encyclopedia

MDPI

Review

Techniques and Developments in Stochastic Streamflow Synthesis—A Comprehensive Review

Shirin Studnicka and Umed S. Panu *

Department of Civil Engineering, Lakehead University, Thunder Bay, ON P7B-5E1, Canada; spiran@lakeheadu.ca
* Correspondence: uspanu@lakeheadu.ca

Abstract

Stochastic streamflow synthesis has long been the cornerstone of water resource planning, enabling the generation of extended hydrological sequences that reflect natural variability beyond the limitations of observed records. This paper presents a comprehensive review of the theoretical foundations, methodological advancements, and evolving trends in synthetic streamflow generation. Historical progression is explored through three distinct eras: the pre-modern formulation era (pre-1960), the era dominated by autoregressive models (1960–2000), and the recent period marked by the rise of data-driven AI/ML approaches. Various modelling paradigms, parametric versus non-parametric, traditional versus AI-based, and single- versus multi-scale approaches, are critically assessed and compared with a focus on their applicability across temporal resolutions and hydrological regimes. This study also categorizes evaluation criteria into four dimensions: preservation of stochastic characteristics, distributional consistency, error-based metrics, and operational performance. In addition, the use and impact of transformation techniques (e.g., log or Box-Cox) employed to normalize streamflow distributions for improved model fidelity are examined. A bibliometric analysis of over 200 studies highlights the global research footprint, showing that the United States leads with 70 studies, followed by Canada with 15, reflecting the growing international engagement in the field. The analysis also identifies the most active journals publishing streamflow synthesis research: *Water Resources Research* (50 publications, since 1967), *Journal of Hydrology* (25 publications, since 1963), and *Journal of the American Water Resources Association* (9 publications, since 1974). This review not only synthesizes past and current practices but also outlines key challenges and future research directions to advance stochastic hydrology in an era of climatic uncertainty and data complexity.

Keywords: stochastic streamflow synthesis; autoregressive models; data-driven AI/ML approaches; traditional models; parametric models; non-parametric models; pattern recognition techniques; textural image analysis; Box-Cox transformation; stochastic hydrology

check for updates

Academic Editor: Andreas N. Angelakis

Received: 21 August 2025
Revised: 16 October 2025
Accepted: 30 October 2025
Published: 21 November 2025

Citation: Studnicka, S.; Panu, U.S. Techniques and Developments in Stochastic Streamflow Synthesis—A Comprehensive Review. *Encyclopedia* **2025**, *5*, 198. https://doi.org/10.3390/encyclopedia5040198

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Streamflow is one of the most important hydrologic phenomena, as both its excess and scarcity can lead to floods and droughts—natural disasters that can cause significant damage and even death. This makes it critically important to model, forecast, and synthesize streamflow to understand its behaviour, anticipate extremes, and support effective water resource management. Towards achieving the foregoing challenges, researchers have developed a wide range of models—from simple regression-based techniques to advanced

machine learning and deep learning approaches. In this section, the fundamental properties of streamflow are first explored. Following this, a historical overview of streamflow synthesis approaches is presented.

1.1. Understanding the Principles of Streamflow Synthesis

Any phenomenon that undergoes continuous changes over time is considered a process. Hydrologic phenomena, such as precipitation, infiltration, evaporation, and runoff (or streamflow), are dynamic and complex, influenced by natural factors (e.g., climate, topography, vegetation), anthropogenic activities (e.g., land use changes, water extraction, urbanization), and unknown physical interactions. This inherent complexity makes hydrologic phenomena *stochastic* rather than deterministic ([1,2]).

Moreover, hydrological processes are continuous in nature. However, to measure, study and model these processes effectively, they are treated as discrete, which means that they are discretized into specific time intervals (e.g., hourly, daily, or monthly) for practical reasons. In this context, the discrete time intervals between successive observations make them *a time series* [3–6].

In the study of streamflow as a *stochastic time series*, the choice of time interval is crucial and can vary between daily, weekly, monthly, or annual observations. This discrete time interval should be determined based on the specific objectives and requirements of the hydrologic project. For example, in the design of drainage structures, the hourly streamflow might be more appropriate, while in long-term water resource management, monthly or annual data can be used. Each of these types of data sets has different behaviour; for instance, annual series tend to lack visible cycles due to aggregation, while monthly series typically show seasonal patterns. Daily streamflow series, in contrast, are characterized by rapid peaks and sharp recessions, reflecting short-term hydrologic dynamics. In general, shorter interval series are more complex and challenging to model, whereas longer interval series (e.g., annual) simplify the modelling and parameter estimation process [2,7,8].

Hydrologic time series can be categorized as either historical, derived from measurements or observations at specific time intervals, or synthesized, generated with a specific objective of applying them in water resource system planning, design, and operations, commonly referred to as *operational hydrology* [2,9,10]. The synthesized realizations carry the information from historical time series and are designed to mimic pre-specified properties derived from historical data [11,12].

From a simple perspective, each synthesized realization of streamflow could represent a series of events that occurred before the actual measurements began, since historical data has only been recorded for almost a century, but the hydrologic processes were already at work for thousands of years before the formal recording started. This means that a severe flood recorded in the historical data might not be the worst ever flood that has occurred, and similarly, the worst recorded drought might not be the most extreme [11,13,14]. The most severe events, those not recorded in the historical data, might still occur in the future. This broader set of possible but unobserved outcomes is referred to as the *population*, the full range of potential streamflow values that could arise under the same governing processes.

Mathematically, it can be said that by streamflow synthesis, the ensemble of m streamflow series would be obtained. If $m \to \infty$, characteristics of the population would be estimated, which means that the probability of the ensemble characteristics being equal to the population characteristics will be equal to 1 ($\Pr\left(x_{sample} \cong X_{population}\right) \approx 1$) [2]. More accurately, an available historical sample is synthesized, and more information from the population is obtained. It all depends on the information extracted from the historical streamflow record, as this information defines the pre-specified properties that must be mimicked in the synthesized series. Based on such knowledge, an appropriate model is

selected, one that can reproduce the identified properties. As more complex characteristics are recognized, more advanced models are required to capture them effectively.

1.2. What Do We Know About Streamflow Characteristics

Before the 1950s, basic statistical moments, particularly measures of central tendency such as mean or median and dispersion such as variance, skewness and kurtosis, were the only information available. These simple metrics were considered sufficient representations of the underlying hydrologic behaviour.

In the 1950s and the early 1960s, the temporal dependence structure of streamflow became a key focus in hydrologic research. Streamflows were commonly characterized as a first-order Markov process, and the strong autocorrelation at the first (and the second) lag was used to quantify short-term dependence, which needs to be preserved by models designed to synthesize streamflow. The findings by Hurst introduced a new dimension to our understanding of streamflow by revealing the presence of long-term persistence, now known as the Hurst phenomenon.

Hurst [15] analyzed around 900 annual time series datasets covering various environmental variables like streamflow, precipitation, lake levels, tree rings, and atmospheric pressure. He observed that the adjusted range represented by R, which is calculated as the difference between the largest excess (P_n) and the greatest deficiency (Q_n) over a steady flow period of n years, varied with the length of the dataset (denoted by n) according to a specific relationship of $R/S \sim n^h$, where R/S is the rescaled range, and S is the sample standard deviation of the dataset. Hurst estimated a coefficient (h) through a formula involving R and S, providing insights into the long-term storage requirements of the Nile River and other hydrological systems. The coefficient (h) serves as a direct measure of long-term persistence within a time series of data [16–19]. A value of $0.5 < h < 1$ indicates a persistent time series, while $h = 0.5$ corresponds to a random walk pattern [20,21].

The identification of long-term persistence introduced new complexity into the understanding of streamflow behaviour and the broader temporal dynamics of hydrologic processes. Klemeš [22] argued that the persistence reflected by the Hurst coefficient might not arise from a single identifiable cause. Instead, it could result from a genuinely long-memory physical process (such as large-scale climatic influences), nonstationarity in the data (like gradual shifts in the mean), the effects of storage and release mechanisms in natural or engineered systems (e.g., soils, lakes, or reservoirs), or a combination of these factors, which are often difficult to understand and disentangle in practice due to data limitations and system complexity. This suggests that the memory structure of hydrologic processes is too complex to be fully explained by a single factor, such as the Hurst coefficient, even though it is widely accepted and used by hydrologists as an indicator of long-term persistence [5,12,16,21–37].

To quantify how patterns and statistical properties change across different temporal scales, revealing underlying structures that are not easily apparent at a single time resolution, the scaling behaviour of streamflow has been investigated in the literature. Several studies have attempted to analyze the scaling properties and underlying statistics [38–42]. The existence of a relationship between scaling behaviour and the Hurst coefficient was first highlighted by Gallant et al. [43] and by Blöschl and Sivapalan [44] as both scaling behaviour and the Hurst coefficient describe the concept of memory in a time series.

To quantify the scaling behaviour of a system, it is widely accepted that the Power Spectrum Analysis (PSA) represents a powerful tool to determine the scaling behaviour of time series [44–50]. In PSA, the power of a signal at various frequencies is plotted on a log-log graph. If the power spectrum follows a form $(1/f^{\alpha})$ then it is considered to be indicative of a scaling behaviour with a scaling exponent (or the power-law exponent) of

α [40,48]. When the scaling exponent, $\alpha = 0$, a time series is considered to exhibit white noise, characterized by random fluctuations with no correlation (and therefore no memory). In contrast, a scaling exponent in the range $0 < \alpha < 2$ corresponds to pink noise. For the scaling exponent, $\alpha = 2$, the behaviour is identified as brown noise or Brownian motion. Finally, for $\alpha > 2$, the series represents a black noise.

Koscielny-Bunde et al. [41] analyzed daily runoff time series from 18 representative rivers in southern Germany and 23 international rivers, and in all cases found that the fluctuations exhibited self-affine scaling behaviour over time scales ranging from weeks to decades. Livina et al. [28] highlighted that the river flow time series exhibits clear scaling behaviour, characterized by strong short-term correlation and reduced long-term correlation, likely due to aquifer (groundwater) storage effects. Dolgonosov et al. [45] discussed that spectral slopes vary across different frequency regions, providing evidence of multi-scale dynamics in streamflow behaviour. Thompson and Katul [49] showed that streamflow exhibits a robust power-law in the scaling behaviour that is largely insensitive to whether the catchment response is modelled linearly or nonlinearly, and to spatial variability in rainfall, provided the rainfall retains its spectral characteristics. They also concluded that multiple hydrologic processes and flow pathways can produce similar spectral exponents, making the mechanistic interpretation of streamflow spectra difficult despite the consistency in observed scaling patterns.

Using the PSA and detrending fluctuation analysis at 11 gauging stations in the Ebro Basin [48], investigated the scaling properties of monthly streamflow (1950–2005) in Spain and revealed the presence of characteristics of pink noise in data sets. Such an observation suggests two things: first, that the α-H relationship for non-stationary series requires further investigations in hydrology, and second, that the monthly streamflow may exhibit all types of coloured noise [51]. Studnicka and Panu [52] developed an approach to the PSA in monthly streamflow, using reported jumps in the literature [47,48] as indicators of the existence of seasonal as well as sub-seasonal cycles, inclusive of a fraction of the cycle and revealed that monthly streamflow exhibited all types of coloured noise in 143 hydrometric stations across Ontario.

The presence of noise can distort scaling behaviour across all time scales, but its influence is particularly strong at shorter time scales, such as monthly [8]. Thus, the non-stationary monthly streamflow should be characterized not only by its scaling properties and memory effects but also by the evolving patterns embedded within the data. The recognition of such patterns, often reflecting seasonal, sub-seasonal, or anthropogenic influences, adds an important dimension to understanding streamflow behaviour. Discussions around patterns as defining characteristics of monthly streamflow date back to the 1960s.

During the 1970s, following the introduction of disaggregation models by Harms and Campbell [53], monthly streamflow began to be treated as a non-stationary time series due to its strong seasonality and periodicity [5]. Panu et al. [54] and Panu and Unny [55] suggested that recurring patterns, such as cyclicity and seasonality, are inherent features of hydrologic time series. Thus, by treating each seasonal segment as an object and grouping similar objects across years, they emphasized that these patterns not only exist but can also be quantitatively analyzed through pattern recognition techniques.

Moreover, streamflow exhibits two interrelated characteristics: emergent behaviour, shaped by external climatic, geological, and anthropogenic influences, and self-organization, which arises from internal hydrologic feedback such as infiltration, storage, and routing processes. These interacting dynamics contribute to the inherent complexity of streamflows. The dual nature of streamflows can be quantitatively assessed using normalized Shannon entropy. Emergence (E) reflects the degree of unpredictability or information content in a time series, while self-organization (S) is defined as the reduction in this entropy, indicating

the underlying structure or order. The product of these two components, scaled by a factor of four, yields a normalized complexity measure (C), which ranges from 0 (complete randomness or total disorder) to 1 (an optimal balance between variability and structure).

Mujumdar and Kumar [56] have shown that this complexity measure can distinguish streamflow dynamics across different catchment types, such as low- and high-snow-dominated basins, and is significantly associated with persistence, particularly in low-snow regions. Piran and Panu [57] indicated that while complexity captures a balance between order and unpredictability in streamflows, it is not directly associated with the persistence or memory-retaining nature of the streamflow process in snow-dominated regions of Ontario. Complexity reflects the ability of a system to generate new information while maintaining underlying patterns; however, whether this behaviour stems from external non-deterministic influences, such as climate variability, or it arises from deterministic chaos, remains a subject for further investigation.

While the presence of recurring patterns, such as seasonality, periodicity, and scaling behaviour, reflects a certain degree of regularity in streamflow time series, these patterns coexist with inherent nonlinearities that confound the dynamics of hydrologic systems, particularly at the discharge rates. Streamflows are governed by a complex interplay of climatic, topographic, and land surface processes, all of which can interact in multiple nonlinear ways. As a result, even small differences in initial conditions, such as antecedent soil moisture, rainfall intensity, or vegetation cover, can lead to disproportionately large differences in hydrologic responses. This phenomenon, known as *sensitive dependence on initial conditions*, underscores the chaotic nature of streamflows [8]. Numerous methods have been developed to identify chaos in time series data, with most grounded in the fundamental concept of phase space reconstruction, a technique used to visualize and analyze the underlying dynamics of complex systems [58–62].

In the context of non-stationary time series influenced by periodicity, the choice between chaotic modelling approaches and pattern recognition systems is guided by the nature of the data set and the specific modelling objectives. Chaotic approaches are often employed when the underlying dynamics exhibit complex, nonlinear behaviour, whereas pattern recognition systems are typically preferred for detecting and adapting to evolving patterns, especially when the patterns are perceived as simpler or when ease of implementation is prioritized [63]. Understanding key characteristics of streamflow informs the selection of appropriate modelling approaches. As more aspects of streamflow are quantified, models are increasingly expected to replicate these features. Figure 1 illustrates key known characteristics of streamflows to date, along with the corresponding methods used to quantify them.

1.3. Role of Streamflow Synthesis in Operational Hydrology

Traditionally, water resource engineers have utilized the Rippl mass curve approach, often in combination with historical streamflow data, to estimate the capacity of reservoirs, which can be used either in the construction or maintenance of storage [64]. The Rippl mass curve is used to analyze and design reservoir storage capacity [65].

Storage-Yield-Reliability (S-Y-R) equations are used to determine a reservoir capacity S (active storage), required to deliver a desired yield Y, with the expected inflow Q, considering losses L at time t ($S_{t-1} + Q_t - Y_t - L_t = S_t$). The analysis involves estimating the reservoir capacity to achieve a certain yield with a specified level of reliability or determining the yield from an existing reservoir of known capacity.

The S-Y-R equations can be used in water resource engineering to estimate the reliability of a reservoir system in meeting water supply demands over a specified period (Reliability = N_s/N; where N_s is the number of time intervals during which demand was

met, and N is the total number of time intervals. Stochastic streamflow models have been recommended to derive the probability distribution of required reservoir capacity for maintaining specified releases. However, estimating parameters of stochastic streamflow models from relatively short hydrological records introduces uncertainty in the derived S-Y-R relationships.

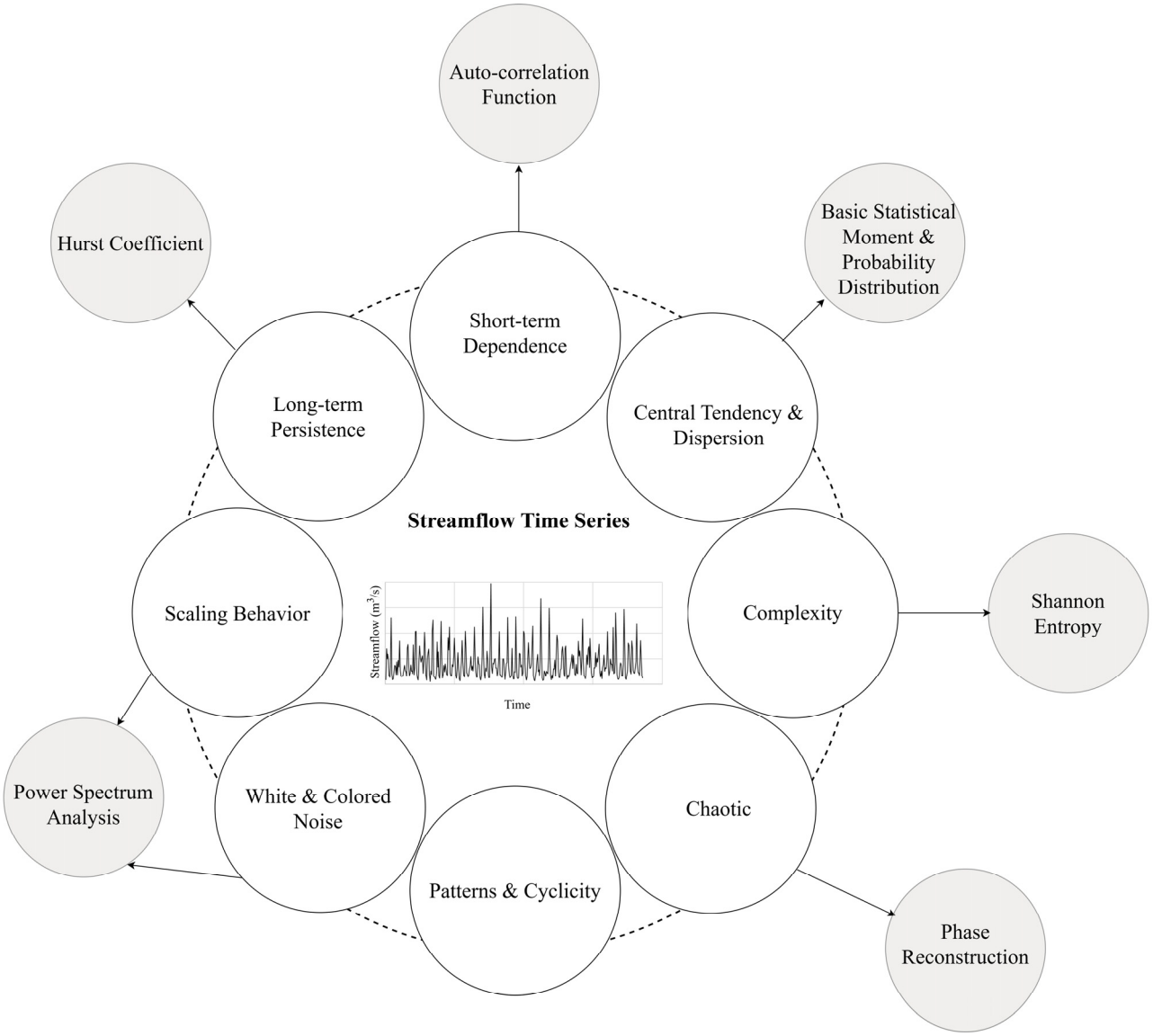


Figure 1. Key known characteristics of streamflow (white circles) and the corresponding methods commonly used to quantify each characteristic (grey circles).

The traditional approaches (deterministic approach) to reservoir analysis faced criticism, leading to a shift towards streamflow synthesis facilitated by the development of stochastic hydrology. This transition to a stochastic approach, in which the aim is to understand how the reservoir system behaves under different scenarios and conditions, such as varying inflow patterns, changing water demand, and operational strategies, requires streamflow synthesis [23].

Vogel and Stedinger [64] examined whether estimated storage capacity based on synthetic streamflow sequences, despite their limitations, offers more accurate estimates of required storage capacity volumes compared to those obtained through traditional drought-of-record analyses. Their results show that fitting an AR(1) lognormal model leads to more

accurate estimates of annual storage requirements than relying solely on historical flows, even in situations where flows were not generated with an AR(1) lognormal model. However, estimates of storage capacity obtained by fitting stochastic annual streamflow models to 80-year samples can exhibit significant variability. Understanding yield and reliability estimates is crucial within the context of typical reservoir system design applications.

2. Evolution of Streamflow Synthesis Techniques

Nearly a century and a half ago, Rippl [66] delved into the challenge of determining reservoir capacity to meet specific water demands. This problem catalyzed the emergence and growth of techniques for generating synthetic streamflow [20]. The Rippl approach deals with mass curve analysis, which determines the minimal reservoir storage based on the combination of a given demand and the historical streamflow record.

Stochastic streamflow models have been developed to estimate the reservoir capacity required to reliably meet a given water demand or release by deriving the probability distribution of that capacity under variable streamflow conditions [64]. However, streamflow synthesis did not originate during this period of hydrological challenges concerning reservoir engineering; rather, it began many years earlier, during 1914–1962, which is here referred to as the *pre-era*.

Towards developing this review of the evolution of streamflow synthesis techniques, relevant studies were identified primarily through Google Scholar by applying time filters (e.g., 1980–1990, 1991–2000) and comprehensive keyword combinations such as "*streamflow synthesis*", "*streamflow generation*" and "*streamflow simulation*".

2.1. Early Developments in Streamflow Synthesis: PRE-ERA (Beginning-1960)

There is a general agreement among hydrologists working on streamflow synthesis that Hazen in 1914 was the first to synthesize streamflow [1,12,67–69]. Hazen [70] endeavoured to extend historical records by creating a 300-year synthetic record from 14 streams, standardizing each annual flow value by dividing each annual flow value by the long-term mean annual flow of that river. The records were then combined in order of increasing variability, as indicated by their coefficient of variation. However, such a synthetic record, while used to compile design charts for reservoir capacities, failed to address limitations in the mass curve procedure and neglected the autocorrelation structure of flows.

In the early days of streamflow synthesis after Hazen, a card-based approach was employed by [71], addressing the storage issue once again. In his approach, each card represented an annual runoff volume, and these cards were shuffled. Cards were drawn one by one, and their values were recorded as they were drawn. Next, a table of random numbers drawn from a normal distribution was utilized by Barnes [72] to create a longer record of annual flows with similar averages and variations as the original data. Although better than using cards, it does not consider any autocorrelation in the annual streamflow time series [67]. The initial streamflow synthesis techniques focused on mimicking the statistical properties of historical series because, at the time, hydrologists primarily understood metrics such as mean, variance, skewness, and kurtosis. As their understanding of streamflow time series deepened, additional characteristics were quantified, and new synthesis techniques were developed to better replicate the observed complexities in historical data.

Two key findings, as noted below, drove a significant shift in streamflow synthesis techniques from the pre-era to Era 1. The first was the autocorrelation function, which revealed the strong temporal dependence within streamflow structures, particularly at immediate lags. The second was the Hurst coefficient, which quantified the memory of a time series, showing how long the effect of an event persists within the system after its

occurrence. As elucidated in the ensuing section, the Era 1 synthesis techniques primarily focused on replicating these two critical findings.

2.2. ERA-1 (1960–2000): The Domination of AR-Family Models

During this period, hydrologists have extensively utilized statistical methods and stochastic processes to analyze and describe streamflows due to their inherent stochastic nature. While statistical advances play a significant role in shaping streamflow synthesis, the literature on statistical hydrology is not rich, with notable exceptions such as Anis and Lloyd, Moran, Gumbel, Bernier, Borgman and Amorocho [1]. Among the pivotal statistical studies are the investigations on a special characteristic inhibited by streamflow records of the Nile River by Hurst [15]. The recognition of this special characteristic in streamflows is now commonly known in hydrology as the *Hurst Coefficient* (or Effect or Exponent).

Research in the 1950s by Hurst, particularly on the Nile River, sparked interest in statistical methods like the rescaled range. Significant findings of Hurst set the stage for further research into modelling hydrological processes.

In the 1960s, hydrologists recognized the importance of preserving key statistical properties in synthesized time series. The autocorrelation function (ACF) at lag 1 (ACF (1)) and lag 2 (ACF (2)), as well as the Hurst coefficient, were particularly significant for streamflow synthesis [5]. These elements became the focus of streamflow synthesis efforts. However, determining which of these properties to prioritize often depended on the specific objectives of researchers. Given the simplicity of models during Era 1, it was generally not feasible to preserve simultaneously all three properties.

Thomas & Fiering [11] were the first to mathematically model streamflow and to consider its dependence structure. In their model, streamflow at any given period is treated as a linear function of flow at the previous period, resulting in the development of a recursive formulation for monthly time intervals in a bivariate model. The model by Thomas & Fiering [11] presents a set of 12 regression equations as $Q_{i,j} = \overline{Q}_j + b_j\left(Q_{i,j-1} - \overline{Q}_{j-1}\right) + t_j S_j\left(1 - {r_j}^2\right)^{1/2}$ where $Q_{i,j}$ is the monthly streamflow in the j^{th} month of the i^{th} year to be simulated by $Q_{i,j-1}$ is the monthly streamflow in the $(j-1)$ month of the year i. The mean monthly streamflow $\overline{Q}_j$ and $\overline{Q}_{j-1}$ are, respectively, for the j and $(j-1)$th months. The term r_j represents the correlation between flow values $Q_{i,j}$ and $Q_{i,j+1}$, while the term S_j is the standard deviation of Q_j; t_j is a normal random deviate with zero mean and unit variance, and b_j is the slope of the regression line between Q_j and Q_{j+1}.

To improve the applicability of the model to skewed hydrologic series, Thomas and Burden [73] suggested a transformation to replace the normal random variate t_j, with a random variate that follows an approximately gamma distribution, making the model adaptable to different distributions. Later, in 2012, an attempt was made to modify the Thomas-Fiering model by extracting the persistence characteristics of a river and analyzing its serial correlation between the random residuals of the historical monthly mean records. The results showed that the new approach to the Thomas-Fiering model was more effective in preserving the monthly and annual standard deviation of the historical data, as well as its frequency distribution [74].

Without a doubt, the model by A. Thomas and B Fiering [11] marked a pioneering step in addressing the issue of dependency within streamflows; however, a notable challenge lay in their approach of treating flows for each month as *individual populations*. Moreover, the Thomas and Fiering model is a first-order Markov process, meaning that the streamflow at a given time step depends only on the previous time step. Therefore, it ignores higher-order dependencies.

Initiated by the work of Thomas and Fiering [11], the historical evolution of AR (Autoregressive) models commenced within hydrology during the 1960s. Box and Jenkins [75] proved a significant catalyst for time series analysis through their classic textbook on Time Series Analysis, Forecasting and Control. Consequently, AR models come in several forms, with the most prevalent being shown as $y_t = \mu + \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + \varepsilon_t$ by Fiering and Jackson [5]. In this model, y_t represents the estimated value of streamflow at time t, using $(t - j)$ lags of the realization time series y_{t-j}, where $\mathrm{E}[y_t] = \mu$ and ϕ_j signifies the autoregressive coefficient, while p denotes the order of the autoregressive process. The stochastic component (ε_t) of the equation is the time-independent series, which is independent of y_t. The values of ϕ_js are derived based on the assumption of normality.

While AR models could describe flow recession based on past values, these models may not fully account for high flow variations caused by rainfall and snowmelt. To better capture the variability in hydrologic time series, the moving average (MA) component is added to the AR models, forming the ARMA model. However, a key limitation of ARMA models is their assumption of stationarity, which may not hold for hydrologic data that exhibit seasonality.

In the 1970s, Hipel et al. [76] used the ARMA model for streamflow synthesis by first removing the seasonal nonstationarity in the data and then developing an ARMA model for the resulting time series. Lawrance and Kottegoda [1] highlighted theoretical challenges in the deseasonalization process, particularly its assumption that the transformed series is stationary. This implies that correlations between flows in different months, such as January and August, remain unchanged, an assumption that contradicts both theoretical models and empirical findings [77]. Despite this limitation, the AR family of models have been widely employed for streamflow synthesis across the globe.

Hirsch [78] used AR and ARMA models, which are quasi-stationary, meaning their statistical properties vary by *season* but remain consistent across years, to synthesize monthly streamflow using three approaches: log-normal, log transform, and normalizing transform. Synthesizing data for six rivers in the Potomac River at Point of Rocks revealed that AR (1) with a normalizing transform performed the best among the methods tested.

During the 1980s, the use of the AR family of models reflected a balance between simplicity, statistical rigour, and practical application. Early in the decade, Stolte [79] emphasized the utility of simple AR models over ARMA due to their ease of use and better alignment with engineering judgment, despite acknowledging their limitations in capturing complex streamflow dynamics. Muzik [80] similarly critiqued the reliability of stochastic models like AR (1) for reservoir design, noting the challenges of model validation and estimation under uncertainty, although the potential of simulations to incorporate variability and risk was recognized.

As the decade progressed, research increasingly turned to more sophisticated methods that addressed the shortcomings of early models. Stedinger and Taylor [34] demonstrated that parameter uncertainty had a greater impact on reservoir reliability estimates than the choice of AR-family models themselves. Lettenmaier and Burges [18], using ARMA-based disaggregation techniques, highlighted limitations in seasonal disaggregation and mass conservation. Further studies, such as that of Stedinger et al. [81], introduced Bayesian approaches to incorporate parameter uncertainty in disaggregated models, while Stedinger et al. [82] found that univariate ARMA(1,1) models with disaggregation could rival more complex multisite models. Meanwhile, Kalman filtering techniques were successfully applied by Bergman and Delleur [83] to enhance AR model forecasting through adaptive parameter estimation.

By the late 1980s, the AR-family modelling landscape had broadened to include more flexible and statistically robust approaches tailored to specific hydrologic and geographic

contexts. Salas et al. [5] explored multivariate AR(1) and contemporaneous ARMA models, showing their effectiveness in preserving key statistics. The use of computer-intensive methods like the Jackknife and Bootstrap by Cover and Unny [84] allowed for deeper insights into parameter sensitivity and model robustness. Bowles et al. [85] proposed model selection strategies based on persistence measures, while Vogel and Stedinger [64] highlighted that AR(1) log-normal models could enhance the precision of reservoir design despite potential misspecification. Toward the end of the decade, Fernandez and Salas [86] validated the Gamma-autoregressive (GAR(1)) model as a practical alternative for modelling skewed, dependent streamflow series across diverse basins, showing that bias-corrected estimators could eliminate the need for normal transformations and improve simulation fidelity.

During the 1990s, hydrologists began addressing the known limitations of traditional AR and ARMA models, such as poor representation of seasonal variability, skewness, and spatial dependence, by developing and applying more advanced variants. Mujumdar and Kumar [56] demonstrated the need for tailored model structures across different river basins, finding that higher-order AR and ARMA models outperformed AR(1) in capturing monthly and ten-day streamflow variability. Despite the lower mean square error of AR(1), more complex models such as AR(4) and ARMA(3,1) were needed to preserve temporal dependencies, and model adequacy was statistically validated across all tested cases. At the same time, as noted earlier, Fernandez and Salas [86] introduced the GAR(1) model to overcome the restrictive Gaussian assumption in traditional ARMA models. Applied to river systems across multiple continents, GAR(1) effectively captured both dependence and skewness without requiring data transformations. This model marked a practical improvement in reproducing higher moments and maintaining statistical fidelity in skewed annual streamflow data.

Progress in the decade continued with innovations targeting seasonal and computational limitations of ARMA-family models. Santos and Salas [87] proposed parsimonious stepwise disaggregation methods that retained autocorrelation and cross-correlation structures while significantly reducing computational demands, making them suitable, especially, for large datasets with bimonthly resolution. Later, Rasmussen et al. [88] advanced the use of Periodic Autoregressive Moving Average (PARMA) models, incorporating seasonality through periodic moment equations, and highlighting challenges in parameter estimation for high-frequency data. To further increase flexibility, Tasker and Dunne [89] applied a nonparametric bootstrap method to the residuals of the PARMA model, enabling the generation of synthetic flow series that retained both spatial and temporal correlation structures without assuming normality. Together, these studies reflect a shift in the 1990s from purely linear stochastic models toward more robust, flexible approaches capable of handling real-world complexities in hydrologic data.

In the 21st century, the AR-family of models were still used in two primary ways: as benchmarks for comparison [90–96] and as components within hybrid modelling frameworks [14,97–100] During the 21st century, Artificial Intelligence (AI) and machine learning (ML) models became increasingly popular and eventually dominant in streamflow synthesis.

2.3. ERA-2 (21st Century): The Rise and Domination of AI/ML Models

The emergence of AI models set the stage for the development of artificial neural networks (ANN). The inception of ANN dates to 1943, when McCulloch and Pitts proposed a theory based on networks of binary processing elements called neurons. However, despite its early development, ANN had not been utilized by hydrologists up to the early 1990s [101].

It is worth noting that prior to this, ANN had been utilized for forecasting purposes as early as 1992. Raman and Sunilkumar [102] noted that many conventional techniques for time series analysis assume linear relationships among variables. However, real-world temporal data often exhibit complex and nonlinear patterns that are challenging to analyze and predict accurately, and linear models and their combinations may prove insufficient in describing the behaviour of such data. Therefore, it was a logical step to employ nonlinear models such as neural networks, which are better suited to tackle the intricacies of the real world exhibiting temporal variation. Hence, twelve distinct neural networks were developed by Raman and Sunilkumar [103], one for each month of the year, to synthesize streamflow. In a comparative analysis with traditional ARMA, it was found that the ANN model yielded comparable results.

Although early efforts in semi-automated pattern recognition during the 19th century demonstrated pseudo forms of AI [31,55], AI had not been widely applied in synthetic hydrology until recently. Early in the 21st century, Jardim et al. [104] applied K-means clustering to monthly streamflow sequences generated by an autoregressive model to reduce the computational burden in mid-term hydroelectric planning—arguably the first clear application of AI in this field.

This period marked the beginning of a transition from classical stochastic models to hybrid and AI-enhanced frameworks. Early hybridization efforts combined traditional time series models like PAR(1) with bootstrapping methods [100] or integrated ARMA-generated flows into adaptive neuro-fuzzy inference systems [90]. Artificial neural networks (ANNs) were increasingly deployed to capture nonlinear dependencies and to simulate complex flow behaviours, outperforming traditional models such as ARMA and Thomas-Fiering in many cases [105,106]. As computational capacity grew, the ability of AI to synthesize hydrologic memory, extreme events, and spatial-temporal patterns more realistically than conventional models established its growing dominance.

Upon confirmation that the performance of artificial neural networks (ANN) in streamflow synthesis surpasses that of traditional models, the attention of researchers naturally shifted towards addressing any limitations, such as those that ANN suffers from overdependence on data size and the possibility of being trapped in local minima. The bootstrap technique has been recommended to overcome the former [107], and the Support Vector Machine (SVM) as a novel solution to the latter [108].

In the Bootstrap techniques, a set of Bootstrap between 50 to 200 is created by sampling with replacement. Each bootstrap sample may include multiple copies of some observations and no copies of other observations. An ANN is then applied to each Bootstrap sample, and the final synthesized streamflow at each time step will be the average of the ANN outcomes [107,109]. As mentioned, in addition to the large data set requirement, ANNs have the limitation of overfitting since they may become trapped in local minima. The main reason why this may happen is the fundamental optimizing method in ANNs-Empirical Risk Minimization (ERM). ANNs seek the minimum error between the actual output vector and the expected one, and Support Vector Machine (SVM) attempts to minimize the upper bound of the generalization error [110]. Apart from global optimization, the main advantage of SVM is that it uses Kernel Tricks to formulate the nonlinear variants.

Another way to overcome overfitting is fuzzy logic. The fuzzy-rule structure and lower parameter complexity often lead to better generalization. Keskin et al. [90] demonstrated that using an Adaptive Neuro-Fuzzy Inference System (ANFIS) for synthetic streamflow generation yields superior performance compared to stand-alone ARMA or ANN models. The fuzzy-rule-based structure of ANFIS enables better generalization, particularly in the presence of nonlinear hydrological data.

The nonlinear nature of streamflow has increasingly driven researchers to adopt AI-based models. Bourdin et al. [111] highlighted a paradigm shift from empirical models to more flexible, ensemble-based and physically informed systems, especially under uncertainty and climate change. Hao and Singh [112] tackled spatial and temporal dependence through a maximum entropy copula approach, accurately capturing multisite monthly streamflow structures. Similarly, Kirsch et al. [13] used the Modified Fractional Gaussian Noise (mFGN) model to simulate seasonal autocorrelations for planning under climate change scenarios, showing how nonlinear variability in inflows directly affects transfer operations.

Advanced models like the MRS—Multimodal Regression-Sampling [113] and M3EB—multi-site multi-season Maximum Entropy Bootstrap [114] combine regression, k-NN resampling, entropy bootstrapping, and transformation techniques to handle both stochasticity and nonlinearity across multiple locations and seasons. You et al. [115] and Marković et al. [116] addressed ecological and hydrologic flow realism using disaggregation and multi-step nonparametric methods that preserved autocorrelation, extreme events, and flow transitions across timescales.

Other innovations include the Hydraulic Mixing-Cell (HMC)-based tracer model by Partington et al. [117], which disaggregates flow contributions in storage-driven basins, and the climate-informed stochastic model by Stagge and Moglen [118] that links GCM predictors to nonlinear streamflow states. In data-limited environments, Patskoski and Sankarasubramanian [103] combined paleo-hydrological reconstructions with Bayesian updating to better estimate extreme drought conditions.

Recently, Long Short-Term Memory (LSTM) networks have gained popularity due to their ability to capture both short- and long-term dependencies in time series. This capability is particularly valuable in hydrology, where seasonal memory, lagged effects, and nonlinear interactions, such as those between rainfall, snowmelt, and soil moisture, play a critical role in streamflow dynamics. Molina et al. [119] proposed an LSTM-based approach to retrospectively estimate daily streamflow in ungauged or data-scarce watershed segments. Their approach models watersheds as interconnected upstream-downstream pairs, allowing the network to leverage downstream flow data to improve upstream estimates.

As a valuable alternative to AI/ML models by directly manipulating the temporal structure of streamflow series, transformation-based methods, particularly wavelet and frequency-domain techniques, for synthetic streamflow generation have also been popular. Wang et al. [120] introduced a nonparametric wavelet-based approach that decomposes and reconstructs daily flows using annual sampling of wavelet coefficients, effectively preserving statistical properties without assuming linearity. Similarly, Niu and Sivakumar [121] used Morlet wavelet decomposition to distinguish between high- and low-flow components, demonstrating strong performance at shorter time scales. In the frequency domain, Brunner et al. [122] employed phase randomization of de-seasonalized streamflow series, combined with a kappa distribution, to simulate realistic daily flows. This was extended by Brunner and Gilleland [123] in the Phase Randomization Simulation using the wavelets (PRSim-wave) model, which accurately captured spatio-temporal dependence and extremes across hundreds of U.S. catchments, proving suitable for regional-scale water planning.

These transformation-based models are increasingly favoured for their capacity to handle complex hydrological behaviours, including non-stationarity and long-memory processes, often without the high parameterization or training data demands of AI models. Abdelaziz et al. [97] compared wavelet and Discrete Fourier Transform (DFT)-based methods, finding both effective at modelling short- and long-term dependencies, with DFT offering marginally better results. Collectively, these studies highlight a shift toward hybrid

or transformation-based modelling strategies that can complement or even outperform machine learning approaches like LSTM or ANFIS in specific contexts, especially when interpretability, spectral fidelity, or data limitations are of concern.

Figure 2 provides a chronological overview of major streamflow synthesis approaches, tracing their evolution from early stochastic and conceptual models to modern AI/ML-based frameworks.

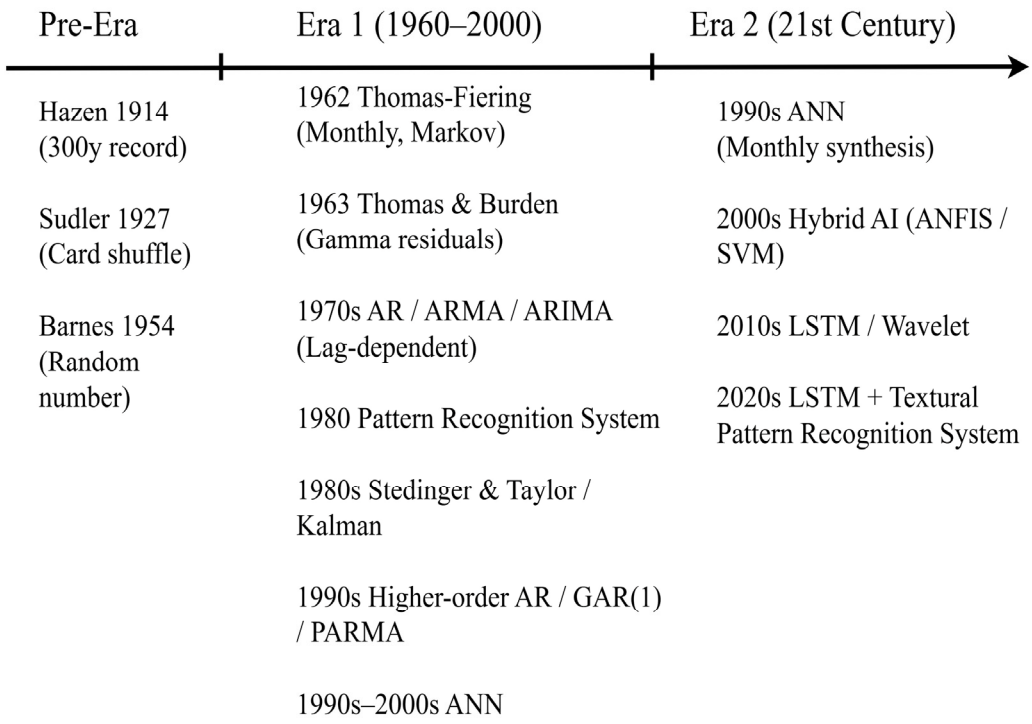


Figure 2. Chronological overview of streamflow synthesis approaches (Data from [11,70–73,81,82]).

As these models become more data-driven and automated, there is a growing risk of rendering hydrologists mere technicians, implementing models without fully understanding their inner workings or limitations. To address this, the community has begun to pursue two complementary strategies. First, hydrologists continue using advanced AI/ML models but with interpretability frameworks integrated, ensuring that the underlying relationships remain transparent and physically meaningful (Section 2.3.1). Second, hydrologists develop semi-automated models that balance computational power with domain expertise, allowing researchers to remain actively involved in the modelling process (as further elaborated in Section 3.1.2).

2.3.1. Interpretability of Hydrological AI/ML Models

The increasing application of AI/ML in hydrology has introduced powerful tools for nonlinear modelling, pattern recognition, and synthetic streamflow generation. ML approaches have demonstrated advantages over traditional physically based and conceptual models in terms of flexibility, predictive accuracy, and generalization. However, despite these benefits, ML models face significant challenges, particularly regarding interpretability [124]. The reliability and transparency of these models depend critically on four interrelated aspects: validation protocols, handling of distribution shifts, uncertainty quantification, and generalization ability.

Validation Protocols

Robust validation is fundamental to ensuring that model performance reflects genuine predictive capability rather than overfitting to historical data. Traditional approaches, such as split-sample validation [90,102], divide datasets into training, validation, and testing subsets to balance model fitting and evaluation. More advanced frameworks employ cross-validation or bootstrap resampling to estimate generalization error and confidence intervals [107]. In hybrid and stochastic models, such as those using the moving block bootstrap [100,109], validation also includes assessing the preservation of temporal and statistical dependencies (mean, variance, autocorrelation, and skewness). In recent deep learning models (e.g., Molina et al. [119]), validation extends to testing on independent datasets and evaluating metrics such as the Nash–Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE) under different flow regimes.

Handling Distribution Shifts

AI/ML hydrological models are often trained under the assumption of stationarity, where statistical properties of the inputs and outputs remain consistent over time. However, distribution shifts—arising from climate change, land-use alteration, or anthropogenic interventions—can degrade model performance. Most early models (e.g., Jardim et al., [104]; Raman & Sunilkumar, [102]) did not explicitly account for nonstationarity. Later developments, such as the mFGN approach [13], introduced mechanisms to simulate future hydro-climatic scenarios by adjusting mean and variance. Modern strategies to handle distribution shifts include data augmentation, transfer learning, and adaptive retraining using updated hydro-climatic records, which can enhance model robustness under changing environmental conditions.

Uncertainty Quantification

Quantifying uncertainty is essential for translating AI/ML predictions into actionable hydrological insights. Early neural network models primarily assessed performance deterministically using statistical indices (e.g., MSE, R^2), with limited consideration of predictive uncertainty. Subsequent frameworks incorporated stochastic sampling, bootstrap ensembles, and nonparametric residual resampling [107,109] to estimate variability in model outputs. More recent methodologies employ probabilistic models and Bayesian neural networks to explicitly represent epistemic (model-related) and aleatory (data-related) uncertainties. Such probabilistic approaches enable the generation of prediction intervals, risk estimates, and confidence bounds—crucial for reservoir management and climate impact studies.

Additionally, explainable AI techniques, such as Shapley Additive Explanations (SHAP), have been increasingly used to identify the input variables driving predictive uncertainty, providing both interpretability and actionable insight into why certain predictions are more uncertain than others [125,126]. Fan et al. [125] applied SHAP values to quantify the contributions of hydrometeorological variables in reservoir inflow modelling. Their analysis revealed that, for snowmelt-dominated reservoirs, past inflow was the dominant predictor, whereas in rainfall-driven reservoirs, both inflow and precipitation were critical.

Generalization Challenges

A persistent challenge in hydrological AI/ML models is ensuring generalization—the ability of a model trained on one basin, time period, or climatic condition to perform reliably in others. Many early models demonstrated satisfactory performance within the calibration basin but exhibited limited cross-basin or temporal transferability [102]. Hybrid approaches such as ANFIS + AR models [90] and ANN + Moving Block Bootstrap (MBB)

frameworks [109] showed improved adaptability by combining deterministic learning with stochastic variability. Recent developments in deep learning architectures, including LSTM networks [119], demonstrate higher generalization potential through dynamic learning of temporal dependencies and flow patterns. Nevertheless, ensuring model robustness under unseen hydrologic conditions remains an open research frontier, requiring both methodological innovation and diverse, high-quality datasets.

To sum up, interpretability is a key consideration in the application of AI/ML models in hydrology, as it directly affects the reliability and actionable value of synthesized streamflow. Therefore, while AI/ML models can be used for streamflow synthesis, their usefulness is limited without proper interpretability, since understanding the drivers of synthesized streamflow and associated uncertainties is essential for informed decision-making.

3. Approaches to Streamflow Synthesis

Streamflow synthesis has been approached through a wide range of modelling strategies, each shaped by different assumptions, data requirements, and computational capabilities. These approaches vary in terms of methodology, time scale, model structure, and the extent to which they incorporate domain knowledge or learning from data. The following subsections provide a structured overview of key distinctions among streamflow synthesis methods.

Figure 3 categorizes streamflow synthesis methods across different perspectives, which are discussed in detail in the following sections. In this figure, methods are classified according to several perspectives: the Paradigm Perspective, distinguishing traditional versus AI-based approaches (Section 3.1); the Pre-specified Probability Distribution Perspective, separating parametric and non-parametric methods (Section 3.2); the Time Scale Perspective, which considers annual, monthly, daily, or hourly resolutions (Section 3.3); the Temporal Generation Process Perspective, differentiating directly synthesized methods from disaggregation models (Section 3.4); and the Feature-Generation Perspective, distinguishing methods that synthesize streamflow directly (without extracting features) from those that first extract features using a pattern recognition system before generating the streamflow series (Section 3.5).

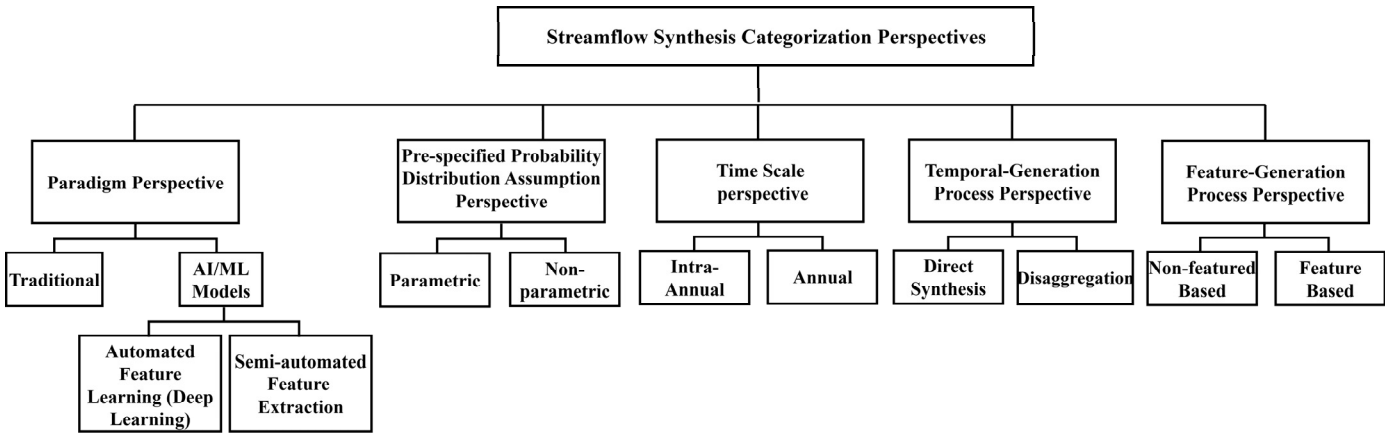


Figure 3. Categorization of streamflow synthesis methods.

3.1. Traditional Versus AI

Streamflow synthesis methods can broadly be categorized into traditional statistical approaches and modern AI techniques. Traditional models, such as the Thomas-Fiering model, Markov chains, and autoregressive (AR) family models, like AR, ARMA, and ARIMA, rely on predefined statistical assumptions and structures. These methods typically model streamflow as a stochastic process and are valued for their simplicity, interpretability,

and ease of implementation. However, they often struggle to capture nonlinear patterns, non-stationarities, and complex dependencies inherent in hydrological systems.

In contrast, AI-based approaches, including ANNs, ANFIS, SVM, and, more recently, LSTM, are data-driven and flexible. These models do not require strict assumptions about data distribution and can learn from the patterns in recorded time series, making them well-suited for modelling nonlinear and dynamic behaviour. Despite their performance advantages, AI models may suffer from overfitting, a lack of interpretability, and a heavy reliance on large, high-quality datasets.

Figure 4 illustrates the evolving research trend in streamflow synthesis, highlighting how traditional models initially dominated the field, while AI-based approaches began to emerge and gain traction, particularly after 2000, reflecting a shift in modelling paradigms.

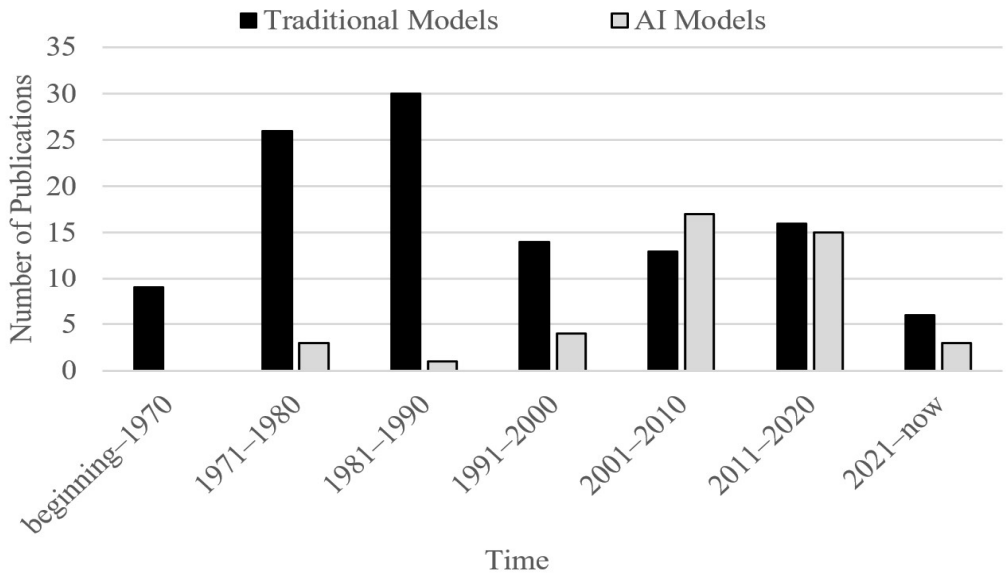


Figure 4. Temporal distribution of publications on streamflow synthesis using traditional and AI-based models.

The continued development of traditional models after 2010 highlights their fundamental simplicity, which makes them attractive for adaptation and improvement. For instance, approaches like the mFGN model [13,127] and hybrid methods such as periodic autoregressive parameters of the linear parametric model with normalizing transformation (PAR(1)-NT) combined with Moving Block Bootstrap [114] have been proposed to enhance their ability to represent dependence structures while maintaining conceptual clarity.

It can be said that traditional models are often more accessible to refine or hybridize due to their clarity and mathematical simplicity. In contrast, AI models offer greater flexibility and performance, particularly in handling nonlinear and dynamic patterns, but they are generally more difficult to interpret and control. Considering LSTM networks, their architecture is specifically designed to capture dependencies across varying time lags, enabling them to implicitly learn aspects of hydrologic memory embedded in streamflow time series. However, this representation remains largely opaque—the model learns and applies these patterns internally without explicitly revealing or encoding them in a human-interpretable form.

Without a doubt, the application of AI in hydrology has steadily been growing. AI methods are attractive due to their ease of implementation and their ability to produce desired outputs—such as synthesized streamflow—without requiring explicit parameter tuning or deep domain knowledge.

Moreover, the key difference between traditional models and AI/ML models lies in how they represent and maintain dependence structures. In traditional models, dependence is structurally encoded, such as through autocorrelation coefficients in AR-family models. In contrast, AI/ML models learn dependence from data via features, whether these features are implicitly learned within the complex architecture of deep learning models like LSTM or explicitly provided in a pattern recognition framework (Section 3.1.2).

3.1.1. Automatic Feature Learning (Deep Learning Models)

In deep learning models, temporal dependence is learned implicitly through internal weights and memory mechanisms, such as the memory cells in LSTM networks. There is no need for manual feature engineering; instead, the network automatically detects and learns from patterns embedded in input sequences.

LSTM architectures, in particular, are designed to capture dependencies across varying time lags, enabling them to implicitly learn aspects of hydrologic memory in streamflow time series. However, this representation remains largely opaque: the model internalizes and applies learned patterns without making them explicitly interpretable or accessible for hydrologic insight [128].

Two key limitations that arise from this approach are as follows.

- **Lack of explicit features**: While omitting feature engineering simplifies the modelling pipeline, it also means the model offers no internal explanation or analysis of the features it relies on. This limits interpretability and reduces the potential for scientific insight (and may also lead to oversight).
- **High data requirements**: Deep learning models generally require large volumes of high-quality training data to perform reliably. This can pose challenges in many hydrologic contexts, where historical records are short, sparse, or discontinuous.

Both limitations, the lack of explicit features and the need for large datasets, can be effectively addressed through a **semi-automated feature extraction approach**, which combines domain knowledge with data-driven pattern recognition. Unlike deep learning models such as LSTM, which often lack visibility into internal decision pathways, pattern recognition approaches provide a more transparent synthesis framework aligned with hydrologic reasoning.

3.1.2. Semi-Automated Feature Extraction (Pattern Recognition Models)

Pattern recognition techniques offer an alternative that can preserve interpretability. Due to their design, these models explicitly encode and reproduce identifiable patterns within streamflow data, allowing researchers to understand and trace how features such as persistence, seasonality, or flow texture are generated.

In a pattern recognition approach, streamflow at time t is synthesized using extracted features, introducing an additional step in the modelling process known as feature extraction. For instance, Panu et al. [54] extracted features by computing the first-order differences between adjacent observations, effectively representing short-term dynamics. Later, Li et al. [129] defined features using a six-day moving window, capturing the slope of a fitted curve to short flow sequences, thus characterizing local trends within the time series.

More recently, Studnicka and Panu [52,130] developed a novel approach by extracting textural features from streamflow time series. They encoded historical flow sequences as grayscale images and defined a simultaneous autocorrelation function that incorporates both the immediate past and corresponding month in the previous year and thus allowing for simultaneous representation of short- and long-term dependencies as features.

It can therefore be said that the semi-automated approach balances complexity and control. It ensures that domain knowledge remains an integral part of the modelling

process, rather than being fully replaced by data-driven algorithms. In addition, the semi-automated approaches offer greater control over the features being modified, allowing for more targeted and interpretable synthesis of streamflow series.

3.2. Parametric Versus Non-Parametric

Parametric approaches assume a specific probability distribution or functional form for the underlying data. These models require parameter estimation and are often more interpretable but may be restrictive if the data deviates from assumed distributions. Examples include:

- Thomas-Fiering model
- Autoregressive (AR), ARMA, and ARIMA models
- Pearson Curve Fitting
- Modified Fractional Gaussian Noise (mFGN)

Non-parametric approaches avoid explicit assumptions about data distribution and are often data-driven and flexible, suitable for complex, nonlinear, or poorly understood systems. Examples include:

- K-nearest neighbours (k-NN)
- Moving Block Bootstrap (MBB)
- Method of Fragments (MoF)
- Monte Carlo resampling without distribution fitting

Streamflow synthesis models can be broadly categorized into parametric and non-parametric approaches, each with distinct assumptions, methodologies, and applications.

Parametric approaches rely on the assumption that streamflow data follow specific probability distributions or time series structures. Such models are often built using statistical formulations such as autoregressive (AR), Markov chains, or multivariate distributions. A key strength of parametric models lies in their simplicity, interpretability, and relatively low data requirements. For example, Pereira et al. [131] utilized a multivariate lognormal AR(1) model for monthly streamflow disaggregation, enabling analytical evaluation of drought severity and economic impacts. Similarly, Kim and Valdes [132] proposed a semi-nonparametric (SNP) model to reproduce multimodal and persistent characteristics of monthly precipitation and temperature.

However, parametric models may struggle to capture the nonlinearity, asymmetry, and extreme behaviour present in many hydrological records. As a result, non-parametric approaches have gained prominence, especially since the 1990s, offering greater flexibility by avoiding rigid distributional assumptions. These models are data-driven and often rely on resampling techniques, such as the Moving Block Bootstrap [133], K-nearest neighbours [134], and kernel density estimation [135,136]. Non-parametric methods are particularly adept at preserving empirical characteristics like skewness, multimodality, long-term dependence, and time irreversibility, features that are difficult to incorporate into traditional frameworks.

Hybrid approaches have also emerged, combining parametric structure with non-parametric flexibility. For instance, Srinivas & Srinivasan [100,137] developed hybrid models that pre-whiten residuals using periodic AR models and resample them using matched or moving block bootstraps. These methods capture both linear autocorrelation and nonlinear variability, showing improved reservoir performance modelling.

Recent advances further extend non-parametric methods through optimization techniques. Borgomeo et al. [138] employed simulated annealing to control persistence and variability in synthetic monthly streamflow, while Mathai and Mujumdar [139] modelled multisite hydrograph asymmetry using a semi-parametric framework.

In summary, parametric models remain attractive for their clarity and computational efficiency, especially when system behaviour is well understood. In contrast, non-parametric and hybrid methods offer superior performance for complex, poorly defined systems, and are particularly effective in capturing diverse hydrologic behaviour across time scales and regimes.

To summarize the discussion above, Table 1 categorizes the major streamflow synthesis approaches according to whether they follow a parametric or non-parametric framework. It can be observed that most methods in the Pre-Era were parametric, while there is a clear tendency toward non-parametric approaches in more recent decades. It should also be noted that pattern recognition methods are generally classified as non-parametric, since they do not assume a specific distribution for the streamflow data itself. However, during the synthesis stage, the modeller often assumes a probability distribution to generate features. This assumption does not make the overall model parametric.

Table 1. Categorization of streamflow synthesis methods based on a parametric versus non-parametric perspective.

Era	Year/Period	Method/Model	Parametric Versus Non-Parametric
Pre-Era (Before 1960)	1914	Hazen	Parametric
	1927	Sudler	Non-parametric
	1954	Barnes	Parametric
Era 1 (1960–2000)	1962	Thomas & Fiering	Parametric
	1963	Thomas & Burden	Parametric
	1970s	AR, ARMA, ARIMA	Parametric
	1978 1980s	Semi-automated Pattern Recognition System	Non-parametric *
	1990s	Higher-order AR/ARMA, GAR (1), PARMA	Parametric
	1990s–2000s	ANN	Non-parametric
Era 2 (21st Century)	Early 2000s	Hybrid AI (ANFIS, ANN + PARMA, SVM, Bootstrap)	Non-parametric
	2010s	LSTM	Non-parametric
	2010s	Wavelet-based	Non-parametric
	2010s	mFGN	Parametric
	2025	Semi-automated Textural Image Pattern Recognition	Non-parametric *

* Features in this model are assumed to follow a specific probability distribution (e.g., normal) during the synthesis stage.

3.3. Timescale

The temporal resolution of streamflow data, whether daily, monthly, or annual, should be selected so that the optimal information can be extracted. Each of these time scales has different behaviour; for instance, in contrast to yearly and monthly streamflow time

series, there exist some rapid peaks and exponential recessions in daily streamflow time series [140].

A review of streamflow time series at various time scales shows the presence of various short- and long-term persistence characteristics, regular and irregular periodicities, linear and non-linear dynamics, and chaotic behaviour. Realizing that the governing force on various systems, including water flow in river systems, is affected by solar energy input. Each year, the solar energy input to the planet Earth is affected by the presence of several sunspots. In addition, other unknown and not well-understood interactions affect the solar energy input to planet Earth, which modulates the hydrological cycle, which, in turn, affects the water flow in our river systems around the globe. All such interactions of planet Earth and the solar system introduce numerous short- and long-term persistence along with regular and irregular dynamics and chaotic behaviour of water flow in rivers throughout the hydrological cycle. Additional exposition on the nature of streamflow data in hydrology has been provided, among others, by Panu & Unny, [30,31,55]; Piran & Panu, [32]; Yevjevich [2].

As streamflow is influenced by complex and interconnected hydrological processes that are not fully understood, consequently some characteristics of streamflow are also not well understood. However, the performance of synthesis models influences the choice of time interval, as well as the overall model design. A review of the literature reveals that most streamflow synthesis studies have focused on **monthly** time scales.

Figure 3 shows a comparison of different time scales in streamflow synthesis studies over time. This figure illustrates temporal trends in the use of various time scales. While monthly and annual models have been equally represented in the earliest period (pre-1970), monthly models have increasingly dominated in subsequent decades. This shift reflects a growing interest in capturing periodicity, seasonality, and the distinctive statistical characteristics associated with monthly streamflow patterns.

On the other hand, finer resolutions, such as daily and hourly streamflow synthesis, have received comparatively less attention over time. As shown in Figure 5, no studies used these resolutions before 1980. Daily models began to appear gradually, with a notable increase in adoption after 2000, peaking between 2001 and 2020. In contrast, hourly models remain rare, with only a few studies using this high-resolution scale, likely due to the complexity of modelling sub-daily hydrologic variability and the limited availability of quality data at such fine time intervals.

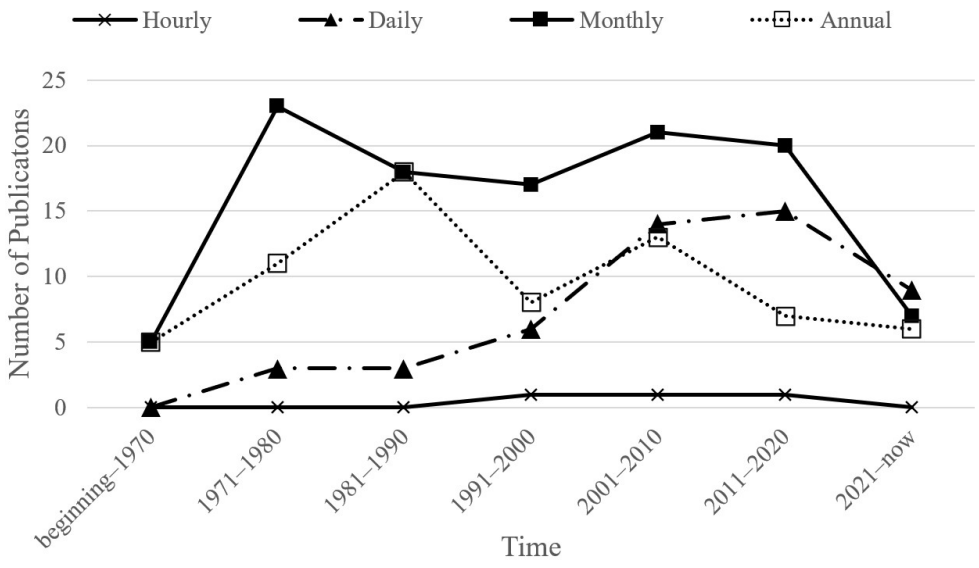


Figure 5. Distribution of streamflow synthesis studies by time scale and decade.

Moreover, instead of developing fully synthetic models at such fine time scales, hydrologists often rely on disaggregation techniques, using coarser flows and disaggregating them to monthly, weekly, daily or hourly resolutions.

3.4. Disaggregation Models

During the 1980s, hydrologists increasingly recognized that traditional stationary models such as simple AR and Markov processes were inadequate for capturing the periodic, non-stationary nature of streamflow data, particularly when modelling monthly and sub-monthly dynamics from annual series. This realization prompted a shift toward the use of disaggregation techniques. Researchers sought to generate realistic, high-resolution streamflow data that preserved important statistical properties and hydrologic dependencies. Several seminal studies emerged during this period.

Stedinger and Taylor [34] highlighted how uncertainties in annual streamflow parameters could significantly distort reservoir system performance metrics. They compared Thomas-Fiering and AR(1) models, showing that disaggregation approaches incorporating fractional Gaussian noise (FGN) and ARMA processes more effectively captured intra-annual and inter-annual dependencies. In parallel, responding to known flaws in the Mejia-Rousselle model [141], Stedinger and Taylor [34] proposed an improved disaggregation framework based on the Valencia-Schaake model [142], modified to include serially correlated innovations, thus enhancing its ability to represent covariance structures across time and space. Later, Lettenmaier [143] critically evaluated multi-site, multi-season models using graphical techniques, revealing biases in higher-order moments and mass balance inconsistencies. His work underscored the need for more robust seasonal disaggregation schemes.

Throughout the decade, practical implementations of the disaggregation approach matured. Stedinger et al. [81] introduced a simplified Bayesian disaggregation approach that accounted for parameter uncertainty, which was validated through applications on major U.S. rivers. Bowles et al. [85] demonstrated how disaggregated monthly flows could inform the operational design and drought resilience for Utah rivers. Grygier and Stedinger [144] expanded the toolkit by comparing multiple multisite disaggregation models, finding the SPIGOT and SPC models to rival the more established Valencia-Schaake method in realism and practicality. Savic et al. [145] illustrated that the accuracy of reservoir design models depended heavily on their ability to capture low-flow extremes, with disaggregation models again playing a central role in bridging the gap between statistical fidelity and operational reliability. Finally, Kuo and Sun [146] developed efficient single- and multi-month models to disaggregate monthly to ten-day flows, showing excellent mean flow preservation.

The evolution of AI also began to influence disaggregation modelling approaches in the 2000s. A notable example is the study by Ochoa-Rivera et al. [147], which implemented three different stochastic models, such as ARMA(1,1), Long-Term Correlated Disaggregation (LCD), and a Multilayer Perceptron Artificial Neural Network (MLP-ANN) to generate synthetic monthly inflows for simulating the operation of the Júcar River Water Resources System (WRS) using the AQUATOOL decision support system. Among the models evaluated, the MLP-ANN showed superior performance in preserving key hydrological features such as memory and drought persistence, outperforming the ARMA model and demonstrating results comparable to the more traditional LCD model. Importantly, the MLP required fewer parameters than LCD, making it both computationally efficient and easier to calibrate.

The need for finer temporal resolution in water resource management and hydrological modelling grew, and thus Nowak et al. [134] and You et al. [115] disaggregated a higher

resolution time scale to a daily streamflow level. Furthermore, You et al. [115] investigated the ability of synthetic streamflow models to reproduce hydro-ecological patterns using daily flow data from the Tamsui River watershed. A range of models was tested, and most of them performed well in simulating monthly/high flows, but low flows remained challenging. Annual to daily disaggregation provided good results for long-term simulation, while modified k-NN with nonparametric disaggregation captured flow variability effectively. Shot noise models are less favourable due to numerical biases.

Nowak et al. [134] used a simple, data-driven nonparametric approach that disaggregates annual flows directly into daily flows at multiple locations by resampling daily flow proportion vectors conditioned on annual flow. The method captures historical statistics well, generates continuous daily flows consistent with lag correlations, and preserves summability (additivity). The approach is computationally efficient and flexible, suitable for natural flow regimes. It compares favourably with other disaggregation techniques in accuracy and complexity.

These developments underscore the growing importance of high-resolution streamflow synthesis for modern water resource planning and ecological analysis. As hydrologists continue to seek realistic representations of flow variability, disaggregation models, especially those incorporating nonparametric and machine learning techniques, are proving essential for bridging temporal scales. The continued refinement of these methods will be crucial in addressing challenges related to climate variability, drought management, and integrated water systems modelling.

3.5. Transfer Learning: Pattern Recognition Approaches

In all approaches discussed thus far, whether parametric, non-parametric, traditional, or transformation-based (i.e., non-pattern recognition approaches), the streamflow (Q_t) at time t is synthesized as a function of its past values $Q_t = f(Q_{t-k})$, where Q_{t-k} represents past streamflow values with a time lag k. In other words, regardless of the model type, the function f is learned entirely from the same dataset on which it will be applied, and thus, it is a non-transfer learning.

On the other hand, *transfer learning* involves first extracting features from one dataset and then using those features for a different but related task, such as synthesis, prediction, or classification. In this context, the streamflow at time t can be represented as $Q_t = f(F)$, where F denotes the vector of extracted features.

Streamflow synthesis using a pattern recognition system, therefore, operates within a transfer learning framework. In pattern recognition techniques, features are first extracted from the streamflow data. These features are then synthesized based on the statistical properties of the features and their intra- and inter-pattern dependence structures. The synthesized features are subsequently transformed back into the streamflow series. Feature extraction stands as a pivotal step in the overall process, as it directly impacts the accuracy and reliability of the synthesized streamflow. In other words, the manner in which features are extracted significantly affects the accuracy of the results [148].

In general, feature extraction methods can be categorized into semi-automatic (manually with machine assistance) and automatic approaches. Semi-automatic feature extraction involves the expertise of domain specialists who, with the help of computational tools, manually select, devise, or design relevant features based on the objective, such as synthesis, forecasting or data infilling, and the contextual understanding of the specialist of the specific hydrologic time series. During the 1970s, a group of researchers [54] utilizing the concepts of pattern recognition pioneered the development of a semi-automated feature extraction model for stochastic streamflow synthesis. Subsequently, through a series of research contributions, Panu and Unny [30,31,55] further enhanced the efficacy of the

feature extraction model. The utility of pattern recognition concepts in the analysis and synthesis of time-dependent hydrologic data was further explored and enhanced through a seminal contribution to advances in hydroscience [140]. More recently, Studnicka and Panu [52,130], utilizing the concepts of textural pattern recognition, developed a novel stochastic streamflow synthesis model designed to capture both short-term and long-term temporal dependencies in streamflow data. The approach begins by encoding streamflow time series into textural grayscale images, where variations in flow magnitudes are represented as pixel intensities. A semi-automated feature extraction method is then applied to identify 2D correlations across both short- and long-term temporal dimensions.

It can therefore be said that the semi-automated approach balances complexity and control. It ensures that domain knowledge remains an integral part of modelling processes, rather than being fully replaced by data-driven algorithms. In addition, the semi-automated approaches offer greater control over the features being designed and/or modified, allowing for more targeted and interpretable synthesis of streamflow series.

To provide a clear overview of the various streamflow synthesis methods discussed above, Table 2 summarizes the main approaches, their key characteristics, and typical applications.

Table 2. Summary of Streamflow Synthesis Methods.

Approach	Key Features	Advantages	Limitations
Traditional Models (Thomas-Fiering, AR-Family, Markovian)	Parametric, linear regression, autoregressive time series, parametric	Simple, interpretable, low data requirement; captures autocorrelation	Limited to linear dependencies; assumes stationarity; may not capture extremes
AI/ML Models (ANN, ANFIS, LSTM, SVR/SVM, K-NN), and hybrid approaches	Data-driven, nonlinear, deep learning + fuzzy logic	Captures nonlinear and long-term dependencies; flexible; handles uncertainty	Requires large datasets; tuning complexity; opaque model (especially LSTM)
Decomposition Input Methods (Wavelet, Fourier/DFT-based)	Transformation-based, frequency domain, nonparametric	Preserves statistical properties; captures patterns at multiple scales	Computationally intensive; less intuitive; may not directly generate predictions
Pattern Recognition–Based Systems (Traditional and Textural)	Feature extraction, classification of flow patterns	Captures complex flow regimes; identifies recurring patterns; flexible for seasonal scale	Requires structured/labelled data

4. Evaluation Approaches for Synthetic Streamflow

Approaches to evaluate synthetic realizations are directly influenced by the understanding of historical streamflow characteristics and the ability to quantify those properties. For example, when the behaviour of a streamflow realization is known to be complex but cannot be quantified, such characteristics cannot be employed as evaluation criteria. Furthermore, the use of synthetic streamflow in operational hydrology has established a practical framework through which streamflow synthesis can be evaluated. Figure 6 presents various evaluation criteria used for assessing synthetic streamflow realizations. Moreover, the evolution of these criteria over time is presented in the section to follow.

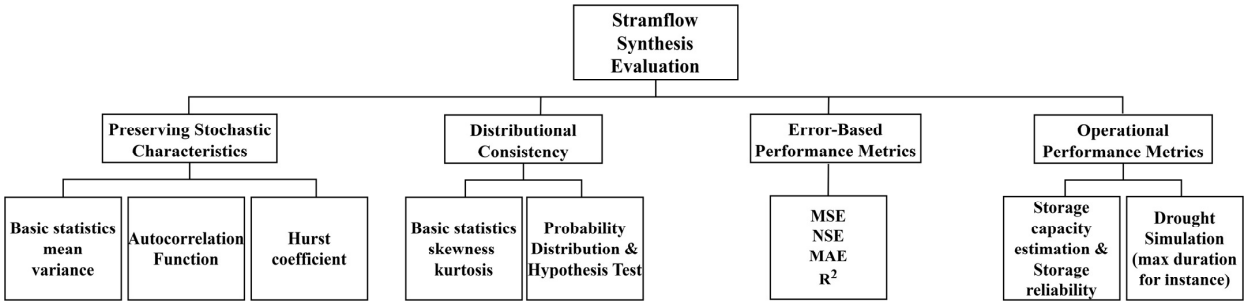


Figure 6. Classification of evaluation criteria commonly used to assess synthetic streamflow realizations.

The operational performance of the reservoir system can be quantitatively assessed using the metrics of reliability, resilience, and vulnerability [149]. *Reliability* measures the frequency with which the system successfully meets demand over the simulation period, reflecting the ability of the system to provide continuous service. *Resilience* quantifies how quickly the system recovers after a failure event, and *vulnerability* represents the average magnitude of deficit when failure occurs. Together, these indicators provide a comprehensive view of system performance under different hydrologic scenarios generated by the stochastic models. Furthermore, extreme event simulation, including drought duration, deficit volume, and flood magnitude, can be used to evaluate how models reproduce critical hydrologic extremes that affect reservoir operations. By integrating these quantitative metrics with the stochastic S–Y–R framework, the analysis extends beyond traditional reliability estimation to assess the practical robustness and adaptability of the reservoir system under variable and changing hydrologic conditions.

The consistent dominance of *Preserving Stochastic Characteristics* as an evaluation criterion is illustrated below in Figure 7, which highlights the foundational importance of maintaining key statistical properties, such as autocorrelation, variance, and long-term persistence, in synthetic streamflow generation. *Distributional Consistency* follows closely, as matching the statistical distribution (e.g., via PDF or CDF comparisons) ensures that synthetic flows reflect variability and extremes present in historical records. In contrast, *Error-Based Performance Metrics* attracted comparatively less attention, particularly in earlier decades, likely because they are primarily suited for calibration and validation in data-driven or machine-learning models. Their increased use after 2000 aligns with the broader adoption of machine learning approaches in hydrology, where minimizing error terms like MSE or NSE became central to model optimization.

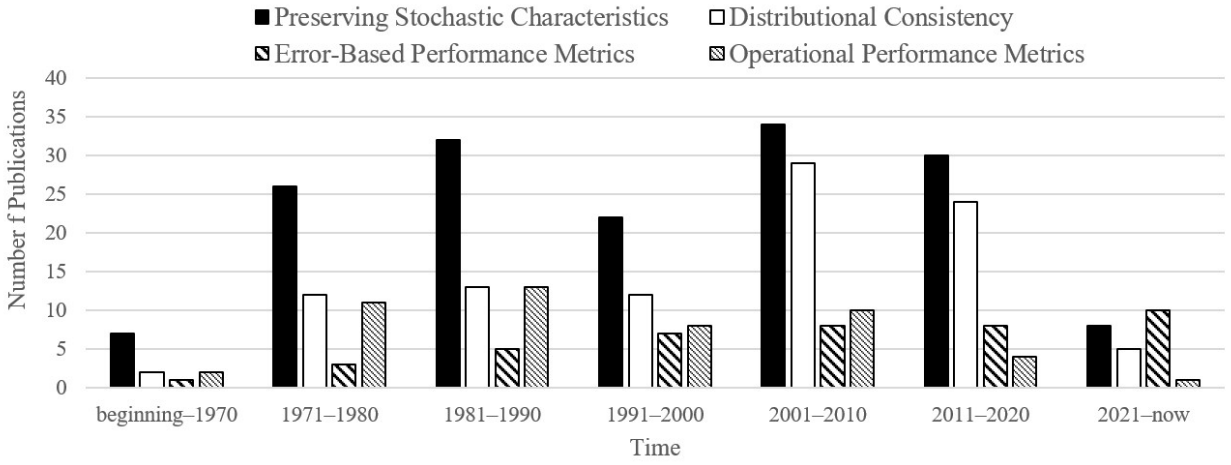


Figure 7. Temporal evolution of evaluation criteria used in synthetic streamflow studies.

The attention to matching probability distributions has led to the application of transformation techniques in streamflow synthesis, primarily rooted in the assumption of normality underlying many parametric approaches.

5. Transformation Techniques in Streamflow Synthesis

In response to the assumption of normally distributed data inherent in AR-family models, transformation techniques have evolved, which involve applying a transformation to convert non-normal data into a normal distribution; that is, the data transformation is converting the format of data and its structure or values to conform to a set of normally distributed data. Tools and technologies used for data transformation can vary widely based on the format, structure, complexity, and volume of the data being transformed. One commonly used transformation method is the Box-Cox transformation, which is defined as $y(\lambda) = \begin{cases} (y^{\lambda} - 1)/\lambda \; if \; \lambda \neq 0 \\ \ln(y) \; if \; \lambda = 0 \end{cases}$, where y represents the original data and λ is the transformation parameter, which varies from -5 to 5. All values of λ need to be considered, and the optimal value for the data should be selected. The optimal value is the one which results in the best approximation of a normal distribution [5]. To find the optimal value of λ, Hinkley [150] proposed a method based on choosing a symmetrizing transformation at which λ is restricted to -1, 0, 0.5, 1, and 2. A symmetrizing transformation is determined to make the degree of asymmetry in the sample ($d_{\lambda} = (mean - median)/sample\ scale$) as small as possible (approximately zero). However, it is important to note that finding the ideal λ value can be challenging, adding complexity to the transformation process. Additionally, the Box–Cox transformation is known to be sensitive to outliers [151]. In a mathematical contribution, Matalas [3] provided a solid foundation for creating synthetic streamflow, where he addressed essential aspects such as operational bias and the necessity for synthetic flows to adhere to a specific distribution. The fundamental generating process for synthetic flows was based on a lag-one Markov process, characterized by mean, variance, and lag-one autocorrelation. The significant focus on Markov chain models, which do not necessitate the transformation of data into a normally distributed realization, is evident in the literature [42,152–155].

The release of the book entitled "Time Series Analysis, Forecasting and Control," by Box and Jenkins in 1970, has been a catalyst for hydrological research to shift its focus toward parameter verification, condition assessment, and assumption validation of models [5]. This book brought increased attention to the study of models themselves, prompting researchers to delve deeper into their insights and provide a more detailed statistical and mathematical infrastructure for parameter estimation and verification [1,34].

5.1. Effects of Transformation in Streamflow Synthesis

Various transformation techniques can be applied as a pre-processing step to improve the statistical properties of streamflow data and enhance the performance of synthesis models. The following subsections summarize key effects of such transformations on streamflow data.

5.1.1. Improved Normality and Distributional Fit

Transformations, particularly log and Box-Cox, were widely used to reduce skewness and approximate normality, which is often assumed by parametric models like ARMA or GAR(1). For instance:

- Bayazit et al. [98] showed that Box-Cox transformations reduced original skewness (~0.74) to near zero, improving the consistency between simulated and historical distributions.

- Siegerstetter & Wahliß [156] successfully reproduced multivariate statistics (mean, SD, correlation) after transforming data using log and Pearson Type III methods.

5.1.2. Preservation of Statistical Characteristics

Many studies reported that transformation allowed for better reproduction of key statistical features:

- Green [157] noted that log transformation enhanced performance, particularly for general synthesis, though it remained inadequate for extremes.
- Marković et al. [116] showed improved fit to observed statistics (e.g., autocorrelation, cross-correlation, extremes) using log-transformed flows.

5.1.3. Reduced Negative Flow Rates

Transforming streamflow data helped reduce the incidence of negative flow, a common issue in untransformed models. For instance, Siegerstetter & Wahliβ [156] reported just 0.05% negative values after transformation.

5.1.4. Improved Model Stability and Parameter Estimation

Transformations contributed to numerical stability, particularly for small sample sizes or skewed data:

- Fernandez & Salas [86] showed that the GAR(1) model performed well without transformation due to bias correction, but most other studies still relied on transformation to avoid parameter bias.
- Guimarães & Santos [158] used Wilson-Hilferty to normalize flows, which helped produce accurate reservoir storage estimates with stable monthly statistics.

5.1.5. Limitations Remain for Extremes and Long-Term Persistence

While transformations generally improved model performance, reproduction of extremes and long-memory characteristics (e.g., persistence, Hurst behaviour) remains challenging:

- Green [157] acknowledged that even after transformation, performance for extreme/high flows was weak.

6. Publication Trends in Streamflow Synthesis Research

The reviewed studies were organized to highlight two key trends in the field of streamflow synthesis: the geographical distribution of applications and the journals most actively publishing related research.

6.1. Geographical Distribution of Streamflow Synthesis Studies

The number of streamflow synthesis studies sorted by country is displayed in Figure 8. The United States dominates the field, contributing 70 studies, more than four times the output of the next leading country, Canada (15 studies). This highlights the long-standing emphasis of various institutions on synthetic hydrology in the United States.

Since the emergence of streamflow synthesis techniques in the mid-20th century, North America has remained a focal point. However, in the 21st century, research activity has become increasingly globalized. Notably, countries such as India (11 studies), Turkey (8 studies) and China (6 studies) have shown growing engagement. In Europe, Germany leads with six contributions, reflecting its active participation in advancing synthetic hydrology. This trend demonstrates the expanding global interest in the synthetic streamflow field.

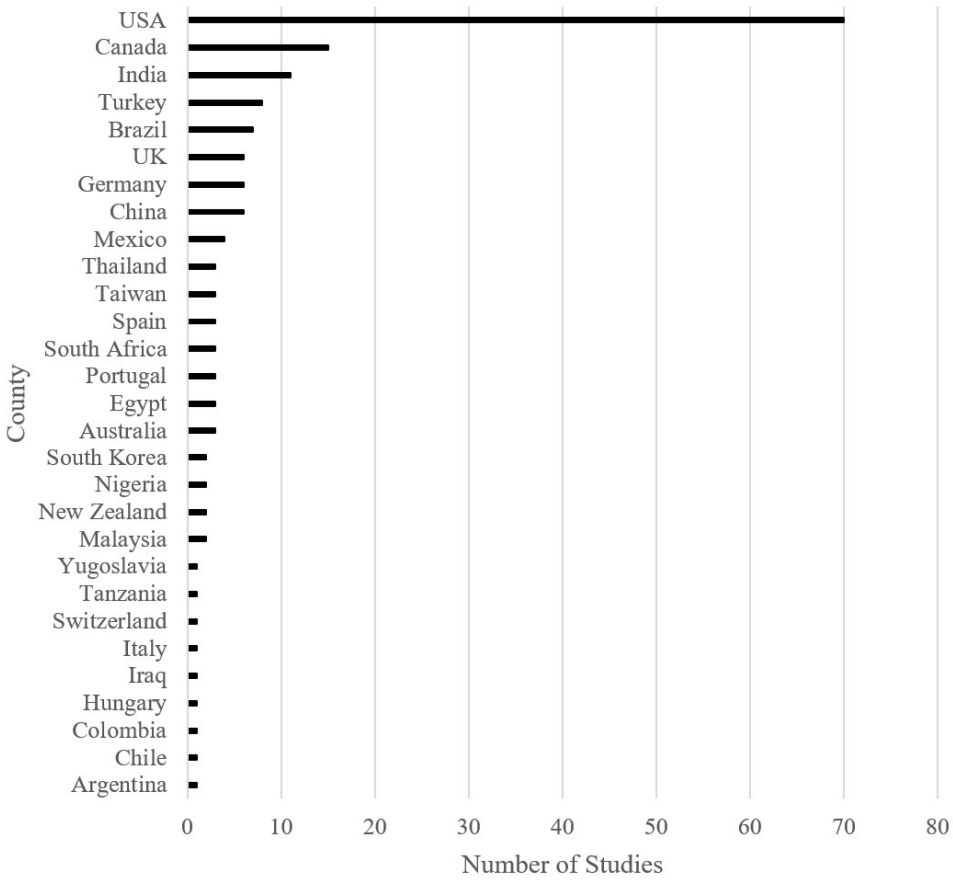


Figure 8. The number of streamflow synthesis studies by country.

6.2. Leading Journals in Streamflow Synthesis Research

This section considers all journals that have published studies related to streamflow synthesis, including those where synthetic methods were applied, discussed, or where the methodology itself was the focus of research.

Table 3 summarizes the number of publications and the first year of appearance for journals that have published research related to streamflow synthesis. The most active journals in this field are *Water Resources Research* (50 publications, first published in 1967), *Journal of Hydrology* (25 publications, since 1963), and the *Journal of the American Water Resources Association* (9 publications, since 1974). The *Journal of Hydrology* was the first to publish research on streamflow synthesis in 1963. Notably, the ***Journal of the Hydraulics Division***, which began publishing streamflow synthesis research in 1965, is no longer in publication.

Table 3. Summary of journals that have published research on streamflow synthesis.

Journal	Number of Publications	First Year of Publication
Water Resources Research	50	1967
Journal of Hydrology	25	1963
Journal of the American Water Resources Association	9	1974
Journal of Hydrologic Engineering	8	2000
Journal of the Hydraulics Division	8	1965
Hydrological Processes	8	1996

Table 3. *Cont.*

Journal	Number of Publications	First Year of Publication
Hydrological Sciences Journal	8	1990
Water Resources Management	7	1974
Stochastic Environmental Research and Risk Assessment	4	2008
Journal of Water Resources Planning and Management	4	1986
Hydrology and Earth System Sciences	3	2002
Canadian Journal of Civil Engineering	3	1980
Developments in Water Science	3	1982
Journal of Hydraulic Engineering	3	1990
Advances in water resources	3	2001
Stochastic Hydrology and Its Use in Water Resources Systems Simulation and Optimization	3	1993
Environmental Modelling & Software	2	2014
Hydrology Research	2	2011
Journal of Hydrology (New Zealand)	2	1968
Water	1	1990
Turkish Journal of Engineering & Environmental Sciences	1	2000
Environment International	1	1995
Regulated Rivers: Research & Management: An International Journal Devoted to River Research and Management	1	1999
Journal of Applied Statistics	1	2004
Journal of Spatial Hydrology	1	2005
Journal of irrigation and drainage engineering	1	2006
Atmosphere-Ocean	1	2012
Water and Environment Journal	1	2012
American Journal of Engineering Research	1	2013
Eur. Water	1	2017
Applied Water Science	1	2019
Authorea Preprints	1	2024

7. Concluding Remark

This review has comprehensively traced the historical evolution, methodological advancements, and evaluation practices in stochastic streamflow synthesis. From the early

reliance on autoregressive models to the contemporary integration of AI and machine learning, the field has demonstrated a dynamic progression aligned with both computational developments and hydrological demands. The increasing diversity in evaluation criteria, ranging from statistical consistency to operational applicability, reflects a maturing discipline that seeks both realism and utility. Moreover, the global expansion of research and the broadening journal presence indicate growing interest and relevance across regions and sectors. Moving forward, addressing challenges such as capturing extremes, incorporating non-stationarity, and enhancing data-driven model interpretability will be crucial in shaping the next era in synthetic hydrology.

8. Future Direction

The rapid advancement of AI/ML, particularly through tools such as ChatGPT and other automated systems, has significantly transformed research workflows. These technologies offer unprecedented support, streamlining tasks that once required substantial time and effort. However, this increasing reliance raises a critical question: **Are we at risk of becoming intellectually passive?**

While it is entirely appropriate and often necessary to leverage AI/ML tools, it is equally important to advocate for the development and use of **semi-automated models** such as the semi-automated textural pattern recognition model recently developed by Studnicka & Panu (2025b, 2025c). These models offer a balanced approach since they enhance efficiency while still requiring human input, interpretation, and critical engagement.

Such frameworks not only ensure that we maintain a deep understanding of our methodologies and data, but they also foster continued learning and encourage innovative problem-solving. It is, in this sense, that semi-automation acts as a safeguard against intellectual complacency and promotes the thoughtful integration of AI into research processes.

Therefore, future research should emphasize critical thinking. This could involve developing more semi-automated models that help uncover new aspects of hydrologic processes or creating enhanced quantification methods that make AI/ML tools more transparent and interpretable.

Author Contributions: S.S. and U.S.P. collaborated on various tasks related to this article, including draft preparation and finalization for journal publication, such as conceptualization and development of the methodology, data collection, and data analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research received partial support from the Natural and Engineering Research Council of Canada.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing does not apply to this article.

Acknowledgments: The authors gratefully acknowledge the partial support for this research by the Natural and Engineering Research Council of Canada.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lawrance, A.; Kottegoda, N. Stochastic modelling of riverflow time series. *J. R. Stat. Soc. Ser. A (Gen.)* **1977**, *140*, 1–31. [CrossRef]
2. Yevjevich, V. *Stochastic Processes in Hydrology*; Water Resources Publication: Fort Collins, CO, USA, 1972.
3. Matalas, N.C. Time series analysis. *Water Resour. Res.* **1967**, *3*, 817–829. [CrossRef]
4. Ledolter, J. *ARIMA Models and Their Use in Modelling Hydrologic Sequences*; IIASA: Laxenburg, Austria, 1976.

5. Salas, J.D.; Delleur, J.W.; Yevjevich, V.; Lane, W.L. *Applied Modelling of Hydrologic Time Series*; Water Resources Publication: Littleton, CO, USA, 1980.
6. Raudkivi, A.J. *Hydrology: An Advanced Introduction to Hydrological Processes and Modelling*; Elsevier: Amsterdam, The Netherlands, 2013.
7. Panu, U.S.; Unny, T. Extension and application of the feature prediction model for the synthesis of hydrologic records. *Water Resour. Res.* **1980**, *16*, 77–96. [CrossRef]
8. Sivakumar, B. *Chaos in Hydrology: Bridging Determinism and Stochasticity*; Springer: Berlin/Heidelberg, Germany, 2016.
9. Mandelbrot, B.B.; Wallis, J.R. Noah, Joseph, and operational hydrology. *Water Resour. Res.* **1968**, *4*, 909–918. [CrossRef]
10. Young, G.K.; Pisano, W.C. Operational hydrology using residuals. *J. Hydraul. Div.* **1968**, *94*, 909–924. [CrossRef]
11. Thomas, J.; Fiering, M.B. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In *Design of Water-Resource Systems: New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning*; Harvard University Press: Cambridge, MA, USA, 1962; pp. 459–493.
12. Fiering, M.B.; Bund, B. *Synthetic Streamflows*; American Geophysical Union: Washington, DC, USA, 1971; Volume 1.
13. Kirsch, B.R.; Characklis, G.W.; Zeff, H.B. Evaluating the impact of alternative hydro-climate scenarios on transfer agreements: Practical improvement for generating synthetic streamflows. *J. Water Resour. Plan. Manag.* **2013**, *139*, 396–406. [CrossRef]
14. Treistman, F.; Maceira, M.E.P.; Penna, D.D.J.; Damázio, J.M.; Rotunno Filho, O.C. Synthetic scenario generation of monthly streamflows conditioned to the El Niño–Southern Oscillation: Application to operation planning of hydrothermal systems. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 331–353. [CrossRef]
15. Hurst, H.E. Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Eng.* **1951**, *116*, 770–799. [CrossRef]
16. Hurst, H.E. The problem of long-term storage in reservoirs. *Hydrol. Sci. J.* **1956**, *1*, 13–27. [CrossRef]
17. Kirkby, M.J. The Hurst effect and its implications for extrapolating process rates. *Earth Surf. Process. Landf.* **1987**, *12*, 57–67. [CrossRef]
18. Lettenmaier, D.P.; Burges, S.J. Operational assessment of hydrologic models of long-term persistence. *Water Resour. Res.* **1977**, *13*, 113–124. [CrossRef]
19. Wallis, J.R.; Matalas, N.C. Small sample properties of H and K—Estimators of the Hurst coefficient h. *Water Resour. Res.* **1970**, *6*, 1583–1594. [CrossRef]
20. O'Connell, P.E. Stochastic Modelling of Long-Term Persistence in Streamflow Sequences. Ph.D. Thesis, University of London, London, UK, 1974.
21. Koutsoyiannis, D. The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrol. Sci. J.* **2002**, *47*, 573–595. [CrossRef]
22. Klemeš, V. The Hurst phenomenon: A puzzle? *Water Resour. Res.* **1974**, *10*, 675–688. [CrossRef]
23. Klemeš, V. One hundred years of applied storage reservoir theory. *Water Resour. Manag.* **1987**, *1*, 159–175. [CrossRef]
24. Boes, D.C.; Salas, J.D. Nonstationarity of the mean and the Hurst phenomenon. *Water Resour. Res.* **1978**, *14*, 135–143. [CrossRef]
25. Jackson, B.B. The use of streamflow models in planning. *Water Resour. Res.* **1975**, *11*, 54–63. [CrossRef]
26. Koirala, S.R.; Gentry, R.W.; Perfect, E.; Mulholland, P.J.; Schwartz, J.S. Hurst analysis of hydrologic and water quality time series. *J. Hydrol. Eng.* **2011**, *16*, 717–724. [CrossRef]
27. Legates, D.R.; Outcalt, S.I. Detection of climate transitions and discontinuities by Hurst rescaling. *Int. J. Clim.* **2021**, *42*, 4753–4772. [CrossRef]
28. Livina, V.; Kizner, Z.; Braun, P.; Molnar, T.; Bunde, A.; Havlin, S. Temporal scaling comparison of real hydrological data and model runoff records. *J. Hydrol.* **2007**, *336*, 186–198. [CrossRef]
29. Markonis, Y.; Moustakis, Y.; Nasika, C.; Sychova, P.; Dimitriadis, P.; Hanel, M.; Máca, P.; Papalexiou, S. Global estimation of long-term persistence in annual river runoff. *Adv. Water Resour.* **2018**, *113*, 1–12. [CrossRef]
30. Panu, U.S.; Unny, T. Stochastic synthesis of hydrologic data based on concepts of pattern recognition: I. General methodology of the approach. *J. Hydrol.* **1980**, *46*, 5–34. [CrossRef]
31. Panu, U.S.; Unny, T. Stochastic synthesis of hydrologic data based on concepts of pattern recognition: III. Performance evaluation of the methodology. *J. Hydrol.* **1980**, *46*, 219–237. [CrossRef]
32. Piran, S.; Panu, U. Encoded-Streamflow Synthesis Using Textural Feature Recognition System. *AGU Fall Meet. Abstr.* **2023**, *2023*, H44G-03c.
33. Sharma, A.; Tarboton, D.G.; Lall, U. Streamflow simulation: A nonparametric approach. *Water Resour. Res.* **1997**, *33*, 291–308. [CrossRef]
34. Stedinger, J.R.; Taylor, M.R. Synthetic streamflow generation: 1. Model verification and validation. *Water Resour. Res.* **1982**, *18*, 909–918. [CrossRef]
35. Suman, A.; Devarajan Sindhu, A.; Nayak, A.K.; Sankaran Namboothiri, A.; Biswal, B. Unveiling the climatic origin of streamflow persistence through multifractal analysis of hydro-meteorological datasets of India. *Hydrol. Sci. J.* **2023**, *68*, 290–306. [CrossRef]
36. Szolgayova, E.; Laaha, G.; Blöschl, G.; Bucher, C. Factors influencing long-range dependence in streamflow of European rivers. *Hydrol. Process.* **2014**, *28*, 1573–1586. [CrossRef]

37. Zhang, Q.; Xu, C.-Y.; Yu, Z.; Liu, C.-L.; Chen, Y.D. Multifractal analysis of streamflow records of the East River basin (Pearl River), China. *Phys. A Stat. Mech. Its Appl.* **2009**, *388*, 927–934. [CrossRef]
38. Balabana, E.; Lub, S. Colour of noise: Comparative analysis of sub-periodic variation in empirical Hurst exponent across foreign currency changes and their pairwise differences. *Preprint* **2018**. Available online: https://www.researchgate.net/profile/Shan-Lu-7/publication/328230754_Color_of_noise_Comparative_analysis_of_sub-periodic_variation_in_empirical_Hurst_exponent_across_foreign_currency_changes_and_their_pairwise_differences/links/5cd0c112458515712e973d7d/Color-of-noise-Comparative-analysis-of-sub-periodic-variation-in-empirical-Hurst-exponent-across-foreign-currency-changes-and-their-pairwise-differences.pdf (accessed on 25 October 2025).
39. Bullmore, E.; Long, C.; Suckling, J.; Fadili, J.; Calvert, G.; Zelaya, F.; Carpenter, T.A.; Brammer, M. Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. *Hum. Brain Mapp.* **2001**, *12*, 61–78. [CrossRef]
40. Dooley, K.J.; Van de Ven, A.H. *A Primer on Diagnosing Dynamic Organizational Processes*; Strategic Management Research Center, University of Minnesota: Minneapolis, MN, USA, 1997.
41. Koscielny-Bunde, E.; Kantelhardt, J.W.; Braun, P.; Bunde, A.; Havlin, S. Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. *J. Hydrol.* **2006**, *322*, 120–137. [CrossRef]
42. Rodriguez-Iturbe, I.; Rinaldo, A. *Fractal River Basins: Chance and Self-Organization*; Cambridge University Press: Cambridge, UK, 1997.
43. Gallant, J.C.; Moore, I.D.; Hutchinson, M.F.; Gessler, P. Estimating fractal dimension of profiles: A comparison of methods. *Math. Geol.* **1994**, *26*, 455–481. [CrossRef]
44. Blöschl, G.; Sivapalan, M. Scale issues in hydrological modelling: A review. *Hydrol. Process.* **1995**, *9*, 251–290. [CrossRef]
45. Dolgonosov, B.; Korchagin, K.; Kirpichnikova, N. Modelling of annual oscillations and 1/f-noise of daily river discharges. *J. Hydrol.* **2008**, *357*, 174–187. [CrossRef]
46. Gu, X.; Sun, H.; Tick, G.R.; Lu, Y.; Zhang, Y.; Zhang, Y.; Schilling, K. Identification and scaling behaviour assessment of the dominant hydrological factors of nitrate concentrations in streamflow. *J. Hydrol. Eng.* **2020**, *25*, 06020002. [CrossRef]
47. Kim, D.H.; Rao, P.S.C.; Kim, D.; Park, J. 1/f noise analyses of urbanization effects on streamflow characteristics. *Hydrol. Process.* **2016**, *30*, 1651–1664. [CrossRef]
48. Telesca, L.; Lovallo, M.; Lopez-Moreno, I.; Vicente-Serrano, S. Investigation of scaling properties in monthly streamflow and Standardized Streamflow Index (SSI) time series in the Ebro basin (Spain). *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 1662–1678. [CrossRef]
49. Thompson, S.E.; Katul, G.G. Multiple mechanisms generate Lorentzian and 1/fα power spectra in daily stream-flow time series. *Adv. Water Resour.* **2012**, *37*, 94–103. [CrossRef]
50. Wen, H.; Liu, Z. Separating fractal and oscillatory components in the power spectrum of a neurophysiological signal. *Brain Topogr.* **2016**, *29*, 13–26. [CrossRef]
51. Cuddington, K.M.; Yodzis, P. Black noise and population persistence. *Proc. R. Soc. Lond.* **1999**, *266*, 969–973. [CrossRef]
52. Studnicka, S.; Panu, U. Streamflow Synthesis Using an Encoded Textural Pattern Recognition System. II: Model Applications. *J. Hydrol. Eng.* **2025**, *30*, 04025040. [CrossRef]
53. Harms, A.A.; Campbell, T.H. An extension to the Thomas--Fiering Model for the sequential generation of streamflow. *Water Resour. Res.* **1967**, *3*, 653–661. [CrossRef]
54. Panu, U.S.; Unny, T.E.; Ragade, R.K. A feature prediction model in synthetic hydrology based on concepts of pattern recognition. *Water Resour. Res.* **1978**, *14*, 335–344. [CrossRef]
55. Panu, U.S.; Unny, T. Stochastic synthesis of hydrologic data based on concepts of pattern recognition: II. Application of natural watersheds. *J. Hydrol.* **1980**, *46*, 197–217. [CrossRef]
56. Mujumdar, P.; Kumar, D.N. Stochastic models of streamflow: Some case studies. *Hydrol. Sci. J.* **1990**, *35*, 395–410. [CrossRef]
57. Piran, S.; Panu, U. *Investigations into the Relationships Between Persistence, Complexity, and Scaling Behaviour in Monthly Streamflow Across Ontario, Canada*; Canadian Society for Civil Engineering (CSCE): Winnipeg, MB, Canada, 2025.
58. Elshorbagy, A.; Simonovic, S.; Panu, U. Estimation of missing streamflow data using principles of chaos theory. *J. Hydrol.* **2002**, *255*, 123–133. [CrossRef]
59. Jayawardena, A.W.; Lai, F. Analysis and prediction of chaos in rainfall and stream flow time series. *J. Hydrol.* **1994**, *153*, 23–52. [CrossRef]
60. Jiang, Y.; Bao, X.; Hao, S.; Zhao, H.; Li, X.; Wu, X. Monthly streamflow forecasting using ELM-IPSO based on phase space reconstruction. *Water Resour. Manag.* **2020**, *34*, 3515–3531. [CrossRef]
61. Li, H.; Bao, S.; Xuan, Y. Parameter selection for phase space reconstruction in hydrological series and rationality analysis of its chaotic characteristics. *EPiC Ser. Eng.* **2018**, *3*, 1171–1183.
62. Liu, Q.; Islam, S.; Rodriguez-Iturbe, I.; Le, Y. Phase-space analysis of daily streamflow: Characterization and prediction. *Adv. Water Resour.* **1998**, *21*, 463–475. [CrossRef]

63. Piran, S.; Panu, U. Textural Image-Based Feature Prediction Model for Stochastic Streamflow Synthesis. *Preprint* **2023**. [CrossRef]
64. Vogel, R.M.; Stedinger, J.R. The value of stochastic streamflow models in overyear reservoir design applications. *Water Resour. Res.* **1988**, *24*, 1483–1490. [CrossRef]
65. Satriani, S.; Lopa, R.; Maricar, F. Storage capacity analysis of Nipa Nipa regulation pond using the Rippl method. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1098*, 022054. [CrossRef]
66. Rippl, W. The capacity of storage-reservoirs for water-slpply. (including plate). In *Minutes of the Proceedings of the Institution of Civil Engineers*; Emerald Publishing Limited: Leeds, UK, 1833. [CrossRef]
67. Boughton, W.; McKerchar, A. Generating synthetic stream-flow records for New Zealand Rivers. *J. Hydrol. (N. Z.)* **1968**, 112–123. Available online: http://www.jstor.org/stable/43944152 (accessed on 25 October 2025).
68. Phien, H.N.; Ruksasilp, W. A review of single-site models for monthly streamflow generation. *J. Hydrol.* **1981**, *52*, 1–12. [CrossRef]
69. Wijayaratne, L.H.; Chan, P.C. Synthetic flow generation with stochastic models. In *Flood Hydrology, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, Louisiana State University, Baton Rouge, LA, USA, 14–17 May 1986*; Springer: Dordrecht, The Netherlands, 1987.
70. Hazen, A. Closure to Storage for Impounding Reservoirs. *Trans. Am. Soc. Civ. Eng.* **1914**, *77*, 1659–1669. [CrossRef]
71. Sudler, C.E. Storage Required for the Regulation of Stream Flow. *Trans. Am. Soc. Civ. Eng.* **1927**, *91*, 622–660. [CrossRef]
72. Barnes, F. Storage required for a city water supply. *J. Inst. Eng. Aust.* **1954**, *26*, 198–203.
73. Thomas, H.; Burden, R.P. *Operations Research in Water Quality Management*; Harvard Water Resources Group, Harvard University: Cambridge, MA, USA, 1963.
74. Arselan, C.A. Stream flow simulation and synthetic flow calculation by the modified Thomas Fiering model. *Al-Rafidain Eng. J. (AREJ)* **2012**, *20*, 118–127. [CrossRef]
75. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
76. Hipel, K.W.; McLeod, A.I.; Lennox, W.C. Advances in Box-Jenkins modelling: 1. Model construction. *Water Resour. Res.* **1977**, *13*, 567–575. [CrossRef]
77. Moss, M.E.; Bryson, M.C. Autocorrelation structure of monthly streamflows. *Water Resour. Res.* **1974**, *10*, 737–744. [CrossRef]
78. Hirsch, R.M. Synthetic hydrology and water supply reliability. *Water Resour. Res.* **1979**, *15*, 1603–1615. [CrossRef]
79. Stolte, W.J. The limitations and usefulness of streamflow generation methods: A case study. *Can. J. Civ. Eng.* **1980**, *7*, 185–191. [CrossRef]
80. Muzik, I. Analysis of capacity requirements for storage reservoirs: A case study. *Can. J. Civ. Eng.* **1980**, *7*, 388–392. [CrossRef]
81. Stedinger, J.R.; Pei, D.; Cohn, T.A. A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. *Water Resour. Res.* **1985**, *21*, 665–675. [CrossRef]
82. Stedinger, J.R.; Lettenmaier, D.P.; Vogel, R.M. Multisite ARMA (1, 1) and disaggregation models for annual streamflow generation. *Water Resour. Res.* **1985**, *21*, 497–509. [CrossRef]
83. Bergman, M.J.; Delleur, J.W. Kalman filter estimation and prediction of daily stream flows: I. Review, algorithm, and simulation experiments. *JAWRA J. Am. Water Resour. Assoc.* **1985**, *21*, 815–825. [CrossRef]
84. Cover, K.A.; Unny, T.E. Application of computer intensive statistics to parameter uncertainty in streamflow synthesis[1]. *JAWRA J. Am. Water Resour. Assoc.* **1986**, *22*, 495–507. [CrossRef]
85. Bowles, D.S.; James, W.R.; Kottegoda, N.T. Initial model choice: An operational comparison of stochastic streamflow models for drought. *Water Resour. Manag.* **1987**, *1*, 3–15. [CrossRef]
86. Fernandez, B.; Salas, J.D. Gamma-autoregressive models for stream-flow simulation. *J. Hydraul. Eng.* **1990**, *116*, 1403–1414. [CrossRef]
87. Santos, E.G.; Salas, J.D. Stepwise Disaggregation Scheme for Synthetic Hydrology. *J. Hydraul. Eng.* **1992**, *118*, 765–784. [CrossRef]
88. Rasmussen, P.F.; Salas, J.D.; Fagherazzi, L.; Rassam, J.; Bobée, B. Estimation and validation of contemporaneous PARMA Models for streamflow simulation. *Water Resour. Res.* **1996**, *32*, 3151–3160. [CrossRef]
89. Tasker, G.D.; Dunne, P. Bootstrap Position Analysis for Forecasting Low Flow Frequency. *J. Water Resour. Plan. Manag.* **1997**, *123*, 359–367. [CrossRef]
90. Keskin, M.E.; Taylan, D.; Terzi, O. Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series. *Hydrol. Sci. J.* **2006**, *51*, 588–598. [CrossRef]
91. Kottegoda, N.; Natale, L.; Raiteri, E. Daily streamflow simulation using recession characteristics. *J. Hydrol. Eng.* **2000**, *5*, 17–24. [CrossRef]
92. Ma, Y.; Zhong, P.-a.; Wang, G.; Xiao, Y. Performance of multisite streamflow stochastic generation approaches for a multi-reservoir system. *Stoch. Environ. Res. Risk Assess.* **2024**, *38*, 2135–2155. [CrossRef]
93. Ochoa-Rivera, J.; García-Bartual, R.; Andreu, J. Multivariate synthetic streamflow generation using a hybrid model based on artificial neural networks. *Hydrol. Earth Syst. Sci.* **2002**, *6*, 641–654. [CrossRef]

94. Pender, D.; Patidar, S.; Pender, G.; Haynes, H. Stochastic simulation of daily streamflow sequences using a hidden Markov model. *Hydrol. Res.* **2016**, *47*, 75–88. [CrossRef]
95. Porto, V.C.; de Souza Filho, F.d.A.; Carvalho, T.M.N.; de Carvalho Studart, T.M.; Portela, M.M. A GLM copula approach for multisite annual streamflow generation. *J. Hydrol.* **2021**, *598*, 126226. [CrossRef]
96. Prairie, J.R.; Rajagopalan, B.; Fulp, T.J.; Zagona, E.A. Modified K-NN model for stochastic streamflow simulation. *J. Hydrol. Eng.* **2006**, *11*, 371–378. [CrossRef]
97. Abdelaziz, S.; Mahmoud Ahmed, A.M.; Eltahan, A.M.; Abd Elhamid, A.M.I. Long-Term Stochastic Modelling of Monthly Streamflow in the River Nile. *Sustainability* **2023**, *15*, 2170. [CrossRef]
98. Bayazit, M.; Önöz, B.; Aksoy, H. Nonparametric streamflow simulation by wavelet or Fourier analysis. *Hydrol. Sci. J.* **2001**, *46*, 623–634. [CrossRef]
99. Pereira, G.; Veiga, A. PAR (p)-vine copula-based model for stochastic streamflow scenario generation. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 833–842. [CrossRef]
100. Srinivas, V.; Srinivasan, K. A hybrid stochastic model for multiseason streamflow simulation. *Water Resour. Res.* **2001**, *37*, 2537–2549. [CrossRef]
101. Khare, S.; Gajbhiye, A. Literature Review on Application of Artificial Neural Network (ANN) in the Operation of Reservoirs. *Int. J. Comput. Eng. Res. (IJCER) IJCER* **1943**, *3*, 63.
102. Raman, H.; Sunilkumar, N. Multivariate modelling of water resources time series using artificial neural networks. *Hydrol. Sci. J.* **1995**, *40*, 145–163. [CrossRef]
103. Patskoski, J.; Sankarasubramanian, A. Improved reservoir sizing utilizing observed and reconstructed streamflows within a Bayesian combination framework. *Water Resour. Res.* **2015**, *51*, 5677–5697. [CrossRef]
104. Jardim, D.; Maceira, M.; Falcao, D. Stochastic streamflow model for hydroelectric systems using clustering techniques. In Proceedings of the 2001 IEEE Porto Power Tech Proceedings (Cat. No. 01EX502), Porto, Portugal, 10–13 September 2001.
105. Ahmed, J.A.; Sarma, A.K. Artificial neural network model for synthetic streamflow generation. *Water Resour. Manag.* **2007**, *21*, 1015–1029. [CrossRef]
106. Awchi, T.A.; Srivastava, D. Artificial Neural Network Model Application in Stochastic Generation of Monthly Streamflows for Mula Project. In Proceedings of the International Conference on Water and Environment, Bhopal, India, 1999. [CrossRef]
107. Jia, Y.; Culver, T.B. Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. *J. Hydrol.* **2006**, *331*, 580–590. [CrossRef]
108. Deka, P.C. *A Primer on Machine Learning Applications in Civil Engineering*; CRC Press: Boca Raton, FL, USA, 2019.
109. Sudheer, K.; Srinivasan, K.; Neelakantan, T.; Srinivas, V. A nonlinear data-driven model for synthetic generation of annual streamflows. *Hydrol. Process. Int. J.* **2008**, *22*, 1831–1845. [CrossRef]
110. Deka, P.C. Support vector machine applications in the field of hydrology: A review. *Appl. Soft Comput.* **2014**, *19*, 372–386. [CrossRef]
111. Bourdin, D.R.; Fleming, S.W.; Stull, R.B. Streamflow modelling: A primer on applications, approaches and challenges. *Atmos.-Ocean* **2012**, *50*, 507–536. [CrossRef]
112. Hao, Z.; Singh, V.P. Modelling multisite streamflow dependence with maximum entropy copula. *Water Resour. Res.* **2013**, *49*, 7139–7143. [CrossRef]
113. Li, C.; Singh, V.P. A multimodel regression-sampling algorithm for generating rich monthly streamflow scenarios. *Water Resour. Res.* **2014**, *50*, 5958–5979. [CrossRef]
114. Srivastav, R.K.; Simonovic, S.P. An analytical procedure for multi-site, multi-season streamflow generation using maximum entropy bootstrapping. *Environ. Model. Softw.* **2014**, *59*, 59–75. [CrossRef]
115. You, G.J.-Y.; Thum, B.-H.; Lin, F.-H. The examination of reproducibility in hydro-ecological characteristics by daily synthetic flow models. *J. Hydrol.* **2014**, *511*, 904–919. [CrossRef]
116. Marković, Đ.; Plavšić, J.; Ilich, N.; Ilić, S. Non-parametric stochastic generation of streamflow series at multiple locations. *Water Resour. Manag.* **2015**, *29*, 4787–4801. [CrossRef]
117. Partington, D.; Brunner, P.; Frei, S.; Simmons, C.T.; Werner, A.D.; Therrien, R.; Maier, H.R.; Dandy, G.C.; Fleckenstein, J. Interpreting streamflow generation mechanisms from integrated surface-subsurface flow models of a riparian wetland and catchment. *Water Resour. Res.* **2013**, *49*, 5501–5519. [CrossRef]
118. Stagge, J.; Moglen, G. A nonparametric stochastic method for generating daily climate-adjusted streamflows. *Water Resour. Res.* **2013**, *49*, 6179–6193. [CrossRef]
119. Molina, A.A.R.; Frame, J.M.; Halgren, J.; Gong, J. A Proof of Concept for Improving Estimates of Ungauged Basin Streamflow Via an LSTM-Based Synthetic Network Simulation Approach. *J. Geophys. Res. Mach. Learn. Comput.* **2024**, *2*, e2024JH000405.
120. Wang, W.; Hu, S.; Li, Y. Wavelet transform method for synthetic generation of daily streamflow. *Water Resour. Manag.* **2011**, *25*, 41–57. [CrossRef]

121. Niu, J.; Sivakumar, B. Scale-dependent synthetic streamflow generation using a continuous wavelet transform. *J. Hydrol.* **2013**, *496*, 71–78. [CrossRef]
122. Brunner, M.I.; Bárdossy, A.; Furrer, R. Stochastic simulation of streamflow time series using phase randomization. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 3175–3187. [CrossRef]
123. Brunner, M.I.; Gilleland, E. Stochastic simulation of streamflow and spatial extremes: A continuous, wavelet-based approach. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 3967–3982. [CrossRef]
124. Yaseen, Z.M. A new benchmark on machine learning methodologies for hydrological processes modelling: A comprehensive review for limitations and future research directions. *Knowl.-Based Eng. Sci.* **2023**, *4*, 65–103. [CrossRef]
125. Fan, M.; Liu, S.; Lu, D.; Gangrade, S.; Kao, S.-C. Explainable machine learning model for multi-step forecasting of reservoir inflow with uncertainty quantification. *Environ. Model. Softw.* **2023**, *170*, 105849. [CrossRef]
126. Mehdiyev, N.; Majlatow, M.; Fettke, P. Quantifying and explaining machine learning uncertainty in predictive process monitoring: An operations research perspective. *Ann. Oper. Res.* **2025**, *347*, 991–1030. [CrossRef]
127. Chadwick, C.; Babonneau, F.; Homem-de-Mello, T.; Letelier, A. Synthetic simulation of spatially-correlated streamflows: Weighted-Modified Fractional Gaussian Noise. *Water Resour. Res.* **2023**, *60*, e2023WR035371. [CrossRef]
128. Girihagama, L.; Naveed Khaliq, M.; Lamontagne, P.; Perdikaris, J.; Roy, R.; Sushama, L.; Elshorbagy, A. Streamflow modelling and forecasting for Canadian watersheds using LSTM networks with an attention mechanism. *Neural Comput. Appl.* **2022**, *34*, 19995–20015. [CrossRef]
129. Li, F.-F.; Cao, H.; Hao, C.-F.; Qiu, J. Daily Streamflow Forecasting Based on Flow Pattern Recognition. *Water Resour. Manag.* **2021**, *35*, 4601–4620. [CrossRef]
130. Studnicka, S.; Panu, U. Streamflow Synthesis Using an Encoded Textural Pattern Recognition System. I: Model Development. *J. Hydrol. Eng.* **2025**, *30*, 04025039. [CrossRef]
131. Pereira, M.; Oliveira, G.; Costa, C.; Kelman, J. Stochastic streamflow models for hydroelectric systems. *Water Resour. Res.* **1984**, *20*, 379–390. [CrossRef]
132. Kim, T.-W.; Valdes, J.B. Synthetic generation of hydrologic time series based on nonparametric random generation. *J. Hydrol. Eng.* **2005**, *10*, 395–404. [CrossRef]
133. Vogel, R.M.; Shallcross, A.L. The moving blocks bootstrap versus parametric time series models. *Water Resour. Res.* **1996**, *32*, 1875–1882. [CrossRef]
134. Nowak, K.; Prairie, J.; Rajagopalan, B.; Lall, U. A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow. *Water Resour. Res.* **2010**, *46*. [CrossRef]
135. Tarboton, D.G.; Sharma, A.; Lall, U. Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resour. Res.* **1998**, *34*, 107–119. [CrossRef]
136. Wang, W.; Ding, J. A multivariate non-parametric model for synthetic generation of daily streamflow. *Hydrol. Process. Int. J.* **2007**, *21*, 1764–1771. [CrossRef]
137. Srinivas, V.; Srinivasan, K. Hybrid matched-block bootstrap for stochastic simulation of multiseason streamflows. *J. Hydrol.* **2006**, *329*, 1–15. [CrossRef]
138. Borgomeo, E.; Farmer, C.L.; Hall, J.W. Numerical rivers: A synthetic streamflow generator for water resources vulnerability assessments. *Water Resour. Res.* **2015**, *51*, 5382–5405. [CrossRef]
139. Mathai, J.; Mujumdar, P. Multisite daily streamflow simulation with time irreversibility. *Water Resour. Res.* **2019**, *55*, 9334–9350. [CrossRef]
140. Unny, T.E.; Panu, U.S.; Macinnes, C.D.; Wong, A.K. Pattern analysis and synthesis of time-dependent hydrologic data. *Adv. Hydrosci.* **1981**, *12*, 195–295.
141. Mejia, J.M.; Rousselle, J. Disaggregation models in hydrology revisited. *Water Resour. Res.* **1976**, *12*, 185–186. [CrossRef]
142. Valencia, R.D.; Schakke, J.C., Jr. Disaggregation processes in stochastic hydrology. *Water Resour. Res.* **1973**, *9*, 580–585. [CrossRef]
143. Lettenmaier, D. Some thoughts about the state-of-the-art in stochastic hydrology and streamflow forecasting. In *Stochastic Hydrology and Its Use in Water Resources Systems Simulation and Optimization*; Springer: Berlin/Heidelberg, Germany, 1993; pp. 209–215.
144. Grygier, J.C.; Stedinger, J.R. Condensed disaggregation procedures and conservation corrections for stochastic hydrology. *Water Resour. Res.* **1988**, *24*, 1574–1584. [CrossRef]
145. Savic, D.A.; Burn, D.H.; Zrinji, Z. A Comparison of Streamflow Generation Models for Reservoir Capacity-Yield Analysis 1. *JAWRA J. Am. Water Resour. Assoc.* **1989**, *25*, 977–983. [CrossRef]
146. Kuo, J.-T.; Sun, Y.-H. An ARMA-type section model for average ten-day streamflow synthesis. *Water Resour. Manag.* **1996**, *10*, 333–354. [CrossRef]
147. Ochoa-Rivera, J.; Andreu, J.; García-Bartual, R. Influence of inflows modelling on management simulation of the water resources system. *J. Water Resour. Plan. Manag.* **2007**, *133*, 106–116. [CrossRef]
148. Dong, G.; Liu, H. *Feature Engineering for Machine Learning and Data Analytics*; CRC Press: Boca Raton, FL, USA, 2018.

149. Hashimoto, T.; Stedinger, J.R.; Loucks, D.P. Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water Resour. Res.* **1982**, *18*, 14–20. [CrossRef]
150. Hinkley, D. On quick choice of power transformation. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1977**, *26*, 67–69. [CrossRef]
151. Vélez, J.I.; Correa, J.C.; Marmolejo-Ramos, F. A new approach to the Box–Cox transformation. *Front. Appl. Math. Stat.* **2015**, *1*, 12. [CrossRef]
152. Denny, J.; Kisiel, C.; Yakowitz, S. Procedures for determining the order of dependence in streamflow records. *Water Resour. Res.* **1974**, *10*, 947–954. [CrossRef]
153. Devries, R.N. Streamflow Simulation Techniques. In *Proceedings of the Oklahoma Academy of Science*; Oklahoma State University: Stillwater, OK, USA, 1970.
154. Rodriguez-Iturbe, I.; Dawdy, D.R.; Garcia, L.E. Adequacy of Markovian models with cyclic components for stochastic streamflow simulation. *Water Resour. Res.* **1971**, *7*, 1127–1143. [CrossRef]
155. Yakowitz, S.J. A nonparametric Markov model for daily river flow. *Water Resour. Res.* **1979**, *15*, 1035–1043. [CrossRef]
156. Siegerstetter, L.A.; Wahliβ, W. Generation of Weekly Streamflow Data for the River Danube-River Main-System Experiences With an Autoregressive Multivariate Multilag Model. *Dev. Water Sci.* **1982**, *17*, 280–291.
157. Green, N. A synthetic model for daily streamflow. *J. Hydrol.* **1973**, *20*, 351–364. [CrossRef]
158. Guimarães, R.C.; Santos, E.G. Principles of stochastic generation of hydrologic time series for reservoir planning and design: Case study. *J. Hydrol. Eng.* **2011**, *16*, 891–898. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.