

Issues in generating stochastic observables for hydrological models

Keith Beven 

Lancaster Environment Centre, Lancaster University, Lancaster, UK

Correspondence

Keith Beven, Lancaster Environment Centre, Lancaster University, Lancaster, UK.
Email: k.beven@lancaster.ac.uk

Abstract

This paper provides a historical review and critique of stochastic generating models for hydrological observables, from early generation of monthly discharge series, through flood frequency estimation by continuous simulation, to current weather generators. There are a number of issues that arise in such models, from uncertainties in the observational data on which such models must be based, to the potential persistence effects in hydroclimatic systems, the proper representation of tail behaviour in the underlying distributions, and the interpretation of future scenarios.

1 | STOCHASTIC OBSERVABLES AS INPUTS TO MODELS

The management of water resources, floods and droughts in hydrology is an exercise in predicting the future, at least in some probabilistic sense. The future is, however, not observable and therefore necessarily unknown. It needs to be predicted, at least in some probabilistic or plausible projection sense, and for that we normally use some form of hydrological model. In doing it is usual to drive, calibrate and evaluate the hindcasts of that model as a hypothesis about how the catchment is functioning making use of some historical observables. The nature of that modelling process has been discussed extensively elsewhere. Here, I wish to concentrate on the nature of the variables used to drive the future discharge projections of the model, in particular catchment discharge, precipitation and evapotranspiration estimates.

The generation of forcing variables will necessarily be directly related to the observables used in calibration and validation. The generation process will, however, be stochastic in both the nature of the observables themselves, and in the uncertain future values. It is worth emphasizing that the observables and their historical time series of values are themselves uncertain, particularly when estimated over a catchment area. We should, therefore, consider the historical data as virtual variables (e.g., Beven et al., 2012; Coxon et al., 2015; Kiang et al., 2018; McMillan et al., 2018; Westerberg et al., 2011). This

creates additional constraints on the modelling process and suggests that hydrology should be treated as one of the inexact sciences (Beven, 2019). Extending the model projections into the future, we cannot necessarily expect the statistics of those stochastic variables to be similar to the past, given the potential for future changes in catchments and climate and persistence in the hydroclimatic system. Such changes will then be necessarily subject to epistemic uncertainties, including those associated with any projections of future conditions.

It has therefore been a fairly natural process in hydrological analyses, particularly since the start of the digital computing age, to generate time series of *potential* future inputs to inform management decisions (although, until recently, it has been much less common to consider the uncertainties in the historical observations). In some cases, this has meant a process of modifying historical records to be compatible with expectations of future conditions. In other cases, stochastic models of inputs have been used to drive hydrological models, in some cases using analytical mathematics but more often using Monte Carlo methods. Reviews of analytical derived distribution methods have been provided by, for example, Eagleson (1972); Klemesš (1978) and Gupta and Waymire (1983). To derive analytical solutions generally requires some strong simplifying assumptions and constraints. In what follows we will concentrate on Monte Carlo methods which are more general and more flexible.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Author. *Hydrological Processes* published by John Wiley & Sons Ltd.

Monte Carlo sampling methods on digital computers date back to the work of Stanislaw Ulam, Nicholas Metropolis, and John von Neumann at Los Alamos in the 1940s where the ENIAC¹ computer was assigned to solve the problem of neutron diffusion using random experiments. Since the project was secret, a code name was required for the work and Nicholas Metropolis suggested the name Monte Carlo (Metropolis & Ulam, 1949). He later contributed to the Metropolis-Hastings Monte Carlo Markov Chain algorithm,² widely used since in Bayesian model conditioning including applications in hydrology (Beven, 2009). John von Neumann also invented a method of generating pseudo-random numbers for these experiments using the middle-square method. This was known to be subject to failures, but it was fast and compact (a critical issue given the capabilities of the ENIAC computer) and von Neumann argued that when it failed it did so obviously.

In the 1950s and 1960s Monte Carlo methods developed rapidly in a variety of fields, including hydrology, as digital computers started to be made more widely accessible. The area of hydrological work we are concerned with in this paper is the generation of input sequences for hydrological models to inform management decisions. This includes the generation of discharge or water yield series for water resource and reservoir management, and precipitation and evapotranspiration inputs for hydrological simulation models. Monte Carlo methods have also been used for other purposes, such as the description of dispersion in water bodies and the generation of spatial fields of parameters for hillslope and groundwater models but these will not be considered further here.

2 | GENERATION OF STOCHASTIC SERIES OF DISCHARGES

The first paper to address the issue of generating series of flows seems to have been that of Brittan (1961) in an application to the Colorado River, but this was preceded by a number of studies that had analysed historical flow records from a probabilistic perspective (e.g., Leopold, 1959; Yevjevich, 1961). It was also followed by the much better-known paper of Thomas and Fiering (1962) that has the title '*Mathematical synthesis of stream flow sequences for the analysis of river basins by simulation*'.³ In that paper Thomas and Fiering provide a method of sampling from a specified distribution of monthly stream flows including the effects of correlation between months. It is formulated as the recursion equation:

$$q_{i+1,j} = \bar{q}'_{i+1} + b_{i+1} \{q_{ij} - \bar{q}'_i\} + t_i s_{i+1} \{1 - r_{i+1}^2\}^{1/2}$$

where $q_{i+1,j}$ and $q_{i,j}$ represent the synthetic monthly stream discharge during year j for month $i - 1$ and month i respectively; \bar{q}'_{i+1} and \bar{q}'_i represent the average monthly flow of the historic streamflow record for month $i - 1$ and month i , respectively; b_{i+1} is the regression coefficient for estimating flow during month $i - 1$ from the flow during month i ; the value s_{i+1} is the SD of the historic streamflow record for month $i + 1$; r_{i+1} is the correlation between flow for month $i + 1$

and month i ; and t_i is a random deviate sampled from a normal distribution with zero mean and unit variance (see also Fiering, 1967). This process therefore has 12 \bar{q}_i parameters, 12 b_i parameters, 12 r_i parameters and 12 s_i parameters all of which can be derived from historical flow records under an assumption of stationarity. Thomas and Fiering generated 510 years of record as a input to a water resource management simulation. Such synthetic records could therefore allow for more extreme conditions than might be available in the historical record but Yagil (1963) gave a proof that this type of equation would preserve the mean, standard deviation and correlation of the original historical record. Any process representation in such a formulation is contained only in the coefficients b_i and r_i . Preservation of the first two moments, of course, is not necessarily an indication that the nature of the extremes of interest in water resources management might be well simulated.

There were a number of extensions to the Thomas and Fiering model proposed. Harms and Campbell (1967) for example suggested sampling a log-normal distribution rather than normal (which also avoids the potential for negative flows in the synthetic series). Based on Yevjevich (1964) they also introduced a correlation from year to year to allow for storage carry-over and historical sequences of wetter and drier years. Colston and Wiggert (1970) developed confidence limits for dependable reservoir flows using stochastic monthly simulations, though Klemes and Bulu (1979) later suggested that such estimates should be considered to have limited confidence. Salas and Smith (1981) show how the Thomas and Fiering model is equivalent to an autoregressive moving-average (ARMA) model of flow and groundwater storage, with structures depending on the structure of the rainfall inputs in applying the water balance. Srikanthan and McMahon (1980) generated monthly flows for ephemeral streams in Australia, comparing six different model structures. An early review of such methods is provided by Yevjevich (1987).

Later work as digital computers became more widely available also led to developments in generating discharge hydrographs (rather than only monthly and annual time series. One of the earliest models of this type was that of Weiss (1977) who used a shot noise concept of impulses occurring as a Poisson process in time. Weiss uses impulses of exponential form, with different time characteristics for surface and sub-surface contributions to the hydrograph, but other assumptions are also possible (e.g., Claps et al., 2005). In that such models normally assume components equivalent to linear stores with different time constants, it is worth noting that there is significant overlap with the data-based mechanistic modelling (DBM) methods of Young (2013, and references therein) as fitted to historical data. and driven by stochastic rainfall series. Such DBM models could also be driven by stochastic assumptions about rainstorms to produce discharge series.

3 | DISCHARGE AS A STOCHASTIC SERIES WITH PERSISTENCE

This concept of stochastic hydrological forcing also led to a body of literature on the hydrological phenomenon as fractal variables,

primarily because the hydrologist Jim Wallis⁴ was working at the IBM Thomas J. Watson Research Center at Yorktown Heights, NY, at the same time as Benoit Mandelbrot, the mathematician who coined the name fractal (see, e.g., Mandelbrot & Wallis, 1968; Mandelbrot, 1971). Persistence effects in hydrological series had been known since the analyses of River Nile data and other geophysical series by Harold Edwin Hurst⁵ (e.g., Hurst, 1957) but fractal scaling behaviour provided a novel way of looking at persistence. Others, however, objected to the fractal description at the time suggesting there were other mechanisms to explain persistence in time series (e.g., Klemeš, 1974; Scheidegger, 1970; O'Connell, 1971). More recently Koutsoyiannis et al. (2018) have argued that what constitutes a fractal is only very vaguely defined but essentially represents some local behaviour as scale tends to zero, whereas the type of persistence exhibited by long time series of discharge is a more global behaviour that is better described as a form of Hurst-Kolmogorov dynamics (see also Koutsoyiannis, 2006, 2011).

Hydrological applications of Hurst-Kolmogorov dynamics are discussed in detail in Koutsoyiannis (2021a). It is an important issue in terms of generating stochastic variables because it can be demonstrated that some rather simple longer term stationary models may exhibit apparently short-term nonstationary characteristics including apparent changes in mean or trends. The controversy in part arises because of the difficulties of identification of such models given only relatively short hydrological records. This remains an issue in determining the nature of change (e.g., Koutsoyiannis, 2013; Serinaldi et al., 2018). Dimitriadis and Koutsoyiannis (2018) have recently proposed a general model for stochastic generation of variables that, for certain structures, can reproduce aspects of intermittency characteristic of some long term series. Koutsoyiannis (2020) has extended this work to time irreversible processes such as series of event hydrographs.

4 | GENERATING STOCHASTIC INPUTS TO HYDROLOGICAL MODELS

These early stochastic models of discharge were primarily intended to replace observed values as inputs to water resource management models. Later developments also started to use stochastic models of the inputs to rainfall-runoff models to replace observed values. There were a number of reasons for doing so. The first was to provide an alternative way of estimating flood frequencies in situations where long discharge records were not available to allow the robust fitting of a flood frequency distribution. It was often the case that longer rainfall time series were available, so that fitting a stochastic model of rainfalls would allow the generation of long rainfall sequences as an input to a rainfall-runoff model calibrated to whatever discharge data were available. This approach had already been used directly on observed rainfall series, for example, using the Hydrocomp version of the Stanford Watershed Model in Fleming and Franz (1971).

The origins of generating stochastic inputs for flood frequency estimation lay in the seminal paper of Pete Eagleson (1972).⁶ His

motivation for developing a method of flood frequency estimation from rainfall distributions was

'to seek an increased understanding of the dynamics of flood frequency through a theoretical development that relates peak streamflow statistics to the statistics of climatic and watershed parameters by using the kinematic equations of runoff as they are derived for homogeneous catchments and storms' (p878).

This involves three components: a model of point rainfalls and the intensity and duration of rainfall excess; a runoff routing model to derive peak discharge from the rainfall excess variables and catchment characteristics; and a transformational component to derive *'the distribution $F(Q_p)$ of peak streamflow from the distributions of the rainfall excess variables by using the analytic relationship provided by the runoff model'* (p.879).

Using simple mathematical descriptions of distributions and kinematic theory for the routing, Eagleson showed that the derived distribution of flood peaks could be obtained analytically. This provided an engineering solution to frequency estimation from rainfall characteristics at a time when the use of digital computers was still not widespread. It was also the start of an extremely fruitful and influential scientific program based on the derived distribution approach, including the sequence of 7 papers on Climate, Soil and Vegetation in Water Resources Research by Eagleson (1978).

The assumption was that such an approach would provide more robust estimates of flood frequencies than the uncertain estimates associated with fitting a distribution to a small number of historical flood peaks. It was also argued that the approach could be applied to ungauged catchments if the parameters of the rainfall-runoff model could be estimated independently based on soil and topographic information. The devil is, of course, in the details of the assumptions necessary to obtain a tractable analytical solution. In Eagleson (1972) it is assumed that both mean rainstorm intensities and durations can be represented as independent one parameter exponential distributions (since a gamma conditional distribution was more difficult to handle) with an areal reduction factor to allow for increasing catchment area. Rainfall excess is then calculated by subtracting a constant loss rate from the mean rainfall intensity (the Φ -index method) leaving a constant excess rate over the duration of the storm. A table of typical loss rate parameters is given but no account is taken of how the loss rate might vary with antecedent wetness prior to an event (Eagleson again notes that would lead to very complicated mathematics). This is then routed over representative slope segments for the catchment. Storm runoff here is very much overland flow, but following the work of Betson (1964), Eagleson allows that this runoff may occur only as a partial area response in the catchment, by letting the rainfall excess to occur on a rectangular area centred on the stream channel will hillslopes of fixed slope and noting that this might result in a partial equilibrium hydrograph depending on the intensity and duration of the excess rainfalls. The routing is completed by routing down the channel itself to derive a flood peak magnitude for an event.

Conditional integration over the distribution of storms that have rainfall excess rates greater than zero results in the required derived distribution of flood peaks.

This paper inspired others that provided alternative models for both the runoff generation and the routing components. As computer power increased, some of the simplifying assumptions about runoff generation and routing required by Eagleson's analytical analysis could be relaxed, including the use of the geomorphic unit hydrograph for routing (e.g., Diaz-Granados et al., 1984; Hebson & Wood, 1982) and more realistic treatments of the effect of antecedent wetness conditions and runoff generation (e.g., De Michele & Salvadori, 2002; Sivapalan et al., 1990). This type of analysis is, however, fundamentally depending on the representation of the driving input variables. More realistic rainstorm generators have been proposed including the random rectangular pulse Bartlett-Lewis and Neyman-Scott models (e.g., Rodriguez-Iturbe et al., 1987; Entekhabi et al., 1989; Cowpertwait & O'Connell, 1992; Onof and Wheeler, 1993). Storms are assumed to be generated as a Poisson process but they differ in how the time positions are generated for cells. In the Bartlett-Lewis model the occurrence of cells is independent and identically distributed, in the Neyman-Scott model the occurrence of cells after the storm origin is independent and identically distributed. The choice of distributions for the occurrences and magnitudes of the cells is also important. For the numbers of cells in a storm both Poisson and geometric distributions have been assumed. These types of model structure will not reproduce the type of scaling behaviour that has been inferred from analyses of rainfall time series. They can, however, reproduce the first few moments of the rainfall distributions. Estimation of the parameters of these types of models can be an issue (e.g., Favre et al., 2004; Velghe et al., 1994). Reviews of such models are provided by Waymire and Gupta (1981) and Bordoy and Burlando (2014).

A further development was to add a distribution of arrival times for events, and an additional evapotranspiration component to then allow the generation of time sequences of inputs to a hydrological model which produces continuous simulations of discharges, including the flood peaks of interest. These models were effectively the first 'weather generators'. Such an approach has less often been used to generate the statistics of droughts given the limitations of many rainfall-runoff models in predicting low flows as connectivities in flow pathways start to break down. One of the first studies to use continuous simulation to estimate flood frequency made use of historical rainfall records and evapotranspiration estimates (Fleming & Franz, 1971). This limits the length of simulations can be made, whereas an approach based on distribution functions can generate records of any required length. A rule of thumb in flood frequency estimation that arose out of some of the early stochastic sampling experiments carried out by Jim Wallis and others (e.g., Wallis & Wood, 1985) was that robust estimates of frequency required records of ten times the return period to be estimated (i.e., 1000 years for events with an annual exceedance probability [AEP] of 0.01 that are commonly used in flood defence designs). For dam safety studies, where event magnitudes with an AEP of

0.0001 might be of interest, this suggests realizations of 100 000 years might be required.

5 | ISSUES IN GENERATING INPUTS FOR RAINFALL-RUNOFF MODELS

Point models of rainfalls are generally based on historical raingauge records, for which daily data are much more commonly available. For larger catchments daily data might be adequate for hydrograph prediction (knowledge of spatial patterns might then be more important, see next section) but for small catchments it can be important to have some knowledge of sub-daily rainfall variation. In both cases, however, what a point raingauge measurement represents may be subject to uncertainty. Many studies have been done about the effects of wind, height and exposure on raingauge catch of different designs, and on the effects of intensity on what is recorded by a tipping bucket gauge. There have also been a number of studies on the small-scale variability of raingauge catches. As far back as Hutchinson (1969) for example, it was demonstrated that the coefficient of variability for a collection of 12 raingauges in a single enclosure at an exposed site in New Zealand decreased with mean storm rainfall but could be 10% for a storm of 10 mm. Averaging to monthly values decreased this to 1%–2%. Such uncertainty is often ignored in rainfall generation.

There is then the further issue of relating point rainfall estimates to the catchment area values required as inputs to hydrological models. Many studies have simply assumed that point rainfall distributions can be used at the catchment scale or have simply defined an areal reduction factor (as used in Eagleson, 1972, above). Others, that point values can be associated with subcatchments, but with a simple cross-correlation structure between subcatchments used in generating catchment scale storms (e.g., Blazkova & Beven, 2009). There have also been developments in space-time stochastic rainfall models to address this problem that aim to preserve the point statistics (see next section); the issue of what is a good estimate of catchment area rainfall, however, remains.

To define a stochastic point rainfall generator, long term rainfall records are generally required, especially in hydrological regimes with strong seasonal or weather type variability. Generating rainfall alone may also not be sufficient where snowfalls can be important, which will require a model that includes both rainfall and temperatures (see next section). Early derived distribution methods tended to fit distributions of storm duration, mean intensity and arrival time of the next storm to the sample values derived from the available rainfall records, checking for correlation between the values. This then requires the definition of what constitutes a storm in the available data (e.g., Beven, 1986a; Restrepo-Posada & Eagleson, 1982) by what minimum period of zero or negligible rainfalls should be required between storms. Such issues arise both for daily and sub-daily rainfall values and will inevitably result in some degree of approximation.

Later studies have also involved the generation of storm profile data. One approach has built on the analyses of the multifractal nature of observed rainfalls in time (e.g., Gupta & Waymire, 1993), by

disaggregating the event total and duration into a sequence of fractal increments for a specified time step (e.g., Puente and Obregón, 1996; Obregon et al., 2002; Maskey et al., 2019). Wavelet analysis has also been proposed as a more robust way of analysing multifractal behaviour and has led to some important insights about the small-scale variability within and between pulses of rainfalls within an event (e.g., Venugopal et al., 2006). This type of scaling behaviour might be expected if the high frequency variations in rainfall rates are linked to atmospheric turbulence (though see the discussion in Veneziano et al., 2006, suggesting that it may not be that simple). A simpler approach, based directly on the observational series, is to sample directly from cumulative profiles taken from the observed storms. This approach was taken by Cameron et al. (1999); Cameron et al. (2000) who classified the sampled profiles, normalized by event volume and duration, into a number of event classes. Given a sampled profile for a generated event, the intensities can be calculated for any required time step.

Another issue, noted earlier, is that some applications require the generation of long time series, for instance in frequency analysis for the design of flood defences or dam safety studies. This raises two problems: one is the issue of how pseudo-random number generators produce extreme values on the tails of the specified distributions (some generators are much better than others in that respect, see Bowman, 1995; Press et al., 2007; Bertoni et al., 2010)⁷; the second is the infinite tails of those distributions which can allow the generation of some very extreme values. In terms of extremes, some consideration needs to be given to reproducing higher moments of the observed data (e.g., Burton et al., 2008; Koutsoyiannis, 2004, 2021a) but for rainfalls, those extreme values might, by random sampling during a long enough run of random numbers, exceed what might be atmospherically possible in a hydroclimatic region. What is atmospherically possible was embodied in the idea of a maximum possible precipitation, that later evolved into a probable maximum precipitation (PMP) (Benson, 1973; Hershfield, 1961). It is, however, difficult to define and its evaluation is subject to multiple assumptions and different sources of epistemic uncertainty, especially in what to assume about the potential advection of moisture and vapour (see, for example, Koutsoyiannis, 1999; Papalexiou & Koutsoyiannis, 2006; Micovic et al., 2015).

Different applications of rainfall generators have dealt with these problems in different ways. Some have simply ignored the problem, particularly when the interest is in defining frequencies that are not too far out on the tails. When the interest is in more extreme values two approaches have been taken. One is to resample if an event exceeds some upper threshold. For rainfalls this has often been some regional definition of PMP (e.g., Blazkova & Beven, 2009). Another approach has been to use a form of sampling distribution that does not have an infinite tail, but rather is asymptotic to some upper limit. Cameron et al. (1999) for example, introduced an upper bound for the distribution of storm intensities based on the UK national envelope of rainfall extremes for different durations. Cameron et al. (2000) compared three different stochastic rainfall models for a number of sites: a version of the original Eagleson (1972) model that had also been used in the continuous flood frequency modelling of Beven (1986a, 1986b, 1987); a data-based model for the Plynlmon rainfalls that had been introduced by Cameron et al. (1999); and the gamma distribution

of the Bartlett-Lewis model developed by Onof and Wheater (1994). Based on the data for that site, they also introduced an upper bound on durations in fitting a Generalized Pareto Distribution to the class of longest duration storms. They concluded that the Eagleson model is the least satisfactory model for that site (though could be improved by relaxing the choice of distributions), that the data-based model had some difficulties in reproducing summer storm profiles for one of the sites where the extreme intensities were associated with longer duration storms; while the gamma Bartlett-Lewis model might require fitting of parameters for sub-annual periods to improve the estimation of extremes at all sites. Cameron et al. (2001) also modified the Bartlett-Lewis model to allow an upper limit to the generation of extremes that was set above the upper envelope of UK rainfall statistics for different durations.

This may, however, be a misguided approach to the problem. Koutsoyiannis (1999) revisits the paper of Hershfield (1961) that made use of data from multiple stations in advocating the probable maximum precipitation concept. He shows that the joint data set used by Hershfield is consistent with a Generalized Extreme Value Type 2 (EV2) distribution out to the very long return periods allowed by concatenating the data from different sites. He also later shows (Koutsoyiannis, 2004, 2021a) that the shape parameter of the EV2 distribution is close to a value of 0.15 for a very wide range of stations across Europe and North America. An EV2 distribution does not, of course, have an upper limit, it will always predict more extreme values than the EV1 (Gumbel) distribution for a given return period (though both will be subject to significant uncertainties at high return periods depending on the length of the data set available for analysis. The statistics therefore do not justify imposition of an upper limit.

In a stochastic generator, however, it remains the case that values of event rainfalls much larger than the envelope curves of extremes for a given duration (often assembled on a national basis) might be randomly generated. These will have low probability, consistent with the distribution from which they are generated, but will clearly generate extremely high flows if used as the input to a hydrological model. What we do not know is whether those values might be beyond the range of atmospheric possibility at these very low probabilities. The implication is that, even if there is no justification for applying any upper limit, sufficiently long realizations should be used, relative to the return period of interest for an application, to allow any 'outlier' values generated, of much lower probability, not to have an effect on the return period of interest; hence the rule of thumb noted earlier that the realizations should be at least ten times longer than the return period of interest. Multiple realizations will also allow the sampling variability at different return periods to be evaluated (see, for example, Blazkova & Beven, 2009, for a dam safety example).

5.1 | Space-time rainfall models

The point rainfall generators of the last section have often been used as if they represented catchment averaged inputs, in the same way that data from single raingauges is also often used as if it represented

the catchment average inputs when there is no other information about the spatial pattern. Data from multiple raingauges can, of course, be averaged or interpolated over the catchment area in a variety of ways to estimate an average input value but there may be epistemic uncertainties associated with such estimates, especially where there is a large elevation range in a catchment or where the precipitation also involves snowfall. Attempts to estimate such uncertainties have been rather rare in the literature but can be derived for methods such as general linear models (e.g., Chandler et al., 2006) or kriging and co-kriging (e.g., Clark & Slater, 2006).

In stochastic generation of precipitation values, going from single sites to a full space–time field is not simple because of the time-varying correlation structures in space and time associated with rainfall fields and the possibility of only partial coverage by some events. Both synoptic and convective rainfall events are structured in cells and bands, with analyses suggesting that this might involve a multifractal scaling (e.g., Schertzer and Lovejoy, 1987; Hubert et al., 1993). An early model of space–time rainfall was suggested by LeCam (1961) within a framework of evolving rainfall cells or ‘showers’, clustered in space and time, with later developments by Waymire et al. (1984) and Cox and Isham (1988).

More recent extensions of point rainfall models to space–time event generation, include the RainSim code of Burton et al. (2008) which is based on a Neyman–Scott clustering process (Cowpertwait et al., 2002; Leonard et al., 2008). Wheeler et al. (2000) present an analysis from the 49 gauges of the HYREX (Hydrological Radar Experiment) study in the UK before presenting a model that combines Bartlett–Lewis clustering of rainfall cells in time, with a Neyman–Scott process for the clustering of cells in space. Cells are born and evolve within a storm moving with a stochastically chosen direction and velocity. Yang et al. (2005) and Chandler et al. (2006) suggest a method based on General Linear Modelling suitable for use in continuous simulation, recently extended to multisite simulation by Chandler (2020). As noted earlier for point rainfall models, all of these approaches depend on the underlying parametric distributions that are used to represent the rainfall generation mechanisms. For the rainfall extremes the results will depend heavily on how the tail behaviour is defined, even if the lower moments of the distributions of observations are well reproduced. This has, it seems, been little studied to date.

5.2 | Issues in generating snow inputs

Snow and snowmelt are important in many catchments, particularly in mountainous areas where at times of transition only part of a catchment may be accumulating or melting snow. Depths of snow accumulation and their water equivalents are difficult to observe and predict, because of the effects of wind and topography on accumulations, the variability of temperature conditions and snow densities, and the potential for changes in albedo as a snow pack ripens. Some of these difficulties were revealed in the WMO snowmelt model intercomparison exercise⁸ (WMO, 1986) and, while there have been

improvements in remote sensing of surface temperatures, albedo and snow-covered areas, there has been little real advance in modelling. There are too many uncertainties in the energy budget method, and too much approximation in the degree-day method. That is why, in real-time snowmelt modelling, data assimilation of snow-covered areas is often used to improve the forecasts during the melt season.

Because of the importance of snow in many regions, however, there have been many attempts to include snow in weather generators, particularly in assessing the impacts of future climate scenarios. In some cases this also involves consideration of the accumulation and melt of snow on glaciers. These approaches are based either on a full energy budget approach or the degree-day method. In both cases, temperatures and precipitation values must be generated (with more meteorological variables for the energy budget method), and a temperature threshold is used to decide whether the precipitation falls as snow or rain as a parameter of the model. Temperatures can, of course, be variable in space, particularly in mountainous terrain so that there is an issue of generating spatial patterns, even if that is approximated to an elevation lapse rate relative to temperature at some point. It also raises the issue of covariation of all the required variables.

5.3 | Issues in converting stochastic precipitation inputs into flows

Peak flows will depend on velocities and rates of flood runoff generation; low flows will depend on storage–discharge characteristics of the subsurface including, at very low flows, the potential for the breakdown of connectivity in some flow pathways. Both are far more difficult problems than how to route that runoff once it is in a river, in part because of the observations of rainfall and runoff under extreme conditions are both limited and uncertain. This will have an important effect on any attempt to validate rainfall–runoff models (Beven, 1986b, 2019).

In continuous simulation studies, rainfalls are not the only input required to drive a hydrological model (and if snow is an important input, then other variables such as temperature and radiation might already be required to estimate melt rates). To allow the simulation of antecedent conditions prior to each event account must also be taken of actual evapotranspiration losses from a catchment. There may also be other losses, such as deep groundwater seepage, but these are generally ignored. Potential evapotranspiration rates can be simulated as part of a weather generator where it is expected that temperatures and humidities might be changing. However, some studies suggest that evapotranspiration rates in a hydroclimatic zone are rather conservative, so that if significant change is not an issue, rather simple models of potential evapotranspiration might be used to give good simulations of soil moisture deficit, even under extreme dry conditions (e.g., Calder et al., 1983).

For extreme wet conditions, a major issue in generating flows from any stochastic model of the inputs is just how good are the process representations of hydrological models, particularly under

extreme conditions? In principle there is an opportunity to use a rainfall-runoff model to reflect dynamic changes in the dominant runoff processes with event magnitude and antecedent conditions. In practice this has proven to be difficult to confirm; though it has to be said that the hydrological modelling community has not really tried that hard to invalidate its model formulations (Beven & Lane, 2019). It has instead relied on calibration techniques to obtain a parameter set giving acceptable results within a model structure (where the definition of acceptable has been rather variable). There has been little attempt to really test the underlying process concepts of model structures, and very few studies where a model formulation has been rejected (but see, for example, Page et al., 2007; Hollaway et al., 2018).

In the case of flood hydrology, there is a long tradition of applying baseflow and stormflow concepts with the continued use of hydrograph separation techniques (e.g., recently Ghotbi et al., 2020). But there is no common definition of what might constitute baseflow (e.g., Beven, 1991) such that different definitions will provide a different balance of volumes. The only physically justifiable method is to estimate what the discharge might have been if the next event had not occurred, but this is not commonly used since that estimation can be difficult when sequences of events occur. Tracer experiments also show that stormflow and baseflow (however defined) are not equivalent to surface and subsurface runoff (see, for example, Kirchner, 2003; Beven, 2012, 2020a, 2021a). The relationship is much more complex.

The identification of model parameters based on historical data also remains an issue, especially when the emphasis is on the higher magnitude flows for flood frequency estimation. Lamb and Kay (2004), for example, applied flood frequency estimation by continuous simulation to a wide variety of catchments in the United Kingdom. They showed that in fitting the model parameters the simulations could often reproduce the observed annual maximum flood frequency statistics (though not in all the catchments studied) but in many cases the highest flood simulated in a year was not the same as the highest flood observed in that year. This is partly a result of model limitations in both representing runoff generation and setting up the antecedent conditions prior to an event, but also a result of the limitations of rainfall and flood discharge observations for extreme (and not so extreme) events.

Indeed it has been suggested that machine learning method might be more successful in converting rainfalls into streamflows than models based on process representations (e.g., Kratzert et al., 2019; Nearing et al., 2021). Such data-based methods are heavily dependent on the range and quality of the training data set available, though they will be able to compensate for inconsistencies in a data set if those inconsistencies show some form of consistent structure (see the discussion in Beven, 2020b). There may also be problems in predicting rare events or extremes using machine learning methods, since by definition such events will be rare in the training data and extrapolation beyond the range of the training data may not be well controlled. There will certainly more attempts to use such techniques with additional constraints (e.g., an event water balance constraint) in future but such studies are still at an early stage.

There is an extensive literature on the uncertainties in hydrological observations and in predicting streamflows from rainfalls and other input data (see the other papers in this volume). This is compounded in the current context by the additional uncertainties associated with any stochastic model of those input data that then cascades through whatever approach is being used to simulate the discharges. The epistemic cascading nature of such uncertainties suggests that all the resulting discharges will be uncertain, and that any estimation of that uncertainty will be subject to the specific assumptions made in an analysis (e.g., Beven & Lamb, 2014).

5.4 | Issues in generating future sequences: Stochastic weather generators and stationarity

Stochastic models are designed to reproduce the statistics of the observational series against which they are calibrated. But this will then depend on the sample support for that calibration and whether the process can be considered as stationary for periods of similar length. In one sense this does not matter in that all the series generated by a stochastic model are necessarily conditional; they will behave strictly according to the assumptions on which they are based (if programmed correctly, and approximately so in the case of, for example, fractional Gaussian generators of finite memory length). This does not mean that they will adequately span the possibilities of future behaviour if either the calibration period is atypical, or if the system shows persistent behaviour that cannot be adequately characterized in the data available, or if the system is changing over time. In all of these cases the generated series might underestimate the potential future variability (see, for example, Thompson & Smith, 2019).

Sometimes it is difficult to determine which of these possibilities might hold. This is illustrated, for example, in the flood frequency estimation by continuous simulation study of Blazkova and Beven (2009). The aim of that study was to get a better estimation of very rare events for assessing dam safety in the Skalka catchment in the Czech Republic. A rainfall-runoff model was calibrated for the catchment using 60 years of discharge records using the GLUE methodology. A rainfall model was also calibrated for covarying subcatchment storm rainfall (including occult precipitation and snow accumulation and melt components), and a simple model of evapotranspiration was used. A variety of limits of acceptability and fuzzy weights were used in fitting the model parameters for different subcatchments within a GLUE framework. Only 39 model realizations out of a sample of more than 600 000 were found that satisfied all the 114 limits of acceptability, but a Pareto method of relaxing limits was then used to obtain a sample of 4000 models for the final simulations. The observed flood frequency curve was well covered by these simulations, but with significant uncertainty. This included a period of 60 years at a site prior to the construction of the Skalka Dam that was not used in the model identification. However, this earlier record showed quite different frequency characteristics to the upstream sites used for the model identification period. That is not to say that the system had changed (though developments in land use and agricultural practice might have resulted in

a change in catchment response), it could have been only exhibiting the characteristics of a stationary stochastic system (e.g., Koutsoyiannis & Montanari, 2015; Montanari & Koutsoyiannis, 2014).

An enormous number of studies have taken the projections from climate models for different future decades to make an evaluation of future change in rainfalls and flows. A very large proportion of those studies have done so without any form of uncertainty estimation of those projections, or only making use of a limited ensemble of projections. This is, in part, still a reflection of computer limitations since there is always a trade-off with global scale models between resolution, process detail, and the number of simulations that can be run given the available computer resource.

Perhaps more importantly, very few have done so while allowing for the potential for persistent stochastic behaviour (though see Tyrallis & Koutsoyiannis, 2017). Most authors of such studies will make the argument that the outcomes are only potential scenarios, and that the projections of climate models are the best indication of future boundary conditions that are available. However, the outcomes from this type of study have rarely been tested, for example, by evaluating such an approach as predicted before the last decade against the actual observations over the last decade. This is not surprising since climate models change over time, so it would not be surprising if a past generation of model did not do well in predicting change in the last decade. The current generation of climate model might do better in predicting change in the next decade but will then also be superseded by the time we have the observations for that decade. In the past, however, they have certainly not done well in predicting hydrologically relevant variables such as rainfalls and near-surface humidities and wind speeds (see, for example, Koutsoyiannis, 2021b); indeed it has been suggested that at the decadal scale simple data-based models can be more successful even for predicting changes in global temperature (Fildes & Kourentzes, 2011; Suckling and Smith, 2013; Young, 2018; Young et al., 2021), but even for local predictions for land surface fluxes (e.g., Nearing et al., 2016).

This problem is generally overcome by only looking at the percentage changes predicted in future decades, and applying those changes to an observed record, or the parameters of a weather generation model. This might also involve downscaling algorithms that allow for the difference in scale from the climate model grid scale to the scale of an application such as a particular catchment. The downscaling algorithms also have to be calibrated and this allows corrections for bias and other characteristics to be made between the climate model and observational statistics. These corrections are then assumed to hold into the future. This is a convenient working procedure, but since changes are being predicted there is no reason why that assumption should hold (see, for example, Koutsoyiannis et al., 2008; Beven, 2011). Indeed, some tests of weather generators suggest that they have not done. This is now cited in the text that well in reproducing the statistics of historical weather (e.g. Semenova et al., 1998), especially for the extremes (e.g. Verhoest et al., 2010).

More recent weather generators produce outputs of high-resolution spatial fields of precipitation and meteorological variables

(e.g., AWE-GEN of Peleg et al., 2017; Peleg et al., 2019). Weather generators are subject to the same issues as stochastic rainfall models, of identifiability of parameters, unverified extreme tail behaviours, and issues of extreme values being generated by chance. There are also issues as to how well covariation between variables can be represented, and Li et al. (2017) suggest that the assumptions about the spatial structure of the generated fields can have an important effect on the outputs of hydrological models. To my knowledge, none of the currently available weather generators used in such studies take any account of the potential for Hurst-Kolmogorov persistence in generating values, albeit that there are algorithms for generating fractional Gaussian noise and other types of persistent behaviour available, including time irreversible processes such as event discharges (Koutsoyiannis, 2020).

6 | CONCLUSIONS

There are enough issues associated with the stochastic generation of observables as inputs to hydrological models that they should be both used with care and the outputs, particularly the extremes, be assessed for realism. This is, of course, easier said than done, since even in the historical record the sample of available extremes might be rather limited. Evaluating future projections is impossible, but that also suggests that such projections should not be used without some effort being made to assess what the potential range of uncertainty associated with the projections might be and how that might propagate through the cascade of models involved in any impact study (e.g., Beven & Lamb, 2014).

Some of the most important issues that remain are:

1. The characteristics of the observations and their uncertainties that underlie stochastic models of inputs. It is still difficult to have an idea of catchment average inputs, because of uncertainties in observations (raingauge, radar or satellite) and interpolations. That makes it difficult to properly calibrate and test any stochastic models of inputs that might be used to produce long term sequences for flood frequency estimation or assessing impacts of change.
2. The limitations of the observational data are also an issue when it comes to representing the extremes (both wet and dry) based on the tail assumptions of the distributions that underlie these stochastic models. This includes the potential for persistence that has been found in time series of hydrological observables. Since we are not going to get this 'right' (it represents an epistemic uncertainty about the nature of the extremes) there can only be a choice of assumptions, but we should make an assessment of the sensitivity of decisions to uncertainty in the assumptions.
3. In terms of assessing impacts of climate change projections it is still difficult to really know how rainfalls will change into the future, or whether there has been any real change in the recent past relative to persistence in natural variability. Downscaling and bias corrections based on historical data, with the same limitations as the

- previous point that have been used to compensate for limitations of climate models are a neat trick to compensate for model deficiencies but will not necessarily hold into the future. Atmospheric modelling of particular storms with convection resolving schemes is improving but is not yet at a stage when it could be used for generating long term sequences of storms because of the dependence on boundary conditions from larger scale (and often coarser resolution) models and the greater sensitivity to land surface boundary conditions, with limitations of representation of surface and subsurface water and energy fluxes.
4. There is still much to understand about the apparent scaling behaviour of rainfall and discharge data, implying some form of long-term persistence in the hydroclimatic system, at least over a certain range of scales, with consequences for the assessment of frequencies. Some studies have made use of this in generating storm profiles and sequences in the short term, but there remains an issue about implications about longer term frequency characteristics, the use of persistence in weather generators, and whether there might be physical constraints on higher magnitude events.
 5. We cannot avoid the general problem that the hydrological models used to convert stochastic sequences of inputs into simulated discharges have not been properly tested for high magnitude events, whether those be models based on process representations or derived by data-based methods or machine learning. More work is required on model validation (or invalidation, see Beven & Lane, 2019; Beven, 2018, 2021b). This also suggests that any such study should include an appropriate assessment of uncertainty in the outcomes, cascading from the model of the inputs through the discharge simulation model, with an expectation that the uncertainty be significant.

ACKNOWLEDGEMENTS

Work on this paper has been supported by the NERC Q-NFM project led by Dr. Nick Chappell (grant no. NE/R004722/1). The paper has greatly benefitted from an excellent review and the recent work of Demetris Koutsoyiannis for which I am most grateful.

DATA AVAILABILITY STATEMENT

This paper does not include any original data.

ORCID

Keith Beven  <https://orcid.org/0000-0001-7465-3934>

ENDNOTES

¹ See <https://en.wikipedia.org/wiki/ENIAC>. ENIAC was first commissioned in 1945. By the end of its operation in 1956, ENIAC contained 18 000 vacuum tubes; 7200 crystal diodes; 1500 relays; 70 000 resistors; 10 000 capacitors; and approximately 5 000 000 hand-soldered joints. It was roughly 2.4 m × 0.9 m × 30 m in size, occupied 167 m² (1800 sq ft) and consumed 150 kW of electricity. Programming involved by a combination of plugboard wiring and three portable function tables (containing 1200 ten-way switches each), so that setting up a new job could take weeks. Several of the early master programmers were women who did not receive proper credit for their work.

² The algorithm was first described by Metropolis et al. (1953) although later accounts (Gubernatis, 2005) suggest that the credit for the original idea should go to Edward Teller and Marshall Rosenbluth, and that the coding was done by Arianna Rosenbluth (all co-authors on the paper). One account suggests that Metropolis' role was only to make the computing time available. The algorithm was extended to a more general form by W. K. Hastings (1970).

³ Harold A. Thomas (1913–2002) was a Professor of Civil and Sanitary Engineering at Harvard University and a member of the interdisciplinary Harvard Water Resources Program that was founded in 1955 and chaired by Arthur Maass from the Department of Public Administration (Maass, 1962; Reuss, 2003). Myron B. Fiering (1934–1992) was a PhD student of Thomas and, after a short spell as an assistant professor at the University of California, returned to spend the rest of his career at Harvard. He was eventually appointed as a Professor of Applied Mathematics.

⁴ http://www.history-of-hydrology.net/mediawiki/index.php?title=Wallis_Jim

⁵ See http://www.history-of-hydrology.net/mediawiki/index.php?title=Hurst_H_E

⁶ See http://www.history-of-hydrology.net/mediawiki/index.php?title=Eagleson_Peter_S

⁷ As far as I know, no hydrological application has used a “true” random number generator based on some external source (e.g., Szczepanski et al., 2004; Yu et al., 2019) rather than a pseudo-random number generator.

⁸ After participating in which, I have to admit, I gave up trying to model snowmelt as far too difficult. Even simple parameters like albedo change in such unpredictable ways. It is far better posed as a data assimilation problem for forecasting during the snowmelt season, but data assimilation is not possible for future scenarios.

REFERENCES

- Benson, M. A. (1973). Thoughts on the design of design floods. In *Floods and droughts, Proceedings of the 2nd international symposium in hydrology* (pp. 27–33). Water Resour. Publ.
- Bertoni, G., Daemen, J., Peeters, M., & Van Assche, G. (2010). Sponge-based pseudo-random number generators. In *International workshop on cryptographic hardware and embedded systems* (pp. 33–47). Springer.
- Betson, R. P. (1964). What is watershed runoff. *Journal of Geophysical Research*, 69, 1541–1552.
- Beven, K. J. (1986a). Hillslope runoff processes and flood frequency characteristics. In A. D. Abrahams (Ed.), *Hillslope processes* (pp. 187–202). Allen and Unwin.
- Beven, K. J. (1986b). Runoff production and flood frequency in catchments of order n: An alternative approach. In V. K. Gupta, I. Rodriguez-Iturbe, & E. F. Wood (Eds.), *Scale problems in hydrology* (pp. 107–131). Reidel.
- Beven, K. J. (1987). Towards the use of catchment geomorphology in flood frequency predictions. *Earth Surface Processes and Landforms*, V12(1), 69–82.
- Beven, K. J. (1991). Hydrograph separation? In *Proc.BHS third National Hydrology Symposium* (pp. 3.1–3.8). Institute of Hydrology.
- Beven, K. J. (2009). *Environmental modelling: An uncertain future?*. Routledge.
- Beven, K. J. (2011). I believe in climate change but how precautionary do we need to be in planning for the future? *Hydrological Processes*, 25, 1517–1520. <https://doi.org/10.1002/hyp.7939>
- Beven, K. J. (2012). *Rainfall-Runoff modelling: The primer* (2nd ed.). Wiley-Blackwell.
- Beven, K. J. (2018). On hypothesis testing in hydrology: Why falsification of models is still a really good idea. *WIREs Water*, 5(3), e1278. <https://doi.org/10.1002/wat2.1278>

- Beven, K. J. (2019). Towards a methodology for testing models as hypotheses in the inexact sciences. *Proceedings Royal Society A*, 475(2224), 20180862. <https://doi.org/10.1098/rspa.2018.0862>
- Beven, K. J. (2020a). A history of the concept of time of concentration. *Hydrology and Earth System Sciences*, 24, 2655–2670. <https://doi.org/10.5194/hess-24-2655-2020>
- Beven, K. J. (2020b). Deep learning, hydrological processes and the uniqueness of place. *Hydrological Processes*, 34(16), 3608–3613. <https://doi.org/10.1002/hyp.13805>
- Beven, K. J. (2021a). The era of infiltration. *Hydrology and Earth System Sciences*, 25(2), 851–866. <https://doi.org/10.5194/hess-25-851-2021>
- Beven, K. J. (2021b). An epistemically uncertain walk through the rather fuzzy subject of observation and model uncertainties. *Hydrological Processes*, 35, e14012. <https://doi.org/10.1002/hyp.14012>
- Beven, K. J., Buytaert, W., & Smith, L. A. (2012). On virtual observatories and modeled realities (or why discharge must be treated as a virtual variable). *Hydrological Processes (HPToday)*, 26(12), 1905–1908. <https://doi.org/10.1002/hyp.9261>
- Beven, K. J., & Lamb, R. (2014). The uncertainty cascade in model fusion. In A. T. Riddick, H. Kessler, & J. R. A. Giles (Eds.), *Integrated environmental modelling to solve real world problems: Methods, vision and challenges. Special Publications* (Vol. 408, pp. 255–266). London: Geological Society. <https://doi.org/10.1144/SP408.3>
- Beven, K. J. and Lane, S., 2019, Invalidation of models and fitness-for-purpose: A rejectionist approach, Chapter 6 in: Beisbart, C. & Saam, N. J. (eds.), *Computer simulation validation—Fundamental concepts, methodological frameworks, and philosophical perspectives*, Springer. 145–171.
- Blazkova, S., & Beven, K. (2009). A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research*, 45, W00B16. <https://doi.org/10.1029/2007WR006726>
- Bordoy, R., & Burlando, P. (2014). Stochastic downscaling of precipitation to high-resolution scenarios in orographically complex regions: Model evaluation. *Water Resources Research*, 50, 540–561. <https://doi.org/10.1002/2012WR013289>
- Bowman, R. L. (1995). Evaluating pseudo-random number generators. *Computers & Graphics*, 19(2), 315–324.
- Brittan, M. R. (1961). *Probability analysis applied to the development of synthetic hydrology for the Colorado River, Rep.4* (p. 99). Bureau of Economic Research.
- Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P., & O'Connell, P. E. (2008). RainSim: A spatial-temporal stochastic rainfall modelling system. *Environmental Modelling & Software*, 23(12), 1356–1369.
- Calder, I. R., Harding, R. J., & Rosier, P. T. W. (1983). An objective assessment of soil moisture deficit models, *J. Hydrology*, 60, 329–355.
- Cameron, D., Beven, K. J., & Tawn, J. (2000). An evaluation of three stochastic rainfall models. *Journal of Hydrology*, 228, 130–149.
- Cameron, D., Beven, K. J., & Tawn, J. (2001). Modelling extreme rainfalls using a modified random pulse Bartlett-Lewis stochastic rainfall model (with uncertainty). *Advances in Water Resources*, 24, 203–211.
- Cameron, D., Beven, K. J., Tawn, J., Blazkova, S., & Naden, P. (1999). Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). *Journal of Hydrology*, 219, 169–187.
- Chandler, R. E. (2020). Multisite, multivariate weather generation based on generalised linear models. *Environmental Modelling & Software*, 134, 104867.
- Chandler, R. E., Isham, V., Bellone, E., Yang, C., & Northrop, P. (2006). Space-time modelling of rainfall for continuous simulation. In B. Finckenstädt, L. Held, & V. Isham (Eds.), *Statistical methods for spatiotemporal systems, monographs on statistics and applied probability* (Vol. 107, pp. 177–216). Chapman and Hall/CRC Press.
- Claps, P., Giordano, A., & Laio, F. (2005). Advances in shot noise modeling of daily streamflows. *Advances in Water Resources*, 28(9), 992–1000.
- Clark, M. P., & Slater, A. G. (2006). Probabilistic quantitative precipitation estimation in complex terrain. *Journal of Hydrometeorology*, 7(1), 3–22.
- Colston, N. V., & Wiggert, J. M. (1970). A technique of generating a synthetic flow record to estimate the variability of dependable flows for a fixed reservoir capacity. *Water Resources Research*, 6(1), 310–315.
- Cowpertwait, P. S. P., Kilsby, C. G., & O'Connell, P. E. (2002). A space-time Neyman-Scott model of rainfall: Empirical analysis of extremes. *Water Resources Research*, 38(8), 1–6.
- Cowpertwait, P. S. P. & O'Connell, P. E. (1992). A Neyman-Scott shot noise model for the generation of daily streamflow time series. In J. P. O'Kane (Ed.), *Advances in Theoretical Hydrology*. European Geophysical Society Series on Hydrological Sciences (pp. 75–94). Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-444-89831-9.50013-2>
- Cox, D. R., & Isham, V. (1988). A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London*, A415(1849), 317–328.
- Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., & Smith, P. J. (2015). A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water Resources Research*, 51(7), 5531–5546.
- De Michele, C., & Salvadori, G. (2002). On the derived flood frequency distribution: Analytical formulation and the influence of antecedent soil moisture condition. *Journal of Hydrology*, 262(1–4), 245–258.
- Diaz-Granados, M. A., Valdes, J. B., & Bras, R. L. (1984). A physically based flood frequency distribution. *Water Resources Research*, 20(7), 995–1002.
- Dimitriadis, P., & Koutsoyiannis, D. (2018). Stochastic synthesis approximating any process dependence and distribution. *Stochastic Environmental Research and Risk Assessment*, 32(6), 1493–1515.
- Eagleson, P. S. (1972a). Dynamics of flood frequency. *Water Resources Research*, 8(4), 878–898.
- Eagleson, P. S. (1978). Climate, soil and vegetation (7 papers). *Water Resources Research*, 14(5), 705–776.
- Entekhabi, D., Rodriguez-Iturbe, I., & Eagleson, P. S. (1989). Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation. *Water Resources Research*, 25(2), 295–302.
- Favre, A. C., Musy, A., & Morgenthaler, S. (2004). Unbiased parameter estimation of the Neyman-Scott model for rainfall simulation with related confidence interval. *Journal of Hydrology*, 286(1–4), 168–178.
- Fiering, M. B. (1967). *Streamflow synthesis*. Harvard University Press.
- Fildes, R., & Kourentzes, N. (2011). Validation and forecasting accuracy in models of climate change. *International Journal of Forecasting*, 27(4), 968–995.
- Fleming, G., & Franz, D. D. (1971). Flood frequency estimating techniques for small watersheds. *Journal of the Hydraulics Division*, 97(9), 1441–1460.
- Ghotbi, S., Wang, D., Singh, A., Blöschl, G., & Sivapalan, M. (2020). A new framework for exploring process controls of flow duration curves. *Water Resources Research*, 56(1), e2019WR026083.
- Gubernatis, J. E. (2005). Marshall Rosenbluth and the Metropolis algorithm. *Physics of Plasmas*, 12(5), 057303.
- Gupta, V. K., & Waymire, E. C. (1993). A statistical analysis of mesoscale rainfall as a random cascade. *Journal of Applied Meteorology and Climatology*, 32(2), 251–267.
- Gupta, V. K., & Waymire, E. D. (1983). On the formulation of an analytical approach to hydrologic response and similarity at the basin scale. *Journal of Hydrology*, 65(1–3), 95–123.
- Harms, A. A., & Campbell, T. H. (1967). An extension to the Thomas-Fiering Model for the sequential generation of streamflow. *Water Resources Research*, 3(3), 653–661.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.
- Hebson, C., & Wood, E. F. (1982). A derived flood frequency distribution using Horton order ratios. *Water Resources Research*, 18(5), 1509–1518.

- Hershfield, D. M. (1961). Estimating the probable maximum precipitation. *Journal of the Hydraulics Division [American Society of Civil Engineers]*, 87(HY5), 99–106.
- Hollaway, M.J., Beven, K.J., Benskin, C.McW.H., Collins, A.L., Evans, R., Falloon, P.D., Forber, K.J., Hiscock, K.M., Kahana, R., Macleod, C.J.A., Ockenden, M.C., Villamizar, M.L., Wearing, C., Withers, P.J.A., Zhou, J. G., Haygarth, P.M., 2018, Evaluating a processed based water quality model on a UKheadwater catchment: What can we learn from a 'limits of acceptability' uncertainty framework?, *Journal of Hydrology* 558: 607–624. Doi: <https://doi.org/10.1016/j.jhydrol.2018.01.063>
- Hubert, P., Tessier, Y., Lovejoy, S., Schertzer, D., Schmitt, F., Ladoy, P., Caronnel, J. P., Violette, S., & Desurosne, I. (1993). Multifractals and extreme rainfall events. *Geophysical Research Letters*, 20(10), 931–934.
- Hurst, H. E. (1957). A suggested statistical model of some time series which occur in nature. *Nature*, 180, 494. <https://doi.org/10.1038/180494a0>
- Hutchinson, P., 1969, A note on random raingauge errors, *Journal of Hydrology (New Zealand)*, 1969, 8 (1), 8–10
- Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., Belleville, A., Sevrez, D., Sikorska, A. E., Petersen-Øverleir, A., & Reitan, T. (2018). A comparison of methods for streamflow uncertainty estimation. *Water Resources Research*, 54(10), 7149–7176.
- Kirchner, J. W. (2003). A double paradox in catchment hydrology and geochemistry. *Hydrological Processes*, 17(4), 871–874.
- Klemeš, V. (1974). The Hurst phenomenon: A puzzle? *Water Resources Research*, 10(4), 675–688.
- Klemeš, V. (1978). Physically based stochastic hydrologic analysis. *Advances in Hydrosience*, 11, 285–356.
- Klemeš, V., & Bulu, A. (1979). Limited confidence in confidence limits derived by operational stochastic hydrological models. *Journal of Hydrology*, 42, 9–22.
- Koutsoyiannis, D. (1999). A probabilistic view of Hershfield's method for estimating probable maximum precipitation. *Water Resources Research*, 35(4), 1313–1322.
- Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrological Sciences Journal*, 49(4), 591–610. <https://doi.org/10.1623/hysj.49.4.591>
- Koutsoyiannis, D. (2006). Nonstationarity versus scaling in hydrology. *Journal of Hydrology*, 324(1–4), 239–254.
- Koutsoyiannis, D. (2011). Hurst-Kolmogorov dynamics and uncertainty. *JAWRA Journal of the American Water Resources Association*, 47(3), 481–495.
- Koutsoyiannis, D. (2013). Hydrology and change. *Hydrological Sciences Journal*, 58(6), 1177–1197.
- Koutsoyiannis, D. (2020). Simple stochastic simulation of time irreversible and reversible processes. *Hydrological Sciences Journal*, 65(4), 536–551. <https://doi.org/10.1080/02626667.2019.1705302>
- Koutsoyiannis, D. (2021a). *Stochastics of hydroclimatic extremes—A cool look at risk* (p. 333). Kallipos open source book, available at Retrieved from <http://www.itia.ntua.gr/2000/>
- Koutsoyiannis, D. (2021b). Rethinking climate, climate change, and their relationship with water. *Water*, 13, 849. <https://doi.org/10.3390/w13060849>
- Koutsoyiannis, D., Dimitriadis, P., Lombardo, F., and Stevens, S., 2018. From fractals to stochastics: Seeking theoretical consistency in analysis of geophysical data. *Advances in Nonlinear Geosciences*. In A.A. Tsonis & A. A. Tsonis (eds.), (pp. 237–278). Cham: Springer. https://doi.org/10.1007/978-3-319-58895-7_14, Springer.
- Koutsoyiannis, D., Efstathiadis, A., Mamassis, N., & Christofides, A. (2008). On the credibility of climate predictions. *Hydrological Sciences Journal*, 53, 671–684. <https://doi.org/10.1623/hysj.53.4.671>
- Koutsoyiannis, D., & Montanari, A. (2015). Negligent killing of scientific concepts: The stationarity case. *Hydrological Sciences Journal*, 60(7–8), 1174–1183.
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, 23(12), 5089–5110.
- Lamb, R., & Kay, A. L. (2004). Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain. *Water Resources Research*, 40, W07501. <https://doi.org/10.1029/2003WR002428>
- LeCam, L. (1961). A stochastic description of precipitation. In J. Neyman (Ed.), *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability* (pp. 165–186). Berkeley, CA.: University of California Press.
- Leonard, M., Lambert, M. F., Metcalfe, A. V., & Cowpertwait, P. S. P. (2008). A space-time Neyman–Scott rainfall model with defined storm extent. *Water Resources Research*, 44(9), W09402. <https://doi.org/10.1029/2007WR006110>.
- Leopold, L. B. (1959). Probability analysis applied to a water supply problem. *U.S. Geological Survey Circular*, 410, 18.
- Li, Z., Lü, Z., Li, J., & Shi, X. (2017). Links between the spatial structure of weather generator and hydrological modeling. *Theoretical and Applied Climatology*, 128(1–2), 103–111.
- Maass, A. (1962). In A. Maass (Ed.), *Design of water resource systems* (p. 602). Harvard University Press.
- Mandelbrot, B. B. (1971). A fast fractional Gaussian noise generator. *Water Resources Research*, 7(3), 543–553.
- Mandelbrot, B. B., & Wallis, J. R. (1968). Noah, Joseph, and operational hydrology. *Water Resources Research*, 4(5), 909–918.
- Maskey, M. L., Puente, C. E., & Sivakumar, B. (2019). Temporal downscaling rainfall and streamflow records through a deterministic fractal geometric approach. *Journal of Hydrology*, 568, 447–461.
- McMillan, H. K., Westerberg, I. K., & Krueger, T. (2018). Hydrological data uncertainty and its implications. *Wiley Interdisciplinary Reviews: Water*, 5(6), e1319. <https://doi.org/10.1002/wat2.1319>
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6), 1087–1092.
- Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, 44(247), 335–341.
- Micovic, Z., Schaefer, M. G., & Taylor, G. H. (2015). Uncertainty analysis for probable maximum precipitation estimates. *Journal of Hydrology*, 521, 360–373.
- Montanari, A., & Koutsoyiannis, D. (2014). Modeling and mitigating natural hazards: Stationarity is immortal! *Water Resources Research*, 50(12), 9748–9756.
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57(3), e2020WR028091. <https://doi.org/10.1029/2020WR028091>
- Nearing, G. S., Mocko, D. M., Peters-Lidard, C. D., Kumar, S. V., & Xia, Y. (2016). Benchmarking NLDAS-2 soil moisture and evapotranspiration to separate uncertainty contributions. *Journal of Hydrometeorology*, 17, 745–759.
- Obregon, N., Puente, C. E., & Sivakumar, B. (2002). Modeling high-resolution rain rates via a deterministic fractal-multifractal approach. *Fractals*, 10(3), 387–394.
- O'Connell, P. E. (1971). A simple stochastic modelling of Hurst's law. In *Proceedings of International Symposium on Mathematical Models in Hydrology, IAHS Warsaw Symposium, Vol. 1* (pp. 169–187). Wallingford, UK: IAHS Press.
- O'Connell, P. E. (1977). Shot noise models in synthetic hydrology. In T. A. Ciriani V. Maione & J.R. Wallis (Eds.), *Mathematical models for surface water hydrology* (pp. 19–26). New York: Wiley InterScience.
- Onof, C., & Wheeler, H. S. (1993). Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model. *Journal of Hydrology*, 149(1–4), 67–95.
- Onof, C., & Wheeler, H. S. (1994). Improvements to the modelling of British rainfall using a modified random parameter Bartlett–Lewis rectangular pulse model. *Journal of Hydrology*, 157, 177–195.

- Page, T., Beven, K. J., & Freer, J. (2007). Modelling the chloride signal at the Plynlimon catchments, Wales using a modified dynamic TOPMODEL. *Hydrological Processes*, 21, 292–307.
- Papalexiou, S. M., & Koutsoyiannis, D. (2006). A probabilistic approach to the concept of probable maximum precipitation. *Advances in Geosciences*, 7, 51–54.
- Peleg, N., Fatichi, S., Paschalis, A., Molnar, P., & Burlando, P. (2017). An advanced stochastic weather generator for simulating 2-D high-resolution climate variables. *Journal of Advances in Modeling Earth Systems*, 9(3), 1595–1627.
- Peleg, N., Molnar, P., Burlando, P., & Fatichi, S. (2019). Exploring stochastic climate uncertainty in space and time using a gridded hourly weather generator. *Journal of Hydrology*, 571, 627–641.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). Cambridge University Press.
- Puente, C. E., & Obregón, N. (1996). A deterministic geometric representation of temporal rainfall: Results for a storm in Boston. *Water Resources Research*, 32(9), 2825–2839.
- Restrepo-Posada, P. J., & Eagleson, P. S. (1982). Identification of independent rainstorms. *Journal of Hydrology*, 55(1–4), 303–319.
- Reuss, M. (2003). Is it time to resurrect the Harvard water program? *Journal of Water Resources Planning and Management*, 129(5), 357–360.
- Rodriguez-Iturbe, I., Cox, D. R., & Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London A*, 410, 269–288.
- Salas, J. D., & Smith, R. A. (1981). Physical basis of stochastic models of annual flows. *Water Resources Research*, 17(2), 428–430.
- Scheidegger, A. E. (1970). Stochastic models in hydrology. *Water Resources Research*, 6(3), 750–755.
- Schertzer, D., & Lovejoy, S. (1987). Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *Journal of Geophysical Research: Atmospheres*, 92(D8), 9693–9714.
- Semenov, M. A., Brooks, R. J., Barrow, E. M., & Richardson, C. W. (1998). Comparison of WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate Research*, 10, 95–107.
- Serinaldi, F., Kilsby, C. G., & Lombardo, F. (2018). Untenable non-stationarity: An assessment of the fitness for purpose of trend tests in hydrology. *Advances in Water Resources*, 111, 132–155.
- Sivapalan, M., Wood, E. F., & Beven, K. J. (1990). On hydrologic similarity, 3. A dimensionless flood frequency distribution. *Water Resources Research*, 26, 43–58.
- Srikanthan, R., & McMahon, T. A. (1980). Stochastic generation of monthly flows for ephemeral streams. *Journal of Hydrology*, 47(1–2), 19–40.
- Suckling, E. B., & Smith, L. A. (2013). An evaluation of decadal probability forecasts from state-of-the-art climate models. *Journal of Climate*, 26(23), 9334–9347.
- Szczepanski, J., Wajnryb, E., Amigó, J. M., Sanchez-Vives, M. V., & Slater, M. (2004). Biometric random number generators. *Computers & Security*, 23(1), 77–84.
- Thomas, I. A., & Fiering, M. (1962). Mathematical synthesis of stream flow sequences for the analysis of river basins by simulation, Ch. 12. In A. Maass (Ed.), *Design of water resource systems*. Harvard University Press.
- Thompson, E. A., & Smith, L. A. (2019). Escape from model-land. *Economics*, 13(2019–40), 1–15. <https://doi.org/10.5018/economics-ejournal.ja.2019-40>
- Tyralis, H., & Koutsoyiannis, D. (2017). On the prediction of persistent processes using the output of deterministic models. *Hydrological Sciences Journal*, 62, 2083–2102. <https://doi.org/10.1080/02626667.2017.1361535>
- Velghe, T., Troch, P. A., de Troch, F. P., & van de Velde, J. (1994). Evaluation of cluster-based rectangular pulses point process models for rainfall. *Water Resources Research*, 30(10), 2847–2857.
- Veneziano, D., Furcolo, P., & Iacobellis, V. (2006). Imperfect scaling of time and space-time rainfall. *Journal of Hydrology*, 322, 105–119. <https://doi.org/10.1016/j.jhydrol.2005.02.044>
- Venugopal, V., Roux, S. G., Fofoula-Georgiou, E., & Arneodo, A. (2006). Revisiting multifractality of high-resolution temporal rainfall using a wavelet-based formalism. *Water Resources Research*, 42(6), W06D14. <https://doi.org/10.1029/2005WR004489>
- Verhoest, N., Vandenbergh, S., Cabus, P., Onof, C., Meca-Figueras, T., & Jameleddine, S. (2010). Are stochastic point rainfall models able to preserve extreme flood statistics? *Hydrological Processes*, 24(23), 3439–3445.
- Wallis, J. R., & Wood, E. F. (1985). Relative accuracy of log Pearson III procedures. *Journal of Hydraulic Engineering*, 111(7), 1043–1056.
- Waymire, E. D., Gupta, V. K., & Rodriguez-Iturbe, I. (1984). A spectral theory of rainfall intensity at the meso- β scale. *Water Resources Research*, 20(10), 1453–1465.
- Waymire, E. D., & Gupta, V. K. (1981). The mathematical structure of rainfall representations: 1. A review of the stochastic rainfall models. *Water Resources Research*, 17(5), 1261–1272.
- Weiss, G. (1977). Shot noise models for the generation of synthetic streamflow data. *Water Resources Research*, 13(1), 101–108.
- Westerberg, I., Guerrero, J.-L., Seibert, J., Beven, K. J., & Halldin, S. (2011). Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. *Hydrological Processes*, 25, 603–613. <https://doi.org/10.1002/hyp.7848>
- Wheeler, H. S., Isham, V. S., Cox, D. R., Chandler, R. E., Kakou, A., Northrop, P. J., Oh, L., Onof, C., & Rodriguez-Iturbe, I. (2000). Spatial-temporal rainfall fields: Modelling and statistical aspects. *Hydrology and Earth System Sciences*, 4(4), 581–601.
- WMO. (1986). *Intercomparison of models of snowmelt runoff*. Operational Hydrology Report no. 23. WMO.
- Yagil, S. (1963). Generation of input data for simulation. *IBM System Journal*, 2, 288–296.
- Yang, C., Chandler, R. E., Isham, V. S., & Wheeler, H. S. (2005a). Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research*, 41, W11415. <https://doi.org/10.1029/2004WR003739>
- Yevjevich, V. (1987). Stochastic models in hydrology. *Stochastic Hydrology and Hydraulics*, 1(1), 17–36.
- Yevjevich, V. M., 1961, Some General Aspects of Fluctuations of Annual Runoff in the Upper Colorado River Basin, Rep. 3, 48 pp., Department of Engineering Research, Colorado State University, Fort Collins.
- Yevjevich, V. M. (1964). *Fluctuation of wet and dry years, part 2, analysis by serial correlation*, hydrology papers no. 4. Colorado State University.
- Young, P. C. (2013). Hypothesis-inductive data-based mechanistic modeling of hydrological systems. *Water Resources Research*, 49(2), 915–935.
- Young, P. C. (2018). Data-based mechanistic modelling and forecasting globally averaged surface temperature. *International Journal of Forecasting*, 34(2), 314–335. <https://doi.org/10.1016/j.ijforecast.2017.10.002>
- Young, P. C., Allen, G. P., & Bruun, J. T. (2021). A re-evaluation of the Earth's surface temperature response to radiative forcing. *Environmental Research Letters in press*. <http://iopscience.iop.org/article/10.1088/1748-9326/abfa50>
- Yu, F., Li, L., Tang, Q., Cai, S., Song, Y., & Xu, Q. (2019). A survey on true random number generators based on chaos. *Discrete Dynamics in Nature and Society*, 2019, 2545123. <https://doi.org/10.1155/2019/2545123>

How to cite this article: Beven, K. (2021). Issues in generating stochastic observables for hydrological models. *Hydrological Processes*, 35(6), e14203. <https://doi.org/10.1002/hyp.14203>