The role of flow regimes for sediment transport and flooding potential of river catchments

presented to the Faculty of Science of the University of Neuchâtel to satisfy the requirements of the degree of Doctor of Philosophy in Science

by

Stefano Basso

Supervisory Committee:
Prof. Dr. Mario Schirmer, Université de Neuchâtel (Director of the thesis)
Prof. Dr. Daniel Hunkeler, Université de Neuchâtel (Co-director of the thesis)
Prof. Dr. Gianluca Botter, Università degli Studi di Padova (Technical supervisor)
Prof. Dr. Peter Molnar, ETH Zürich (External reviewer)

Defended:
04.05.2016

Université de Neuchâtel
2016

UNIVERSITÉ DE NEUCHÂTEL

Faculté des sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel - Suisse
Tél: + 41 (0)32 718 2100
E-mail: secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Monsieur Stefano BASSO

Titre:

"The role of flow regimes for sediment transport and flooding potential of river catchments"

sur le rapport des membres du jury composé comme suit:

- Prof. Mario Schirmer, directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Daniel Hunkeler, co-directeur de thèse, Université de Neuchâtel, Suisse
- Prof. Gianluca Botter, Università degli Studi di Padova, Italie
- Prof. Peter Molnar, ETH Zürich, Suisse

Neuchâtel, le 20 mai 2016

Le Doyen, Prof. B. Colbois

SUMMARY

The river flow regime is the main driver of several processes occurring in riverine environments, which are relevant for the sustainable management of water resources. It also determines the sensitivity of the flow distribution itself to a changing climate, and possibly affects catchments' behaviors with respect to extreme flows. For this reason, understanding links among hydrologic and eco-morphological processes which shape riverine environments is pivotal to ensure safety against flood hazards and the protection of human-valued ecosystem services. In order to reach these goals, quantitative investigations on the links between the flow regime and commonly used metrics in river geomorphology and engineering were pursued in this study.

The variety of the effective discharge for sediment transport observed in different river catchments is here related to the underlying heterogeneity of flow regimes. The main climatic and landscape drivers of the effective discharge are identified through an analytic framework, which links the effective ratio (i.e. the ratio between effective discharge and mean streamflow) to the empirical exponent of the sediment rating curve and to the streamflow variability. The analysis shows that different streamflow dynamics intrinsic to erratic and persistent flow regimes (respectively characterized by high and low flow variability) cause the emergence of diverse effective ratios, with larger values associated to erratic regimes. The provided formulation predicts patterns of effective ratios versus streamflow variability observed in a set of catchments of the continental United States, and may support the estimate of effective discharge in rivers belonging to diverse climatic areas.

The capability of a mechanistic analytical model of streamflow distributions to capture statistical features of high flows is also investigated. The model builds on a stochastic description of soil moisture dynamics and a simplified hydrologic response, described through different catchment-scale storage-discharge relations. The results show that non-linear relations are needed for a proper characterization of high flows frequencies and to explain the emergence of heavy-tailed streamflow distributions, which is mechanistically linked to the degree of non-linearity of the catchment hydrologic response.

Finally, a novel physically-based analytic expression of the seasonal flood-frequency curve is proposed. The expression is derived from a stochastic model of daily discharge dynamics, whose parameters embody climate and landscape attributes of the contributing catchment and can be estimated from daily rainfall and streamflow data. Only one parameter, which is related to the antecedent wetness condition in the watershed, requires calibration on the observed maxima. Applications in two rivers featured by contrasting daily flow regimes (erratic and persistent) are used to illustrate the potential of the method, which is able to capture diverse shapes of flood-frequency curves emerging in different climatic settings. The model provides reliable estimates of the seasonal maximum flows associated to a set of recurrence intervals, and its performances do not significantly decrease for return times longer than the available sample size. This result is due to the model structure, which allows for an efficient exploitation of the information contained in the entire range of daily flows experienced by rivers. Therefore, the proposed approach may be especially valuable in data scarce regions of the world.

RÉSUMÉ

Le régime d'écoulement de la rivière est le principal moteur de plusieurs processus qui se produisent dans des environnements fluviaux. Ces processus sont importants pour la gestion durable des ressources en eau. Le régime d'écoulement détermine aussi la sensibilité de la distribution du flux lui-même à une variation climatique, et affecte les comportements des bassins versants dans le cas de flux élevés. Pour cette raison, la compréhension des liens entre les processus hydrologiques et éco-morphologiques qui façonnent les environnements fluviaux est essentielle pour assurer la sécurité contre les risques d'inondation et la protection des services écosystémiques humains. Afin d'atteindre ces objectifs, il est proposé ici une étude quantitative sur les liens entre le régime d'écoulement et les métriques couramment utilisées en géomorphologie de la rivière et en ingénierie.

La variété de la décharge efficace pour le transport des sédiments observés dans différents bassins de rivière est ici liée à l'hétérogénéité sous-jacente des régimes d'écoulement. Les principaux éléments climatiques et du paysage responsables de la décharge effective sont identifiés grâce à un cadre analytique, qui relie le rapport effectif (i.e. le rapport entre le flux efficace et flux moyen) à l'exposant empirique des courbes de notation des sédiments et à la variabilité des flux. L'analyse montre que différentes dynamiques intrinsèques aux flux erratiques et aux flux persistants (caractérisées par la variabilité élevée et faible du flux) provoquent l'émergence de différentes ratios effectifs, avec des valeurs plus élevées associées à des régimes de flux erratiques. Le modèle prédit bien les rapport entre le ratio effectif et la variabilité des flux d'un ensemble de bassins des Etats-Unis continentaux, et peut appuyer l'estimation de la décharge effective dans des rivières qui appartiennent à différentes zones climatiques.

La capacité d'un modèle mécanique analytique des distributions des flux à capturer les propriétés des flux élevés est également étudiée. Le modèle se fonde sur une description stochastique de la dynamique de l'humidité du sol, et sur une réponse hydrologique simplifiée, décrite par des différentes relations de stockage-décharge. Les résultats montrent que des relations non-linéaires sont nécessaires pour la caractérisation correcte des fréquences des flux élevés et pour expliquer l'émergence de distributions a queue lourde du flux, ce qui est mécaniquement lié au degré de non-linéarité de la réponse hydrologique du bassin.

Enfin, une nouvelle expression analytique pour expliquer les courbes inondations-fréquence saisonnière est proposée. L'expression est dérivé d'un modèle stochastique de la dynamique quotidienne de décharge, dont les paramètres représentent les attributs du climat et de paysage du bassin et peuvent être estimés à partir des données quotidiennes des précipitations et des flux. Un seul paramètre, qui est lié à l'antécédent état d'humidité dans le bassin, nécessite une calibration sur les maxima observés. Les modèles sont appliqués dans deux rivières présentées en comparant les régimes quotidiens d'écoulement (erratique et persistante) pour montrer lefficacité de la méthode, qui est capable de capturer différentes formes de courbes de fréquence d'inondation émergentes dans différents contextes climatiques. Le modèle fournit des bonnes estimations des

flux maximaux saisonniers associées à un ensemble d'intervalles de récurrence, et les performances du modèle ne diminuent pas de manière significative pour des temps de retour plus longs que la taille de l'échantillon disponible. Ce résultat est dû à la structure du modèle, qui permet une exploitation efficace de l'information contenue dans l'ensemble des flux quotidiens des rivières. Par conséquent, l'approche peut être particulièrement utile dans les régions du monde pour lesquelles les données sont limitées.

Keywords: effective discharge, streamflow variability, river flow regimes, stochastic analytical model, suspended sediment transport, flow dynamics, streamflow distributions, high flows, heavy-tails, ecohydrology, recession analysis, storage-discharge, flood-frequency curves, physically-based, data scarcity.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The heavy hydrological modifications faced by rivers during the 20th century in many regions of the world, especially due to damming and intense human exploitation of water resources, triggered interest in understanding the influence of streamflow fluctuations on riverine environments and their capability to provide human-valued ecosystem services.

Poff et al. [1997], by introducing the paradigm of natural flow regime, offered a benchmark to assess the dynamic state of rivers and acknowledged the driving role of flow dynamics for ecological and morphological processes. The control exerted by the flow regime on e.g. availability of habitats for biological species, riparian vegetation growth, transport of solutes and sediments, inundation dynamics of river floodplains and hydropower production is mediated by physical, chemical and biological variables (e.g. water depth and velocity, temperature, turbidity, oxygen content) which influence the concerned processes.

In order to provide a quantitative description of the streamflow dynamics associated to distinct flow regimes, lists of relevant hydrologic metrics have been compiled [*Richter et al.,* 1996]. These metrics are usually derived from data series or distributed physically-based models of catchments, whose parameters are calibrated in order to reproduce the observed hydrograph. Distributed models perform well for the study and prediction of punctual events, such as the erosion of specific river reaches. However, their intense data and time requirements make them less suited when large scale patterns and river management issues are investigated, or for comparing the driving role of flow regimes for processes encompassing different fields (e.g. geomorphology and ecology) and several case studies.

In these cases, a quantitative description of the flow regime through the probability distribution of streamflows provides an alternative framework, where ecological and morphological features can be directly linked to the underlying streamflows and the effect of their overall variability can be assessed using a derived distribution approach. Examples of this approach are represented by the works of *Vogel and Fennessey* [1995], *Camporeale and Ridolfi* [2006] and *Doyle and Shields* [2008].

Also the characterization of flooding by rivers has been traditionally based on high dimensional models (e.g., for predicting timing and peak values of high flows) or statistical methods (e.g. *Castellarin et al.* [2012]), which profit from the available observations of extreme events to estimate frequency and magnitude of high flows, disregarding their links to the underlying hydrologic regime. These approaches proved successful, and allowed for large improvements in the evaluation of hazard posed by flooding to lives and properties.

However, natural and human-induced changes of the driving processes (e.g. precipitation and streamflow generation) [*Milly et al.,* 2005] and data scarcity intrinsic to most regions of the world demand for an efficient exploitation of all available information, included those provided by intermediate and low flows. Indeed, the flow regime may contain signatures of the flooding

potential of a river, and provide precious information that can be used to lower the related risk. First steps in this direction have been made by *Sivapalan et al.* [2005] and *Guo et al.* [2014], who investigated the relation between catchment water balance and flooding behavior through experimental and modeling approaches.

Recently, *Botter et al.* [2007a] proposed a lumped mechanistic-stochastic model of daily flow dynamics, which provides an analytic expression of the probability distribution of streamflows in terms of climatic and landscape attributes of the contributing basin. The framework applies to rain-fed catchments during seasonal time frames, and allows for a classification of river flow regimes based on the underlying daily flow variability. Erratic and persistent regimes are identified, respectively characterized by high and low flow variability. The approach builds on a catchment-scale balance of the soil moisture in the root zone, driven by stochastic increments due to infiltration from rainfall and losses due to evapotranspiration and events of instantaneous release of the water exceeding a certain soil moisture threshold (i.e. the water holding capacity of the soil). These events constitute the effective rainfall and recharge the hydrologic storage of the catchment. From here the water is released to the river according to a deterministic storage-discharge relation, which can be either linear or non-linear.

The introduction of this framework, which is especially suited to investigate similarity and differences among rivers and to identify general behaviors associated to distinct hydrologic regimes, fostered studies on the role of the flow regime as driver of e.g. riparian vegetation dynamics [*Doulatyari et al.,* 2015], hydropower production [*Basso and Botter,* 2012; *Lazzaro et al.,* 2013] and grazing by benthic organisms [*Ceola et al.,* 2013]. The set of riverine natural processes studied within this framework is further expanded in the work illustrated in the following, where issues related to river geomorphology (namely the estimation of the effective discharge) and flooding potential of catchments are investigated.

Aims and objectives

The general goal of this work is to provide a deeper understanding of the driving role of climatic and morphologic attributes of river catchments for a set of riverine processes ranging from sediment transport to flooding potential. The focus is on commonly used metrics and tools applied in practice, as the effective discharge and the flood-frequency curve. This objective is pursued through the novel development of stochastic physically-based analytical tools, which are then tested against observational data. These approaches are especially designed to deal with data scarcity, in order to allow for their widespread application also in sparsely gauged regions of the world.

The specific research objectives of this work are:

- advancing the capability of an existing stochastic model of streamflow to characterize high flows and tails of flow distributions;
- deriving physically-based analytic models to estimate effective discharge and flood-frequency curves in catchments belonging to different climatic regions;
- investigating the capability of the selected classification of flow regimes to unravel the observed varieties of the effective discharge and the flood-frequency curve;

- disentangling climatic and landscape controls on sediment transport and flooding behaviors in river catchments;

Particular attention is devoted to the development of tools suited to be applied in context characterized by data scarcity.

The achievement of these goals constitutes an essential step to study the role of high flows in river ecosystems and develop robust approaches providing a quantitative characterization of riverine dynamics.

Thesis structure

Results of the outlined research are presented in this manuscript as follows:

Chapter 2 discusses the causes of the observed heterogeneity of the effective discharge for sediment transport. An analytic framework is introduced, which links the effective discharge to river flow variability and the underlying climatic and landscape controls. The effects of distinct streamflow dynamics typical of erratic and persistent regimes are identified, and results checked through data gathered in 18 case studies.

Chapter 3 investigates the capability of a mechanistic stochastic model of flow duration curves to resemble statistical properties of high flows. 16 catchments are used as case studies. Different relations between catchment water storage and discharge, and alternative parameter estimation methods are evaluated. Conditions leading to the emergence of heavy-tailed streamflow distributions are also identified, and their implications for high flows characterization debated.

Chapter 4 presents a novel physically-based analytical model of flood-frequency curves. The analytical derivation of the extremal distribution of streamflows is described, and outcomes of the established link between daily flow dynamics and hydrologic extremes are discussed through a set of applications in two catchments. The link between flow regime and shape of the flood-frequency curve and the model performance in data scarce contexts are especially discussed.

Chapter 5 summarizes pivotal findings and conclusions. Future challenges and research perspectives on these topics are also delineated.

2. CLIMATIC AND LANDSCAPE CONTROLS ON EFFECTIVE DISCHARGE

Basso, S.[123], A. Frascati[4], M. Marani[35], M. Schirmer[12], G. Botter[5] (2015), Climatic and landscape controls on effective discharge, *Geophys. Res. Lett.*, 42, 84418447, doi:10.1002/2015GL066014.

Abstract

The effective discharge constitutes a key concept in river science and engineering. Notwithstanding many years of studies, a full understanding of the effective discharge determinants is still challenged by the variety of values identified for different river catchments. The present paper relates the observed diversity of effective discharge to the underlying heterogeneity of flow regimes. An analytic framework is proposed, which links the effective ratio (i.e. the ratio between effective discharge and mean streamflow) to the empirical exponent of the sediment rating curve and to the streamflow variability, as resulting from climatic and landscape drivers. The analytic formulation predicts patterns of effective ratio versus streamflow variability observed in a set of catchments of the continental United States, and helps in disentangling the major climatic and landscape drivers of sediment transport in rivers. The findings highlight larger effective ratios of erratic hydrologic regimes (characterized by high flow variability) compared to those exhibited by persistent regimes, which are attributable to intrinsically different streamflow dynamics. The framework provides support for the estimate of effective discharge in rivers belonging to diverse climatic areas.

2.1 Introduction

The concept of dominant or channel-forming discharge is central to geomorphological sciences, river engineering and restoration practices. Described as the discharge which shapes the cross section of natural rivers or transports the most sediments over long time periods, the dominant flow is used in the design of stable morphological configuration of channels [*Shields et al.,* 2003; *Doyle et al.,* 2007] and to estimate sedimentation rate and lifespan of reservoirs [*Podolak and Doyle,* 2015]. Moreover, dominant flows summarize the hydrologic forcing in models studying long-term evolution of rivers (see e.g. *Frascati and Lanzoni* [2009]) and the related incision patterns (see *Lague et al.* [2005] and references therein).

[1] Department of Water Resources and Drinking Water, Eawag - Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland.
[2] Centre for Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Neuchâtel, Switzerland.
[3] Department EOS, Duke University, Durham, NC, USA.
[4] Shell Global Solutions International, Rijswijk, The Netherlands.
[5] Department ICEA and International Center for Hydrology "Dino Tonini", University of Padova, Padua, Italy.

Several methods have been proposed in the literature to estimate the magnitude of the channel-forming discharge. In some cases, the dominant flow is identified with the bankfull discharge, evaluated from field surveys as the break point between main channel and floodplain [*Andrews,* 1980]. Other studies suggest equivalence between the channel-forming discharge and flows characterized by suitable return times (e.g. 1.5 years) [*Simon et al.,* 2004]. In order to address the role of both morphological and hydrological factors, *Wolman and Miller* [1960] proposed to estimate the dominant flow by combining information contained in the frequency distribution of streamflows and the sediment rating curve, giving rise to the concept of effective discharge. Thereafter, empirical, theoretical and numerical studies on effective discharge flourished in the literature, providing sometimes contrasting results (see e.g. *Bunte et al.* [2014]). Effective discharge was found to vary significantly among catchments as a function of climate and sediment rating characteristics, which encapsulate differences of morphology, bed sediment composition, hydrodynamic conditions and, as a consequence, between dominant transport mechanisms (i.e. suspended versus bedload) [*Emmett and Wolman,* 2001; *Simon et al.,* 2004]. Previous investigations (e.g. *Vogel et al.* [2003]; *Goodwin* [2004]; *Doyle and Shields* [2008]; *Quader and Guo* [2009]) also suggest the effective discharge to be significantly affected by the variability of river flows. For example, *Bolla Pittaluga et al.* [2014] showed that fluctuations of the hydrodynamic forcing implied by hydrologic variability do not prevent rivers from achieving a quasi-equilibrium morphodynamic state linked to a steady effective forcing, which differs however from typical channel-forming estimates (e.g. the bankfull discharge). Notwithstanding many years of studies, a consistent framework that enables separating hydrologic and landscape controls on the effective discharge and explains the observed patterns of sediment delivery across a gradient of climatic and landscape conditions is still lacking.

In this work, a framework is presented which links the effective discharge to the degree of non-linearity of the sediment rating curve and to streamflow variability, as resulting from climatic and landscape drivers. The framework builds on a physically-based stochastic representation of streamflow dynamics and a lumped description of the sediment transport capacity, and is tested using suspended sediment data observed in a set of catchments in the continental United States. The proposed framework gives insight into the major climatic and morphologic controls of sediment transport in rivers.

2.2 Observed effective discharge and streamflow variability

In this paper, the effective discharge is defined as the constant water flow rate resulting in the same long-term sediment load generated by the complete frequency distribution of streamflows. *Doyle and Shields* [2008] termed it functional-equivalent discharge q_f, to distinguish it from alternate definitions proposed in the literature (e.g. *Wolman and Miller* [1960] and *Vogel et al.* [2003]). Mathematically, the functional-equivalent discharge q_f is defined as:

$$\int_0^{+\infty} q_s p_s(q_s) dq_s = \beta \, {q_f}^{\delta} \tag{2.1}$$

where q_s is a stochastic variable representing the flow of sediments at-a-station and $p_s(q_s)$ its probability density function. The empirical coefficients δ and β describe the instantaneous power law relation between water and sediment flows, $q_s = \beta q^{\delta}$ (hereafter termed sediment rating curve), whose physical interpretation has been the goal of extensive research (e.g. *Syvitski et al.* [2000]). The validity of such a relation is a basic assumption of this work. Therefore, cases

where the power law relation between water and sediment flows may be significantly distorted (e.g. supply-limited catchments, rivers with extensive floodplains or strongly varying sediment storage) are not considered.

In order to introduce dimensionless quantities, we define the functional-equivalent ratio (R_f) as:

$$R_f = \frac{q_f}{< q >} = \frac{\left[\frac{1}{\beta}\int_0^{+\infty} q_s p_s(q_s) dq_s\right]^{\frac{1}{\delta}}}{< q >} \tag{2.2}$$

R_f scales the functional-equivalent discharge q_f by the average discharge $< q >$ (see also *Tucker and Bras* [2000]), which is a common statistic of river flows, relatively simple to estimate also in poorly gauged regions based on climatic data (e.g. *Hrachowitz et al.* [2013]). Since the sediment flow q_s is a non-linear function of q, its probability density function, $p_s(q_s)$ (right-hand side of eq. (2.2)), should strongly depend on the underlying streamflow distribution. As a consequence, the streamflow variability constitutes a major driver of R_f. The effective ratio can be profitably used to infer long-term (e.g. annual or decadal) sediment transport rates from instantaneous load-discharge relations, correctly accounting for the actual range of streamflows experienced by rivers.

The relation between R_f and streamflow variability is here analyzed by using a data set of 18 catchments in the continental US (Figure 2.1a), for which synchronous data of water and suspended sediment flows were available (summary information for the case studies are reported in Supporting Information, Tables S2.1 and S2.2). Morphological and climatic attributes of the catchments are diverse, ranging from the semi-arid South-Central to the humid Eastern US, and spanning rugged mountainous terrains, gently rolling hills and flatter plains. The land cover is primarily agricultural and forest, with some catchments including urbanized areas. The analyzed rivers are not impacted by significant flow regulation. In some cases, river reaches have undergone straightening and channelization, and topsoil or bank and streambed erosion is a known issue. The mean grain size of sediments transported in suspension (d_{50}) ranges from less than 0.001 to 0.029 mm.

Figure 2.1b displays (white dots) observed R_f as a function of the coefficient of variation of streamflows, CV_q (defined as the ratio between standard deviation and mean of the observed flows), for every seasons of the considered case studies. Despite some scattering, a clear pattern emerges from the plot shown in Figure 2.1b: low values of the functional-equivalent ratio are associated to low CV_q, whereas R_f increases significantly for increasing values of the coefficient of variation of flows. Accordingly, the functional-equivalent discharge q_f is only slightly larger than the average discharge for rivers exhibiting weak variability of flows, whereas it increases (up to 5 times the average flow) for rivers characterized by pronounced streamflow variability.

2.3 An explanatory framework

The observed variations of the functional-equivalent ratio as a function of streamflow variability are explained by adopting a lumped framework recently proposed by *Botter et al.* [2013] to characterize and classify flow regimes. According to this framework, the variability of river flows results from the interplay between the frequency of effective (i.e. streamflow-producing) rainfall events (λ) and the mean catchment response time ($1/k$). Censoring of rainfall events by soil moisture deficit controls the value of λ, which is bounded from above by the rainfall

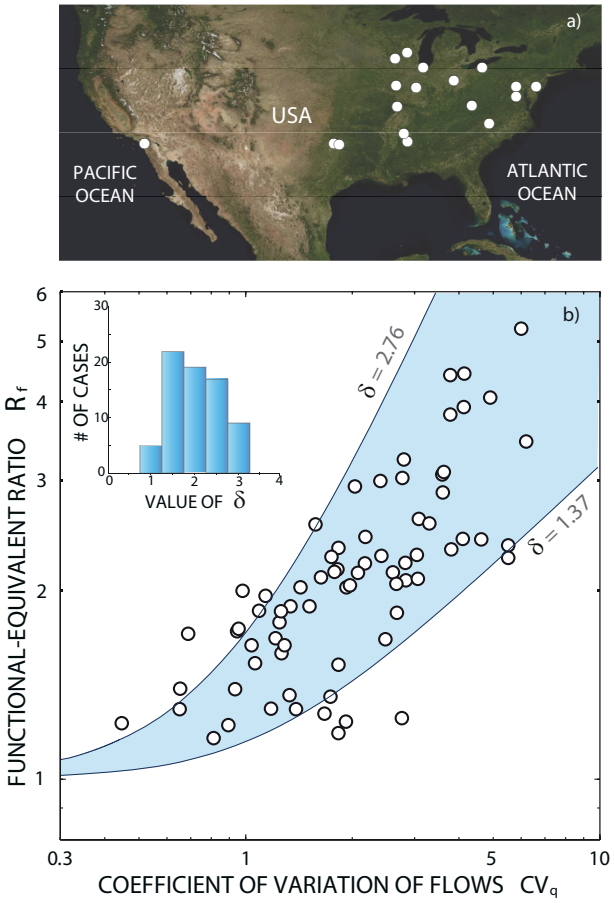

Fig. 2.1: a) Outlets locations of catchments considered in this study. b) The observed functional-equivalent ratio R_f scales with the coefficient of variation of streamflows CV_q. Solid lines containing most of the observations (shadowed area) represent estimate of the analytical model (eq. (2.3)) for δ equal to the 0.1 ($\delta = 1.37$) and 0.9 ($\delta = 2.76$) quantiles of its observed distribution (shown in the inset).

frequency and chiefly depends on catchment-scale evapotranspiration rates (in turn determined by vegetation cover and climate). The parameter k expresses the flow decay rate when recessions are assumed exponential, and its value embeds catchment-scale morphological and hydrological features, like the mean length of hydrologic pathways and soil conductivity [*Botter et al.,* 2007b].

When the mean interarrival time of flow-producing rainfall events is shorter than the mean catchment response time ($\lambda > k$), the river is continually fed by pulses delivered from the contributing catchment, and the range of streamflows observed between pulses is reduced. River flows are weakly variable around the mean, and the arising flow regime is termed persistent. When $\lambda < k$, effective rainfall events are interspersed in between long periods of flow recession, and a wider range of streamflows is observed. In this case an erratic regime emerges, characterized by high flow variability. The ratio λ/k (termed persistency index) fully determines the coefficient of variation of streamflows, since $CV_q = \sqrt{k/\lambda}$ in this framework (see Supporting Information).

Coupling the analytical expression for the probability distribution of streamflows derived by *Botter et al.* [2007a] (see Supporting Information) and the sediment rating curve $q_s = \beta\, q^{\delta}$, analytical expressions for the probability distribution of sediment flows, $p_s(q_s)$, and its statistical moments can be derived (eqs.(S2.4) and (S2.5) in Supporting Information). The integral in eq.(2.2) can thus be expressed as a function of climate and landscape attributes of the catchment. Accordingly, the functional-equivalent ratio R_f can be expressed as:

$$R_f = \frac{k}{\lambda}\left[\frac{\Gamma\left(\frac{\lambda}{k}+\delta\right)}{\Gamma\left(\frac{\lambda}{k}\right)}\right]^{\frac{1}{\delta}} = CV_q^2\left[\frac{\Gamma\left(CV_q^{-2}+\delta\right)}{\Gamma\left(CV_q^{-2}\right)}\right]^{\frac{1}{\delta}} \tag{2.3}$$

Notice that the functional-equivalent ratio only depends on the persistency index λ/k (or, alternatively, on the coefficient of variation of streamflows) and on the exponent δ of the sediment rating curve. The ratio λ/k embeds the effects of rainfall variability and soil drainage, whereas δ summarizes sediments size (i.e. suspended or bedload transport), erodibility of hillslopes and local conditions of river bed (e.g. armoring) [*Bunte et al.,* 2014].

When δ is integer, eq.(2.3) can be written in a simpler way:

$$R_f = \begin{cases} 1 & \text{if } \delta = 1 \\ \sqrt{1+\frac{k}{\lambda}} & \text{if } \delta = 2 \\ \sqrt[3]{\left(1+\frac{k}{\lambda}\right)\left(1+2\frac{k}{\lambda}\right)} & \text{if } \delta = 3 \end{cases} \tag{2.4}$$

Eq. (2.4) clarifies that when the catchment hydrologic response is flashy (high k), the rainfall frequency is low or the soil water deficit in the root zone is pronounced (i.e. low λ, as it might happen in semi-arid climates) the value of R_f increases (i.e. the functional-equivalent discharge is larger than the mean). However, the sensitivity to λ/k is modulated by the value of δ, as discussed later.

In Figure 2.1b, the analytical expression for R_f (eq. (2.3)) is plotted as a function of CV_q, by assuming δ equal to the 0.1 ($\delta = 1.37$) and 0.9 ($\delta = 2.76$) quantiles of its empirical distribution across the case studies (shown in the inset). The analytical curves (contours of the shadowed area) contain most of the observations, thereby suggesting that the proposed model is able to capture the first-order controls on the effective discharge.

Figure 2.2a displays (continuous lines) theoretical patterns of the functional-equivalent ratio R_f as a function of the persistency index, for different values of δ. R_f tends to one for very

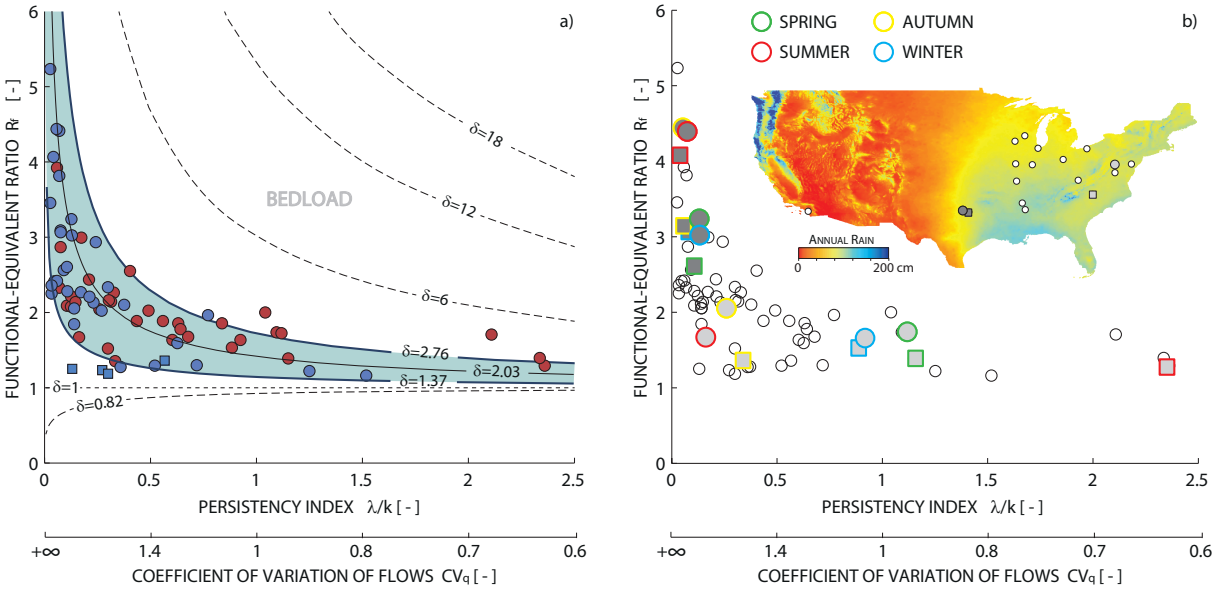

Fig. 2.2: a) Functional-equivalent ratio R_f as a function of the persistency index λ/k. Thick solid lines represent estimate of the analytical model for δ equal to the 0.1 ($\delta = 1.37$) and 0.9 ($\delta = 2.76$) quantiles of its observed distribution, and contain most of the observations (shadowed area), which refer to suspended sediment transport. Blue squares are cases with $\delta \sim 1$ ($\delta < 1.20$). Blue dots tag catchments with observed δ lower than the mean value ($\delta < 2.03$), while basins with $\delta > 2.03$ are marked by red dots. The analytical estimate for the discriminating δ value is represented with a thin solid line. Analytical estimates for other values of δ usually associated to bed or dissolved loads are also displayed with dotted lines. b) Clustering of R_f values of catchments belonging to different climatic areas. Dark gray dots and squares represent basins in South-Central US (respectively Elm Fork and Little Elm Creek in Texas), while light gray dots and squares tag catchments in Eastern US (respectively Conococheague Creek in Maryland and South Yadkin River in North Carolina). Locations of their outlets are displayed in the map, which also shows total annual rainfall throughout the United States. Green, red, yellow and blue contours of markers respectively indicate spring, summer, autumn and winter seasons.

stable (persistent) regimes, and significantly increases for low values of the persistency index (i.e. high values of CV_q characteristic of erratic flow regimes), approaching infinity for extremely variable flows ($\lambda/k \mapsto 0$). The higher the exponent of the sediment rating curve the higher the effective ratio and the smoother its increment with decreasing λ/k (increasing flow variability). High values of δ ($\delta > 3$, see Figure 2.2a) are typically associated to bedload transport [*Emmett and Wolman,* 2001], and correspond to highly non-linear sediment rating curves which mimic the effect of a minimum threshold of movement. Interestingly, *Bunte et al.* [2014] and *Lanzoni et al.* [2015] suggested that the effective discharge for gravel transport should correspond to the maximum observed discharge, which implies $R_f >> 1$ (in agreement with the theoretical analysis presented here). On the other hand, $\delta < 1$ causes a decrease of the functional-equivalent ratio with increased flow variability. Values of δ smaller than one are uncommon in sediment transport [*Sholtes et al.,* 2014], but may be appropriate for characterizing the flushing of geogenic and anthropogenic solutes [*Neal et al.,* 2012], where in fact upscaled load-discharge relations need to be suitably adjusted to account for hydrologic variability (e.g. *Basu et al.* [2010]).

Values of δ in the catchments analyzed in this study well represent the range of values reported in the literature for suspended sediment flows ($\delta \in [1, 3]$) [*Nash,* 1994]. Case studies displaying $\delta \sim 1$ ($\delta < 1.20$, tagged with squares in Figure 2.2a) lay near the corresponding analytical curve ($R_f = 1$) and thus confirm that in these circumstances the functional-equivalent discharge is almost insensitive to flow variability, being $q_f \sim< q >$ regardless of the type of flow regime. Colored dots in Figure 2.2a represent observed R_f versus $CV_q^{-2} = \lambda/k$ for the same set of catchments shown in Figure 2.1. Most of the observations fall within the range predicted by the analytical model (shadowed area in Figure 2.2a). Case studies exhibiting δ lower than the average observed value ($\delta < 2.03$) are tagged with blue dots, while red dots correspond to cases with $\delta > 2.03$. Though considerable scatter appears for high flow variability (left side of the plot), the stratification of the two groups of observations seems to support the layering predicted by the model for different exponents of the sediment rating curve, with larger values of R_f associated to larger values of δ.

The existence of distinct values of the functional-equivalent ratio for erratic ($CV_q > 1.1$, 59 cases) and persistent ($CV_q < 0.9$, 6 cases) regimes has been quantitatively checked by means of statistical hypothesis testing. Intermediate regimes ($0.9 < CV_q < 1.1$) have been excluded from the analysis, in agreement with *Botter et al.* [2013]. The null hypothesis that observed R_f values in persistent and erratic regimes are sampled from distributions with equal medians has been tested through the Wilcoxon rank sum test. The null hypothesis was rejected at the 5% significance level, thereby implying that the median functional-equivalent ratios of erratic and persistent rivers are statistically different. Because of the different size of the two samples (erratic and persistent regimes), the test has been repeated by dividing cases with CV_q lower or higher than the observed median value across the catchments/seasons considered ($CV_q = 1.92$), with analogous results.

Distinct streamflow dynamics exhibited by erratic and persistent regimes are the physical driver of the different behaviors of these systems. When flows weakly vary around their mean value (persistent regimes), the probability of high flows and even higher magnitudes of sediment transport (implied by non-linearity of the sediment rating curve) is very low. The magnitude of such events is overshadowed by the corresponding low probability of occurrence, and high flows weakly contribute to long-term sediment transport. Erratic regimes, instead, are composed of a sequence of high flows interspersed in between prolonged periods of low flows (i.e. droughts), which may not be effective in mobilizing sediments (e.g. the shear stress they exercise is under

the threshold of movement). As a result, only high flows are responsible for sediment transport, and the functional-equivalent discharge increases significantly.

Spatial patterns of river flow regimes are the complex by-product of large scale climatic drivers and local heterogeneities (e.g. soil, vegetation, geology). Nevertheless, when climatic attributes like seasonal rainfall and potential evapotranspiration are the primary controls on streamflow variability [*Botter et al.,* 2013], functional-equivalent ratios of rivers could mirror the underlying climatic patterns. An example of climatic clustering is shown in Figure 2.2b, which highlights R_f values of two groups of catchments subject to very diverse climatic conditions (map shows total annual precipitation). Light gray dots tag basins in Eastern US (Maryland and North Carolina), while dark gray dots indicate South-Central catchments (Texas). Because of their semi-arid climate, Texas catchments are characterized by extremely erratic flow regimes throughout the year, and are located in the left-upper part of the plot. For these rivers q_f is several times larger than the mean streamflow in all cases. R_f is instead less than 2 in every season for rivers of the Eastern Coast of US (right-lower part of the plot in Figure 2.2b), because of the more frequent rainfall input, which leads the seasonal regime to shift from persistent to slightly erratic.

These results, jointly with recent advances in the prediction of flow regimes using climatic and morphologic data [*Doulatyari et al.,* 2015], support the possibility of applying the present framework for first-order estimates of functional equivalent discharge in rivers belonging to different geographic and climatic areas, based on rainfall and landscape attributes. This represents the goal of ongoing research.

2.4 Conclusions

The diversity of effective discharges observed in rivers is here explained in terms of the underlying heterogeneity of flow regimes. The ratio between effective discharge and mean streamflow (effective ratio) is analytically expressed as a function of the exponent of the sediment rating curve and the coefficient of variation of daily flows, which in turn depends on streamflow-producing rainfall frequency and mean catchment response time. The analytic expression captures the first-order controls on the effective ratio for suspended sediment in a set of 18 case studies in the continental US. High values of the effective ratio are associated to larger exponents of the sediment rating curve and to more erratic flow regimes (high flow variability). Instead, the effective discharge is only slightly higher than the average flow in persistent regimes (weak flow variability). This is the by-product of distinct streamflow dynamics, which cause high flows to be the main responsible for sediment transport in erratic regimes. Conversely, the highest discharges weakly contribute to long-term load in persistent regimes. Values of the effective ratio can exhibit climatic signatures because of the strong control of evapotranspiration and rainfall regimes on flow variability. The formal linkage between effective ratio and flow regimes may constitute a valuable tool for preliminary estimates of the effective discharge in rivers belonging to different geographic and climatic settings.

Acknowledgments

The US Geological Survey (http://waterdata.usgs.gov) and the National Climatic Data Center (http://cdo.ncdc.noaa.gov/) are acknowledged for providing hydrologic and climatic data.

This study was funded by the Swiss National Science Foundation (SNF, Project No. 200021-149126). Additional support was provided by the Competence Center Environment and Sustainability (CCES) of the ETH domain in the framework of the RECORD and RECORD Catchment projects. The first author acknowledges a SNF Mobility Fellowship. Riccardo Sprocati (University of Padova, Italy) is gratefully acknowledged for his help in data manipulation and analysis. Thanks to Eric Deal (Universit Joseph Fourier, Grenoble, France), Michele Bolla Pittaluga (University of Genova, Italy) and an anonymous reviewer for their comments and suggestions.

2.5 Supplementary Information

2.5.1 Derivation of the analytical expression of the functional-equivalent ratio R_f (eq.(2.3)).

The probability density function of sediment flows, $p_s(q_s)$, is required to derive an analytic expression of the functional-equivalent ratio R_f defined in eq.(2.2). $p_s(q_s)$ is here obtained by applying the derived distribution rule:

$$p_s(q_s) = p(q(q_s))\frac{dq}{dq_s} \tag{S2.1}$$

The relation between q and q_s is described by a sediment rating curve, $q_s = \beta q^{\delta}$ (with β and δ empirical parameters), while the analytic distribution of streamflows, $p(q)$, is [*Botter et al.,* 2007a]:

$$p(q) = \frac{\Gamma(\lambda/k)^{-1}}{\alpha k}\left(\frac{q}{\alpha k}\right)^{\frac{\lambda}{k}-1}\exp\left(-\frac{q}{\alpha k}\right) \tag{S2.2}$$

where α is the mean rainfall depth, λ the frequency of effective rainfall, k the decay rate of recession flows and $\Gamma(\cdot)$ is the complete gamma function.

Mean and variance of the distribution above are respectively equal to $\lambda\alpha$ and $\lambda k\alpha^2$ [*Botter et al.,* 2013]. As a consequence, the analytic expression of the coefficient of variation of streamflows is:

$$CV_q = \frac{\sqrt{\lambda k\alpha^2}}{\lambda\alpha} = \sqrt{\frac{k}{\lambda}} \tag{S2.3}$$

By solving eq.(S2.1), an analytic expression of the probability distribution of sediment flows is obtained:

$$p_s(q_s) = \frac{\Gamma(\lambda/k)^{-1}}{\beta\delta(\alpha k)^{\delta}}\left(\frac{q_s}{\beta(\alpha k)^{\delta}}\right)^{\frac{\lambda}{k\delta}-1}\exp\left[-\left(\frac{q_s}{\beta(\alpha k)^{\delta}}\right)^{\frac{1}{\delta}}\right] \tag{S2.4}$$

The mean of the distribution above is:

$$< q_s >= \int_0^{+\infty} q_s p_s(q_s) dq_s = \beta(\alpha k)^{\delta}\frac{\Gamma(\lambda/k+\delta)}{\Gamma(\lambda/k)} \tag{S2.5}$$

Substituting eq.(S2.5) in eq.(2.2), and recalling that the average streamflow is $< q >= \alpha\lambda$ [*Botter et al.,* 2007a], the dependencies on β and on the mean rainfall depth α disappear and eq.(2.3) is obtained.

Catchment	State (Country)	Streamflow gauging station	Rainfall gauging stations	Area [km^2]	Period
San Juan Creek	California	Capistrano (CA)	Laguna Beach (CA)	303	1971-1985
Brandywine Creek	Delaware	Wilmington (DE)	Wilmington Porter (DE)	813	1947-1961
Des Plaines River	Illinois	Riverside (IL)	Oak Brook (IL)	1632	2003-2013
Kaskaskia River	Illinois	Cooks Mills (IL)	Tuscola (IL)	1225	1979-1997
McKee Creek	Illinois	Chambersburg (IL)	Perry (IL)	883	2002-2008
Johns Creek	Kentucky	Meta (KY)	Meta (KY)	145	1974-1982
Conococheague Creek	Maryland	Fairview (MD)	Chambersburg (PA)	1279	1966-1980
Hickahala Creek	Mississippi	Senatobia (MS)	Senatobia (MS)	313	1987-2000
Yalobusha River	Mississippi	Derma (MS)	Calhoun City (MS)	414	1998-2003
St. Francis River	Missouri	Saco (MO)	Fredericktown (MO)	1683	1989-1996
South Yadkin River	North Carolina	Mocksville (NC)	Mocksville (NC)	793	1958-1967
Grand River	Ohio	Painesville (OH)	Chardon (OH)	1774	1978-1991
Stillwater River	Ohio	Pleasant Hill (OH)	Pleasant Hill (OH)	1303	1963-1974
Elm Fork	Texas	Muenster (TX)	Muenster (TX)	119	1957-1968
Little Elm Creek	Texas	Celina (TX)	Gunter (TX)	121	1966-1975
Rappahannock River	Virginia	Remington (VA)	Culpeper (VA)	1603	1953-1990
Grant River	Wisconsin	Burton (WI)	Lancaster (WI)	699	1977-2004
Yahara River	Wisconsin	Windsor (WI)	Arlington University (WI)	192	1990-2014

Tab. S2.1: Summary information about the catchments used in this study.

2.5.2 Derivation of the simplified analytical expressions of the functional-equivalent ratio R_f (eq.(2.4)).

The simplified analytical expressions of the functional-equivalent ratio R_f (eq.(2.4)) are derived by substituting in eq.(2.3) suited integer values of δ, and by taking advantage of the recurrence relation of the gamma function, $\Gamma(z+n) = \Gamma(z) \cdot \prod_{i=1}^{n}(z+n-i)$, with n integer.

For $\delta = n = 1$ the recurrence relation is $\Gamma(z+1) = \Gamma(z) \cdot z$ and eq.(2.3) becomes:

$$R_f = \frac{k}{\lambda}\left[\frac{\Gamma\left(\frac{\lambda}{k}+1\right)}{\Gamma\left(\frac{\lambda}{k}\right)}\right] = \frac{k}{\lambda}\left[\frac{\Gamma\left(\frac{\lambda}{k}\right)\cdot\frac{\lambda}{k}}{\Gamma\left(\frac{\lambda}{k}\right)}\right] = \frac{k}{\lambda}\cdot\frac{\lambda}{k} = 1 \tag{S2.6}$$

The same procedure stands for the case $\delta = n = 2$ and $\delta = n = 3$, for which the recurrence relation is respectively $\Gamma(z+2) = \Gamma(z)\cdot(z+1)z$ and $\Gamma(z+3) = \Gamma(z)\cdot(z+2)(z+1)z$.

Catchment	Season	$< q >$ $[cm/d]$	CV_q [-]	δ [-]	$\beta \cdot 10^{-3}$ $[(cm/d)^{1-d}]$	d_{50} $[mm]$
San Juan Creek	Spring	0.05	4.14	2.75	7.12	0.001
	Summer	0.00	1.92	0.82	0.01	0.001
	Autumn	0.00	3.78	1.72	1.77	¡0.001
	Winter	0.04	5.99	2.00	2.33	0.002
Brandywine Creek	Spring	0.21	0.69	3.38	0.23	0.014
	Summer	0.11	1.34	2.62	0.34	0.007
	Autumn	0.09	1.57	2.58	0.21	0.014
	Winter	0.17	0.98	3.23	0.31	0.019
Des Plaines River	Spring	0.19	0.81	1.78	0.08	-
	Summer	0.11	1.18	1.66	0.09	-
	Autumn	0.07	1.39	1.57	0.07	-
	Winter	0.11	0.89	1.76	0.11	-
Kaskaskia River	Spring	0.16	1.33	1.17	0.03	-
	Summer	0.06	1.83	1.00	0.04	0.001
	Autumn	0.04	2.77	1.09	0.02	-
	Winter	0.11	1.67	1.20	0.02	-
McKee Creek	Spring	0.05	2.08	1.80	0.75	-
	Summer	0.03	6.20	1.36	0.28	-
	Autumn	0.04	4.63	1.60	0.52	-
	Winter	0.07	3.30	1.65	0.67	-
Johns Creek	Spring	0.19	1.75	2.31	0.38	0.013
	Summer	0.06	3.61	2.36	1.63	0.005
	Autumn	0.05	2.18	2.46	1.54	0.003
	Winter	0.19	1.78	2.14	0.39	0.007
Conococheague Creek	Spring	0.21	0.95	2.47	0.08	0.002
	Summer	0.09	2.48	2.57	0.70	0.001
	Autumn	0.10	1.97	2.23	0.11	0.002
	Winter	0.17	1.04	2.62	0.14	0.003
Hickahala Creek	Spring	0.18	2.83	2.25	0.74	-
	Summer	0.10	3.81	2.39	1.62	0.029
	Autumn	0.06	2.83	2.29	1.40	-
	Winter	0.22	2.60	2.23	0.61	0.016
Yalobusha River	Spring	0.26	2.67	1.60	0.20	-
	Summer	0.07	4.10	1.37	0.22	-
	Autumn	0.09	5.51	1.45	0.21	-
	Winter	0.28	2.67	1.63	0.20	-
St. Francis River	Spring	0.24	1.83	1.96	0.03	-
	Summer	0.03	2.17	1.28	0.01	-
	Autumn	0.09	5.51	1.61	0.04	-
	Winter	0.18	2.03	1.89	0.03	-
South Yadkin River	Spring	0.15	0.93	3.07	1.54	0.006
	Summer	0.08	0.65	2.77	4.29	-
	Autumn	0.09	1.73	2.76	1.90	-
	Winter	0.14	1.06	3.04	1.14	-
Grand River	Spring	0.20	1.26	2.09	0.10	0.011
	Summer	0.06	2.42	1.75	0.15	0.005
	Autumn	0.11	1.63	1.84	0.07	-
	Winter	0.22	1.25	2.10	0.09	-
Stillwater River	Spring	0.15	1.51	2.28	0.24	¡0.001
	Summer	0.04	1.93	1.97	0.60	0.001
	Autumn	0.03	3.05	1.72	0.14	¡0.001
	Winter	0.10	1.82	2.07	0.17	0.001
Elm Fork	Spring	0.06	2.80	1.64	0.23	0.003
	Summer	0.02	3.79	1.37	0.16	0.006
	Autumn	0.04	4.14	1.58	0.30	0.002
	Winter	0.03	2.77	1.38	0.07	0.002
Little Elm Creek	Spring	0.13	3.08	1.40	0.32	¡0.001
	Summer	0.04	4.90	1.40	0.32	0.001
	Autumn	0.09	3.63	1.44	0.30	¡0.001
	Winter	0.06	3.60	1.53	0.31	¡0.001
Rappahannock River	Spring	0.15	1.09	2.77	0.27	-
	Summer	0.06	3.06	2.61	0.93	-
	Autumn	0.06	2.40	2.22	0.19	0.006
	Winter	0.13	1.43	2.56	0.25	0.008
Grant River	Spring	0.08	0.96	2.84	4.27	0.011
	Summer	0.07	1.29	2.83	9.19	0.004
	Autumn	0.05	0.45	2.69	3.45	-
	Winter	0.05	1.14	1.95	0.38	-
Yahara River	Spring	0.05	1.21	2.27	0.70	-
	Summer	0.04	1.83	2.14	0.91	-
	Autumn	0.03	0.65	2.08	0.38	-
	Winter	0.03	1.26	1.82	0.16	-

Tab. S2.2: Seasonal features of the analyzed catchments: average streamflow $< q >$, coefficient of variation of flows CV_q, exponent δ and coefficient β of the sediment rating curve, mean grain size d_{50} of sediments transported in suspension. d_{50} has been calculated for each catchment and season as the median value among limited field measurements available.

3. ON THE EMERGENCE OF HEAVY-TAILED STREAMFLOW DISTRIBUTIONS

Basso, S.[12], M. Schirmer[12], G. Botter[3] (2015), On the emergence of heavy-tailed streamflow distributions, *Adv. Water. Resour.*, 82, 98-105, doi:10.1016/j.advwatres.2015.04.013.

Abstract

The right tail of streamflow distributions quantifies the occurrence probability of high flows, which play an important role in the dynamics of many eco-hydrological processes and eventually contribute to shape riverine environments. In this paper, the ability of a mechanistic analytical model for streamflow distributions to capture the statistical features of high flows has been investigated. The model couples a stochastic description of soil moisture dynamics with a simplified hydrologic response based on a catchment-scale storage-discharge relationship. Different types of relations between catchment water storage and discharge have been investigated, and alternative methods for parameter estimation have been compared using informal performance metrics and formal model selection criteria. The study highlights the pivotal role of non-linear storage-discharge relations in reproducing observed frequencies of high flows, and reveals the importance of analyzing the behavior of individual events for a reliable characterization of recession parameters. The emergence of heavy-tailed streamflow distributions is mechanistically linked to the degree of non-linearity of the catchment hydrologic response, with implications for the understanding of rivers' flooding potential and related ecologic and morphological processes.

3.1 Introduction

Heavy-tailed distributions are characteristic of many variables used in the description of natural and anthropogenic systems, including for example city populations and earthquake intensities [*Clauset et al.,* 2009]. These variables can assume values orders of magnitude greater than their averages, and are characterized by markedly skewed distributions which assign significant probabilities to extreme events [*Newman,* 2005].

Daily river flows have been previously recognized as a heavy-tail distributed variable [*Katz et al.,* 2002; *Bowers et al.,* 2012]. Several statistical and physically-based models have been developed to characterize runoff dynamics and estimate streamflow distributions (e.g. *Castellarin et al.* [2007]; *Botter et al.* [2007a]; *Kirchner* [2009]; *Yokoo and Sivapalan* [2011]), and rainfall-runoff

[1] Department of Water Resources and Drinking Water, Eawag - Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland.
[2] Centre for Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Neuchâtel, Switzerland.
[3] Department ICEA and International Center for Hydrology "Dino Tonini", University of Padova, Padua, Italy.

models have been specifically designed to incorporate heavy-tailed components (e.g. *Carreau et al.* [2009]).

Indeed, reliably modeling frequencies of high flows and identifying mechanisms promoting the emergence of heavy-tailed flow distributions entail important consequences for the characterization of a number of hydrological and ecological processes. For example, the enhanced frequencies of high flows associated with heavy-tailed distributions may result in wider areas of the river transect being affected by flooding, implying a wider aquatic/terrestrial transition zone within river corridors [*Tockner et al.,* 2000]. An accurate description of the tail of flow distributions is also important to characterize catchment-scale sediment transport, because of the role of high flows in mobilizing sediments and driving morphodynamic processes [*Surian,* 1999]. High flows are especially important for the assessment of formative discharges and in long-term landscape evolution models [*Snyder et al.,* 2003], and entail broad implications for different fields of geosciences that include reservoir sedimentation, geomorphology and river restoration [e.g. *Schirmer et al.,* 2014]. Furthermore, the features of the tail of streamflow distributions may strongly impact the probabilistic structure of extreme flows, whose characterization is an important task for hydrologists and engineers. In fact, the extreme value theory postulates that different shapes of the tail of flow distributions result in different types of extremal distributions [*Fisher and Tippett,* 1928], which denote distinct characters of flooding. Hence, factors controlling the emergence of heavy tails should be identified to link expected changes of climate and land use to resulting modifications of flow frequency and magnitude.

This work, which represents a preliminary step towards a physically-based characterization of flood frequencies and catchment-scale sediment flow dynamics, aims at analyzing performances of a mechanistic model of streamflow dynamics [*Botter et al.,* 2007a; 2009] in the range of high flows. Recently, the model has been extensively applied to characterize river flow regimes and a variety of biotic and anthropogenic riverine processes (see e.g. *Botter et al.* [2010]; *Botter et al.* [2013]; *Ceola et al.* [2013]; *Pumo et al.* [2013]; *Schaefli et al.* [2013]; *Doulatyari et al.* [2014]; *Lazzaro et al.* [2013]; *Mejia et al.* [2014]; *Müller et al.* [2014]). The ability of the model to reproduce observed probability distributions in the range of low to medium streamflows has been the object of previous studies [*Botter et al.,* 2007c; *Ceola et al.,* 2010]. However, model performance for high flows has never been investigated before. In order to improve the model ability to reproduce the frequencies of high flows, a new method for parameter estimation is also tested. Moreover, conditions that promote the emergence of heavy tails in streamflow distributions are investigated. The paper is organized as follows: Section 3.2 describes the probabilistic characterization of streamflow dynamics, case studies and parameter estimation procedures. In Section 3.3, results of the application of the model to a set of catchments are described with a specific focus on the highest streamflows. The main conclusions of the paper are summarized in Section 3.4.

3.2 Methods

3.2.1 Probabilistic characterization of streamflows

The primary tool used in this investigation is the mechanistic-stochastic model for streamflow dynamics developed by *Botter et al.* [2007a; 2009]. The model builds on a catchment-scale balance of the soil moisture in the root zone, as resulting from the following processes: (1) stochastic increments due to infiltration from rainfall, which is assumed to be a Poisson process

with frequency $\lambda_p[T^{-1}]$ and exponentially distributed depths with average $\alpha[L]$; (2) losses due to evapotranspiration; (3) instantaneous deep percolation above a certain soil moisture threshold. These deep percolation events, occurring with frequency $\lambda < \lambda_p[T^{-1}]$, recharge the subsurface storage of the catchment and produce effective rainfall which contributes to streamflow. The release of water is modelled by assuming a suitable storage-discharge relation which can be either linear or non-linear. In the former case, the dynamics of specific (per unit catchment area) streamflows q are described by:

$$\frac{dq}{dt} = -k\,q + \xi_1(t) \tag{3.1}$$

where $\xi_1(t)$ describes a sequence of random (exponentially distributed) streamflow increments due to percolation events and k is the hydrograph recession rate. In the latter case, instead, the dynamics of q are defined by:

$$\frac{dq}{dt} = -K\,q^a + \xi_2(t) \tag{3.2}$$

In analogy with eq. (3.1), the term $\xi_2(t)$ in eq. (3.2) represents (state dependent) streamflow increments due to percolation events; K and a (hereafter termed recession coefficient and exponent, respectively) represent the coefficient and exponent of the power law relation describing flow recessions.

Parameters k, K and a in equations (3.1) and (3.2) not only describe recession patterns, but also influence the increase of streamflow following an effective rainfall event (i.e. ξ_1 and ξ_2). While in eq. (3.1) a single catchment response time is assumed, thus implicitly referring to a specific component of the hydrologic response (e.g. subsurface runoff, see *Botter et al.* [2007a]), the non-linear model allows the catchment hydrologic response to vary as a function of the overall water storage. As a consequence, high saturation of the catchment results in higher streamflows, with no need to explicitly differentiate runoff generation mechanisms triggered by crossing of humidity (or rainfall) thresholds. Eq. (3.2) thus provides a consistent representation of the catchment hydrologic response, incorporating the effect of different flow components (deep, subsurface and surface runoff).

The model allows an analytical expression of the steady-state probability distribution (pdf) of specific streamflows, which reads [*Botter et al.,* 2007a; 2013]:

$$p(q) = \frac{\Gamma(\lambda/k)^{-1}}{\alpha k}\left(\frac{q}{\alpha k}\right)^{\frac{\lambda}{k}-1}\exp\left(-\frac{q}{\alpha k}\right) \tag{3.3}$$

in the linear case, and [*Botter et al.,* 2009; *Ceola et al.,* 2010]:

$$p(q) = C\,q^{-a}\exp\left(-\frac{q^{2-a}}{\alpha K(2-a)} + \frac{\lambda\,q^{1-a}}{K(1-a)}\right) \tag{3.4}$$

in the non-linear case. C in eq. (3.4) is a normalizing constant. The original formulation (see eq.(2) of *Ceola et al.* [2010]) includes an atom of probability for $q = 0$, possibly arising when $0 < a < 1$, which has not been produced here because in most cases $a > 1$ [*Biswal and Marani,* 2010].

The exceedance probability of q is obtained by integrating equations (3.1) or (3.2) as:

$$D(q) = \int_q^{+\infty} p(x)dx \tag{3.5}$$

where $D(q)$ represents the flow duration curve at-a-station.

3.2.2 Case studies

The models described in Section 3.2.1 have been applied to estimate the seasonal distribution of streamflows and the related cumulative exceedance probability (i.e., the flow duration curve) in 16 catchments located in northeastern Italy, Switzerland and the United States, for a total of 43 different combinations of catchments and seasons. The selected basins are characterized by different flow regimes, and their areas span two orders of magnitude (from 10 to 10^3 km^2). To comply with the basic assumptions of the models, rivers affected by anthropogenic regulation as well as regimes significantly impacted by snow dynamics have been excluded from the analysis. Table 3.1 reports the main features of the investigated catchments.

3.2.3 Parameter estimation

The characterization of equations (3.3) or (3.4) relies on the specification of three (α, λ, k) or four (α, λ, K, a) model parameters, which can be estimated based on hydrologic, climatic and geomorphologic information.

The mean rainfall depth α is estimated by using daily rainfall data recorded in several stations located within or nearby the considered basin. When synchronous data from different stations are available for the same catchment, rainfall records are first averaged to obtain spatially averaged rainfall series.

For the sake of simplicity, the frequency of effective rainfall λ is here estimated by equaling the observed mean specific discharge, $< q >$, and the analytical mean of q according to the stochastic model (i.e., $\lambda =< q > /\alpha$). An alternative estimation method exists, based solely on soil and climatic features of basins (see *Botter et al.* [2013]).

Parameters defining the storage-discharge relations are set by using two alternative methods, whose outcomes are compared. (1) The ensemble of points derived by plotting time derivatives of q (dq/dt) against the corresponding flow values is fitted by a linear or power-law curve, from which the parameters k, a and K are estimated (dashed black line in Figure 3.1). This is the standard procedure introduced by *Brutsaert and Nieber* [1977] and previously used by e.g. *Botter et al.* [2007c; 2009; 2013] and *Ceola et al.* [2010] to assess performance of the model described in Section 3.2.1. (2) A linear or power-law curve is fitted to pairs $dq/dt - q$ of each recession curve (colored lines in Figure 3.1) to estimate k, a and K of individual recessions. By keeping a constant and equal to the median of the set of observed values, K is then re-calculated for each recession curve. The median values of the parameters across all the observed recessions are finally considered for modeling purposes. This new procedure is justified by observations of *Biswal and Marani* [2010], who showed that each recession curve is typically featured by a steeper slope and a different intercept compared to the estimate performed on the ensemble of recessions (Figure 3.1).

In spite of the recognized inter-event variability of recession properties, constant (effective) parameters are used in the model to represent flow recessions. In fact, no analytical solutions are available for the streamflow pdf in the general case. Moreover, it has been shown that the variability of characteristic recession timescales affects only the third and higher moments of the streamflow distribution, without altering its mean and variance [*Botter,* 2010].

Table 3.2 shows estimated values of parameters λ_P, λ, α, k, a, K for the set of case studies.

Catchment	State (Country)	Streamflow gauging station	Rainfall gauging stations	Area [km^2]	Period	Seasons
Sand Run	West Virginia (USA)	Buckhannon (WV)	Buckhannon (WV)	37	1946-2012	Spring, Summer, Autumn, Winter
Rio Tanama	Puerto Rico (USA)	Utuado (PR)	Adjuntas (PR)	48	1970-2011	Spring, Summer, Autumn, Winter
Johns Creek	Kentucky (USA)	Meta (KY)	Meta (KY)	146	1974-1982	Spring, Summer, Autumn, Winter
Youghiogheny River	Maryland (USA)	Oakland (MD)	Oakland (MD)	347	1941-2012	Spring, Summer, Autumn, Winter
Glatt River	Switzerland	Herisau	Herisau	16	1970-2011	Spring, Summer, Autumn, Winter
Sitter River	Switzerland	Appenzell	Brülisau, Säntis	74	1970-2011	Spring, Summer, Autumn, Winter
Murg River	Switzerland	Wängi	Eschlikon	79	1970-2011	Spring, Summer, Autumn, Winter
Thur River	Switzerland	Jonschwil	Ricken, St. Peterzell, Starkenbach	493	1970-2011	Spring, Summer, Autumn, Winter
Thur River	Switzerland	Halden	Appenzell, Bischofszell, Eschlikon, Flawil, Herisau, Ricken, St. Peterzell, Starkenbach, Säntis, Teufen, Urnäsch	1085	1970-2011	Spring, Summer, Autumn, Winter
Thur River	Switzerland	Andelfingen	Affeltrangen, Andelfingen, Appenzell, Bischofszell, Eschlikon, Flawil, Frauenfeld, Herisau, Illhart, Kalchrain, Niederneunforn, Ricken, St. Peterzell, Starkenbach, Sulgen, Säntis, Teufen, Urnäsch, Weinfelden	1696	1970-2011	Spring, Summer, Autumn, Winter
Boite Creek	Italy	Podestagno	Podestagno	82	1992-2008	Summer
Boite Creek	Italy	Cancia	Podestagno, Faloria, Borca, Cortina	313	1986-2008	Summer
Cordevole River	Italy	La Vizza	P.sso Pordoi	8	1985-2008	Summer
Cordevole River	Italy	Saviner	P.sso Pordoi, Arabba, P.sso Falzarego	109	1990-2008	Summer
Fiorentina Creek	Italy	Sottorovei	P.sso Falzarego, Caprile, Pescul	58	1993-2008	Summer
Padola Creek	Italy	S. Stefano di Cadore	P.sso M. C. Comelico, Casamazzagno	130	1987-2007	Summer
Piave River	Italy	Ponte della Lasta	P.sso M. C. Comelico, Malga Campobon, S. Stefano di Cadore, Casamazzagno, Costalta, Cimacanale, Sappada	355	1990-2006	Summer

Tab. 3.1: Summary information about the catchments used in this study.

Catchment	**Season**	$\lambda_p[d^{-1}]$	$\lambda[d^{-1}]$	$\alpha[cm]$	$k[d^{-1}]$	$a[-]$	$K[cm^{1-a}/d^{2-a}]$
Sand Run	Spring	0.49	0.38	0.72	0.26	1.74	0.79
	Summer	0.42	0.09	0.95	0.44	1.54	1.49
	Autumn	0.38	0.12	0.76	0.32	1.76	1.86
	Winter	0.55	0.51	0.56	0.28	1.84	0.91
Rio Tanama	Spring	0.51	0.21	1.00	0.27	3.58	9.03
	Summer	0.56	0.23	1.00	0.24	4.20	9.99
	Autumn	0.70	0.34	1.20	0.22	4.21	2.03
	Winter	0.47	0.38	0.47	0.09	5.04	17.63
Johns Creek	Spring	0.36	0.23	0.85	0.20	1.79	0.81
	Summer	0.35	0.05	1.11	0.36	1.80	3.31
	Autumn	0.26	0.06	0.93	0.37	2.01	3.76
	Winter	0.33	0.24	0.80	0.23	1.82	0.98
Youghiogheny River	Spring	0.50	0.45	0.73	0.20	1.68	0.45
	Summer	0.42	0.13	0.89	0.30	1.71	0.99
	Autumn	0.39	0.15	0.74	0.23	1.93	1.35
	Winter	0.52	0.50	0.63	0.23	1.82	0.57
Glatt River (Herisau)	Spring	0.53	0.47	0.72	0.30	1.37	0.34
	Summer	0.54	0.32	1.03	0.49	1.53	0.49
	Autumn	0.47	0.34	0.74	0.51	1.37	0.39
	Winter	0.51	0.51	0.55	0.39	1.70	0.55
Sitter River (Appenzell)	Spring	0.63	0.58	0.86	0.25	2.08	0.50
	Summer	0.65	0.46	1.14	0.39	2.20	0.79
	Autumn	0.52	0.32	1.02	0.35	2.31	1.26
	Winter	0.57	0.28	0.92	0.32	1.98	0.76
Murg River (Wängi)	Spring	0.51	0.35	0.66	0.15	2.41	1.03
	Summer	0.51	0.19	0.86	0.23	2.39	1.68
	Autumn	0.49	0.24	0.67	0.21	2.61	2.45
	Winter	0.51	0.41	0.59	0.15	2.46	1.07
Thur River (Jonschwil)	Spring	0.59	0.59	0.78	0.20	2.25	0.48
	Summer	0.62	0.37	1.07	0.32	2.14	0.83
	Autumn	0.51	0.31	0.90	0.29	2.07	1.06
	Winter	0.54	0.37	0.76	0.22	2.14	0.87
Thur River (Halden)	Spring	0.66	0.60	0.62	0.21	2.20	0.55
	Summer	0.69	0.39	0.85	0.33	2.15	0.99
	Autumn	0.56	0.34	0.71	0.27	2.23	1.42
	Winter	0.60	0.43	0.58	0.21	2.17	0.95
Thur River (Andelfingen)	Spring	0.64	0.54	0.56	0.17	2.57	1.08
	Summer	0.66	0.33	0.75	0.28	2.21	1.33
	Autumn	0.56	0.32	0.60	0.23	2.43	2.29
	Winter	0.60	0.45	0.50	0.19	2.28	1.09
Boite Creek (Cancia)	Summer	0.65	0.45	0.65	0.06	4.09	2.63
Boite Creek (Podestagno)	Summer	0.56	0.38	0.81	0.09	3.17	1.56
Cordevole River (La Vizza)	Summer	0.56	0.43	0.85	0.13	2.24	0.27
Cordevole River (Saviner)	Summer	0.63	0.33	0.72	0.09	2.99	0.67
Fiorentina Creek	Summer	0.64	0.32	0.69	0.09	2.78	1.16
Padola Creek	Summer	0.58	0.32	0.84	0.07	2.85	0.45
Piave River	Summer	0.73	0.36	0.68	0.07	4.56	6.00

Tab. 3.2: Estimates of mean rainfall frequency (λ_P), mean frequency of effective rainfall (λ), mean rainfall depth (α), recession rate (k, linear model), recession exponent (a) and coefficient K (non-linear model) for the study catchments.

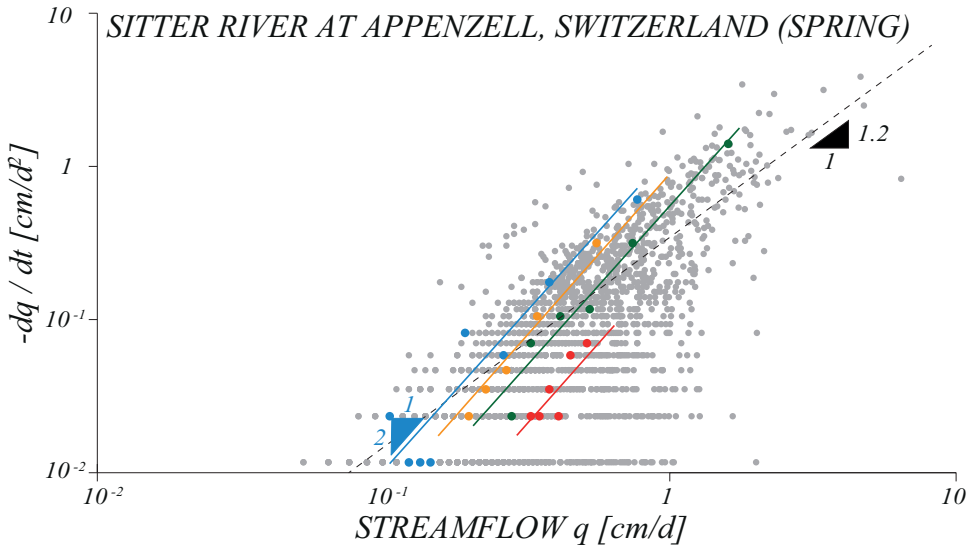

Fig. 3.1: Flow recession analysis for the Sitter River at Appenzell in the spring season. Grey dots represent the cloud of points obtained by plotting flow q versus time derivative of the flow $-dq/dt$ for decreasing branches of the hydrograph. Colored points and lines display four individual recessions and their regression lines. Notice that regression lines of individual recessions are parallel and shifted, and exhibit steeper slopes than the regression line of the ensemble of points (dotted black line).

Combining different versions of the model (linear or non-linear storage-discharge relations) and different methods for parameter estimation (fitting the ensemble of recessions or fitting each recession separately), we identified three model versions, whose performances have been compared. These versions are: (a) a linear version of the stochastic model (hereafter named model L) where the parameter k is set by first computing its value separately for each recession curve and then by selecting the median value of the observed set (method (2) above); (b) a non-linear version of the model (named model $NL1$ in the following), where parameters K and a are estimated from the ensemble of recessions, similarly to what was done by *Botter et al.* [2009] and *Ceola et al.* [2010] (method (1) above); (c) a non-linear version of the model (hereafter named model $NL2$) where a and K are estimated separately for each recession curve and then their median values are used (method (2) above).

Comparisons of alternative models with frequency distributions obtained from observed data series are carried out by displaying cdfs of streamflows in a log-log scale, which emphasizes the right tail of the distributions, thereby focusing on the highest flows which are the primary subject of this paper.

3.3 Results and Discussion

Results of the application of the model to a subset of the study catchments are displayed in Figures 3.2 and 3.3. All model versions prove able to reproduce quite well the observed exceedance probability in the range of low to medium streamflows, as already shown by previous studies (see e.g. *Botter et al.* [2007c]; *Ceola et al.* [2010]). However, their performances markedly differ for high flows, where model $NL2$ outperforms the others thanks to the new method of parameter estimation applied in this context.

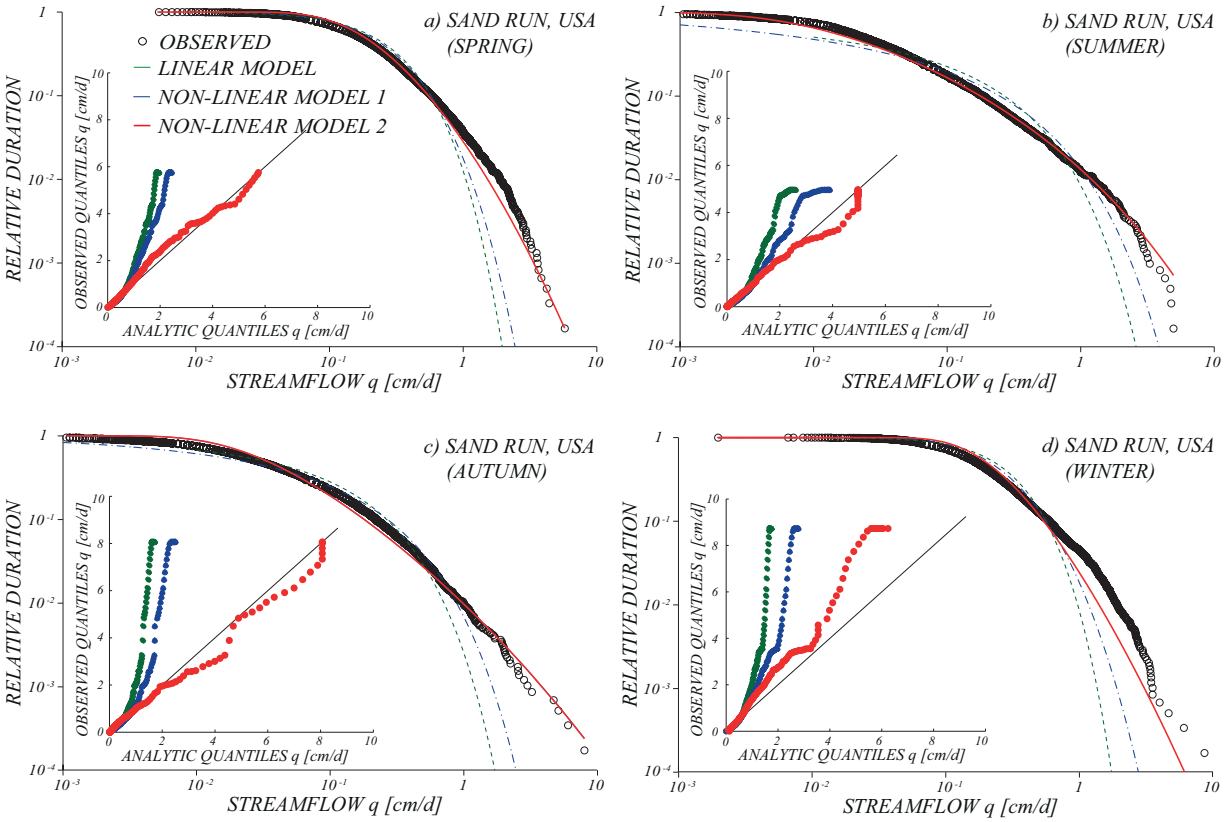

Fig. 3.2: Flow duration curves for the Sand Run (WV) catchment during spring (a), summer (b), autumn (c) and winter (d) seasons. Plots are on a log-log scale. Dots represent duration curves estimated from observed data series. Green dotted lines illustrate results of the linear model, while estimates of the non-linear model are represented with dashed-dotted blue lines ($NL1$) and solid red lines ($NL2$). Insets show q-q plots for the case studies.

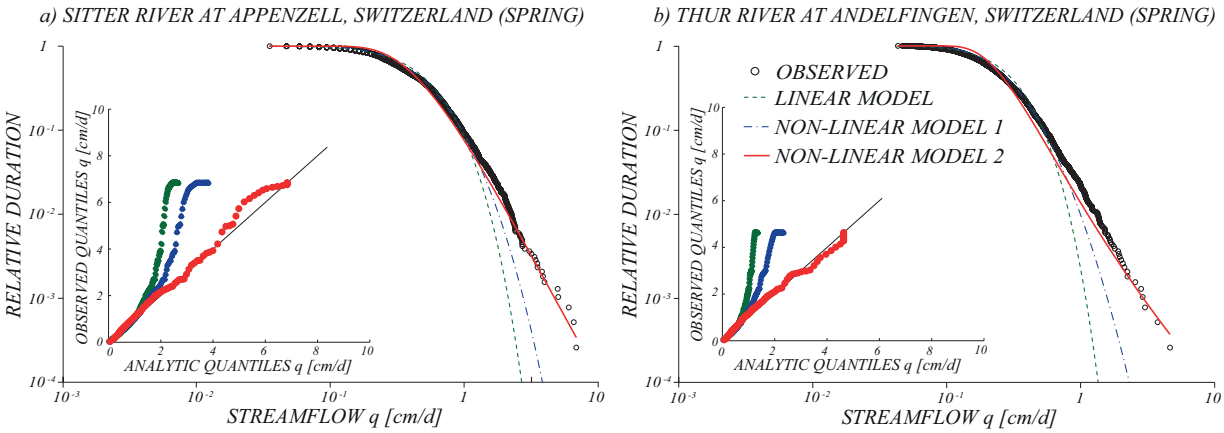

Fig. 3.3: Flow duration curves for (a) the Sitter river at Appenzell and (b) the Thur river at Andelfingen (Switzerland) during the spring season. Plots are on a log-log scale. Dots represent duration curves estimated from observed data series. Green dotted lines illustrate results of the linear model, while estimates of the non-linear model are represented with dashed-dotted blue lines ($NL1$) and solid red lines ($NL2$). Insets show q-q plots for the case studies.

Figure 3.2 displays flow cdfs for the Sand Run (WV) catchment in different seasons. The linear model L (dotted green lines) correctly reproduces the lower flows, but is unable to describe the frequency distribution of discharges greater than $1\,cm/d$. The non-linear model $NL1$, displayed with dashed-dotted blue lines, allows for a slight improvement in performance, but still fails to reproduce the shape of the tail of the observed cdfs. Model $NL2$, instead, guarantees in every season significantly improved estimates of the cumulative probability of high flows. Insets of Figure 3.2 represent q-q plots for the case studies. In order to emphasize performance for high flows, 100 log-spaced quantiles are selected in the range of cumulated probability $[10^{-4}, 1]$. These plots provide clear evidence of the superior results guaranteed by model $NL2$ as compared to model L and $NL1$, for which deviations from the 45° line are significant.

Similar behaviors have been obtained in most of the case studies, including the Sitter and Thur rivers shown in Figure 3.3. A summary of the performance of alternative models in representing observed cumulative distribution functions of streamflows is shown in Table 3.3 and Figure 3.4. Table 3.3 reports average values among all considered cases of the Nash-Sutcliffe coefficients (NSC), computed on the same quantiles used for q-q plots. These values indicate good performance of model $NL2$ ($NSC > 0.75$), which deteriorates to fair ($0.5 < NSC < 0.75$) and poor ($NSC < 0.5$) respectively for models $NL1$ and L [*Castellarin et al.,* 2004; *Müller et al.,* 2014]. Analogous results are obtained by using linearly spaced quantiles.

The goodness of fit is also assessed through the probability distribution of the root mean squared error ($RMSE$) between modeled and observed cdfs in all considered cases (Figure 3.4a). The linear model (green line) exhibits the highest mean $RMSE$. When a linear model is used, the $RMSE$ is greater than 0.3 in 44% of the cases, whereas this percentage drops to zero when a non-linear model is used. In particular, the $RMSE$ associated with model $NL2$ is less than 0.2 and its probability peaks around 0.1.

The Akaike's Information Criterion (AIC), properly rescaled to eliminate the effect of sample size [*Cudeck and Browne,* 1983; *Marsh et al.,* 1988], is also used. The AIC allows for a proper comparison of alternative models discounting the effects of different numbers of parameters. In particular, the AIC Differences (ΔAIC) and the Akaike Weights (AW) are used for a formal assessment of the best model. Table 3.3 lists mean ΔAIC and AW values for the alternative models. Increasing ΔAIC values identify less plausible models, while models with $\Delta AIC \leq 2$ are believed to have substantial evidence [*Burnham and Anderson*, 2002]. Based on the values of ΔAIC obtained for each combination of catchment/season, the number of times each model outperforms the others in reproducing the observed cdfs has been quantified (Figure 3.4b). The linear model L never provides the best performance, while model $NL1$ provides the best result only in 12 out of 43 cases. Model $NL2$ outperforms the others in 31 out of 43 cases. The Akaike Weights AW express (on a scale from 0 to 1) the weight of evidence in favor of a certain model to describe the analyzed phenomenon [*Burnham and Anderson*, 2002]. Figure 3.4c displays the probability distribution of the weights of models obtained from all the considered cases. While model L and $NL1$ are quite unlikely (mainly due to a misrepresentation of the tails), the observational evidence clearly supports model $NL2$, whose AW peaks around 0.6. Therefore, the additional parameter of the non-linear model proves to be significant, provided that it is estimated by separately analyzing each recession event (model $NL2$). The new estimation procedure may guarantee a better characterization of physical attributes of basins summarized by recession exponent and coefficient, namely the receding of the active drainage network and the average catchment humidity conditions [*Biswal and Marani*, 2010; *Biswal and Kumar,* 2014].

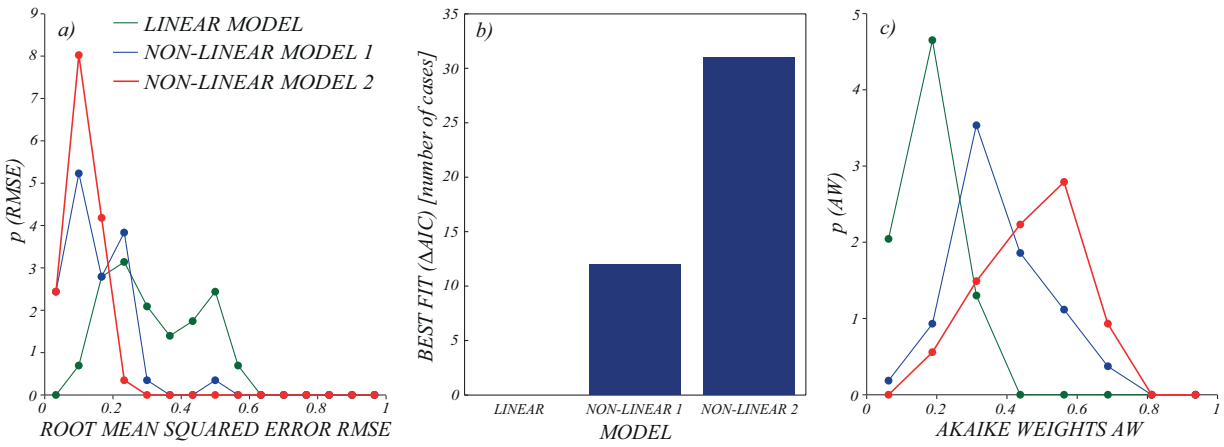

Fig. 3.4: Indices summarizing the goodness of fit of the alternative models with respect to observed data. (a) Probability distribution of the Root Mean Squared Error ($RMSE$) between modeled and observed flow cdfs; (b) Number of cases each model gives the best performance, measured by using Akaike Differences (ΔAIC); (c) Probability distribution of the Akaike Weights (AW) assigned to each model. In (a) and (c) green lines display the performance of model L, while blue and red lines the performance of models $NL1$ and $NL2$.

	Model L	Model $NL1$	Model $NL2$
NSC	0.24	0.60	0.88
$RMSE$	0.32	0.15	0.11
AIC_c	-2.47	-4.07	-4.51
ΔAIC	2.40	0.80	0.36
AW	0.17	0.37	0.46

Tab. 3.3: Goodness of fit of linear (L) and non-linear ($NL1$ and $NL2$) models. The table reports the mean values, obtained by averaging the results of the application of models to all the considered catchments and seasons, of Nash-Sutcliffe Coefficient (NSC), Root Mean Square Error ($RMSE$), rescaled Akaike Information Criterion (AIC_c), Akaike Differences (ΔAIC) and Akaike Weight (AW). $NSC = 1 - \frac{\sum_{f=1}^{u}(\hat{q}_f - q_f)^2}{\sum_{f=1}^{u}\left(q_f - \frac{1}{u}\sum_{f=1}^{u} q_f\right)^2}$, where f is the index of $u = 100$ log-spaced values of cumulated probability selected in the range $10^{-4}-1$, and q and $\hat{q}$ the corresponding observed and analytic flow values (quantiles). $RMSE = \sqrt{RSS/n}$, where $RSS = \sum_{i=1}^{n}(\log D_i - \log \hat{D}_i)^2$ is the residual sum of squares, and D_i and $\hat{D}_i$ are observed and analytic estimates of duration of a series of n flow values (sample size). $AIC_c = \frac{2(g+1)}{n} + ln(RSS/n) + \frac{2g(g+1)}{n-g-1}$, where g is the number of parameters of each model. $\Delta AIC = AIC_{c,m} - AIC_{c,min}$, where $AIC_{c,min}$ is the minimum value of AIC_c among those obtained for the r different models ($m \in [1, r]$). $AW = \frac{\exp(-\Delta AIC_m/2)}{\sum_{m=1}^{r}\exp(-\Delta AIC_m/2)}$.

When cumulative distribution functions are plotted on a log-log scale, power-law tails appear as straight lines that can be easily detected by visual inspection [*Newman*, 2005]. Figure 3.3 shows heavy-tailed streamflow distributions of two case studies, namely the Sitter river at Appenzell and the Thur river at Andelfingen (Switzerland). Also in these cases, model *NL*2 suitably reproduces the shape of the cdfs, showing superior performance as compared with models *L* and *NL*1.

The ability of model *NL*2 to accurately describe the frequency of high flows when $p(q)$ is heavy-tailed provides mathematical insight on the reasons underlying the emergence of power-law tails in streamflow distributions. The analytical expression of the streamflow pdf (eq. (3.4)) is a combination of an exponential term, which is responsible for the fast decay to zero of the cdf, and a power-law function, which guarantees instead relatively high probabilities for high flows:

$$p(q) \propto q^{-a} \exp\left(-q^{2-a} + q^{1-a}\right) \tag{3.6}$$

When the parameter a is less than 2, the exponential function in eq. (3.6) dominates the asymptotic behaviour of the pdf for $q \mapsto +\infty$, thereby dominating the power-law term q^{-a}, and the cdf quickly drops to zero. Conversely, if a is higher than 2 the exponential term tends to a constant for $q \mapsto +\infty$ and the asymptotic behaviour of $p(q)$ is dominated by q^{-a}.

The emergence of heavy-tailed flow distributions when $a > 2$ is linked to distinct streamflow dynamics. Indeed, higher flow jumps followed by faster recessions occur due to the pronounced non-linearity of the storage-discharge relation, leading to an enhanced probability of high flows. This is the case for the Sitter and Thur river catchments at the outlets of Appenzell and Andelfingen, where the estimated values of a for the spring season are 2.08 and 2.57, respectively. Consistently, the streamflow cdfs of the Sand Run catchment, for which the estimated values of a are 1.74 (spring), 1.54 (summer), 1.76 (autumn), and 1.84 (winter), do not display any power-law behavior.

Previous works (see *Gaume* [2006] for a review) identified the rainfall features as the main control of the statistical properties of high flows. Our analysis suggests that heavy-tailed streamflow distributions can emerge in response to enhanced non-linearities of the catchment hydrologic response, regardless of the underlying rainfall regime and the partitioning of precipitation into evapotranspiration and discharge. The physical origin of such non-linear responses is explained by alternative theories. The hydraulic interpretation [*Brutsaert and Nieber,* 1977] is based on the Boussinesq equation, which describes saturated flow over a sloping plane. In this case, the power law recession results from the time evolution of the water table gradient which drives flow through a hillslope. According to this theory, the catchment responds as a hillslope whose properties effectively represent the variety of responses exhibited by catchment landforms. A different explanation is given by *Harman et al.* [2009]. In his work, catchment-scale power law recessions emerge from the composition of varied hydraulic properties of hillslopes, each one assumed to behave like a linear reservoir, and diverse pattern of recharge events faced by catchments. Other studies connect recession properties to stream network morphology, especially focusing on expansion and contraction over time of the fraction of stream network actively contributing to the flow at the outlet (e.g. *Biswal and Marani* [2010]; *Mutzner et al.* [2013]; *Godsey and Kirchner* [2014]). The present finding, by relating the emergence of power-law tailed streamflow distributions to the recession exponent, allows the emergence of different types of extreme distributions to be linked to measurable features of the considered catchments, thereby providing some support in the choice of the most suitable distribution type in flood-frequency analysis [*Papalexiou and Koutsoyiannis,* 2013].

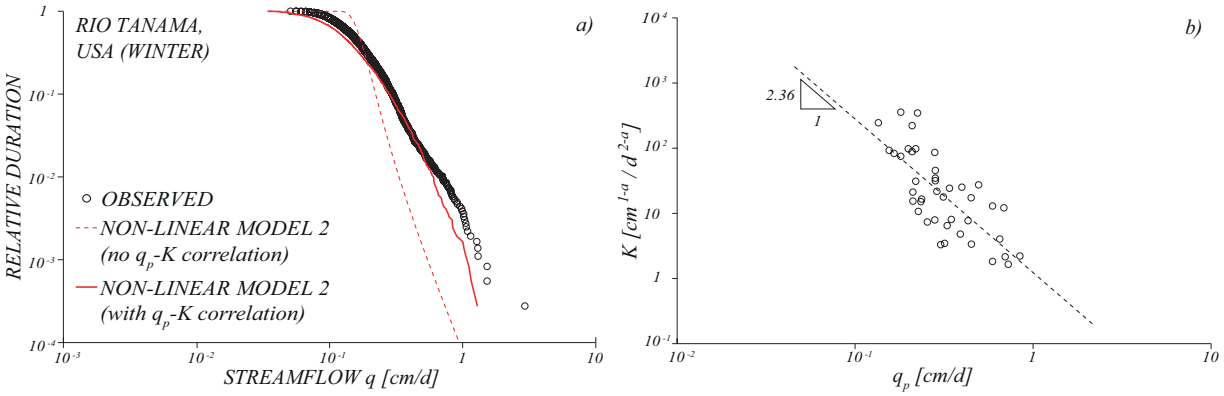

Fig. 3.5: (a) Flow duration curve for the Rio Tanama (PR) catchment during the winter season. Plots are on a log-log scale. Dots represent the duration curve estimated from the observed data series. The dotted red line displays the estimate of model *NL2* when a fixed K is used. The solid red line instead shows the estimate of model *NL2* when the observed correlation between q_p and K (b) is numerically accounted for in the model. For this purpose, the numerical counterpart of the analytical model has been run by selecting a different value of K from the regression (dotted line in Figure 5b) for each runoff event.

In a few cases, the non-linear model *NL2* does not perform properly in reproducing the observed flow cdf (e.g. the case of Rio Tanama, displayed in Figure 3.5a). In such cases model *NL2* is not able to reproduce the observed streamflow cdf for both low and high flows. The poor performance of the non-linear model is explained by the observed correlation between the peak value of the recession q_p and the recession coefficient K (Figure 3.5b), which makes using the median observed K ineffective. Numerically including the relationship between q_p and K in model *NL2* significantly improves its performance (solid red line in Figure 3.5a). An analogous correlation has not been detected between q_p and a. Physically, the variability of K has been related to seasonal evapotranspiration rates [*Shaw and Riha,* 2012] or to varying subsurface water storage and soil moisture conditions in catchments [*Biswal and Kumar,* 2014]. The correlation observed in Figure 3.5b may express the key role played by a varying humidity state in determining peak flows, and supports the explanation which links shifting of the recession coefficient K with varying basin-scale soil moisture storage [*Shaw et al.,* 2013]. Notice that the analogous $q_p - K$ relations in the remaining study catchments are quite weak, thereby enabling the use of the analytical model given by eq. (3.4) to describe the frequency of high flows. The contrasting behaviors exhibited by different catchments pose questions on the climatic and morphologic features which enhance or prevent the emergence of such a correlation. This is the object of ongoing research.

3.4 Conclusions

Different versions of a stochastic analytical model for streamflow distribution and the related flow duration curve have been compared to assess their ability to reproduce statistical properties of high flows. Methods of parameter estimation used in previous work have also been compared to a new procedure. The analysis shows that the non-linearity of the storage-discharge relation is a key ingredient to reproduce observed frequencies of high flows and to explain the emergence of heavy-tailed streamflow distributions. Regardless of the underlying hydro-climatic regime,

the power-law behaviour can arise only when the recession exponent a is higher than 2, i.e. for relatively high degree of non-linearity of the catchment hydrologic response. The slope of the tail can be also affected by the negative correlation between recession coefficients and corresponding peak flows, which may express the role of varying moisture conditions in determining peak flows. Furthermore, the analysis emphasizes the importance of studying recession patterns for individual events as opposed to the classical approach according to which all the available recessions are merged together prior to the fitting of the recession exponent.

Acknowledgments

This study was funded by the Swiss National Science Foundation (SNF, Project No. 200021-149126). Additional support was provided by the Competence Center Environment and Sustainability (CCES) of the ETH domain in the framework of the RECORD and RECORD Catchment projects. The Regional Agency for Environmental Protection of the Veneto Region, the Swiss Federal Offices for the Environment and for Meteorology and Climatology, the US Geological Survey (http://waterdata.usgs.gov) and the National Climatic Data Center (http://cdo.ncdc.noaa.gov/) are acknowledged for providing hydrologic and climatic data. John Molson (Laval University, Quebec, Canada), Behnam Doulatyari (Eawag, Switzerland) and three anonymous reviewers are gratefully acknowledged for their comments and suggestions.

4. A PHYSICALLY-BASED ANALYTICAL MODEL OF FLOOD-FREQUENCY CURVES

Basso, S.[1,2], M. Schirmer[1,2], G. Botter[3], A physically-based analytical model of flood-frequency curves, *Geophys. Res. Lett.*, Under Review.

Abstract

Predicting magnitude and frequency of floods is a key issue in hydrology, with implications in many fields ranging from river science and geomorphology to the insurance industry. In this paper, a novel physically-based approach is proposed to estimate the recurrence intervals of flow maxima. The method links the extremal distribution of streamflows to the stochastic dynamics of daily discharge, providing an analytical expression of the seasonal flood-frequency curve. The parameters involved in the formulation embody climate and landscape attributes of the contributing catchment, and can be estimated from daily rainfall and streamflow data. Only one parameter, which is linked to the antecedent wetness condition in the watershed, needs to be calibrated on the observed maxima. The potential of the method is discussed through a set of applications in two rivers featured by contrasting daily flow regimes (erratic and persistent). The model provides reliable estimates of seasonal maximum flows in different climatic settings, and is able to capture diverse shapes of flood-frequency curves emerging in erratic and persistent regimes. The proposed method embodies experimental information on the full range of discharges experienced by rivers. As a consequence, model performances do not deteriorate significantly when the magnitude of events with return times longer than the available sample size is estimated. The proposed approach provides a framework for the prediction of floods based on climate and landscape attributes, that may be especially valuable in data scarce regions of the world.

4.1 Introduction

Assessing magnitude and frequency of high flows in river basins is pivotal for many research and applied disciplines, encompassing urban planning and infrastructures' design, geomorphology, water resources management and the insurance industry (e.g. *Lague et al.* [2005]; *Horn and McShane* [2013]; *Poff et al.* [2016]).

Extremes are traditionally estimated exploiting observations and probabilistic tools based on asymptotic statistical theories [*Gnedenko,* 1943; *Coles,* 2001]. Many studies deal with the

[1] Department of Water Resources and Drinking Water, Eawag - Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland.
[2] Centre for Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Neuchâtel, Switzerland.
[3] Department ICEA and International Center for Hydrology "Dino Tonini", University of Padova, Padua, Italy.

application of these methods for characterizing extreme flows (see *Castellarin et al.* [2012]). Relevant topics of the recent literature include multivariate analysis of floods (e.g. *Grimaldi and Serinaldi* [2006]), parameters estimation methods (e.g. *Morrison and Smith* [2002]) and regional analysis of flood distributions (e.g. *Villarini and Smith* [2010]; *Salinas et al.* [2014]).

Large data samples are required to constrain statistical methods, especially when the magnitude of rare events is concerned. Since long-term flow records are unavailable in most locations, tools have been developed which capitalize on a variety of information to predict flood-frequency curves [*Merz and Blöschl,* 2008]. Approaches that exploit the information content of daily flow series [*Claps and Laio,* 2003], and frameworks that link catchment water balance and flooding potential [*Sivapalan et al.,* 2005; *Guo et al.,* 2014] have also been proposed to improve the reliability of predicted flood statistics.

A physically-based stochastic approach for estimating seasonal flood-frequency curves is proposed in this work. The method links the features of the extremal distribution of daily flows to a limited number of parameters embodying climate and landscape attributes, which are estimated exploiting information on daily rainfall and discharge. The potential of the proposed method is illustrated through a set of applications in two catchments featured by contrasting flow regimes, and the model performance to estimate the magnitude of events with return times longer than the available sample size is discussed.

4.2 Analytic expression of the seasonal flood-frequency curve

The seasonal flood-frequency curve is derived building on a mechanistic-stochastic model of daily streamflows introduced by *Botter et al.* [2009]. Daily flow dynamics result from a catchment-scale balance of the soil moisture in the root zone [*Laio et al.,* 2001] driven by: (1) stochastic increments due to infiltration from Poissonian rainfall; (2) losses due to evapotranspiration; (3) effective rainfall pulses triggered by the exceedance of a critical soil moisture threshold (say, the water holding capacity of the soil). The latter are modeled as a Poisson process with frequency $\lambda[T^{-1}]$ and exponentially distributed depths with average $\alpha[L]$.

In this framework, the dynamics of the specific (per unit catchment area) excess storage V (i.e., the fraction of water storage exceeding the holding capacity) are driven by the following equation:

$$\frac{dV}{dt} = -q + \xi(t) \tag{4.1}$$

where $\xi(t)$ is a Poisson sequence of uncorrelated storage increments due to effective rainfall events and $q = \rho(V)$ is a deterministic power law relation between specific discharge and excess storage (see supporting information), that quantifies how the excess water is drained from the catchment and becomes streamflow. Under the above assumptions, the probability density function (pdf) of the excess storage, $p(V)$, can be derived analytically as a function of λ, α and the parameters of $\rho(V)$ (see supporting information and *Botter et al.* [2009]). The existence of a monotonous storage-discharge relation enables a formal link between the statistical features of V and q. Storage dynamics described by eq. (4.1) (and the corresponding discharge dynamics) are represented in Figure 4.1.

The plot emphasizes that seasonal maxima of daily discharge belong to the subset of flows occurring immediately after jumps (grey dots in Figure 4.1), hereafter termed peakflows. Therefore, only the stochastic processes represented by the series of peakflows and peak storages (i.e. the excess storages right after streamflow jumps) are considered in the following derivations.

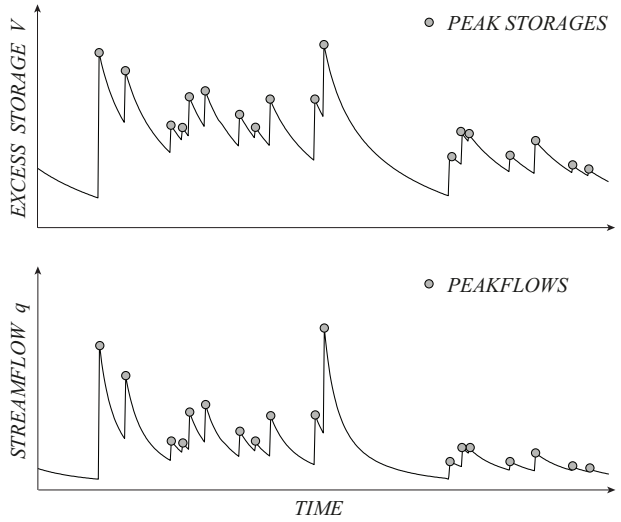

Fig. 4.1: Excess storage V and streamflow q dynamics as produced by the mechanistic-stochastic model used in this study [*Botter et al.,* 2009]. Flows (excess storages) occurring immediately after increments are termed peakflows (peak storages) and labeled with grey dots.

The probability distribution of peak storages, $p_j(V)$, is obtained as a convolution between the pdf of the excess storage immediately before jumps and the pdf of the storage increments, which is an exponential distribution with average α. Provided that the jump process is Markovian, the pdf of the excess storage before jumps is equal to the pdf of the excess storage itself, $p(V)$. By solving the convolution (see supporting information), the following expression for the probability distribution of peak storages is obtained:

$$p_j(V) = C_1 \exp\left(-\frac{V}{\alpha}\right) \exp\left[\frac{\lambda(2-a)}{(1-a)[K(2-a)]^{\frac{1}{2-a}}}(V-V_0)^{\frac{1-a}{2-a}}\right] \tag{4.2}$$

where C_1 is a normalization constant, a and K are exponent and coefficient of the power law that defines the decay of q through time, and V_0 represents a minimum or a maximum storage, depending on the storage-discharge relation chosen (see supporting information).

The probability distribution of peakflows (improperly termed $p_j(q)$ for ease of notation) is obtained by applying the derived distribution approach to eq. (4.2) (see supporting information):

$$p_j(q) = C_2 q^{1-a} \exp\left[-\frac{q^{2-a}}{\alpha K(2-a)} + \frac{\lambda q^{1-a}}{K(1-a)}\right] \tag{4.3}$$

where C_2 is a suitable normalization constant. The integration of eq.(4.3) provides the non-exceedance cumulative probability of peakflows, $P_j(q) = \int_0^q p_j(s)ds$.

In order to derive the probability distribution of seasonal maxima, let us assume a constant number m of independent peakflows in a given reference period τ (e.g. one season). Since peakflows are assumed to be independent, the probability that the maximum q among the m peakflows is not exceeded is the product of the probability of each peakflow to have a magnitude lower than q, i.e. $P_j[q_1 \leq q] \cdot ... \cdot P_j[q_m \leq q] = P_j(q)^m$. Provided that the number of peakflows in the considered timespan is distributed according to a Poisson distribution (because effective rainfall events are Poisson distributed), and recalling that $e^z = \sum_{m=0}^{+\infty} \frac{z^m}{m!}$, the following expression

for the non-exceendance cumulative probability of seasonal maxima is obtained:

$$P_{M,s}(q) = \sum_{m=0}^{+\infty} \frac{(\lambda\tau)^m e^{-\lambda\tau}}{m!} P_j(q)^m = \exp\{-\lambda\tau[1 - P_j(q)]\} = \exp[-\lambda\tau D_j(q)] \tag{4.4}$$

where $D_j(q) = 1 - P_j(q)$ represents the duration curve of peakflows. The same result is obtained through a peak over threshold method (i.e. keeping only the peakflows exceeding a suitable threshold).

The recurrence interval T_r can be calculated as the inverse of the exceedance cumulative probability of flow maxima. Therefore, the following analytical expression of the seasonal flood-frequency curve is obtained:

$$T_r(q) = \frac{1}{1 - P_{M,s}(q)} = \frac{1}{1 - \exp[-\lambda\tau D_j(q)]} \tag{4.5}$$

where $\lambda\tau$ represents the (dimensionless) average number of effective rainfall events during the considered season.

4.3 Case studies and parameters estimation

Two river basins characterized by contrasting daily flow regimes have been selected to illustrate the main features of the method. The case studies are the Sand Run (WV, area=37 km^2), which exhibits reduced flow variability typical of a persistent regime during spring and winter seasons, and the Big Eau Plein River (WI, area=580 km^2), displaying an erratic regime with enhanced streamflow variability throughout the year.

The characterization of the seasonal flood-frequency curve relies on specifying the four parameters (α, λ, a, K) of the mechanistic-stochastic model of daily streamflows (Table S4.1). The estimation methods for α, λ and a adopted in this study employ both daily rainfall and flow data, and are detailed in *Basso et al.* [2015a]. The recession coefficient K, instead, is obtained by calibrating the analytic flood-frequency curve on the observed seasonal maxima through maximum likelihood estimate. This is required by the pronounced sensitivity of high flows to antecedent catchment-scale soil moisture conditions (portrayed by the value of K) [*Sivapalan et al.,* 2005], which is further emphasized by focusing on maxima.

4.4 Results

Results of the model application are displayed in Figure 4.2, which also shows confidence intervals of the model predictions (see supporting information for details regarding the uncertainty analysis). Although the performance varies for different catchments and seasons, estimates of the flood-frequency curve obtained through Weibull plotting position of observed maxima (dots) mostly fall in the predicted range of uncertainty. The model performs better for short and intermediate return periods ($T_r < 10$ years), while results for the longest recurrence intervals are poor only in two cases (panels e and g).

The comparison of normalized flood-frequency curves of Sand Run and Big Eau Plein River sheds light on the link between the shape of the flood-frequency curve and the daily flow regime

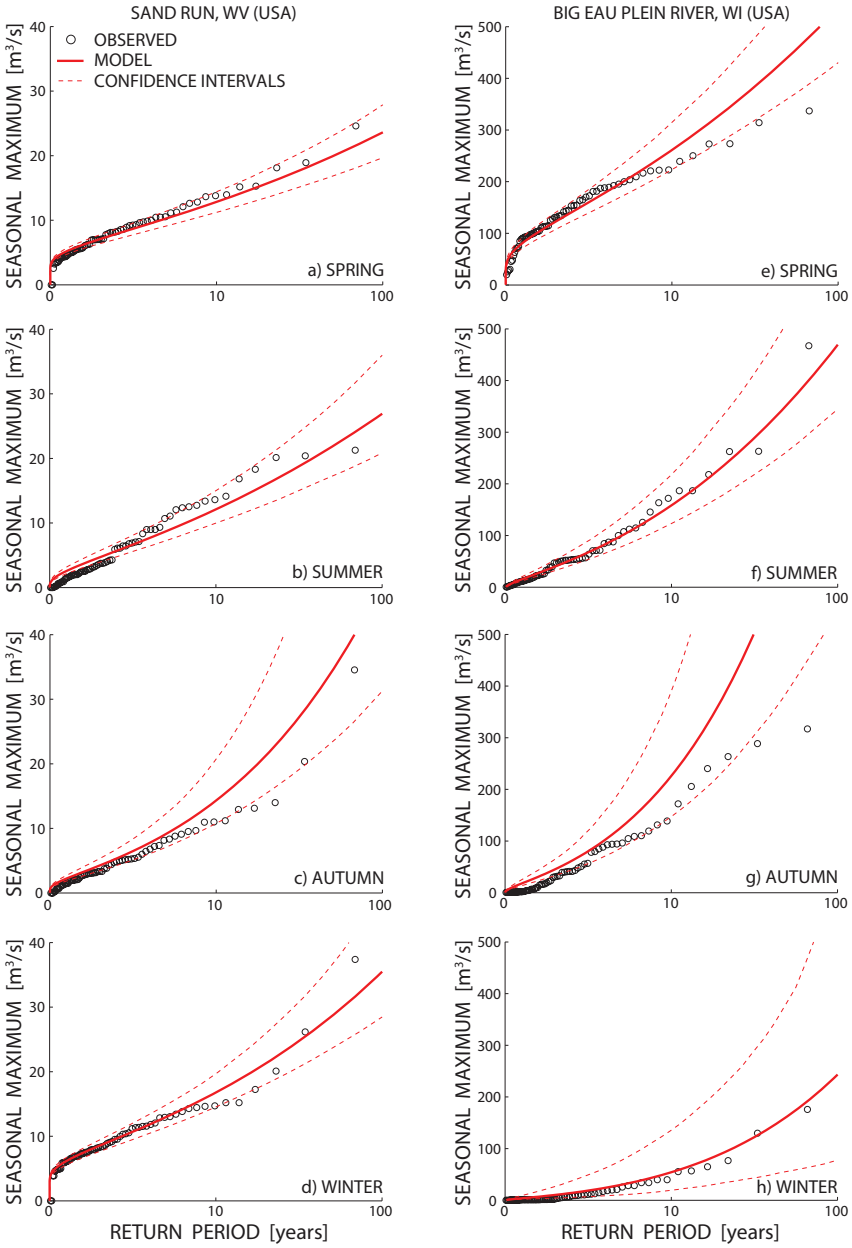

Fig. 4.2: Seasonal flood-frequency curves for the Sand Run (West Virginia) and the Big Eau Plein River (Wisconsin). Curves obtained from the entire series of observed data through Weibull plotting position are represented with dots, while solid red lines display estimates of the proposed analytical model. Confidence intervals of the model estimates are plotted with dotted red lines.

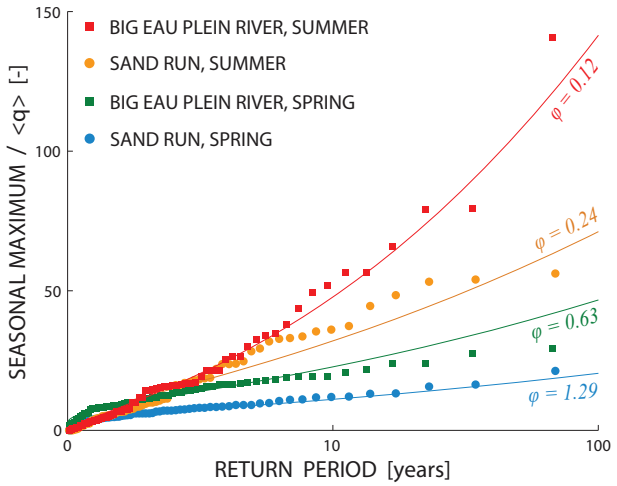

Fig. 4.3: Normalized flood frequency curves (i.e., seasonal maximum divided by the average daily flow, $< q >$) in four case studies characterized by decreasing persistency index ϕ. For the calculation of the persistency index, K has been estimated as in *Basso et al.* [2015a]. Values of the coefficient of variation of daily flows associated to the four case studies are 1.30 (blue), 2.22 (green), 2.95 (yellow) and 4.13 (red). A decrease of the persistency index associated to the daily flow regime results in lower magnitude of events characterized by short return periods and higher magnitude of rare events in erratic regimes.

(see also *Guo et al.* [2014]). Flow regimes are here classified using the persistency index ϕ, which represents the ratio between frequency of flow producing events and mean recession rate [*Basso et al.,* 2015b]. ϕ is inversely proportional to the variability of streamflows (see Figure 4.3 and *Botter et al.* [2013]), and can be expressed as a function of the model parameters as $\phi = \frac{\lambda}{K(\alpha\lambda)^{a-1}}$ (see supporting information).

Figure 4.3 displays observed and modeled flood-frequency curves of spring and summer seasons for the two case studies. Persistent regimes (blue) produce concave flood-frequency curves in a wide range of return periods, with a moderate increase of the magnitude of events associated to longer return periods. On the contrary, for erratic regimes (yellow and red) the flood-frequency curve becomes convex for moderate values of T_r, which implies a sharp increase of the flows associated to long return periods. The increasing erraticity of the flow regime (from blue to red) is associated with a more pronounced change of concavity of flood-frequency curves, with step changes [*Rogger et al.,* 2012] emerging in more erratic regimes. The observed link between the shape of the flood-frequency curve and the underlying persistency index allows to establish a formal relation between the behavior of the extremes and a set of climatic (λ, α) and geomorphic (a, K) features of the catchment [*Doulatyari et al.,* 2015], that are summarized by the value of ϕ.

Figure 4.4 illustrates the reliability of the proposed model with respect to decreased lengths of the available data series. Sliding windows with fixed length of 20, 10 or 5 years are applied to the available data (subdivided by seasons) to generate shorter samples. These samples are used to constrain the model, and its predictions for a set of return periods ranging from 10 to 60 years are compared with the estimates obtained from the entire dataset available through Weibull plotting position. The percent error (ϵ) between model and data based estimates is computed each time. Panels a and b of Figure 4 represent, for each return period, the fraction of cases providing errors lower than 25% (arbitrarily assumed as an acceptable error threshold).

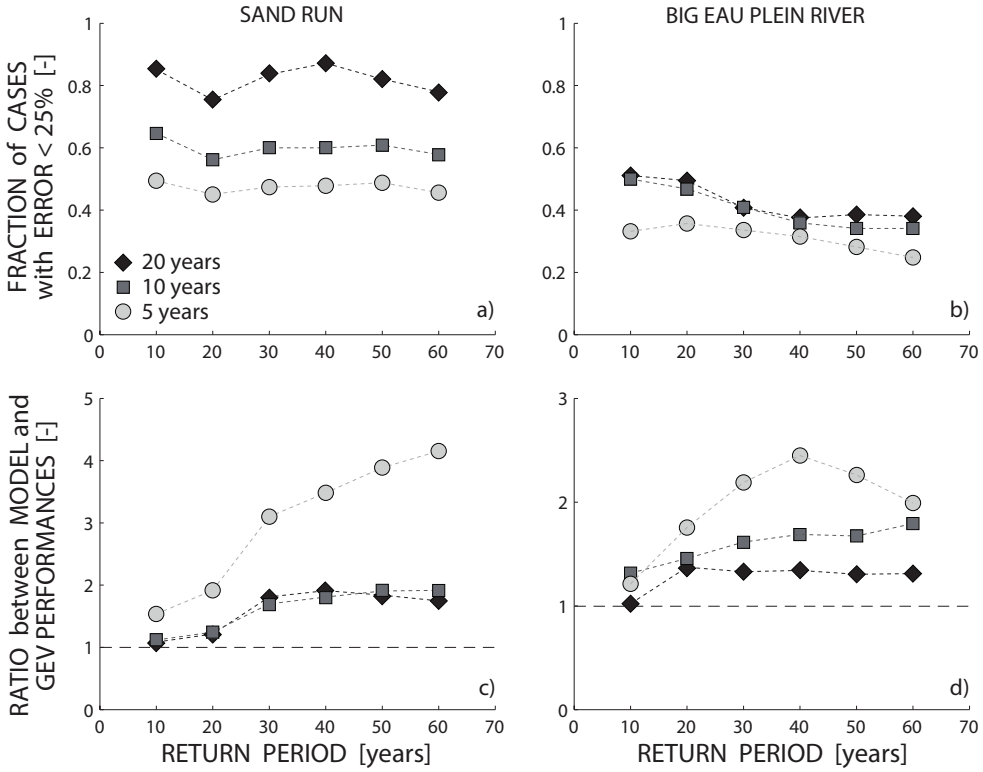

Fig. 4.4: a,b) Performance of the proposed analytical model in predicting flood-frequency curves with short data series. Samples of seasonal data (20, 10 or 5 years) selected from the entire dataset through a moving window are used to estimate analytic seasonal flood-frequency curves. Model results are then compared with the estimate obtained from the complete data series through Weibull plotting position. Errors are computed for a set of return periods (from 10 to 60 years). The plot shows the fraction of cases displaying an error (ϵ) lower than 25%. The shorter the number of years (seasons) of available data, the lower the fraction of cases displaying $\epsilon < 25\%$. The results seem instead quite constant for increasing return periods. The same behavior is found by setting different thresholds for the acceptable error. c,d) Comparison between performances of the proposed analytical model and of generalized extreme value (GEV) distributions calibrated through maximum likelihood on the same samples. Performances are comparable for long data series available (e.g., 20 years) and short return periods (e.g., 10 years). The proposed method exhibits better performances for rare events (high return periods), particularly when only short observed data series (e.g., 5 years) are available to constrain the models.

As expected, the fraction of cases characterized by $\epsilon < 25\%$ diminishes with decreasing length of the sample used for calibration. The fraction of acceptable cases is lower in the erratic than in the persistent regime. This is due to the higher streamflow variability of erratic regimes, which makes them intrinsically unpredictable [*Botter et al.,* 2013]. Interestingly, the performance is quite stable with increasing return periods. This result hints to a structural robustness of the model, whose prediction weakly depends on the specific flood events (and return periods) sampled in the observed data series used for calibration.

In order to better appreciate the significance of these results, panels c and d of Figure 4.4 compare the performance of the proposed approach with that of a generalized extreme value distribution whose parameters are calibrated by maximum likelihood estimate on the same samples used for the calibration of the physically-based approach. The plots evidence systematically higher performances of the proposed method for return periods longer than the sample size. These performances stem from the model's capability to exploit experimental information on the full range of discharges experienced by rivers, as may happen with alternative methods applied in concert with the generalized extreme value distribution (e.g. *Viglione et al.* [2013]).

The proposed framework may prove particularly useful to estimate relatively long return periods (as required e.g. to evaluate long-term sediment delivery [*Kirchner et al.,* 2001] and erosion rates in bedrock rivers [*Lague et al.,* 2005]), or when only short records are available, for example due to temporary or newly installed flow gauges. This is the case for the majority of the catchments worldwide [*Müller and Thompson,* 2015].

4.5 Conclusions

A novel physically-based analytic expression of the seasonal flood-frequency curve is presented. The expression is grounded on a mechanistic-stochastic model of daily streamflow dynamics, whose parameters embody hydroclimatic and landscape attributes of the contributing catchment. Only one parameter related to the antecedent wetness state of the catchment needs to be fitted on observed flow maxima. The application of the method in two rivers featured by contrasting flow regimes (erratic and persistent) is discussed. The model provides reliable estimates of the flows associated to a set of return periods, and capture the diverse shape of flood-frequency curves emerging in erratic and persistent regimes. Model performances do not deteriorate significantly when flows characterized by return times longer than the available sample size are estimated. This represents a structural advantage of the proposed method, that exploits information on the whole range of flows occurring in rivers. The proposed framework is a first step toward the prediction of high flow statistics based on climate and landscape attributes, and may be especially worth in data scarce regions of the world.

Acknowledgments

The US Geological Survey (http://waterdata.usgs.gov) and the National Climatic Data Center (http://cdo.ncdc.noaa.gov/) are acknowledged for providing hydrologic and climatic data. This study was funded by the Swiss National Science Foundation (SNF, Project No. 200021-149126). Additional support was provided by the Competence Center Environment and Sustainability (CCES) of the ETH domain in the framework of the RECORD Catchment project. The first author acknowledges a SNF Mobility Fellowship. The Marani Lab at Duke University, Durham, NC, USA is acknowledged for providing computational resources. Roberto Corvaja and Nicola Pozzobon are gratefully acknowledged for the useful discussions on mathematical derivations.

4.6 Supplementary Information

4.6.1 Derivation of the probability distribution of peak storages, $p_j(V)$ (eq. (4.2)).

The probability density function (pdf) of peak storages, $p_j(V)$, is derived by convoluting the pdf of the (specific) excess storage immediately before jumps and the pdf of the (specific) storage increments, $f(x)$, which is an exponential distribution with average α [*Botter et al.,* 2009]. Provided that the jump process is markovian, the pdf of the excess storage immediately before jumps is equal to the pdf of the excess storage, $p(V)$, which can be derived from the solution of the master equation associated to eq.(4.1) in the main text [*Botter et al.,* 2009]:

$$p(V) = C\frac{1}{\rho(V)}\exp\left[-\frac{V}{\alpha} + \lambda I(V)\right] \tag{S4.1}$$

where C is a normalization constant, ρ is a monotonous non linear storage-discharge relation which describes how the excess water is drained from the catchment, and I is the primitive of its reciprocal. The storage-discharge relation ρ is typically assumed as a power law (see e.g. eq.(15) of *Botter et al.* [2009]):

$$V = V_0 + \hat{a}q^{\hat{b}} \tag{S4.2}$$

that, expressing the streamflow q as the dependent variable, becomes (see eq.(14) of *Botter et al.* [2009]):

$$q = \rho(V) = \left(\frac{1}{\hat{a}}\right)^{\frac{1}{\hat{b}}}(V - V_0)^{\frac{1}{\hat{b}}} = \hat{K}(V - V_0)^{\hat{n}} \tag{S4.3}$$

where $\hat{K} = \left(\frac{1}{\hat{a}}\right)^{\frac{1}{\hat{b}}}$ and $\hat{n} = \frac{1}{\hat{b}}$. These expressions are valid for both of the two alternative models (termed convex and hyperbolic) frequently used in the literature [*Kirchner,* 2009].

It is worth clarifying the relation between parameters $\hat{a}$, $\hat{b}$, $\hat{K}$, $\hat{n}$ in equations (S4.2) and (S4.3) and the parameters K and a in eq.(4.2) (termed recession coefficient and exponent in the following). K and a can be estimated from recession analysis and characterize the decay of discharge in between flow-producing events ($\frac{dq}{dt} = -Kq^a$). The relation between the parameters, obtained by studying the mass conservation equation during recession events (i.e. setting $\xi(t) = 0$ in eq.(4.1)), is:

$$a = 2 - \hat{b} = 2 - \frac{1}{\hat{n}} \tag{S4.4}$$

$$K = \frac{1}{\hat{a}\hat{b}} = \hat{n}\hat{K}^{\frac{1}{\hat{n}}} \tag{S4.5}$$

Accordingly, the storage-discharge relation ρ (eq. (S4.3)) can be expressed as a function of a and K as:

$$\rho(V) = \hat{K}(V - V_0)^{\hat{n}} = [K(2-a)]^{\frac{1}{2-a}}(V - V_0)^{\frac{1}{2-a}} \tag{S4.6}$$

which leads to the following expression for $I(V)$ in eq.(S4.1):

$$I(V) = \frac{a-2}{a-1}[K(2-a)]^{\frac{1}{a-2}}(V - V_0)^{\frac{a-1}{a-2}} \tag{S4.7}$$

The convolution between the pdf of the excess storage $p(V)$ and the pdf of the storage increments $f(x)$, which provides the probability distribution of peak storages $p_j(V)$, can thus be written as:

$$p_j(V) = \int_A^B p(x) f(V-x) dx \tag{S4.8a}$$

$$= \int_A^B C \frac{1}{\rho(x)} \exp\left[-\frac{x}{\alpha} + \lambda I(x)\right] \frac{1}{\alpha} \exp\left(-\frac{V-x}{\alpha}\right) dx \tag{S4.8b}$$

$$= \frac{C}{\alpha} \exp\left(-\frac{V}{\alpha}\right) \int_A^B \frac{(x-V_0)^{-\frac{1}{2-a}}}{[K(2-a)]^{\frac{1}{2-a}}} \exp\left[\frac{\lambda(2-a)(x-V_0)^{\frac{1-a}{2-a}}}{(1-a)[K(2-a)]^{\frac{1}{2-a}}}\right] dx \tag{S4.8c}$$

where A and B are suitable integration boundaries (see below). Performing the following change of variable:

$$u = \frac{\lambda(2-a)(x-V_0)^{\frac{1-a}{2-a}}}{(1-a)[K(2-a)]^{\frac{1}{2-a}}} \tag{S4.9}$$

we obtain:

$$p_j(V) = \frac{C}{\alpha\lambda} \exp\left(-\frac{V}{\alpha}\right) \int_{\hat{A}}^{\hat{B}} \exp(u) du \tag{S4.10}$$

In order to solve eq.(S4.10), the lower and upper boundaries of the integral in eq. (S4.8a) must be specified. Both for the convex and hyperbolic models, the magnitude of the storage increments ranges between 0 and $+\infty$.

For the convex model, the excess storage can vary between V_0 (minimum storage) and $+\infty$. Therefore, the lower and upper boundaries of the integral in eq.(S4.8a) are $A = V_0$ and $B = V$. Expressing them in terms of the new variable u we obtain:

$$\hat{B} = \frac{\lambda(2-a)}{(1-a)[K(2-a)]^{\frac{1}{2-a}}} (V-V_0)^{\frac{1-a}{2-a}} \tag{S4.11}$$

and

$$\hat{A} = -\infty \tag{S4.12}$$

Solving eq.(S4.10) with the integration boundaries expressed by equations (S4.11) and (S4.12), the probability distribution of peak storages for the convex model is obtained:

$$p_j(V) = \frac{C}{\alpha\lambda} \exp\left(-\frac{V}{\alpha}\right) \exp\left[\frac{\lambda(2-a)}{(1-a)[K(2-a)]^{\frac{1}{2-a}}} (V-V_0)^{\frac{1-a}{2-a}}\right] \tag{S4.13}$$

In the hyperbolic model the excess storage can instead vary between $-\infty$ and V_0 (maximum storage), and the lower and upper boundaries of integral (S4.8a) becomes $A = -\infty$ and $B = V$. Expressing them in terms of the new variable u, we obtain:

$$\hat{B} = \frac{\lambda(2-a)}{(1-a)[K(2-a)]^{\frac{1}{2-a}}} (V-V_0)^{\frac{1-a}{2-a}} \tag{S4.14}$$

and

$$\hat{A} = -\infty \tag{S4.15}$$

Solving eq.(S4.10) with the integration boundaries expressed by equations (S4.14) and (S4.15), the probability distribution of peak storages for the hyperbolic model is obtained:

$$p_j(V) = \frac{C}{\alpha\lambda} \exp\left(-\frac{V}{\alpha}\right) \exp\left[\frac{\lambda(2-a)}{(1-a)[K(2-a)]^{\frac{1}{2-a}}}(V-V_0)^{\frac{1-a}{2-a}}\right] \tag{S4.16}$$

which is the same as eq.(S4.13). Hence, the same solution (eq.(4.2) of the main text) holds for both the convex and hyperbolic models.

4.6.2 Derivation of the probability distribution of peakflows, $p_j(q)$ (eq. (4.3)).

The probability distribution of peakflows is obtained from the probability distribution of peak storages (equation (S4.13) or (S4.16)) by applying the derived distribution rule:

$$p_j(q) = p_j(V(q))\frac{dV}{dq} \tag{S4.17}$$

where the expression of V as a function of q is obtained by combining equations (S4.2), (S4.4) and (S4.5):

$$V = V_0 + \frac{q^{2-a}}{K(2-a)} \tag{S4.18}$$

Solving eq.(S4.17) we finally obtain the probability distribution of peakflows:

$$p_j(q) = \frac{C}{\alpha K\lambda} \exp\left(-\frac{V_0}{\alpha}\right) \cdot q^{1-a} \cdot \exp\left[-\frac{q^{2-a}}{\alpha K(2-a)} + \frac{\lambda q^{1-a}}{K(1-a)}\right] \tag{S4.19}$$

which is the same as eq.(4.3) in the main text.

4.6.3 Uncertainty of the physically-based analytic expression of the seasonal flood-frequency curve

In order to assess the uncertainty of the proposed physically-based model for seasonal flood-frequency curves, information on the uncertainty involved in the estimation of each model parameter is needed. For this purpose, the probability distributions of the parameters α, λ, a, K involved in eq.(4.5) are calculated.

Mean rainfall depth α. α is here estimated as the mean rainfall depth during wet days, identified from the sample of daily rainfall data recorded in a reference rainfall station belonging to the contributing catchment. The distribution of the sample mean of a stochastic variable (in this case the rainfall depth h) is Gaussian, with mean:

$$\mu_\alpha = \bar{h} = \frac{1}{N}\sum_{i=1}^{N} h_i \tag{S4.20}$$

and standard deviation:

$$\sigma_\alpha = \sqrt{\frac{\frac{1}{N-1}\sum_{i=1}^{N}(h_i-\bar{h})^2}{N}} \tag{S4.21}$$

where N is the sample size (in this case the number of observed rainfall events).

Effective rainfall frequency λ. The effective rainfall frequency λ is estimated as the ratio between the sample mean of streamflows, $\bar{q}$, and the sample mean of rainfall depths, $\bar{h}$ ($\lambda = \bar{q}/\bar{h}$) [*Basso et al.,* 2015a]. Recalling that eq. (S4.20) and (S4.21) also hold for the sample mean of streamflows, and that the distribution of the ratio between sample means of independent stochastic variables is Gaussian, the mean of the distribution of λ is [*Schafer,* 1997]:

$$\mu_\lambda = \frac{\mu_{\bar{q}}}{\mu_\alpha} \tag{S4.22}$$

and its standard deviation is [*Schafer,* 1997]:

$$\sigma_\lambda = \sqrt{\sigma_\alpha^2 \left(\frac{\mu_{\bar{q}}}{\mu_\alpha^2}\right)^2 + \sigma_{\bar{q}}^2 \left(\frac{1}{\mu_\alpha^2}\right)} \tag{S4.23}$$

In the above derivation $\bar{q}$ and $\bar{h}$ are assumed to be independent variables. Notice that this is a precautionary assumption, since the variance of λ would be lower if the existing correlation between $\bar{q}$ and $\bar{h}$ were taken into account.

Recession exponent a. The recession exponent is calculated by fitting a power-law curve ($\frac{dq}{dt} = -Kq^a$) to pairs $dq/dt - q$ of each observed recession curve. The median values among all the estimates is then selected. The distribution of the sample median a of the stochastic variable recession exponent is Gaussian, with mean μ_a equal to the median of the set of recession exponents estimated from the observed recession curves, and standard deviation [*Schafer,* 1997]:

$$\sigma_a = \frac{1}{4Nf(\mu_a)^2} \tag{S4.24}$$

where N is the sample size (in this case the number of recessions considered) and $f(\mu_a)$ is the probability density function of the recession exponent evaluated in μ_a. A Gaussian distribution fitted to the empirical distribution obtained from the set of observed recession exponents has been used to calculate $f(\mu_a)$ in eq.(S4.24).

Recession coefficient K. K is fitted on the observed seasonal maxima using a maximum likelihood technique, and thus its posterior pdf is asymptotically normal with variance equal to the reciprocal of the Fisher information number. The distribution of K is here obtained through a Monte Carlo approach, based on the assumption that α, λ and a are mutually independent variables. The procedure is specified in the following:

- random values of the parameters α, λ and a are extracted from their corresponding Gaussian distributions, while K is uniformly sampled in a pre-defined range;
- for each set of α, λ and a and for each value of K, the physically-based analytical distribution of seasonal maxima is computed together with the likelihood of the considered set of parameters. The likelihood of the physically-based analytical distribution is:

$$l(\Theta) = \prod_{i=1}^{N} p_M(x_i|\Theta) \tag{S4.25}$$

  where $p_M(x_i|\Theta)$ expresses the likelihood of the parameters set $\Theta = (\alpha, \lambda, a, K)$ with respect to the observed seasonal maxima x_i, p_M is the probability density function of seasonal maxima (i.e. the derivative of eq. (4.4) in the main text) and N is the number of observed seasonal maxima;

- each K value is now associated with a set of likelihoods corresponding to different (randomly extracted) values of the other model parameters (α, λ, a);
- in order to compute the pdf of K conditional to the values assumed by the other parameters (α, λ, a), their values are divided in three ranges. As a result, 3^3 conditional pdfs of K are obtained;
- the marginal likelihood of K is computed as the average of the set of likelihoods associated to each K value;
- since the prior distribution of K is uniform, its posterior probability distribution is proportional to the marginal likelihood of K.

The uncertainty of the physically-based analytical distribution of the seasonal maxima is then computed through the following procedure:

- random values of α, λ, a and K are extracted from their corresponding Gaussian distributions;
- the physically-based analytical distribution is computed with the extracted set of parameters, and return times for a range of flow values are evaluated from the distribution;
- by reiterating the previous steps, different return times (arising from different sets of model parameters) are obtained for each flow value;
- finally, the 0.05 and 0.95 quantiles of the set of return times associated to each flow value are computed, and the confidence intervals of the flood-frequency curve are drawn.

4.6.4 Derivation of the persistency index ϕ

River flow regimes can be classified based on the underlying variability of daily flows, as proposed by *Botter et al.* [2013]. If the drainage of water from the contributing catchment is described through a linear storage-discharge relation, the coefficient of variation of streamflows is $CV_q = \sqrt{k/\lambda}$, where λ is the frequency of the effective rainfall and k is the (constant) hydrograph recession rate (see *Botter et al.* [2013] for details). This allows to define a persistency index λ/k [*Basso et al.,* 2015b], which can be used to distinguish persistent regimes with reduced flow variability ($\lambda/k > 1$) from erratic regimes with enhanced flow variability ($\lambda/k < 1$).

When the storage-discharge relation is non linear (as it is the case for the model considered in this study), the hydrograph recession rate varies during the recession period, and a characteristic value must be chosen to define an analogous persistency index. With this goal in mind, let us rewrite the equation for the hydrograph's recession as [*Kirchner,* 2009]:

$$\frac{dq}{dt} = -Kq^a = -Kq^{a-1}q = -g(q) \cdot q \tag{S4.26}$$

where $g(q) = Kq^{a-1}$ is the time-variant recession rate.

An effective characteristic value of the hydrograph recession rate is represented by the average of $g(q)$:

$$< g(q) >= \int_0^\infty g(q)p(q)dq = K \int_0^\infty q^{a-1}p(q)dq \sim K < q >^{a-1} = K(\alpha\lambda)^{a-1} \tag{S4.27}$$

	Sand Run				Big Eau Plein River			
Season	α	λ	a	K	α	λ	a	K
Spring	0.72	0.38	1.75	0.48	0.70	0.25	1.67	0.76
Summer	0.95	0.09	1.57	0.66	1.00	0.05	1.64	0.76
Autumn	0.76	0.12	1.84	0.90	0.79	0.07	1.82	1.15
Winter	0.56	0.51	1.89	0.66	0.31	0.05	1.81	2.29

Tab. S4.1: Estimates of mean rainfall depth ($\alpha[cm]$), frequency of effective rainfall ($\lambda[d^{-1}]$), recession exponent ($a[-]$) and coefficient ($K[cm^{1-a}/d^{2-a}]$) for the study catchments.

Notice that the above approximation holds for values of a around 2 (which is a quite common circumstance, see *Doulatyari et al.* [2015]).

The persistency index ϕ for the non linear case is then defined as:

$$\phi = \frac{\lambda}{< g(q) >} = \frac{\lambda}{K(\alpha\lambda)^{a-1}} \tag{S4.28}$$

Investigations carried out in a sample of catchments reveal that the linear and non linear formulations of the persistency index typically provide consistent results.

5. CONCLUSIONS AND PERSPECTIVES

The main results described in the previous chapters are summarized in this section, and potential future areas of research are presented.

5.1 Conclusions

The chief goal of this thesis was to deepen the understanding of the driving role of streamflow regimes for sediment transport and flooding potential of river catchments, which requires an adequate characterization of high flow statistics.

The variety of effective discharges observed for rivers of different climatic areas was explained by the underlying heterogeneity of flow regimes, characterized by the coefficient of variation of daily flows. The ratio between effective discharge and mean streamflow (effective ratio) was found to depend on the frequency of streamflow-producing rainfall events, the mean catchment response time and the exponent of the sediment rating curve. An analytic expression to quantitatively described these dependencies was provided, and successfully tested in a set of 18 case studies in the continental US. Larger effective ratios arise with more erratic flow regimes (high flow variability) and larger exponents of the sediment rating curve, while persistent regimes (weak flow variability) display effective discharge not much higher than the average flow. This is the outcome of distinct streamflow dynamics in the two types of flow regime. The established link between effective ratio and flow regimes and the analytic tool provided may be valuable for preliminary estimates of the effective discharge for sediment transport.

In order to improve the characterization of high flows, especially important for e.g. bedload sediment transport, different versions of a stochastic analytical model for flow duration curves have been compared. The analysis highlighted the importance of studying recession patterns for single events, and revealed the pivotal role of a non-linear storage-discharge relation to reproduce observed frequencies of high flows. Moreover, the emergence of heavy-tailed streamflow distributions was linked to the degree of non-linearity of the catchment hydrologic response, regardless of the underlying rainfall regime.

These advances was the basis to derive a novel physically-based analytic expression of seasonal flood-frequency curves, whose parameters embody climatic and landscape attributes of the contributing catchment. The framework unveiled the diverse shape of flood-frequency curves associated to erratic and persistent regimes, which was confirmed by flow records from two rivers adopted as case studies. The model provided reliable estimates of the flows associated to a set of return periods, with almost steady performances for return times longer than the available sample size. These results stem from the model's capability of efficiently using the information content of the whole range of flows experienced by rivers, which is obtained by constraining three model parameters using daily rainfall and streamflow data. The proposed framework open the avenue for a prediction of high flow statistics based on climate and landscape attributes, and may be especially valuable where data scarcity is an issue.

5.2 Outlook

The novel modeling tools developed and presented in this thesis proved effective, and allowed for satisfactory results to be achieved in different case studies encompassing a variety of climatic settings. As a result, research goals set at the beginning of this work and discussed in the previous chapters have been achieved. Nevertheless, unanswered research questions exist that may be addressed by further investigations and by improving the proposed methods. Possible directions for the continuation of this research are discussed in the following.

Advances on the characterization of high flow statistics achieved in Chapter 3 and 4 may be applied to improve sediment transport and effective discharge estimates, especially for bedload, where only high flows exceeding a suited threshold are capable of triggering the transport. Classic bedload transport formulas (e.g. *Meyer-Peter and Müller* [1948]) may be integrated in the proposed framework, and the patterns of effective ratio against flow variability studied when these experimental relations are considered instead of the empirical sediment rating curve.

The impacts of heavy tailed flow distributions on the probability density function of sediment flows (and on its statistical moments) may be also examined, and the emergence of diverse tails of the distribution of sediment flows linked to the availability of sediments ready to be transported in the catchment. In fact, the actual variability of the coefficient β of the sediment rating curve (which is assumed constant in the current modeling framework) may be interpreted as an expression of sediment supply limitations, which may occur on an event base or during the considered time period (e.g. seasonally). Its effect on the estimated effective discharge values may be studied.

The variability of β may be integrated in the model by considering its probability distribution. However, this change would probably prevent the derivation of closed-form expressions for the probability distribution of sediment flows and the effective ratio. Aside, the link between the variability of β and the features of the main sediment source in a catchment (e.g. localized or uniform supply, connectivity between sources and river channels) may be investigated.

Coupling the framework proposed in Chapter 2 with studies that explain the parameters of the sediment rating curve in terms of physical attributes of the contributing catchments (see e.g. *Syvitski et al.* [2000]) would finally provide a method to estimate the effective discharge from soft data (i.e. without using observations of sediment flow), thus broadening the applicability of the methods.

Another major area of study regards the influence of catchment wetness conditions for the emergence of high flows. The wetness state of a watershed is summarized in the proposed framework by the recession coefficient K, as previously suggested by *Shaw et al.* [2013] and *Biswal and Kumar* [2014]. Results shown in Chapter 3 reveal the efficacy of constant K values for the estimation of high flows in most of the catchments studied, but also highlight the need to account for the correlation between wetness conditions and resulting high flows in specific cases. This correlation and its effects seem to be particularly strong in certain settings. Climatic and morphological attributes leading to a pivotal role of the instantaneous wetness state for the estimate of high flow statistics may be identified.

The analyses presented in Chapter 4 further emphasize the relevance of time-specific soil moisture deficits for seasonal maxima, as expressed by the need of calibrating K on the observed series of extremes. The variability of the recession coefficients estimated by separately analyzing

hydrograph recession curves, and the uncertainty of the calibrated K values, seem to differ in rivers featuring contrasting flow regimes. The relation between effective values of the recession coefficient and type of flow regime may therefore be investigated. This step would provide a method to estimate the recession coefficient based on measurable attributes of the contributing catchment (the same used to characterize the flow regime).

Alternatively, other proxies for the catchment wetness state needs to be found in order to estimate the recession coefficient in absence of flow records. New generations of satellite data, which provide large-scale snapshots of soil moisture content, are particularly apt for this purpose [*Brown et al.,* 2013]. Long-term statistics derived from these data may indeed provide relevant metrics (e.g. spatial average and variance) of catchment wetness at the appropriate scale, and their value may be linked to the likelihood of soil storage deficits significant for runoff generation.

Coupling a method for the estimation of K which does not require series of observed flow maxima with a previously presented framework for the assessment of flow duration curves, which is based on the same mechanistic-stochastic model used here [*Doulatyari et al.,* 2015], would finally open the way for a physically-based characterization of flood statistics in absence of discharge data.

Extensive application of the model for flood-frequency curves proposed in Chapter 4 constitutes another important line of future research. This is envisioned in order to assess its performance for a wider set of case studies, at the same time advancing the comprehension of the controls of flow regimes on catchment flooding potential and shape of the flood-frequency curve.

An investigation of the relevance of snow melting and other snow related processes (e.g. rain-on-snow) in the generation of extreme discharges may suggest to extend the model to regions where flows are affected by snow dynamics. This development may be particularly challenging, because of the need to consider in a coupled way the stochasticity of both rainfall and air temperature, and the additional water input coming from the snow stored in the catchment. In some cases, treating the snow accumulation and subsequent melting as a delayed hydrologic response, whose parameters can be estimated by additionally analyzing the statistics of the temperature signal (see e.g. *Schaefli et al.* [2013]), may provide a solution.

Finally, analysis not included in this thesis reveal the emergence of tails of flow distributions heavier than the power law for particular combinations of the model parameters. These special cases may be further investigated, in view of their relevance for flooding hazard assessment.

Eventually, tools developed and tested in this thesis may be integrated in a modeling framework (geo-database and WebGis platform) recently proposed by *Doulatyari et al.* [under review], which is specifically designed for data scarce regions of the world. Integration of the methods here described would further increase the appeal of this platform for practical applications.

BIBLIOGRAPHY

Andrews, E. D. (1980), Effective and bankfull discharges of streams in the Yampa River basin, Colorado and Wyoming, *J. Hydrol.*, 46(3-4), 311-330, doi:10.1016/0022-1694(80)90084-0.

Basso, S., and G. Botter (2012), Streamflow variability and optimal capacity of run-of-river hydropower plants, *Water Resour. Res.*, 48, W10527, doi:10.1029/2012WR012017.

Basso, S., M. Schirmer, and G. Botter (2015a), On the emergence of heavy-tailed streamflow distributions, *Adv. Water. Resour.*, 82, 98-105, doi:10.1016/j.advwatres.2015.04.013.

Basso, S., A. Frascati, M. Marani, M. Schirmer, and G. Botter (2015b), Climatic and landscape controls on effective discharge, *Geophys. Res. Lett.*, 42, 84418447, doi:10.1002/2015GL066014.

Basu, N. B., G. Destouni, J. W. Jawitz, S. E. Thompson, N. V. Loukinova, A. Darracq, S. Zanardo, M. Yaeger, M. Sivapalan, A. Rinaldo, and P. S. C. Rao (2010), Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity, *Geophys. Res. Lett.*, 37, L23404, doi:10.1029/2010GL045168.

Biswal, B., and D. N. Kumar (2014), Study of dynamic behaviour of recession curves, *Hydrol. Process.*, 28, 784792, doi:10.1002/hyp.9604.

Biswal, B., and M. Marani (2010), Geomorphological origin of recession curves, *Geophys. Res. Lett.*, 37, L24403, doi:10.1029/2010GL045415.

Bolla Pittaluga, M., R. Luchi, and G. Seminara (2014), On the equilibrium profile of river beds, *J. Geophys. Res.*, 119, 317332, doi:10.1002/2013JF002806.

Botter, G., A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2007a), Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: slow, leaching-prone components of the hydrologic response, *Water Resour. Res.*, 43, W02417, doi: 10.1029/2006WR005043.

Botter, G., A. Porporato, E. Daly, I. RodriguezIturbe, and A. Rinaldo (2007b), Probabilistic characterization of base flows in river basins: roles of soil, vegetation, and geomorphology, *Water Resour. Res.*, 43, W06404, doi:10.1029/2006WR005397.

Botter, G., F. Peratoner, A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2007c), Signatures of large-scale soil moisture dynamics on streamflow statistics across U.S. climate regimes, *Water Resour. Res.,* 43, W11413, doi: 10.1029/2007WR006162.

Botter, G., A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2009), Nonlinear storage-discharge relations and catchment streamflow regimes, *Water Resour. Res.*, 45, W10427, doi:10.1029/2008WR007658.

Botter, G. (2010), Stochastic recession rates and the probabilistic structure of stream flows, *Water Resour. Res.*, 46, W12527, doi:10.1029/2010WR009217.

Botter, G., S. Basso, A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2010), Natural stream-flow regime alterations: Damming of the Piave river basin (Italy), *Water Resour. Res.*, 46, W06522, doi:10.1029/2009WR008523.

Botter, G., S. Basso, I. Rodriguez-Iturbe, and A. Rinaldo (2013), Resilience of river flow regimes, *PNAS*, 110, 32, 12925-12930, doi:10.1073/pnas.1311920110

Bowers, M. C., W. W. Tung, and J. B. Gao (2012), On the distributions of seasonal river flows: Lognormal or power law?, *Water Resour. Res.*, 48, W05536, doi:10.1029/2011WR011308.

Brown, M. E., V. Escobar, S. Moran, D. Entekhabi, P. E. O'Neill, E. G. Njoku, B. Doorn, and J. K. Entin (2013), NASA's Soil Moisture Active Passive (SMAP) mission and opportunities for applications users, *B. Am. Meteorol. Soc.*, 94(8), 1125-1128, doi: 10.1175/BAMS-D-11-00049.1.

Brutsaert, W., and J.L. Nieber (1977), Regionalized drought flow hydrographs from a mature glaciated plateau, *Water Resour. Res.*, 13(3), 637-643, doi:10.1029/WR013i003p00637.

Bunte, K., S. R. Abt, K. W. Swingle, and D. A. Cenderelli (2014), Effective discharge in Rocky Mountain headwater streams, *J. Hydrol.*, 519(B), 21362147, doi:10.1016/j.jhydrol.2014.09.080.

Burnham, K. P., and D. R. Anderson (2002), Model selection and multimodel inference, *Springer*, USA.

Camporeale, C., and L. Ridolfi (2006), Riparian vegetation distribution induced by river flow variability: a stochastic approach, *Water Resour. Res.*, 42, W10415, doi:10.1029/2006WR004933.

Carreau, J., P. Naveau, and E. Sauquet (2009), A statistical rainfall-runoff mixture model with heavy-tailed components, *Water Resour. Res.*, 45, W10437, doi:10.1029/2009WR007880.

Castellarin, A., G. Galeati, L. Brandimarte, A. Montanari, and A. Brath (2004), Regional flow-duration curves: reliability for ungauged basins, *Adv. Water Res.*, 27, 953-965, doi:10.1016/j.advwatres.2004.08.005.

Castellarin, A., G. Camorani, and A. Brath (2007), Predicting annual and long-term flow-duration curves in ungauged basins, *Adv. Water Res.*, 30, 937-953, doi:10.1016/j.advwatres.2006.08.006.

Castellarin, A., S. Kohnova, L. Gaal, A. Fleig, J. L. Salinas, A. Toumazis, T. R. Kjeldsen, and N. Macdonald (2012), Review of applied-statistical methods for flood-frequency analysis in Europe, *NERC Centre for Ecology and Hydrology*, UK.

Ceola, S., G. Botter, E. Bertuzzo, A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2010), Comparative study of ecohydrological streamflow probability distributions, *Water Resour. Res.*, 46, W09502, doi:10.1029/2010WR009102.

Ceola, S., I. Hödl, M. Adlboller, G. Singer, E. Bertuzzo, L. Mari, G. Botter, J. Waringer, T. J. Battin, and A. Rinaldo (2013), Hydrologic Variability Affects Invertebrate Grazing on Phototrophic Biofilms in Stream Microcosms, *PLoS ONE*, 8(4), e60629, doi:10.1371/journal.pone.0060629.

Claps, P., and F. Laio (2003), Can continuous streamflow data support flood frequency analysis? An alternative to the partial duration series approach, *Water Resour. Res.*, 39(8), 1216, doi:10.1029/2002WR001868.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009), Power-Law Distributions in Empirical Data, *SIAM Rev.*, 51(4), 661703, doi:10.1137/070710111.

Coles, S. (2001), An introduction to statistical modeling of extreme values, *Springer-Verlag*, London, UK.

Cudeck, R., and M. W. Browne (1983), Cross-validation of covariance structures, *Multivar. Behav. Res.*, 18, 147-167, doi:10.1207/s15327906mbr1802-2.

Doulatyari, B., S. Basso, M. Schirmer, and G. Botter (2014), River flow regimes and vegetation dynamics along a river transect, *Adv. Water Res.*, 73, 30-43, doi:10.1016/j.advwatres.2014.06.015.

Doulatyari, B., A. Betterle, S. Basso, B. Biswal, M. Schirmer, and G. Botter (2015), Predicting streamflow distributions and flow duration curves from landscape and climate, *Adv. Water. Resour.*, 83, 285-298, doi:10.1016/j.advwatres.2015.06.013.

Doulatyari, B., A. Betterle, D. Radny, E. A. Celegon, P. Fanton, M. Schirmer, and G. Botter, Patterns of streamflow regimes along the river network: the case of the Thur river, *Environ. Modell. Softw.*, Under Review.

Doyle, M. W., D. Shields, K. F. Boyd, P. B. Skidmore, and D. Dominick (2007), Channel-Forming Discharge Selection in River Restoration Design, *J. Hydraul. Eng.*, 133(7), 831837, doi:10.1061/(ASCE)0733-9429(2007)133:7(831).

Doyle, M. W., and C. A. Shields (2008), An alternative measure of discharge effectiveness, *Earth Surf. Process. Landforms*, 33, 308316, doi:10.1002/esp.1543.

Emmett, W. W., and M. G. Wolman (2001), Effective discharge and gravel-bed rivers, *Earth Surf. Process. Landforms*, 26, 13691380, doi: 10.1002/esp.303.

Fisher, R. A., and L. H. C. Tippett (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Math. Proc. Cambridge*, 24(2), 180-190, doi:10.1017/S0305004100015681.

Frascati A., and S. Lanzoni (2009), Morphodynamic regime and long-term evolution of meandering rivers, *J. Geophys. Res.*, 114, F02002, doi:10.1029/2008JF001101.

Gaume, E. (2006), On the asymptotic behavior of flood peak distributions, *Hydrol. Earth Syst. Sci.*, 10, 233-243, doi:10.5194/hess-10-233-2006, 2006.

Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une serie aleatoire, *Ann. Math.*, 44(3), 423-453, doi:10.2307/1968974.

Godsey S.E., and Kirchner J.W. (2014), Dynamic, discontinuous stream networks: hydrologically driven variations in active drainage density, flowing channels and stream order, *Hydrol. Process.*, 28, 57915803, doi: 10.1002/hyp.10310.

Goodwin, P. (2004), Analytical Solutions for Estimating Effective Discharge, *J. Hydraul. Eng.*, 130(8), 729738, doi:10.1061/(ASCE)0733-9429(2004)130:8(729).

Grimaldi, S., and F. Serinaldi (2006), Asymmetric copula in multivariate flood frequency analysis, *Adv. Water. Resour.*, 29(8), 11551167, doi:10.1016/j.advwatres.2005.09.005.

Guo, J., H.-Y. Li, L. R. Leung, S. Guo, P. Liu, and M. Sivapalan (2014), Links between flood frequency and annual water balance behaviors: A basis for similarity and regionalization, *Water Resour. Res.*, 50, 937953, doi:10.1002/2013WR014374.

Harman, C. J., M. Sivapalan, and P. Kumar (2009), Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics, *Water Resour. Res.*, 45, W09404, doi:10.1029/2008WR007392.

Horn, D., and M. McShane (2013), Flooding the market, *Nat. Clim. Change*, 3, 945947, doi:10.1038/nclimate2025.

Hrachowitz, M., H. H. G. Savenije, G. Blschl, J.J. McDonnell, M. Sivapalan, J. W. Pomeroy, B. Arheimer, T. Blume, M. P. Clark, U. Ehret, F. Fenicia, J. E. Freer, A. Gelfan, H. V. Gupta, D. A. Hughes, R. W. Hut, A. Montanari, S. Pande, D. Tetzlaff, P. A. Trocho, S. Uhlenbrook, T. Wagener, H. C. Winsemius, R.A. Woods, E. Zehe, and C. Cudennec (2013), A decade of predictions in ungauged basins (PUB) - a review, *Hydrol. Sci. J.*, 58(6), 1198-1255, doi:10.1080/02626667.2013.803183.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, 25, 1287-1304, doi:10.1016/S0309-1708(02)00056-8.

Kirchner, J. W., R. C. Finkel, C. S. Riebe, D. E. Granger, J. L. Clayton, J. G. King, and W. F. Megahan (2001), Mountain erosion over 10 yr, 10 k.y., and 10 m.y. time scales, *Geology*, 29(7), 591594, doi:10.1130/0091-7613(2001)029¡0591:MEOYKY¿2.0.CO;2.

Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, 45, W02429, doi:10.1029/2008WR006912.

Lague, D., N. Hovius, and P. Davy (2005), Discharge, discharge variability, and the bedrock channel profile, *J. Geophys. Res.*, 110, F04006, doi:10.1029/2004JF000259.

Laio, F., A. Porporato, L. Ridolfi, and I. Rodriguez-Iturbe (2001), Plants in water-controlled ecosystems: Active role in hydrologic processes and response to water stress. II. Probabilistic soil moisture dynamics, *Adv. Water Resour.*, 24(7), 707723, doi:10.1016/S0309-1708(01)00005-7.

Lanzoni, S., R. Luchi, and M. Bolla Pittaluga (2015), Modeling the morphodynamic equilibrium of an intermediate reach of the Po River (Italy), *Adv. Water. Resour.*, 81, 95-102, doi:10.1016/j.advwatres.2014.11.004.

Lazzaro, G., S. Basso, M. Schirmer, and G. Botter (2013), Water management strategies for run-of-river power plants: Profitability and hydrologic impact between the intake and the outflow, *Water Resour. Res.*, 49, 82858298, doi:10.1002/2013WR014210.

Marsh, H. W., J. R. Balla, and R. P. McDonald (1988), Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size, *Psychol. Bull.*, 103(3), 391410, doi:10.1007/BF01102761.

Mejia, A., E. Daly, F. Rossel, T. Jovanovic, and J. Gironas (2014), A stochastic model of streamflow for urbanized basins, *Water Resour. Res.*, 50, 19842001, doi:10.1002/2013WR014834.

Merz, R., and G. Blöschl (2008), Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, 44, W08432, doi:10.1029/2007WR006744.

MeyerPeter, E., and R. Müller (1948), Formulas for bedload transport, *Proc. 2nd Meeting IAHR*, Stockholm, Sweden, 3964.

Milly, P. C. D., K. A. Dunne, and A. V. Vecchia (2005), Global pattern of trends in streamflow and water availability in a changing climate, *Nature*, 438(7066), 347350, doi:10.1038/nature04312.

Morrison, J. E., and J. A. Smith (2002), Stochastic modeling of flood peaks using the generalized extreme value distribution, *Water Resour. Res.*, 38(12), 1305, doi:10.1029/2001WR000502.

Müller, M. F., D. N. Dralle, and S. E. Thompson (2014), Analytical model for flow duration curves in seasonally dry climates, *Water Resour. Res.*, 50, 55105531, doi:10.1002/2014WR015301.

Müller, M. F., and S. E. Thompson (2015), Stochastic or statistic? Comparing flow duration curve models in ungauged basins and changing climates, *Hydrol. Earth Syst. Sci. Discuss.*, 12, 9765-9811, doi:10.5194/hessd-12-9765-2015.

Mutzner, R., E. Bertuzzo, P. Tarolli, S. V. Weijs, L. Nicotina, S. Ceola, N. Tomasic, I. Rodriguez-Iturbe, M. B. Parlange, and A. Rinaldo (2013), Geomorphic signatures on Brutsaert base flow recession analysis, *Water Resour. Res.*, 49, 54625472, doi:10.1002/wrcr.20417.

Nash, D. B. (1994), Effective sediment-transporting discharge from magnitude-frequency analysis, *J. Geol.*, 102(1), 79-95, doi:10.1086/629649.

Neal, C., B. Reynolds, P. Rowland, D. Norris, J. W. Kirchner, M. Neal, D. Sleep, A. Lawlor, C. Woods, S. Thacker, H. Guyatt, C. Vincent, K. Hockenhull, H. Wickham, S. Harman, and L. Armstrong (2012), High-frequency water quality time series in precipitation and streamow: from fragmentary signals to scientic challenge, *Sci. Total Environ.*, 434, 3-12, doi:10.1016/j.scitotenv.2011.10.072.

Newman, M. E. J. (2005), Power laws, Pareto distributions and Zipf's law, *Contemporary Physics*, 46, 323-351, doi:10.1016/j.cities.2012.03.001.

Papalexiou, S. M., and D. Koutsoyiannis (2013), Battle of extreme value distributions: A global survey on extreme daily rainfall, *Water Resour. Res.*, 49, doi:10.1029/2012WR012557.

Podolak, C. J. P., and M. W. Doyle (2015), Reservoir sedimentation and storage capacity in the United States: management needs for the 21st Century, *J. Hydraul. Eng.*, 141(4), 02515001, doi:10.1061/(ASCE)HY.1943-7900.0000999.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg (1997), The Natural Flow Regime, *Bioscience*, 47, 11, 769-784, doi:10.2307/1313099.

Poff, N. L., C. M. Brown, T. E. Grantham, J. H. Matthews, M. A. Palmer, C. M. Spence, R. L. Wilby, M. Haasnoot, G. F. Mendoza, K. C. Dominique, and A. Baeza (2016), Sustainable water management under future uncertainty with eco-engineering decision scaling, *Nat. Clim. Change*, 6, 2534, doi:10.1038/nclimate2765.

Pumo, D., L. Noto, and F. Viola (2013), Ecohydrological modelling of flow duration curve in Mediterranean river basins, *Adv. Water Resour.*, 52(0), 314327, doi:10.1016/j.advwatres.2012.05.010.

Quader, A. and Y. Guo (2009), Relative importance of hydrological and sediment-transport characteristics affecting effective discharge of small urban streams in Southern Ontario, *J. Hydrol. Eng.*, 14(7), 698710, doi:10.1061/(ASCE)HE.1943-5584.0000042.

Richter, B. D., J. V. Baumgartner, J. Powell, and D. P. Braun (1996), A method for assessing hydrologic alteration within ecosystems, *Conserv. Biol.*, 10, 11631174, doi: 10.1046/j.1523-1739.1996.10041163.x.

Rogger, M., H. Pirkl, A. Viglione, J. Komma, B. Kohl, R. Kirnbauer, R. Merz, and G. Blöschl (2012), Step changes in the flood frequency curve: Process controls, *Water Resour. Res.*, 48, W05544, doi:10.1029/2011WR011187.

Salinas, J. L., A. Castellarin, A. Viglione, S. Kohnov, and T. R. Kjeldsen (2014), Regional parent flood frequency distributions in Europe Part 1: Is the GEV model suitable as a pan-European parent?, *Hydrol. Earth Syst. Sci.*, 18, 4381-4389, doi:10.5194/hess-18-4381-2014.

Schaefli, B., A. Rinaldo, and G. Botter (2013), Analytic probability distributions for snow-dominated streamflow, *Water Resour. Res.*, 49, 27012713, doi:10.1002/wrcr.20234.

Schafer, J. L. (1997), Analysis of incomplete multivariate data, *Chapman & Hall*, London.

Schirmer, M., J. Luster, N. Linde, P. Perona, E. A. D. Mitchell, D. A. Barry, J. Hollender, O. A. Cirpka, P. Schneider, T. Vogt, D. Radny, E. Durisch-Kaiser (2014), Morphological, hydrological, biogeochemical and ecological changes and challenges in river restoration The Thur River case study, *Hydrol. Earth Syst. Sc.*, 18, 1-14, doi:10.5194/hess-18-2449-2014.

Shaw, S. B., T. M. McHardy, and S. J. Riha (2013), Evaluating the influence of watershed moisture storage on variations in base flow recession rates during prolonged rain-free periods in medium-sized catchments in New York and Illinois, USA, *Water Resour. Res.*, 49, 60226028, doi:10.1002/wrcr.20507.

Shaw, S. B., and S. J. Riha (2012), Examining individual recession events instead of a data cloud: A modified interpretation of streamflow recession data applied to glaciated watersheds in a humid, temperate climate, *J. Hydrol.*, 434435, 4654, doi:10.1016/j/jhydrol.2012.02.034.

Shields, F. Jr., R. Copeland, P. Klingeman, M. Doyle, and A. Simon (2003), Design for stream restoration, *J. Hydraul. Eng.*, 129(8), 575-584, doi:10.1061/(ASCE)0733-9429(2003)129:8(575).

Sholtes, J., K. Werbylo, and B. Bledsoe (2014), Physical context for theoretical approaches to sediment transport magnitude-frequency analysis in alluvial channels, *Water Resour. Res.*, 50, 79007914, doi:10.1002/2014WR015639.

Simon, A., W. Dickerson, and A. Heins (2004), Suspended-sediment transport rates at the 1.5-year recurrence interval for ecoregions of the United States: transport conditions at the bankfull and effective discharge?, *Geomorphology*, 58(14), 243262, doi:10.1016/j.geomorph.2003.07.003.

Sivapalan, M., G. Blöschl, R. Merz, and D. Gutknecht (2005), Linking flood frequency to long-term water balance: Incorporating effects of seasonality, *Water Resour. Res.*, 41, W06012, doi:10.1029/2004WR003439.

Snyder, N. P., K. X Whipple, G. E. Tucker, and D. J. Merritts (2003), Importance of a stochastic distribution of floods and erosion thresholds in the bedrock river incision problem, *J. Geophys. Res.*, 108, 2117, doi:10.1029/2001JB001655, B2.

Surian, N. (1999), Channel changes due to river regulation: the case of the Piave River, Italy, *Earth Surf. Process. Landforms*, 24, 11351151, doi: 10.1002/(SICI)1096-9837(199911)24:12<1135::AID-ESP40>3.0.CO;2-F.

Syvitski, J. P., M. D. Morehead, D. B. Bahr, and T. Mulder (2000), Estimating fluvial sediment transport: The rating parameters, *Water Resour. Res.*, 36(9), 27472760, doi:10.1029/2000WR900133.

Tockner, K., F. Malard, and J. V. Ward (2000), An extension of the flood pulse concept, *Hydrol. Process.*, 14, 28612883, doi: 10.1002/1099-1085(200011/12)14:16/17<2861::AID-HYP124>3.0.CO;2-F.

Tucker, G. E., and R. L. Bras (2000), A stochastic approach to modeling the role of rainfall variability in drainage basin evolution, *Water Resour. Res.*, 36(7), 19531964, doi:10.1029/2000WR900065.

Viglione, A., R. Merz, J. L. Salinas, and G. Blöschl (2013), Flood frequency hydrology: 3. A Bayesian analysis, *Water Resour. Res.*, 49, doi:10.1029/2011WR010782.

Villarini, G., and J. A. Smith (2010), Flood peak distributions for the eastern United States, *Water Resour. Res.*, 46, W06504, doi:10.1029/2009WR008395.

Vogel, R. M., and N. M. Fennessey (1995), Flow duration curves II : a review of applications in water resources planning, *J. Am. Water Resour. Assoc.*, 31(6), 10291039, doi:10.1111/j.1752-1688.1995.tb03419.x.

Vogel, R. M., J. R. Stedinger, and R. P. Hooper (2003), Discharge indices for water quality loads, *Water Resour. Res.*, 39, 1273, doi:10.1029/2002WR001872.

Wolman, M. G., and J. P. Miller (1960), Magnitude and frequency of forces in geomorphic processes, *J. Geol.*, 68, 5474, doi:10.1086/626637.

Yokoo, Y., and M. Sivapalan (2011), Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis, *Hydrol. Earth Syst. Sci.*, 15, 2805-2819, doi:10.5194/hess-15-2805-2011.