Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: https://www.tandfonline.com/loi/thsj20

An effective three-step algorithm for multi-site generation of stochastic weekly hydrological time series

Nesa Ilich

To cite this article: Nesa Ilich (2014) An effective three-step algorithm for multi-site generation of stochastic weekly hydrological time series, Hydrological Sciences Journal, 59:1, 85-98, DOI: 10.1080/02626667.2013.822643

To link to this article: https://doi.org/10.1080/02626667.2013.822643

Published online: 16 Dec 2013.

http://dx.doi.org/10.1080/02626667.2013.822643

An effective three-step algorithm for multi-site generation of stochastic weekly hydrological time series

Nesa Ilich

Principal Optimal Solutions Ltd, Calgary, Canada
nilich@optimal-solutions-ltd.com

Received 30 November 2012; accepted 11 June 2013; open for discussion until 1 July 2014

Editor D. Koutsoyiannis

Citation Ilich, N., 2014. An effective three-step algorithm for multi-site generation of stochastic weekly hydrological time series. *Hydrological Sciences Journal*, 59 (1), 85–98.

Abstract A new method is presented to generate stationary multi-site hydrological time series. The proposed method can handle flexible time-step length, and it can be applied to both continuous and intermittent input series. The algorithm is a departure from standard decomposition models and the Box-Jenkins approach. It relies instead on the recent advances in statistical science that deal with generation of correlated random variables with arbitrary statistical distribution functions. The proposed method has been tested on 11 historic weekly input series, of which the first seven contain flow data and the last four have precipitation data. The article contains an extensive review of the results.

Key words stochastic hydrology; correlated random variables; time series; empirical distribution

Un algorithme efficace en trois étapes pour la génération stochastique multi-sites de séries chronologiques hydrologiques hebdomadaires

Résumé Nous présentons une nouvelle méthode de génération de séries chronologiques hydrologiques multi-sites stationnaires. La méthode proposée peut gérer différentes durées du pas de temps et elle peut être appliquée à des séries d'entrée continues ou intermittentes. L'algorithme est différent des modéles de décomposition standard et de l'approche de Box-Jenkins. Il s'appuie au contraire sur les derniéres avancées statistiques traitant de la génération de variables aléatoires corrélées de fonctions de distribution statistique arbitraires. La méthode proposée a été testée sur 11 séries d'entrée historiques hebdomadaires, sept étant des données de débits, et quatre des données de précipitations. L'article présente une analyse approfondie des résultats.

Mots clefs hydrologie stochastique ; variables aléatoires corrélées ; séries chronologiques ; distribution empirique

INTRODUCTION

Ability to generate random times series that closely represent hydrological processes has been the main raison d'être of stochastic hydrology. Possible applications for using the stochastic series as alternative inputs abound, ranging from studying operation of reservoirs, drought management, alternative input into water quality studies, or design of new hydraulic structures within the multi-purpose stakeholder framework associated with modern water resources systems. Stochastic hydrology holds out a promise of providing input data variations that are statistically likely to occur, but are also more challenging to manage than those seen in the historic record, in terms of either the magnitude of individual events or their duration, as in the example of back-to-back dry years. This added challenge is implicitly contained in up to 1000 hypothetical years of generated data, which provide more reliability for designing new system components and for analysing performance of the entire system in various management scenarios (see e.g. Nalbantis and Koutsoyiannis 1997, Langousis and Koutsoyiannis 2006).

Some of the pioneering advances in the development of stochastic hydrology go back to the 1960s (Thomas and Fiering 1962, Matalas 1967). A turning point was the publication by Box and Jenkins (1970) on time series analyses, which has subsequently had a profound impact on research in this field, although this work was primarily motivated by applications of random time series in finance. Since both the cyclical and seasonal statistics of the generated time series should be similar to historic series, applications of Box–Jenkins models in hydrology were combined with the so-called *disaggregation* models. In this approach, annual flow series are generated first, making sure the annual statistics are on target, and then they are broken into seasonal (typically monthly) time steps using various disaggregation algorithms (Valencia and Schaake 1973, Mejia and Rousselle 1976, Koutsoyiannis 2001). A comprehensive review of the history of previous efforts is provided by Srinivas and Srinivasan (2005).

There is currently no universally accepted methodology or model for generation of stochastic time series that has gained widespread acceptance among hydrologists. The contending issues can be listed as follows:

(a) flexible time-step resolution: it would be ideal to have a model that can develop either monthly, weekly or daily stochastic series on the basis of the same historic daily input flow series;
(b) handling of dry regions which have legitimate zero flows in some periods and in some years, while, in other years, flows may be positive all year round;
(c) detection and inclusion of long-term annual cycles into modelling, originally propounded by Hurst (1957), known as the Hurst phenomenon and subsequently studied worldwide on the basis of associated indicators (see e.g. Langousis and Koutsoyiannis 2006). Closely related to this is the unresolved argument among researchers and practitioners on the importance of stochastic modelling of non-stationary processes and their proper representation (see e.g. Koutsoyiannis 2003, 2006, Koutsoyiannis *et al.* 2009).

This study presents a model that is restricted to the generation of stationary time series. While the existence of long-term cycles is acknowledged, especially in view of the recent studies on the effects of climate change, most hydraulic structures have a design lifetime of less than 150 years. A range of assumptions on the possible changes to the statistical distribution of historic flows can be incorporated in the generation of a family of stationary stochastic series, each representing a possible state with respect to the anticipated flow conditions in the near future. Inclusion of a trend, such as the steady reduction of runoff in a stochastic series that is 1000 years long, may actually lead to complete absence of any flows well before the 1000-year limit is reached. Consequently, modelling of stationary series adjusted to statistically represent the anticipated flows during the lifetime of a contemplated structure is typically preferred among practitioners.

Even when discussion is limited to stationary models, one cannot help but notice the lack of a generally accepted approach that would handle both the issues outlined under (a) and (b) above. After almost 50 years of research in this field and numerous publications, a stochastic model that is easy to understand, use, and widely accepted as the industry standard, does not yet exist. In other words, there is no stochastic time series generation model which would represent to hydrologists what HEC-RAS represents to river engineers. It could be argued that part of the problem may have been a too narrow focus on the combination of the decomposition principle coupled with the Box–Jenkins modelling approach. This approach has its limitations, while other radically different and promising ideas have perhaps been explored with a fraction of the effort so far. Multi-site generation models presented in the past have mainly been restricted to monthly time steps, typically using decomposition in combination with the AR(1) or ARMA(1) monthly model due to the difficulties of modelling auto-regressive dependence of higher orders. As documented by Bras and Rodriguez-Iturbe (1985), classical time series models involve significant effort and knowledge to identify the appropriate model and estimate its parameters, as well as to assess the shape of the multivariate probabilities and their transformations from normal to skewed distributions. Srinivas and Srinivasan (2005) point out that, in spite of the numerous reports on models in stochastic hydrology, none has gained universal acceptance. In fact, the US Bureau of Reclamation (USBR) uses a simple approach of recycling subsets of historical flow series, an approach known as the index sequential sampling (ISM) method (Kendall and Dracup 1991), rather than rely on any of the complex models which have been the subject of so much research in the past few decades. Lee and Salas (2008) experimented with using copulas in stochastic streamflow generation; however, they limited their work to the generation of annual series. While some of the ideas they presented were similar to the ideas in this article, they miss the opportunity to use the full potential of handling multi-site generation of monthly or weekly flows. Similarly, Tasker and Dunne (1997) have proposed a non-parametric method for multi-site generation of streamflow series, although they have restricted it to modelling of monthly residuals. A general, non-parametric approach independent of the time-step length or the type of the hydrological series to be modelled (continuous or intermittent) and free from cumbersome calibration and the disaggregation step has yet to emerge. The work presented here relies on recent advances in statistical science related to the generation of random variables with skewed distribution functions and a given correlation matrix which represents their statistical interdependence. This article extends these developments by adding one more step, which enables these models to generate random time series with given seasonality and periodicity, while preserving the desired skewed distributions and desired statistical dependence among them. Initially published in 2008 (Ilich and Despotovic 2008), the ideas were further improved in all three phases of the algorithm and verified on test runs that are significantly larger than those reported in 2008. The following section provides an introduction to the necessary statistical theory that serves as the background of the algorithm. This is followed by a more detailed description of each phase of the algorithm, while the final section provides an overview of two test runs which were based on hydrological data from southern Alberta, Canada.

THEORETICAL BACKGROUND

Simulation models used in a variety of industries, ranging from engineering to finance, often generate random variables with a desired statistical distribution function that are an essential component of the modelling process. Moreover, it is frequently necessary to generate several random variables that have different statistical distribution functions, while they are also statistically dependent, as defined by their correlation matrix. Iman and Conover (1982) published an article which provided a breakthrough in this kind of modelling. Numerous refinements to the idea followed (Cario and Nelson 1996, 1997), and the most recent one provides more control of the accuracy of the fit for each individual random variable (Ilich 2009). However, all of these algorithms are based on the notion that random variables with normal distribution and a desired correlation structure can be generated and that such random variables can then serve as a key to re-ordering the previously generated random variables with the appropriate skewed distribution, such that they too comply with the desired correlation structure. In other words, the desired correlation structure is achieved by permutation of the previously generated skewed random variables, while the initial matrix with uniform distribution serves as a guide to permutations. This is essentially an iterative process, since some accuracy related to matching the target correlation matrix is lost in the process of replacing the initially generated uniform variables by the respective random variables with skewed distribution. Like any other iterative process, the solutions are not exact and are subject to user-defined convergence criteria. However, the idea is simple and powerful, and reasonable solutions can be obtained for hundreds of random variables simultaneously within minutes.

In their work, Iman and Conover (1982) describe an algorithm for re-ordering elements of random vectors with desired skewed distributions such that their Spearman rank correlation matrix is the same as the desired target. In this application, the work of Iman and Conover has been modified to handle Pearson's correlation matrix as the desired target, rather than the Spearman rank correlations. The principle is similar, with a somewhat larger loss of accuracy when fitting the Pearson correlations. There are two ideas which are essentially the basis of the proposed algorithm. The first is a theorem originally proved by Whitt (1976), which states that the highest correlation among two independent random vectors is obtained if they are both re-arranged in a sorted order. This can be used effectively in hydrology. For example, if a desired probability distribution function of historic flows y_i in week i for a particular site is known based on historic data, we can then generate n independent random variables Y_i of these flows that fit this distribution. Assume that n random sample flows X_i have already been generated for the previous week and for the same site. It is possible to generate n estimates of hypothetical flows Y_i' by using the known regression parameters in an autocorrelation function developed on the basis of historic flows x_i and y_i in two subsequent weeks, as shown in equation (1):

$$Y_i' = a_o X_i + a_1 + N[0, \varepsilon_y] \tag{1}$$

There are two shortcomings when it comes to autoregressive flows Y'_i. The first is that, for a large sample size, Y'_i has a normal distribution, instead of the desired known distribution which is typically skewed. The other is that there is a possibility that some sampled values of Y'_i are negative, due to the normalized random term N, which can be negative. Neither of those two outcomes is desirable in hydrology. However, one useful property of autoregressive flows Y'_i is that they have a desired correlation to flows X_i from the previous week. If we re-arrange the elements of the original flows Y_i such that they have the same order as the elements of Y'_i, then the correlation structure between X_i and Y_i will be close to the target developed from historic data, while the resulting flows Y_i will preserve their original distribution which excludes negative values. Re-arranging the elements of Y_i is a simple process of finding the rank k for the minimum value in Y'_I, placing the minimum value of Y_i in the kth position, and proceeding in the same way for all other elements, from the second smallest to the largest. In other words, the elements of Y_i are permuted using the rank order of the elements of Y'_i as their permutation key. Since some of the dispersion may be increased once all elements of normally distributed Y'_i have been replaced by the elements of a skewed distribution Y_i, it may be necessary to assume a small reduction in the initial value of the random term N, such that the final correlation between X_i and Y_i fits the desired target. The proper value of the standard deviation of N can be found iteratively. Once the desired permutation of X_i and Y_i has been achieved, the model can use multiple regression to proceed, such that flows X_i and Y_i from the first two weeks are independent variables, while the flow Z'_i in the third week is dependent, and the transformation between Z_i and Z'_i can proceed in the same manner as for Y_i and Y'_i. This is a flexible approach that has been tested successfully on a number of skewed distributions and a combination of both positive and negative correlations (Ilich 2009). However, the method of Iman and Conover (1982) is more robust and provides quicker solutions to large-scale problems, although it does not have the same kind of flexibility for achieving fitness of each individual random variable. In spite of this, the proposed method has several advantages over the earlier method of Ilich (2009):

(a) Each iteration resets all target correlations within the variables that have not managed to converge in the previous iteration, as opposed to the much slower convergence process that had to tackle each variable (weekly flow) on an individual basis. This results in solution times that on large-scale test runs are up to two orders of magnitude faster than the previous version.

(b) This approach does not ignore correlations that are below the user-specified threshold, as opposed to the earlier method, since this method automatically attempts to fit the entire correlation matrix simultaneously. This avoids the sudden drop between the correlations of the simulated series that are above the user-defined threshold (e.g. 0.5) and below, which was a shortcoming of the previous algorithm.

(c) Other improvements in Phase 3 of the previous algorithm allow more flexibility when it comes to the selection of the final sequence of years, such as the new weight factors that can give more importance to stations that exhibit slower convergence, or parameters that alter the significance of annual auto-correlation *versus* the correlation of weekly flows that are associated with transition from year i to year $i + 1$.

Instead of dealing with each random variable individually, the method of Iman and Conover proceeds on the premise of the development of an entire random matrix with a desired correlation structure, and then uses the columns of this matrix as a key to re-ordering individual random variables that had previously been generated independently on the basis of their distributions. They begin by defining an n-dimensional random vector $\boldsymbol{X}$ as a vector whose correlation matrix is $\boldsymbol{I}$ (that is, elements on the main diagonal are 1, while all other elements are zero, implying that elements of $\boldsymbol{X}$ are not correlated). Let $\boldsymbol{C}$ be the desired correlation matrix of some transformation of $\boldsymbol{X}$. A new matrix $\boldsymbol{X'} = \boldsymbol{XP'}$ can be created since $\boldsymbol{P'}$ can be determined from the desired correlation structure $\boldsymbol{C}$ using a Cholesky factorization. This results from a known theorem in statistical science which served as the basis for the algorithm proposed by Iman and Conover (1982). This theorem is referenced in their article and covered in detail by the background work cited therein. Matrix $\boldsymbol{X'}$ can then be used as a key for re-ordering any matrix of the same size to fit the desired correlation structure. This is the basis of the algorithm, and it implies an iterative procedure, since after recalculating the correlation matrix of the original elements that had been permuted, the resulting correlation matrix may not be exactly the same as the target correlation matrix. This may require additional

Year	STATION 1					STATION 2					STATION 3				
	Weeks					Weeks					Weeks				
	1	2	.	.	52	1	2	.	.	52	1	2	.	.	52
1933	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	.	$X_{1,52}$	$X_{1,53}$	$X_{1,54}$	.	.	$X_{1,104}$	$X_{1,105}$	$X_{1,106}$	$X_{1,107}$	.	$X_{1,156}$
1934	.														.
1935	.														.
1936	.														.
1937	.														.
1938	.		.	.	.	.	.	$X_{i,j}$	.	.	.	.	.	.	.
.	.														.
.	.														.
.	.														.
1999	.														.
2000	.														.
2001	.														.
2002	.	.	.	.	.	.	.	.	.	.	.	.	.	.	$X_{m,n}$

Correlation Matrix

variable	STATION 1					STATION 2					STATION 3				
	Variable number					Variable number					Variable number				
	1	2	3	.	52	53	54	.	.	104	105	106	.	155	156
1	1	$\sigma_{1,2}$	$\sigma_{1,3}$		$\sigma_{1,52}$	$\sigma_{1,53}$	$\sigma_{1,54}$			$\sigma_{1,104}$	$\sigma_{1,105}$	$\sigma_{1,106}$	$\sigma_{1,107}$		
2	$\sigma_{1,2}$	1	$\sigma_{2,3}$		$\sigma_{2,52}$	$\sigma_{2,53}$	$\sigma_{2,54}$			$\sigma_{2,104}$	$\sigma_{2,105}$	$\sigma_{2,106}$	$\sigma_{2,107}$		
3	$\sigma_{1,3}$	$\sigma_{2,3}$	1										$\sigma_{3,107}$		
4				1											
5					1										
6						1									
.							1								
.									.						
.									.						
.									.						
.									.						
155														1	
156															1

Fig. 1 Matrix representation of generated data.

iterations such that the target correlations that were not fitted well initially are modified to ensure better fitness. Consider the hypothetical results of generated weekly flow data for three sites and their target correlation matrix in Fig. 1, which depicts the generated data in a matrix format along with the corresponding correlation matrix below. Note that since it is well known that the correlation matrix is symmetrical around the main diagonal whose elements are set to 1, it is common practice in the literature to show the correlation matrix as the upper or lower triangular matrix for reasons of clarity.

Columns X_j in this matrix represent flows generated individually for each week and for each of the three stations, based on their respective statistical distribution functions. How these functions are constructed is explained in more detail in the next section of this article. Assume that the flows in each of the weeks have been successfully permuted such that the resulting correlation matrix is close to the target correlation matrix, which is calculated on the basis of historic natural flow data. This has several profound implications:

(a) A good match of correlation $\sigma_{1,2}$ implies preserved auto-correlation between flows in week 1 and 2 at the first station. Similarly, a good match of correlations $\sigma_{53,54}$ and $\sigma_{105,106}$ implies preserved auto-correlation of the first order for weeks 1 and 2 at Stations 2 and 3, respectively.

(b) A good match of correlation $\sigma_{1,3}$ and $\sigma_{2,3}$ implies preserved auto-correlation of both first and second order for the first station, i.e. flows in week 3 for Station 1 are auto-correlated to flows in both weeks 1 and 2 for the same station. The same principle holds for the other two stations, and this observation extends to as many significant lags that may be found in the historic data based on the values of the correlation coefficients.

(c) In addition to auto-correlation, cross-correlation between different stations is also preserved. For

> example, a good match of $\sigma_{1,53}$ and $\sigma_{1,105}$ implies preserved cross-correlation for week 1 between flows on Station 1 and 2 and between Stations 1 and 3, respectively. Moreover, cross-correlations are also preserved with any significant lag that may be found in the historic series data, a feature that has been very difficult to fit in other algorithms.

It is apparent that the above algorithms can generate weekly flows with a desired statistical distribution that is unique for each week, thus capturing the seasonal nature of hydrological data, while at the same time preserving all statistical interdependence that may exist between the data within a year. The resulting variables do not represent time series yet, since the statistical dependence of flows at the end of a particular year and the beginning of the subsequent year have not yet been established. The principal premise of the proposed algorithm is to find the appropriate permutation of the entire rows X_i of the above matrix such that this condition can also be satisfied. For a 1000 synthetic years of data, there are 1000! combinations of possible sequences of rows, so there are likely many solutions that would fit this condition. A suitable sequence of years should fit both the weekly statistics for all stations, as well as the annual auto-correlation functions for any significant lag. This essentially is the final step of the proposed algorithm. More details on all three steps are available in the following section. It should be noted that the original work of Iman and Conover (1982) has been modified here, by using the Pearson product moment correlation instead of the Spearman rank order correlation. This does not alter the theoretical basis of the approach, which is applicable to any type of correlation matrix.

PROCEDURE

The basic idea of the algorithm involves three distinct steps which are independent of each other, such that each subsequent step builds on the solution from the previous step without any negative impacts on it. These steps are explained below.

Step 1: generate random variables

The first step generates randomly 1000 years of weekly stochastic flows for each station that has the desired statistical distributions. This could be done by treating the data for each week as a sample for an independent variable and by fitting a theoretical statistical distribution function using the maximum likelihood approach and then running a Monte Carlo simulation for the chosen distribution functions and its estimated parameters. However, theoretical functions do not always fit the available data well. A non-parametric approach that uses empirical distribution functions (Lall 1995) has been emerging recently, with a potential to eliminate possible problems in fitting theoretical functions to historical data samples. One of the most popular is the kernel density function, which is defined as a weighted moving average of the empirical frequency distribution of the available data (Sharma *et al.* 1997). The result is a distribution function guaranteed to fit the historical data in the probability range for which the data are available, and this virtually eliminates the need to run the goodness-of-fit tests required for fitting theoretical distributions. Much of the on-going research in non-parametric density functions has been focused on the handling of their tail ends. Lall (1995) provides a summary of several reported approaches, of which the one proposed by Moon *et al.* (1993) has been used in this work with minor modifications. This approach employs theoretical distributions commonly used for rare events, such as extreme value or log-normal, to fit the tail ends of the statistical distribution function. The fitting algorithm proposed by Moon was adjusted in this case to ensure smooth transition from the empirical to the theoretical part of the curve, by finding the intersection of the two curves (theoretical and empirical) and by smoothing out the region in the vicinity of the intersecting point if necessary.

Figure 2 shows a comparison between log-normal and empirical distributions. The central region of the graph, between the probabilities of 0.3 and 0.7, shows a reasonably good fit with the observed data for both distributions. However, log-normal overestimates flows for regions below probability of 0.3 and the flows are slightly underestimated for probabilities between 0.7 and 0.9. On the high-flow side, the log-normal distribution would provide a good tail end of the empirical curve. On the low-flow tail end of the curve, the fit is based on using the extreme value fit for low flows with return periods of 100, 200, 500 and 1000 years, which are based on the available data (there were 84 years of historic data for the numerical examples presented in this article). Hence, to generate 1000 years of hypothetical flows for week 20 at this station, the model generates 1000 random numbers (probabilities) with uniform distribution between [0,1] and reads the appropriate flow value for each of the generated probabilities.

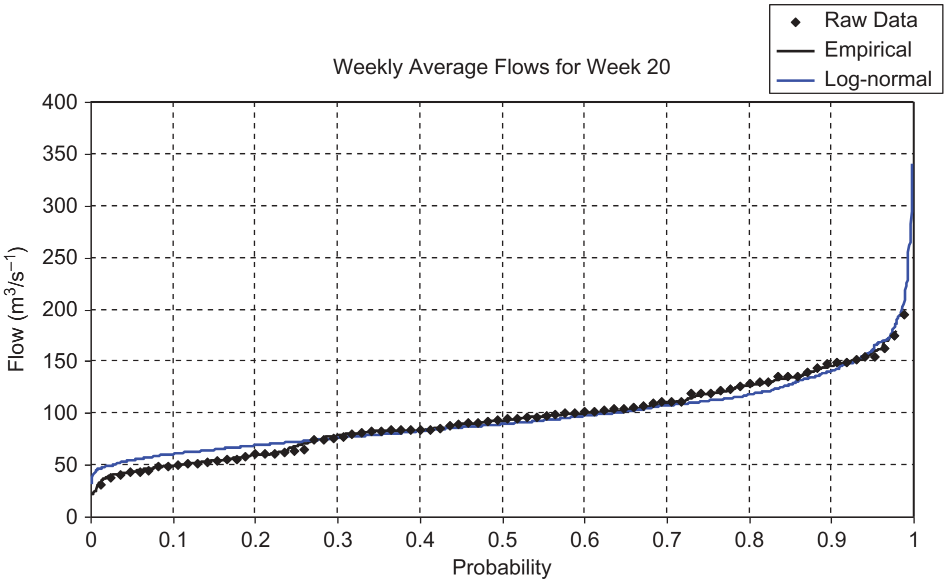

Fig. 2 Empirical and theoretical distributions of weekly flows.

Perhaps the most significant advantage of using the empirical distribution is when it comes to generating stochastic data on rivers that may legitimately dry out in some years—assume, for example, a case where the flows are zero 30% of the time for a particular week. While empirical distributions handle this case with ease, eliminating the need to engage in any functional fitting of the lower tail end, theoretical distributions routinely fail to address this situation. Step 1 is finished when 1000 years of hypothetical flows are generated for all weeks and for all stations using the above approach. These flows are independent of each other. The following two steps will introduce the necessary permutations of these flows to induce the necessary statistical dependence. Permutations only happen within the data generated in a particular week, which implies that the statistical distributions of the generated samples remain intact.

It should be noted that a close match in terms of weekly statistics, such as mean, standard deviation, skew and the overall fit of the weekly probability distribution functions, also implies a close match between related annual statistics. Along with the mechanisms incorporated in Step 3 explained below, this eliminates the need for separate generation of annual time series and subsequent use of decomposition.

Step 2: induce desired correlation structure

In this step, individual weekly flows are permuted until a permutation for all flows with the desired correlation structure is found. Assume that the weekly flows for 1000 synthetic years created in Step 1 have been saved in matrix $\boldsymbol{R}[n,1000]$, where n is the product of the number of time steps within a year (52) and the total number of stations where flows are generated. Step 2 is then based on the variation of the algorithm by Iman and Conover (1982), and it consists of the following:

(a) Generate a uniform random matrix $\boldsymbol{X}[n,1000]$ with all elements that are uniformly distributed, i.e. $X[i, j] \sim U[0,1]$. Elements of this matrix have no mutual correlation.

(b) Calculate the target correlation matrix $\boldsymbol{C}$ based on the available historic weekly flows for all weeks and all stations. Skip the weeks with missing data or zeros for those correlations that are affected (in other words, use all available years of data to calculate individual correlation coefficients).

(c) Calculate the elements of matrix $\boldsymbol{P'}$ where $\boldsymbol{C} = \boldsymbol{PP'}$. This amounts to finding the square root of the correlation matrix $\boldsymbol{C}$ using Cholskey factorization. There are commercial DLL libraries that can be called by the model to provide smooth execution of this transformation.

(d) Create matrix $\boldsymbol{X'} = \boldsymbol{XP'}$.

(e) Create matrix $\boldsymbol{R'}$ by re-arranging the orders of the elements of matrix $\boldsymbol{R}$ created in Step 1 such that they have the same ordering as matrix $\boldsymbol{X'}$. This is achieved in the following way: (i) find the rank of each element along each column of matrixes $\boldsymbol{X'}$ and $\boldsymbol{R}$; and (ii) re-arrange elements of matrix ***X*** along each column such that the rank order within each column of matrix ***X*** is the same as the rank order of the elements in the respective column of matrix ***R′***. In other words, if an element in the first column of matrix ***X′*** had a rank *j*, then find the element in the first column of matrix ***R*** with the same rank *j* and place it in the same row that corresponds to the row number of the smallest element in the first column of matrix ***X***. Repeat the process for all elements of matrix ***X*** in the first column, and then proceed with all other columns in this fashion.

This procedure is used by Iman and Conover as part of their algorithm; they refer to this step as "the use of elements of matrix ***X′*** as a *key to re-ordering* the positions of the elements of matrix ***R***, resulting in matrix ***R′***". Due to the theorem of Whitt (1976), mentioned earlier in this article, the correlation matrix calculated among the variables of matrix ***R′*** should have a close structure to the target correlation matrix ***C***.

(f) Since the column elements of ***R*** are not distributed uniformly, as are the elements of matrix ***X′***, some deviation from the target correlations is inevitable. In the final step of the algorithm, the level of deviation is determined, and if it is not acceptable a new target correlation matrix is set with some correlations that are higher than the original targets.

Then steps (b)–(f) are repeated. Hence, calculate the correlation matrix ***C′*** of the transformed matrix ***R′***. If individual correlations are found with a difference from the target correlations, create a new target correlation matrix ***C″*** which is a merger of the original target correlations from matrix ***C*** with some of its elements modified using a procedure explained by the following pseudo code:

$$\text{if}\{C'[i,j] < (C[i,j] - \alpha)\}$$

$$\text{then } C''[i,j] = \min\{C[i,j] + (\mathrm{C}[i,j] - C'[i,j]), 0.99\} \tag{2}$$

$$\text{else } C''[i,j] = C[i,j]$$

The above threshold, *α*, is set by the user; the suggested value is typically 0.05, while the upper bound on setting the target is set to 0.99 given that 1.0 is the maximum possible target correlation. Steps (c)–(f) are executed repeatedly and the process ends when no new improvements are possible or after a user-specified number of iterations.

As with most other numerical procedures, simultaneous convergence on a large number of variables is an issue. Most articles published in statistical and mathematical journals cover this issue in greater detail (Cario and Nelson 1996, 1997), but provide no universal recipe on how to ensure smooth convergence. The problem arises because some of the correlations that have converged in previous iterations may violate the convergence criteria in subsequent iterations. This remains an area of active research in mathematics and statistics. A typical threshold is the minimum correlation examined for convergence of 0.5 or 0.4, since correlations below those levels are not considered significant in hydrology, and they are usually satisfied in this algorithm by default since they are easy to achieve. Another user-defined model parameter is the convergence criterion, which represents the difference between the target correlation calculated from historic data and the final correlation achieved by the permutation of the generated weekly flows. This parameter was usually set to 0.05, implying successful convergence if the difference between the two correlations is less than 5%. The entire execution of Step 2 for this problem takes about 10 s per iteration, and the number of iterations is typically less than 5. Step 1 executes in less than 3 s. This approach to Step 2 is much more efficient than in the previous algorithm (Ilich and Despotovic 2008), since the iterative steps are applied on all variables at once (i.e. the entire matrix), as opposed to each individual weekly flow, as was the case earlier. The solution times for this step are now faster than before by an order of magnitude, even for medium-size problems.

Step 3: convert generated random variables to time series

Once the transformations in Step 2 have been completed, statistical dependence is established for all random variables generated by the model within the 52-week periods. Consequently, flows in week 52 are correlated to the flows in all previous weeks, including both auto-correlations and cross-correlation among the stations. What remains to be imposed is the transition from year to year. In other words, the entire years of generated data have to be rearranged such that their sequence guarantees that the relevant weekly and annual statistics will be preserved for the weeks that are associated with crossing from year to year. In addition to the weekly statistics, the annual auto-correlations should also be preserved, although these auto-correlations usually show much lower statistical dependence, as will be demonstrated by the numerical example in the subsequent section. The problem can be described mathematically as combinatorial optimization that minimizes the following objective function:

$$D_k = \sum_{p=1}^{L}\sum_{q=n-p}^{n-L}(\rho^k_{q,p} - \sigma^k_{q,p})^2 + \sum_{l=1}^{m}(\mathrm{ACH}^k_l - \mathrm{ACG}^k_l)^2 \qquad (3)$$

where n is the total number of time steps in a year (52 for weekly flows); $\rho^k_{q,p}$ is the historic weekly auto-correlation between weeks q and p for station k; $\sigma^k_{q,p}$ is the generated weekly auto-correlation between weeks q and p for station k; L is the weekly lag taken into account in transition from year to year; ACH^k_l is the historic annual auto-correlation function for lag-l and station k; and ACG^k_l is the generated annual auto-correlation function for lag-l and station k.

A visual example of the correlations that are included in the first summation term in equation (3) is shown by the shaded "window" in Table 1. For example, the first week of the subsequent year has correlations of 0.72, 0.37, 0.27 and 0.28 with weeks 52, 51, 50 and 49, respectively. For generated series, only correlations located within the Table 1 were fixed in Step 2 and they remain unchanged in Step 3.

To extend the above objective function to multiple stations, introduce a composite statistic D as a sum of all individual terms D_k:

$$D = \sum_{k=1}^{r} D_k \qquad (4)$$

where r is the total number of stations being modelled simultaneously. The objective of Step 3 is to minimize D. It is also possible to define expression (3) in other ways, without using the square function; the absolute value or a simple difference between correlations of the generated series and their corresponding historic targets may work as well. Performance of the proposed algorithm is demonstrated using a numerical example.

Table 1 Historic correlations between two subsequent years for Station C5AE27.

Week	49	50	51	52	1	2	3	4
49	1.00	0.96	0.88	0.58	0.28	0.30	0.37	0.41
50		1.00	0.91	0.59	0.27	0.27	0.36	0.42
51			1.00	0.73	0.37	0.33	0.39	0.43
52				1.00	0.72	0.59	0.50	0.50
1					1.00	0.94	0.65	0.56
2						1.00	0.71	0.60
3							1.00	0.88
4								1.00

NUMERICAL EXAMPLE

The example presented in this article provides 1000 years of synthetic weekly flows at seven sites and synthetic rainfall intensities on four sites in the Oldman River basin in the Province of Alberta, western Canada. This basin has been subject to many studies because intense water abstractions and the requirement to pass 50% of natural flow to the downstream province of Saskatchewan cannot be simultaneously satisfied in dry years. Consequently, basin management that includes hedging of demand in combination with the optimal reservoir operating rules can benefit from stochastic inflow series that contain 1000 years of statistically possible inflows. These include challenging periods with back-to-back dry years, as well as extreme droughts and wet years that have not been seen in the historic record. Using the stochastic series as an alternative input in basin management models provides more opportunities to examine how the entire system should respond to a variety of runoff conditions in the basin. A schematic model of the basin is shown in Fig. 3, with inflow locations shown as sites where the stochastic flow series were generated. The study region involves the Oldman River in southern Alberta, with two of its major tributaries, the Waterton River and St Mary River. There are three storage reservoirs and several irrigation districts, represented as five composite water-demand blocks in Fig. 3. Two test runs are presented, one with seven flow series and the other with 11 series in total, of which seven are flow series and four are precipitation series available for the

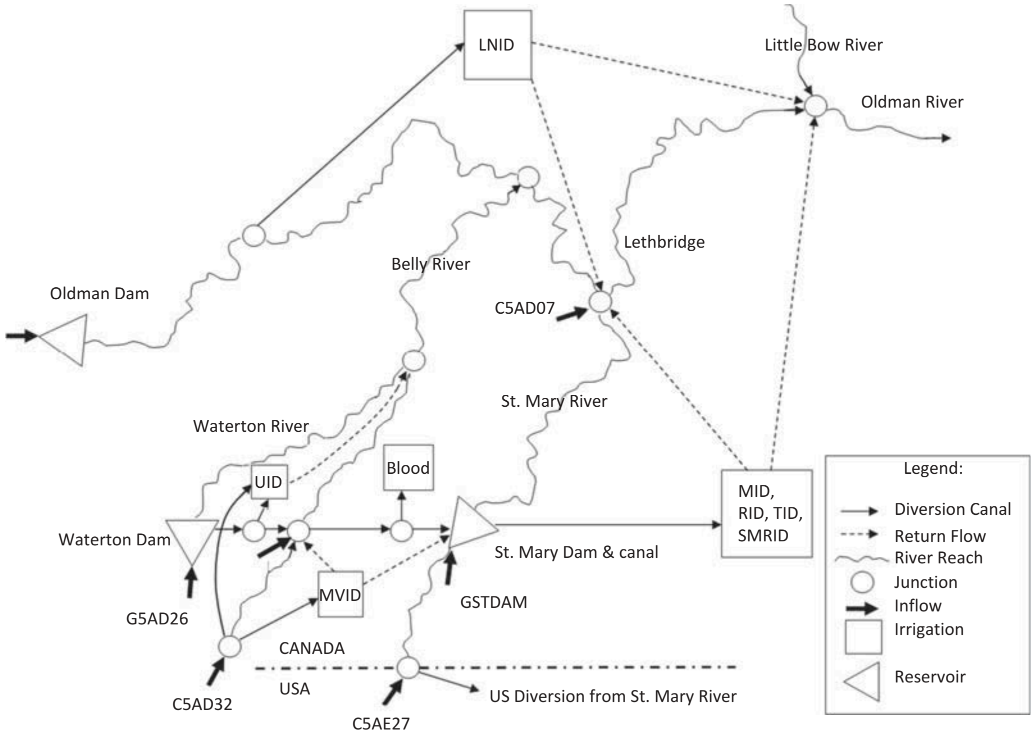

Fig. 3 Schematic of the Oldman River basin in southern Alberta.

same period. While there are no visible differences in the results between the two runs, the results are presented in a composite form for both test runs. For example, the annual flow statistics in both runs are virtually identical in both runs, as seen in Table 2.

It should be noted that the proposed algorithm did not include any direct fitting of annual statistics; their fit is only achieved as a by-product of fitting the weekly statistics. Yet, the annual means are reasonably close. Standard deviations in the generated series are slightly higher, especially for rainfall series, due to significantly longer generated series that include rare events which increase standard deviations. However, deviations from the historic values are modest. In addition to these statistics, Tables 3 and 4 confirm that annual cross-correlation among the stations and annual auto-correlations are similar in the generated and historic data. This effectively removes the need to generate the annual series separately and subsequently apply a decomposition principle, which has so far been the usual practice. Again, note that the algorithm does not directly fit annual cross-correlations between the stations. Those come as a by-product of a high fit of all historic weekly correlations. Table 3 shows annual cross-correlations between all 11 data series. It should be

Table 2 Annual statistics of historic and generated data series.

	Flow statistics ($m^3 s^{-1}$)							Precipitation statistics (mm)			
Historic data											
Max	39.53	48.79	10.71	16.68	45.47	75.83	222.93	962	714	1195	663
Min	14.06	10.60	3.80	5.35	11.35	14.06	45.10	158	131	401	187
Mean	25.10	27.28	6.63	9.47	22.09	39.66	109.19	497	396	710	355
SD	5.69	7.45	1.47	2.30	5.97	12.96	35.94	147	101	168	96
Generated data											
Max	57.15	70.45	13.87	22.06	53.11	106.94	308.20	1177	950	1791	889
Min	10.99	10.24	3.36	4.57	8.11	12.61	35.72	118	111	199	73
Mean	25.01	27.08	6.59	9.43	21.97	39.23	108.46	486	387	696	347
SD	5.97	7.71	1.48	2.33	6.19	12.69	35.81	186	136	247	123

Table 3 Annual cross-correlations of historic and generated data.

	Flow stations							Precipitation stations			
	C5AE27	GSTDAM	C5AD32	C5AD41	G5AD26	C5AA24	C5AD07	Cardstone	Lethbr.	M.View	Taber
Historic data											
1	1.000	0.964	0.966	0.957	0.944	0.831	0.881	0.542	0.403	0.482	0.347
2		1.000	0.924	0.958	0.948	0.865	0.914	0.582	0.493	0.539	0.372
3			1.000	0.968	0.926	0.822	0.869	0.561	0.409	0.530	0.384
4				1.000	0.948	0.873	0.930	0.616	0.482	0.586	0.436
5					1.000	0.906	0.941	0.571	0.480	0.535	0.391
6						1.000	0.966	0.604	0.522	0.547	0.430
7							1.000	0.598	0.513	0.561	0.431
8								1.000	0.699	0.801	0.778
9									1.000	0.631	0.696
10										1.000	0.644
11											1.000
Generated series											
1	1.000	0.957	0.959	0.949	0.937	0.827	0.873	0.536	0.394	0.437	0.369
2		1.000	0.913	0.951	0.940	0.855	0.906	0.559	0.442	0.474	0.390
3			1.000	0.959	0.922	0.820	0.861	0.566	0.410	0.481	0.407
4				1.000	0.944	0.865	0.922	0.581	0.447	0.494	0.427
5					1.000	0.902	0.937	0.568	0.459	0.483	0.417
6						1.000	0.954	0.598	0.497	0.492	0.444
7							1.000	0.582	0.471	0.481	0.431
8								1.000	0.764	0.787	0.799
9									1.000	0.705	0.805
10										1.000	0.735
11											1.000

noted that the flow series are all mutually correlated with high correlation coefficients. Somewhat lower but still significant is the mutual correlation between the precipitation stations, while the correlations between mean annual flows and annual precipitations are the weakest.

Similarly, Table 4 shows the annual auto-correlations for the historic and generated series. Since the historic annual auto-correlation of precipitation stations is very close to zero for all inspected lags, it has not been included in Table 4. Long-term annual cycles have been subject to much controversy in hydrology, and especially in stochastic hydrology. However, as attested by the historic annual auto-correlations in Table 4, their level of significance in this basin does not appear to be an important

Table 4 Annual auto-correlations of historic and generated data.

	C5AE27	GSTDAM	C5AD32	C5AD41	G5AD26	C5AA24	C5AD07
Historic data							
Lag-1	0.116	0.203	0.102	0.138	0.145	0.082	0.174
Lag-2	–0.040	0.049	–0.063	0.029	0.014	0.013	0.084
Lag-3	0.064	0.074	–0.005	0.013	0.092	0.085	0.073
Lag-4	0.074	0.056	0.116	0.069	0.087	0.051	0.020
Lag-5	0.111	0.089	0.083	0.081	0.133	0.174	0.129
Lag-6	0.039	0.035	0.057	0.036	0.092	0.016	0.039
Lag-7	–0.014	–0.022	–0.006	–0.052	–0.083	–0.097	–0.105
Lag-8	0.035	0.021	0.042	–0.023	0.048	–0.024	–0.037
Generated series							
Lag-1	0.104	0.115	0.112	0.128	0.121	0.087	0.103
Lag-2	–0.017	–0.006	–0.020	–0.018	–0.003	–0.006	0.000
Lag-3	0.007	0.020	0.005	0.012	0.016	0.023	0.028
Lag-4	–0.001	–0.002	0.019	0.012	–0.012	–0.021	–0.012
Lag-5	0.011	0.037	0.012	0.019	0.024	0.044	0.029
Lag-6	0.046	0.000	0.049	0.008	0.006	–0.021	–0.019
Lag-7	–0.050	–0.069	–0.062	–0.060	–0.051	–0.027	–0.056
Lag-8	–0.034	–0.011	–0.040	–0.014	–0.017	0.004	0.009

factor. Still, Table 4 is presented to show that the algorithm is capable of addressing proper sequences of years when re-arranging the final sequence of years in Step 3. In fact, the model offers enough flexibility to assign a user-defined parameter that may give higher weight to some stations within the objective function formulation, if it is found that convergence for those stations lags behind other stations in the process of fitting the target annual auto-correlation statistics.

Presentation of weekly statistics is more challenging, due to the sheer volume of data. The model achieves a very close match of weekly means and standard deviations between the historic and generated series, along with the weekly distribution functions used to generate data in Step 1, based on the previously explained concepts. Showing the correlation match between the historic target correlation matrix and the correlation matrix of generated series is not possible here (the matrix rank is 52 × 11 = 572). However, it is interesting to see the correlations that cross over from year to year. They are similar for all flow stations (and close to zero for precipitation stations), so it is sufficient to present correlations for one flow station here. Since the historic correlations between the last four weeks of the previous year and the first four weeks of the subsequent year are already given in Table 1, Table 5 shows the same correlations that show statistical dependence for generated data for the same station series.

Note that the correlations placed in the window in Table 5 have been induced by permutation of the entire years of generated data in Step 3, while the other correlations shown in Table 5 have been induced in Step 2.

Comparison of auto-correlation functions between historic and generated series is a standard way to test stationary stochastic series. As seen in Table 6, the differences shown for a lag of 15 weeks are usually on the third decimal. Such small differences would hardly be visible in a graphical form on a correlogram that is typically used to present comparisons of target and generated auto-correlations.

Finally, since there were 1000 years of weekly data generated for 11 stations, very rare events with return flow periods of 100, 200 or 500 years found in 1000 hypothetical years should have similar values to the weekly values that would be derived from statistical distributions that are used to fit extreme hydrological events. To this end, annual weekly maximums have been extracted from each year of the generated series and ranked in ascending order, thus allowing estimates of the flows for specific return flow periods based on the standard Weibull plotting position formula $n/(m+1)$, where n is the rank in the sorted series and m is the maximum number of data in the series (in this case $m = 1000$). These flows are then compared with the output of a standard frequency analysis model that uses historic annual weekly maximums as input and estimates high weekly flows with the same return flow periods of 100, 200 and 500 years. The results are presented in Table 7.

The percentage of rainy weeks in the generated series is also similar to the historic series. For three precipitation stations (Cardstone, Mountain View and Taber), this duration is between 27% and 29% of the time, while for the precipitation at Lethbridge this duration is 20% in the historic series and 22% of the time in the generated series. The slight increase in the generated series may be due to counting very small generated values (less than 0.1 mm) as rainy days, while the historic measurements would record such rainfalls with zeros.

Note that high weekly flows with the return periods of 100, 200 and 500 years are within the expected range that was obtained by fitting the historical annual maximums with the three statistical

Table 5 Correlations of weekly stochastic series for Station C5AE27.

Generated	Week 49	Week 50	Week 51	Week 52	Week 1	Week 2	Week 3	Week 4
Week 49	1.00	0.92	0.82	0.51	0.30	0.30	0.33	0.39
Week 50		1.00	0.86	0.52	0.31	0.31	0.34	0.40
Week 51			1.00	0.67	0.42	0.39	0.40	0.44
Week 52				1.00	0.69	0.60	0.47	0.48
Week 1					1.00	0.81	0.52	0.49
Week 2						1.00	0.56	0.52
Week 3							1.00	0.83
Week 4								1.00

Table 6 Comparison of historic (*H*) and generated (*G*) weekly auto-correlations.

Lag	C5AE27		GSTDAM		C5AD32		C5AD41		G5AD26		C5AA24		C5AD07	
	H	*G*	*H*	*G*	*H*	*G*	*H*	*G*	*H*	*G*	*H*	*G*	*H*	*G*
1	0.908	0.893	0.896	0.879	0.874	0.865	0.886	0.866	0.886	0.868	0.827	0.818	0.876	0.860
2	0.800	0.784	0.787	0.769	0.773	0.764	0.768	0.763	0.768	0.752	0.696	0.695	0.751	0.736
3	0.691	0.682	0.683	0.671	0.676	0.674	0.656	0.673	0.656	0.648	0.594	0.593	0.645	0.637
4	0.576	0.571	0.577	0.572	0.567	0.570	0.531	0.575	0.531	0.529	0.481	0.490	0.534	0.534
5	0.456	0.458	0.465	0.467	0.457	0.466	0.403	0.470	0.403	0.408	0.358	0.374	0.418	0.427
6	0.333	0.340	0.352	0.361	0.340	0.352	0.281	0.360	0.281	0.290	0.250	0.267	0.311	0.323
7	0.218	0.229	0.242	0.253	0.230	0.245	0.168	0.256	0.168	0.180	0.154	0.170	0.208	0.221
8	0.118	0.130	0.149	0.161	0.133	0.148	0.078	0.162	0.078	0.091	0.074	0.086	0.124	0.135
9	0.033	0.044	0.065	0.078	0.049	0.061	0.000	0.077	0.000	0.012	0.009	0.018	0.052	0.061
10	–0.037	–0.027	–0.003	0.008	–0.022	–0.011	–0.061	0.006	–0.061	–0.051	–0.043	–0.038	–0.008	–0.001
11	–0.094	–0.084	–0.058	–0.049	–0.079	–0.071	–0.108	–0.054	–0.108	–0.099	–0.083	–0.081	–0.055	–0.051
12	–0.139	–0.130	–0.104	–0.096	–0.126	–0.120	–0.143	–0.103	–0.143	–0.136	–0.112	–0.113	–0.092	–0.090
13	–0.174	–0.166	–0.143	–0.135	–0.163	–0.159	–0.170	–0.144	–0.170	–0.164	–0.135	–0.138	–0.123	–0.122
14	–0.199	–0.193	–0.172	–0.164	–0.191	–0.188	–0.188	–0.175	–0.188	–0.183	–0.153	–0.156	–0.147	–0.147
15	–0.219	–0.213	–0.195	–0.188	–0.213	–0.210	–0.200	–0.199	–0.200	–0.196	–0.165	–0.169	–0.165	–0.165

Table 7 Comparison of extreme values with frequency analysis model outputs.

Return period (years)	Weekly flow series							Precipitation series			
	C5AE27	GSTDAM	C5AD32	C5AD41	G5AD26	C5AA24	C5AD07	Cardstone	Leth.	M.View	Taber
Generated series: based on the Weibull plotting position probability											
100	280.1	337.5	69.0	111.6	286.0	645.9	1598.9	159	123	204	123
200	321.2	366.2	78.1	130.1	342.4	739.7	1733.4	179	131	213	133
500	337.8	428.4	83.2	146.3	379.1	896.6	2301.0	188	150	252	172
Historic series: 3-parameter log-normal distribution, MLM fit											
100	263.0	327.4	71.4	110.5	289.6	722.1	1708.0	149	129	204	129
200	287.8	365.3	78.8	123.4	318.9	822.3	1942.1	161	141	227	145
500	321.0	417.4	88.8	141.2	358.3	962.3	2269.5	176	156	258	166
Historic series: extreme value distribution, MLM fit											
100	274.0	334.8	73.1	118.9	293.5	793.5	1781.0	149	128	206	134
200	304.0	378.0	81.4	137.0	324.8	940.8	2071.2	159	138	231	152
500	345.2	438.8	92.9	164.0	366.9	1166.1	2500.8	172	152	264	179
Historic series: log-Pearson III distribution, MLM fit											
100	268.0	335.0	75.0	120.4	288.0	739.3	1731.7	146	126	202	131
200	294.8	376.7	83.9	138.4	315.9	845.5	1980.6	156	136	224	147
500	331.2	435.0	96.4	165.0	353.0	994.7	2333.9	167	149	255	171

distributions commonly used to model rare events (three-parameter log-normal, extreme value and log-Pearson III distributions). All distributions were fitted using the maximum likelihood method (MLM). A commercial frequency analysis model was obtained from Golder Associates Ltd of Calgary to obtain statistical estimates of these flows.

CONCLUSIONS AND RECOMMENDATIONS

This article presents a general approach to modelling stochastic hydrological and meteorological time series that relies on recent scientific advances in statistics. The relevant statistical dependences found in the historic series that have been preserved in the generated series include statistical distributions of weekly flows, both annual and weekly auto-correlations, as well as cross-correlations with all significant lags. The method does not require extensive calibration. The current research is focused on testing the model using daily time steps.

Acknowledgements Dr Shuhong Wu of Golder Associates' office in Calgary is acknowledged for allowing the use of Golder's frequency analysis software in this study.

FUNDING

Alberta Innovates, a provincial research funding agency in Canada, is acknowledged for its kind financial support. In its earlier phases, this development was jointly funded by Optimal Solutions of Calgary and the MITACS research agency from Western Canada.

REFERENCES

Box, G.E.P. and Jenkins, G., 1970. *Time series analysis, forecasting and control*. 1st ed. San Francisco, CA: Holden Day.

Bras, R.L. and Rodríguez-Iturbe, I., 1985. *Random functions and hydrology*. Reading, MA: Addison-Wesley.

Cario, M.C. and Nelson, B.L., 1996. Autoregressive to anything: time series input processes for simulation. *Operations Research Letters*, 19, 51–58.

Cario, M.C. and Nelson, B.L., 1997. Numerical methods for fitting and simulating autoregressive-to-anything processes. *INFORMS Journal of Computing*, 10, 72–81.

Hurst, H., 1957. A suggested statistical model for some time series that occur in nature. *Nature*, 180, 494.

Ilich, N., 2009. A matching algorithm for generation of correlated random variables with arbitrary distribution functions. *European Journal of Operational Research*, 192, 468–478.

Ilich, N. and Despotovic, J., 2008. A simple method for effective multi-site generation of stochastic hydrologic time series. *Journal of Stochastic Environmental Research and Risk Assessment*, 22 (2), 265–279.

Iman, R. and Conover, W., 1982. A distribution free approach to inducing a rank correlation among input variables. *Communications in Statistics—Simulation and Computation*, 11 (3), 311–334. doi:10.1080/03610918208812265

Kendall, D.R. and Dracup, J.A., 1991. A comparison of Index-Sequential and AR(1) generated hydrologic sequences. *Journal of Hydrology*, 122 (1–4), 335–352.

Koutsoyiannis, D., 2001. Coupling stochastic models of different time scales. *Water Resources Research*, 37 (2), 379–392.

Koutsoyiannis, D., 2003. Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrological Sciences Journal*, 48 (1), 3–24.

Koutsoyiannis, D., 2006. Nonstationarity *versus* scaling in hydrology. *Journal of Hydrology*, 324, 239–254.

Koutsoyiannis, D., *et al.*, 2009. Climate, hydrology, energy, water: recognizing uncertainty and seeking sustainability. *Hydrology and Earth System Sciences*, 13, 247–257.

Lall, U., 1995. Recent advances in non-parametric function estimation: hydrology—applications. *Review of Geophysics,* Supplement, July. US National Report to International Union of Geodesy and Geophysics 1991–1994, 1093–1102. Available from: http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291944-9208/homepage/Contact.html.

Langousis, A. and Koutsoyiannis, D., 2006. A stochastic methodology for generation of seasonal time series reproducing over year scaling behaviour. *Journal of Hydrology*, 322, 138–154.

Lee, T. and Salas, J. D., 2008. *Using copulas for stochastic streamflow generation*. *World Environmental and Water Resources Congress 2008*. Honolulu: ASCE Publication.

Matalas, N.C., 1967. Mathematical assessment of synthetic hydrology. *Water Resources Research*, 3 (4), 937–945.

Mejia, J.M. and Rousselle, J., 1976. Disaggregation models in hydrology revisited. *Water Resources Research*, 12 (2), 185–186.

Moon, Y.U., Lall, U., and Bosworth, K., 1993. A comparison of tail probability estimators for flood frequency analyses. *Journal of Hydrology*, 151, 343–363.

Nalbantis, I. and Koutsoyiannis, D., 1997. A parametric rule for planning and management of multiple reservoir systems. *Water Resources Research*, 33 (9), 2165–2177.

Sharma, A., Tarboton, D.G., and Lall, U., 1997. Streamflow simulation: a nonparametric approach. *Water Resources Research*, 33 (2), 291–308.

Srinivas, V.V. and Srinivasan, K., 2005. Hybrid moving block bootstrap for stochastic simulation of multi-site streamflows. *Journal of Hydrology*, 302, 307–330.

Tasker, R.C. and Dunne, J., 1997. Bootstrap position analysis for forecasting low flow frequency. *Journal of Water Resources Planning and Management*, 123 (6), 359–367.

Thomas, H.A. Jr. and Fiering, M.B., 1962. Mathematical synthesis of streamflow sequences for analyses of river basins by simulation. *In*: A. Maas., *et al.*, eds. *The design of water resources systems.* Cambridge, MA: Harvard University Press, 459–493.

Valencia, D.R. and Schaake, J.C., 1973. Disaggregation processes in stochastic hydrology. *Water Resources Research*, 9 (3), 580–585.

Whitt, W., 1976. Bivariate distributions with given marginals. *The Annals of Statistics*, 4, 1280–1289.