A Fokker–Planck–Kolmogorov equation-based inverse modelling approach for hydrological systems applied to extreme value analysis

Thomas Rosmann and Efraín Domínguez

ABSTRACT

This paper takes a stochastic approach to identify uncertainties in hydrological systems that can be applied to the study of hydrological extremes. The system to be identified is supposed to be governed by a stochastic differential equation of the Langevin type, whose parameters are found through the inverse solution of the equivalent Fokker–Planck–Kolmogorov equation. The study presents the algorithmic and numerical implementation for the inverse modelling process, along with the implementation of this approach in three study areas. Results showed a flexible method that made it possible to consider hydrological variability and seasonality during system identification. The identified system parameters rely on the internal–external driving factors of the analysed river basin and provide indications about the behaviour of extreme events in possible future climate scenarios or situations where internal system parameters are altered. While the study cases presented refer to non-stationary Markov processes that allow for one-dimensional systems identification only, the proposed methodological approach is a step in the right direction when it comes to identifying n-dimensional Markov processes/systems.

Thomas Rosmann (corresponding author)
Pontificia Universidad Javeriana,
Maestria en Hidrosistemas,
Carrera 7 No. 42-00, Bogota,
Colombia
E-mail: trosmann@javeriana.edu.co

Efraín Domínguez
Departamento de Ecologia y Territorio,
Pontificia Universidad Javeriana,
Bogota,
Colombia

INTRODUCTION

In recent years, extreme events have been associated with global climate change and have had an impact on human society (Parry *et al.* 2007; UNESCO 2014). According to these reports, worldwide water consumption will increase by about 55% until 2050, a fact which highlights factors of uncertainty like hydrometeorological extremes (floods, droughts or heavy precipitation events), especially in light of the fact that they have changed in recent decades.

Changes in hydrological time series have been studied in a wide range of investigations using a number of different methods, with trend analysis being the most frequent one. A vast number of studies has been conducted on the topic,

all of which would be impossible to mention here. However, all of the following studies show an indication of change in the examined parameters. Trends in temperature time series are almost exclusively positive, both in mean values and extreme value series and on all latitudes (Aguilar *et al.* 2005; Nicholson *et al.* 2013; Tao *et al.* 2014). Some studies were able to prove that changes in temperatures impacted other variables, as well (Hayhoe *et al.* 2007; Xu *et al.* 2010). Trends in precipitation series were also usually positive and while some studies found precipitation events to intensify (Vargas *et al.* 2002; Cantet *et al.* 2011; Min *et al.* 2011), others observed decreasing trends (Tabari *et al.* 2015). For

discharge series, there were many negative trends, although in a number of cases an increase in extreme events like floods was observed (Dai *et al.* 2009; Zhao *et al.* 2015). However, trend studies for discharge series have to take into account human activity as a major factor. Furthermore, in most of the above-mentioned studies, seasonal trends were found.

With the results of trend studies in mind, the assumption is that some kind of change is happening on the planet, which gives reason to further study the topic in depth. One of the crucial questions that needs to be asked is as follows: Is the nature of extreme events changing and how can humans adapt to that change? To answer this question, it is not enough to state that extreme events are changing. Rather, explanations are needed to also show how their underlying processes are working and which influences are causing the change (Rosmann 2014). A more profound understanding of the statistical characteristics of the time series under investigation can be achieved by taking a stochastic approach to studying the random processes that define them (Maldonado 2009; Kovalenko 2012).

Stochastic models have been applied successfully in various hydrological practices. The long-term variations in annual discharges and runoff dynamics were described by dynamic stochastic models for various Russian river basins (Frolov 2006; Dolgonosov & Korchagin 2007). Naidenov proposed a physical explanation of probabilistic characteristics of hydrological processes (Naidenov & Podsechin 1992; Naidenov & Shveikina 2002, 2005), as well as Koutsoyiannis (2008) in the Nile River basin, while Kovalenko developed a new modelling framework heavily based on the concepts of the theory of stochastic processes for the simulation and forecasting of complex systems (Kovalenko 1986, 2012; Kovalenko *et al.* 1993). In the field of extreme events, Naghettini *et al.* (2012) proposed a model to relate extreme rainfall events to flood volumes, and Lawal *et al.* (1997) applied a stochastic model of annual low flow hydrographs to Canadian river basins, among others.

Domínguez & Rivera (2010) proposed a stochastic model using the Fokker–Planck–Kolmogorov (FPK) equation for predicting the monthly affluent to the Betania hydropower plant in the Upper Magdalena River basin. Unami *et al.* (2010) developed a model for the assessment of flood and drought risk based on an Ornstein–Uhlenbeck process, which is numerically solved by the backward Kolmogorov equation, and Dong & Wu (2013) used a Fokker–Planck equation model to study the evolution of shorelines.

All of these previous studies show the ability of stochastic models to relate the probabilistic characteristics of hydrological processes with the system input signal and physio-geographic and other characteristics of the watersheds, in which they are located.

With the results of previous investigations in mind, the goal of this study was to develop an inverse modelling approach using the FPK equation to identify uncertainties in hydrological systems and the driving factors that influence their extreme values. A model based on the Langevin equation was developed by solving its equivalent FPK equation and therefore finding the drift and diffusion terms that describe the structure of the random process. It also tries to establish a relationship between the calibrated model parameters and the geomorphometrical basin parameters, as well as rainfall input to the system. This way, a relationship between the abstract model parameters with the physical and hydrological parameters of the river basin can be developed allowing a better understanding of the possible mechanisms that govern the dynamics of the underlying process. This, again, could indicate new practices in watershed management, prevention and mitigation of extreme events.

METHODS

Stochastic model

To extract information from a hydrologic random variable, Yevjevich (1987) proposed the use of four of its properties: (1) estimation of parameters and other characteristic values, (2) estimation of probability distribution functions, (3) modelling of processes in time and (4) the determination of properties of random variables.

The main parameter in the proposed model is discharge at a given point and is represented by a random process $Q(t)$. The process consists of 12 independent random variables q, indexed from 1 to 12, which represent the monthly mean, maximal or minimal discharges for each month of the

year. $Q(t)$ takes the form of

$$Q(t) = \{q_1, q_2, \ldots, q_{12}\} \tag{1}$$

The monthly mean, maximal or minimal discharges were calculated from the daily values of the obtained discharge time series.

Several authors reported that the rainfall–runoff process, at certain time scales can be considered as a Markov process (Naidenov & Podsechin 1992; Kovalenko *et al.* 1993; Druzhinin & Sikan 2001; Naidenov & Shveikina 2005) because the Markovianity condition is met (Equation (2)).

$$p(Q_k|Q_1, Q_2, \ldots, Q_{k-1}) = p(Q_k|Q_{k-1}) \tag{2}$$

Annual and monthly extreme runoff series' autocorrelation functions have correlation radii of lag one, which is the required autocorrelation radius for a process to be considered as Markovian (Rozhdenstvenskiy & Chevotariov 1974; Haan 1977; Kazakievich 1989). If Markovianity is already established, then for the full identification of the hydrological system, seasonality and random components should be derived. The seasonality of the process can be described through an ordinary differential equation, usually referred to as the deterministic kernel, while the random part needs to be represented by a white noise modulated process or by its respective noise intensities (Sveshnikov 1966; Kovalenko *et al.* 1993; Gardiner 2004). Combining the deterministic kernel with a random component leads to a stochastic differential equation of the Langevin type which completely describes the hydrological system. At the same time, the ordinary differential equation defining the deterministic kernel requires the form of the right part expression to be defined while the random part needs the noise intensity or the modulation function to be calibrated.

For this reason, $Q(t)$ is modelled using a Langevin-type equation, which has already been applied successfully in previous studies (Denisov *et al.* 2009), taking the form of

$$\frac{dQ(t)}{dt} = \psi[Q(t), t] + \zeta[Q(t), t]\xi(t) \tag{3}$$

where $\psi[Q(t), t]$ is the function describing the deterministic kernel of the process and $\zeta[Q(t), t]$ the deterministic

modulating function of the white noise term $\xi(t)$. It is well known that there is an equivalent FPK equation for any Langevin equation (Gardiner 2004; Denisov *et al.* 2009), and since it was shown by Dominguez & Rivera (2010) that it is possible to solve the FPK equation numerically, enabling a bidirectional shift, an FPK equation approach was implemented. The FPK equation models the evolution of probability density in time, which is one of the most important statistical characteristics of random differential equations and can be seen as a closed system if it is produced by a noise-generating function (Denisov *et al.* 2009). As described later on, Markovianity was proven for the data used in this investigation and therefore it was possible to implement a model using the one-dimensional FPK equation for the solution of the Langevin equation.

The one-dimensional FPK equation derived from the Langevin equation (Equation (3)) takes the following form (Equation (4)) and was used to solve the inverse problem described in the following section:

$$\frac{\partial p(Q, t)}{\partial t} = \frac{-\partial}{\partial x}[A(Q, t)p(Q, t)] + \frac{1}{2}[B(Q, t)p(Q, t)] \tag{4}$$

where $A(Q, t)$ and $B(Q, t)$ are deterministic equations describing the drift and diffusion terms, and $p(Q, t)$ represents the probability density of Q at the moment t (Dominguez & Rivera 2010). $A(Q, t)$ and $B(Q, t)$ are linked to the parameters of the system that is described and in this case is the river basin (Kovalenko *et al.* 1993), although the nature of these relationships was initially not known.

An important goal of this paper was to gain a better understanding of the dynamics of the underlying process, therefore deterministic functions that were sufficiently easy to implement were used to create the drift and diffusion functions. $A(Q, t)$ and $B(Q, t)$ were chosen to be second order polynomials, since this form of representation takes into account linear and quadratic components for drift and diffusion functions, such as those presented in Kovalenko (2012) for one analytical derivation of drift and diffusion coefficients.

$$A(Q, t) = k_1 Q + k_2 Q^2 + k_3 \tag{5}$$

$$B(Q, t) = g_1 Q + g_2 Q^2 + g_3 \tag{6}$$

with k_1, k_2, k_3, g_1, g_2, and g_3 being the initially unknown parameters of the system linked to drift (k_i) and diffusion (g_i).

In order to establish a relationship between the parameters of the Langevin and FPK equation, Sveshnikov (1966) proposed the following:

$$A(Q, t) = \psi[Q(t), t] + \frac{1}{2}\zeta[Q(t), t]\frac{\partial\zeta(Q, t)}{\partial Q} \tag{7}$$

$$B(Q, t) = \zeta^2[Q(t), t] \tag{8}$$

Substituting Equation (6) into Equation (8) gives Equation (9), as well as substituting Equations (5) and (9) into Equation (7) gives Equation (10):

$$\zeta[Q(t), t] = \sqrt{g_1 Q + g_2 Q^2 + g_3} \tag{9}$$

$$\psi[Q(t), t] = k_1 Q + k_2 Q^2 + k_3 - \frac{1}{4}(g_1 + 2g_2 Q) \tag{10}$$

and Equation (3) becomes

$$\frac{dQ(t)}{dt} = k_1 Q + k_2 Q^2 + k_3 - \frac{1}{4}(g_1 + 2g_2 Q) \\ + \sqrt{g_1 Q + g_2 Q^2 + g_3}\xi(t) \tag{11}$$

Since all of the above-presented parameters k_1, k_2, k_3 and g_1, g_2, g_3 are initially unknown, one of the main tasks of this paper was to establish a connection between these and the known and measurable parameters of the basin. These could be external parameters like precipitation or temperature and internal parameters such as the morphometry or the land cover of the basin.

Implementation of the inverse modelling FPK approach

To identify the stochastic mathematical operator that rules the hydrological response of a basin, an inverse modelling approach based on the FPK equation was implemented. This approach includes five steps as shown in Figure 1.
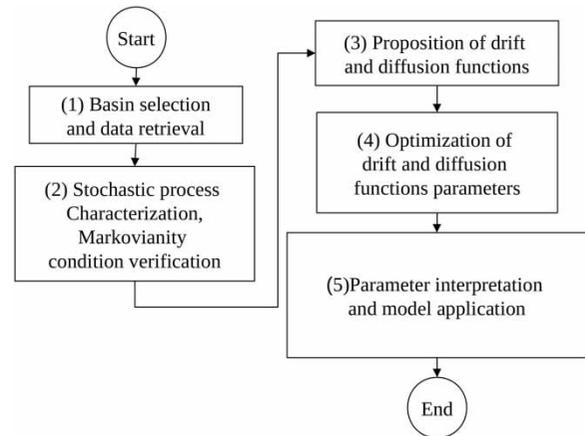

Figure 1 | Inverse modelling flow chart.

In the first step, a basin of interest was selected and related hydrometeorological and geomorphometrical data were retrieved. Precipitation and temperature stations well correlated with the discharges at the basin outlet were selected. As system output, mean and extreme discharges were calculated from the observed daily discharge data as the mean, maxima and minima of daily values. These mean and extreme values were applied as the basin's responses to the monthly mean and extreme event reference precipitation. The reference precipitation was calculated as the mean, maxima and minima value from the number of days previous to the extreme events that had the highest correlation with the discharge values. A more detailed description of the data selection criteria can be found in the data description section.

In the second step, the stochastic process was characterized at the monthly time level. Characterization included randomness tests for each time index and evaluating the best fit of probability distribution functions to the discharge data among 12 theoretical distribution models (Normal, Lognormal, Gamma, Loggamma, Gumbel with positive skew, Gumbel with negative skew, Weibull Min, Weibull Max, Powerlaw, Pareto, Exponential, Logistic). Subsequently, the best fitted probability distribution was used as the initial condition for the FPK equation optimization problem for monthly mean, maximum and minimum discharge, respectively. The best fit distribution for each month was used, instead of inferring the best fit over all of the 12 months of the same random process and applying it to each month. The distribution functions were adjusted to

fit the monthly means, maxima and minima and as a first approximation, the stochastic process characterization can be represented as the sequence of marginal probability density distribution functions for each of the random variables.

In order to implement the one-dimensional FPK equation presented above, a Markovian structure was needed for all the runoff processes, therefore the Markovianity condition was tested and verified by comparing the autocorrelation coefficient of different time lags against their confidence interval. As proposed in Druzhinin & Sikan (2001), the autocorrelation function of the process was compared against the standard error function (Equation (12)) of the autocorrelation coefficient (Figure 2). The critical autocorrelation radius of the process is the x-value of the autocorrelation function, at which it intersects the standard error function of the autocorrelation coefficient. The results of this analysis indicated that for all the basins described later, a Markovian structure could be assumed. The standard error function is given as follows:

$$e = \frac{1 - r^2}{\sqrt{n-1}} t_{1-\alpha/2} \tag{12}$$

where r is the autocorrelation coefficient for lagged values, n is the total number of observations in each random variable and t is the critical value of the Student's t distribution with significance level α, which was chosen to be 5%.

To understand the process, it is very important to determine the form of the Langevin equation that describes the process of the observed river basin. This means that it is necessary to find the drift governing expression $\psi(Q(t), t)$ and also the noise modulating function $\zeta(Q(t), t)$ of Equation (3). To achieve this goal, in the third step of the implementation the FPK equation was solved numerically as described in the previous section and implemented in the Python scientific programming language (van Rossum & Drake 2003). The drift and diffusion functions were proposed to be second order polynomials (Equations (5) and (6)), whose parameters were initially unknown. The base for the inverse problem was set, since these were the parameters whose values should be found in the optimization process in order to obtain the best simulation of the PDF translation from one month to the next.

In the fourth step, a numerical scheme was implemented to be used for the optimization of the unknown model parameters. The numerical scheme that makes it possible to fully solve the FPK equation and, therefore, also includes an implicit solution is described in detail in Dominguez & Rivera (2010) and was also implemented in the Python programming language, making use of the scientific libraries SciPy, NumPy and Pandas. This scheme offers a solution including a bidirectional approach for the drift term, therefore allowing a drift in both directions. The solution includes directional weights (φ_L and φ_R) for this purpose
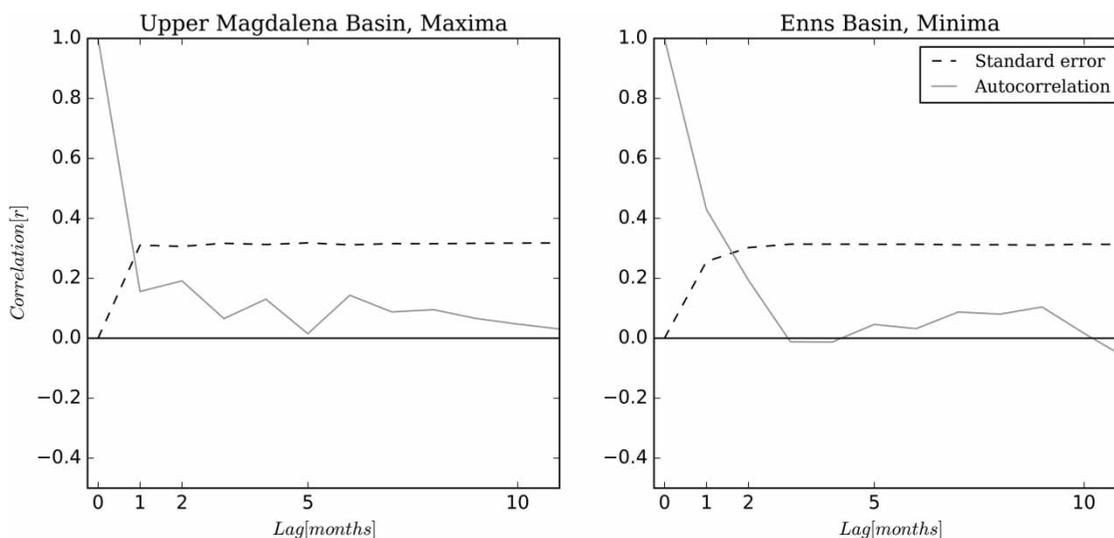

Figure 2 │ Autocorrelation function of monthly maxima and minima that fulfil Markovianity.

and a time layer weight (λ) that controls whether the equation is solved explicitly (when equal to 0) or implicitly (when equal to 1).

$$\frac{p_j^{i+1} - p_j^i}{\Delta t}$$

$$= -\left\{ \lambda \left[\frac{\varphi_L \left(A_{j+1}^{i+1} p_{j+1}^{i+1} - A_j^{i+1} p_j^{i+1} \right)}{\Delta Q} + \frac{\varphi_R \left(A_j^{i+1} p_j^{i+1} - A_{j-1}^{i+1} p_{j-1}^{i+1} \right)}{\Delta Q} \right] \right.$$

$$+ (1-\lambda) \left[\frac{\varphi_L \left(A_{j+1}^i p_{j+1}^i - A_j^i p_j^i \right)}{\Delta Q} + \frac{\varphi_R \left(A_j^i p_j^i - A_{j-1}^i p_{j-1}^i \right)}{\Delta Q} \right]$$

$$+ \frac{\lambda}{2} \left[\frac{\left(B_{j+1}^{i+1} p_{j+1}^{i+1} - 2 B_j^{i+1} p_j^{i+1} + B_{j-1}^{i+1} p_{j-1}^{i+1} \right)}{\Delta Q^2} \right]$$

$$\left. + \frac{(1-\lambda)}{2} \left[\frac{\left(B_{j+1}^i p_{j+1}^i - 2 B_j^i p_j^i + B_{j-1}^i p_{j-1}^i \right)}{\Delta Q^2} \right] \right\} \tag{13}$$

In Equation (13), i denotes the interval descriptor for the time step and j for discharge. Due to the input data used, the increments of discharge j remain constant at $1 \text{ m}^3/\text{s}$, and therefore, to guarantee the stability of the system, it is necessary for the increment of the time step to fulfil a Courant–Friedrichs–Lewy stability condition (Dominguez & Rivera 2010).

$$max[|B(Q,t)|] \frac{\Delta t}{\Delta Q^2} < \frac{1}{2} \tag{14}$$

An absorbing-type boundary was used, which assumes that any value outside the upper and lower boundary of the PDF has a probability of 0. Therefore, it was necessary to scale the boundary values α and β sufficiently in order not to lose probability density in the tails (Gardiner 2004):

$$p(Q,t) = 0, \ if \ Q \notin \ \{\alpha \ldots \beta\} \tag{15}$$

All of the parameters were known except the drift and diffusion vectors, which were determined using an iterative optimization approach for Equation (13). First, initial guesses for the six parameters k_1, k_2, k_3, g_1, g_2 and g_3 were provided to the model to calculate the translation of probability density from one time step to the next. The mean average error between the observed and the modelled data of the translation's final month was calculated and the initial guesses were adapted by iteratively optimizing them and thereby minimizing the error. For this task, the *curvefit* optimization method in SciPy's *optimize* package was chosen, which uses the Levenberg–Marquardt algorithm (Moré 1978), applying the least squares method and returns a set of parameters corresponding to the minimized error. As an additional control, a mean average error of less than 10% had to be achieved to accept the resulting set of parameters.

In the fifth and final step, the obtained optimal values of k_1, k_2, k_3, g_1, g_2 and g_3 were validated by establishing a relationship between them and the physical (morphometric and climatological) parameters of the basin, whose values could be deduced from observed data, such as precipitation and temperature time series. For this purpose, multiple regression analysis was applied, which was conducted between the optimal parameters and precipitation and temperature data to test the external influences on the basin on one side. On the other side, this was more difficult to do for internal basin parameters because no data were available about them. Some previous studies have related the coefficients of variation and skewness to internal parameters. For example, Andrews (1980) showed that the skew of streamflow is inversely proportional to the basin size and Dodov & Foufoula-Georgiou (2005) showed correlations between streamflow variability and channel vegetation. In a more recent work, Calle *et al.* (2017) gave a measure of correlation between the regulation coefficient, also a measure of variability, and different morphometric parameters. For this reason, the coefficients of variation and skewness were used as an indicator for the behaviour of internal basin parameters.

Data used

The aforementioned inverse modelling approach was implemented in four study areas, which should be located in different climatological regimes and at the same time count on high-quality data. For each of the selected basins, the lowermost discharge station available before its outlet was used for discharge data. All precipitation stations within the basin, which counted on a correlation of over

60% with the discharge data, as well as the closest temperature station to the discharge station, were also used for the analysis. The precipitation value was estimated as the mean of all the selected precipitation stations. The temporal range of data was chosen according to the availability of data in each river basin, but was at least 40 years up to the year 2012. The location of the four basins is shown in Figure 3.

The Enns River basin is located in central Austria, has an area of approximately 6,100 square kilometres and includes parts of the northern range of the Alps in its upper reaches. The Enns River has a length of 253 kilometres, the highest elevation within the basin is approximately 2,600 metres, and the outflow into the Danube River is located at 240 metres above sea level. At this point, the average discharge is over 200 m^3/s. Thirty-four precipitation stations meeting the above-mentioned criteria are located in the basin.

The Magdalena River is the biggest river in Colombia with a total length of 1,530 kilometres. The Betania reservoir is located about 200 kilometres from the source of the river, and the last discharge station before it was chosen to delimit the study area, which has an area of 5,800 square kilometres. It ranges from approximately 700 to almost 3,700 metres in altitude. It includes five precipitation stations that meet the selection criteria.

The third basin analysed was the Upper Great Miami River in the United States, located in the states of Ohio and Indiana, delimited at Dayton, OH with a total area of 6,500 square kilometres. Eight precipitation stations with data available between 1948 and 2012 were used.

The Brisbane River basin is located in the Australian territory of Queensland with the last discharge station before the outflow into the Pacific Ocean being Savages Crossing with a basin of 10,000 square kilometres. It is located 18 kilometres downstream of the Wivenhoe dam, which delimits the reservoir of the same name and regulates the river. Thirteen precipitation stations were used within the basin. The Brisbane basin was used principally to verify whether the fact that the model is based on basin characteristics holds for a station that is regulated. The expectation was that there would be problems implementing the model in this basin.

For each basin, the necessary discharge, precipitation and temperature data were collected and prepared as described above and presented in Figure 4. The initial conditions shown in this figure were the input data for the model optimization runs.

RESULTS

As mentioned above, the FPK equation model was used to simulate the transformations of probability density from

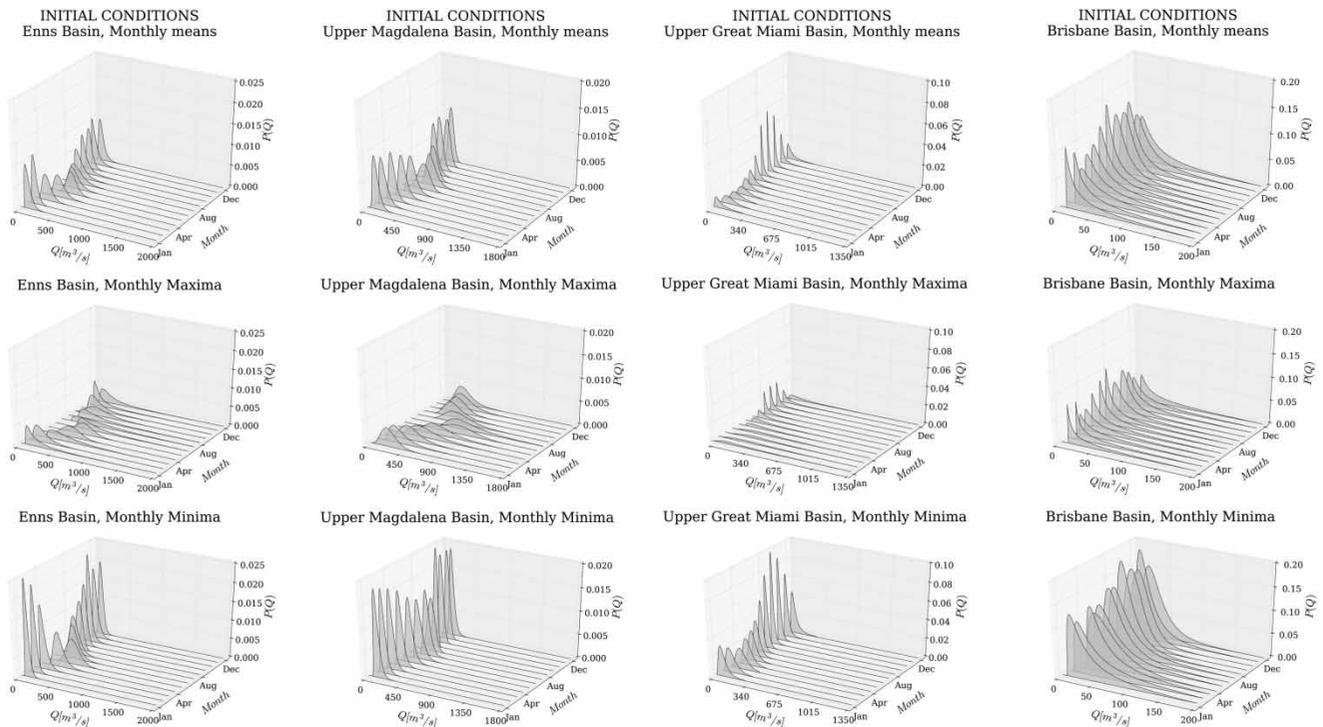**Figure 3** | Location of the study areas.

Figure 4 | Stochastic process characterization with marginal probability density functions.

one month to the next, which means that the probability density curve of one month was used to model that of the next. For this reason, the optimal model parameters were determined in order to provide the best possible simulation. After the relationship to the physical parameters was established, these model parameters could be modified to simulate changed conditions and the system's response to them. As described in Equations (5), (6) and (11), these were the parameters k_1, k_2, k_3, g_1, g_2 and g_3.

In the first set of model runs, unsupervised optimization was conducted, which did not take any assumptions and tried to optimize all six parameters. The results of this first run showed that it was impossible to find a global minimum of optimization and, therefore, a perfectly optimized solution. It further led to the conclusion that a very large number of local minima could be obtained and therefore the procedure was very much dependent on the initial guesses that were used for the parameters. Changing these initial guesses in many cases led to completely different results. On the other hand, it was shown that an extraordinarily good fit could be achieved with a large number of different parameter sets, taking into account that it was quite possible to find a fit

with a mean average error below 1% for all transformations from one month to the next for all four basins.

The ordinary least squares technique (Feldman & Valdez-Flores 2009) was used to establish a linear regression between the model and watershed parameters described above, but it was not possible to find any results that relate satisfactorily to the physical parameters of the basin.

For this reason, it could be concluded that an unsupervised optimization was not the most adequate approach to this problem, and that some kind of supervision had to be implemented. This supervised modelling approach meant reducing the number of equation parameters to be optimized and deducting some of their values from observed data. Without knowing the detailed characteristics of the river basin, this was more easily done for external basin parameters than for internal ones.

To decide how model parameters could be provided by available data, the following reasoning was applied. The analytical forms of the drift and diffusion coefficients in Equations (16) and (17) show that the drift term describes the instantaneous rate of change of the mean and the diffusion term the instantaneous rate of change of the squared

fluctuations of the process at a given Q (Dominguez & Rivera 2010). This leads to the suggestion that a variable which is correlated to Q might influence the drift and diffusion coefficients.

$$A(Q, t) = \lim_{\Delta t \to 0} \frac{E[\Delta Q | Q]}{\Delta t} \tag{16}$$

$$B(Q, t) = \lim_{\Delta t \to 0} \frac{E[\Delta Q^2 | Q]}{\Delta t} \tag{17}$$

Further data analysis showed that there was a correlation between the runoff and precipitation data, usually ranging from 0.6 to 0.8, as well as a slightly lower one between temperature and runoff minima. Therefore, monthly precipitation was chosen as the variable to provide the data used as an external basin parameter. The assumption was that the independent terms of the drift and diffusion equations, k_3 and g_3, should be most closely related to the external parameters of the basin. The remaining four parameters, k_1, k_2, g_1 and g_2, were optimized and their correlation to the discharge series' coefficients of variation and skewness was tested after an optimal result was obtained.

Leading from Equations (16) and (17), the following equations were proposed to calculate these fixed independent terms, k_3 and g_3:

$$k_3 = \Delta X = X_{final} - X_{initial} \tag{18}$$

$$g_3 = \Delta var[X] = var[X_{final}] - var[X_{initial}] \tag{19}$$

where X represents monthly precipitation for mean data and reference precipitation for extreme value data of the initial and final month of the transformation.

For each transformation, the values for the fixed parameters were calculated for their use in the model and, consequently, the optimization procedure reduced to only four parameters to be optimized. Due to the sensibility to the initial guesses, the optimization procedure was run with three sets of initial guesses for the four parameters, and the best approximation was used, requiring a 5% margin between the expected and the simulated PDF of the final month as an acceptable error. While it was possible to find a satisfactory

approximation for each of the unsupervised optimizations, it was now only possible in about 70% of all the transformations.

As expected, a supervised optimization in the Brisbane basin was not possible at all, which is most probably due to the Wivenhoe dam, which regulates the inflow from the majority of the river basin before the Savages Crossing station. It could therefore be concluded that the regulatory nature of the dam does not permit a reasonable application of the model in the Brisbane basin. Since the changes in discharge do not respond to the natural variation of the major part of the watershed, but are principally controlled by the artificial conditions imposed by the amount of water passing by the dam, one cannot apply natural operators to the model. For this reason, the Brisbane basin was not used in further analysis.

Regression analysis was applied to the optimal parameters of the remaining three basins, which indicated a relationship of the parameters k_1, g_1 and g_2 with the coefficients of variation and skewness, therefore these three parameters were related to the internal parameters of the basin. The coefficient k_2 does not have a significant correlation with any coefficient, which makes sense, since the parameters k_2 and g_2 are associated with the squared nature of the process (see Equation (17)). Yet the expected value, to which the parameter k_2 is linked, does not produce polynomials of an order larger than 1. For extreme values, the changes in the coefficients could only be related to the parameter k_1, which suggests that they are less influenced by the fluctuations and more by direct changes to the kernel of the system.

Model application

The calibrated model was applied to the observed data for the three watersheds of the Enns, Upper Magdalena and Upper Great Miami Rivers to simulate future changes of extreme events. The observed discharge data were used as initial conditions as before, and future time series were created for precipitation by extrapolating linear regressions calculated for the available data for a period of 20 years after the last record. The Kullback–Leibler divergence criterion (Kullback & Leibler 1951) was used to compare the difference between the actual and the simulated probability density curves.

The intensification and weakening of the extreme events could be estimated by the movement of the probability density curve in either direction, or by the way the area contained under the tails of the distributions changed. In most of the cases, the curve was moved to one direction by the changes, but it also sharpened or flattened and the tails of the distribution rose or lowered.

The model runs were executed using the calculated future precipitation values to represent the external parameters of the basin, as well as modifying the parameters of the model associated with the internal parameters, k_1, g_1 and g_2. Since no association could be made with the observed data, the values of these three parameters were increased or decreased by 10%, which is approximately equal to the average change in external parameters. Finally, a simulation was conducted applying changes to both external and internal parameters.

Observations show that the effects were strongest when applying changes to both types of parameters, but it also became apparent that the effects caused by changes to internal parameters of the watershed were greater than those caused by changes to external parameters. The results indicate the biggest impacts in the Great Miami and the least significant ones in the Upper Magdalena basin. Minima were affected more strongly than maxima, except for the Magdalena basin.

Figure 5 shows how the shape of the PDF for minima is altered by changes to internal parameters (the values of the three parameters were reduced by 10%) for the month of April in the Enns basin. What can be seen is that the curve is sharpened, so the region around the mean is fortified and the tails are lowered. The difference between the two curves is shown in the lower part of the graph. This behaviour might indicate that the number of events showing a discharge amount similar to the expected value of the distribution become more frequent, and those that are more or less intense, less frequent.

Figure 6 shows the effects on the PDF for the Magdalena basin for maxima in the transformation from September to October, suggesting an intensification and increase of maxima in the month of October due to the move of the curve to the right and the rise of the upper tail. This might be an important observation, since September and October coincide with the rainy season in this watershed.
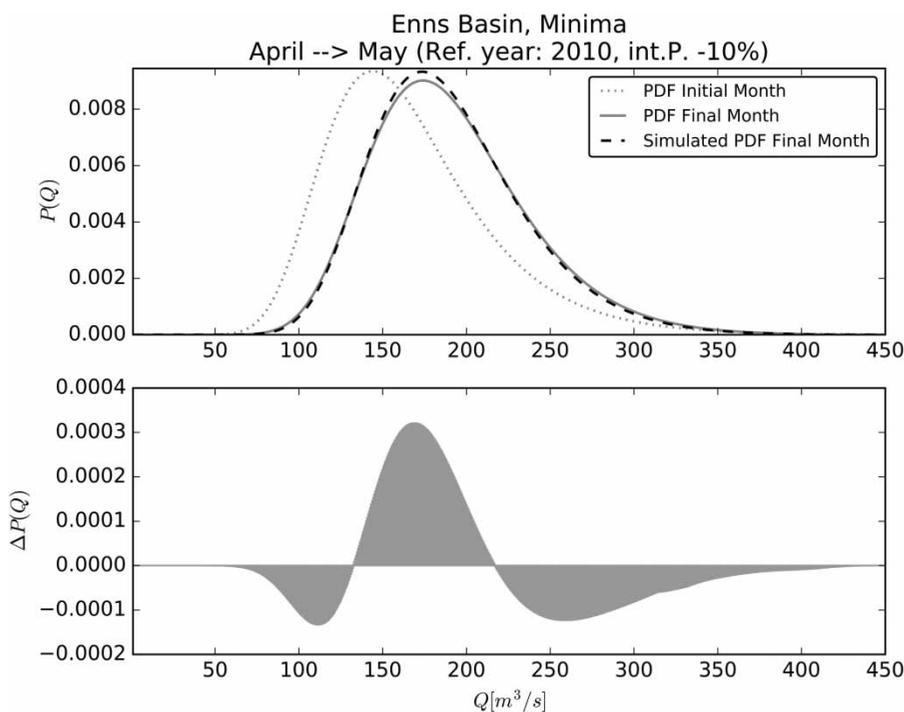

Figure 5 │ Simulation of the model applying change to internal parameters in the Enns basin, using the reference rainfall from 2010 and a change to internal parameters of -10%.
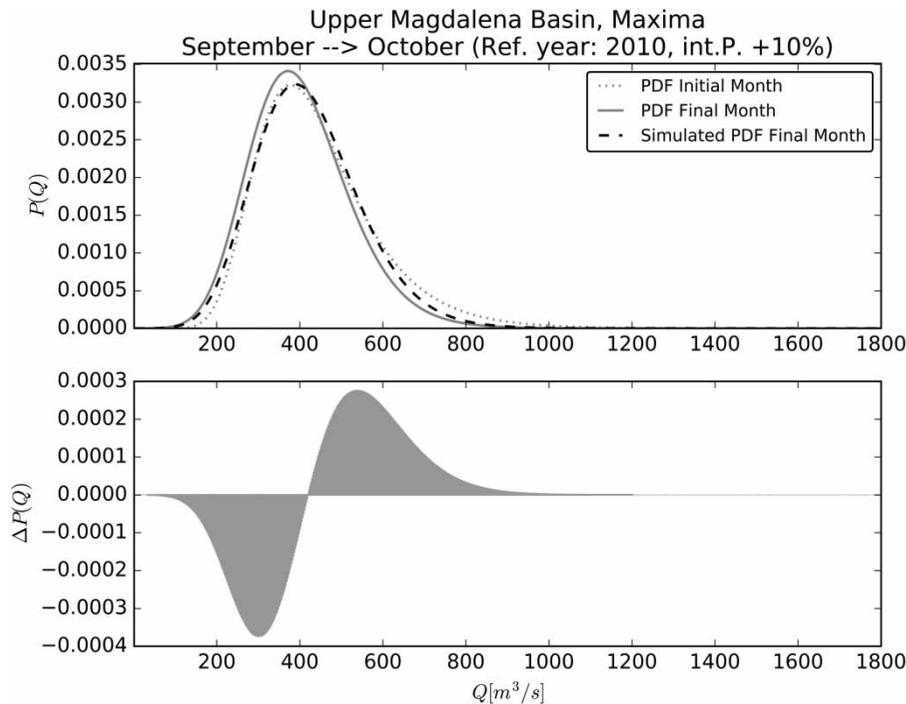
Figure 6 | Simulation of the model applying change to internal parameters in the Magdalena basin, using the reference rainfall from 2010 and a change to internal parameters of +10%.

Figure 7 shows that combining the simulated changes for internal and external parameters causes a greater impact on the result, due to the clearly visible move of the curve to the right. This indicates an increase of mean monthly discharge in the month of August in the Upper Magdalena basin and probably an increase of extreme events in the sense that they are defined for the existing data.

CONCLUSIONS

This paper was an attempt to stochastically model the regimes of extreme events in hydrometeorological time series and possible mechanisms of alterations using the FPK equation with a main focus on discharge data. It provided insight into many of its details, but at the same time opened new possibilities for further research on the topic.

It was possible to develop a stable algorithm that permits the simulation of the rainfall–runoff model with the one-dimensional FPK equation. The optimization procedure successfully completed the computation of the parameters for all variables in all basins with a natural runoff mechanism, although it was found that the model was extremely

sensitive to initial guesses. In a possible next step, a pre-optimization procedure could be applied to find more suitable initial guesses that provide better results.

The results of optimization and application of data to the model indicated very well that it can be used for river basins with a natural rainfall–runoff cycle. For all of the basins that rely on this characteristic, it was possible to calibrate the model and simulate future scenarios. The only exception was the Brisbane basin, where not even a calibration was possible. This implied that the model works well for natural basins, but cannot be applied to regulated ones, since the discharge data do not respond to the natural variations of the system anymore.

The proposed methodology of inverse modelling proved to be a powerful tool to effectively implement the studied problem. This way, the task of modelling a complicated physical process could be achieved without initially knowing its detailed structure. One conclusion is that this technique can most likely be applied to numerous other hydrological modelling problems, for example, in the assessment of hydrological risks, the study of hydrometeorological variability or further topics related to global climate change.
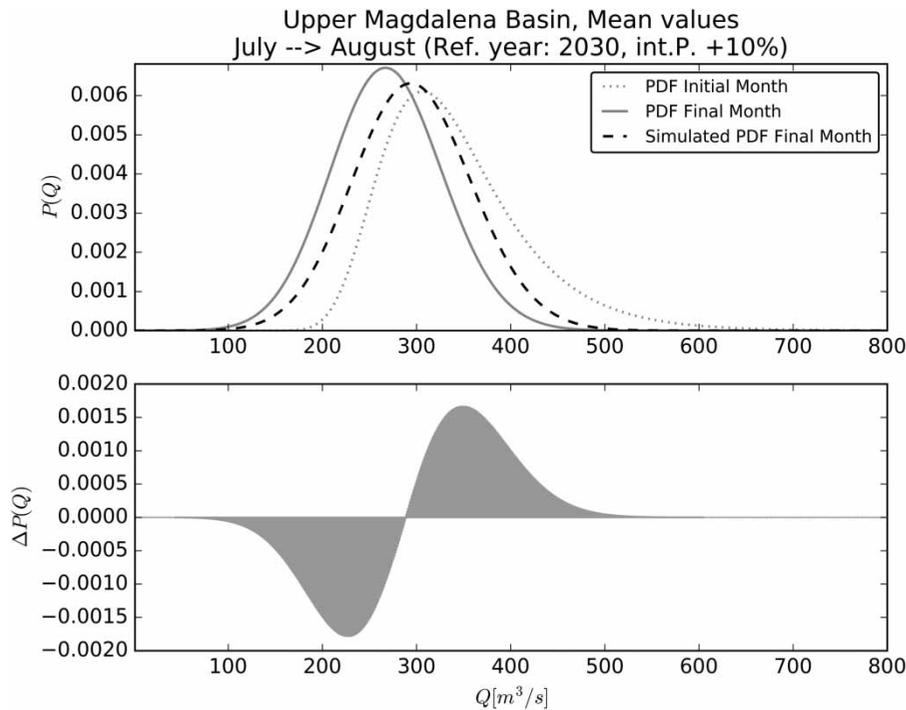
Figure 7 | Simulation of the model applying change to internal and external parameters in the Magdalena basin, using the extrapolated reference rainfall from 2030 and a change to internal parameters of +10%.

Finally, the results of simulation of future scenarios showed that it is possible to simulate alterations for changes in the basin's external and internal parameters and produced one key indication. It became evident that the changes to internal basin parameters have a bigger impact on the behaviour of the runoff process than those to external parameters, especially for extreme events.

These results and the fact that a large number of changes to internal parameters are caused by human activity reiterates the fact that it is up to us to control future extreme events by better understanding the internal mechanisms of river basins and taking the right steps to control our impact on them. It became apparent that some impact is caused by changes in external parameters, for example, precipitation changes which might be attributed to the influence of global climate change, but they are not as strong.

To completely understand all of the mechanisms driving the alterations in the behaviour of extreme events, further investigations are necessary that focus more in-depth on the topic. Especially studies on the relationship between the drift and diffusion parameters and the basin's land use

parameters as well as the geomorphometric characteristics should be encouraged, since no clear allocation of the model parameters to those internal parameters of the basins could be made in this study.

REFERENCES

Aguilar, E., Peterson, T. C., Ramírez Obando, P., Frutos, R., Retana, J. A., Solera, M., Soley, J., González García, I., Araujo, R. M., Rosa Santos, A., Valle, V. E., Brunet, M., Aguilar, L., Álvarez, L., Bautista, M., Castañón, C., Herrera, L., Ruano, E., Sinay, J. J., Sánchez, E., Hernández Oviedo, G. I., Obed, F., Salgado, J. E., Vázquez, J. L., Baca, M., Gutiérrez, M., Centella, C., Espinosa, J., Martínez, D., Olmedo, B., Ojeda Espinoza, C. E., Núñez, R., Haylock, M.,

Benavides, H. & Mayorga, R. 2005 Changes in precipitation and temperature extremes in Central America and Northern South America, 1961–2003. *Journal of Geophysical Research: Atmospheres (1984–2012)* **110** (D23). Available at: http://onlinelibrary.wiley.com/doi/10.1029/2005JD006119/full (accessed 10, November 2013).

Andrews, E. D. 1980 Effective and bankfull discharges of streams in the Yampa River basin, Colorado and Wyoming. *Journal of Hydrology* **46** (3–4), 311–330.

Calle, E. A. D., Miranda, J. A. M., Rodríguez, M. H. O., Martínez, J. F., Agudelo, C. A. R., Madriñan, L. F., Girón, J. B. & Baez, S. E. L. 2017 Objective assessment of ecosystem hydrological services in tropical areas: a Colombian experience in arid and semi-arid zones. *Revista Ambiente & Água* **12** (3), 365–379.

Cantet, P., Bacro, J.-N. & Arnaud, P. 2011 Using a rainfall stochastic generator to detect trends in extreme rainfall. *Stochastic Environmental Research and Risk Assessment* **25** (3), 429–441.

Dai, A., Qian, T., Trenberth, K. E. & Milliman, J. D. 2009 Changes in continental freshwater discharge from 1948 to 2004. *Journal of Climate* **22** (10), 2773–2792.

Denisov, S. I., Horsthemke, W. & Hänggi, P. 2009 Generalized Fokker-Planck equation: derivation and exact solutions. *The European Physical Journal B* **68** (4), 567–575.

Dodov, B. & Foufoula-Georgiou, E. 2005 Fluvial processes and streamflow variability: interplay in the scale-frequency continuum and implications for scaling. *Water Resources Research* **41** (5), W05005.

Dolgonosov, B. M. & Korchagin, K. A. 2007 A nonlinear stochastic model describing the formation of daily and mean monthly water flow in river basins. *Water Resources* **34** (6), 624–634.

Dominguez, E. & Rivera, H. 2010 A Fokker-Planck-Kolmogorov equation approach for the monthly affluence forecast of Betania hydropower reservoir. *Journal of Hydroinformatics* **12** (4), 486–501.

Dong, P. & Wu, X. Z. 2013 Application of a stochastic differential equation to the prediction of shoreline evolution. *Stochastic Environmental Research and Risk Assessment* **27** (8), 1799–1814.

Druzhinin, V. & Sikan, A. 2001 Stochastic processes. *Statistical Methods for the Treatment of Hydrometeorological Information* (A. M. Vladimirov, ed.). RGGMU, Saint Petersburg.

Feldman, R. M. & Valdez-Flores, C. 2009 *Applied Probability and Stochastic Processes.* Springer, Dordrecht.

Frolov, A. V. 2006 Dynamic-stochastic modeling of long-term variations in river runoff. *Water Resources* **33** (5), 483–493.

Gardiner, C. W. 2004 *Handbook of Stochastic Methods: for Physics, Chemistry & the Natural Sciences.* Springer, Berlin.

Haan, C. T. 1977 *Statistical Methods in Hydrology.* Iowa State University Press, Ames, Iowa.

Hayhoe, K., Wake, C. P., Huntington, T. G., Luo, L., Schwartz, M. D., Sheffield, J., Wood, E., Anderson, B., Bradbury, J.,

DeGaetano, A., Troy, T. J. & Wolfe, D. 2007 Past and future changes in climate and hydrological indicators in the US Northeast. *Climate Dynamics* **28** (4), 381–407.

Kazakievich, D. I. 1989 *Osnovi teoriy sluchainij funktsiiv v zadachax guidrometeorologuii (Fundamentals of the Theory of Random Functions in Hydrometeorological Problems).* Guidrometeoizdat, Leningrad, p. 230.

Koutsoyiannis, D. 2008 *Probability and Statistics for Geophysical Processes.* National Technical University of Athens, Athens.

Kovalenko, V. 1986 *Hydrometrical Assessment of Streamflow with an Stochastic Approach.* Leningradskiy Politekhnicheskiy Institut, Leningrad, p. 61.

Kovalenko, V. 2012 *Method of Characteristic Applied in Fractionally Infinite Hydrology.* RGGMU, Saint Petersburg, p. 134.

Kovalenko, V., Viktorova, N. V. & Gaidukova, E. 1993 *Modelling of Hydrological Processes.* Guidrometeoizdat, Saint Petersburg, p. 255.

Kullback, S. & Leibler, R. A. 1951 On information and sufficiency. *The Annals of Mathematical Statistics* **22**, 79–86.

Lawal, S. A., Watt, W. E. & Watts, D. G. 1997 A stochastic model of low flows. *Stochastic Hydrology and Hydraulics* **11** (4), 303–321.

Maldonado, C. E. 2009 *Complejidad: revolución científica y teoría (Complexity: Scientific Revolution and Theory).* Editorial Universidad del Rosario, Bogota.

Min, S.-K., Zhang, X., Zwiers, F. W. & Hegerl, G. C. 2011 Human contribution to more-intense precipitation extremes. *Nature* **470** (7334), 378–381.

Moré, J. J. 1978 The Levenberg-Marquardt algorithm: implementation and theory. In: *Numerical Analysis* (G. A. Watson, ed.). Springer, Berlin, pp. 105–116. Available at: http://link.springer.com/content/pdf/10.1007/BFb0067700.pdf (accessed 11 December 2015).

Naghettini, M., Gontijo, N. T. & Portela, M. M. 2012 Investigation on the properties of the relationship between rare and extreme rainfall and flood volumes, under some distributional restrictions. *Stochastic Environmental Research and Risk Assessment* **26** (6), 859–872.

Naidenov, V. I. & Podsechin, V. P. 1992 A nonlinear mechanism of water level fluctuations of inland reservoirs. *Water Resources* **6**, 5–11.

Naidenov, V. I. & Shveikina, V. I. 2002 A nonlinear model of level variations in the Caspian Sea. *Water Resources* **29** (2), 160–167.

Naidenov, V. I. & Shveikina, V. I. 2005 Hydrological theory of global warming of the Earth's climate. *Russian Meteorology and Hydrology* **12**, 46–56.

Nicholson, S. E., Nash, D. J., Chase, B. M., Grab, S. W., Shanahan, T. M., Verschuren, D., Asrat, A., Lézine, A.-M. & Umer, M. 2013 Temperature variability over Africa during the last 2000 years. *The Holocene* **23** (8), 1085–1094. Available at: http://hol.sagepub.com/content/early/2013/04/23/0959683613483618, abstract (accessed 8 November 2013).

Parry, M., Canziani, O., Paluitikof, J., van der Linden, P. & Hanson, C. 2007 *Climate Change 2007: Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Fourth Assessment Report of the IPCC*, Cambridge University Press. Available at: http://books.google.com.co/books?hl=en&lr=&id=TNo-SeGpn7wC&oi=fnd&pg=PA81&dq=Working+Group+II+Report+%22Impacts,+Adaptation+and+Vulnerability%22&ots=vO8BvfVusD&sig=m-QmHI2qZ9Me3PDx3iHvGrZaBgM (accessed 6 May 2013).

Rosmann, T. 2014 *Stochastic Modeling of Hydrometeorological Extremes and Their Possible Relation with Global Change*. Pontificia Universidad Javeriana, Bogota. Available at: http://repository.javeriana.edu.co/bitstream/10554/13103/1/RosmannThomas2014.pdf.

Rozhdenstvenskiy, A. V. & Chevotariov, A. I. 1974 Analysis of hydrological time series. *Statistical Methods in Hydrology* (Z. M. Kozhina, ed.). Guidrometeoizdat, Leningrad, p. 386.

Sveshnikov, A. A. 1966 *Applied Methods of the Theory of Random Functions,* Pergamon, Oxford.

Tabari, H., Taye, M. T. & Willems, P. 2015 Statistical assessment of precipitation trends in the upper blue Nile River basin. *Stochastic Environmental Research and Risk Assessment* **29**, 1751–1761.

Tao, H., Fraedrich, K., Menz, C. & Zhai, J. 2014 Trends in extreme temperature indices in the poyang lake basin, China.

Stochastic Environmental Research and Risk Assessment **28** (6), 1543–1553.

Unami, K., Abagale, F. K., Yangyuoru, M., Alam, A. H. M. B. & Kranjac-Berisavljevic, G. 2010 A stochastic differential equation model for assessing drought and flood risks. *Stochastic Environmental Research and Risk Assessment* **24** (5), 725–733.

UNESCO 2014 *The United Nations World Water Development Report 2014 – Water and Energy*. Paris.

van Rossum, G. & Drake, F. L. 2003 *An Introduction to Python*. Network Theory Ltd, Bristol.

Vargas, W. M., Minetti, J. L. & Poblete, A. G. 2002 Low-frequency oscillations in climatic and hydrological variables in Southern South America's tropical-subtropical regions. *Theoretical and Applied Climatology* **72** (1–2), 29–40.

Xu, Z., Liu, Z., Fu, G. & Chen, Y. 2010 Trends of major hydroclimatic variables in the Tarim River basin during the past 50 years. *Journal of Arid Environments* **74** (2), 256–267.

Yevjevich, V. 1987 Stochastic models in hydrology. *Stochastic Hydrology and Hydraulics* **1** (1), 17–36.

Zhao, G., Li, E., Mu, X., Wen, Z., Rayburg, S. & Tian, P. 2015 Changing trends and regime shift of streamflow in the Yellow River basin. *Stochastic Environmental Research and Risk Assessment* **29**, 1331–1343.