_1, x_2, x_3, x_4) =& x_1^4 + 6x_1^2x_2 + 3x_2^2 + 4x_1x_3 + x_4. \end{aligned}$$

Proposition A.1. Let B_n be the n-th complete exponential Bell polynomial. Then,

$$B_n\left(\{(-1)^i x_i\}_{i=1}^{n}\right) = (-1)^n B_n(\{x_i\}_{i=1}^{n}) \tag{A-18}$$

Proof.

$$\begin{aligned}
B_n\left(\{(-1)^i x_i\}_{i=1}^n\right) =& \sum_{k=1}^{n} B_{n,k}(\{(-1)^i x_i\}_{i=1}^{n-k+1}) \\
=& \sum_{k=1}^{n} n! \sum_{\mathbf{j}\in I(n,k)} \prod_{i=1}^{n-k+1} \frac{1}{j_i!}(-1)^{ij_i}\left(\frac{x_i}{i!}\right)^{j_i} \\
=& \sum_{k=1}^{n} n! \sum_{\mathbf{j}\in I(n,k)} (-1)^{\sum_{i=1}^{n-k+1} ij_i} \prod_{i=1}^{n-k+1} \frac{1}{j_i!}\left(\frac{x_i}{i!}\right)^{j_i} \\
=& \sum_{k=1}^{n} n! \sum_{\mathbf{j}\in I(n,k)} (-1)^{n} \prod_{i=1}^{n-k+1} \frac{1}{j_i!}\left(\frac{x_i}{i!}\right)^{j_i} \\
=& (-1)^n \sum_{k=1}^{n} B_{n,k}(\{x_i\}_{i=1}^{n-k+1}) \\
=& (-1)^n B_n(\{x_i\}_{i=1}^{n})
\end{aligned}$$

□

Definition A.4.2. For any positive integer n, the gamma function is defined as

$$\Gamma(n) = (n-1)!. \tag{A-19}$$

For any real number x, except the non-positive integers, the gamma function can be defined as an infinite product

$$\Gamma(x) = \frac{1}{x}\prod_{k=1}^{\infty} \frac{\left(1+\frac{1}{k}\right)^x}{1+\frac{x}{k}}. \tag{A-20}$$

By this definition, the gamma function satisfies $\Gamma(x) = (x-1)\Gamma(x-1)$ for all real numbers x except the non-positive integers. If we use this recurrence property, for a real number x and a positive integer i we get

$$\begin{aligned}
\Gamma(x+i) =& (x+i-1)\Gamma(x+i-1) \\
=& (x+i-1)(x+i-2)\Gamma(x+i-2) \\
=& (x+i-1)(x+i-2)\dots(x+1)\,x\,\Gamma(x) \\
=& \Gamma(x)\prod_{k=0}^{i-1} x+k.
\end{aligned} \tag{A-21}$$

Bibliography

Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Applied Mathematical Sciences. Springer-Verlag.

Álvarez-Villa, O. D., Vélez, J. I., and Poveda, G. (2011). Improved long-term mean annual rainfall fields for Colombia. *International Journal of Climatology*, 31(14):2194–2212.

Basso, S., Frascati, A., Marani, M., Schirmer, M., and Botter, G. (2015). Climatic and landscape controls on effective discharge. *Geophysical Research Letters*, 42(20):8441–8447.

Beerends, R., Morsche, H., van den Berg, J., and van de Vrie, E. (2003). *Fourier and Laplace Transforms*. Fourier and Laplace Transforms. Cambridge University Press.

Beven, K. J. (2011). *Rainfall-runoff modelling: the primer*. John Wiley & Sons.

Bhattacharya, R. N. and Waymire, E. C. (2009). *Stochastic processes with applications*, volume 61. Siam.

Bhunya, P., Panda, S., and Goel, M. (2011). Synthetic unit hydrograph methods: a critical review. *The Open Hydrology Journal*, 5(1).

Bingham, N., Goldie, C., and Teugels, J. (1989). *Regular Variation*. Number no. 1 in Encyclopedia of Mathematics and its Applications. Cambridge University Press.

Botter, G. (2010). Stochastic recession rates and the probabilistic structure of stream flows. *Water Resources Research*, 46(12).

Botter, G., Basso, S., Rodriguez-Iturbe, I., and Rinaldo, A. (2013). Resilience of river flow regimes. *Proceedings of the National Academy of Sciences*, 110(32):12925–12930.

Botter, G., Porporato, A., Daly, E., Rodriguez-Iturbe, I., and Rinaldo, A. (2007a). Probabilistic characterization of base flows in river basins: Roles of soil, vegetation, and geomorphology. *Water resources research*, 43(6).

Botter, G., Porporato, A., Rodriguez-Iturbe, I., and Rinaldo, A. (2007b). Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response. *Water resources research*, 43(2).

Botter, G., Porporato, A., Rodriguez-Iturbe, I., and Rinaldo, A. (2009). Nonlinear storage-discharge relations and catchment streamflow regimes. *Water resources research*, 45(10).

Botter, G., Zanardo, S., Porporato, A., Rodriguez-Iturbe, I., and Rinaldo, A. (2008). Ecohydrological model of flow duration curves and annual minima. *Water Resources Research*, 44(8).

Cho, H.-K., Bowman, K. P., and North, G. R. (2004). A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission. *Journal of Applied meteorology*, 43(11):1586–1597.

Chung, K. L. (2001). *A course in probability theory*. Academic press.

Claps, P., Giordano, A., and Laio, F. (2005). Advances in shot noise modeling of daily streamflows. *Advances in Water Resources*, 28(9):992–1000.

Davis, M. H. (1984). Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 353–388.

Dooge, J. C. (1959). A general theory of the unit hydrograph. *Journal of geophysical research*, 64(2):241–256.

Durrett, R. (1999). *Essentials of stochastic processes*, volume 1. Springer.

Eagleson, P. S. (1972). Dynamics of flood frequency. *Water Resources Research*, 8(4):878–898.

Gupta, V. K., Troutman, B. M., and Dawdy, D. R. (2007). Towards a nonlinear geophysical theory of floods in river networks: an overview of 20 years of progress. In *Nonlinear dynamics in geosciences*, pages 121–151. Springer New York, New York, NY.

Gupta, V. K. and Waymire, E. (1990). Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research: Atmospheres*, 95(D3):1999–2009.

Gupta, V. K. and Waymire, E. (1998). Spatial variability and scale invariance in hydrologic regionalization. *Scale dependence and scale invariance in hydrology*, pages 88–135.

Gupta, V. K., Waymire, E., and Wang, C. (1980). A representation of an instantaneous unit hydrograph from geomorphology. *Water resources research*, 16(5):855–862.

Hrachowitz, M. and Clark, M. P. (2017). Hess opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrology and Earth System Sciences*, 21(8):3953–3973.

Jacobsen, M. (2006). *Point process theory and applications*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA.

James, W. P., Winsor, P. W., and Williams, J. R. (1987). Synthetic unit hydrograph. *Journal of Water Resources Planning and Management*, 113(1):70–81.

Kallenberg, O. (2002). *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition.

Koch, R. W. (1985). A stochastic streamflow model based on physical principles. *Water Resources Research*, 21(4):545–553.

Konecny, F. (1992). On the shot-noise streamflow model and its applications. *Stochastic Hydrology and Hydraulics*, pages 1–15.

Leopold, L. B. and Maddock, T. (1953). *The hydraulic geometry of stream channels and some physiographic implications*, volume 252. US Government Printing Office.

Lindgren, G. (2006). Lectures on stationary stochastic processes. *PhD course of Lund's University*.

McGuire, K., McDonnell, J. J., Weiler, M., Kendall, C., McGlynn, B., Welker, J., and Seibert, J. (2005). The role of topography on catchment-scale water residence time. *Water Resources Research*, 41(5).

Menabde, M. and Sivapalan, M. (2001). Linking space–time variability of river runoff and rainfall fields: a dynamic approach. *Advances in Water Resources*, 24(9-10):1001–1014.

Moradkhani, H. and Sorooshian, S. (2009). General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In *Hydrological modelling and the water cycle*, pages 1–24. Springer.

Morlando, F., Cimorelli, L., Cozzolino, L., Mancini, G., Pianese, D., and Garofalo, F. (2016). Shot noise modeling of daily streamflows: A hybrid spectral-and time-domain calibration approach. *Water Resources Research*, 52(6):4730–4744.

Nakagawa, K. (2005). Tail probability of random variable and laplace transform. *Applicable Analysis*, 84(5):499–522.

Nash, J. (1957). The form of the instantaneous unit hydrograph. *International Association of Scientific Hydrology, Publ*, 3:114–121.

Nash, J. (1959). Systematic determination of unit hydrograph parameters. *Journal of Geophysical Research*, 64(1):111–115.

Nguyen, P., Thorstensen, A., Sorooshian, S., Hsu, K., and AghaKouchak, A. (2015). Flood forecasting and inundation mapping using hiresflood-uci and near-real-time satellite precipitation data: The 2008 iowa flood. *Journal of Hydrometeorology*, 16(3):1171–1183.

Peckham, S. D. and Gupta, V. K. (1999). A reformulation of horton's laws for large river networks in terms of statistical self-similarity. *Water Resources Research*, 35(9):2763–2777.

Quintero, F., Krajewski, W. F., Mantilla, R., Small, S., and Seo, B.-C. (2016). A spatial–dynamical framework for evaluation of satellite rainfall products for flood prediction. *Journal of Hydrometeorology*, 17(8):2137–2154.

Ramirez, J. M. and Constantinescu, C. (2020). Dynamics of drainage under stochastic rainfall in river networks. *Stochastics and Dynamics*, page 2050042.

Reggiani, P., Sivapalan, M., and Hassanizadeh, S. M. (1998). A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics. *Advances in Water Resources*, 22(4):367–398.

Rice, S. O. (1944). Mathematical analysis of random noise. *Bell System Technical Journal*, 23(3):282–332.

Rinaldo, A., Marani, A., and Rigon, R. (1991). Geomorphological dispersion. *Water Resources Research*, 27(4):513–525.

Rodriguez-Iturbe, I., Porporato, A., Ridolfi, L., Isham, V., and Coxi, D. (1999). Probabilistic modelling of water balance at a point: the role of climate, soil and vegetation. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 455, pages 3789–3805. The Royal Society.

Rodríguez-Iturbe, I. and Valdes, J. B. (1979). The geomorphologic structure of hydrologic response. *Water resources research*, 15(6):1409–1420.

Saco, P. M. and Kumar, P. (2002). Kinematic dispersion in stream networks 1. coupling hydraulic and network geometry. *Water Resources Research*, 38(11).

Salisu, D., Supiah, S., Azmi, A., et al. (2010). Modeling the distribution of rainfall intensity using hourly data. *American Journal of Environmental Sciences*, 6(3):238–243.

Sato, K. and Yamazato, M. (1984). Operator-selfdecomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type. *Stochastic processes and their applications*, 17(1):73–100.

Snyder, F. F. (1938). Synthetic unit-graphs. *Eos, Transactions American Geophysical Union*, 19(1):447–454.

Suweis, S., Bertuzzo, E., Botter, G., Porporato, A., Rodriguez-Iturbe, I., and Rinaldo, A. (2010). Impact of stochastic fluctuations in storage-discharge relations on streamflow distributions. *Water resources research*, 46(3).

Te Chow, V., Maidment, D. R., and Mays, L. W. (1962). Applied hydrology. *Journal of Engineering Education*, 308:1959.

Urrea, V., Ochoa, A., and Mesa, O. (2019). Seasonality of rainfall in colombia. *Water Resources Research*.

Van der Tak, L. D., Bras, R. L., et al. (1989). Incorporating hillslope effects into the geomorphologic instantaneous unit hydrograph. In *Hydrology and Water Resources Symposium 1989: Comparisons in Austral Hydrology; Preprints of Papers*, page 399. Institution of Engineers, Australia.

Walsh, J. B. (2012). *Knowing the odds: an introduction to probability*, volume 139. American Mathematical Soc.

Zakian, V. (1969). Numerical inversion of laplace transform. *Electronics Letters*, 5(6):120–121.