FDCFIT MANUAL

FDCFIT: A MATLAB Toolbox of Closed-form Parametric Expressions of the Flow Duration Curve

Jasper A. Vrugt[a,b], Mojtaba Sadegh[a]

[a] *Department of Civil and Environmental Engineering, University of California Irvine, 4130 Engineering Gateway, Irvine, CA 92697-2175*
[b] *Department of Earth System Science, University of California Irvine, Irvine, CA*

Abstract

The flow duration curve (FDC) is a signature catchment characteristic that depicts graphically the relationship between the exceedance probability of streamflow and its magnitude. This curve is relatively easy to create and interpret, and is used widely for hydrologic analysis, water quality management, and the design of hydroelectric power plants (among others). Several mathematical formulations have been proposed to mimic the FDC. Yet, these efforts have not been particularly successful, in large part because classical functions are not flexible enough to portray accurately the functional shape of the FDC for a large range of catchments and contrasting hydrologic behaviors. In a recent paper, *Sadegh et al.* (2015) introduced several commonly used models of the soil water characteristic as new class of closed-form parametric expressions for the flow duration curve. These soil water retention functions are relatively simple to use, contain between two to five parameters, and mimic closely the empirical FDCs of watersheds. Here, we present a simple MATLAB toolbox for the fitting of FDCs. This toolbox, called FDCFIT implements the different expressions introduced by *Sadegh et al.* (2015) and returns the optimized values of the coefficients of each model, along with graphical output of the fit. This toolbox is particularly useful for diagnostic model evaluation (*Vrugt and Sadegh*, 2013), as the optimized coefficients can be used as summary metrics. Two different case studies are used to illustrate

Email address: `jasper@uci.edu` (Mojtaba Sadegh)
URL: `http://faculty.sites.uci.edu/jasper` (Mojtaba Sadegh), `http://scholar.google.com/citations?user=zkNXecUAAAAJ&hl=en` (Mojtaba Sadegh)

the main capabilities and functionalities of the FDCFIT toolbox.

Keywords: Flow duration curve, Watershed hydrology, Discharge, Numerical modeling, Diagnostic model evaluation, Bayesian inference

1. Introduction and Scope

The flow duration curve (FDC) is a widely used characteristic signature of a watershed, and is one of the three most commonly used graphical methods in hydrologic studies, along with the mass curve and the hydrograph (*Foster*, 1934). The FDC relates the exceedance probability (frequency) of streamflow to its magnitude, and characterizes both the flow regime and the streamflow variability of a watershed. It is closely related to "survival" function in statistics (*Vogel and Fennessey*, 1994), and is interpreted as a complement to the streamflow cumulative distribution function (cdf). The FDC is frequently used to predict the distribution of streamflow for water resources planning purposes, to simplify analysis of water resources problems, and to communicate watershed behavior to those who lack in-depth hydrologic knowledge. One should be particularly careful to rely solely on the FDC as main descriptor of catchment behavior (*Vogel and Fennessey*, 1995) as the curve represents the rainfall-runoff as disaggregated in the time domain and hence lacks temporal structure (*Searcy*, 1959; *Vogel and Fennessey*, 1994).

The first application of the FDC dates back to 1880 and appears in the work by Clemens Herschel (*Foster*, 1934). Ever since, the FDC has been used in many fields of study including (among others) the design and operation of hydropower plants (*Singh et al.*, 2001), flow diversion and irrigation planning (*Chow*, 1964; *Warnick*, 1984; *Pitman*, 1993; *Mallory and McKenzie*, 1993), streamflow assessment and prediction (*Tharme*, 2003), sedimentation (*Vogel and Fennessey*, 1995), water quality management (*Mitchell*, 1957; *Searcy*, 1959; *Jehng-Jung and Bau*, 1996), waste-water treatment design (*Male and Ogawa*, 1984), and low-flow analysis (*Wilby et al.*, 1994; *Smakhtin*, 2001; *Pfannerstill et al.*, 2014). Recent studies have used the FDC as a benchmark for quality control (*Cole et al.*, 2003), and signature or metric for model calibration and evaluation (*Refsgaard and Knudsen*, 1996; *Yu and Yang*, 2000; *Wagener and Wheater*, 2006; *Son and Sivapalan*, 2007; *Yadav et al.*, 2007; *Yilmaz et al.*, 2008; *Zhang et al.*, 2008; *Blazkova and Beven*, 2009; *Westerberg et al.*, 2011; *Vrugt and Sadegh*, 2013; *Pfannerstill et al.*, 2014; *Sadegh and Vrugt*, 2014; *Sadegh et al.*, 2015b). For instance, *Vrugt and Sadegh* (2013) used the fitting coefficients of a simple parametric expression of the FDC as summary statistics in diagnostic model evaluation with approximate Bayesian computation (ABC).

Application of FDCs for hypothesis testing (*Kavetski et al.*, 2011) can improve identifiability and help attenuate the problems associated with tra-

ditional residual-based objective (likelihood) functions (e.g. Nash-Sutcliffe, sum of squared residuals, absolute error, relative error) that emphasize fitting specific parts of the hydrograph, such as high or low flows (*Schaefli and Gupta*, 2007; *Kavetski et al.*, 2011; *Westerberg et al.*, 2011), and thereby lose important information regarding the structural inadequacies of the model (*Gupta et al.*, 2008, 2012; *Vrugt and Sadegh*, 2013). The FDC is a signature watershed characteristic that along with other hydrologic metrics, can help shed lights on epistemic (model structural) errors (*Euser et al.*, 2013; *Vrugt and Sadegh*, 2013). For example, *Son and Sivapalan* (2007) used the FDC to highlight the reasons of model malfunctioning and to propose improvements to the structure of their conceptual water balance model for the watershed under investigation. Indeed, a deep groundwater flux was required to simulate adequately dominant low flows of the hydrograph. *Yilmaz et al.* (2008) in a similar effort to improve simulation of the vertical distribution of soil moisture in the HL-DHM model, used the slope of the FDC as benchmark for model performance. The FDC was deemed suitable for this purpose due to its strong dependence on the simulated soil moisture distribution, and relative lack of sensitivity to rainfall data and timing errors. However, the proposed refinements of the HL-DHM model were found inadequate and this failure was attributed to the inherent weaknesses of the conceptual structure of HL-DHM.

The usefulness of duration curves (e.g. precipitation (*Yokoo and Sivapalan*, 2011), baseflow (*Kunkle*, 1962) and streamflow (flow) (*Hughes and Smakhtin*, 1996)) depends in large part on the temporal resolution of the data (e.g. quarterly, hourly, daily, weekly, and monthly) these curves are constructed from. FDCs derived from daily streamflow data are commonly considered to warrant an adequate analysis of the hydrologic response of a watershed (*Vogel and Fennessey*, 1994; *Smakhtin*, 2001; *Wagener and Wheater*, 2006; *Zhao et al.*, 2012). For example, a FDC with a steep mid section (also referred to as slope) is characteristic for a watershed that responds quickly to rainfall, and thus has a small storage capacity and large ratio of direct runoff to baseflow. A more moderate slope, on the contrary, is indicative of a basin whose streamflow response reacts much slower to precipitation forcing with discharge that is made up in large part of baseflow (*Yilmaz et al.*, 2008).

The shape of the FDC is determined by several factors including (amongst others) topography, physiography, climate, vegetation cover, land use, and storage capacity (*Singh*, 1971; *Lane et al.*, 2005; *Zhao et al.*, 2012), and can be used to perform regional analysis (*Wagener and Wheater*, 2006; *Masih et al.*,

2010) or to cluster catchments into relatively homogeneous groups that exhibit a relatively similar hydrologic behavior (*Coopersmith et al.*, 2012; *Sawicz et al.*, 2011). Different studies have appeared in the hydrologic literature that have analyzed how the shape of the FDC is affected by physiographic factors and/or vegetation cover. Despite this progress made, interpretation of the FDC can be controversial if an insufficiently long streamflow data record is used (*Vogel and Fennessey*, 1994). The lower end of the FDC (low flows) is particularly sensitive to the period of study, and to whether the streamflow data includes severe droughts or not (*Castellarin et al.*, 2004a). If the available data is spare and does not warrant an accurate description of the FDC, then the use of an annual duration curve is advocated (*Searcy*, 1959; *Vogel and Fennessey*, 1994; *Castellarin et al.*, 2004a,b). This curve describes the relationship between the magnitude and frequency of the streamflow for a "typical hypothetical year" (*Vogel and Fennessey*, 1994). To construct an annual FDC, the available data is divided into n years and individual FDCs are constructed for each year. Then, for each exceedance probability a median streamflow is derived from these n different FDCs and used to create the annual FDC. *Vogel and Fennessey* (1994) used this concept to associate confidence and recurrence intervals to FDCs in a nonparametric framework. One should note that the FDC of the total data record is, in general, more accurate than the annual FDC (*Leboutillier and Waylen*, 1993).

To better analyze and understand the physical controls on the FDC, it is common practice to divide the total FDC (TFDC) into a slow (SFDC) and fast (FFDC) flow component (*Yokoo and Sivapalan*, 2011; *Cheng et al.*, 2012; *Coopersmith et al.*, 2012; *Yaeger et al.*, 2012; *Ye et al.*, 2012). For example, *Yokoo and Sivapalan* (2011) concluded from numerical simulations with a simple water balance model that the FFDC is controlled mainly by precipitation events and timing, whereas the SFDC is most sensitive to the storage capacity of the watershed and its baseflow response. This type of analysis is of particular value in regionalization studies, and prediction in ungauged basins. Indeed, much effort has gone towards prediction of the FDC in ungauged basins using measurements of the rainfall-runoff response from hydrologically similar, and preferably geographically nearby, gauged basins (*Holmes et al.*, 2002; *Sivapalan et al.*, 2003).

In this context, one approach has been to cluster catchments into classes with similar physiographic and climatic characteristics, and then to estimate dimensionless (non-parametric) FDCs for gauged basins which in turn are then applied to ungauged basins (*Niadas*, 2005; *Ganora et al.*, 2009).

The FDCs are normalized by an index value (e.g. mean annual runoff) to generate dimensionless curves (*Ganora et al.*, 2009). A detailed review on methods for clustering of homogeneous catchments appears in *Sauquet and Catalogne* (2011) and *Booker and Snelder* (2012), and interested readers are referred to these publications for more information. Another approach has been to mimic the empirical (observed) FDC with a mathematical/probabilistic model and to correlate the fitting coefficients of such parametric expressions to physical and climatological characteristics of the watershed using regression techniques (*Fennessey and Vogel*, 1990; *Yu et al.*, 1996, 2002; *Croker et al.*, 2003; *Castellarin et al.*, 2004a,b, 2007; *Li et al.*, 2010; *Sauquet and Catalogne*, 2011). Such pedotransfer functions can then be used to predict the FDC of ungauged basins from simple catchment data (e.g. soil texture, topography, vegetation cover, etc.).

Models that emulate the FDC can be grouped in two main classes: 1. Physical models that use physiographic and climatic characteristics of the watersheds (e.g. drainage area, mean areal precipitation, soil properties, etc.) as parameters of the FDC (*Singh*, 1971; *Dingman*, 1978; *Yu et al.*, 1996; *Holmes et al.*, 2002; *Yu et al.*, 2002; *Lane et al.*, 2005; *Botter et al.*, 2008; *Mohamoud*, 2008); and, 2. Probabilistic/mathematical functions that use between two to five fitting coefficients to mimic the empirical FDC as closely and consistently as possible (*Quimpo et al.*, 1983; *Mimikou and Kaemaki*, 1985; *Fennessey and Vogel*, 1990; *Leboutillier and Waylen*, 1993; *Franchini and Suppo*, 1996; *Cigizoglu and Bayazit*, 2000; *Croker et al.*, 2003; *Botter et al.*, 2008; *Li et al.*, 2010; *Booker and Snelder*, 2012).

The early work of *Dingman* (1978) is the first study that used physical models to mimic empirical FDCs. Topography maps were used as signature of catchment behavior to predict the FDC using relatively simple first-order polynomial functions. Two more recent studies by *Yu et al.* (1996, 2002) used similar regression functions to predict the FDC of catchments in Taiwan but considered the drainage area as main proxy of the rainfall-runoff transformation.

Probabilistic methods include the use of lognormal and lognormal mixture (*Fennessey and Vogel*, 1990; *Leboutillier and Waylen*, 1993; *Castellarin et al.*, 2004a; *Li et al.*, 2010), generalized Pareto (*Castellarin et al.*, 2004b), generalized extreme value, gamma and Gumbel (*Booker and Snelder*, 2012), beta (*Iacobellis*, 2008), and logistic distributions (*Castellarin et al.*, 2004b). Unfortunately, the presence of temporal correlation between successive (e.g. daily) streamflow observations violates some of the basic assumptions of these

classical statistical distributions, hence casting doubt on the validity of such models to describe closely and consistently the FDC (*Vogel and Fennessey*, 1994). Other mathematical models include (amongst others) the use of exponential, power, logarithmic (*Quimpo et al.*, 1983; *Franchini and Suppo*, 1996; *Lane et al.*, 2005; *Booker and Snelder*, 2012), and polynomial functions (*Mimikou and Kaemaki*, 1985; *Yu et al.*, 2002). In another line of work, *Cigizoglu and Bayazit* (2000) used convolution theory to predict the FDC as a product of periodic and stochastic streamflow components. What is more, *Croker et al.* (2003) used probability theory to combine a model that predicts the FDC of days with non-zero streamflow with a distribution function that determines randomly the probability of dry days.

Notwithstanding this progress made, the probabilistic and mathematical functions used in the hydrologic literature fail, usually, to properly mimic all parts of the FDC when benchmarked against watersheds with completely different hydrologic behaviors. Consequently, some researchers focus only on a specific portion of the FDC, commonly the low flows (*Fennessey and Vogel*, 1990; *Franchini and Suppo*, 1996), whereas others prefer to use several percentiles of the FDC rather than the entire curve (*Lane et al.*, 2005; *Mohamoud*, 2008; *Blazkova and Beven*, 2009). While complex "S" shaped FDCs can only be modeled adequately if a sufficient number of parameters (say five to seven) are used (*Ganora et al.*, 2009), one should be particularly careful using such relatively complex FDC models for regionalization and prediction in ungauged basins as parameter correlation and insensitivity can complicate and corrupt the inference and results.

In this manual, we introduce a new class of closed-form mathematical expressions which are capable of describing the FDCs of a very large number of watersheds with contrasting hydrologic behaviors. In a recent paper, *Sadegh et al.* (2015) extended on the ideas presented in *Vrugt and Sadegh* (2013), and proposed several commonly used functions of the soil water characteristic as mathematical models for the observed (empirical) FDCs. These models have between two to five parameters and describe closely the FDCs of a very large number of watersheds.

This paper describes a MATLAB toolbox for fitting of the FDC. This toolbox implements the various expressions of *Sadegh et al.* (2015) and returns their optimized coefficients along with graphical output of the quality of the fit. The built-in functions are illustrated using discharge data from two contrasting watersheds in the United States. These example studies are easy to run and adapt and serve as templates for other data sets. The present man-

ual has elements in common with the toolboxes of DREAM (*Vrugt*, 2015) and AMALGAM (*Vrugt*, 2015) and is specifically developed to help users implement diagnostic model evaluation (*Vrugt and Sadegh*, 2013).

The remainder of this paper is organized as follows. Section 2 discusses different parametric expressions of the FDC that are available to the user. This is followed in section 3 with a description of the MATLAB toolbox FDCFIT. In this section we are especially concerned with the input and output arguments of FDCFIT and the various utilities and options available to the user. Section 4 discusses two case studies which illustrate how to use the toolbox. The penultimate section of this paper (section 5) highlights recent research efforts aimed at further improving the fitting of flow duration curves with specific emphasis on the mathematical description of their spatial variability using the scaling framework. Finally, section 6 concludes this manual with a summary of the main findings.

2. Inference of the Flow Duration Curve

The flow duration curve (FDC) is a signature catchment characteristic that depicts graphically the relationship between the exceedance probability of streamflow and its magnitude. Figure 1 shows the empirical (observed) FDCs of eight watersheds of the MOPEX data set. Note that the streamflow values on the y-axis are normalized so that different watersheds are more easily compared.

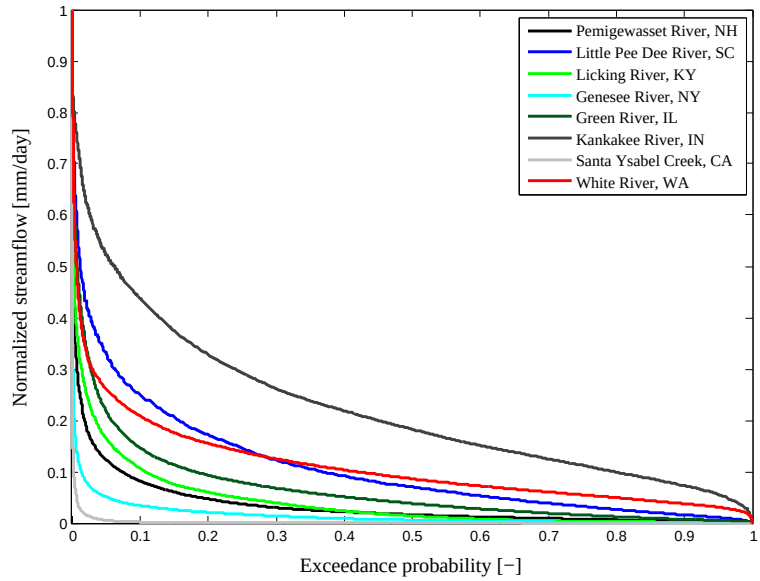

Figure 1: Flow duration curves of eight watersheds of the MOPEX data set. The watersheds exhibit quite contrasting hydrologic behavior - some retain water much better than others and hence the discharge response to rainfall is more delayed. The streamflow values are normalized so that the FDCs of the different watersheds are easily compared.

The plotted FDCs differ quite substantially from each other - a reflection of differences among the watersheds in their transformation of rainfall into runoff emanating from the catchment outlet. Ideally, we would have available a single parametric expression that can fit very closely the empirical FDCs of each of these watersheds.

2.1. Parametric expressions of FDC: Literature models

We briefly review the most widely used functions in the hydrologic literature to fit the observed FDCs. The Gumbel distribution is one of the most widely used functions to mimic the observed FDCs (*Booker and Snelder*, 2012). The cdf of the Gumbel distribution is given by

$$\mathcal{F}_{\mathbf{Y}}(y_t|a_{\mathrm{G}}, b_{\mathrm{G}}) = \exp\left\{-\exp\left(-\frac{y_t - a_{\mathrm{G}}}{b_{\mathrm{G}}}\right)\right\}, \tag{1}$$

where $\mathcal{F}_{\mathbf{Y}}$ denotes the cumulative density, a_{G} signifies the location, and b_{G} measures the scale of y_t. The Gumbel distribution is a particular case of the

generalized extreme value (GEV) distribution, whose cdf is given by

$$\mathcal{F}_{\mathbf{Y}}(y_t|a_{\mathrm{GEV}}, b_{\mathrm{GEV}}, c_{\mathrm{GEV}}) = \exp\left\{-\left[1 + c_{\mathrm{GEV}}\left(\frac{y_t - a_{\mathrm{GEV}}}{b_{\mathrm{GEV}}}\right)\right]^{-1/c_{\mathrm{GEV}}}\right\}, \quad (2)$$

where a_{GEV}, b_{GEV}, and c_{GEV} are the location, scale, and shape parameters, respectively. The shape parameter, c_{GEV}, controls the skewness of the distribution, and enables fitting of tailed streamflow distributions. The GEV distribution is widely used in the field of hydrology to analyze extremes (floods and droughts) (*Katz et al.*, 2002), as well to mathematically describe FDCs (*Booker and Snelder*, 2012). Note that if $c_{\mathrm{GEV}} = 0$, the GEV distribution simplifies to the Gumbel distribution. The exceedance probability, g_t, of a streamflow datum, y_t, can be derived from Equations 1 and 2 using

$$g_t = 1 - \mathcal{F}_{\mathbf{Y}}(y_t|\cdot). \quad (3)$$

Another popular formulation of the FDC is the 2-parameter exponential function of *Quimpo et al.* (1983) which is given by

$$y_t = a_{\mathrm{Q}} \exp(-b_{\mathrm{Q}} g_t), \quad (4)$$

where a_{Q} (mm/day) and b_{Q} (-) are fitting coefficients. This function predicts streamflow, y_t, based on exceedance probability, g_t, while the cdf predict exceedance probability, g_t, based on the observed discharge, y_t. We can simply invert Equation 4 to derive a mathematical expression for the exceedance probability, g_t, which in turn can then be compared directly to Equation 3

$$g_t = -\frac{1}{b_{\mathrm{Q}}} \log\left(\frac{y_t}{a_{\mathrm{Q}}}\right). \quad (5)$$

Another formulation of the FDC was proposed by *Franchini and Suppo* (1996). Their 3-parameter model is defined as

$$y_t = b_{\mathrm{FS}} + a_{\mathrm{FS}}(1 - g_t)^{c_{\mathrm{FS}}}, \quad (6)$$

in which a_{FS} (mm/day), b_{FS} (mm/day) and c_{FS} (-) are fitting coefficients. This function was originally proposed to only fit the low flows of the FDC, yet some authors (*Sauquet and Catalogne*, 2011) have used this formulation to fit the entire curve. If we invert Equation 6, then we have available an expression for the exceedance probability

$$g_t = 1 - \left(\frac{y_t - b_{\mathrm{FS}}}{a_{\mathrm{FS}}}\right)^{1/c_{\mathrm{FS}}}, \quad (7)$$

which allows us to compare the simulated values directly against those of the other two models.

Practical experience with these existing models suggests that they are not flexible enough to describe accurately the full FDC for the entire range of exceedance probabilities, let alone capture as closely and consistently as possible the large variability that exists among the FDCs of different watersheds. In their work on diagnostic model evaluation, *Vrugt and Sadegh* (2013) therefore introduced a new class of parametric expressions for the FDC. These models are derived from the field of soil physics, and describe closely the FDCs of the MOPEX data set (*Sadegh et al.*, 2015).

2.2. Parametric expressions of FDC: Proposed formulations

The functional shape of the FDC has many elements in common with that of the soil water characteristic (SWC). This is graphically illustrated in Figure 2 which plots the water retention function of five different soils presented in *van Genuchten* (1980). These curves depict the relationship between the volumetric moisture content, θ (x-axis) and the corresponding pressure head, h (y-axis) of a soil and are derived by fitting the following equation

$$\theta = \theta_r + (\theta_s - \theta_r)\left[1 + (\alpha|h|)^n\right]^{-m}, \tag{8}$$

to experimental (θ, h) data collected in the laboratory. This equation is also known as the van Genuchten (VG) model and contains five different parameters, where θ_s (cm^3/cm^3) and θ_r (cm^3/cm^3) denote the saturated and residual moisture content, respectively, and α (1/cm), n (-) and m (-) are fitting coefficients that determine the air-entry value and slope of the SWC. In most studies, the value of m is set conveniently to $1 - 1/n$ which not only reduces the number of parameters to four, but also provides a closed-form expression for the unsaturated soil hydraulic conductivity function (*van Genuchten*, 1980).

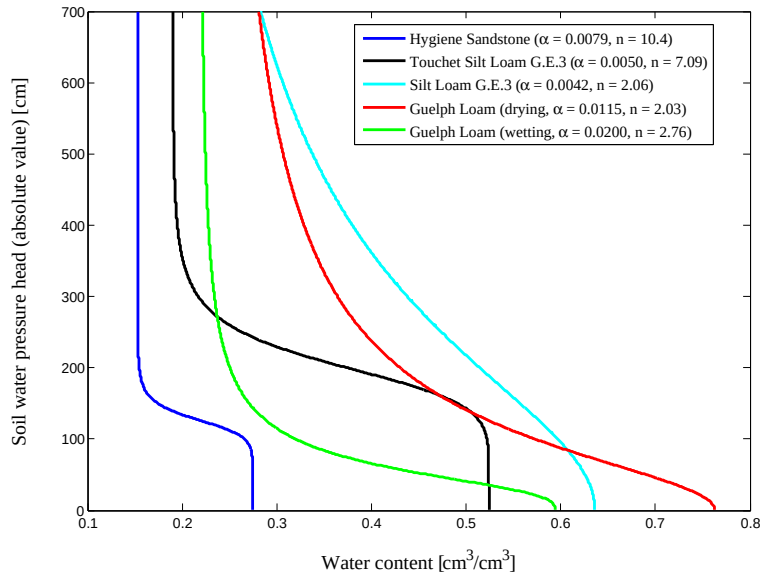

Figure 2: Water retention functions of five different soil types (derived from *van Genuchten* (1980)). This curve depicts the relationship between the water content, θ (cm^3/cm^3), and the soil water potential, h (cm). This curve is characteristic for different types of soil, and is also called the soil moisture characteristic.

The shape of the SWCs plotted in Figure 2 show great similarity with the FDCs displayed previously in Figure 1. This suggests that Equation 8 might be an appropriate parametric expression to describe quantitatively the relationship between the exceedance probability of streamflow and its magnitude. To make sure that the exceedance probability is bounded exactly between 0 and 1, we fix $\theta_r = 0$ and $\theta_s = 1$, respectively. This leads to the following 2-parameter VG formulation of the FDC proposed by *Vrugt and Sadegh* (2013)

$$g_t = \left(1 + (a_{\mathrm{VG}} y_t)^{b_{\mathrm{VG}}}\right)^{(1/b_{\mathrm{VG}}-1)}, \tag{9}$$

where the parameters a_{VG} (day/mm), and b_{VG} (-) need to determined by fitting against the observed FDC. If deemed appropriate, we can further increase the flexibility of Equation 9 by defining $c_{\mathrm{VG}} = 1/b_{\mathrm{VG}} - 1$. The MATLAB toolbox described herein allows the user to select both the 2- and 3-parameter VG formulations of the FDC.

The VG model is used widely in porous flow simulators to describe numerically variably saturated water flow. Yet, many other hydraulic models

have been proposed in the vadose zone literature to characterize the retention and unsaturated soil hydraulic conductivity functions. We consider here the lognormal SWC of *Kosugi* (1994, 1996)

$$g_t = \begin{cases} \frac{1}{2}\mathrm{erfc}\left\{\frac{1}{\sqrt{2}b_{\mathrm{K}}}\log\left(\frac{y_t - c_{\mathrm{K}}}{a_{\mathrm{K}} - c_{\mathrm{K}}}\right)\right\} & \text{if } y_t > c_{\mathrm{K}} \\ 1 & \text{if } y_t \leq c_{\mathrm{K}} \end{cases}, \tag{10}$$

where "erfc" denotes the complimentary error function, and a_{K} (mm/day), b_{K} (-) and c_{K} (mm/day) are fitting coefficients that need to be determined by calibration against the empirical FDC of each watershed. We can simplify this 3-parameter Kosugi formulation of the FDC by setting $c_{\mathrm{K}} = 0$. This assumption is appropriate for experimental data of the soil water characteristic, and the MATLAB toolbox considers both formulations.

The main reason to use Kosugi's SWC rather than other commonly used parametric expressions of the SWC such as *Brooks and Corey* (1964), is that the parameters can be related directly to the pore size distribution and hence exhibit a much better physical underpinning than their counterparts of the VG model. This might increase the chances of a successful regionalization.

The 2- and 3-parameter formulations of the Kosugi and VG models of the SWC assume implicitly a unimodal pore size distribution. This assumption might not be realistic for some soils and can lead to a relatively bad fit of these unimodal pore size distribution models to the observed water retention data. Soils with a heterogeneous (e.g. bimodal) pore size distribution can be much better described if we use a mixture of two or more SWCs. We follow *Durner* (1994) and propose the following mixture FDCs for watersheds with a complex hydrologic response

$$\begin{aligned} g_t &= \sum_{i=1}^{k} w_{i,\mathrm{VG}}\left[1 + (a_{i,\mathrm{VG}}y_t)^{b_{i,\mathrm{VG}}}\right]^{(1/b_{i,\mathrm{VG}}-1)} \\ g_t &= \frac{1}{2}\sum_{i=1}^{k} w_{i,\mathrm{K}}\mathrm{erfc}\left\{\frac{1}{\sqrt{2}b_{i,\mathrm{K}}}\log\left(\frac{y_t}{a_{i,\mathrm{K}}}\right)\right\}, \end{aligned} \tag{11}$$

where k denotes the number of constituent modes, $w_{i,\cdot}$ signifies the weight of each individual SWC ($w_{i,\cdot} \geq 0$ and $\sum_{i=1}^{k} w_{i,\cdot} = 1$), and $a_{i,\cdot}$ and $b_{i,\cdot}$ are the fitting coefficients of the ith SWC. The subscript "$\cdot$" stands either for VG or K. If fitting of the FDC is of main importance, then one could use a relatively large value for k (e.g. $k = 5$). Yet, for diagnostic analysis and

regionalization one should be particularly careful not to overfit the empirical FDCs. Indeed, considerable trade-off exists between the quality of fit of the FDC and the identifiability and correlation of the FDC model coefficients. The more parameters that are used to emulate the FDC, the better the fit to the empirical curve but at the expense of an increase in parameter uncertainty and correlation. Here, we assume $k = 2$ and hence mimic the empirical FDCs with a mixture of two VG or two Kosugi SWCs. For both classes of models, this mixture formulation has five parameters.

A bimodal FDC would seem appropriate for catchments with two or more discerning flow paths. For example, if the catchment exhibits two dominant modes (e.g. surface runoff and baseflow) then a bimodal mixture formulation of the FDC would seem appropriate. Appendix A summarizes the 2-, 3- and 5-parameter formulations of VG and Kosugi used in the MATLAB toolbox to model empirical FDCs.

3. FDCFIT

Now we have discussed the different models for the FDC (see Table 5), we are left with inference of their coefficients. We have developed a MATLAB program called FDCFIT which automatically determines the best values of the fitting coefficients of the FDC models described herein for a given streamflow data record. Graphical output is provided as well. FDCFIT implements a multi-start gradient-based local optimization approach with Levenberg Marquardt (LM) (*Marquardt*, 1963) to rapidly locate the "best" values of the FDC model parameters. A classical sum of squared error (SSE) objective function is used to summarize the distance between the observed and simulated FDCs,

$$\mathrm{SSE}(\mathbf{x}|\tilde{\mathbf{Y}}) = \sum_{t=1}^{N} \left(\tilde{G}_t - G_t(\mathbf{x}|\tilde{y}_t)\right)^2, \tag{12}$$

where $\mathbf{x}$ is the vector of fitting parameters (differs per FDC model), $\tilde{\mathbf{Y}} = \{\tilde{y}_1, \ldots, \tilde{y}_N\}$ denotes the observed streamflow data, and $\tilde{G}_t = \{\tilde{g}_1, \ldots, \tilde{g}_n\}$ and $G_t = \{g_1, \ldots, g_n\}$ signify the observed and FDC modeled exceedance probabilities, respectively. To maximize LM search efficiency, the Jacobian matrix is computed using analytical expressions for $\partial G_t / \partial x_i$ ($i = \{1, \ldots, d\}$), the partial sensitivity of the output (exceedance probability) predicted by each FDC model with respect to its individual model parameters. Moreover,

we recommend at least 20 different LM trials with different starting points drawn randomly from $\mathcal{U}_d[0,1]$, the d-dimensional uniform distribution on the interval between zero and one.

The LM search method is computationally efficient and much faster than global optimization methods. What is more, it does not require specification of a prior parameter space (except that all parameters are at least zero). Nevertheless, in some cases ($< 5\%$) the LM method was unable to reduce substantially the fitting errors of the (randomly chosen) starting points and the (derivative-free) Nelder-Mead Simplex algorithm (*Nelder and Mead*, 1965) was used instead to minimize Equation 13. Numerical results for the 3- and 5-parameter VG and Kosugi FDC models demonstrate (not shown) that this hybrid two-step approach provides better results for some of the watersheds than commonly used global optimizers. This is in large part due to the rather peculiar properties of the SSE-based response surfaces (flat with one small solution pocket).

3.1. FDCFIT: MATLAB implementation

The basic code of FDCFIT was written in 2014 and some changes have been made recently to support the needs of users. The FDCFIT code can be executed from the MATLAB prompt by the command

```
[x,RMSE] = FDCFIT(FDCPar,Meas_info,options)
```

where `FDCPar` (structure array), `Meas_info` (structure array) and `options` (structure array) are input arguments defined by the user, and `x` (vector) and `RMSE` (scalar) are output variables computed by FDCFIT and returned to the user. To minimize the number of input and output arguments in the FDCFIT function call and related primary and secondary functions called by this program, we use MATLAB structure arrays and group related variables in one main element using data containers called fields, more of which later. The third input variable, `options` is optional. We will now discuss the content and usage of each variable.

The FDCFIT function uses the built-in Levenberg Marquardt (LM) (*Marquardt*, 1963) and Nelder-Mead Simplex (NMS) (*Nelder and Mead*, 1965) algorithms to derive the coefficients of the different parametric expressions of the FDC. The functions that are used in the MATLAB package of FDCFIT are briefly summarized in Appendix B. In the subsequent sections we will discuss the MATLAB implementation of FDCFIT. This, along with prototype

case studies presented herein and template examples listed in RUNFDCFIT should help users apply the toolbox to their own data set and models (for diagnostic model evaluation).

3.2. Input argument 1: FDCPar

The structure FDCPar defines the name of the function and its formulation used to fit the empirical FDC, and the number of trials with the LM and NMS algorithms to optimize its coefficients. Table 1 lists the three different fields of FDCPar, their options, and default settings.

Table 1: Main algorithmic variables of FDCFIT: Description and corresponding fields of FDCPar and their options and default settings.

Description	Field FDCPar	Options	Default
	Problem dependent		
model class	model_class	'G'/'Q'/'GEV'/'FS'/'VG'/'K'	'VG'
formulation	formulation	'2'/'3'/'5' (if 'VG' or 'K')	'2'
number of trials	N	≥ 1	20

The field model_class of FDCPar allows the user to specify which of the different classes of FDC models (G: Gumbel, Q: Quimpo, GEV: GEV distribution, FS: Franchini and Suppo, VG: van Genuchten or K: Kosugi) to use. The field formulation pertains to the class of functions of Kosugi and van Genuchten, and defines which of their variants to use - either with 2, 3 or 5-parameters. Once the user has defined the model class and its formulation (if 'VG' or 'K') then the main FDCFIT script knows which parameters to estimate and their feasible search ranges. Finally, the field N of FDCPar stores the number of successive trials with LM and NMS used to calibrate the coefficients in the parametric expressions of VG and K.

3.3. Input argument 2: Meas_info

The second input argument Meas_info of the FDCFIT function has two fields that summarize the empirical flow duration curve. Table 2 summarizes these two different fields of Meas_info, their content and corresponding variable type.

Table 2: Content of input structure `Meas_info`.

Field of `Meas_info`	Description	Type
Y	Observed discharge data	$n \times 1$-vector
G	Exceedance probability	$n \times 1$-vector

The field Y of `Meas_info` stores the observed streamflow data which is used, together with the corresponding exceedance probability contained in field G to derive the coefficients of the parametric model of the FDC selected by the user. The number of elements of Y and G should match exactly, otherwise a warning is given and the code terminates prematurely.

3.4. (Optional) input argument 3: options

The structure `options` is optional and passed as third input argument to FDCFIT. The fields of this structure are passed directly to the LM and NMS algorithms. The following settings are recommended

```
options = optimset('MaxFunEvals',1e5,'TolFun',1e-4,'Display','off');
```

We refer to introductory textbooks and/or the MATLAB "help" utility for the built-in 'optimset' utility. This function is used to define the different fields of `options` and used by standard optimization routines.

3.5. Output arguments

We now briefly discuss the two output (return) arguments of FDCFIT including `x`, and `RMSE`. These two variables summarize the results of FDCFIT and are used for plotting of the results, and diagnostic analysis.

The variable `x` is a vector of size $1 \times d$ with the optimized values for the coefficients of the FDC model selected by the user. The root mean square error (RMSE) of the model fit (as measured in exceedance probability space) is stored as scalar in `RMSE`.

The following MATLAB command

```
FDC_Plot(FDCPar,Meas_info,x,RMSE,type);
```
(13)

generates the graphical output of FDCFIT.

4. Numerical examples

We now demonstrate the application of the MATLAB FDCFIT package to two different streamflow data sets. These two studies involve a dry and wet watershed, and as such cover different hydrologic behaviors.

4.1. Case Study I: French Broad River

The first case study involves daily streamflow data (mm/day) from the French Broad river. The file '08167500.dly' was downloaded from the MOPEX ftp site `ftp://hydrology.nws.noaa.gov/pub/gcip/mopex/US_Data/` and the following script is used in MATLAB to run the FDCFIT package.

```
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% %                          SETUP OF PROBLEM                            % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

% Set options for LM and Simplex search methods
options = optimset('MaxFunEvals',1e5,'TolFun',1e-4,'Display','off');

% Which class of model and formulation? (defined by user)
FDCPar.class = 'VG';            % Model class, 'G' (Gumbel) / 'Q' (Quimpo) / 'GEV' (GEV distribution) /
                                % 'FS'  (Franchini and Suppo) / 'VG' (van Genuchten) / 'K' (Kosugi)
FDCPar.formulation = '2';       % '2' (2-par) / '3' (3-par) / '5' (5-par), if model class is 'VG' or 'K'
FDCPar.N = 20;                  % How many different trials with Levenberg-Marquardt / Simplex)

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% %                           DEFINE DATA                                % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

% Load data of watershed
ID_watershed = '08167500';

% Load the data
data = load_data_dly(ID_watershed);

% Which column has the streamfow values?
column = 6;

% What type of FDC do we want: daily/weekly/monthly/yearly
type = 'daily';

% Now compute the empirical FDC
FDC = CalcFDC ( data , type , column );

% Now extract the streamflow values and exceedance probabilities
Meas_info.Y = FDC(:,1); Meas_info.G = FDC(:,2);

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% %                      DERIVE FITTING COEFFICIENTS                      % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
[ x , RMSE ] = FDCFIT ( FDCPar , Meas_info , options );

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% %                           PLOT RESULTS                               % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
FDCPlot ( FDCPar , Meas_info , x , RMSE , type );
```

Figure 3: Case study I: Daily streamflow data from the French Broad river basin in the USA.

The model of choice is the 2-parameter van Genuchten model (*Vrugt and Sadegh*, 2013) - and this model is fitted against the empirical FDC using 20 independent trials with the LM and NMS algorithms.

Figure 4 plots the observed (red dots) and fitted (blue line) FDC using a (A) linear and (B) logarithmic scale of the streamflow values.

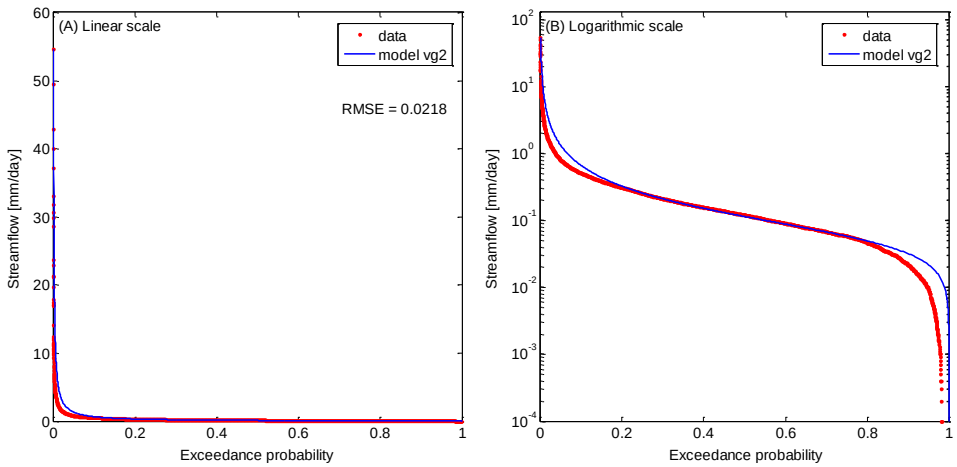

Figure 4: Comparison of the observed (red dots) and fitted (blue line) daily FDC using the 2-parameter van Genuchten expression. The plot on the left-hand side uses a linear scale whereas a logarithmic y-scale is used at the right-hand side to better visualize the results in the tails of the FDC.

The fit to the data is overall acceptable - yet deviations from the observed FDC are clearly visible in the tails of the FDC at low and high streamflow values respectively. A much improved fit to the empirical FDC is possible if the 5-parameter formulation of van Genuchten or Kosugi is used. We refer the reader to the work of *Sadegh et al.* (2015) for a comprehensive investigation of the proposed parametric expressions.

4.2. Case Study II: Gaudalupe River basin

The second case study involves daily streamflow data (mm/day) from the Guadalupe river basin. The file '03443000.dly' was downloaded from the MOPEX ftp site `ftp://hydrology.nws.noaa.gov/pub/gcip/mopex/US_Data/` and used to calculate the daily FDC. The following setup is used in MATLAB.

```
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% %                        SETUP OF PROBLEM                              % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

% Set options for LM and Simplex search methods
options = optimset('MaxFunEvals',1e5,'TolFun',1e-4,'Display','off');

% Which class of model and formulation? (defined by user)
FDCPar.class = 'K';              % Model class, 'G' (Gumbel) / 'Q' (Quimpo) / 'GEV' (GEV distribution) /
                                 % 'FS'  (Franchini and Suppo) / 'VG' (van Genuchten) / 'K' (Kosugi)
FDCPar.formulation = '5';        % '2' (2-par) / '3' (3-par) / '5' (5-par), if model class is 'VG' or 'K'
FDCPar.N = 20;                   % How many different trials with Levenberg-Marquardt / Simplex)

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% %                         DEFINE DATA                                  % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

% Load data of watershed
ID_watershed = '03443000';

% Load the data
data = load_data_dly(ID_watershed);

% Which column has the streamfow values?
column = 6;

% What type of FDC do we want: daily/weekly/monthly/yearly
type = 'daily';

% Now compute the empirical FDC
FDC = CalcFDC ( data , type , column );

% Now extract the streamflow values and exceedance probabilities
Meas_info.Y = FDC(:,1); Meas_info.G = FDC(:,2);

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% %                      DERIVE FITTING COEFFICIENTS                     % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
[ x , RMSE ] = FDCFIT ( FDCPar , Meas_info , options );

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% %                         PLOT RESULTS                                 % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
FDCPlot ( FDCPar , Meas_info , x , RMSE , type );
```

Figure 5: Case study II: Streamflow data from the Gaudalupe river basin in the USA.

The 5-parameter Kosugi model is used to describe the empirical FDC. A total of 20 trials is used with LM and NMS to optimize the fitting coefficients.

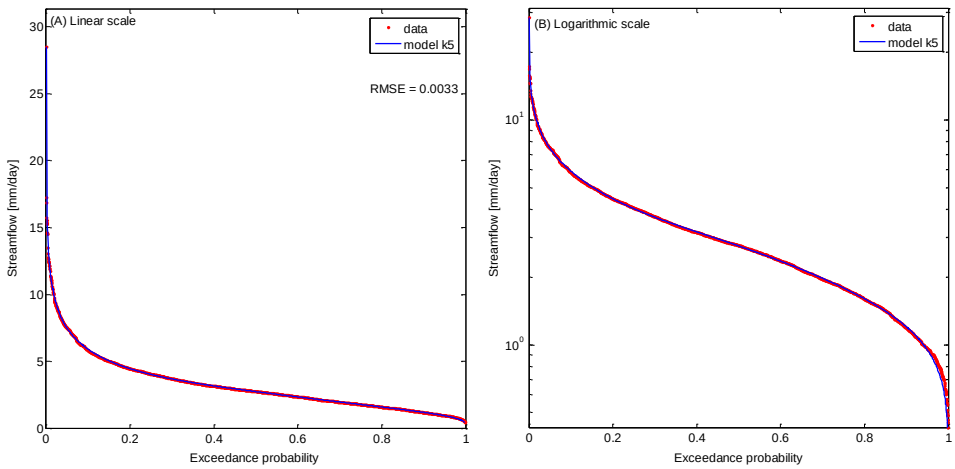

Figure 6: Comparison of the observed (red dots) and fitted (blue line) monthly FDC using the 5-parameter Kosugi expression. The plot on the left-hand side uses a linear scale whereas a logarithmic y-scale is used at the right-hand side to better visualize the results in the tails of the FDC.

Whereas some systematic deviations in the tails of the FDC were apparent in the first case study, the 5-parameter formulation of Kosugi almost perfectly matches the observed FDC of the Gaudalupe river basin. Our results demonstrate that the bimodal SWC formulations of the FDC offer a particularly large improvement in fit for watersheds that exhibit two dominant water flow pathways. Preferential flow at moisture contents of the watershed close to saturation, and baseflow during periods without extended rainfall. In between these events the watershed switches from one dominant transport mechanism to another.

Table 3 summarizes the quality of fit (RMSE) of each of the models of the MATLAB toolbox for the weekly FDC of the Gaudalupe watershed.

Table 3: Results of the different models for the Gaudalupe data set.

Model	class	formulation	RMSE (-)
Gumbel	'g'	n/a	0.0148
Quimpo	'q'	n/a	0.0599
GEV	'gev'	n/a	0.0055
Franchini and Suppo	'fs'	n/a	0.0659
Van Genuchten	'VG'	'2'	0.0143
Van Genuchten	'VG'	'3'	0.0056
Van Genuchten	'VG'	'5'	0.0034
Kosugi	'K'	'2'	0.0071
Kosugi	'K'	'3'	0.0071
Kosugi	'K'	'5'	0.0033

The Kosugi model with 5-parameters generates the closest fit to the observed FDC. The Franchini and Suppo and Quimpo formulations are particularly deficient - and unable to closely mimic the empirical weekly FDC. *Sadegh et al.* (2015) evaluates the different parametric expressions of FDCFIT for a large suite of watersheds of the MOPEX data set. Readers are referred to this publication for further details about the fitting results - and the hydrologic insights the proposed FDC functions offer. The work of *Vrugt and Sadegh* (2013) and *Sadegh et al.* (2015b) illustrates how to use the FDC parameter expressions for diagnostic model evaluation and detection/diagnosis of of hydrologic nonstationarity.

5. Scaling of flow duration curves

The use of equation 10 enables the use of physically based scaling (*Tuli et al.*, 2001) to coalesce the FDCs into a single reference curve using scaling factors that describe the set as a whole. This opens up new ways for catchment classification, geostatistical analysis and regionalization.

Figure 7 compares the original (unscaled) and scaled FDCs of the MOPEX data set derived by application of the scaling theory of *Tuli et al.* (2001). The solid line represents the reference curve.

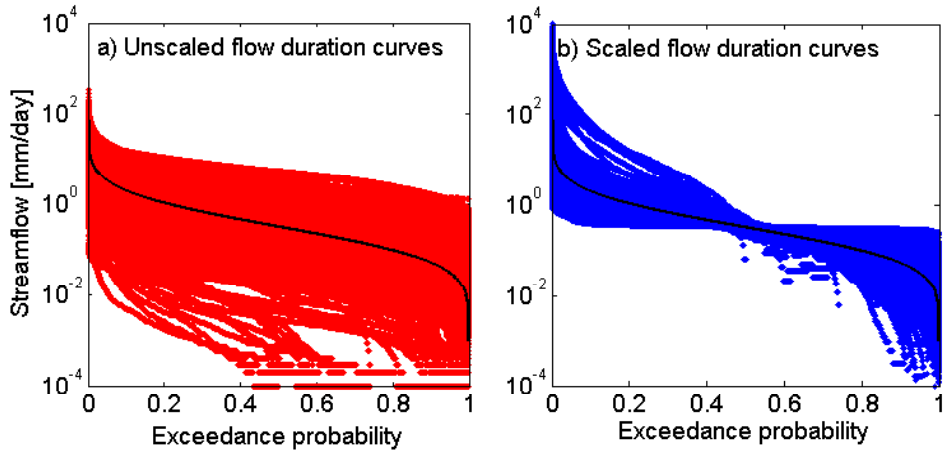

Figure 7: Application of physically-based scaling to flow duration curves of MOPEX data set: a) unscaled data, b) scaled data.

The scaled data groups much closer around the reference curve - which demonstrates that we can define with a single scaling factor each observed FDC. A correlation coefficient of 0.87 is found (not shown) between scaling factors and basic watershed properties. This provides new opportunities for regionalization. The FDC reference curves of different continents (countries) can serve as benchmark for prediction in ungaged basins. A publication on scaling and regionalization of FDCs is forthcoming (*Naeini and Vrugt*, 2015).

6. Printing of FDCFIT to screen

Figures 4 and 6 presented in case study I and II are generated automatically by the function `FDCPlot`. Note that the legend in both graphs adapts automatically to the model class and parametric formulation used. The FDCFIT code also returns to the user (in the MATLAB command window) the optimized values of the coefficients. Figure 8 displays a screen shot of the MATLAB command window after the program FDCFIT has terminated its calculations.

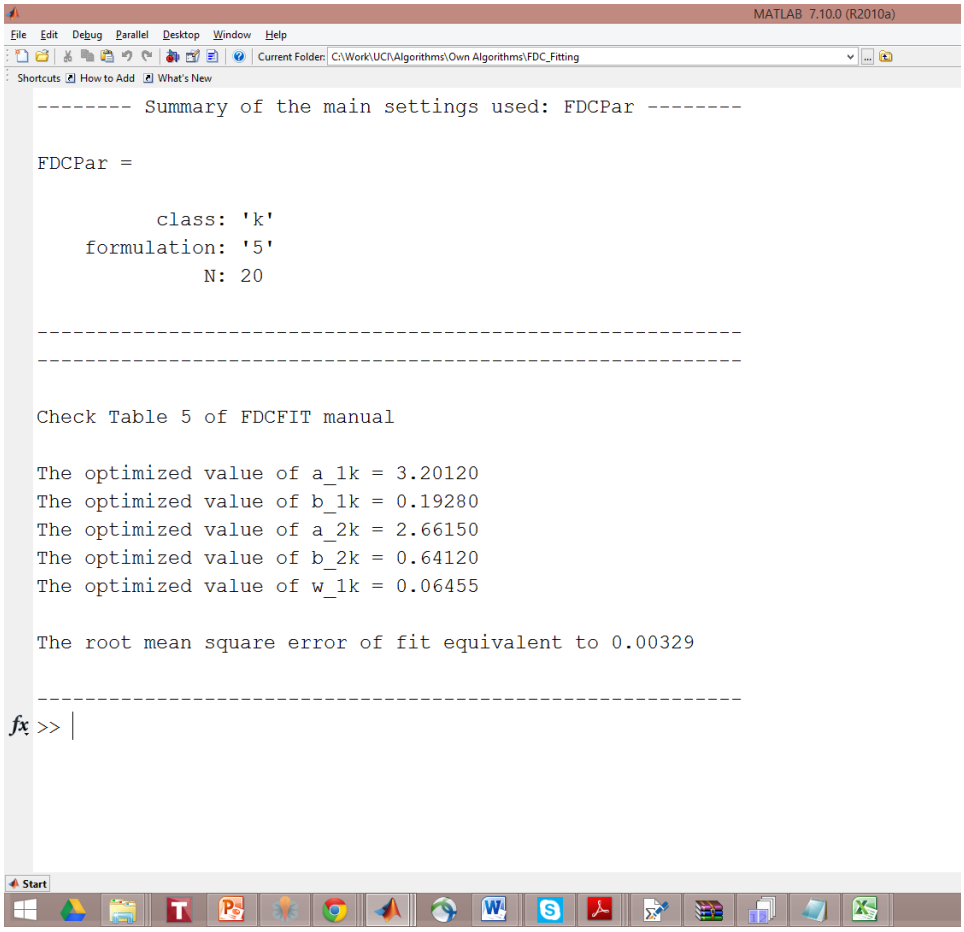

Figure 8: A print of the MATLAB screen for case study 2 after FDCFIT has terminated. The optimized values of the coefficients are listed using notation that corresponds to the Equations presented herein (summarized in Table 5 in Appendix B. For convenience, the print out also displays the RMSE of the least squares fit.

Notation (symbol use) is consistent with the different parametric expressions listed in section 2, and summarized in Table 5 of Appendix B. Once the parameters have been determined, they can (among others) be used within their respective parametric expression to predict the streamflow for a given exceedance probability. This requires inverting the different equations of Table 5 - or alternatively linear interpolation (table lookup) from a simulated record of streamflow values and corresponding exceedance probabilities. The

optimized parameter values can also be used for scaling and geostatistical analysis.

7. Summary

In this paper we have introduced a MATLAB package, entitled FDCFIT, which provides hydrologist with a new class of parametric functions of the flow duration curve. The coefficients (parameters) in these expressions are fitted automatically using data from an empirical FDC. Graphical output is provided as well. Two different case studies were used to illustrate the main capabilities and functionalities of the MATLAB toolbox. These example studies are easy to run and adapt and serve as templates for other modeling problems and watershed data sets.

The toolbox allows for determination of the daily, weekly, monthly and annual FDC - yet in our work we have not analyzed the relationship between these different curves and their optimized parameter values. Also, the parametric expressions of the FDC used herein apply directly to fitting of the annual peak flow curve as well - a powerful alternative to the log-Pearson type-III distribution advocated by the USGS in their 1982 contribution (*IACWD*, 1982) and used worldwide by many researchers to model flood flow frequencies. Much additional work is required to adopt this new methodology - with the advantage that it is easy to implement and provides estimates of flood-flow frequency estimates as key element to flood damage control.

8. Acknowledgements

The MATLAB toolbox of FDCFIT is available upon request from the first author, `jasper@uci.du`.

9. Appendix A

Table 4 summarizes, in alphabetic order, the different function/program files of the AMALGAM package in MATLAB. The main program runFDCFIT contains two prototype studies which involve the discharge response of the driest and wettest watershed on record of the original MOPEX data set. These example studies provide a template for users to setup their own case study. The last line of each example study involves a function call to FDCFIT, which uses all the other functions listed above to derive the fitting coefficients of the parametric expressions selected by the user. Each example problem of runFDCFIT has its own directory which stores the discharge data.

The function FDCPlot visualizes the results (output arguments) of FDCFIT. This includes two figures (linear and log-space) with a comparison of the observed and fitted FDC.

Those users that do not have access to the optimization toolbox need an alternative search algorithm to determine the values of the fitting coefficients of each parametric expression. The next version of this toolbox will include such utility. Alternatively, the user can combine this toolbox with the DREAM algorithm (*Vrugt*, 2015) - something that is straightforward to do and as byproduct also provides estimates of parameter and model uncertainty. Knowledge of these uncertainties is key for diagnostic model evaluation - in which the parameters are used as summary metrics.

Table 4: Description of the MATLAB functions and scripts (.m files) used by AMALGAM, version 1.4.

Name of function	Description
CalcFDC	Calculates the flow duration curve (daily/weekly/monthly/yearly) from a time series of discharge data
FDCFIT	Calls different functions and derives the fitting coefficients of a given parametric expression
FDCFIT_check	Checks the settings supplied by the user and if inconsistent returns a warning
FDCFIT_setup	Defines the ranges of the coefficients in each of the parametric expressions
FDCJac	Calculate the Jacobian (sensitivity) matrix for a given parametric expression
FDCPlot	Graphical output - plots the fitted flow duration curve in linear and log-scale and compares with data
FDCSWC	Returns the simulated exceedance probabilities of parametric FDC model selected by user
fminsearch	Built-in MATLAB function for multidimensional unconstrained nonlinear minimization (Nelder-Mead)
lsqnonlin	Built-in MATLAB function that solves nonlinear least squares optimization problems (Levenberg-Marquardt)
runFDCFIT	Main script of toolbox - setup of problem and call of main function

10. Appendix B

This Appendix summarizes the different parametric expressions of the FDC presented in section 2 and available in the MATLAB toolbox.

Table 5: Summary of the literature and proposed functions of the FDC. The fitting coefficients $a.$, $b.$, $c.$ and $w_{i.}$ are estimated from the empirical FDC using a multi-start local optimization method.

Model class	formulation
	2-parameter
Gumbel	$g_t = 1 - \exp\left\{-\exp\left(-\frac{y_t - a_{\mathrm{G}}}{b_{\mathrm{G}}}\right)\right\}$
Quimpo	$g_t = -\frac{1}{b_{\mathrm{Q}}} \log\left(\frac{y_t}{a_{\mathrm{Q}}}\right)$
VG	$g_t = \left(1 + (a_{\mathrm{VG}} y_t)^{b_{\mathrm{VG}}}\right)^{(1/b_{\mathrm{VG}} - 1)}$
Kosugi	$g_t = \frac{1}{2}\mathrm{erfc}\left\{\frac{1}{\sqrt{2} b_{\mathrm{K}}} \log\left(\frac{y_t}{a_{\mathrm{K}}}\right)\right\}$
	3-parameter
GEV	$g_t = 1 - \exp\left\{-\left[1 + c_{\mathrm{GEV}}\left(\frac{y_t - a_{\mathrm{GEV}}}{b_{\mathrm{GEV}}}\right)\right]^{-1/c_{\mathrm{GEV}}}\right\}$
Franchini and Suppo	$g_t = 1 - \left(\frac{y_t - b_{\mathrm{FS}}}{a_{\mathrm{FS}}}\right)^{1/c_{\mathrm{FS}}}$
VG	$g_t = \left(1 + (a_{\mathrm{VG}} y_t)^{b_{\mathrm{VG}}}\right)^{-c_{\mathrm{VG}}}$
Kosugi	$g_t = \begin{cases} \frac{1}{2}\mathrm{erfc}\left\{\frac{1}{\sqrt{2} b_{\mathrm{K}}} \log\left(\frac{y_t - c_{\mathrm{K}}}{a_{\mathrm{K}} - c_{\mathrm{K}}}\right)\right\} & \text{if } y_t > c_{\mathrm{K}} \\ 1 & \text{if } y_t \leq c_{\mathrm{K}} \end{cases}$
	5-parameter
VG	$g_t = \sum_{i=1}^{2} w_{i,\mathrm{VG}} \left[1 + (a_{i,\mathrm{VG}} y_t)^{b_{i,\mathrm{VG}}}\right]^{(1/b_{i,\mathrm{VG}} - 1)}$
Kosugi	$g_t = \frac{1}{2}\sum_{i=1}^{2} w_{i,\mathrm{K}}\mathrm{erfc}\left\{\frac{1}{\sqrt{2} b_{i,\mathrm{K}}} \log\left(\frac{y_t}{a_{i,\mathrm{K}}}\right)\right\}$

11. References

S. Blazkova, and K. Beven, "A limits of acceptability approach to model evaluation and uncertainty estimation in ïňĆood frequency estimation by continuous simulation: Skalka catchment, Czech Republic," *Water Resources Research*, vol. 45, W00B16, doi:10.1029/2007WR006726, 2009.

D.J. Booker, and T.H. Snelder, "Comparing methods for estimating flow duration curves at ungauged sites," *Journal of Hydrology*, vol. 434-435, pp. 78-94, 2012.

G. Botter, S. Zanardo, A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo, "Ecohydrological model of flow duration curves and annual minima," *Water Resources Research*, vol. 44, W08418, doi:10.1029/2008WR006814, 2008.

R.H. Brooks, and A.T. Corey, "Hydraulic properties of porous media," Hydrol. Pap. 3, Colo. State Univ., Fort Collins, 1964.

A. Castellarin, G. Galeati, L. Brandimarte, A. Montanari, and A. Brath, Regional flow-duration curves: reliability for ungauged basins," *Advances in Water Resources*, vol. 27, pp. 953-965, 2004a.

A. Castellarin, R.M. Vogel, and A. Brath, "A stochastic index flow model of flow duration curves," *Water Resources Research*, vol. 40, W03104, ffcccc-cdoi:10.1029/2003WR002524, 2004b.

A. Castellarin, G. Camorani, and A. Brath, "Predicting annual and long-term flow-duration curves in ungauged basins," *Advances in Water Resources*, vol. 30, pp. 937-953, doi:10.1016/j.advwatres.2006.08.006, 2007.

V.T. Chow, "Handbook of applied hydrology," Mc-Graw Hill Book Co., New York, NY, 1964.

L. Cheng, M. Yaeger, A. Viglione, E. Coopersmith, S. Ye, and M. Sivapalan, "Exploring the physical controls of regional patterns of flow duration curves âĂŞ Part 1: Insights from statistical analyses," *Hydrology and Earth System Sciences*, vol. 16, pp. 4435-4446, doi:10.5194/hess-16-4435-2012, 2012.

H.K. Cigizoglu, and M. Bayazit, "A generalized seasonal model for flow duration curve," *Hydrological Processes*, vol. 14, pp. 1053–1067, 2000.

R.A.J. Cole, H.T. Johnston, and D.J. Robinson, "The use of flow duration curves as a data quality tool," *Hydrological Sciences Journal*, vol. 48, no. 6, pp. 939-951, 2003.

E. Coopersmith, M.A. Yaeger, S. Ye, L. Cheng, and M. Sivapalan, "Exploring the physical controls of regional patterns of flow duration curves âĂŞ Part 3: A catchment classification system based on regime curve indicators," *Hydrology and Earth System Sciences*, vol. 16, pp. 4467-4482, doi:10.5194/hess-16-4467-2012, 2012.

K.M. Croker, A.R. Young, M.D. Zaidman, and H.G. Rees, "Flow duration curve estimation in ephemeral catchments in Portugal," *Hydrological Sciences Journal*, vol. 48, no. 3, pp. 427-439, doi: 10.1623/hysj.48.3.427.45287, 2003.

S.L. Dingman, "Sythesis of Flow-Duration Curves for Unregulated Streams in New Hampshire," *Water Resources Bulletin*, vol. 14, no. 6, pp. 1481-1502, 1978.

W. Durner, "Hydraulic conductivity estimation for soils with heterogeneous pore structure," *Water Resources Research*, vol. 30, no. 2, pp. 211-223, 1994.

T. Euser, H.C. Winsemius, M. Hrachowitz, F. Fenicia, S. Uhlenbrook, and H.H.G. Savenije, "A framework to assess the realism of model structures using hydrological signatures," *Hydrology and Earth System Sciences*, vol. 17, pp. 1893-1912, doi:10.5194/hess-17-1893-2013, 2013.

N. Fennessey, and R.M. Vogel, "Regional flow-duration curves for ungauged sites in Massachusetts, *Journal of Water Resources Planning and Management*, vol. 116, pp. 530-549, 1990.

H.A. Foster, "Duration curves," *ASCE Transactions*, vol. 99, pp. 1213-1267, 1934.

M. Franchini, and M. Suppo, "Regional analysis of flow duration curves for a limestone region," *Water Resources Management*, vol. 10, pp. 199-218, 1996.

D. Ganora, P. Claps, F. Laio, and A. Viglione, "An approach to estimate nonparametric flow duration curves in ungauged basins," *Water Resources Research*, vol. 45, W10418, doi:10.1029/2008WR007472, 2009.

H.V. Gupta, T. Wagener, and Y. Liu, "Reconciling theory with observations: elements of a diagnostic approach to model evaluation," *Hydrological Processes*, vol. 22, no. 18, pp. 3802-3813, 2008.

H.V. Gupta, M.P. Clark, J.A. Vrugt, G. Abramowitz, and M. Ye, "Towards a comprehensive assessment of model structural adequacy," *Water Resources Research*, vol. 48, W08301, doi:10.1029/2011WR011044, 2012.

M.G.R. Holmes, A.R. Young, A. Gustard, and R. Grew, "A region of influence approach to predicting flow duration curves within ungauged catchments," *Hydrology and Earth System Sciences*, vol. 6, pp. 721-731, doi:10.5194/hess-6-721-2002, 2002.

D.A. Hughes, and V. Smakhtin, "Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves," *Hydrological Sciences Journal*, vol. 41, no. 6, pp. 851-871, doi: 10.1080/02626669609491555, 1996.

V. Iacobellis, "Probabilistic model for the estimation of T year flow duration curves," *Water Resources Research*, vol. 44, W02413, doi:10.1029/2006WR005400, 2008.

K. Jehng-Jung, and S.F. Bau, "Risk analysis for flow duration curve based seasonal discharge management Programs," *Water Research*, vol. 30, no. 6, pp. 1369-1376, 1996.

R.W. Katz, M.B. Parlange, and P. Naveau, "Statistics of extremes in hydrology," *Advances in Water Resources*, vol. 25, no. 8âĂŞ12, pp. 1287-1304, 2002.

D. Kavetski, F. Fenicia, and M.P. Clark, "Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment," *Water Resources Research*, vol. 47, W05501, doi:10.1029/2010WR009525, 2011.

K. Kosugi, "Three-parameter lognormal distribution model for soil water retention," *Water Resources Research*, vol. 30, no. 4, pp. 891-901, 1994.

K. Kosugi, "Lognormal distribution model for unsaturated soil hydraulic properties," *Water Resources Research*, vol. 32, no. 9, pp. 2697-2703, 1996.

G.R. Kunkle, "The baseflow-duration curve, a technique for the study of groundwater discharge from a drainage basin," *Journal of Geophysical Research*, vol. 67, no. 4, pp. 1543-1554, 1962.

P.N.J. Lane, A.E. Best, K. Hickel, and L. Zhang, "The response of flow duration curves to afforestation," *Journal of Hydrology*, vol. 310, pp. 253-265, 2005.

D.W. Leboutillier, and P.R. Waylen, "A stochastic model of flow duration curves," *Water Resources Research*, vol. 29, no. 10, pp. 3535-3541, 1993.

M. Li, Q. Shao, L. Zhang, and F.H.S. Chiew, A new regionalization approach and its application to predict flow duration curve in ungauged basins," *Journal of Hydrology*, vol. 389, no. 1âĂŞ2, pp. 137-145, 2010.

J.W. Male, and H. Ogawa, "Tradeoffs in water quality management," *Water Resources Planning and Management*, vol. 110, no. 4, pp. 434-444, 1984.

S.J.L. Mallory, and R.S. McKenzie, "Water resources modeling of flow diversions," *Proceedings of the Sixth South African National Hydrology Symposium*, Pietermarietzburg, South Africa, vol. 1, pp. 429-436, 1993.

D.W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial & Applied Mathematics*, vol. 11, no. 2, pp. 431-441, 1963.

I. Masih, S. Uhlenbrook, S. Maskey, and M.D. Ahmad, "Regionalization of a conceptual rainfallâĂŞrunoff model based on similarity of the flow duration curve: A case study from the semi-arid Karkheh basin, Iran," *Journal of Hydrology*, vol. 391, pp. 188-201, 2010.

M. Mimikou, and S. Kaemaki, "Regionalization of Flow Duration Characteristics," *Journal of Hydrology*, vol. 82, pp. 77-91, 1985.

W.D. Mitchell, "Flow duration curves of Illinois streams," Illinois Department of Public Works and Buildings, Division of Waterways, 1957.

Y.M. Mohamoud, "Prediction of daily flow duration curve and streamflow for ungauged catchments using regional flow duration curves," *Hydrological Sciences Journal*, vol. 53, pp. 706-724, doi:10.1623/hysj.53.4.706, 2008.

M. Naeini, and J.A. Vrugt, "Miller similarity and scaling of flow duration curves: Theory, numerical implementation and regionalization," *Water Resources Research*, vol. XX, pp. XX–XX, doi:10.1002/wrcr.XXXX, 2015.

J.A. Nelder, R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308-313, doi:10.1093/comjnl/7.4.308, 1965.

I.A. Niadas, "Regional flow duration curve estimation in small ungauged catchments using instantaneous flow measurements and a censored data approach," *Journal of Hydrology*, vol. 314, pp. 48-66, 2005.

M. Pfannerstill, B. Guse, and N. Fohrer, "Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, *Journal of Hydrology*, vol. 510, pp. 447-458, 2014.

W.V. Pitman, "Simulation of run-of-river schemes using monthly data," *Proceedings of the Sixth South African Hydrology Symposium*, Pietermaritzburg, South Africa, vol. 1, 445-452, 1993.

R.G. Quimpo, A.A. Alejandrino, and T.A. McNally, "Regionalised flow duration curves for Philippines," *Journal of Water Resources Planning and Management*, vol. 109, no. 4, pp. 320âĂŞ330, 1983.

J.C. Refsgaard, and J. Knudsen, "Operational validation and intercomparison of different types of hydrological models," *Water Resources Research*, vol. 32, pp. 2189âĂŞ2202, 1996.

M. Sadegh, and J.A. Vrugt, "Approximate Bayesian Computation using Markov Chain Monte Carlo simulation: $\mathrm{DREAM}_{(\mathrm{ABC})}$, *Water Resources Research*, vol. 50, pp. 6767âĂŞ6787, doi:10.1002/2014WR015386, 2014.

M. Sadegh, J.A. Vrugt, and H.V. Gupta, "The soil water characteristic as new class of closed-form parametric expressions for the flow duration curve," *Water Resources Research*, vol. XX, no. XX, pp. XX-XX, 2015.

M. Sadegh, J.A. Vrugt, C. Xu, and E. Volpi (2015b), Diagnosing nonstationary watershed behavior from process-based model evaluation using approximate Bayesian computation, *Water Resources Research*, Under review.

E. Sauquet, and C. Catalogne, "Comparison of catchment grouping methods for flow duration curve estimation at ungauged sites in France, *Hydrology and Earth System Sciences*, vol. 15, pp. 2421âĂŞ2435, 2011.

K. Savicz, T. Wagener, M. Sivapalan, P.A. Troch, and G. Carrillo, "Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA," *Hydrology and Earth System Sciences*, vol. 15, pp. 2895-2911, 2011.

B. Schaefli, and H.V. Gupta, "Do Nash values have value?," *Hydrological Processes*, vol. 21, pp. 2075-2080, 2007.

J.K. Searcy, "Flow-duration curves," *Geological water supply paper: Methods and practices of the Geological Survey*, pp. 1-33, 1959.

K.P. Singh, "Model flow duration and streamflow variability," *Water Resources Research*, vol. 7, no. 4, pp. 1031-1036, 1971.

R.D. Singh, S.K. Mishra, and H. Chowdhary, "Regional flow-duration models for large number of ungauged Himalayan catchments for planning microhydro projects, *Journal of Hydrological Engineering*, vol. 6, no. 4, pp. 310-316, 2001.

M. Sivapalan, K. Takeuchi, S.W. Franks, V.K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J.J. McDonnell, E.M. Mendiondo, P.E. O'Connell, T. Oki, J.W. Pomeroy, D. Schertzer, S. Uhlenbrook, and E. Zehe, "IAHS decade on predictions in ungauged basins (PUB), 2003âĂŞ2012: Shaping an exciting future for the hydrological sciences," *Hydrological Sciences Journal*, vol. 48, pp. 857-880, doi:10.1623/hysj.48.6.857.51421, 2003.

V.U. Smakhtin, "Low flow hydrology: a review," *Journal of Hydrology*, vol. 240, pp. 147-186, 2001.

K. Son, and M. Sivapalan, "Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data," *Water Resources Research*, vol. 43, W01415, doi:01410.01029/02006wr005032, 2007.

R.E. Tharme, "A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers," *River Research Applications*, vol. 19, pp. 397âĂŞ441, doi:10.1002/rra.736, 2003.

A. Tuli, K. Kosugi, and J.W. Hopmans, "Simultaneouls scaling of soil water retention and unsaturated hydraulic conductivity functions assuming log-normal pore-size distribution," *Advances in Water Resources*, vol. 24, pp. 677-688, 2001.

M.T van Genuchten, "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils," *Soil Science Society of America Journal*, vol. 44, no. 5, pp. 892-898, 1980.

R.M. Vogel, and N.M. Fennessey, "Flow-duration curves I: New interpretation and confidence intervals, *Journal of Water Resources Planning and Management*, vol. 120, no. 4, pp. 485-504, 1994.

R.M. Vogel, and N.M. Fennessey, "Flow-duration curves II: A Review of applications in water resources planning, *Water Resources Bulletin*, vol. 31, no. 6, pp. 1029-1039, 1995.

J.A. Vrugt, and M. Sadegh, "Toward diagnostic model calibration and evaluation: Approximate Bayesian computation," *Water Resources Research*, vol. 49, doi:10.1002/wrcr.20354, 2013.

J.A. Vrugt, "Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB Implementation," *Environmental Modeling & Software*, vol. XX, no. XX, pp. XX-XX, doi:10.1016/j.envsoft.2014.XX.XXX, 2015a.

J.A. Vrugt, "Multi-criteria Optimization using the AMALGAM software package: Theory, concepts, and MATLAB Implementation, "*Manual, Version 1.0*, pp. 1-53, 2015b.

T. Wagener, and H.S. Wheater, "Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty," *Journal of Hydrology*, vol. 320, pp. 132-154, 2006.

C.C. Warnick, "Hydropower engineering," Prentice-Hall, Inc., Englewood Cliffs, NJ, pp. 59-73, 1984.

I.K. Westerberg, J.L. Guerrero, P.M. Younger, K.J. Beven, J. Seibert, S. Halldin, J.E. Freer, and C.Y. Xu, "Calibration of hydrological models using flow-duration curves," *Hydrology and Earth System Sciences*, vol. 15, pp. 2205âĂŞ2227, doi:10.5194/hess-15-2205-2011, 2011.

R. Wilby, B. Greenfield, and C. Glenny, "A coupled synoptic hydrological model for climate change impact assessment," *Journal of Hydrology*, vol. 153, pp. 265-290, 1994.

Interagency advisory commitee on water data, "Guidelines for determining flood flow frequency", Bulletin #17B of the hydrology subcommittee, US Department of the Interior, Geological Survey, Reston, VA 22092, 1982.

M. Yadav, T. Wagener and H.V. Gupta, "Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins," *Advances in Water Resources*, vol. 30, pp. 1756-1774, 2007.

M. Yaeger, E. Coopersmith, S. Ye, L. Cheng, A. Viglione, and M. Sivapalan, "Exploring the physical controls of regional patterns of flow duration curves âĂŞ Part 4: A synthesis of empirical analysis, process modeling and catchment classification," *Hydrology and Earth System Sciences*, vol. 16, pp. 4483âĂŞ4498, doi:10.5194/hess-16-4483-2012, 2012.

S. Ye, M. Yaeger, E. Coopersmith, L. Cheng, and M. Sivapalan, "Exploring the physical controls of regional patterns of flow duration curves âĂŞ Part 2: Role of seasonality, the regime curve, and associated process controls," *Hydrology and Earth System Sciences*, vol. 16, pp. 4447-4465, doi:10.5194/hess-16-4447-2012, 2012.

K.K. Yilmaz, H.V. Gupta, and T. Wagener, "A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model," *Water Resources Research*, vol. 44, W09417, doi:10.1029/2007WR006716, 2008.

Y. Yokoo, and M. Sivapalan, "Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis," *Hydrology and Earth System Sciences*,vol. 15, pp. 2805âĂŞ2819, doi:10.5194/hess-15-2805-2011, 2011.

P.S. Yu, and T.C. Yang, "Synthetic regional flow duration curve for southern Taiwan, *Hydrological Processes*, vol. 10, pp. 373âĂŞ391, 1996.

P.S. Yu, and T.C. Yang, "Using synthetic flow duration curves for rainfall-runoff model calibration at ungauged sites," *Hydrological Processes*, vol 14, pp. 117-133, 2000.

P.S. Yu, T.C. Yang, and Y.C. Wang, "Uncertainty analysis of regional flow duration curves," *Journal of Water Resources Planning and Management*, vol. 128, no. 6, pp. 424-430, 2002.

L. Zhang, N. Potter, K. Hickel, Y. Zhang, and Q. Shao, "Water balance modeling over variable time scales based on the Budyko framework âĂŞ Model development and testing," *Journal of Hydrology*, vol. 360, pp. 117âĂŞ131, doi:10.1016/j.jhydrol.2008.07.021, 2008.

F. Zhao, Z. Xu and L. Zhang, "Changes in streamflow regime following vegetation changes from paired catchments," *Hydrological Processes*, vol. 26, pp. 1561-1573, doi:10.1002/hyp.8266, 2012.