# Development of Stochastic Models for Streamflow Synthesis Based on Colored-Noise–Dominated Hydrologic Systems: Textural Pattern Modelling for Black- and Pink-Noise Systems and Brownian Motion Modelling for Brown-Noise Systems

A dissertation submitted to the Civil Engineering Department, Lakehead University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Ph.D.) in Civil Engineering
(ENGI-9900)

By: Shirin Studnicka
Ph.D. Candidate, Civil Engineering Department, Lakehead University
Spiran@lakeheadu.ca

Supervisor: Dr. Umed Panu, P.Eng.
Professor, Civil Engineering Department, Lakehead University
uspanu@lakeheadu.ca

Co-Supervisor: Dr. Deli Li
Professor, Mathematical Sciences Department, Lakehead University
dli@lakeheadu.ca



April 2026

Dedicated to the souls of all brave people who sacrificed their lives for the freedom of Iran while I was writing this dissertation. Forever in my heart.

# Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Panu, for his guidance, support, and valuable feedback, as well as for the many weekends he devoted to our research meetings and for always being available to support me. I also extend my sincere thanks to my co-supervisor, Dr. Li, for his support and helpful insights throughout this research.

I would like to thank my committee members, Dr. Cui, Dr. Bajwa, Dr. Akilan, and Dr. Singh, for their valuable suggestions and guidance. I also thank the faculty and staff of the Department of Civil Engineering at Lakehead University for their support during my studies.

I am grateful to all researchers whose work is cited in my thesis; their research provided valuable guidance and inspiration for developing new ideas, and this work would not have been possible without their contributions.

I would like to thank my mom and dad, who never stopped supporting me despite the borders, oceans, mountains, and continents between us.

Finally, as I was deeply immersed in this research, my husband, Peter, and I often had little time for ourselves. I am deeply grateful to him for his patience and understanding throughout this work.

# Abstract

Civil engineering structures are often designed based on flood design criteria; however, there is no guarantee that these design floods represent worst-case scenarios. It is because streamflow has been recorded for only a few hundred years, while rivers have been running for over the past thousands of years. Consequently, the available records may not include the most extreme flood events. This limitation creates a need for synthesizing additional possible streamflow scenarios. Over the decades, hydrologists have developed various methods to synthesize streamflow, ranging from early stochastic models in the 1960s to AI-based approaches of today. The evolution of streamflow synthesis methods relies on two equally important considerations: a deeper understanding of streamflow characteristics and advances in computational techniques. In some cases, however, the first consideration is overlooked under the assumption that more complex models can better synthesize possible scenarios. Therefore, this research focuses equally on these two crucial considerations: (1) improving the understanding of monthly streamflow characteristics, and (2) developing models according to the identified characteristics. To improve the understanding of streamflow characteristics, this research quantifies scaling behaviour as a representation of the memory of hydrological systems using an approach based on identifying jumps in the power spectral density of the system being analyzed. Using this framework, 143 hydrometric stations in Ontario were analyzed to identify pink-, brown-, and black-noise–dominated monthly streamflow behaviour. Of these, two stations were characterized by the dominance of brown noise, 39 by pink noise, and the remaining 102 by black noise. Based on the identified dominant noise, the development of modelling approaches is then initiated in this research. For hydrometric stations dominated by the brown noise, an Extended Geometric Brownian Motion (EGBM) model has been developed. Compared with a multiplicative ARIMA model, the EGBM better reproduces the statistical properties of historical streamflow and yields residuals with statistically insignificant lag-1 autocorrelation. For stations exhibiting pink- or black-noise dominance, a textural pattern recognition system (TPRS) has been developed. The comparison of TPRS and ARIMA/SARIMA models shows that for pink-noise–dominated watersheds, although TPRS performs slightly better, the improvement is not sufficient to justify the added model complexity in cases where high accuracy gains are not required. In contrast, black-noise–dominated watersheds exhibit stronger temporal dependence, which is more appropriately captured using the proposed TPRS framework. Further comparison of the proposed TPRS with traditional pattern recognition methods demonstrates its effectiveness in reproducing autocorrelation structures and higher-order statistical features. The results show that the TPRS improves the representation of the autocorrelation function (ACF) up to 100 lags, compared with 24 lags for traditional pattern recognition methods. Moreover, the Hurst coefficient analysis confirms that the TPRS model provides a slightly enhanced representation of statistical characteristics exhibited by the historical time series. These findings support more informed model selection by reducing the mismatch between model structure and underlying hydrological behaviour, ultimately improving the reliability of synthesized streamflow scenarios for flood frequency analysis and flood risk assessment

# Table of Contents

# List of Figures

# List of Tables

# Notations

$q_{i,j}^{k}$: historical streamflow time series, segmented into k seasons and arranged in a matrix format (i,j), where i denotes the month and j denotes the year.

$x_{i,j}^{k}$ : normalized historical segmented streamflow between [0,1].

$G^{k}$ : Seasonal textural image.

$g_{i,j}^{k}$: Grayscale intensity of the textural pixel corresponding to the k$^{th}$ seasonal textural image.

$SACF_{i,j}^{k}$Simultaneous Autocorrelation function of a textural pixel, i.e., the extracted textural features in the time domain.

$P_{i,j}^{k}$: Power spectrum of a textural pixel, which is the transformed textural features in the frequency domain.

$PH_{i,j}^{k}$: high-power components of $P_{i,j}^{k}$.

$PL_{i,j}^{k}$: Low-power components of $P_{i,j}^{k}$.

$R_{PH}^{k}$: Reference matrix of $PH_{i,j}^{k}$.

$R_{PL}^{k}$: Reference matrix of $PL_{i,j}^{k}$.

All the above elements have a synthesized version, denoted by a prime. For example,

$g_{i,j}^{k}{}'$ is the synthesized grayscale intensity of the textural pixel corresponding to the k$^{th}$ synthesized seasonal textural image ($G^{k\prime}$).

Textural Feature Extraction and Transformation Model (Summary in brief):

$q_{i,j}^{k} \rightarrow x_{i,j}^{k} \rightarrow g_{i,j}^{k} \rightarrow SACF_{i,j}^{k} \rightarrow F^{k}(u,v) \rightarrow P_{i,j}^{k} \rightarrow (PH_{i,j}^{k}$ **&** $PL_{i,j}^{k})$.

Image Synthesis Model (Summary in brief):

$P_{i,j}^{k}{}' \rightarrow F^{k}(u,v)' \rightarrow SACF_{i,j}^{k}{}' \rightarrow g_{i,j}^{k}{}' \rightarrow x_{i,j}^{k}{}' \rightarrow q_{i,j}^{k}{}'$

# 1. Introduction to Streamflow Synthesis

This chapter provides a brief introduction to the nature of streamflow time series. It begins by describing streamflow as a time series and outlining the key considerations in its analysis. It then discusses the importance and definition of streamflow synthesis, followed by an overview of the applications of synthetic streamflow in operational hydrology.

## 1.1 Nature of Streamflow Time Series – A Perspective

Analysis of streamflow time series at various time scales reveals a range of short- and long-term persistence, regular and irregular periodicities, linear and non-linear dynamics, and chaotic behaviour. The governing forces of river systems are influenced by solar energy, which varies annually due to sunspots and other complex, not fully understood interactions. These variations modulate the hydrological cycle and, in turn, affect river flow worldwide.

Capturing all types of persistence and periodicity in streamflow synthesis is challenging. Therefore, methods for generating synthetic streamflow at appropriate time scales must strive to incorporate as many of these dynamics as possible. The complexities of the Earth-Solar system highlight the need for comprehensive synthesis approaches that address diverse persistence patterns and periodicities (Yevjevich, 1972; Panu et al., 1978; Salas et al., 1980; Stedinger & Taylor, 1982; Sivakumar, 2016).

As streamflow is influenced by complex and interconnected hydrological processes that are not fully understood, some characteristics of streamflow are also not well understood. Furthermore, streamflow in many streams is modified due to human interventions such as the construction of reservoirs, the discharge of pumped groundwater into streams, and the diversion of water for various uses. Such alterations disrupt the natural flow regime and pose challenges in defining flow characteristics at specific points along the stream (Benson & Carter, 1973). Despite these complexities and human interventions, some of the inherent and salient characteristics of streamflow have been identified and studied extensively by Riggs (1985) and Dettinger and Diaz (2000). Key characteristics of streamflow time series are:

- **Randomness:** Streamflow results from a random process whose outcomes vary over time and are governed by probabilistic structures and uncertainties (Brunner et al., 2019).

- **Persistence:** Streamflow time series tend to remain consistent or exhibit similar patterns over consecutive time intervals (Studnicka & Panu, 2025a, 2025b). As noted by Ghimire and Krajewski (2020) persistence reflects the idea that *"tomorrow will be similar to today,"* exploiting the memory of the hydrological system.
- **Correlation:** Describes the relationship between streamflow at a given time and previous observations at earlier time steps (autocorrelation), or with other hydrological variables such as rainfall, snowfall, and temperature (cross-correlation) (Yue & Wang, 2002).
- **Trend:** A deterministic component in streamflow time series that exhibits a slow and gradual change over time, often associated with climate change, land-use alterations, or long-term watershed modifications (Yue & Wang, 2002; Dastour & Hassan, 2023).
- **Periodicity, Cyclicity, and Seasonality:** Deterministic components indicating repetitive or cyclic behaviour in streamflow, often linked to seasonal climatic drivers such as snowmelt cycles, monsoon patterns, or annual precipitation variability (Studnicka & Panu, 2025b, 2025c).
- **Nonlinearity:** Changes in factors influencing streamflow do not produce proportional changes in discharge. This reflects the nonlinear transformation of rainfall (input) into streamflow (output) due to interacting processes such as soil moisture dynamics, evapotranspiration, and basin storage (Sivakumar, 2016).
- **Sensitivity to Initial Conditions:** Small differences in initial hydrological states (e.g., soil moisture or antecedent flow) can lead to significantly different streamflow outcomes. This concept, often referred to as the *"butterfly effect,"* has been highlighted in hydrology by Sivakumar (2016), arising due to the nonlinear and interconnected nature of the system.
- **Stochasticity and Chaoticity:** Stochasticity arises from random and complex natural processes influencing streamflow. Chaoticity is associated with nonlinear interdependence, hidden determinism, and sensitivity to initial conditions, making streamflow behaviour highly complex and difficult to predict (Elshorbagy et al., 2001).
- **Noise Characteristics:** Streamflow time series often contain different types of noise that influence their variability and predictability. In this context, white noise and coloured noise are not treated as separate defining categories; rather, white noise represents the absence of temporal correlation, while coloured noise refers to a broader class of noise processes characterized by frequency-dependent power spectra (Studnicka & Panu, 2025d).

Therefore, given the complex, nonlinear, and stochastic nature of streamflow, its key characteristics must first be identified and quantified before reliable streamflow synthesis can be performed.

## 1.2 Characterization of Streamflow Prior to Synthesis

A fundamental step in realistic streamflow synthesis is the proper characterization of the temporal structure and memory behaviour inherent in streamflow time series.

### 1.2.1 Persistence

Persistence is a fundamental characteristic of streamflow time series and reflects the tendency of hydrologic systems to retain information from past states. In simple terms, streamflow at a given time is influenced by previous conditions, indicating the presence of memory within the watershed. This dependence may arise from storage mechanisms such as soil moisture, groundwater, snowpack, and channel routing processes. Persistence plays a central role in shaping streamflow variability and governs how quickly a system responds to changes in climatic inputs. Long-term persistence in streamflow has been widely recognized in hydrology and is commonly associated with the Hurst phenomenon, which reflects the tendency of hydrologic processes to exhibit sustained departures from the mean value over extended periods. The presence of Hurst behaviour highlights that streamflow variability is not purely random but structured across multiple time scales. Ignoring persistence can lead to unrealistic representations of streamflow behaviour, particularly when reproducing variability and extreme events in synthetic records.

### 1.2.2 Memory and Scaling Behaviour

Memory in streamflow time series refers to the way hydrologic processes retain influence across multiple temporal scales. Rather than behaving as a sequence of independent events, streamflow often exhibits structured dependence that extends beyond short time horizons, reflecting cumulative effects of storage, climate variability, and basin processes. This behaviour is closely related to scaling, which describes how statistical properties of streamflow remain consistent or systematically change when observed at different temporal resolutions. Importantly, such scaling characteristics are intrinsic to the system and are not tied to a specific point in time; shifting a time series does not alter its underlying correlation structure or memory. The presence of scaling

behaviour suggests that streamflow dynamics are governed by multi-scale interactions and long-range dependence, commonly associated in hydrology with the Hurst phenomenon. Recognizing memory and scaling behaviour is therefore essential prior to streamflow synthesis, as failure to preserve these properties may result in synthetic records that misrepresent long-term variability and persistence observed in natural streamflow.

### 1.2.3 Periodicity

Periodicity refers to the presence of recurring patterns in streamflow over regular intervals and is commonly associated with climatic seasonality. Many river systems exhibit characteristic annual cycles driven by precipitation regimes, temperature variations, and snowmelt processes. These repeating patterns introduce a structured form of temporal dependence that influences both the magnitude and timing of streamflow. Periodicity provides a natural framework for interpreting streamflow behaviour across time scales and is closely linked to the concept of persistence within cycles. Proper recognition of periodic behaviour is essential for ensuring that the synthesized streamflow reflects the natural rhythm observed in historical records (Unny et al., 1981).

## 1.3 Streamflow Synthesis: Definition, Importance, and Role in Water Resources Management

Hydrologic time series may be broadly classified into two categories: historical time series, which are derived from direct measurements or observations at regular time intervals, and synthesized time series, which are artificially generated to support water resources planning, design, and operational decision-making. Streamflow synthesis is therefore a process of generating multiple realizations of streamflow that reproduce key characteristics observed in historical records, with the objective of representing the range of hydrologic variability relevant to operational hydrology.

Synthesized streamflow time series are designed to carry forward information embedded in historical observations while extending beyond the limitations imposed by finite record lengths. From a conceptual perspective, each synthesized realization can be viewed as a plausible sequence of streamflow events that could have occurred before systematic measurements began or that may occur in the future under similar governing processes. Since most instrumental streamflow records span only about a century, they are unlikely to capture the full range of hydrologic variability, particularly rare and extreme events such as severe floods or prolonged droughts. As a result,

historical observations represent only a limited sample of a much larger population of possible streamflow outcomes.

Streamflow synthesis provides a framework for exploring this broader population by generating an ensemble of streamflow realizations that collectively reflect the variability, dependence structure, and statistical properties inferred from observed data. These synthetic ensembles enable more robust estimation of hydrologic risks and system performance by accounting for events that may not be present in the historical record but remain physically plausible. The effectiveness of streamflow synthesis depends critically on the extent to which the synthesized series preserves the defining characteristics of the historical data, including variability, persistence, and temporal structure.

The selection of an appropriate synthesis model is therefore guided by the characteristics identified in the historical streamflow record. As understanding of streamflow behaviour has evolved, so too has the complexity of synthesis models. Early approaches relied primarily on basic statistical descriptors such as mean and variance, which were once considered sufficient representations of hydrologic behaviour. Subsequent advances highlighted the importance of temporal dependence, leading to the incorporation of short-term persistence through autocorrelation structures. The recognition of long-term persistence further expanded this perspective, revealing that streamflow variability may exhibit memory extending across long time scales.

These developments underscore the importance of streamflow synthesis in hydrologic analysis and water resources management. By enabling the generation of realistic streamflow scenarios that reflect both observed behaviour and unobserved but plausible extremes, streamflow synthesis plays a critical role in reservoir design, flood risk assessment, drought management, and the evaluation of water system reliability under uncertainty. As hydrologic understanding deepens, synthesis approaches continue to evolve to better capture the complex and multi-scale nature of streamflow processes.

### 1.3.1 Applications of Synthetic Streamflow in Operational Hydrology

One of the most important applications of synthetic streamflow in operational hydrology is reservoir storage analysis. Traditionally, reservoir capacity has been estimated using historical streamflow records, often through deterministic methods such as the Rippl mass curve approach.

While these methods have been widely used in both reservoir design and operation, their reliability is inherently limited by the length of available hydrologic records.

Reservoir design requires estimates of storage capacity that can reliably meet specified water demands under variable inflow conditions. However, historical streamflow records typically represent only a small sample of the full range of hydrologic variability. Extreme droughts or prolonged low-flow periods may not be adequately captured, leading to potential underestimation of required storage. As a result, deterministic approaches based solely on observed records have been criticized for their inability to quantify risk and uncertainty (Stedinger & Taylor, 1982; Vogel & Stedinger, 1988).

The development of stochastic hydrology and streamflow synthesis enabled a shift toward probabilistic reservoir analysis. By generating multiple synthetic streamflow realizations that preserve key statistical and dependence characteristics of observed flows, reservoir performance can be evaluated under a wider range of plausible hydrologic scenarios. This approach allows storage–yield–reliability relationships to be assessed probabilistically rather than deterministically, providing a more robust basis for decision-making in water resources planning and management (Stedinger & Taylor, 1982; Vogel & Stedinger, 1987).

Previous studies (Vogel & Stedinger, 1987; Studnicka & Panu, 2025a) have demonstrated that reservoir storage estimates derived from synthetic streamflow sequences can outperform traditional drought-of-record analyses, even when simplified stochastic models are used. Although uncertainty remains due to limited data and model assumptions, synthetic streamflow has become an essential tool for understanding reservoir behaviour under hydrologic uncertainty and for improving the reliability of water supply system design.

## 1.4 Critical Considerations in the Application of Synthetic Models for Forecasting

It is crucial to note that a synthetic model is not a forecasting model. In synthesis, the model generates data starting from an arbitrary initial condition. For instance, if the model produces 50 data points, the first 20 are typically discarded because, after some iterations, the influence of the starting point disappears (Salas et al., 1980; Kendall & Dracup, 1991; You et al., 2014). The goal

here is to produce statistically realistic sequences, and not to predict future values based on evolving conditions.

In contrast, forecasting requires careful consideration of changing external factors such as land use changes, climate change, and other relevant predictors. These factors must be quantified and explicitly fed into the model in order to make meaningful predictions about the future. Simply applying a synthetic model for forecasting without accounting for such changes is fundamentally flawed. In this case, the only variation comes from the starting point, which implicitly assumes that nothing in the system will change over time, a clearly unrealistic and misleading assumption. In other words, using a synthetic model for forecasting would only work under the unrealistic assumption that future conditions remain identical to those of the past, which is scientifically unjustifiable.

## 1.5 Objectives and Scope of the Dissertation

Despite decades of progress in streamflow synthesis, a key challenge remains in systematically linking streamflow characteristics to appropriate modelling approaches, particularly when selecting models that preserve the underlying statistical behaviour of hydrological systems. In addition, while modern AI and machine learning methods provide powerful tools for capturing complex temporal patterns, their reliance on large datasets and limited interpretability restricts their applicability in many hydrological contexts. These limitations motivate the need for approaches that explicitly incorporate hydrological understanding into the modelling process.

The primary objective of this dissertation is to analyze and characterize the key statistical and temporal properties of streamflow, including persistence, memory, periodicity, and noise behaviour. By understanding these fundamental characteristics, it becomes possible to identify the strengths and limitations of current modelling approaches and to quantify the aspects of streamflow that must be preserved in any synthetic series. This step provides a solid foundation for developing reliable and realistic streamflow synthesis techniques.

Building on this analysis, the dissertation addresses the growing reliance on Artificial Intelligence and Machine Learning (AI/ML) approaches in hydrology. While these methods offer powerful

tools for prediction and pattern recognition, their fully automated nature can make them passive, limiting control of the user over the generated outputs and the ability to incorporate domain knowledge. To overcome such considerations, a semi-automated framework is proposed that integrates AI/ML capabilities with user-guided control, allowing both short-term and long-term dependencies in streamflow to be captured accurately while maintaining transparency and interpretability. This approach enables the generation of synthetic streamflow series that are not only statistically consistent with historical observations but also flexible enough to explore alternative scenarios for water resources planning and management, especially in climate change scenarios.

# 2. Literature Review

This chapter first reviews the methods used to quantify the fundamental characteristics of streamflow, including persistence, memory, periodicity, and noise behaviour. It then discusses the evolution of streamflow synthesis methodologies, from traditional approaches focused on preserving basic correlation structures to modern techniques capable of capturing stochastic and multi-scale patterns. As our understanding of streamflow behaviour grows, there is a parallel expectation for synthesis methods to become increasingly complex, incorporating both statistical rigour and flexibility to generate realistic hydrologic scenarios.

## 2.1 Studies on Quantifying Streamflow Characteristics

The accurate characterization of streamflow is fundamental for understanding hydrologic processes and developing realistic synthesis models. Streamflow exhibits complex temporal behaviour, including persistence, scaling, memory, and noise patterns, which have been extensively studied in the hydrologic literature.

### 2.1.1 Scaling Behaviour and Long-Term Memory

Scaling behaviour in streamflow describes how statistical properties change across temporal scales, revealing multi-scale correlations and persistent memory in the system (Hurst, 1951; Fiering & Bund, 1971; Klemeš, 1974). The Hurst exponent (H) is widely used in hydrology to quantify long-term memory and persistence, with values between 0.5 and 1 indicating a persistent time series (Salas et al., 1980; Koutsoyiannis, 2002). Power Spectrum Analysis (PSA) is commonly employed to assess scaling behaviour, where the slope of the power spectrum in log-log space indicates the type of scaling and associated noise characteristics (Dooley & Van de Ven, 1997; Telesca et al., 2012). Studies have highlighted the variability of λ across frequencies, pointing to the existence of multiple scaling regimes within hydrological time series (Thompson & Katul, 2012; Kim et al., 2016).

### 2.1.2 Noise Characteristics in Hydrological Time Series

Hydrologic variables contain intrinsic noise, reflecting the complexity of natural systems (Sivakumar, 2016). Noise can be categorized by colour based on the spectrum exponent λ: white

noise ($\lambda = 0$), pink noise ($0 < \lambda < 2$), brown noise ($\lambda = 2$), and black noise ($\lambda > 2$) (Dooley & Van de Ven, 1997; Stoyanov et al., 2011). These different noise types relate to system memory: for example, hydrologic pink noise is persistent, while brown and black noise reflect long-term trends and low-frequency dominance (Kim et al., 2016). Understanding noise characteristics is critical, as they influence model selection and the ability to capture the temporal dependencies of streamflow accurately (Rodriguez-Iturbe & Rinaldo, 1997).

### 2.1.3 Memory and Persistence

Persistence reflects the tendency of streamflow to retain information from past states, influenced by mechanisms such as soil moisture, groundwater, snowpack, and channel routing processes. Short-term persistence is commonly measured through lag-one autocorrelation, whereas long-term persistence can be quantified via the Hurst exponent (Hurst, 1951; Fiering & Bund, 1971). Different studies have emphasized that both short- and long-term persistence must be considered when synthesizing streamflow, as ignoring either can lead to unrealistic representations of variability and extremes (Panu & Unny, 1980a; Salas et al., 1980).

### 2.1.4 Summary of Insights

Overall, literature demonstrates that quantifying streamflow characteristics requires an integrated assessment of scaling behaviour, noise structure, memory, and persistence. Various stochastic and semi-stochastic models have been developed to capture individual streamflow characteristics. Accordingly, the next section reviews the progression of streamflow data synthesis approaches from traditional stochastic models to more modern techniques.

## 2.2 Studies on Streamflow Synthesis

Nearly 150 years ago, Rippl (1883) investigated the problem of estimating the reservoir capacity required to satisfy specified water demands. This work helped stimulate the development and advancement of methods for generating synthetic streamflow (O'Connell, 1974). The Rippl method is based on mass curve analysis, which is used to determine the minimum reservoir storage needed, given a particular demand and the historical streamflow record.

The development of reservoir analysis and streamflow modelling has evolved significantly over time. Early deterministic approaches, such as the mass curve method introduced by Hazen and

later refined by Rippl, focused on critical-period analysis using observed records. Subsequent advancements incorporated probabilistic thinking, including Markovian models and autoregressive (ARMA) frameworks, which enabled synthetic streamflow generation and stochastic reservoir simulation. In more recent decades, data-driven approaches and pattern recognition techniques have further expanded the analytical toolbox, allowing for nonlinear modelling and complex dependence structures. This progression reflects the shift from purely deterministic design toward increasingly sophisticated representations of hydrologic variability and uncertainty (Studnicka & Panu, 2025a).

Stochastic streamflow models have been developed to estimate the reservoir capacity needed to consistently satisfy a specified water demand or release by deriving the probability distribution of required capacity under variable streamflow conditions (Vogel & Stedinger, 1988). However, streamflow synthesis did not originate during this later period of hydrological challenges related to reservoir engineering; instead, it emerged much earlier, between 1914 and 1962, a timeframe referred to here as the pre-era.

### 2.2.1 Early Developments in Streamflow Synthesis: PRE-ERA (Beginning-1960)

There is broad agreement among hydrologists involved in streamflow synthesis that Hazen (1914) was the first to generate synthetic streamflow records (Boughton & McKerchar, 1968; Fiering & Bund, 1971; Lawrance & Kottegoda, 1977; Phien & Ruksasilp, 1981; Wijayaratne & Chan, 1987). Hazen (1914) sought to extend historical data by producing a 300-year synthetic record based on 14 streams, in which each annual flow was normalized by dividing it by the long-term mean annual flow of the corresponding river. The resulting records were then arranged and combined according to increasing variability, as measured by their coefficient of variation. However, although this synthetic dataset was used to develop design charts for reservoir capacity estimation, it did not overcome the limitations of the mass curve method and did not account for the autocorrelation structure of streamflow data.

In the period following the work of Hazen, early streamflow synthesis methods included a card-based technique introduced by Sudler (1927), which once again addressed the storage problem. In this method, each card represented an annual runoff volume, and the cards were randomly shuffled. Cards were then drawn sequentially, and the corresponding values were recorded. Later, Barnes

(1954) used a table of random numbers drawn from a normal distribution to generate a longer annual streamflow record with similar mean and variability to the original data. Although this approach represented an improvement over the card method, it did not account for autocorrelation in the annual streamflow time series (Boughton & McKerchar, 1968). Early streamflow synthesis techniques primarily aimed to reproduce the statistical properties of historical records, as hydrologists at the time were mainly familiar with measures such as mean, variance, skewness, and kurtosis. As understanding of streamflow time series advanced, additional characteristics were identified and quantified, leading to the development of new synthesis methods that more effectively captured the complexities observed in historical data.

Two major findings, discussed below, prompted a substantial shift in streamflow synthesis methods from the pre-era to Era 1. The first was the identification of the autocorrelation function, which demonstrated strong temporal dependence in streamflow behaviour, particularly at short time lags. The second was the Hurst coefficient, which measures the long persistence, or memory, of a time series by indicating how long the influence of an event remains within the system after it occurs. As described in the following section, streamflow synthesis techniques developed during Era 1 primarily aimed to reproduce the above-noted two key characteristics.

### 2.2.2 ERA-1 (1960-2000): The Domination of AR-Family Models

During this period, hydrologists made extensive use of statistical methods and stochastic processes to analyze and characterize streamflow data, reflecting their inherently stochastic behaviour. Although statistical developments have been important in advancing streamflow synthesis, the body of literature in statistical hydrology remains relatively limited, with notable contributions by Anis and Lloyd, Moran, Gumbel, Bernier, Borgman, and Amorocho (Lawrance & Kottegoda, 1977). Among the most influential statistical studies are those of Hurst (1951) on the unique characteristics exhibited by streamflow records from the Nile River. The identification of this distinctive behaviour in streamflow data is now widely recognized in hydrology as the Hurst coefficient (also referred to as the Hurst effect or exponent).

The pioneering research investigations of Hurst in the 1950s, particularly his work on the Nile River, generated considerable interest in statistical techniques such as the rescaled range method. Hurst examined approximately 900 annual time series datasets representing a wide range of

environmental variables, including streamflow, precipitation, lake levels, tree rings, and atmospheric pressure. Hurst found that the adjusted range, denoted by R and calculated as the difference between the maximum excess ($P_n$) and the greatest deficiency ($Q_n$) relative to steady streamflow over n years, varied with record length according to a specific relationship, expressed as $R/S \sim n^h$, where R/S is the rescaled range, and S is the sample standard deviation of the series. By estimating the coefficient h using R and S, Hurst was able to gain insights into long-term storage requirements for the Nile River and other hydrological systems. These important findings laid the foundation for subsequent research on modelling hydrological processes.

During the 1960s, hydrologists came to recognize the importance of maintaining key statistical characteristics in synthesized time series. In particular, the autocorrelation function at lag 1 (ACF(1)) and lag 2 (ACF(2)), along with the Hurst coefficient, were identified as especially important for streamflow synthesis (Salas et al., 1980). As a result, these properties became central to streamflow generation efforts. However, the selection of which characteristics to emphasize often depended on the specific goals of individual studies. Owing to the simplicity of models developed during Era 1, it was generally not possible to preserve all three properties simultaneously.

Thomas and Fiering (1962) 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 (1962) presents a set of 12 regression equations as

$$Q_{i,j} = \bar{Q}_j + b_j\left(Q_{i,j-1} - \bar{Q}_{j-1}\right) + t_j S_j (1 - {r_j}^2)^{1/2}, \tag{2.1}$$

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 $\bar{Q}_j$ and $\bar{Q}_{j-1}$ are respectively for the $j$ and $(j-1)^{\text{th}}$ 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 enhance the applicability of the model to skewed hydrologic series, Thomas and Burden (1963) proposed a transformation that replaces the normal random variate $t_j$with a random variate that approximately follows a gamma distribution, thereby allowing the model to accommodate different distributional forms. Subsequently, in 2012, efforts were made to modify the Thomas–Fiering model by extracting the persistence characteristics of a river and examining the serial correlation among the random residuals of historical monthly mean records. The findings indicated that this revised version of the Thomas–Fiering model more effectively preserved the monthly and annual standard deviations of the historical data, as well as the associated frequency distribution (Arselan, 2012).

Undoubtedly, the model developed by Thomas and Fiering (1962) represented a significant advancement in addressing dependencies within streamflow data. However, one limitation of their approach was the treatment of flows for each month as separate populations. Additionally, since the Thomas–Fiering model is a first-order Markov process, the streamflow at any given time step depends solely on the preceding time step, thereby neglecting higher-order dependencies.

Initiated by the work of Thomas and Fiering (1962), the historical evolution of AR (Autoregressive) models commenced within hydrology during the 1960s. Box and Jenkins (1970) proved a significant catalyst for time series analysis through their classic textbook on Time Series Analysis, Forecasting and Control. Consequently, AR models came in several forms, with the most prevalent being Fiering and Jackson (1971) and later discussed by Salas et al. (1980), as shown in Eq. 2.2:

$$y_t = \mu + \sum_{j=1}^{p} \phi_j \left(y_{t-j} - \mu\right) + \varepsilon_t. \tag{2.2}$$

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 $\phi_j$ signifies the autoregressive coefficient, while $p$ denotes the order of the autoregressive process. The stochastic component $(\varepsilon_t)$ of the equation is the time-independent series, which is independent of $y_t$. The values of $\phi_j$ parameters are derived based on the assumption of normality.

Although autoregressive (AR) models can describe flow recessions based on past observations, they may not fully capture the high-flow fluctuations caused by rainfall and snowmelt. To better represent the variability in hydrologic time series, a moving average (MA) component is

incorporated into AR models, resulting in the ARMA model. A notable limitation of ARMA models, however, is their assumption of stationarity, which may not be valid for hydrologic data that exhibit seasonal patterns.

Hipel et al. (1977) applied the ARMA model for streamflow synthesis by first removing seasonal non-stationarity from the data and then fitting an ARMA model to the resulting series. Lawrance and Kottegoda (1977) pointed out theoretical issues with this de-seasonalization approach, particularly the assumption that the transformed series is stationary. This assumption implies that correlations between flows in different months—for example, January and August—remain constant, which contradicts both theoretical models and empirical observations (Moss and Bryson, 1974). Despite this limitation, AR-based models have been widely used for streamflow synthesis in many regions worldwide.

Hirsch (1979) applied AR and ARMA models, which are quasi-stationary—having statistical properties that vary seasonally but remain consistent across years—to synthesize monthly streamflow using three different approaches: log-normal, log transformation, and normalizing transformation. Analysis of streamflow data from six rivers in the Potomac River at Point of Rocks indicated that the AR (1) model combined with a normalizing transform outperformed the other methods tested.

In the 1980s, the adoption of AR-type models reflected a compromise between simplicity, statistical rigour, and practical applicability. Early in the decade, Stolte (1980) highlighted the advantages of simple AR models over ARMA models, citing their ease of use and better alignment with engineering judgment, while acknowledging their limitations in capturing complex streamflow behaviour. Similarly, Muzik (1980) critiqued the reliability of stochastic models such as AR(1) for reservoir design, pointing out the difficulties in model validation and parameter estimation under uncertainty, while also recognizing the potential of simulations to account for variability and risk.

As the decade progressed, research increasingly focused on more advanced methods that addressed the limitations of earlier models. Stedinger and Taylor (1982) showed that uncertainty in model parameters had a greater effect on reservoir reliability estimates than the specific choice of AR-family model. Lettenmaier and Burges (1977), using ARMA-based disaggregation techniques, highlighted challenges related to seasonal disaggregation and mass conservation. Subsequent

studies, including Stedinger et al. (1985), introduced Bayesian approaches to account for parameter uncertainty in disaggregated models, while Stedinger et al. (1985) demonstrated that univariate ARMA (1,1) models with disaggregation could perform comparably to more complex multisite models. At the same time, Kalman filtering techniques were successfully applied by Bergman and Delleur (1985) to improve AR model forecasting through adaptive estimation of parameters.

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. (1980) 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 (1986) allowed deeper insights into parameter sensitivity and model robustness. Bowles et al. (1987) proposed model selection strategies based on persistence measures, while Vogel and Stedinger (1988) 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 (1990) 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 increasingly focused on overcoming the known limitations of conventional AR and ARMA models, including their limited ability to represent seasonal variability, skewness, and spatial dependence, by developing and applying more advanced model variants. Mujumdar and Kumar (1990) demonstrated the importance of basin-specific model structures, showing that higher-order AR and ARMA models performed better than AR (1) in capturing monthly and ten-day streamflow variability. Although AR (1) exhibited lower mean square error, more complex models such as AR (4) and ARMA (3,1) were required to adequately preserve temporal dependencies, and model adequacy was statistically confirmed across all cases that were examined. At the same time, as noted earlier, Fernandez and Salas (1990) introduced the GAR(1) model to address the restrictive Gaussian assumption inherent in traditional ARMA models. When applied to river systems across multiple continents, the GAR(1) model successfully captured both dependence and skewness without the need for data transformations, representing a

practical improvement in reproducing higher-order moments and maintaining statistical fidelity in skewed annual streamflow series.

Progress during the decade continued through innovations aimed at addressing the seasonal and computational limitations of ARMA-family models. Santos and Salas (1992) introduced parsimonious stepwise disaggregation techniques that preserved autocorrelation and cross-correlation structures while substantially reducing computational requirements, making them particularly suitable for large datasets with bimonthly resolution. Subsequently, Rasmussen et al. (1996) advanced the application of Periodic Autoregressive Moving Average (PARMA) models by incorporating seasonality through periodic moment equations, while also highlighting difficulties in parameter estimation for high-frequency data. To further enhance model flexibility, Tasker and Dunne (1997) applied a nonparametric bootstrap approach to the residuals of the PARMA model, allowing the generation of synthetic flow series that preserved both spatial and temporal correlation structures without assuming normality. Collectively, these studies illustrate a shift during the 1990s away from purely linear stochastic models toward more robust and flexible methods capable of accommodating the complexities of real-world hydrologic data.

In the 21st century, the AR-family of models were still used in two primary ways: as benchmarks for comparison (Kottegoda et al., 2000; Ochoa-Rivera et al., 2002; Keskin et al., 2006; Prairie et al., 2006; Pender et al., 2016; Porto et al., 2021; Ma et al., 2024) and as components within hybrid modelling frameworks (Bayazit et al., 2001; Srinivas & Srinivasan, 2001; Pereira & Veiga, 2018; Treistman et al., 2020; Abdelaziz et al., 2023). During the 21st century, Artificial Intelligence (AI) and machine learning (ML) models became increasingly popular and eventually dominant in streamflow synthesis.

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

The emergence of AI models paved the way for the development of artificial neural networks (ANNs). The origins of ANNs can be traced back to 1943, when McCulloch and Pitts introduced a theoretical framework based on networks of binary processing elements, referred to as neurons. However, despite this early foundation, ANNs were not widely adopted by hydrologists until the early 1990s (Khare & Gajbhiye, 1943).

It is also noteworthy that, prior to this period, ANNs had already been applied for forecasting purposes as early as 1992. Raman and Sunilkumar (1995) observed that many conventional time series analysis techniques are based on assumptions of linear relationships among variables. In contrast, real-world temporal data often display complex and nonlinear patterns that are difficult to analyze and predict accurately, making linear models and their combinations insufficient for representing such behaviour. Consequently, the use of nonlinear models, such as neural networks, represents a logical advancement for addressing the complexities of temporally varying real-world systems. Accordingly, Raman and Sunilkumar (1995) developed twelve separate neural network models, one for each month of the year, to synthesize streamflow. Their comparative analysis with traditional ARMA models showed that the ANN-based approach produced comparable results.

Early work on semi-automated pattern recognition in the 19th century demonstrated pseudo forms of artificial intelligence (Panu & Unny, 1980a, 1980b). Later, in the early 21st century, Jardim et al. (2001) applied K-means clustering to monthly streamflow sequences generated from an autoregressive model in order to reduce computational demands in mid-term hydroelectric planning, representing what is arguably the first clear application of AI in this domain.

This period marked the onset of a shift from classical stochastic models toward hybrid and AI-enhanced frameworks. Early hybrid approaches combined traditional time series models, such as PAR(1), with bootstrapping techniques (Srinivas & Srinivasan, 2001) or integrated ARMA-generated flows within adaptive neuro-fuzzy inference systems (Keskin et al., 2006). Artificial neural networks (ANNs) were increasingly employed to capture nonlinear dependencies and to simulate complex flow behaviours, and in many cases outperformed conventional models such as ARMA and Thomas–Fiering (Awchi & Srivastava, 2004. ; Ahmed & Sarma, 2007). With continued growth in computational capacity, AI-based methods demonstrated an enhanced ability to reproduce hydrologic memory, extreme events, and spatiotemporal patterns more realistically than traditional approaches, reinforcing their increasing prominence.

Following confirmation that ANNs outperform traditional models in streamflow synthesis, research attention shifted toward addressing their limitations, including sensitivity to data size and the tendency to become trapped in local optima. The bootstrap technique was recommended to mitigate the former issue (Jia & Culver, 2006), while Support Vector Machines (SVMs) were proposed as a novel approach to address the latter (Deka, 2014).

In bootstrap techniques, a set of bootstrap samples, typically ranging from 50 to 200, is generated through sampling with replacement. Each bootstrap sample may contain repeated observations while excluding others. An ANN is then trained on each bootstrap sample, and the finally synthesized streamflow at each time step is obtained by averaging the outputs of all ANN models (Jia & Culver, 2006; Sudheer et al., 2008). As noted previously, in addition to requiring large datasets, ANNs are prone to overfitting and may become trapped in local optima. This behaviour is largely attributed to the underlying optimization approach used in ANNs, namely Empirical Risk Minimization (ERM), which seeks to minimize the error between the predicted and observed output vectors. In contrast, Support Vector Machines (SVMs) aim to minimize an upper bound on the generalization error (Deka, 2019). Beyond their global optimization capability, a key advantage of SVMs is the use of kernel functions to construct nonlinear variants, enabling more flexible modelling of complex relationships (Deka, 2019).

Another approach to mitigating overfitting involves the use of fuzzy logic. The fuzzy-rule structure and reduced parameter complexity often contribute to improved generalization performance. Keskin et al. (2006) demonstrated that employing an Adaptive Neuro-Fuzzy Inference System (ANFIS) for synthetic streamflow generation results in superior performance compared to stand-alone ARMA or ANN models. The fuzzy rule–based framework of ANFIS facilitates enhanced generalization, particularly when modelling nonlinear hydrological data.

The inherently nonlinear behaviour of streamflow has increasingly motivated researchers to adopt AI-based modelling approaches. Bourdin et al. (2012) identified a paradigm shift from traditional empirical models toward more flexible, ensemble-based and physically informed systems, particularly in the context of uncertainty and climate change. Hao and Singh (2013) addressed spatial and temporal dependence using a maximum entropy copula framework, successfully capturing multisite monthly streamflow structures. Likewise, Kirsch et al. (2013) employed the Modified Fractional Gaussian Noise (mFGN) model to reproduce seasonal autocorrelation patterns for planning under climate change scenarios, demonstrating how nonlinear variability in inflows can directly influence transfer operations.

Advanced models such as MRS (Multimodal Regression–Sampling) (Li & Singh, 2014) and M3EB (multi-site, multi-season Maximum Entropy Bootstrap) (Srivastav & Simonovic, 2014) integrate regression, k-NN resampling, entropy-based bootstrapping, and transformation

techniques to address both stochasticity and nonlinearity across multiple locations and seasons. In addition, You et al. (2014) and Marković et al. (2015) focused on improving ecological and hydrologic flow realism through disaggregation and multi-step nonparametric methods that preserved autocorrelation, extreme events, and flow transitions across multiple time scales.

Additional innovations include the Hydraulic Mixing-Cell (HMC)–based tracer model developed by Partington et al. (2013), which disaggregates flow contributions in storage-dominated basins, and the climate-informed stochastic model proposed by Stagge and Moglen (2013), which links GCM predictors to nonlinear streamflow states. In data-limited settings, Patskoski and Sankarasubramanian (2015) integrated paleo-hydrological reconstructions with Bayesian updating to improve estimation of extreme drought conditions.

More recently, Long Short-Term Memory (LSTM) networks have gained increasing attention due to their ability to represent both short- and long-term dependencies in time series data. This capability is particularly important in hydrology, where seasonal memory, delayed responses, and nonlinear interactions—such as those among rainfall, snowmelt, and soil moisture—play a critical role in governing streamflow dynamics. Molina et al. (2024) proposed an LSTM-based framework to retrospectively estimate daily streamflow in ungauged or data-scarce watershed segments. Their method represents watersheds as interconnected upstream–downstream pairs, enabling the model to utilize downstream flow information to enhance upstream streamflow estimates.

As a complementary alternative to AI/ML approaches, transformation-based methods that directly manipulate the temporal structure of streamflow series—particularly wavelet and frequency-domain techniques—have also been widely used for synthetic streamflow generation. Wang et al. (2011) proposed a nonparametric wavelet-based method that decomposes and reconstructs daily flows through annual sampling of wavelet coefficients, thereby preserving key statistical properties without relying on linearity assumptions. Similarly, Niu and Sivakumar (2013) applied Morlet wavelet decomposition to separate high- and low-flow components, demonstrating strong performance at shorter time scales. In the frequency domain, Brunner et al. (2019) utilized phase randomization of de-seasonalized streamflow series in combination with a kappa distribution to simulate realistic daily flows. This approach was later extended by Brunner and Gilleland (2020) through the Phase Randomization Simulation using wavelets (PRSim-wave) model, which

successfully captured spatiotemporal dependence and extremes across hundreds of U.S. catchments, making it suitable for regional-scale water resources planning.

These transformation-based approaches are increasingly favoured for their ability to represent complex hydrological behaviours, including non-stationarity and long-memory processes, often without the high levels of parameterization or large training data requirements associated with AI-based models. Abdelaziz et al. (2023) compared wavelet- and Discrete Fourier Transform (DFT)–based techniques, showing that both were effective in capturing short- and long-term dependencies, with DFT providing slightly superior performance. Taken together, these findings indicate a growing shift toward hybrid or transformation-based modelling strategies that can complement, or in some cases outperform, machine learning approaches such as LSTM or ANFIS in specific applications, particularly when interpretability, spectral accuracy, or data availability are key considerations.

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 such important considerations, the hydrologic community has begun to pursue two complementary strategies. First, hydrologists continue to use advanced AI/ML models, but with interpretability frameworks integrated, ensuring that the underlying relationships remain transparent and physically meaningful. Second, hydrologists develop semi-automated models that balance computational power with domain expertise, allowing researchers to remain actively involved in the modelling process.

### 2.2.4 Interpretability of Hydrological AI/ML Models

The growing use of AI/ML in hydrology has introduced powerful tools for nonlinear modelling, pattern recognition, and synthetic streamflow generation. Machine learning approaches have shown advantages over traditional physically based and conceptual models in terms of flexibility, predictive performance, and generalization. However, despite these strengths, ML models also face important challenges, particularly with respect to interpretability (Yaseen, 2023). The reliability and transparency of these models depend strongly on four closely related factors: validation protocols, treatment of distribution shifts, uncertainty quantification, and generalization capability.

#### 2.2.4.1 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 (Raman & Sunilkumar, 1995; Keskin et al., 2006), 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 (Jia & Culver, 2006). In hybrid and stochastic models, such as those using the moving block bootstrap (Srinivas & Srinivasan, 2001; Sudheer et al., 2008), 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. (2024)), 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.

#### 2.2.4.2 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 (Raman & Sunilkumar, 1995; Jardim et al., 2001) did not explicitly account for non-stationarity. Later developments, such as the mFGN approach (Kirsch et al., 2013), 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.

#### 2.2.4.3 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* (Jia & Culver, 2006; Sudheer et al., 2008) 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 (Fan et al., 2023; Mehdiyev et al., 2025). Fan et al. (2023) 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.

#### 2.2.4.4 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 (Raman & Sunilkumar, 1995). Hybrid approaches such as ANFIS + AR models (Keskin et al., 2006) and ANN + Moving Block Bootstrap (MBB) frameworks (Sudheer et al., 2008) showed improved adaptability by combining deterministic learning with stochastic variability. Recent developments in deep learning architectures, including LSTM networks (Molina et al., 2024), 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 recap, 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.

### 2.2.5 From Black-Box AI to Interpretable Streamflow Synthesis

In deep learning models, temporal dependence is learned implicitly through internal weights updating 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 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 (Girihagama et al., 2022).

Two key limitations that arise from the LSTM approach are as follows.

- **Lack of explicit features**: While omitting feature engineering simplifies the modelling pipeline, it also means that the model offers no internal explanation or analysis of the features upon which it explicitly relies. 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 effectively be 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 and logic. Appendix 2A provides a summary of AI/ML approaches potentially viable for streamflow synthesis.

Pattern recognition techniques offer an alternative that can preserve interpretability. Due to their design, these models explicitly encode and reproduce identifiable dominant 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. (1978) extracted features by computing the first-order differences between adjacent observations, effectively representing short-term dynamics. Later, Li et al. (2021a) 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.

It can therefore be said that the semi-automated approach balances complexity and control and also allows for the capture of long-term interactions (or dynamism) among patterns representing short-term dynamics. 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. In a nutshell, one may envision developing a hybrid approach where LSTM is utilized to generalize and optimize a more robust feature set within each type of seasonal segments under the strict supervision and control of a professional in the specific field of study.

## 2.3 Spatial Considerations in Streamflow Synthesis

Spatial considerations in streamflow synthesis refer to the ways in which streamflow patterns vary across different locations within a river network or catchment. One common form of these considerations in the hydrological literature is through spatial disaggregation models, in which information from an upstream station is used to generate finer-resolution streamflow data at downstream locations. Unlike purely temporal analysis, which focuses on how streamflow changes over time at a single location, spatial analysis accounts for the correlation of flow across multiple sites.

During the 1990s, hydrologists began addressing the known limitations of traditional AR and ARMA models, such as their poor representation of spatial dependence. Tasker and Dunne (1997) 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. Later, Woods and Sivapalan (1999) presented an analytical framework to quantify the relative importance of spatial and temporal variability in rainfall,

landscape properties, and hydrological processes during storm events. By integrating runoff generation and routing mechanisms, the method identifies dominant controls on catchment response and variability across basins. The framework provides a systematic way to assess how different sources of spatial heterogeneity influence hydrograph behaviour.

As computational capacity grew, AI-based approaches demonstrated the ability to synthesize hydrologic memory, extreme events, and spatial–temporal patterns more realistically than conventional models, establishing their growing dominance. Further, Hao and Singh (2013) addressed spatial and temporal dependence through a maximum entropy copula approach, accurately capturing multisite monthly streamflow structures. Spatial considerations are fundamental to realistic streamflow synthesis, as they govern the interactions between hydrologic processes across locations. It should be noted that the primary focus of the present research is on temporal characteristics rather than spatial variability.

## 2.4 So, Where is the Research Gap?

Despite nearly a century of research on streamflow synthesis, two critical gaps remain. First, while individual streamflow characteristics such as persistence, memory, periodicity, and noise colour are well documented, there is no systematic framework for selecting or developing synthesis models based on these quantified characteristics.

Second, modern AI and deep learning models, although capable of capturing complex temporal patterns, are often opaque and require large volumes of quality historical data to generate realistic synthetic series. In many hydrological contexts, such extensive quality datasets are not available and even non-existent, and the lack of interpretability seriously hinders scientific insight and control over the synthesis process.

This research endeavours to address such gaps by first analyzing and quantifying the fundamental characteristics of streamflow and then developing a semi-automated, interpretable modelling approach that explicitly accounts for the dominant noise type.

## Supplementary Material for Chapter 2:

## Appendix 2A: Summary of AI/ML Approaches in Streamflow Synthesis

This appendix provides a concise summary of key artificial intelligence and machine learning (AI/ML) approaches that have been applied to streamflow synthesis.

*Table 2A.1. Summary of representative AI/ML-based approaches for synthetic streamflow generation.*

| **Author** | **Model/Method** | **Uncertainty quantification** | **Validation Protocol** | **Handling Distribution Shifts** | **Generalization Ability** |
|---|---|---|---|---|---|
| Raman, H., & Sunilkumar, N. (1995). | Artificial Neural Network (Feedforward, Backpropagation) for multivariate inflow synthesis. | Not explicitly quantified; model performance assessed using MSE and comparison of statistical moments (mean, variance, skewness), | Split-sample validation: 10 years for training, 2 years for cross-validation, and 2 years for testing. Validation is used for early stopping and model selection to prevent overfitting. | Not addressed; assumes stationarity and consistent data distribution. | Evaluated via train–validation–test split to prevent overfitting; limited cross-site generalization. |
| Jardim, D. L. D. D., Maceira, M. E. P., & Falcao, D. M. (2001). | Periodic Autoregressive (PAR) stochastic model + K-means clustering for inflow sequence generation. | Residuals modelled with a 3-parameter Lognormal distribution; stochastic sampling captures variance, skewness, and spatial correlation. | Validation through preservation of historical marginal distributions and frequency analysis between simulated and observed streamflows (Furnas and Sobradinho reservoirs). | Limited — assumes stationarity of seasonal and spatial correlations; no explicit adaptation to non-stationary or shifting climate patterns. | Moderate — validated within-sample for two major reservoirs; no cross-basin or temporal generalization tested. |
| Srinivas, V. V., & Srinivasan, K. J. W. R. R. (2001). | **Hybrid Model (HM):** Combines a periodic autoregressive model (PAR (1)) for partial pre-whitening with moving block bootstrap (MBB) for resampling residuals, | Residual uncertainty is represented through nonparametric resampling (MBB) of block-structured residuals, capturing both linear and nonlinear dependence and | **Split-sample validation** on Weber River (Utah): model calibrated on first 55 years, validated on last 28 years. Evaluated for the ability to reproduce summary statistics, serial | Assumes stationarity of historical dependence and variability; limited in capturing nonstationary behaviour or climate-induced shifts. | Demonstrated effectiveness in replication of statistical structure and reservoir storage on an independent validation set (Weber River) and reasonable transferability to |

| | | | | | |
|---|---|---|---|---|---|
| | followed by post-blackening. | preserving skewness and marginal distributions. | correlations, and reservoir storage statistics. | | the San Juan River; no explicit multisite generalization. |
| Keskin, M. E., Taylan, D., & Terzi, O. (2006). | Adaptive Neuro-Fuzzy Inference System (ANFIS) coupled with AR(2) stochastic model. | Implicitly addressed through ensemble synthetic data generation; uncertainty captured indirectly via variability in AR(2)-generated flow series used for ANFIS training. | Split-sample validation using 36 years of monthly streamflow (1961–1996); first 34 years for training (1961–1994), last 2 years for testing (1994–1996); performance metrics: $R^2$ and MSE. | AR(2)-based data augmentation extends the input domain to unobserved flow conditions, partially mitigating bias from limited historical data. | High — demonstrated improved prediction of both low and high flows when trained on synthetic + observed data; strong adaptability to unseen scenarios due to hybrid learning. |
| Sarma, K., Ahmed, B. (2007). | Feed-forward Artificial Neural Network (ANN) for synthetic streamflow generation. | Incorporated indirectly through the addition of a random component based on the standard deviation of historical streamflow (ξt σt) to simulate natural variability. | Model trained on 40 years of data (1957–1996); performance evaluated by comparing synthetic and observed series using five statistics: monthly mean, monthly SD, overall mean, overall SD, and skew coefficient. | Partially addressed through stochastic noise addition, which allows the ANN to generate new flow sequences beyond the historical record. | Demonstrated strong generalization by accurately reproducing observed streamflow statistics and generating realistic 100-year synthetic series. |
| Jia, Y., & Culver, T. B. (2006). | Bootstrapped Artificial Neural Network (BANN). | Bootstrap method used to estimate generalization error and construct confidence intervals for each synthesized series. | Cross-validation using out-of-bag (left-out) samples from bootstrap; performance evaluated over observed flow periods and the full 4-year period. | Limited; BANN trained on small datasets struggles to extrapolate beyond observed hydrologic conditions; temporal variability in predictors may reduce accuracy. | High within the period covered by data; lower beyond the observed data period; performance improves with longer data spanning multiple years; can approximate nonlinear relationships. |
| Sudheer, K. P., | ANN-based Hybrid Model | Non-parametric resampling of | Data split: 75% training, 25% | Moderate: the model | High within observed range: |

| | | | | | |
|---|---|---|---|---|---|
| Srinivasan, K., Neelakantan, T. R., & Srinivas, V. V. (2008). | (ANNHM: ANN + Moving Block Bootstrap). | residuals using MBB to generate innovations; multiple replicates (1000) used to quantify variability in predictions (e.g., storage, drought characteristics). | validation; performance evaluated using RRMSE, preservation of summary statistics, marginal distribution, autocorrelation, drought/run characteristics, and storage predictions. | captures weak dependence structure and distributional information in residuals via MBB, but limited smoothing and low extrapolation capability; block size tuning helps address some variability. | RBF ANN captures non-linear trends; MBB adds stochasticity; similar RRMSE in training and validation indicates good generalization; less effective for extrapolation beyond historical flows. |
| Bourdin, D. R., Fleming, S. W., & Stull, R. B. (2012). | AR, ARMA, ARMAX, ARX, linear/non-linear regression, ANNs, SVM. | Error in input data, parameter uncertainty, model structure; back-propagation and ensemble methods for probabilistic outputs. | Calibration with observed streamflow; splitting data for training/testing; comparing pre- and post-urbanization hydro-meteorological data. | Limited; relies on the choice of the modeller for predictors and input variables; can include scenario-based forcing (urbanization, climate). | Moderate; black-box models can capture physical relevance but may fail outside the training range. |
| Hao, Z., & Singh, V. P. (2013). | Maximum Entropy Copula for Multisite Monthly Streamflow Simulation. | Uncertainty arises from input data variability (historical streamflow), approximations in marginal distributions, and conditional copula sampling; 100 sequences were generated to capture variability. | Comparison of simulated versus observed streamflow using scatterplots, Spearman rank correlations, and boxplots; evaluation of temporal and spatial dependence preservation. | Limited explicit treatment: can handle shifts in monthly flow patterns indirectly through conditional distributions, but large changes in climate/land use are not directly modelled. | Moderate; preserves temporal and spatial rank correlations; methodology can be extended to higher dimensions and similar problems (rainfall simulation, geostatistical interpolation), but downstream flow summation constraints are not ensured. |
| Kirsch, B. R., Characklis, G. W., & Zeff, H. B. (2013). | Modified Fractional Gaussian Noise (mFGN) integrated with a stochastic multi-reservoir | Monte Carlo simulation with 5,000 realizations; evaluation of expected outcomes and | Comparison of mFGN-generated synthetic streamflows to historic 82-year record; PACF | Un-whitening procedure allows adjustment of mean (mj) and standard deviation (sj) | Demonstrated ability to generalize over multiple climate scenarios; preserves seasonal |

| | | | | | |
|---|---|---|---|---|---|
| | water supply model. | extreme events (CDFs of lowest observed reservoir storage). | and ACF checks; extreme event frequency (droughts). | to simulate future hydro-climate scenarios; scenarios include increased inflow volatility (SD 10, SD 20) and mean reduction (MR10). | correlation structure and interannual autocorrelation, allowing simulation of conditions beyond the historical record and evaluation of risk-based transfer agreements under changed climates. |
| Li, C., & Singh, V. P. (2014). | Multi-model Regression-Sampling (MRS) Algorithm. | Implicit uncertainty captured through gamma kernel density estimation (KDE) of residuals and an ensemble of multiple regression models; probabilistic streamflow generation without assuming fixed parametric distributions. | Leave-One-Out Cross-Validation (LOOCV) for hyperparameter tuning; comparison with MKNN, NPL, and KGKA across 17 gauges (split-sample testing). | Limited; model assumes stationarity and authors note inability to simulate nonstationary streamflow; suggest inclusion of external covariates for future work. | Demonstrates improved reproduction of historical variability, extremes, and distributional characteristics; capable of generating richer and more diverse synthetic scenarios compared to single-model frameworks but restricted to single-site applications. |
| Ramirez Molina, A. A., Frame, J. M., Halgren, J., & Gong, J. (2025). | Long Short-Term Memory (LSTM) networks combined with synthetic "Leaky Bucket" hydrologic models (DeepBucketLab framework) | Implicitly assessed via model performance under varying flow regimes (normal, low-flow, high-flow); no formal probabilistic uncertainty quantification (e.g., Bayesian or ensemble methods) applied. | Hyperparameter optimization using validation datasets; independent test datasets used for evaluation; performance assessed with NSE, RMSE, and MAE. | Explored through experiments under low-flow and high-flow extremes; model performance declines slightly under extremes, indicating moderate robustness to distribution shifts. | Demonstrated strong generalization within synthetic watershed networks; able to reproduce upstream flow patterns from downstream signals; generalizability to real-world basins proposed but not yet validated. |

| | | | | | |
|---|---|---|---|---|---|
| Wang, W., Hu, S., & Li, Y. (2011). | Wavelet Transform Method (nonparametric stochastic approach for synthetic daily streamflow generation). | No explicit uncertainty quantification: uncertainty addressed indirectly through ensemble generation (100 synthetic samples) and variability analysis of statistical parameters. | Multi-level validation using 100 generated samples; comparison of means, standard deviations, skewness, autocorrelation, flood volumes, monthly runoff, and seasonal variation against observed data. | Not explicitly tested for non-stationarity or climate-induced distribution shifts; method assumes stationarity of historical records. | Demonstrated strong reproduction of observed statistical characteristics; applicable to both stationary and non-stationary hydrological series in principle, but not externally tested beyond a case study. |

# 3. Methodology

As this research aims to develop an interpretable streamflow synthesis framework grounded in quantified hydrological characteristics, the methodology is structured in two main stages: systematic characterization of streamflow properties (Section 3.1) and development of a semi-automated synthesis approach (Section 3.2).

## 3.1 Quantification of Streamflow Characteristics

To effectively quantify the key characteristics of streamflow, it is essential to analyze its temporal dynamics and underlying statistical properties. The following subsections detail specific aspects of streamflow behaviour, including persistence and memory, periodicity and seasonal structure, and noise characteristics.

### 3.1.1 Persistence and Memory Analysis

Hurst (1951) introduced a pioneering method for characterizing long-term persistence in time series, known as Rescaled Range (R/S) analysis, which quantifies the self-similarity and long-range dependence in a natural streamflow process. The procedure to estimate the Hurst exponent, H, follows several steps, as outlined in Appendix 3A.

### 3.1.2 Periodicity and Seasonal Structure

Streamflow time series exhibit distinct periodic and seasonal characteristics that vary across temporal scales. At an annual scale, streamflow patterns emphasize long-term persistence, including sustained flood and drought periods extending over decades and, in some cases, reflecting broader periodic influences such as sunspot cycles. At finer temporal resolutions, particularly the monthly scale, streamflow behaviour is governed by pronounced seasonality, characterized by recurring patterns within and across years. These seasonal structures may manifest as 2-month, 3-month, 4-month, 6-month, or 12-month periodic components, necessitating explicit consideration in synthetic streamflow modelling.

To adequately capture these periodic and seasonal features, it is essential to examine streamflow characteristics across multiple temporal scales, including annual, monthly, weekly, and daily resolutions. Monthly-scale analysis is particularly important for identifying and incorporating seasonality, as it reflects the cyclic response of hydrologic systems to climatic forcing. A comprehensive understanding of such seasonal variations is therefore required to formulate models capable of accommodating diverse seasonal patterns. This consideration naturally leads to specialized representations, such as 6-monthly structures corresponding to two dominant seasons within a geophysical year, as demonstrated by Panu and Unny (1980) for Canadian river systems.

The representation of seasonal periodicities in hydrological time series commonly involves segmentation of the data into distinct seasonal components. Such segmentation does not necessarily need to coincide with conventional geophysical seasons or uniform temporal divisions. Instead, seasonal boundaries are more effectively determined by examining inherent data patterns and hydrologic behaviour. The number of seasonal segments required depends on the variability, diversity, and distinctiveness of environmental conditions experienced within a geophysical year at a given site.

A straightforward example of seasonal structuring involves classifying the hydrologic year into two dominant seasons, typically wet and dry periods, which together describe the primary seasonal behaviour of the streamflow time series. In regions with more complex climatic transitions, however, one, two, three, four, or even six seasonal segments may be required to adequately represent the underlying variability. While increasing the number of seasons provides finer temporal resolution, it also results in smaller pattern vectors. Consequently, selecting a limited number of seasons, such as two seasonal segments, represents a strategic balance between capturing essential seasonal variability and minimizing information loss due to excessive segmentation.

### 3.1.3 Noise Colour Identification

Time series can often be viewed as a combination of multiple sinusoidal waves with varying frequencies or periods. Based on such an assumption, a time series, when analyzed in the frequency domain using the Fourier transform, the spectral density function describes the contribution of

each frequency to the overall variance of the series. The Fourier transform of a hydrologic time series, $x(t)$, into its frequency domain representation, G($f$), is expressed as follows:

$$G(f) = \int_{-\infty}^{+\infty} x(t) e^{-2\pi i f t} dt, \quad (3.1)$$

where $t$ represents time, $f$ is the frequency, and $i = \sqrt{-1}$. Typically, a PSA is used to assess the periodic strength in a time series across various timescales, such as monthly, annual, and decadal. However, the existence of a linear relationship between a frequency and a power spectrum on a log-log scale indicates scale invariance in the temporal domain. Such a relationship is expressed as follows:

$$P(f) \propto 1/f^{\lambda}, \quad (3.2)$$

where $P(f)$ is the power spectral density (PSD) at that frequency, and the scaling exponent λ is the slope of the line of best fit on a log-log plot. This scale-invariance characteristic of the power spectrum is known as $1/f^{\lambda}$ Noise, where the colour of the noise varies based on the value of λ.

In analyzing the scaling behaviour of streamflow time series, previous studies have utilized a cross-point in frequency to divide different regimes of streamflow (Telesca et al., 2012; Thompson & Katul, 2012; Kim et al., 2016). Likewise, a cross-point in this study is defined as 1 year, as utilized by Telesca et al. (2012) for monthly streamflow. Therefore, for a frequency window of $f$ >1/12 months (= 0.833 months$^{-1}$) (1 year), the power law can be analyzed.

In this window of $f$ > 0.833 months$^{-1}$, Telesca et al. (2012) observed notable jumps at both 12- and 6-month intervals in the power spectrum of monthly streamflow data. They ascribed these peaks in the power spectrum to the frequencies at which cycles and patterns dominate. Such an observation was further investigated in more detail by Wen and Liu (2016), who also concluded that oscillatory patterns in a signal appear as distinct peaks in the PSD at the corresponding oscillation frequencies. The height and sharpness of such distinct peaks depend on the amplitude and purity of oscillations.

Considering jumps in the PSD correspond to frequencies where certain patterns dominate, it can be expected that a jump will occur at the frequency of 12-month intervals, as this represents the frequency at which a hydrological cycle repeats. Additionally, if there are sub-patterns within the cycle, jumps can be expected at frequencies that match the durations of these sub-patterns. Telesca

et al. (2012) reported two jumps at frequencies of $f$ = 1/12 and $f$ = 1/6 months$^{-1}$. The presence of different seasons (2, 3 and maybe 4) and sub-patterns within the monthly streamflow has been discussed by Panu and Unny (1980a); therefore, it is reasonable to expect peaks in PSD at frequencies corresponding to fractions of a cycle, such as 6 months, 4 months, 3 months, and 2 months. Figure 3.1 presents a conceptual diagram of PSD for monthly streamflow, where, within a cycle of 12 months, there exist different seasons and sub-patterns. The points labelled in Figure 3.1, corresponding to frequencies of $f$ = 0.0833, 0.1667, 0.25, 0.3333, and 0.5, represent periods of 12 months, 6 months, 4 months, 3 months, and 2 months, respectively. At each of these frequencies, a notable increase or 'jump' in PSD is anticipated, which is consistent with observations reported by (Telesca et al., 2012; Kim et al., 2016).

Should the power spectrum be expected to follow its scale-invariance characteristic of $1/f^{\lambda}$ Noise than the rule for frequencies $f$ > 0.0833, as recommended by Telesca et al. (2012) is that each of these observed jumps should also align with the relationship of $1/f^{\lambda}$.

Based on the foregoing discussion, in the case of monthly streamflow data, a jump in the PSD at a certain frequency indicates dominant patterns within the time series. Therefore, focusing on such jumps, where dominant patterns are most pronounced, can lead to a more accurate estimation of the scaling exponent, λ. Such considerations, in turn, enhance the estimation of the Hurst coefficient, H, in our quest to adequately account for both short-term and long-term dependence in hydrologic data series for improved modelling practices. It is in this vein that this study proposes a new methodology, as illustrated in Figure 3.2, to isolate jumps in PSD, which in turn, imbue the $1/f^{\lambda}$model to all such frequencies or patterns. By isolating jumps in PSD, and considering $P(f) \propto 1/f^{\lambda}$, the logarithm of both sides will be

$$\mathrm{Log}\big(\mathrm{P(f)}\big) = \mathrm{Log1} - \mathrm{Log}\ (\mathrm{f}^{\lambda}). \tag{3.3}$$

Therefore,

$$\mathrm{Log}\big(\mathrm{P(f)}\big) = \mathrm{C} - \lambda\,\mathrm{Log}\ (\mathrm{f}), \tag{3.4}$$

alternatively expressed as

$$\mathrm{Log}\big(\mathrm{P(f)}\big) = -\lambda\,\mathrm{Log}\ (\mathrm{f}) + \mathrm{C}. \tag{3.5}$$

This is used to describe the line of best fit to high-frequency points in a log-log plot, as shown in Figure 3.2. In Eq. 3.5, the constant C serves as a vertical shift in the log-log plot.

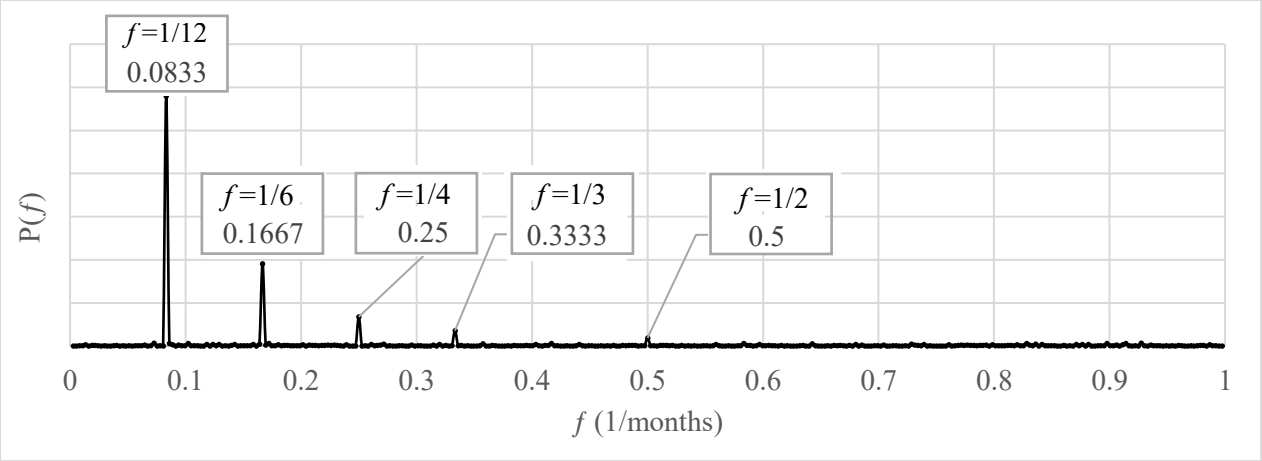
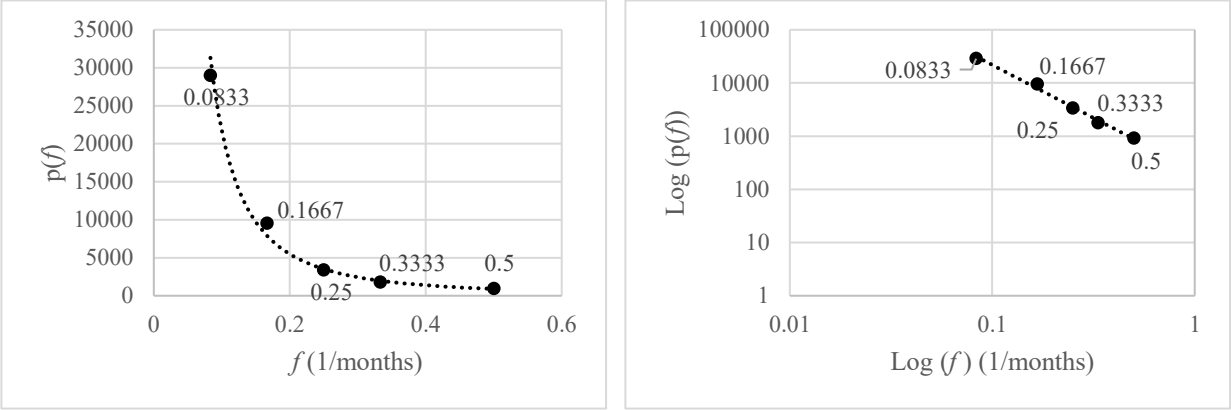



*Figure 3.1. Power spectral density (PSD) of a hypothetical monthly streamflow time series.*




*(a)* *(b)*

*Figure 3.2. Depiction of isolated peaks in the power spectral density (PSD) of a hypothetical monthly streamflow time series on (a) a linear scale and (b) a log–log scale.*

It is important to acknowledge that the number of these isolated peaks (presented in *Figure 3.2*), as well as their intensity, is a function of hydrologic conditions and the number of seasons within a year. Therefore, the identified scaling behaviour and corresponding classification may vary across different river systems.

The scaling exponent, λ, determines the colour of noise and describes several behavioural characteristics of a time series, as summarized in Table 3.1. It is to be noted that, as indicated in Table 1, the black noise is characterized by low-frequency dominance. Since the focus of this study

is on high frequencies above 0.0833 months$^{-1}$, hence when analyzing a signal with a frequency window of $f >$ 0.0833 months$^{-1}$, any classification as black noise implies that the lowest frequency in this range (0.0833 months$^{-1}$) dominates. Consequently, the high PSD value at this frequency results in a steeper slope in the log-log PSD plot, leading to a scaling exponent, λ greater than 2.

*Table 3.1. Categorization of various types of noises based on values of the scaling exponent "λ".*

| Scaling Exponent, "λ" | Noise Type | Behaviour | Memory |
|---|---|---|---|
| λ = 0 | White | Random | No Memory |
| 0 < λ < 2 | Pink | Mean Reverting | Long-term memory* |
| λ = 2 | Brown | Brownian Motion Process | No long-term memory (independent increments): Changes over non-overlapping time intervals are independent |
| λ > 2 | Black | Structured with Low-frequency Dominance | Long-term memory** |

Note: *Pink noise and **black noise both exhibit memory; however, pink noise has a shorter memory compared to black noise. In financial literature, pink noise is often ascribed to having a short memory, whereas in hydrologic studies, it is characterized by a long memory. The pink noise does possess memory, but its memory is not as persistent as that of the black noise.

## 3.2 Streamflow Synthesis Models

The streamflow synthesis framework developed in this thesis is designed to accommodate the diverse temporal behaviours observed in hydrological systems by integrating both spectral noise characteristics and pattern recognition principles. Streamflow records reflect the combined influence of basin storage, climatic forcing, and catchment response mechanisms, which manifest as distinct persistence, memory, and periodic structures in the time series. Accordingly, the synthesis model is formulated to adapt its structure based on the dominant noise regime governing streamflow variability, while simultaneously preserving recurring temporal patterns and seasonal organization through pattern recognition techniques. This dual-perspective formulation ensures that both stochastic persistence and deterministic temporal structure are retained in the synthesized streamflow sequences. As shown in *Figure 3.3*, the identified dominant noise colours are used as the basis for selecting appropriate modelling approaches.

The expectation from a synthesized dataset is to capture, retain, and closely reproduce the statistical characteristics exhibited in historical streamflow realization. In all developed models of

this section, the initial step is segmentation into seasons. At the monthly scale, it could be the seasonal behaviour may include 2-month, 3-month, 4-month, 6-month, and 12-month patterns.

For demonstrating seasons in hydrological time series, the concept of seasonal periodicities or segmentation of a time series into distinct seasons is essential. Such segmentation does not necessarily need to correspond strictly to geophysical seasons or equal time intervals. Instead, the most appropriate way to define segments is by examining underlying data patterns rather than relying on predefined calendar-based or uniform temporal divisions.

The selection of an appropriate number of seasonal patterns depends on the diversity, variability, and distinctiveness of environmental conditions experienced within a geophysical year at a given region or site. A simple example of such classification includes hydrological seasons divided into two groups (i.e., seasonal segments), namely dry and wet seasons, which collectively represent the seasonal structure of the hydrological time series (Panu et al., 1978).

In regions characterized by complex climatic variability or transitional periods, the number of required seasonal segments may vary, potentially including one, two, three, four, or even six seasons. In general, increasing the number of seasons leads to more specific temporal representation; however, it also results in smaller pattern vectors and reduced statistical stability within each segment.

Therefore, the use of a limited number of seasons, such as two seasons, in streamflow pattern recognition systems is a strategic choice that balances the need to capture significant variability while minimizing information loss associated with excessive segmentation. This approach also supports the optimization of entropic measures (Kojiri et al., 1994) by identifying an appropriate seasonal resolution. Although more than two segments may better represent detailed seasonal variability in some cases, a two-season representation provides a simplified and effective framework for streamflow synthesis models.

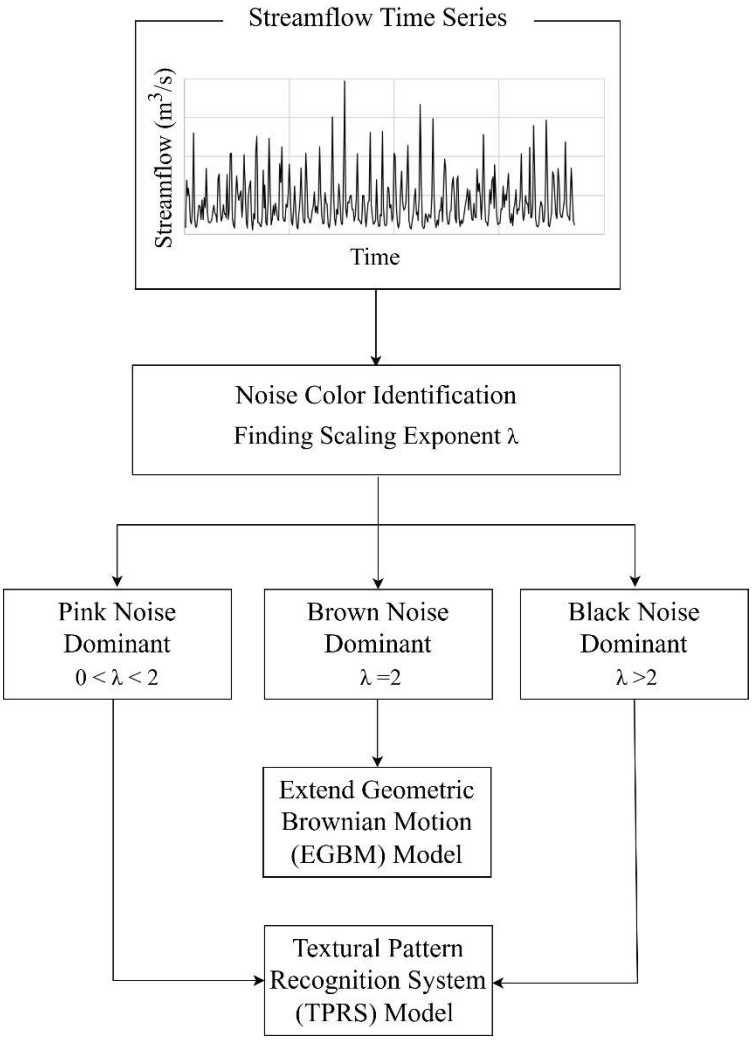



*Figure 3.3. Framework for selecting appropriate synthesis models based on dominant noise colours (pink, brown, and black) in a monthly time series.*

### 3.2.1 Streamflow Synthesis Model Based on Noise Characteristics

The spectral properties of streamflow time series provide critical insight into basin persistence, memory, and storage behaviour, which in turn inform the appropriate structure of the synthesis model. Depending on the dominant noise regime, different modelling strategies are required to adequately reproduce the observed temporal dynamics.

#### 3.2.1.1 Pink Noise Dominance

In systems dominated by *pink noise*, streamflow exhibits moderate persistence and a balance between low-frequency and high-frequency variability. Such behaviour is commonly associated with urban or flashy basins where storage capacity is limited and hydrologic response is rapid. For these systems, synthesis models must emphasize short-memory structures and event-scale variability while retaining sufficient continuity to reproduce observed correlations.

ARIMA models are well-suited for pink-noise-dominated basins, as they can capture the moderate persistence through autoregressive terms and reproduce event-scale variability through moving-average components, providing a practical framework for generating realistic synthetic streamflow sequences.

#### 3.2.1.2 Brown Noise Dominance

*Brown noise dominance* is characterized by a steep spectral decay and exceptionally strong low-frequency persistence, indicating high system inertia and substantial storage control. Although less common at the monthly scale, the presence of brown noise suggests a very slow hydrologic response and long memory. Brown noise is characterized by a scaling exponent $\lambda \approx 2$, indicating strong low-frequency persistence and long-term memory. To support this classification, a diagnostic check based on first-order differencing is also employed. Since brown noise transforms into white noise after first-order differencing (Dooley & Van de Ven, 1997), the autocorrelation function (ACF) at lag 1 of the differenced series is expected to be approximately zero. Therefore, in addition to the spectral criterion ($\lambda \approx 2$), stations are confirmed as brown noise when the lag-1 ACF of the differenced series is close to zero.

In such cases, synthesis models must prioritize the preservation of long-term persistence and ultra-low-frequency variability to avoid artificial attenuation of storage-driven dynamics.

*Brownian motion/Wiener process* is a stochastic process in which all samples ($W_t$) have a multivariate normal distribution with a mean of zero and a variance of t. In other words, the longer the time horizon into the future, the higher the variance will be. The joint distribution can be written as

$$F(t_1, t_2, \dots, t_n) \leftrightarrow N(0, (t_1, t_2, \dots, t_n)). \tag{3.6}$$

Therefore, a standard n-dimensional Brownian motion is a vector-valued stochastic process $\{W_t; t \in T\}$ in which the increments of each component $W_t$ are independent and normally distributed, thus resulting in one-dimensional Brownian motions.

Brownian motion, as a continuous-time random walk, is visualized as a walk along a line, beginning at $W_0 = 0$ and moving at each incremental time interval, dt, either up or down by an amount $\sqrt{dt}$. If the position of the walk is denoted by $W_n$, then after the $n^{th}$ move, the position can be expressed as follows:

$$W_n = W_{n-1} \pm \sqrt{dt}. \tag{3.7}$$

On the other hand, the *standard Brownian motion* process has a drift rate of zero and a variance of one. The drift rate of zero means that the expected value of $W_t$ at any future time is equal to the current value. The variance of one means that the variance of the change in $W_t$ over a time interval of length T is equal to T. The Brownian motion process is the basis for a collection of more general stochastic processes known as continuous-time random walks.

Standard Brownian motion is centred around zero. To model a time series, a Brownian motion with drift—obtained by scaling and shifting a standard Brownian motion—is often used. Suppose that $[W_t;\ t \geq 0]$ is a standard Brownian motion, and that $\mu \in R$ and $\sigma \in (0, \infty)$. And subsequently letting $Y_t = \mu t + \sigma W_t$ for $t \geq 0$. Then $[Y_t;\ t \geq 0]$ is a Brownian motion with a drift parameter μ and a scale parameter σ.

A GBM is a continuous stochastic process in which the logarithm of a randomly varying quantity follows a Brownian motion with a drift. A stochastic process $X_t$ is said to follow a GBM if it satisfies the following stochastic differential equation:

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \tag{3.8}$$

where $W_t$ is the Wiener process or Brownian motion. For an arbitrary initial value $X_0$, the above stochastic differential equation has the following solution:

$$X_t = X_0 e^{\left(\mu + \frac{\sigma^2}{2}\right)t + \sigma W_t}, \tag{3.9}$$

which is also called Exponential Brownian Motion. Taking the natural logarithm of both sides of

the GBM equation:

$$\ln(X_t) = \ln(X_0) + \left(\mu + \frac{\sigma^2}{2}\right)(t) + \sigma W_t. \quad (3.10)$$

As shown in Equation 3.10, exponential growth is a key characteristic of the GBM model. This feature becomes a limitation in the field of hydrology, where such unbounded growth may not be realistic for certain long-term environmental processes. To address this limitation, approaches such as a suitable segmentation process and an extended GBM model, discussed below in the following section, could potentially be utilized. It should be noted that the GBM model itself, defined as a potential solution of the stochastic differential equation, is obtained by solving via a lemma by Itô (Petters & Dong, 2016), states that $[X(t) = X_0\, e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B(t)}]$, where the term $(-\sigma^2/2)$ is the Itô correction, which ensures that the stochastic path satisfies the stochastic differential equation. However, when taking the expectation, the term $(+\sigma^2/2)$ appears due to the properties of the log-normal distribution.

To overcome the main limitation of GBM process, which lies in its inability to model distant future events due to the inherent randomness and unpredictability of a time series over extended time periods, a natural logarithmically transformed monthly streamflow time series is represented as a matrix with 12 rows (corresponding to months) and N columns (corresponding to years), which is further divided into $K$ segments representing $K$ seasons.

One possible approach is to use the Lyapunov Time (LT), or time horizon, which represents the timescale over which a system remains predictable (Mihailović et al., 2023). Beyond the LT, the system exhibits a strong tendency toward unpredictability due to chaos or sensitivity to initial conditions. Considering a case where LT ≥ 6 months, the accuracy of modelling is high within a period equal to or greater than 6 months, making it suitable for segmentation into two seasons (6 months per season). Therefore, a case with two seasons (K=2) is considered, where each season k=1 and k=2 is represented by a matrix of dimension $6 \times N$. Therefore, each season consists of six months (I = 1 to 6) from year j = 1 to N.

In season k, the natural logarithm of streamflow, ln $(X_{i,j})^k$ of i$^{th}$ month within the j$^{th}$ year, can be modelled as follows.

$$\ln (X_{i,j})^k = \ln (X_{0,j})^k + \left(\mu^k + \frac{(\sigma^k)^2}{2}\right)(i) + \sigma^k W_{i,j}^k, \quad (3.11)$$

where $\mu^k$ and $\sigma^k$ respectively, are the drift and volatility parameters of season k, and $\ln (X_{0,j})^k$ is an initial value or a seed value. Eq. 3.11 is another form of GBM (discussed earlier in Eq. 3.10) applied to a matrix form of the segmented monthly streamflow data over different years. The application of Eq. 3.10 is suited to a time series, whereas Eq. 3.11 is suited for a matrix form of segmented monthly data.

In the next step, an extended GBM is developed by multiplying a constant coefficient ($\ln (\Omega_i^k)$) to each seasonal element (i.e., to the i$^{th}$ month within a season $k$), which controls the inherent growth of a GBM process for higher lags while retaining the essential Brownian noise pattern(s).

$$\ln(X_{i,j})^k = \ln(X_{0,j})^k + [\left(\mu^k + \frac{(\sigma^k)^2}{2}\right)(i) + \sigma^k W_{i,j}^k] \times \ln(\Omega_i^k), \quad (3.12)$$

which is called an extended GBM (EGBM). The first part $[\ln\left(X_{0,j}\right)^k + \left(\mu^k + \frac{(\sigma^k)^2}{2}\right)(i) + \sigma^k W_{i,j}^k]$ of Eq. 3.12 is the GBM (also represented in Eq. 3.11) and the second part $[\ln(\Omega_i^k)]$ in Eq. 3.12, the seasonal parameter for $i^{th}$ month, which, in turn, is determined using an optimization algorithm, GA.

Each monthly value $(\ln(X_{i,j})^k)$ in Eq. 3.12 is a function of the initial value, drift parameter and volatility of season k, and the Wiener process. In AR-family models, such as AR (1), each value is a function of an autoregressive parameter and the previous value. In this context, however, the key difference with the GBM process is that each value is a function of an initial value. That is why a certain time duration cannot be exceeded, which can be identified using a suitable value of LT for different geographical regions.

The use of the proposed EGBM model for streamflow synthesis requires the determination and optimization of seasonal parameters $(\ln\left(\Omega_i^k\right))$ for each season. One potential approach for optimization would be GA.

For operational reasons, the adoption of the proposed EGBM model for streamflow synthesis requires one additional value than the seasonal length (in the present study: six values) to be

generated in each season. The additional value, thus generated for each season, (i.e., $7^{th}$ value), serves as a seed value ($\ln(X_{0,j})^k$) for the subsequent season. To commence the monthly streamflow synthesis process, a random seed value is generated using the statistical parameters of the month before the first month of the first season in the first year of the monthly streamflow synthesis process, as shown in *Figure 3.4*.

#### 3.2.1.3 Black Noise Dominance

Systems exhibiting *Black noise dominance* reflect highly persistent behaviour governed by substantial storage in soils, groundwater, and wetlands. These natural basins require synthesis models capable of representing long-memory processes and sustained autocorrelation across extended time scales. Storage-based and state-dependent formulations are therefore more appropriate to reproduce the observed temporal structure of streamflow in such environments.

ARIMA can be used for black-noise dominated basins, and AR coefficients provide a quantitative way to analyze and interpret persistence, but for true black-noise, Auto Regressive Fractionally Integrated Moving Average (ARFIMA**)** or storage-based models are often more accurate in capturing the very long memory. Moreover, pattern recognition techniques are employed to extract and preserve the structured temporal features of the streamflow, such as seasonal cycles and multi-month wet/dry periods.

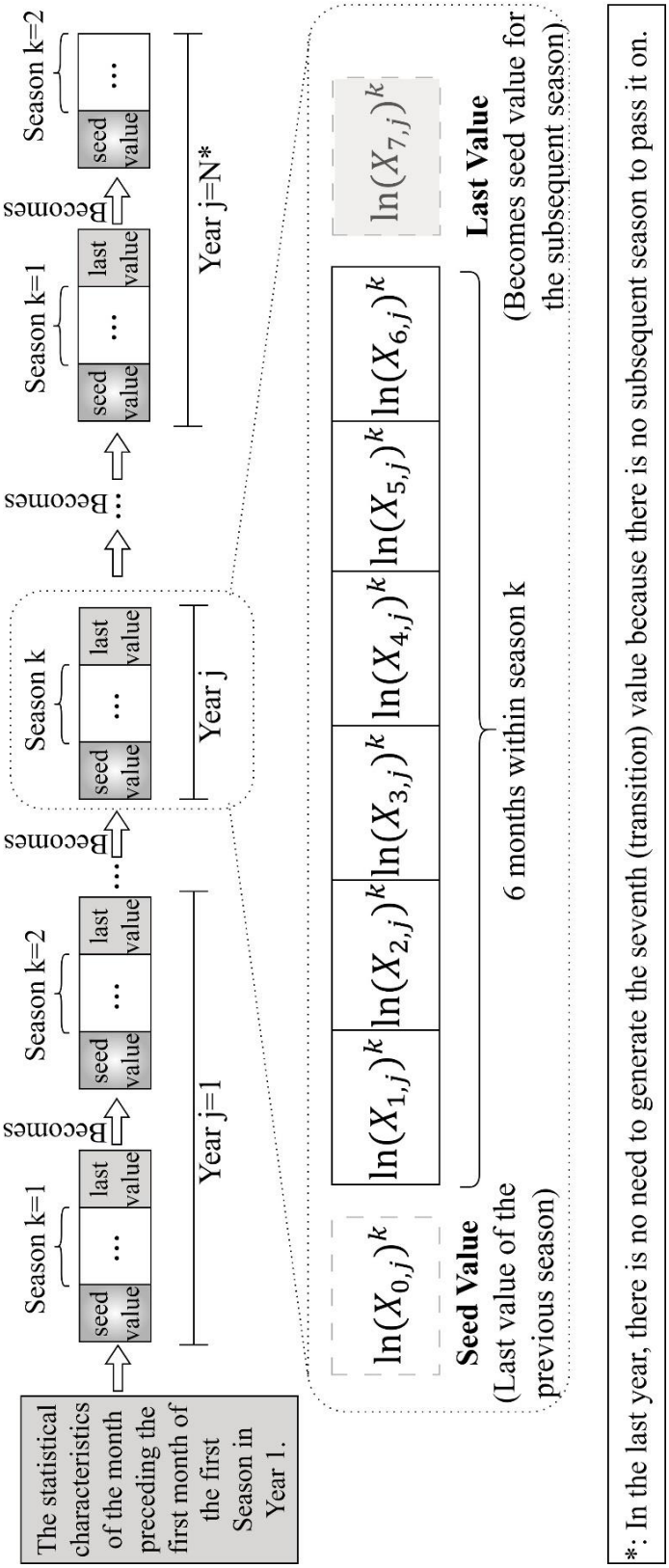



*Figure 3.4. A schematic streamflow synthesis diagram for the EGBM (Extended Geometric Brownian Motion) model for monthly streamflow series dominated by brown noise.*

### 3.2.2 Streamflow Synthesis Model Based on Pattern Recognition

The expectation from the synthesized dataset is to capture and retain or closely match statistical parameters and temporal patterns exhibited in the historical streamflow data. As noted above, pattern recognition techniques can serve as a valuable approach in the synthesis of datasets, particularly when aiming to replicate or emulate statistical parameters and temporal patterns observed in the historical streamflow data. Pattern recognition can be extended to various time scales, which otherwise necessitates different approaches because of the uniqueness of characteristics exhibited by each time scale. For example, annual streamflow patterns emphasize persistence in flood and drought periods over longer periods, monthly patterns highlight seasonality within each year and across multiple years, while daily patterns evolve around rapid fluctuations driven by immediate cause-and-effect relationships between rainfall and streamflow over shorter time intervals. Streamflow exhibits complex, chaotic, and nonlinear attributes across various time scales. Such complexities prompt exploration beyond conventional methodologies, thus paving a path for research in emerging domains such as image analysis and synthesis. Transitioning from streamflow time series to image synthesis widens the scope of analysis, enhances visualization capabilities, and leverages advanced computational methods to extract and comprehend complex patterns present in hydrological datasets. This study aims to develop textural image processing techniques to expand and enhance the performance of the feature prediction model in synthetic hydrology, as advocated by Panu et al. (1978), and further advanced by recent developments in stochastic hydrology (Parasuraman, 2007; Li et al., 2021b; Tang et al., 2023; Liu et al., 2025).

It is essential to consider diverse characteristics of streamflow time series across various temporal scales, such as annual, monthly, weekly, and daily. At an annual scale, the focus should be on overarching features such as long-term persistence. Shifting to monthly scales, attention can be directed towards identifying and incorporating seasonality aspects, encompassing 2-month, 3-month, 4-month, 6-month, and 12-month patterns. It is crucial to establish a comprehensive understanding of these seasonal variations and to articulate a model that accommodates diverse seasonal patterns. Such a consideration leads to the introduction of some specialized cases, such as the development of a model tailored to 6-monthly patterns, that is, the case of two seasons in a geophysical year, as was utilized by Panu and Unny (1980) on three rivers across Canada. Using the same data set on these three rivers will provide a unique opportunity for a comparative analysis

to test the efficacy of the proposed method. This approach ensures a systematic exploration of streamflow characteristics and formulation of a synthetic model capable of handling a wide array of temporal patterns.

In this process, the initial step involves segmentation of the monthly streamflow time series ($X = [x_1, x_2, \cdots, x_N\ ]^T$) into $N$ distinct geophysical years ($X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,N} \\ \vdots & \ddots & \vdots \\ x_{12,1} & \cdots & x_{12,N} \end{bmatrix}_{12\times N}$), each of such a geophysical year is further segmented into $K$ sections (here, $K = 2$) representing the number of seasons in a geophysical year. For demonstrating pattern recognition in hydrological time series, the concept of seasonal periodicities or their representation in the form of segmentation of a time series into distinct seasons is necessary. Segmentation into seasons need not strictly match the geophysical seasons or have equal durations, and the best way to slice such segments/sections is by examining the data patterns rather than following a specific geophysical or equal time-based divisions. The requirement for different numbers of seasonal patterns is contingent upon the diversity, variability, and distinctiveness of environmental conditions experienced throughout a geophysical year in a particular region or a specific site. A reasonable, straightforward example of a dataset includes hydrologic seasons categorized into two groups (i.e., seasonal segments), the dry season and the wet season. These categories describe all the seasonal segments within the hydrological time series (Panu et al. 1978).

In regions characterized by diverse transitional periods or intricate climatic variations, the necessity might arise for one, two, three, four, or even six seasons. Generally, with a higher number of seasons, each representing a more specific time frame, the resulting pattern vectors tend to be smaller. Opting for a limited number of seasons, such as K=2, in streamflow pattern recognition systems, is strategic to strike a balance between capturing substantial variations within the data and avoiding information loss, particularly stemming from increased segmentation. This decision seeks to optimize the entropic measures (Kojiri et al., 1994) by pinpointing the most suitable length for seasonal patterns. While in some cases, more than two sections might better represent the actual seasonal variations, using two segments (K=2) could be an initial approach and a simplified representation that fits the purpose of the pattern recognition and provides an opportunity to compare the efficacy of the proposed method to the feature extraction model of Panu et al (1978). In turn, each segmented seasonal pattern is converted into a grayscale textural image ($G^k, k = 1, 2(= K)$)as described in section 2.1. A logical flow diagram describing various

elements of the innovative method of streamflow synthesis based on concepts of textural image encoding techniques is illustrated in *Figure* 3.5.

For the extraction of seasonal features, a novel textural image feature extraction method is developed, which in turn, is employed on the matrices of seasonal textural images ($G^k, k = 1, 2(= K)$). After extracting seasonal textural features, a seasonal reference matrix is generated (section 2.3). Subsequently, a textural image feature synthesizer is developed for the synthesis of textural images. These synthesized textural images are then decoded back into streamflow time series and are assessed through image-based criteria and time series-based criteria.

#### 3.2.2.1 Encoding Monthly Streamflow into Grayscale Textural Images

A grayscale textural pixel ($g_{i,j}^k$) representing a streamflow rate over time ($x_{i,j}^k$) is created by encoding into shades of gray. The process involves normalization within the range [0,1] using the normalization formula of $[(x_{i,j}^k - \min(x_{i,j}^k))/(\max(x_{i,j}^k) - \min(x_{i,j}^k))]$ within each season (Pei & Lin, 1995; Zhang et al., 2020) followed by a scaling process to fit within 0 to 255 (i.e., 256 shades of grayscale pixels). Collectively, these grayscale pixels corresponding to each seasonal pattern are arranged in the form of seasonal pattern matrices ($G^k$) as follows.

$$G^k = \begin{bmatrix} g_{1,1}^k & \cdots & g_{1,N}^k \\ \vdots & g_{i,j}^k & \vdots \\ g_{6,1}^k & \cdots & g_{6,N}^k \end{bmatrix}_{6\times N} \quad \text{for i= 1,2, …, I; j = 1,2, …, N; and k=1, 2(=K),} \tag{3.13}$$

where, I is the number of elements in the seasonal pattern (I = (number of months in year/number of seasons in a year) = (12 / (K=2)) = 6; N is the number of years of streamflow data; and K is the number of seasons in a year = (number of months in a year/ number of elements in a season) = (12 / (I = 6)) = 2.

Such a method ensures that each streamflow rate (i.e., signal) is digitally represented as a textural pixel with a brightness value ranging from 0 to 255, providing a visual depiction of the individual flow rate $x_{i,j}^k$ into $g_{i,j}^k$. A group of textural pixels (here, six pixels) corresponding to a streamflow seasonal pattern is called a pattern-textural-image. A collection (or collage) of pattern-textural-images of the same season, hereafter referred to as a seasonal-textural-image (*Figure* 3.6). In all

such textural images, brighter pixels (255) indicate higher brightness and darker pixels (0) indicate lower brightness.

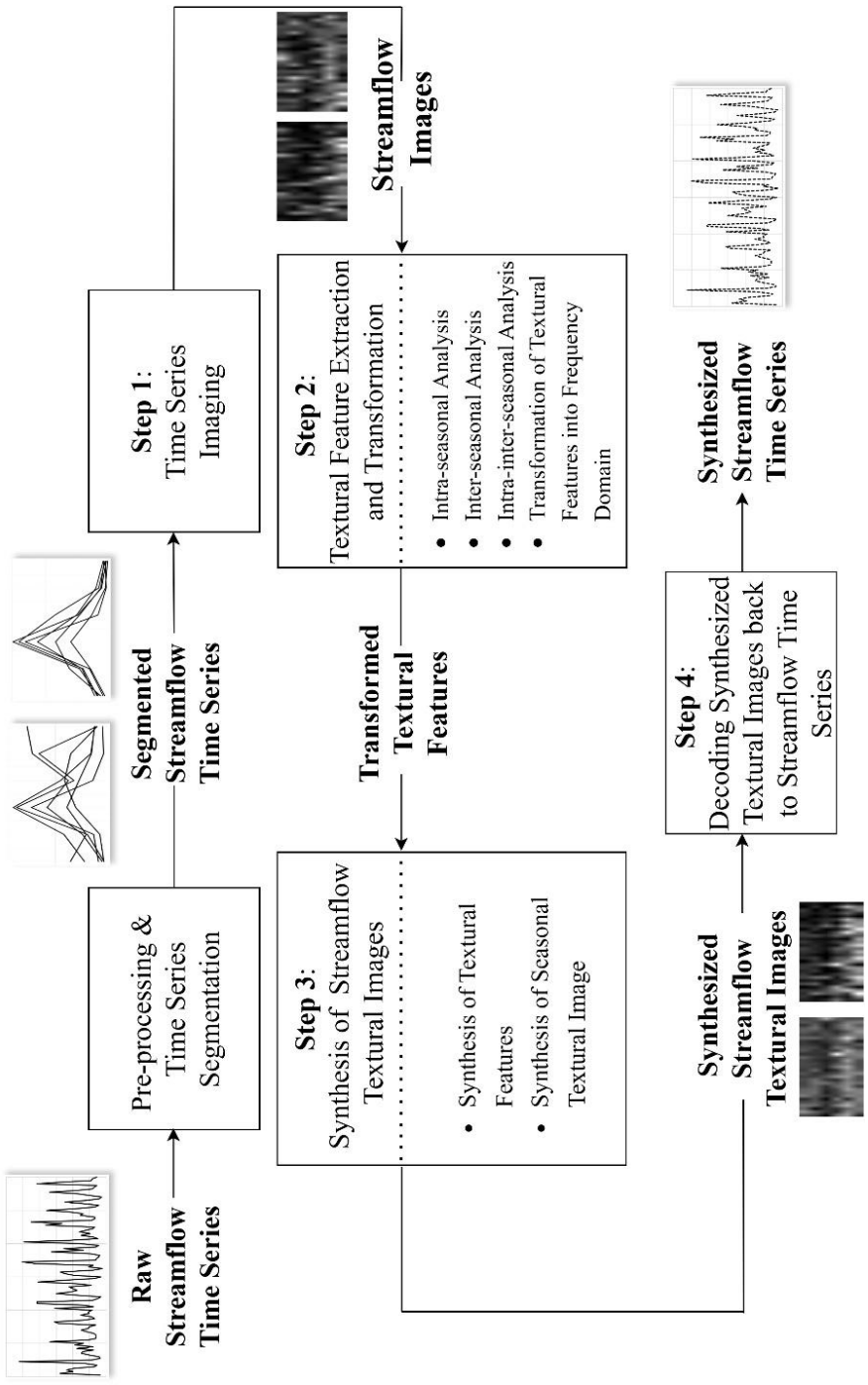



*Figure 3.5. A logical flow diagram for stochastic streamflow synthesis using an encoded textural feature recognition system.*

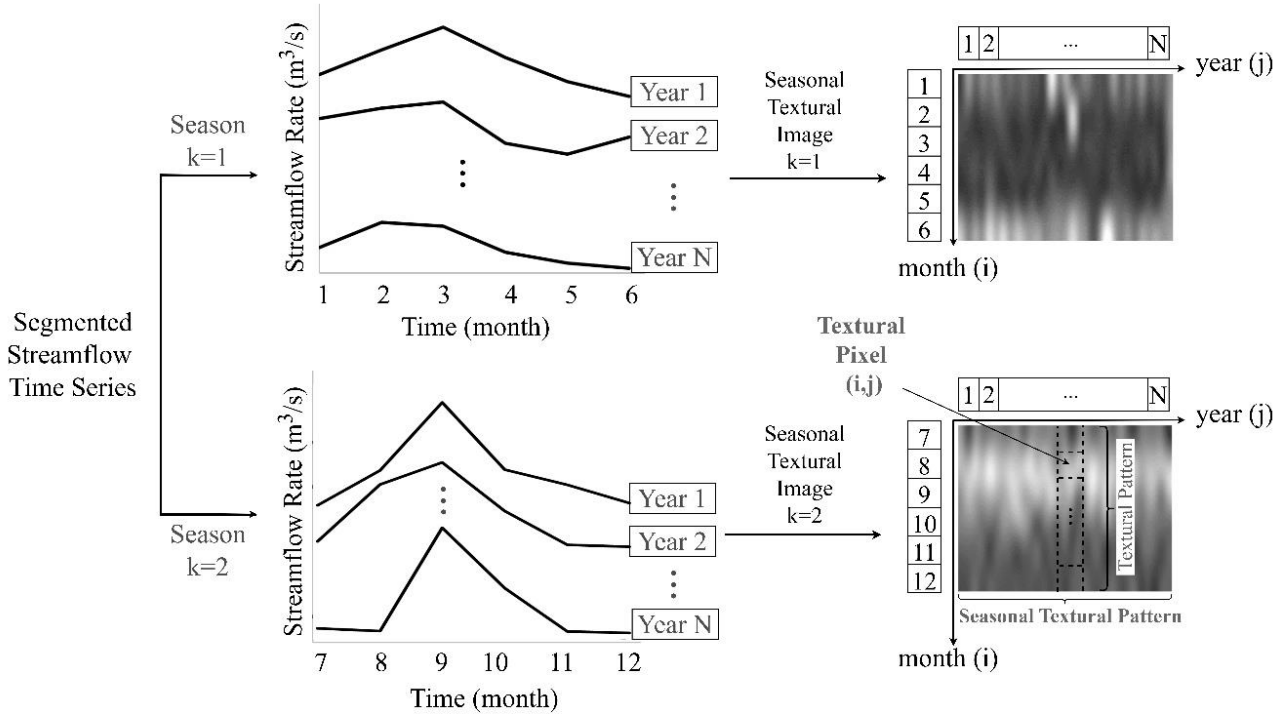



*Figure 3.6. The process of encoding monthly streamflow data into grayscale textural images.*

### 3.2.2.2 Textural Feature Extraction and Transformation

For clarity, it is noted that the auto-correlation function (ACF) in digital image processing captures the relationship between the grayscale intensity of a pixel relative to its neighbouring pixels based on their relative Cartesian locations (or positions). While the ACF has been valuable in analyzing fine-grained materials such as clays and micrites, but while expressed in conventional mathematical formulation involving $f(x,y) \otimes f(x,y)$ where $f(x,y)$ is a two-dimensional grayscale brightness function defining the textural composition of an image, but it may not directly be applicable when the principal goal is to synthesize textural images (Heilbronner, 1992).

One of the aims of this thesis is to synthesize textural images, and thus, a different approach is required to effectively describe the relationships and statistical properties of such unique textural images. Moreover, in an encoded textural image representing a time series, the conventional ACF, which typically involves correlating each pixel with its neighbouring pixels, poses limitations. In such a scenario, especially involving analysis of textural images of time series, the relevance of neighboring pixels may differ since textural images are sequential by nature due to their formulation in a time series (i.e., the development of a textural image involves the arrangement of

pixels in a sequential fashion in vertical and horizontal directions), thus emphasizing the importance of *previous pixels* in a time series rather than the entire neighboring set of pixels in an ordinary image or picture, where time remain the same for all pixels.

A monthly streamflow time series exhibits periodicity and correlations between flow values at a time '$t$' and the previous month $(t-1)$ are evident, and a notably strong correlation exists with the corresponding month in the previous year $(t-12)$. A visualization of this 2D-correlation in a time series plot can be intricate and challenging, but the representation of such a data set in a matrix format can simplify this 2D-correlation, where the previous row represents the immediate lag $(t-1)$, and the adjacent values in the same row correspond to a lag of 12 time points ahead (*Figure 3.7*). Encoding a streamflow time series into grayscale in a matrix format offers a reasonably effective representation of its intricate patterns and correlations defining such patterns.

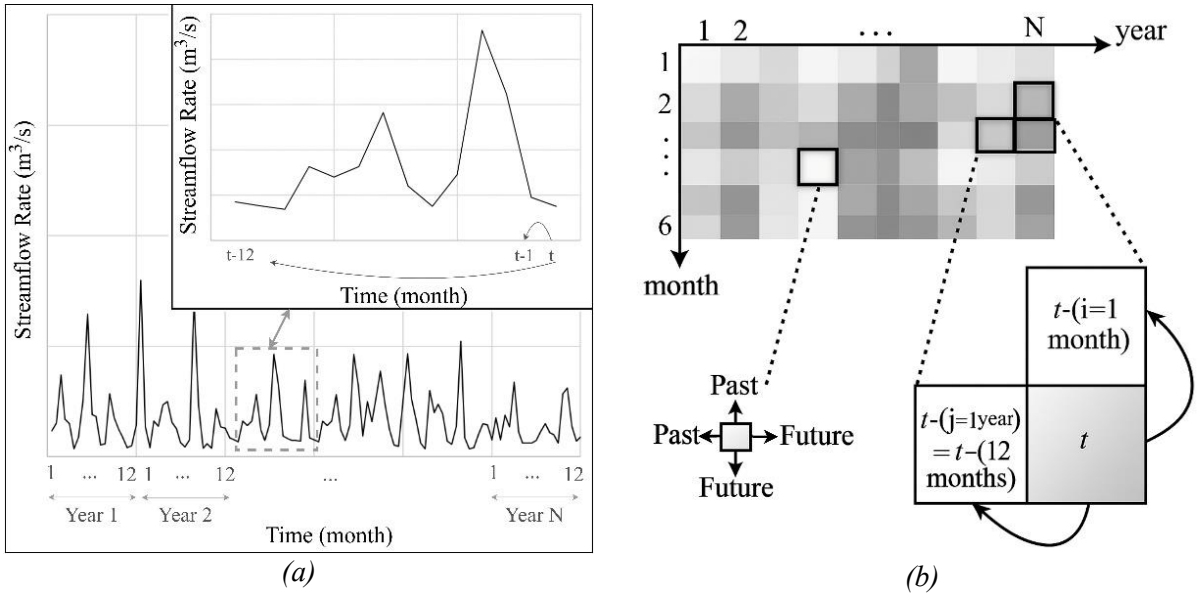



*Figure 3.7. 2D-correlation representation of a monthly streamflow time series using (a) the original time series and (b) an encoded textual-seasonal-image.*

Knowing that the nature of time series data only entertains relationship(s) to past (or previous) occurrences of the variable(s) of interest. Also, recalling that the development of a pattern-textural-image from a time series data of the variable of interest is segmented (or arranged) to develop a two-dimensional time representation, thus involves time in both horizontal and vertical directions. In other words, each pixel (of a pattern-textural-image) in such a representation

involves four directions of time. Two-directions relate to the past time, and likewise two-directions relate to the future time. That is, each pixel is related to two neighbouring pixels that had previously occurred (i.e., a pixel occurring in the horizontal direction pointing towards the west (i.e., towards the left of the pixel of interest) and another pixel occurring in the vertical direction pointing upwards to the north (Fig 3). Since correlative relationships and analyses always relate to past observations, a two-dimensional correlation structure for analysis of the (pattern or seasonal)-textural-image is required. In this case, two correlations are simultaneously occurring between the grayscale intensity of a specific pixel and the grayscale intensities of its two neighbouring pixels, hereafter referred to as the simultaneous auto-correlation function (SACF).

A grayscale textural image feature extraction model is developed upon obtaining a grayscale intensity (or brightness) matrix ($G^k$) for each season $k$. Since the grayscale intensity matrix of a streamflow time series data is divided through a segmentation process, specific intra- and inter-pattern relationships are essential for extracting features of the seasonal-textural-image. The *intra-season-pattern* analysis *within* a seasonal-textural-image comprising many pattern-textural-images in a season ($G^k$, k =1, 2(=K)) and the *inter-seasonal-pattern* analysis *between* two different seasonal-textural-images $G^{k=1}$ and $G^{k=2}$. Additional analytical detail on intra-seasonal-pattern and inter-seasonal-pattern relationships is respectively depicted in *Figure 3.7* and *Figure 3.8*.

- **Intra-seasonal-pattern analysis:** To develop and encompass relationships within seasonal patterns, the grayscale textural image of each season is separately analyzed by developing suitable correlations between grayscale intensity values of a specific pixel to its adjacent pixels (that have previously occurred) in patterns (that is a pixel vertically upward in a column and another pixel horizontally to the left in a row of the specific pixel of interest). In other words, relationships within pixels of seasonal patterns are called simultaneous autocorrelations of the intra-seasonal patterns. The word "autocorrelations" is used because these correlations are simultaneously occurring among pixels of patterns in the same season. It is further noted that these simultaneous autocorrelations are always occurring with pixels that have already occurred in relation to the pixel of interest. Such an analysis provides insight(s) into temporal relationships of seasonal patterns in a streamflow time series that may not be easily apparent in a traditional time series analysis, since the grayscale textural image of each season ($G^{k=1}$ and $G^{k=2}$) can be considered in a 2-dimensional matrix of streamflow time series, where 2D-correlation is depicted (*Figure 3.7*).

To understand and analyze the intra-pattern relationship between grayscale intensity of a pixel in a seasonal pattern (i.e., pixels in a column) to grayscale intensities of its adjacent and previously occurring pixels in a vertical, as well as, a horizontal direction (i.e., a previous pixel in the same column and a previous pixel in a row), a simultaneous auto-correlation function (SACF) is developed. The $SACF_{i,j}^{k}$ value of a pixel (i,j) in a pattern of a season (say k$^{th}$ season) is defined as a relationship to its $(i-\beta \; and \; j-\alpha)$ adjacently occurring pixels ($'\beta'$ pixel in the vertical direction, i.e., in a column and another $'\alpha'$ pixel in the horizontal direction, i.e., in a row) as depicted in *Figure 3.8*.

In the grayscale matrix $G^{k}$ of the k$^{th}$ season, the simultaneous deviation ($Sdev_{i,j}^{k}$) for any pixel (i,j), can be defined as a matrix representing the product of deviations in both horizontal and vertical directions for all i, j, and k (i =1, 2, …, I=6 and j=1, 2, …, N=30, and k =1, 2 (=K) as follows:

$$Sdev_{i,j}^{k} = [(g_{i,j}^{k} - \mu_{j}^{k})(g_{i,j-\alpha}^{k} - \mu_{i}^{k})][(g_{i,j}^{k} - \mu_{i}^{k})(g_{i-\beta,j}^{k} - \mu_{j}^{k})], \quad (3.14)$$

where $\mu_{i}^{k}$ and $\mu_{j}^{k}$ respectively denotes the expected value of pixels in each row (same month) and each column (same year) of the season k. By dividing $Sdev_{i,j}^{k}$ of a specific pixel ($g_{i,j}^{k}$) by the variance of its row ($(S_{i}^{k})^{2}$) and column ($(S_{j}^{k})^{2}$), the matrix of simultaneous correlation for all i, j, and k (i=1, 2, …, 6(=I) and j=1, 2, …, N=30, and k =1, 2 (=K)) is obtained as follows:

$$SACF_{i,j}^{k} = Sdev_{i,j}^{k} / (S_{i}^{k})^{2}(S_{j}^{k})^{2}. \quad (3.15)$$

In the computation of $SACF_{i,j}^{k}$, at lag-ζ (ζ=1) requires that the values of $'\alpha' \; and \; '\beta'$ be set equal to ζ=1 such that the simultaneous auto-correlation function for each pixel is calculated using the previously occurring pixels in both horizontal (row) as well as vertical (column) direction; therefore, the matrix ($SACF_{i,j}^{k}$) has one row and one column less (usually equal to the number of lags, in this thesis, the lag is equal to one) than the matrix ($G^{k}$). The reduction in dimensions occurs because the process involves computing correlations in the historical matrix ($G^{k}$) with itself but shifted by lag-ζ (ζ=1) in both horizontal and vertical directions (lagged) version (*Figure 3.9*). Thus, the process of conducting intra-season-pattern analysis involves one essential piece of evidence, which is the potential loss of information due to lagging (i.e., shifting) operation in the computation of the simultaneous autocorrelation function.

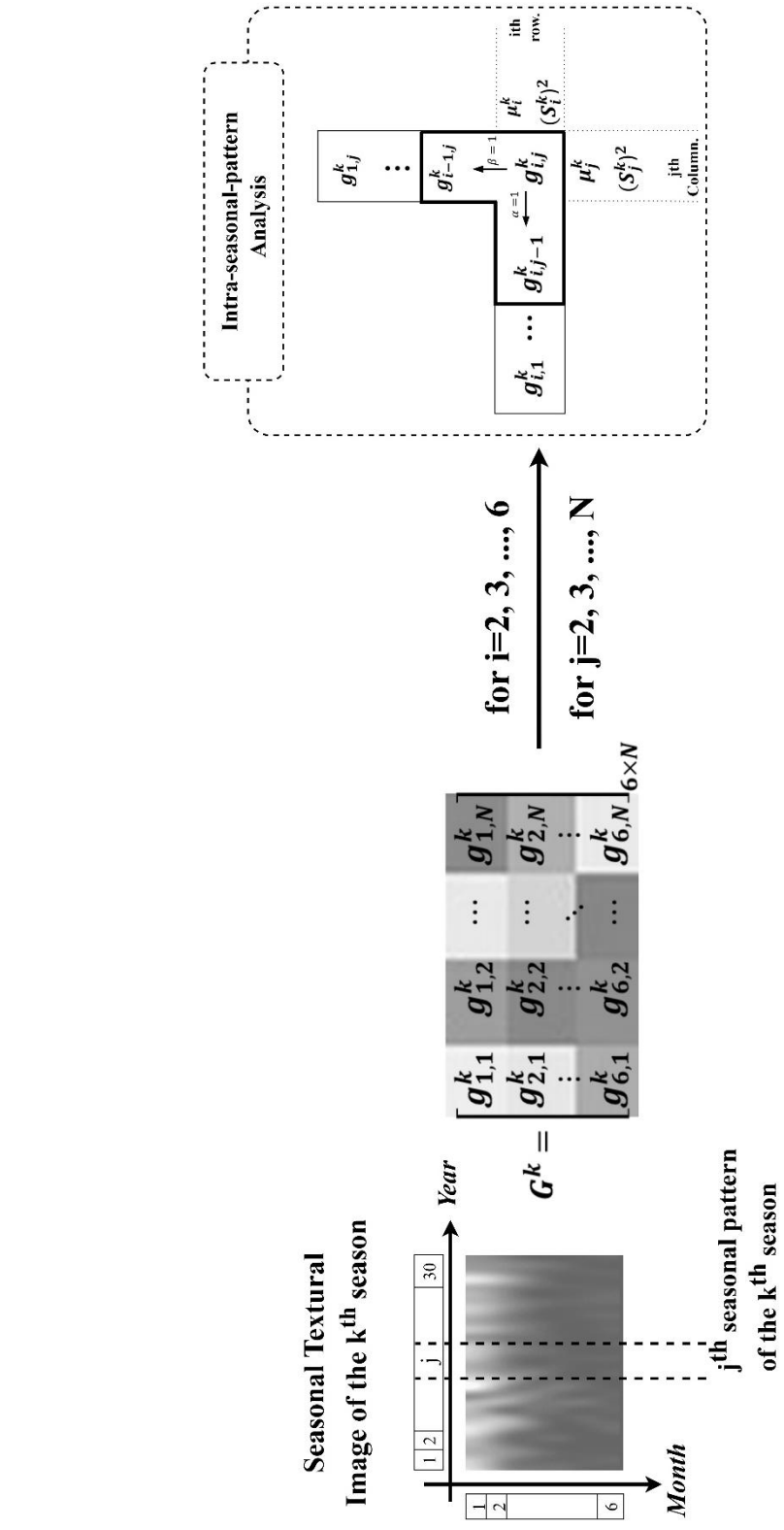



*Figure 3.8. Intra-seasonal-pattern analysis for a pixel at Lag-$\zeta$ ($\alpha = \beta = (\zeta = 1)$) of an 8-bit grayscale textural seasonal image $G^k$.*

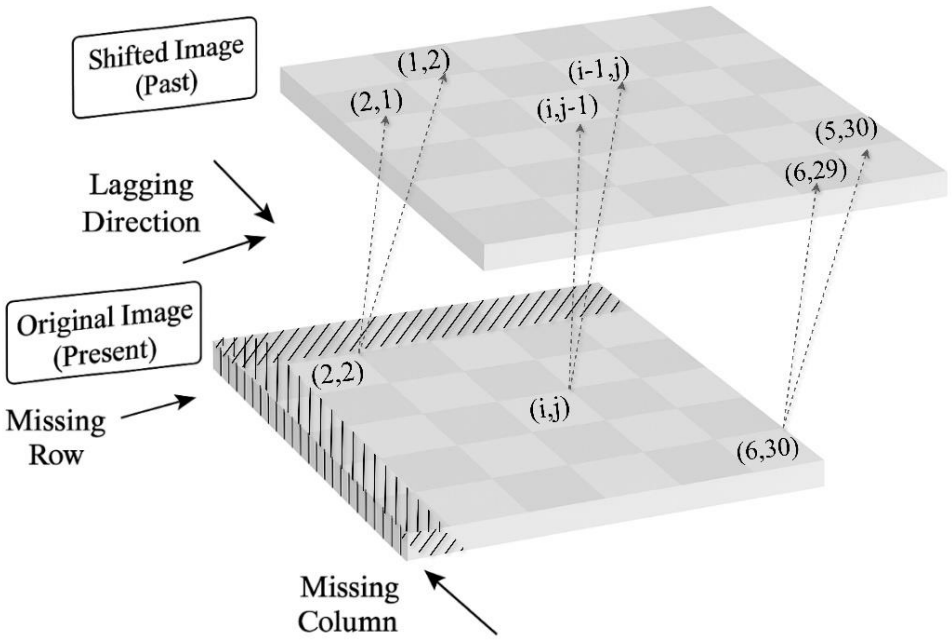



*Figure 3.9. A diagram illustrating $G^k$ and its lag-ζ (ζ=1) shift (to the right and downward) for calculating elements within a seasonal $SACF^k$ matrix.*

To reconstruct (or estimate) the unavailable information in *Figure 3.9* due to the shifting (or lagging) process of the matrix ($G^k$) in columns and rows; for example, in columns, the length of N years should be adjusted to ($N$ + ζ), and since ζ=1, therefore one should include (i.e., append) data for an additional year from the immediate past streamflow records for the analysis (or where such past streamflow records do not exist, then the assumption of circularity should be invoked). Similarly, to add more rows (months) to a season $k$, one needs to invoke the assumption of repeatability of hydrological seasons (as the hydrologic cycle is known to repeat); the last row of the previous season ($k$ − 1) can be utilized to establish and compute the required information in the form of correlations.

The first row of $SACF^k$ matrix that is missing due to the shifting/lagging process of the $G^k$, k =1, 2 = K matrices by one month lag (= ζ=1) can be estimated based on the assumption of Markovian dependence between successive occurrences (or repeatability of seasons in a year) of seasons (also given that K=2, hence both k-1 and k+1 refer to the same season (equal k=2)). Therefore, for extracting features of the $SACF^k$ matrix from the initial (i.e., starting) or first row of $G^{k=1}$and the last (or ending) row of $G^{k=2}$is used, and likewise for extracting features from the last (or ending) row of $G^{k=2}$and the starting (or first) row of $G^{k=1}$ is used.

- **Inter-seasonal-pattern analysis:** It is noted that the values of $SACF^k$, k=1, 2 (= K) matrices were developed in the intra-seasonal-pattern analysis utilizing the simultaneous deviation, $Sdev_{i,j}^k$, (Eq. 3.20) where the differences in the grayscale intensity of each pixel with the corresponding expected grayscale intensity ($\mu_i^k$ and $\mu_j^k$) respectively of pixels in each row (same month) and each column (same year) of the season k were considered.

The dependence structure between successive occurrences of seasonal patterns of K (here, 2) seasons is assumed to follow the Markovian concept. Therefore, the intersection between the first row of the $SACF^k$ matrix of the k[th] season and the last row of the $SACF^{k-1}$ matrix of the previous season (k-1) (since only two seasons are considered, hence it will be ((k=1) -1= 0) season, which means k=2). In other words, the presence of Markovian dependence between successive seasonal patterns from season k=1, 2 (=K) indicates that seasonal patterns successively and regularly follow each other. However, what is desired is the magnitude (in terms of high or low) of grayscale intensity of the first pixel of a season (say k=1), given that the magnitude (in terms of high or low) of grayscale intensity of the last pixel of a previous season (say k=2, since only two seasons are considered for brevity). Because the magnitude of grayscale intensity of a pixel (either being the first or the last involves two states of the magnitude (high or low) and therefore would involve two transition matrices of size (2×2). In other words, one transition matrix between the last pixel of the seasonal pattern (say season, k=1) to the first pixel of the succeeding seasonal pattern (say season, k=2). Likewise, another transition matrix between the first pixel of the seasonal pattern (say season, k=2) to the first pixel of the succeeding seasonal pattern (say season, k=1). The two transition matrices stated above are formally defined as follows.

The first matrix $[p(g_{1,j}^2|g_{6,j}^1)]$ describes the transition from the last row of season k=1 to the first row of season k=2 (*Figure 3.10 a*) and the second matrix $[p(g_{1,j}^1|g_{6,j}^2)]$ describing transitioning from the last row of season k=2 to the first row of season k=1 (*Figure 3.10 b*). This transition, as stated earlier, involves the magnitude of grayscale intensity of the pixel, either being higher than or lower than the mean grayscale magnitude of the $\mu_{i,j}^k$ (i = row and j = month) as follows:

$$p(g_{1,j}^2|g_{6,j}^1) = \begin{bmatrix} p(g_{1,j}^2 < \mu_1^2 \mid g_{6,j}^1 < \mu_6^1) & p(g_{1,j}^2 \geq \mu_1^2 \mid g_{6,j}^1 < \mu_6^1) \\ p(g_{1,j}^2 < \mu_1^2 \mid g_{6,j}^1 \geq \mu_6^1) & p(g_{1,j}^2 \geq \mu_1^2 \mid g_{6,j}^1 \geq \mu_6^1) \end{bmatrix}_{2\times 2} \quad (3.16a)$$

$$\xrightarrow{P(a|b)=\frac{P(a\cap b)}{P(b)}} p\left(g_{1,j}^{2}\middle|g_{6,j}^{1}\right)=\begin{bmatrix}\frac{p(g_{1,j}^{2}<\mu_1^2\,\cap g_{6,j}^{1}<\mu_6^1)}{p(g_{6,j}^{1}<\mu_6^1)} & \frac{p(g_{1,j}^{2}\geq\mu_1^2\,\cap g_{6,j}^{1}<\mu_6^1)}{p(g_{6,j}^{1}<\mu_6^1)}\\ \frac{p(g_{1,j}^{2}<\mu_1^2\,\cap g_{6,j}^{1}\geq\mu_6^1)}{p(g_{6,j}^{1}\geq\mu_6^1)} & \frac{p(g_{1,j}^{2}\geq\mu_1^2\,\cap g_{6,j}^{1}\geq\mu_6^1)}{p(g_{6,j}^{1}\geq\mu_6^1)}\end{bmatrix}_{2\times2},$$

$$p\left(g_{1,j}^{1}\middle|g_{6,j-1}^{2}\right)=\begin{bmatrix}p\left(g_{1,j}^{1}<\mu_1^1\mid g_{6,j-1}^{2}<\mu_6^2\right) & p\left(g_{1,j}^{1}\geq\mu_1^1\middle|\, g_{6,j-1}^{2}<\mu_6^2\right)\\ p\left(g_{1,j}^{1}<\mu_1^1\mid g_{6,j-1}^{2}\geq\mu_6^2\right) & p\left(g_{1,j}^{1}\geq\mu_1^1\mid g_{6,j-1}^{2}\geq\mu_6^2\right)\end{bmatrix}_{2\times2} \tag{3.16b}$$

$$\xrightarrow{P(a|b)=\frac{P(a\cap b)}{P(b)}} p\left(g_{1,j}^{1}\middle|g_{6,j-1}^{2}\right)=\begin{bmatrix}\frac{p(g_{1,j}^{1}<\mu_1^1\,\cap g_{6,j-1}^{2}<\mu_6^2)}{p(g_{6,j-1}^{2}<\mu_6^2)} & \frac{p(g_{1,j}^{1}\geq\mu_1^1\,\cap\, g_{6,j-1}^{2}<\mu_6^2)}{p(g_{6,j-1}^{2}<\mu_6^2)}\\ \frac{p(g_{1,j}^{1}<\mu_1^1\,\cap g_{6,j-1}^{2}\geq\mu_6^2)}{p(g_{6,j-1}^{2}\geq\mu_6^2)} & \frac{p(g_{1,j}^{1}\geq\mu_1^1\,\cap\, g_{6,j-1}^{2}\geq\mu_6^2)}{p(g_{6,j-1}^{2}\geq\mu_6^2)}\end{bmatrix}_{2\times2}.$$

It is noted that since segmentation does not follow the calendar year, the month of January, when the new year starts, may either be in season (k=1) or season (k=2). Irrespective of where (or when) the month of January occurs, the j[th] seasonal pattern starts in a year and always ends in the following year. Therefore, while calculating the inter-seasonal-pattern relationship of the starting month of season (k=1), season (k-1 (here k=2) is shifted to the right by lag-1.

As shown in *Figure 3.10*, the Markovian probability ($p$) provides the dependence information in the vertical direction. In intra-seasonal-pattern analysis, since a 2D-approach (i.e., horizontal as well as vertical directions) has been considered, and likewise, it is expected to have the same 2D-approach in inter-seasonal-pattern analysis. Hence, there is a need for another Markovian probability ($\hat{p}$) to describe the horizontal transitioning in the starting row of textural images as depicted in *Figure 3.10*. The probability $\hat{p}$ determines the probability of transition from a textural pixel to its next textural pixel in the horizontal direction (i.e., the same row within the textural image of the same season $k$) and is expressed as $\hat{p}\ (g_{i,j}^{k}|\ g_{i,j-1}^{k})$:

$$\hat{p}\left(g_{1,j}^{1}\middle|g_{1,j-1}^{1}\right)=\begin{bmatrix}p(g_{1,j}^{1}<\mu_1^1\,|g_{1,j-1}^{1}<\mu_1^1) & p(g_{1,j}^{1}\geq\mu_1^1\,|g_{1,j-1}^{1}<\mu_1^1)\\ p(g_{1,j}^{1}<\mu_1^1\,|g_{1,j-1}^{1}\geq\mu_1^1) & p(g_{1,j}^{1}\geq\mu_1^1\,|g_{1,j-1}^{1}\geq\mu_1^1)\end{bmatrix}_{2\times2} \tag{3.17a}$$

$$\xrightarrow{P(a|b)=\frac{P(a\cap b)}{P(b)}} \hat{p}\left(g_{1,j}^{1}\middle|g_{1,j-1}^{1}\right)=\begin{bmatrix}\frac{p(g_{1,j}^{1}<\mu_1^1\,\cap g_{1,j-1}^{1}<\mu_1^1)}{p(g_{1,j-1}^{1}<\mu_1^1)} & \frac{p(g_{1,j}^{1}\geq\mu_1^1\,\cap\, g_{1,j-1}^{1}<\mu_1^1)}{p(g_{1,j-1}^{1}<\mu_1^1)}\\ \frac{p(g_{1,j}^{1}<\mu_1^1\,\cap g_{1,j-1}^{1}\geq\mu_1^1)}{p(g_{1,j-1}^{1}\geq\mu_1^1)} & \frac{p(g_{1,j}^{1}\geq\mu_1^1\,\cap\, g_{1,j-1}^{1}\geq\mu_1^1)}{p(g_{1,j-1}^{1}\geq\mu_1^1)}\end{bmatrix}_{2\times2},$$

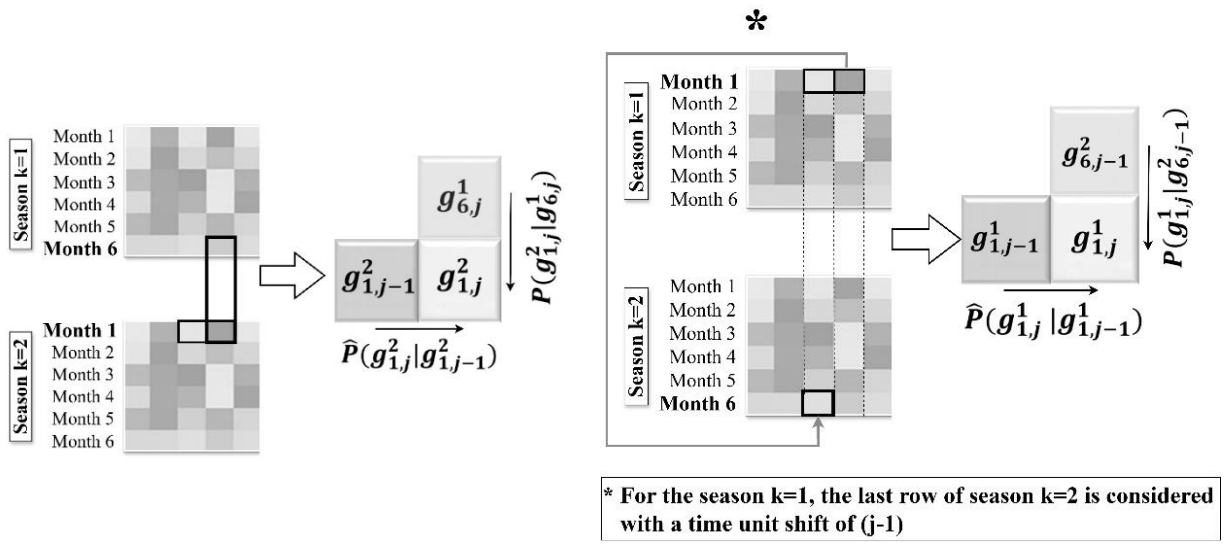
$$\hat{p}\left(g_{1,j}^{2}\middle|g_{1,j-1}^{2}\right)=\begin{bmatrix} p\left(g_{1,j}^{2}<\mu_1^2 \mid g_{1,j-1}^{2}<\mu_1^2\right) & p\left(g_{1,j}^{2}\geq\mu_1^2 \mid g_{1,j-1}^{2}<\mu_1^2\right) \\ p\left(g_{1,j}^{2}<\mu_1^2 \mid g_{1,j-1}^{2}\geq\mu_1^1\right) & p\left(g_{1,j}^{2}\geq\mu_1^2 \mid g_{1,j-1}^{2}\geq\mu_1^2\right) \end{bmatrix}_{2\times 2} \tag{3.17b}$$

$$\xrightarrow{P(a|b)=\frac{P(a\cap b)}{P(b)}} \hat{p}\left(g_{1,j}^{2}\middle|g_{1,j-1}^{2}\right)=\begin{bmatrix} \frac{p\left(g_{1,j}^{2}<\mu_1^2\cap g_{1,j-1}^{2}<\mu_1^2\right)}{p\left(g_{1,j-1}^{2}<\mu_1^2\right)} & \frac{p\left(g_{1,j}^{2}\geq\mu_1^2\cap g_{1,j-1}^{2}<\mu_1^2\right)}{p\left(g_{1,j-1}^{2}<\mu_1^2\right)} \\ \frac{p\left(g_{1,j}^{2}<\mu_1^2\cap g_{1,j-1}^{2}\geq\mu_1^2\right)}{p\left(g_{1,j-1}^{2}\geq\mu_1^2\right)} & \frac{p\left(g_{1,j}^{2}\geq\mu_1^2\cap g_{1,j-1}^{2}\geq\mu_1^2\right)}{p\left(g_{1,j-1}^{2}\geq\mu_1^2\right)} \end{bmatrix}_{2\times 2}.$$

In a Markovian process, where future states depend solely on the current state, the presence of correlation between successive observations is inherent. Mathematical relationships involving correlations, projections, and spaces related to past innovations in a system were formally expounded by Akaike (1975). Simultaneous Markovian Correlation ($SM$C) can, therefore, be defined as the correlation between textural pixel(s) ($g_{i,j}^{k}$) located in the starting row of season k with the textural image of the previous season ($k-1$), as shown in *Figure 3.10*a, where the correlation depends on two independent transition probabilities (p) between seasons $k$ and ($k-1$), and $\hat{p}$ within the season $k$. Hence, the elements of $SMC_{i,j}^{k}$are obtained as follows:

$$SMC_{i,j}^{k}=\mathrm{p}\left(g_{i,j}^{k}\middle|g_{i,j}^{k-1}\right)\times\hat{p}\left(g_{i,j}^{k}\middle|g_{i,j-1}^{k}\right) \text{ for i=1; j=1,2,...,N; and k=1, 2(=K),} \tag{3.18}$$

which results in ($1\times N$) dimensions and will be used to restore the elements of the missing row in the $SACF^{k}$.




*(a)* *(b)*

*Figure 3.10. Inter-seasonal pattern analysis: the process of computing values of SMC for (a) season k=2 ($SMC^2_{1,j}$) and (b) season k=1 ($SMC^1_{1,j}$).*

- **Intra-inter-seasonal pattern analysis:** Upon establishing both the intra-seasonal pattern and inter-seasonal pattern relationships between seasonal grayscale images to obtain $SMCs$ for both seasons k=1 and k=2, the first row in $SACF^1_{1,j}$ and $SACF^2_{1,j}$ will respectively be attained by $SMC^1_{1,j}$ and $SMC^2_{1,j}$ ($SACF^k_{1,j} = SMC^k_{1,j}$). In other words, the dimension of $SACF$ has been regained to its original size of $6 \times N$ and is ready for use by the textural image feature extraction model. For brevity, the feature extraction model is explained in the section to follow:

$$\xrightarrow{SACF^k_{1,j}=SMC^k_{1,j}} SACF^k = \begin{bmatrix} SACF^k_{1,1} & SACF^k_{1,2} & \cdots & SACF^k_{1,N} \\ SACF^k_{2,1} & SACF^k_{2,2} & \cdots & SACF^k_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ SACF^k_{6,1} & SACF^k_{6,2} & \cdots & SACF^k_{6,N} \end{bmatrix}_{6\times N}.$$

To extract features of any seasonal textural image, the matrix of $SACF^k$ of that specific season is used by incorporating the Fourier analysis to enhance the extraction of features. Additionally, the Fourier analysis can provide a unique and complementary perspective when combined with $SACF^k$, thus making it a valuable tool for textural feature extraction in image analysis (Gonzalez, 2009). The Fourier analysis also provides information about the frequency components of a signal

and hence can be used to compute the square of a power spectrum to represent the energy associated with each frequency component.

In addition, the Discrete Fourier Transform (DFT) is a powerful tool for analyzing the frequency content of signals (Chong et al., 2019) or images (Tuceryan & Jain, 1993; Gonzalez, 2009). By computing the Fourier transform of $SACF^k$, the frequency components that represent the underlying textural features of an image can be extracted. By applying the DFT to the $SACF^k$, one can extract the frequency components that contribute to the autocorrelation of the seasonal textural image, and in turn, this information can be used to synthesize a seasonal textural image. Since $SACF^k$ is a 2D-matrix, a 2D-Fourier-transform is accomplished. The 2D-Fourier-transform of a matrix $F(u, v)$ can be computed using the expression (Gonzalez, 2009) as follows:

$$F^k(u,v) = \sum_{i=0}^{6-1} \sum_{j=0}^{N-1} SACF_{i,j}^k \, e^{-2\pi(\frac{u_i}{6} + \frac{v_j}{N})}, \tag{3.19}$$

where $F^k(u,v)$ is a 2D-Fourier coefficient matrix of the k[th] season, and $u$ and $v$ are respectively the frequency components in horizontal and vertical directions. A general form of the output of $F^k(u,v)$ is a complex number of the form $(a + bj)$, where $a$ and $b$ are real numbers, and $j$ is an imaginary unit $(\sqrt{(-1)})$. For a given frequency, the real part $'a'$ represents the magnitude (or amplitude) of the cosine component, and the imaginary part $'b'$ represents the amplitude of the sine component. Once $F^k(u,v)$ has been obtained, the power spectrum can be computed as the features of the seasonal textural image. A power spectrum is a tool for understanding the energy distribution of an image in the frequency domain and allows for the extraction of features of the seasonal textural image, which can be obtained by computing the squared magnitude of the Fourier coefficients matrix. Mathematically, $P\left(F^k(u,v)\right) = |F^k(u,v)^2|$ is defined as the power spectrum $(P^k)$ of the k[th] season.

The extracted textural image features $(P^k)$ can then be used to synthesize a new seasonal textural image with similar frequency characteristics (Tuceryan & Jain, 1993). The DFT is employed because the autocorrelation function of an image is directly related to the power spectrum obtained via the Fourier transform, as shown in *Figure 3.11*.

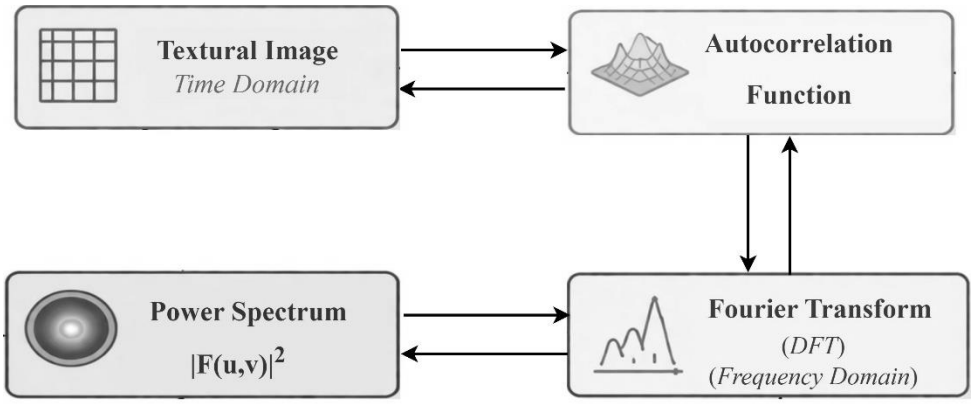



*Figure 3.11. Interrelationship between the textural image, extracted time-domain features (simultaneous autocorrelation), and transformed frequency-domain features (power spectrum) obtained using the Fourier transform.*

Once the power spectrum is obtained, it is divided into high-power and low-power components using a cut-off point defined as the mean ($\overline{P^k}$) and is expressed as follows.

$$\text{For High-Power Component } (PH_{i,j}^k): \begin{cases} if\ P_{i,j}^k \geq \overline{P^k}\ then\ PH_{i,j}^k = P_{i,j}^k, \\ \quad Otherwise\ PH_{i,j}^k = \emptyset. \end{cases} \tag{3.20}$$

$$\text{For Low-Power Component } (PL_{i,j}^k): \begin{cases} if\ P_{i,j}^k < \overline{P^k}\ then\ PL_{i,j}^k = P_{i,j}^k, \\ \quad Otherwise\ PL_{i,j}^k = \emptyset. \end{cases} \tag{3.21}$$

It is apparent from Eqs. 3.20 and 3.21 that $PL_{i,j}^k$ and $PH_{i,j}^k$ together constitute the entire set of power components in a row (month), indicating the number of years ($PL_{i,j}^k + PH_{i,j}^k = N$). For each group ($PH^k\ and\ PL^k$), a reference vector ($R^k$) is then selected as a representative. A reference vector ($R_{PH}^k and\ R_{PL}^k$) for the High-power group and Low-power group, respectively, in the $i^{th}$ month will be obtained using Eqs. 3.22a and 3.22b.

$$R_{PH}^k = \frac{1}{N_{PH_i}} \sum_{j=1}^{N_{PH_i}} PH_{i,j}^k \tag{3.22a}$$

$$= [\frac{1}{N_{PH_1}} \sum_{j=1}^{N_{PH_1}} PH_{1,j}^k\ , \frac{1}{N_{PH_2}} \sum_{j=1}^{N_{PH_2}} PH_{2,j}^k\ , \dots , \frac{1}{N_{PH_6}} \sum_{j=1}^{N_{PH_6}} PH_{6,j}^k\ ]^T,$$

$$R_{PL}^k = \frac{1}{N_{PL_i}} \sum_{j=1}^{N_{PL_i}} PL_{i,j}^k \tag{3.22b}$$

$$= [\frac{1}{N_{PL_1}} \sum_{j=1}^{N_{PL_1}} PL_{1,j}^k \ , \frac{1}{N_{PL_2}} \sum_{j=1}^{N_{PL_2}} PL_{2,j}^k \ , \ldots , \frac{1}{N_{PL_6}} \sum_{j=1}^{N_{PL_6}} PL_{6,j}^k \ ]^T,$$

where $N_{PH_i}$ and $N_{PL_i}$ corresponds to the number occurrences falling into the high-power-spectrum group (hereafter referred to as HPS-group) and low-power-spectrum group (hereafter referred to as LPS-group), respectively, for each i[th] month with $1 \leq \ i \leq 6$. The flow diagram for extracting a reference vector from a power-spectrum in the k[th] season is displayed in *Figure 3.12*. The division of a power spectrum matrix into high and low power components is based on a cutting point defined as the average-power-spectrum that helps distinguish between frequency components with higher and lower contributions to the overall variability or strength within the image. Components of the high-power spectrum often correspond to the more pronounced patterns, structures, or variations, while components of the low-power spectrum might correspond to less pronounced or subtle patterns, noise, or minor variations.

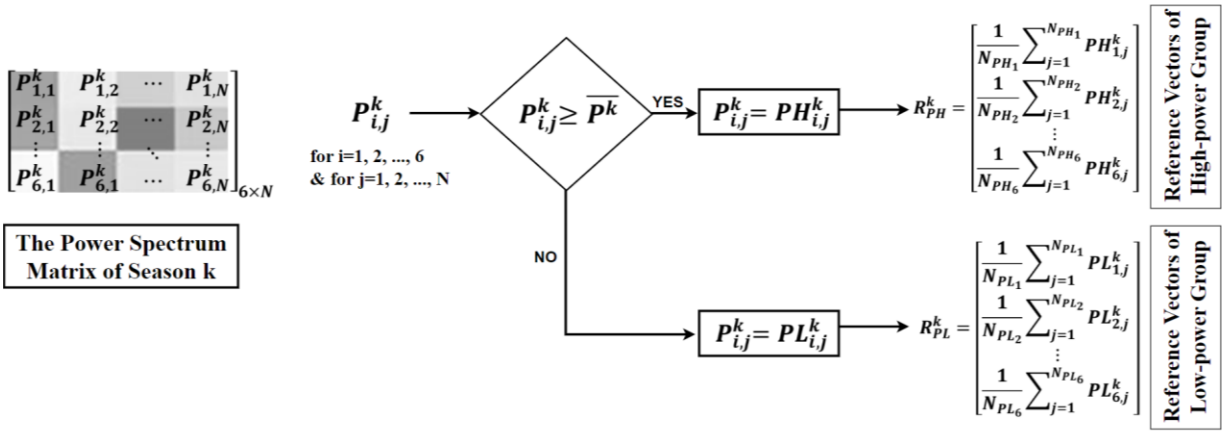



*Figure 3.12. A conceptual diagram exhibiting the process of computing the reference vector of the k[th] season in an encoded textural feature recognition system.*

The reference vectors serve to confirm whether the synthesized values of the power spectrum belong to high and low power groups. This process is further explained in the next section.

The approach of separating the power spectrum into high- and low-power components is effective. However, using the mean as the threshold may not be suitable for all cases, particularly for rivers with highly skewed flow distributions. Exploring alternative thresholding methods could improve the flexibility and generalizability of the model. To address this issue, a statistical analysis needs

to be conducted. The median or percentile-based thresholds could serve as suitable alternatives in cases where the power spectrum distribution is highly skewed.

It should be noted that physically, each grayscale intensity in the figure represents the magnitude of streamflow, with brighter pixels corresponding to higher values. Each SACF reflects the correlation of a given time step with its immediate past (previous month) and with the same month in the previous year. A high SACF value indicates strong correlation; however, the method and model used do not distinguish whether this correlation arises from the immediate past or the same month in the previous year. These SACF features were then transformed into the frequency domain using the discrete Fourier transform, and the corresponding power spectra were computed. While in image analysis, the power spectrum is interpreted as the energy of an image, in this research, it represents the frequency-domain representation of the SACF, providing a quantitative measure of the strength of correlations at different temporal scales. In other words, the power spectrum shows whether streamflow is more influenced by short-term memory (last month) or long-term/seasonal memory (same month in the previous year), rather than representing energy in the conventional hydrological sense.

#### 3.2.2.3 Synthesis of Streamflow Encoded Grayscale Seasonal Textural Images

To form comparable textural features within the synthesized seasonal grayscale textual images, it is essential to generate random complex numbers in the form of an identical size $(6 \times N)$ matrix for each season, along with a matching power spectrum distribution. For this purpose, two separate generators have been envisioned to generate the magnitude and phase of complex numbers. To avoid loss of the phase information, which otherwise would lead to unattainable synthesis of textural images, it is essential to use two generators to ensure that the process of taking the square root of a power spectrum does not result in a complex number. Therefore, it is necessary to use a Gaussian distribution with the probability of occurrence of either the HPS-group or LPS-group in each row of a season to ensure that a synthesized power spectrum is generated.

The probability of occurrence for the HPS-group is calculated as follows:

$$P\left(x \geq \overline{P^k}\right) = \left[\frac{N_{PH_i}}{N}\right];\ \text{for } i = 1, 2, \ldots, 6$$
$$= \left[\frac{N_{PH_1}}{N},\ \frac{N_{PH_2}}{N},\ \ldots\ , \frac{N_{PH_6}}{N}\right]^T. \tag{3.23}$$

For any given i[th] row (= month), $\left(N_{PH_i} + N_{PL_i}\right)$ = N, and it can also be concluded that $\left[P\left(x \geq \overline{P^k}\right) + P\left(x < \overline{P^k}\right)\right] = 1$; therefore, the probability of occurrence for the HPS-group in the i[th] month can be obtained as follows:

$$p\left(x < \overline{P^k}\right) = 1 - P\left(x \geq \overline{P^k}\right) = \left[N_{PL_i}/N\right] = [1 - (N_{PH_i}/N)]; \text{ for } i = 1,2,\ldots,6. \quad (3.24)$$

For any pixel (i,j), the process of image synthesis starts with the generation of a random value ($r_{i,j}$) between 0 and 1 from a uniform distribution. For a random value thus generated to be less than the probability of occurrence of the LPS-group ($(N_{PL_i}/N)$ obtained from Eq. 3.22b, a Gaussian distribution with mean and standard deviation corresponding to the LPS-group is used to generate a synthesized power spectrum $(P^k)'$. Otherwise, a Gaussian distribution with mean and standard deviation corresponding to the HPS-group is invoked to obtain a value of $(P^k)'$.

For a value of $(P^k)'$ belonging to the HPS-group, then it is paired with a reference vector ($R_{PH}^k$) (obtained from Eq. 3.22a) of the HPS-group, and it is expected to lie within $R_{PH}^k \pm C \times St_{PH}^k$, where $St_{PH}^k$ is the standard deviation of the HPS-group with C=1.96 at the 95% confidence level for a data set that follows a Gaussian distribution. Otherwise, it is paired with a reference vector ($R_{PL}^k$) (obtained from Eq. 3.22b) of the LPS-group, and it is expected to lie within $R_{PL}^k \pm C \times St_{PL}^k$, where $St_{PL}^k$ is the standard deviation of the LPS-group. The process of randomly generating a synthesized power spectrum $(P^k)'$ including $(PH^k)'$ and $(PL^k)'$is illustrated below in *Figure 3.13*. Once the output of the first generator is accepted, the second generator gets activated.

In *Figure 3.13*, the subsequent phase involves deriving the $(F^k(u, v))'$ matrix of synthesized Fourier coefficients. This is achieved through the multiplication of the square root of the synthesized power spectrum (either $\sqrt{(PH^k)'}$ or $\sqrt{(PL^k)'}$ ) by a Gaussian distribution, incorporating the exponential of the generated sine amplitude ($\sqrt{(P^k)'} \times e^{1i \times b'}$).

Given that the Fourier coefficient matrix has the form of $(a + bj)$, the synthesized Fourier coefficient matrix must have the same format as a complex number. However, the root square of the synthesized power spectrum $\sqrt{(P^k)'}$ will provide the magnitude (or amplitude) of the cosine component 'a', and therefore, the amplitude of the sine component 'b' will be lost. To avoid this situation, there must be two independent generators. The first generator (Generator 1.1 or Generator 1.2) generates the cosine components, and the sine components are generated by the second generator. It is noted that, in the analysis of periodic signals, the dominant part of a signal

is characterized by the cosine function (Ersoy, 1994) because the contribution of the sine function is relatively smaller. Hence, only the output of the first generator is tested to determine if it belongs to the reference vector.

The inverse DFT of the synthesized Fourier coefficients matrix is then developed as follows:

$$(SACF_{i,j}^{k})' = Inverse\ DFT\ ((F^{k}\ (u,v))'\ ) \tag{3.25}$$
$$= \frac{1}{6\times N}\sum_{i=1}^{5}\sum_{j=1}^{N-1}(F^{k}\ (\mathrm{u},\mathrm{v}))'\ e^{-2\pi(\frac{ui}{6}+\frac{Vj}{N})}.$$

To obtain the intensity values of each pixel, $(SACF_{i,j}^{k})'$ must be transformed to the grayscale image intensity as follows:

$$(SACF_{i,j}^{k})' = \frac{(g_{i,j}^{k\,\prime}-\mu_j^k)(g_{i,j}^{k\,\prime}-\mu_i^k)(g_{i-1,j}^{k\,\prime}-\mu_j^k)(g_{i,j-1}^{k\,\prime}-\mu_i^k)}{(S_i^k)^2(S_j^k)^2}. \tag{3.26}$$

Therefore, the numerator is expressed as follows:

$$\left(g_{i,j}^{k\,\prime}-\mu_j^k\right)\left(g_{i,j}^{k\,\prime}-\mu_i^k\right)\left(g_{i-1,j}^{k\,\prime}-\mu_j^k\right)\left(g_{i,j-1}^{k\,\prime}-\mu_i^k\right) = (SACF_{i,j}^{k})' \times (S_i^k)^2(S_j^k)^2. \tag{3.27}$$

Solving Eq. 3.27 for $g_{i,j}^{k\,\prime}$, one proceeds as follows:

$$\left(g_{i,j}^{k\,\prime}-\mu_i^k\right)\left(g_{i,j}^{k\,\prime}-\mu_j^k\right) = \frac{(SACF_{i,j}^{k})' \times (S_i^k)^2(S_j^k)^2}{\left(g_{i-1,j}^{k\,\prime}-\mu_j^k\right)\left(g_{i,j-1}^{k\,\prime}-\mu_i^k\right)}, \tag{3.28}$$

$$(g_{i,j}^{k\,\prime})^2 - \left(\mu_i^k+\mu_j^k\right)g_{i,j}^{k\,\prime} + [\mu_i^k\times\mu_j^k - \frac{(SACF_{i,j}^{k})'\times (S_i^k)^2(S_j^k)^2}{\left(g_{i-1,j}^{k\,\prime}-\mu_j^k\right)\left(g_{i,j-1}^{k\,\prime}-\mu_i^k\right)}] = 0. \tag{3.29}$$

Eq. 3.29 is a quadratic equation in its classic form of $(ax^2 + bx\ + c = 0)$ with two roots. To solve a quadratic equation with two roots, the quadratic formula is used, which is given by $x = -b \pm \sqrt{b^2 - 4ac}/2a$. The ± sign indicates that there are two possible roots (or solutions), one with a plus sign and another with a minus sign. By expressing the given information in a general form of the quadratic equation, the following relationship is obtained.

$$g_{i,j}^{k\,\prime} = \frac{-(-\left(\mu_i^k+\mu_j^k\right))\pm\sqrt{\left(\mu_i^k+\mu_j^k\right)^2-4\times1\times[\mu_i^k\times\mu_j^k-\frac{(SACF_{i,j}^{k})'\times(S_i^k)^2(S_j^k)^2}{\left(g_{i-1,j}^{k\,\prime}-\mu_j^k\right)\left(g_{i,j-1}^{k\,\prime}-\mu_i^k\right)}]}}{2\times1}. \tag{3.30}$$

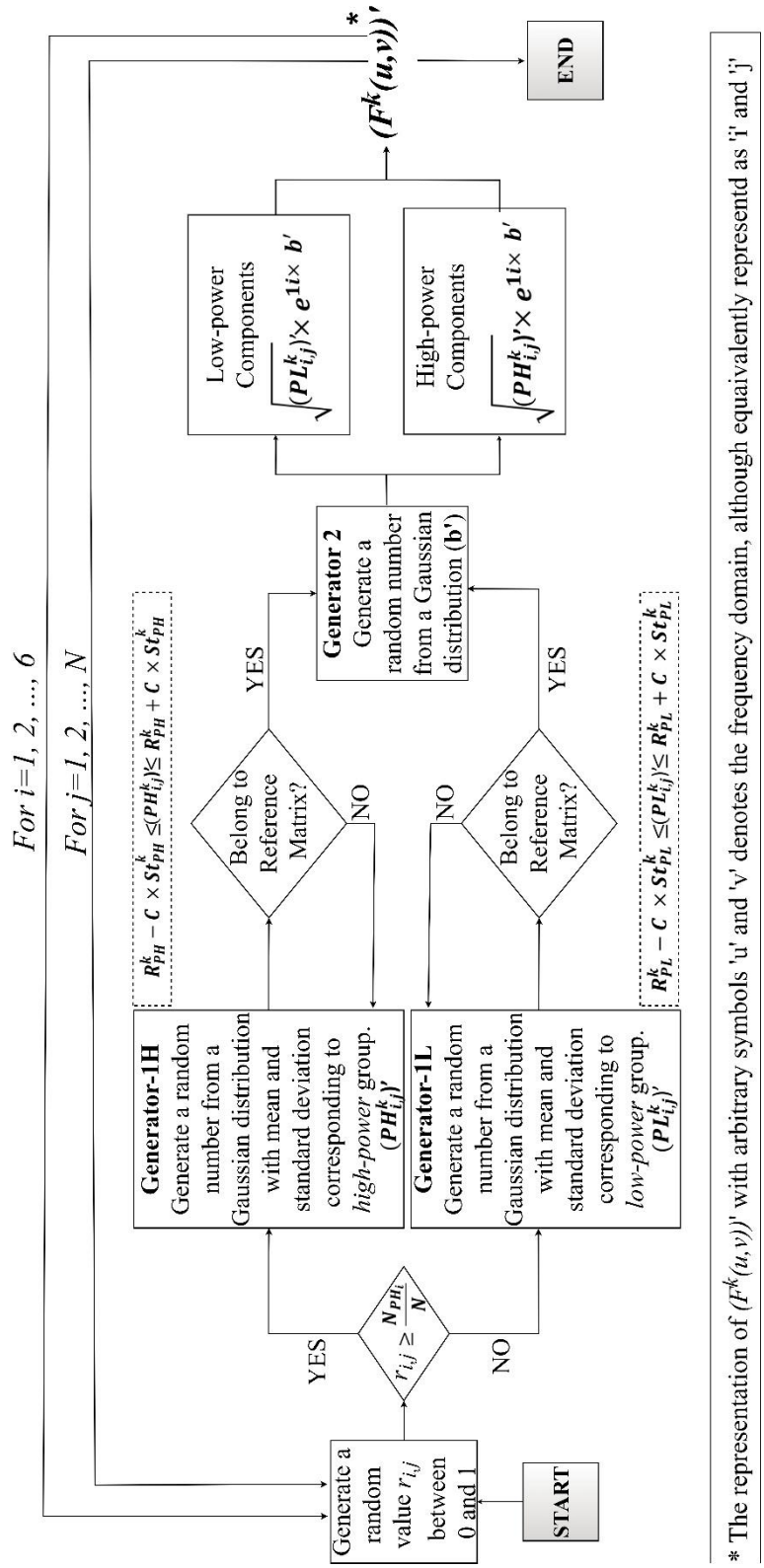



*Figure 3.13. A logical flow diagram for synthesizing a textural image in season k using an encoded textural feature recognition system.*

Since streamflow synthesis is the goal, a new constraint has been identified when the original image contains a pixel value that exceeds the mean intensity of all pixels within seasonal textural images of the historical streamflow data; then the root with a positive sign is selected; otherwise, the root with a negative sign is chosen. Such an approach ensures that synthesized streamflow values mimic and likewise remain relevant to the historical data, while also considering the overall intensity of seasonal textural images of the historical streamflow time series.

Once the synthesis process of textural image(s) corresponding to a realization of the streamflow time series is complete, then an important, vital and final step involves the back transformation and sequential arrangement of the resulting collages of seasonal textural images into a realization of the streamflow time series using the following formula:

$$\frac{g_{i,j}^{k\,\prime}}{255} \times [\max(x_{i,j}^{k}) - \min(x_{i,j}^{k})] + \min\,(x_{i,j}^{k}), \tag{3.31}$$

where $\min(x_{i,j}^{k})$ and $\max(x_{i,j}^{k})$ respectively represent the minimum and maximum values of the historical streamflow record for each season. Once the synthesized images $(G^{k})'$, for all seasons that have been generated, they are concatenated together sequentially, with $(G^{k=1})'$ followed by $(G^{k=2})'$, and so on.

A more general model incorporating additional lags is provided in Appendix 3B.

### 3.2.3 Assessing Model Performance

Encoding the streamflow time series into grayscale textural images can best be assessed using the Structural Similarity (SSIM) index between the synthetic and original time series. In addition, hypothesis tests were conducted to assess the comparability of the statistical properties of synthetic and historical streamflow.

#### 3.2.3.1 Structural Similarity (SSIM) Index

The Structural Similarity (SSIM) index measures the similarity between two images. It combines three components: luminance distortion, contrast distortion, and correlation to produce a single index. The first two components account for non-structural distortion, while the third component

accounts for structural distortion (i.e., the absence of correlation) in the image. By choosing certain parameters, the SSIM index can be expressed in a simplified form as follows:

$$SSIM(x,y) = [l(x,y) \times c(x,y) \times s(x,y)], \tag{3.32}$$

where, $l(x,y) = [\frac{2\mu_x\mu_y+C_1}{{\mu_x}^2+{\mu_y}^2+C_1}]$; $c(x,y) = [\frac{2\sigma_x\sigma_y+C_2}{{\sigma_x}^2+{\sigma_y}^2+C_2}]$; and $s(x,y) = [\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}]$.

In these expressions, $x$ and $y$ are the two images being compared, respectively, the original encoded image of streamflow, and the synthesized encoded image. The $\mu_x$ and $\mu_y$ are the local mean values of the respective images, the ${\sigma_x}^2$ and ${\sigma_y}^2$ are the local variances, the $\sigma_{xy}$ is the local covariance, and $C_1$, $C_2$, and $C_3$ are small constants used for numerical stability when the denominators are close to zero. The terms $l(x,y)$, $c(x,y)$, and $s(x,y)$, respectively, represent the luminance distortion, the contrast distortion, and structural distortion (absence of correlation). The SSIM index is typically computed on a local window and produces an SSIM map, and a single index is obtained by pooling together all computed local errors (Wang et al., 2004).

#### 3.2.3.2 Assessment of Time Series Similarity in Terms of Statistical Properties

The historical and synthesized time series can be compared by examining their relevant statistical parameters, such as the mean, standard deviation, skewness, and kurtosis. These parameters provide a quantitative summary of the time series properties. To determine the significance of differences between historical and synthesized time series, hypothesis testing, such as a t-test or an ANOVA, can be performed.

Although a comparison of relevant statistical parameters can provide some insights into the similarities or differences between the observed and synthesized time series, it may not capture the underlying patterns or dynamics of a streamflow time series. On the other hand, using entropy can help to capture the complexity of a time series because it measures the amount of uncertainty in a time series. In turn, such information can provide a measure of the degree of similarity or dissimilarity between the historical and synthesized time series.

The permutation entropy is widely used for data analysis in many fields. It refers to the Shannon entropy of the distribution of ordinal patterns within a time series. It is based on the concept of representing a time series as a sequence of ordinal patterns, which essentially implies the orderings of values in non-overlapping windows of the time series (Bandt & Pompe, 2002). In other words, the permutation entropy measures the degree of difficulty in the prediction of the next value in a

time series, $\{X_t; 1 \leq t \leq T\}$ based on previous values and can be computed for order $n \geq 2$ as follows:

$$H(n) = -\sum p(\pi) \log p(\pi), \tag{3.33}$$

where the relative frequency number (#) for each $\pi$ is determined as follows:

$$p(\pi) = \#\{t/(t \leq T - n, (X_{t+1} + \cdots + X_{t+n}))\ has\ type\ \pi\}/(T - n + 1). \tag{3.34}$$

Permutation entropy specifically is used as a complexity measure that quantifies the amount of regularity or predictability in a time series. Therefore, it is useful for assessing the significance of pattern recognition in time series data.

### 3.2.4 Interpretability of Textural Pattern Recognition Approach

As discussed in Section 2.2.5, machine learning approaches, particularly deep learning models, often lack interpretability due to their reliance on implicitly learned features. Also, developing an interpretable modelling approach is one of the primary objectives of this research (as discussed in section 2.3).

The proposed approach offers improved interpretability for two main reasons. First, it is based on explicitly defined and physically meaningful features. Specifically, the extracted textural features represent temporal dependency structures, providing clear insight into the underlying streamflow dynamics, unlike methods such as convolutional neural networks (CNNs), where features are learned implicitly and are often difficult to interpret.

Second, the user retains control over key modelling steps. For example, the selection of the threshold (cut-off point) in the power spectrum can be adjusted based on the statistical characteristics of the data; if the distribution is not well represented by the mean, alternative measures such as the median may be used. Additionally, the model formulation allows for controlled selection between multiple mathematical solutions (e.g., positive or negative roots).

Overall, these transparent and user-guided steps enhance the interpretability of the proposed framework by enabling direct understanding and adjustment of the model components.

## 3.3 Concluding Remark

This chapter established a unified framework for streamflow synthesis that explicitly accounts for the dominant noise characteristics, persistence, and periodic structures inherent in hydrological time series. By classifying basins according to their spectral behaviour, ranging from pink- and black-noise dominance to less common brown-noise regimes, the chapter demonstrated that streamflow synthesis models must be adapted to reflect basin-specific memory, storage, and response dynamics. The integration of noise-based approaches, including ARIMA and ARFIMA models, with pattern recognition techniques enables the preservation of both stochastic persistence and organized temporal structures such as seasonality and multi-scale periodicity. Collectively, this framework provides a physically interpretable and flexible basis for synthetic streamflow generation, ensuring consistency with observed hydrologic behaviour across diverse basin types and laying the foundation for the methodological developments presented in subsequent Chapter 4.

# Supplementary Material for Chapter 3:

## Appendix 3A: Calculation of Hurst Coefficient

Below is the procedure used to calculate the Hurst coefficient, which quantifies the long-term persistence and scaling behaviour of the time series analyzed in this study.

**Step-1: Segmentation:** The time series of y[n] is broken up into equally sized, non-overlapping subseries of length s, denoted as:

$$Y_s = \{y_{s_m}[n] \mid 1 \leq m \leq N_s\} \tag{3A.35}$$

where y[n] is the original time series value at time step $n$ (streamflow in m³/s), $s$ is the length of each segment in number of time steps, $m$ is the segment index, $N_s$ is the total number of segments, and $y_{s_m}[n]$ is the $k^{\text{th}}$ segment of $y[n]$, containing $'s'$ consecutive values.

**Step-2: Range Calculation:** For each segment, $m$, of length, s, the range $R_{m,s}$, which is the difference between the maximum and minimum values within the segment being calculated:

$$R_{m,s} = \max[y_{s_m}] - \min[y_{s_m}] \tag{3A.36}$$

**Step-3: Variance Calculation:** The variance, $S_{m,s}$ of the original time series values within each segment, m, of length, s, is computed by:

$$S_{m,s} = \sigma_x[y_{s_m}]$$
(3A.37)

**Step-4: Mean Values:** The mean values of the range, $R_s$, and the standard deviation, $S_s$ across all segments are determined as follows:

$$\bar{R}_s = \frac{1}{N_s}\sum_{m=1}^{N_s} R_{m,s}$$
(3A.38)

$$\bar{S}_s = \frac{1}{N_s}\sum_{m=1}^{N_s} S_{m,s}$$
(3A.39)

**Step-5: Ratio and Scaling:** The ratio $\bar{R}_s/\bar{S}_s$ exhibits a power-law scaling as a function of segment length, s, which can be described by the equation 3A.6:

$$\frac{\bar{R}_s}{\bar{S}_s} = S^H \quad (3A.40)$$

where $H$ is the Hurst exponent.

## Appendix 3B: Model Development for Additional Lags

Considering $\alpha$ lags in the horizontal and $\beta$ lags in the vertical direction, Eq. 2 for further lags will be as follows.

$$Sdev_{i,j}^{k} = \prod_{m=1}^{\alpha}[(g_{i,j}^{k} - \mu_i^k)(g_{i,j-m}^{k} - \mu_i^k)] \times \prod_{n=1}^{\beta}[(g_{i,j}^{k} - \mu_j^k)(g_{i-n,j}^{k} - \mu_j^k)] \quad (3B\text{-}1)$$

The matrix of $(SACF_{i,j}^{k})$-which is obtained using Eq. 3, has $\beta$ rows and $\alpha$ columns fewer compared to the matrix $(G^k)$. To reconstruct the unavailable information, the length of N years should be adjusted to $(N+\alpha)$ additional years, and in the inter-seasonal analysis, $\beta$ of last rows of the previous season are utilized.

In the synthesis stage, $(SACF_{i,j}^{k})'$ is transformed to the grayscale image intensity as follows:

$$(SACF_{i,j}^{k})' = \frac{\prod_{m=1}^{\alpha}[(g_{i,j}^{k\,\prime} - \mu_i^k)(g_{i,j-m}^{k\,\prime} - \mu_i^k)] \times \prod_{n=1}^{\beta}[(g_{i,j}^{k\,\prime} - \mu_j^k)(g_{i-n,j}^{k\,\prime} - \mu_j^k)]}{(S_i^k)^2 (S_j^k)^2} \quad (3B\text{-}2)$$

the gray intensity value of each pixel $(g_{i,j}^{k\,\prime})$ is calculated considering various scenarios for $\alpha$ and $\beta$ as the number of lags determines the degree of the Equation.

**Scenario 1**: if $\alpha = \beta = 1$: already explained in the manuscript.

**Scenario 2**: if $\alpha + \beta = 3$ (either $\alpha = 1, \beta = 2$ or $\alpha = 2, \beta = 1$)

$$\left(g_{i,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i-1,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i,j}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j-1}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j-2}^{k\,\prime} - \mu_i^k\right) = (SACF_{i,j}^{k})' \times (S_i^k)^2 (S_j^k)^2 \quad (3B\text{-}3)$$

$$\left(g_{i,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i,j}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j}^{k\,\prime} - \mu_i^k\right) = \frac{(SACF_{i,j}^{k})' \times (S_i^k)^2 (S_j^k)^2}{\left(g_{i-1,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i,j-1}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j-2}^{k\,\prime} - \mu_i^k\right)} \quad (3B\text{-}4)$$

The left side of Eq. 3B-4 is:

$$\left(g_{i,j}^{k\,\prime} - \mu_j^k\right)((g_{i,j}^{k\,\prime})^2 + \left(\mu_i^k\right)^2 - 2g_{i,j}^{k\,\prime}\mu_i^k) \quad (3B\text{-}5)$$

Which expands to:

$$(g_{i,j}^{k\,\prime})^3 - g_{i,j}^{k\,\prime}\left(\mu_i^k\right)^2 - 2(g_{i,j}^{k\,\prime})^2\mu_i^k - (g_{i,j}^{k\,\prime})^2\mu_j^k - \mu_j^k\left(\mu_i^k\right)^2 - 2g_{i,j}^{k\,\prime}\mu_i^k\mu_j^k \quad (3B\text{-}6)$$

And therefore, the whole Eq. 3B-4 is:

$$(g_{i,j}^{k\,\prime})^3 - \left(2\mu_i^k + \mu_j^k\right)(g_{i,j}^{k\,\prime})^2 - \left(\left(\mu_i^k\right)^2 + 2\mu_i^k\mu_j^k\right)g_{i,j}^{k\,\prime} - \frac{(SACF_{i,j}^{k})' \times (S_i^k)^2 (S_j^k)^2}{\left(g_{i-1,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i,j-1}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j-2}^{k\,\prime} - \mu_i^k\right)} - \mu_j^k\left(\mu_i^k\right)^2 = 0 \quad (3B\text{-}7)$$

Which has the form of a cubic equation ($ax^3+bx^2 + cx + d = 0$). This equation can be solved using Numerical Methods such as Newton's Method.

**Scenario 3**: if $\alpha + \beta = 4$ ($\alpha = \beta = 2$)

$$\left(g_{i,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i-1,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i-2,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i,j}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j-1}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j-2}^{k\,\prime} - \mu_i^k\right) = (SACF_{i,j}^k)' \times (S_i^k)^2 (S_j^k)^2 \quad \text{(3B-8)}$$

$$\left(g_{i,j}^{k\,\prime} - \mu_i^k\right)^2 \left(g_{i,j}^{k\,\prime} - \mu_j^k\right)^2 = \frac{(SACF_{i,j}^k)' \times (S_i^k)^2 (S_j^k)^2}{\left(g_{i-1,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i-2,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i,j-1}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j-2}^{k\,\prime} - \mu_i^k\right)} \quad \text{(3B-9)}$$

$$\left(g_{i,j}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j}^{k\,\prime} - \mu_j^k\right) = \pm \sqrt{\frac{(SACF_{i,j}^k)' \times (S_i^k)^2 (S_j^k)^2}{\left(g_{i-1,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i-2,j}^{k\,\prime} - \mu_j^k\right)\left(g_{i,j-1}^{k\,\prime} - \mu_i^k\right)\left(g_{i,j-2}^{k\,\prime} - \mu_i^k\right)}} \quad \text{(3B-10)}$$

The left-hand side is the same as in Eq. 3.29, but with a different right-hand side. Therefore, it should be solved using the quadratic equation.

**Scenario 4:** if $\alpha + \beta \geq 5$

This generates an equation of degree 5 or higher, which **requires numerical methods** for solving.

# 4. Results and Discussion

This chapter presents the results of applications of the Textural Feature Pattern Recognition System developed as a framework for streamflow synthesis and discusses their implications across different noise regimes and basin characteristics. The analysis begins with noise recognition for all streamflow stations in Ontario. Appendix 4A provides a summary of relevant information on hydrometric stations used for noise recognition analysis.

Following this regional assessment, brown-noise-dominated systems are examined in detail by presenting results from two representative stations, highlighting their unique persistence and long-memory behaviour. For pink- and black-noise-dominated basins, ARIMA-based models were fitted, and the resulting autoregressive structures are analyzed to assess persistence, memory, and model performance for comparative purposes with the TPRS developed in this thesis. This analysis also serves to confirm the appropriateness of the methodology used to estimate the scaling exponent λ, as explained in Section 3.1.3.

To evaluate the capability of the proposed textural pattern recognition approach, three stations with the dominance of black noise and three stations with the dominance of pink noise were randomly selected for synthetic streamflow generation. The synthesized streamflow time series obtained using the textural pattern recognition method are then compared against those synthesized by ARIMA models. This comparison is used to demonstrate the effectiveness of the textural pattern recognition framework developed in this thesis in preserving both persistence and seasonal structure.

In addition, six streamflow stations were selected for synthetic streamflow generation, as identified by Panu et al. (1978). This comparative analysis provides a benchmark for evaluating the performance of the TPRS developed in this thesis relative to established pattern-based synthesis techniques.

## 4.1 Study Area Description

The streamflow stations analyzed in this thesis are divided into two groups: (i) stations in Ontario that were used to identify and analyze coloured noise characteristics, as summarized in Table 4A.1, and (ii) stations previously used by Panu and Unny (1980) for streamflow synthesis using a pattern recognition system, as summarized in Table 4.1.

**Group I:** In Ontario, out of 1122 hydrometric stations, discontinued stations were excluded because their outdated data sets may not account for the effects of urbanization on the scaling exponent (Kim et al., 2016). Additionally, continuous hydrometric stations with data gaps exceeding two years (24 monthly observations) were excluded, as gap-filling could potentially affect the Hurst coefficient. As a result, 143 hydrometric stations were chosen, and relevant hydrologic information of the selected hydrometric stations is presented in Appendix 4A.

**Group II:** The Textural Pattern Recognition System was applied to streamflow datasets in Table 4.1. This involved utilizing the same streamflow data employed by Panu and Unny (1980), which included the streamflow time series segmentation approach.

*Table 4.1. Summary of relevant information on rivers used for streamflow synthesis using the textural pattern* recognition system.

| River Name | Hydrometric Gauging Station | Geographical Location (Lat: & Long:) | Years of Records |
|---|---|---|---|
| South Saskatchewan | 05HG001 at Saskatoon | [Lat: 52° 08' 26" N & long:106° 38' 39" W] | Nov 1912 - Oct 1942 |
| Fraser River | 08MF005at Hope | [Lat:49° 23' 09" N & Long: 121° 27' 15" W] | May 1913 - Apr 1943 |
| Black River | 02EC002 near Washago | [Lat: 44° 42' 49" N & Long: 79° 16' 53" W] | Oct 1916 - Sep 1946 |
| Northeast Margaree River at Margaree Valley, N. S (N.E.M) | 01FB001 | [Lat: 46° 22' 10" N & Long: 68° 35' 35" W] | Sep. 1930-Aug. 1960 |
| St. John River at Fort Kent, N.B. (St. John) | 01AD002 | [Lat: 47° 15' 25" N & Long: 60° 58' 36" W] | May 1927-Apr 1957 |
| Grand River at Loch Lomond, N. S. (Grand) | 01FH001 | [Lat: 45° 43' 48" N & Long: 60° 58' 36" W] | June 1921-May 1951 |

## 4.2 Temporal Structure Analysis of Monthly Streamflow

In this section, the results of the scaling behaviour analysis are first presented, with the stations in Ontario classified into pink, brown, and black noise types. Next, the relationship between the Hurst

coefficient and the scaling exponent λ is examined to assess the memory characteristics of the series. Finally, the behaviour of the lag-1 autoregressive coefficient ($\varphi_1$) across various AR-family models is analyzed to evaluate short-term persistence in the different noise-dominated hydrometric stations.

### 4.2.1 Power Spectrum Analysis for estimation of the scaling exponent, λ

The results of applying the methodology developed in this thesis for identifying dominant noise colour, as described in Section 3.1.3, are presented in this section. All Ontario streamflow stations summarized in Table 4A.1 were analyzed using power spectral analysis. Spectral jumps were identified, and a line was fitted to the jump regions to estimate the spectral slope, λ. Based on the estimated values of λ, the stations were classified into brown-, pink-, and black-noise categories. The classification thresholds based on the scaling exponent λ are adopted from established definitions in noise analysis and are widely used in the literature (Dooley & Van de Ven, 1997; Stoyanov et al., 2011).

Two stations were characterized as brown noise, 39 as pink noise, and the remaining 102 as black noise, as shown in

Figure 4.1.


*Figure 4.1. Spatial distribution of stations in Ontario classified according to the dominant colored noise in their monthly streamflow series, showing stations dominated by (a) pink noise, (b) brown noise, and (c) black noise.*

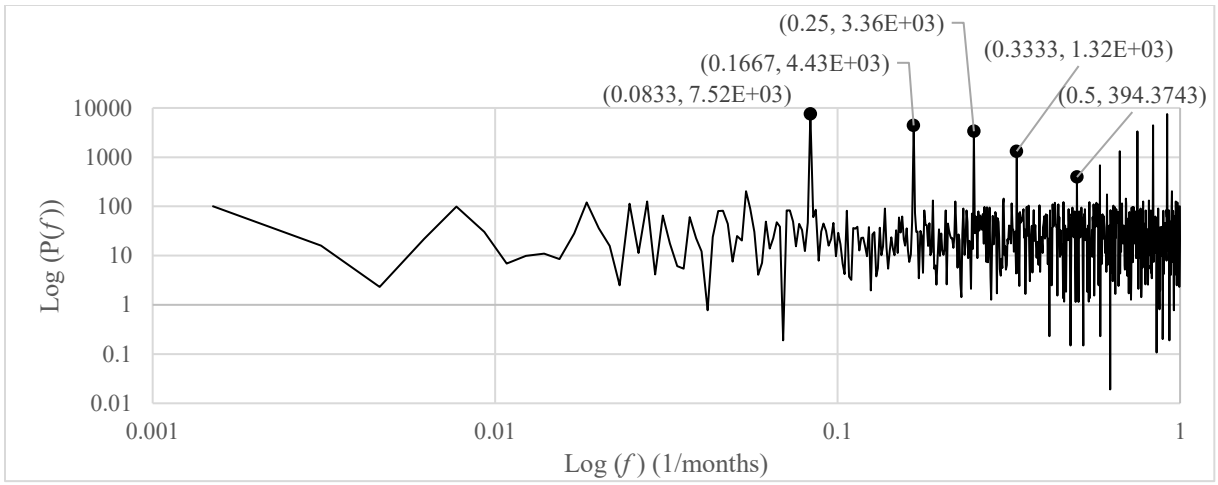
For clarity and focus, one representative station from each coloured-noise category was randomly chosen, and the corresponding results are presented and discussed. For the other stations, the Hurst

and scaling exponents are summarized in Appendix 4A. Figure 4.2 shows the power spectral density (PSD) for the Jock River (02LA007), Basswood River (05PA012), and English River (05QA002), with identified spectral jumps indicated. As shown, distinct discontinuities occur in the power spectrum at frequencies corresponding to fractional cycles. Similar spectral increases were also observed across all other analyzed stations.

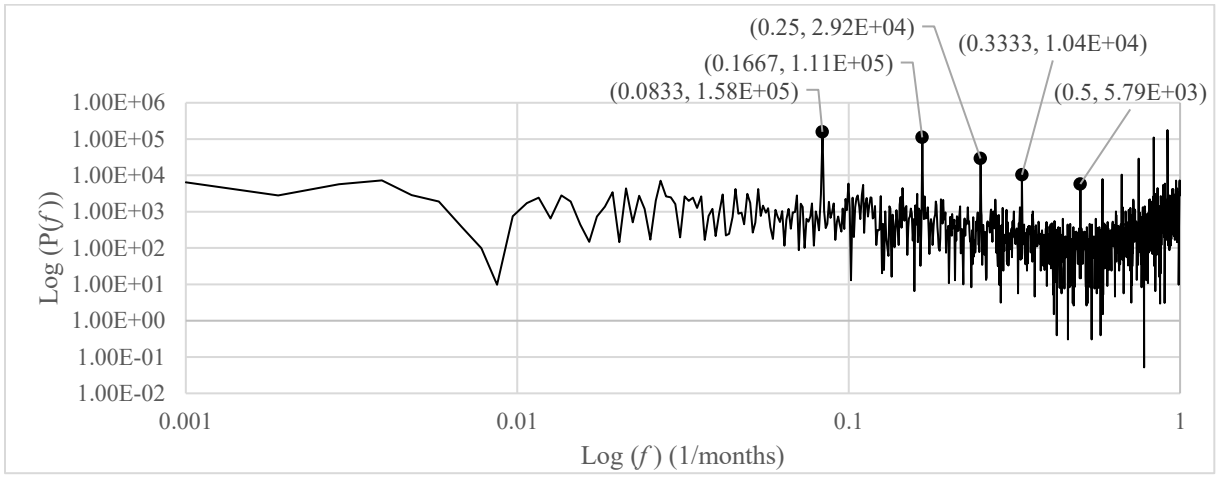
At frequencies greater than f = 0.5, a pronounced rise in spectral power is evident. Consequently, the analysis was restricted to the frequency interval $0.0833 < f < 0.50$. The detected jumps in the power spectrum, associated with frequencies of 1/12, 1/6, 1/4, 1/3, and 1/2 months, were extracted and are presented in Figure 4.3, together with the fitted power-law relationships and their corresponding $R^2$values.

Based on the scaling exponent values of 1.574, 2.008, and 2.6 shown in Figure 4.3, the estimated λ values indicate that the Jock River (Figure 4.3a), Basswood River (Figure 4.3b), and English River (Figure 4.3c) are respectively characterized by pink, brown, and black noise behaviour.
The identification of brown-noise-dominated stations among the full set of 143 sites required careful consideration, since λ values in the range of approximately 1.9 to 2.1 may easily be rounded to 2 based on significant digit considerations and therefore classified as brown noise. As summarized in Table 4.2, eight stations were found to have λ values within this interval.

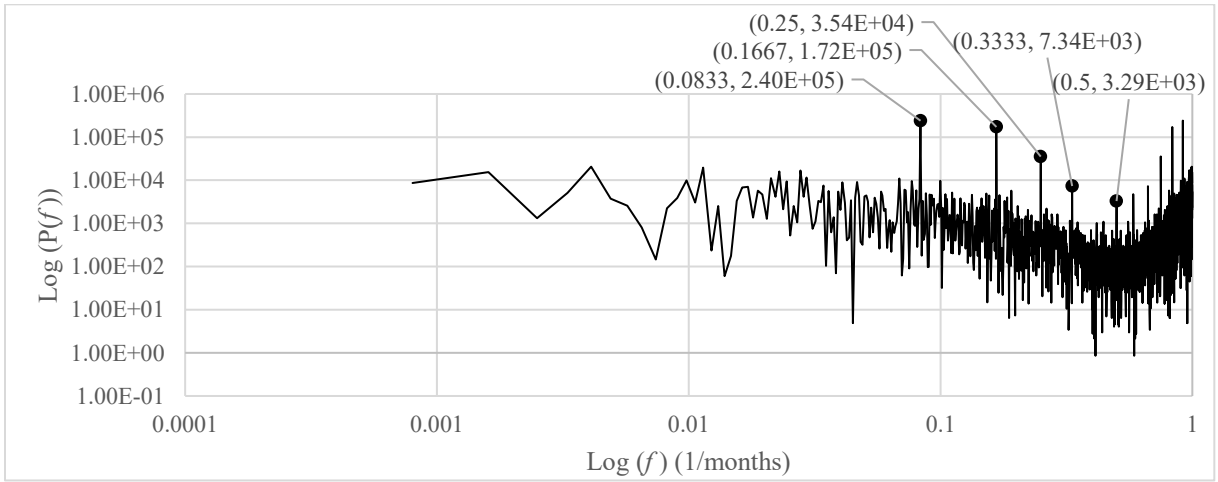
Since brown noise becomes white noise after first-order differencing (Dooley & Van de Ven, 1997), the autocorrelation function (ACF) at lag 1 of the differenced series is expected to be approximately zero. As reported in Table 2, only two stations satisfy this criterion: Basswood River near Winton (05PA012), with an λ value of 2.008 and a lag-1 ACF of 0.003 for the differenced series, and Black Sturgeon River at Highway No. 17 (02AC002), with an λ value of 2.048 and a lag-1 ACF of −0.003. Therefore, only two stations are characterized as brown noise, as shown in Figure 4.1b.




*(a)*

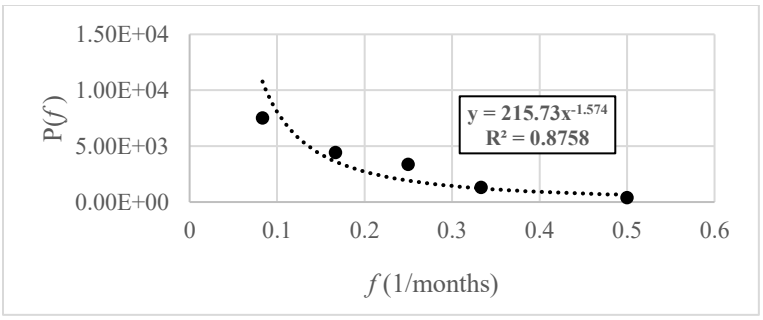



*(b)*

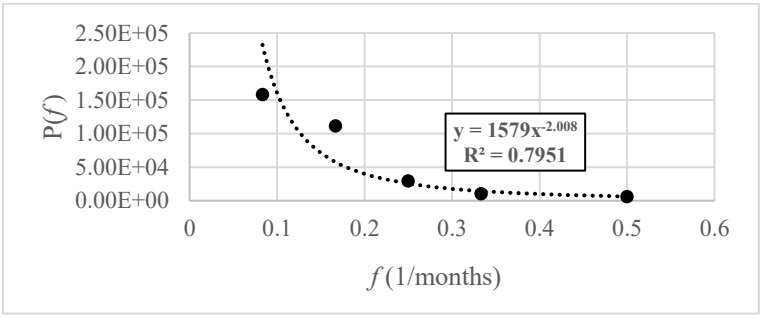



*(c)*

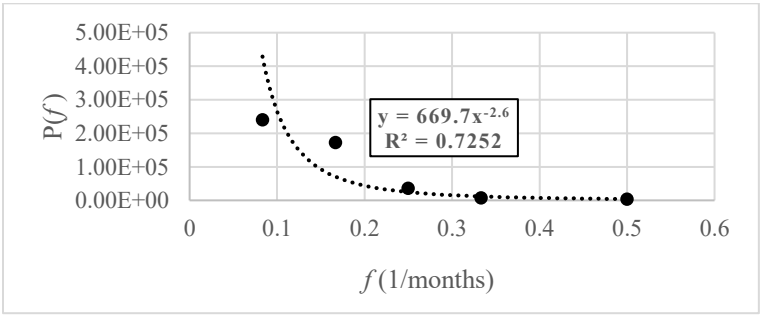
*Figure 4.2. Power spectral density (PSD) plots of rivers with labelled Jumps: (a) Jock River, (b) Basswood River, and (c) English River.*




*(a)*




*(b)*




*(c)*

*Figure 4.3. Isolated jumps in PSD plots of rivers along with the fitted exponential function: (a) Jock River, (b) Basswood River, and (c) English River.*

*Table 4.2. Summary of watersheds exhibiting values of the scaling exponent, λ, in the range of 1.9 and 2.1 for the existence of brown noise.*

| Station Name | Scaling Exponent (λ) | ACF (1) of a differenced series (d = 1) |
|---|---|---|
| Black River near Actinolite | 1.958 | -0.18074 |
| Nagagami River at Highway No. 11 | 1.963 | -0.09993 |
| Black River near Washago | 1.98 | -0.12368 |
| Moira River near Deloro | 1.983 | -0.20980 |
| **Basswood River near Winton** | **2.008** | **0.00285** |
| **Black Sturgeon River at Highway No. 17** | **2.048** | **-0.00212** |
| North Branch Muskoka River at Port Sydney | 2.074 | -0.20216 |
| Moira River near Foxboro | 2.075 | -0.14014 |

**Note**: The rivers exhibiting dominance of the brown noise are highlighted by bolded letters.

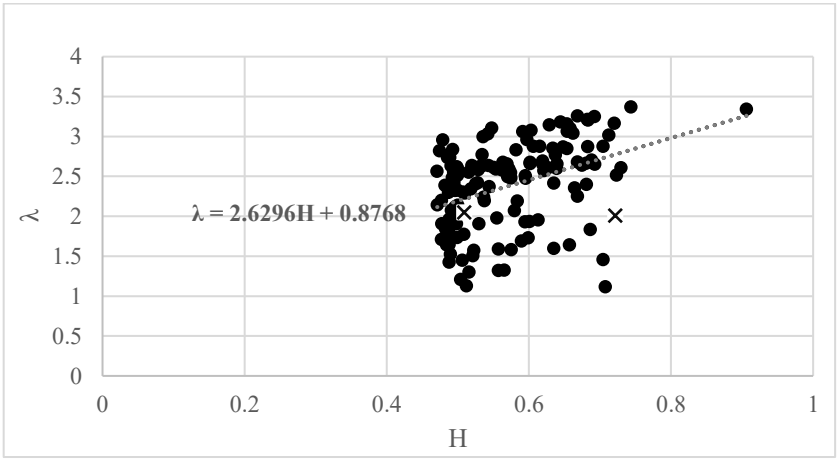
### 4.2.2 Relationship between the Hurst Coefficient and Scaling Exponent

Further analysis was conducted using the relationship between the scaling exponent (λ) and the Hurst coefficient (H) to better characterize the memory and persistence properties of the streamflow time series. While the scaling exponent provides information on the spectral behaviour and dominant noise colour, the Hurst coefficient offers a complementary measure of long-term dependence and persistence in the time domain. Estimating H therefore allows for a more complete assessment of memory behaviour, distinguishing between short-range dependence and long-memory processes.

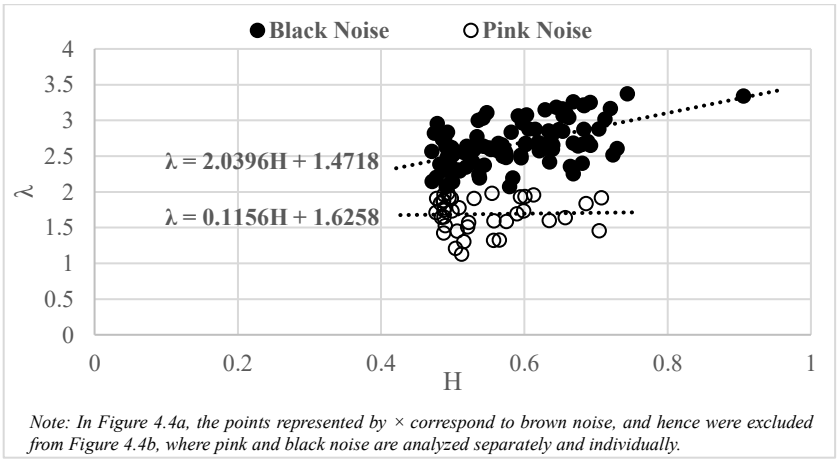
The Hurst coefficient, as described and discussed in Section 3.1.1, was subsequently estimated for all stations. Figure 4.4 displays the relationship between the Hurst coefficient (H) on the x-axis and the scaling exponent λ on the y-axis for the complete dataset: (a) without distinguishing among coloured noise types, and (b) with the data grouped according to coloured noise classification.
The best-fit regression lines are shown along with their corresponding equations. Because only two stations were classified as exhibiting brown-noise characteristics, these stations were excluded from the analysis presented in Figure 4.4b. Including them would have resulted in a zero value of the mean squared error (MSE), which would not be informative for comparative purposes. Table 4.3 summarizes the associated MSE values of different cases considered in this thesis. For the combined noise model, where all coloured noise (pink, brown, and black) is treated together, the linear fit is given by the equation $\lambda = 2.6296H + 0.8768$, with an MSE value of 0.236. However, a relatively high value of MSE suggests that the equation lacks an optimal fit.
When the coloured noise is separated by pink and black noise consideration, it shows an improved best fit with a lower value of MSE. For Pink noise, the linear relationship of $\lambda = 0.1156H + 1.6258$ resulted in a significantly lower value of MSE of 0.050, and for black noise, the linear model is expressed as $\lambda = 2.0396H + 1.4718$ with an MSE value of 0.053.
A relatively very gradual slope of 0.1156 in the λ -H relationship for pink noise indicates that the scaling exponent is less dependent on variations in the Hurst coefficient. This is suggestive of the fact that the level of pink noise remains relatively stable despite changes in the Hurst coefficient of a time series. On the other hand, a sharp increase in slope of 2.0396 in the λ -H relationship for black noise is indicative of the fact that the level of black noise is highly dependent on H. These

findings are aligned with the fact that pink noise has lower memory than black noise as described below.

It is noted that when streamflow data from a hydrometric station is characterized by a specific type of coloured noise, it does not imply the absence of other types of coloured noise. Instead, it is indicative of the fact that the characteristics of the dominant-coloured noise overshadow those of the others. When modelling such a time series, it is important to consider that preserving the characteristics of one noise type (e.g., black noise) may compromise the preservation of characteristics associated with other noise types (e.g., pink noise).




*(a)*




*Note: In Figure 4.4a, the points represented by × correspond to brown noise, and hence were excluded from Figure 4.4b, where pink and black noise are analyzed separately and individually.*

*(b)*

*Figure 4.4. Hurst coefficient H versus scaling exponent λ values for the hydrometric stations: (a) without separating different levels of noise, and (b) with noise levels separated.*

*Table 4.3. Approaches and corresponding models for fitting 'λ' values as a function of the Hurst coefficient.*

| Case | Model (λ = f(H)) | MSE* |
|---|---|---|
| Linear fit without separating noise colour (for pink, brown, and black) | λ = 2.6296H + 0.8768 | 0.236 |
| Linear fit for pink noise | λ = 0.1156H + 1.6258 | 0.050 |
| Linear fit for black noise | λ = 2.0396H + 1.4718 | 0.053 |

***Note:*** **MSE =* $1/n \sum_{i=0}^{n}(\lambda_i - \widehat{\lambda_i})^2$*, where n is the number of data points,* $\lambda_i$ *and* $\widehat{\lambda_i}$ *are respectively, the actual and predicted values of* λ*.*

To further examine the strength of the λ-H relationship, beyond the fitted linear trends, a cross-correlation analysis was performed at lag zero. For black-noise-dominated stations, the cross-correlation at lag zero is 0.5456, indicating a moderate to strong positive dependence between λ and H. In contrast, pink-noise-dominated stations exhibit a very weak cross-correlation of 0.033, suggesting nearly independence between the two parameters. Moreover, the statistical structure underlying the λ–H relationship and the contribution of higher-order terms were evaluated using sequential analysis of variance for each coloured noise category.

Table 4.4 reveals that for pink-noise-dominated stations, neither the linear nor quadratic terms are statistically significant, while the cubic term is significant ($p = 0.044$), indicating that the relationship between λ and H is governed by higher-order nonlinear behaviour. This finding is consistent with the previously observed weak cross-correlation (0.033), suggesting that the interaction between λ and H is complex and not well captured by simple linear dependence. In contrast, black-noise-dominated stations exhibit a highly significant linear term ($p < 0.001$), with negligible contributions from higher-order terms. This indicates that the λ–H relationship in black-noise systems is predominantly linear and structurally simple. The moderate cross-correlation value (0.5456) observed for black noise further supports this result. Overall, these findings demonstrate that pink-noise systems are characterized by weak, nonlinear, and more complex relationships between scaling behaviour and persistence, whereas black-noise systems exhibit strong, linear, and more predictable dependence structures.

*Table 4.4. Sequential analysis of variance for the λ–H relationship across pink- and black-noise-dominated stations.*

| ***Noise Type*** | ***Source*** | ***F**** | ***P**** |
|---|---|---|---|
| ***Pink Noise*** | *Linear* | *0.04* | *0.842* |
| | *Quadratic* | *0.77* | *0.386* |
| | *Cubic* | *4.37* | *0.044* |
| ***Black Noise*** | *Linear* | *42.50* | *0.000* |
| | *Quadratic* | *0.01* | *0.906* |
| | *Cubic* | *0.02* | *0.890* |

*Note: *F indicates how much variance is explained by a model term relative to the unexplained (residual) variance, whereas **P represents the probability that the observed effect (or a more extreme one) could occur by chance if the term had no real effect.*

A model selection framework based on the λ –H relationship can help identify the most appropriate streamflow model by linking spectral behaviour, λ, with persistence and memory structure, H. A model-selection strategy based on the λ –H relationship distinguishes whether a watershed behaves as a memory-driven natural system or a flashy urban system, guiding which modelling approaches are most appropriate.

The relationship for black noise-dominated streamflow, $\lambda = 2.0396H + 1.47$, shows a very strong and nonlinear link between the spectral slope, λ and the Hurst exponent, H. Even a small increase in H produces a large increase in λ, meaning the spectrum decays much faster than classical fractional Brownian motion (fBm). This implies a highly persistent and very smooth process dominated by strong long-range dependence. Hydrologically, such a relationship characterizes natural basins with substantial infiltration and storage, where low-frequency components dominate streamflow behaviour. These watersheds exhibit strong memory over months to years, reflecting slow hydrologic processes and well-organized internal dynamics. When λ and H are strongly coupled ($\lambda = 2H + 1.5$), the streamflow exhibits strong long-term memory, persistent low-frequency fluctuations, and multi-scale fractal behaviour, typical of natural basins with substantial storage from soils, groundwater, and wetlands. These systems require models that capture long memory and storage dynamics, such as ARFIMA, HBV, SAC-SMA, GR4J/GR5J, or other conceptual storage-based models, along with long calibration periods (20–30 years) to properly represent slow hydrologic processes and climate teleconnections.

The relationship for pink noise-dominated streamflow, $\lambda = 0.11H + 1.6$, shows a very weak coupling between λ and H. In such cases, λ remains nearly constant even when H changes significantly, indicating that long-range dependence plays a minimal role in shaping the spectral slope. This suggests that other external or fast processes dominate the behaviour of the system. Hydrologically, this pattern corresponds to urbanized or highly impervious basins with quick rainfall–runoff responses and minimal storage. Such systems are flashy, dominated by high-frequency variations, and possess very weak hydrologic memory. When λ depends only weakly on H ($\lambda = \mathbf{0.11H + 1.6}$), the watershed shows minimal memory and fast rainfall–runoff response, characteristic of urban or highly impervious basins. These systems are best modelled using short-memory models, including ARMA (1,1), AR (2), transfer-function models, SWMM, or event-based runoff approaches, supported by high-resolution rainfall data and periodic recalibration due to parameter instability caused by land-use changes and infrastructure evolution.

### 4.2.3 Lag-1 Autoregressive Coefficient ($\varphi_1$) Analysis

Various AR and ARMA models were fitted to all stations, and the lag-1 autoregressive coefficient ($\varphi_1$) was plotted in Figure 4.5, as this parameter provides a consistent measure of short-term persistence in time series. The coefficient $\varphi_1$ quantifies the dependence of the current value on its immediately preceding value. A value of $\varphi_1$ close to 1 indicates strong persistence and long-memory behaviour, whereas values near zero indicate weak or negligible autocorrelation. Furthermore, $|\varphi_1| < 1$ corresponds to a stationary time series, in which the memory of the process decays over time. In contrast, $\varphi_1 > 1$ or $\varphi_1 < -1$ indicates non-stationarity, where the time series tends to diverge or oscillate without bound.

Figure 4.5 shows that when moving from AR (1) to ARMA (2,2), $\varphi_1$ increases for black-noise-dominated watersheds, whereas for pink-noise-dominated watersheds, $\varphi_1$ values remain relatively stable across different model structures. This indicates that black-noise series are more sensitive to model order in terms of short-term persistence, while pink-noise series exhibit more consistent memory characteristics irrespective of the complexity of the ARMA model.

Following the analysis of the temporal structure of monthly streamflow using various AR and ARMA models, the textural pattern recognition system described in Section 3.2.2 was applied to

three watersheds dominated by pink noise and three watersheds dominated by black noise, which were randomly selected for synthetic streamflow generation. Additionally, the Extended Brownian Motion model, as explained in Section 3.2.1.2, was fitted to two watersheds exhibiting characteristics of brown-noise dominance. The results of these applications are presented and discussed in the subsequent section.

## 4.3 Results of Streamflow Synthesis

For optimal presentation of results and discussion, firstly, the models are fitted based on the identified dominant noise type. For brown-noise-dominated watersheds, the Extended Geometric Brownian Motion model (described in Section 3.2.1.2) is applied, while for pink- and black-noise-dominated watersheds, the textural pattern recognition system (described in Section 3.2.2) is used. In both cases, the results are compared with those obtained using multiplicative ARIMA models.

Secondly, the performance of the textural pattern recognition system is compared with that of the traditional pattern recognition system to evaluate improvements in synthetic streamflow generation.

### 4.3.1 Streamflow Synthesis for Brown-Noise Watersheds: EGBM Model Performance

As discussed earlier, two watersheds were identified as exhibiting brown-noise-dominated characteristics. The Extended Geometric Brownian Motion (EGBM), being a GBM-based model, is particularly suited for shorter time durations, making it appropriate for seasonal-scale modelling. Through preliminary and goal-oriented experimentation, a seasonal length of six months was selected. This choice is further supported by the Lupanov Time (LT > 6 months), which represents the characteristic time of predictable dynamics in chaotic systems of the Great Lakes region of North America (Mihailović et al., 2023). Therefore, the monthly streamflow series for these watersheds were then segmented accordingly for analysis.

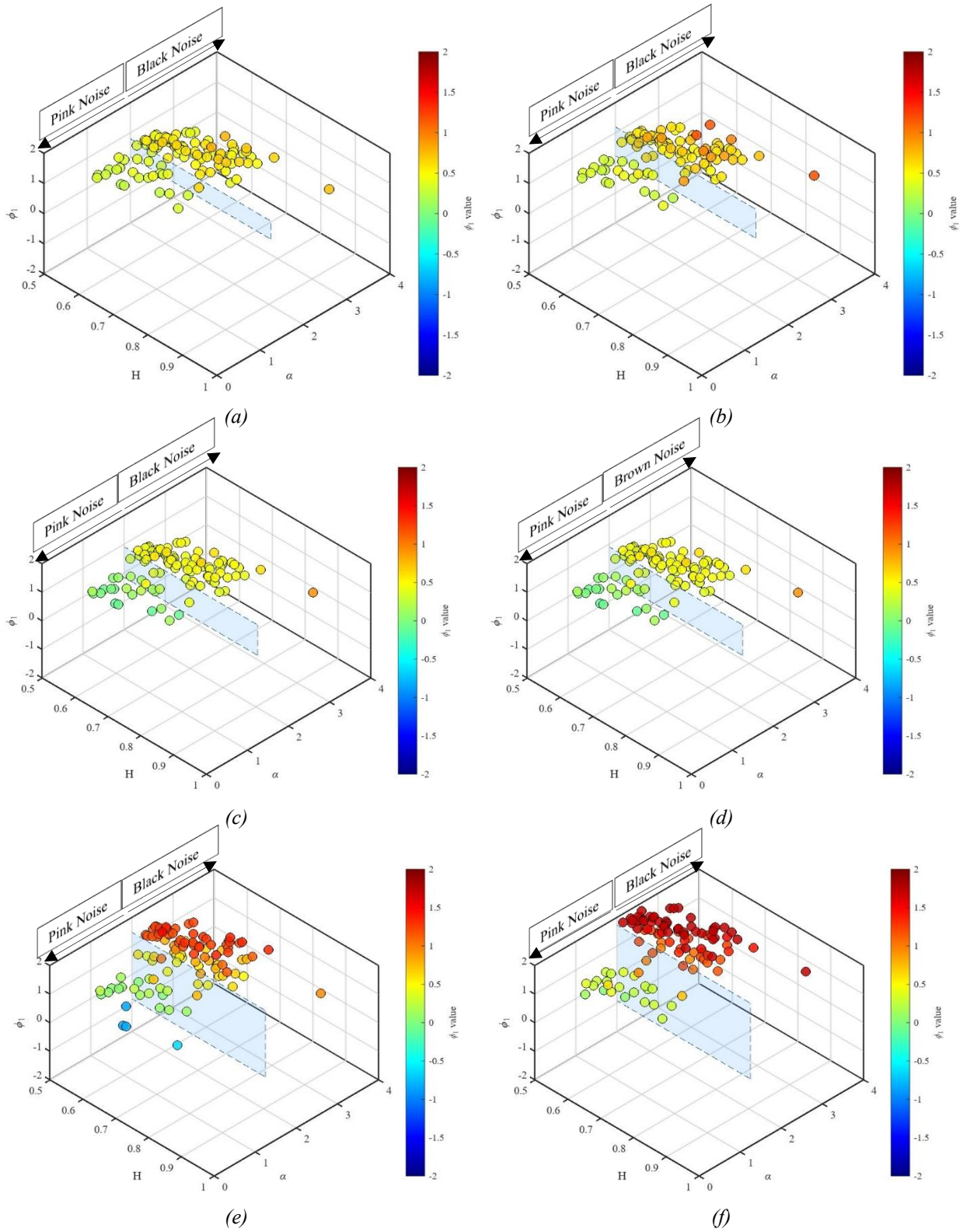
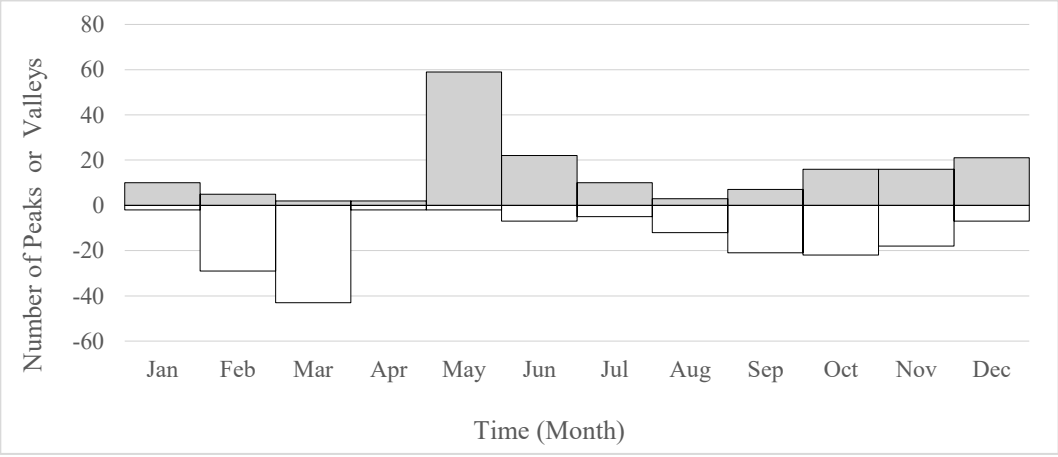


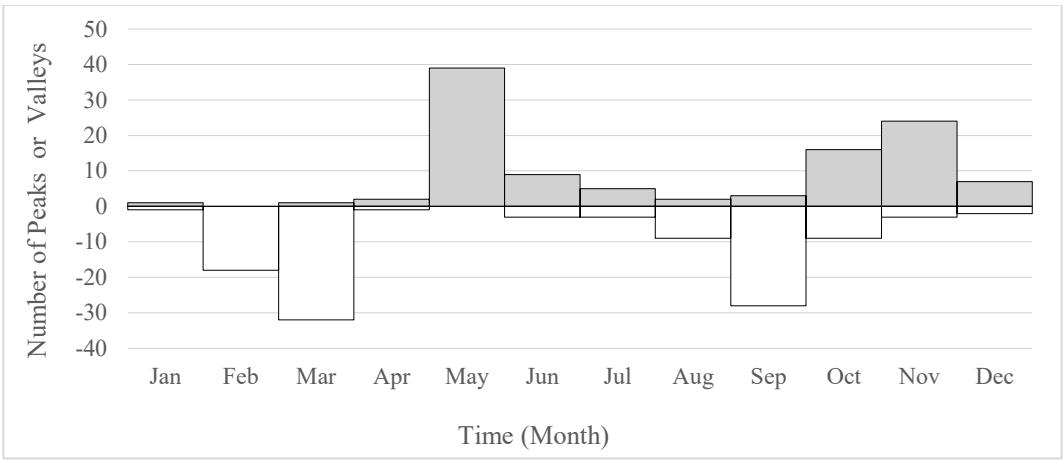

*Figure 4.5. Lag-1 autoregressive coefficient ($\varphi_1$) for all watersheds estimated using (a) AR (1), (b) AR (2), (c) ARMA (1,1), (d) ARMA (2,1), (e) ARMA (1,2), and (f) ARMA (2,2).*

In view of the characteristic time of predictable dynamics in chaotic domains, the LT > 6 months, the segmentation of log-natural transformed monthly time series data is performed for both watersheds, with each segment spanning six months. The location of the beginning and end of a segment corresponding to a season was accomplished using a peak-and-valley plot (Panu and Unny, 1980). To ensure that the data within each season exhibits an increasing or decreasing trend, the segmentation should occur at points where peaks or valleys are at their corresponding maximum and minimum. Upon completion of the segmentation process of the log-natural transformed monthly streamflow time series into two seasons (k=1 and k=2), that is, the data is considered to belong to two diverse classes, and hence the EGBM model with distinct values of parameters corresponding to each season is considered and applied separately to each class of segmented data. However, it is important to note that these two classes or seasons (or seasonal segments) belong to the same dataset of a river.

Figure 4.6 shows the peaks and valleys plot for both the Basswood River and the Black Sturgeon River. For both rivers, season k = 1 (from May to October) and season k = 2 (from November to April) are identified as illustrated in Figure 4.7.

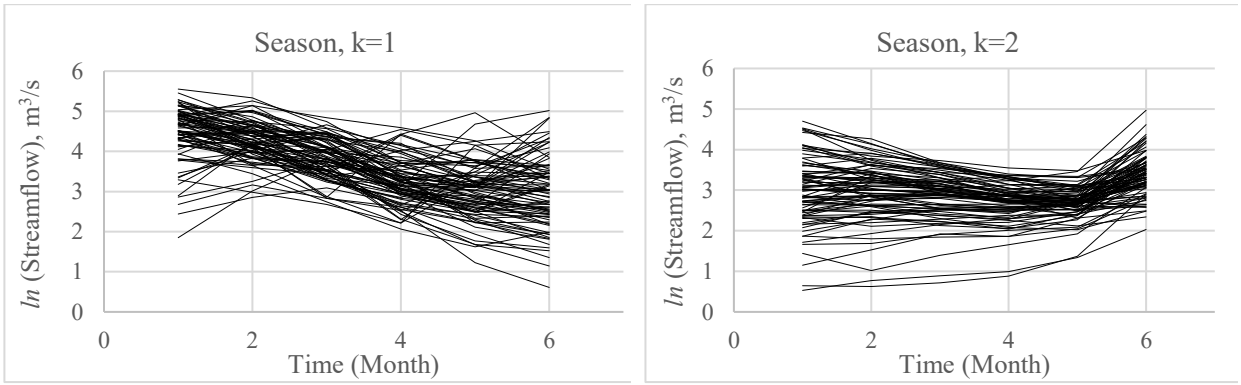



*(a)*

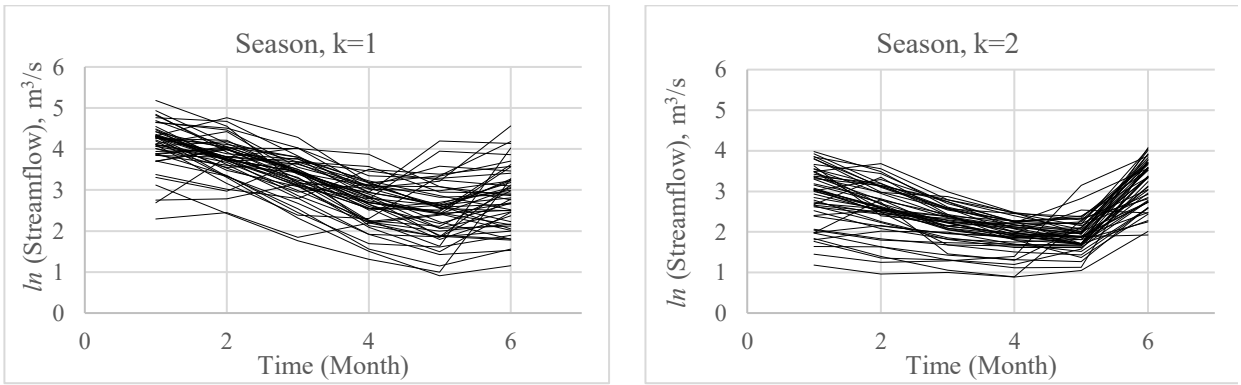



*(b)*

*Figure 4.6. Peak and valley histograms: (a) Basswood River and (b) Black Sturgeon River, where peaks are defined as points in the monthly streamflow series satisfying $Q_t > Q_{t-1}$ and $Q_t > Q_{t+1}$, and valleys are defined as points satisfying $Q_t < Q_{t-1}$ and $Q_t < Q_{t+1}$.*




*(a)*




*(b)*

*Figure 4.7. Pictorial representation of the segmented and transformed streamflow data: (a) Basswood River and (b) Black Sturgeon River with season, k =1 (May-Oct) and season. k =2 (Nov-Apr).*

Separate EGBM models were developed for each of the two identified seasons to synthesize streamflow. Accordingly, within each season, the EGBM parameters were optimized using a Genetic Algorithm (GA). Table 4.5 presents the optimized seasonal parameter values $\ln(\Omega_i^k)$ for each month of both watersheds.

*Table 4.5. Summary of optimized $\ln(\Omega_i^k)$ values in the EGBM model.*

| Hydrometric Station (Watershed) | Month | Season | | Hydrometric Station (Watershed) | Season | |
|---|---|---|---|---|---|---|
| | | k = 1 | k = 2 | | k = 1 | k = 2 |
| Basswood River Winton (05PA012) | 1 | 0.278 | 0.001 | Black Sturgeon River (02AC002) | 0.152 | 0.036 |
| | 2 | 0.114 | -0.001 | | 0.045 | -0.044 |
| | 3 | 0.038 | -0.013 | | 0.007 | -0.083 |
| | 4 | -0.008 | -0.023 | | -0.039 | -0.086 |
| | 5 | -0.020 | -0.015 | | -0.044 | -0.068 |
| | 6 | -0.018 | 0.023 | | -0.030 | 0.026 |

The optimized EGBM models for each season at both watersheds were then used to synthesize streamflow time series. The performance of the proposed EGBM model was evaluated and compared with that of a multiplicative Seasonal Autoregressive Integrated Moving Average model, ARIMA $(p,d,q) \times (P,D,Q)_{12}$, where $p,d,q$ denote the order of the non-seasonal model and $P,D,Q$ represent the seasonal components with a periodicity of $w = 12$ months.

To identify the most appropriate ARIMA $(p,d,q) \times (P,D,Q)_{12}$ model for the monthly streamflow data of the Basswood River near Winton (05PA012) and the Black Sturgeon River at Highway No. 17 (02AC002), multiple candidate models were assessed using Log-Likelihood, AICc (corrected Akaike Information Criterion), AIC, and BIC criteria (Salas et al., 1980). Based on the minimum value of AICc, the selected model for the Basswood River turned out to be ARIMA $(2,2,0) \times (0,1,1)_{12}$, with autoregressive parameters $(\phi_1, \phi_2)$ and a seasonal moving average parameter $(\Theta_1)$. For the Black Sturgeon River, the optimal model was ARIMA $(2,2,0) \times (1,1,1)_{12}$, with autoregressive parameters $(\phi_1, \phi_2)$ and seasonal moving average parameters $(\Theta_1, \Theta_2)$.

Table 4.6 summarizes the model parameters and their associated standard error (SE), which indicates the precision of the parameter estimates. A lower SE value suggests higher confidence in the estimated parameter; conversely, a higher SE value implies greater uncertainty. The ARIMA model for the Basswood River near Winton (05PA012) shows significant negative autocorrelation with $\phi_1$ = - 0.4884 with SE = 0.0287 and $\phi_2$ = - 0.4026 with SE = 0.0288, while

the strong seasonal component is captured by the MA (12) parameter $\Theta_1$ = 0.9738 with SE = 0.0084. Similarly, the ARIMA model for the Black Sturgeon River exhibits significant negative autocorrelation ($\phi_1$ = - 0.5772 with SE = 0.0371 and $\phi_2$ = - 0.4068 with SE = 0.0372), and strong seasonal components. Further details regarding the selection of the ARIMA model are provided in Appendix 4B.

*Table 4.6. A summary of the optimized parametric estimates of ARIMA for both Rivers.*

| Hydrometric Station | Model* | Parameter | SE Coefficient |
|---|---|---|---|
| Basswood River near Winton (05PA012) * | AR (1) | $\phi_1$=-0.4884 | 0.0287 |
| | AR (2) | $\phi_2$=-0.4026 | 0.0288 |
| | Seasonal MA (12) | $\Theta_1$=0.9738 | 0.0084 |
| Black Sturgeon River at Highway No. 17 (02AC002) ** | AR (1) | $\phi_1$=-0.5772 | 0.0371 |
| | AR (2) | $\phi_2$=-0.4068 | 0.0372 |
| | Seasonal AR (12) | $\Phi_1$=0.0367 | 0.0426 |
| | Seasonal MA (12) | $\Theta_1$=0.9579 | 0.0164 |

Note*: * The model used for Basswood River near Winton is ARIMA* $(2,2,0)\times(0,1,1)_{12}$ *with two autoregressive parameters ($\phi_1$ and $\phi_2$) and one seasonal moving average parameter ($\Theta_1$) whereas**The model used for Black Sturgeon River at Highway No. 17 is ARIMA* $(2,2,0)\times(1,1,1)_{12}$ *with two autoregressive parameters of ($\phi_1$ and $\phi_2$) and one seasonal autoregressive parameter ($\Phi_1$) and one seasonal moving average parameter ($\Theta_1$).*

After determining the parameters of the ARIMA model, the model was used to synthesize streamflow time series. A comparison with the historical streamflow time series is presented in Table 4.7, which summarizes the statistical characteristics of the average of 30 synthetic realizations by both the ARIMA and EGBM models. For both the Basswood and Black Sturgeon watersheds, the synthesized series by EGBM exhibit mean values that are closer to the historical mean values than those produced by the ARIMA model.

Specifically, for the Basswood River, the historical mean flow is 39.08 $m^3/s$, while the EGBM model yields a mean of 39.78 $m^3/s$, which corresponds to an increase of 1.8%. In contrast, the ARIMA model produces a higher mean of 43.86 $m^3/s$, representing an increase of 12.23% relative to the historical value. Similarly, for the Black Sturgeon River, the historical mean is 24.22 $m^3/s$, compared to an EGBM mean of 25.81 $m^3/s$ (an increase of 6.6%) and an ARIMA mean of 26.98 $m^3/s$ (an increase of 11.4%). Both skewness and kurtosis are also better captured by the EGBM model, as they more accurately reflect the distribution of historical streamflow data. Such observations on the performance of the EGBM model are indicative of its ability to model and

replicate not only the central tendency and variance but also the overall shape characteristics of the data.

To evaluate model adequacy, the autocorrelation function (ACF) of the residual series synthesized by both the EGBM and ARIMA models is presented in Figure 4.8, together with the 95% confidence bounds ($\pm 1.96/\sqrt{N}$, where N denotes the number of data points). As shown in Figure 4.8, the residuals from the EGBM model for the Basswood River indicate that ACF (1) lies within the confidence limits, suggesting no significant autocorrelation at lag 1, while ACF (2) exceeds the bounds, implying remaining correlation at lag 2. Lags 3 through 7 remain within the confidence limits, indicating no additional significant residual correlation.

For the Black Sturgeon River, the EGBM residual series shows that ACF values for all lags from 1 to 12 fall within the confidence limits, indicating that the model effectively captures the temporal dependence structure. In contrast, for the ARIMA model, the Basswood residuals exhibit significant remaining autocorrelation at ACF (1) and ACF (3), while the Black Sturgeon residuals show uncaptured correlation at ACF (1), with the remaining lags within the confidence bounds.

Overall, a key finding is that the ARIMA model fails to adequately capture lag-1 dependence for both watersheds, with ACF (1) values of − 0.141 for the Basswood River and − 0.180 for the Black Sturgeon River, whereas the EGBM model demonstrates improved performance in removing short-term correlation, particularly for the Black Sturgeon River.

*Table 4.7. A Summary of statistical parameters of historical and synthesized samples for both rivers.*

| **Hydrometric Station (Watershed)** | **Sample** | | **Statistical Properties** | | | |
|---|---|---|---|---|---|---|
| | | | **Mean ($m^3/s$)** | **Std. Dev. ($m^3/s$)** | **Skewness** | **Kurtosis** |
| Basswood River near Winton (05PA012) | Historical | | 38.08 | 36.27 | 2.13 | 5.29 |
| | Synthesized | EGBM | 39.78 | 37.82 | 2.42 | 5.87 |
| | | ARIMA | 43.86 | 49.14 | 4.18 | 8.01 |
| Black Sturgeon River at Highway No. 17 (02AC002) | Historical | | 24.22 | 23.14 | 2.15 | 6.38 |
| | Synthesized | EGBM | 25.81 | 25.13 | 2.77 | 7.17 |
| | | ARIMA | 26.98 | 32.59 | 3.16 | 9.05 |

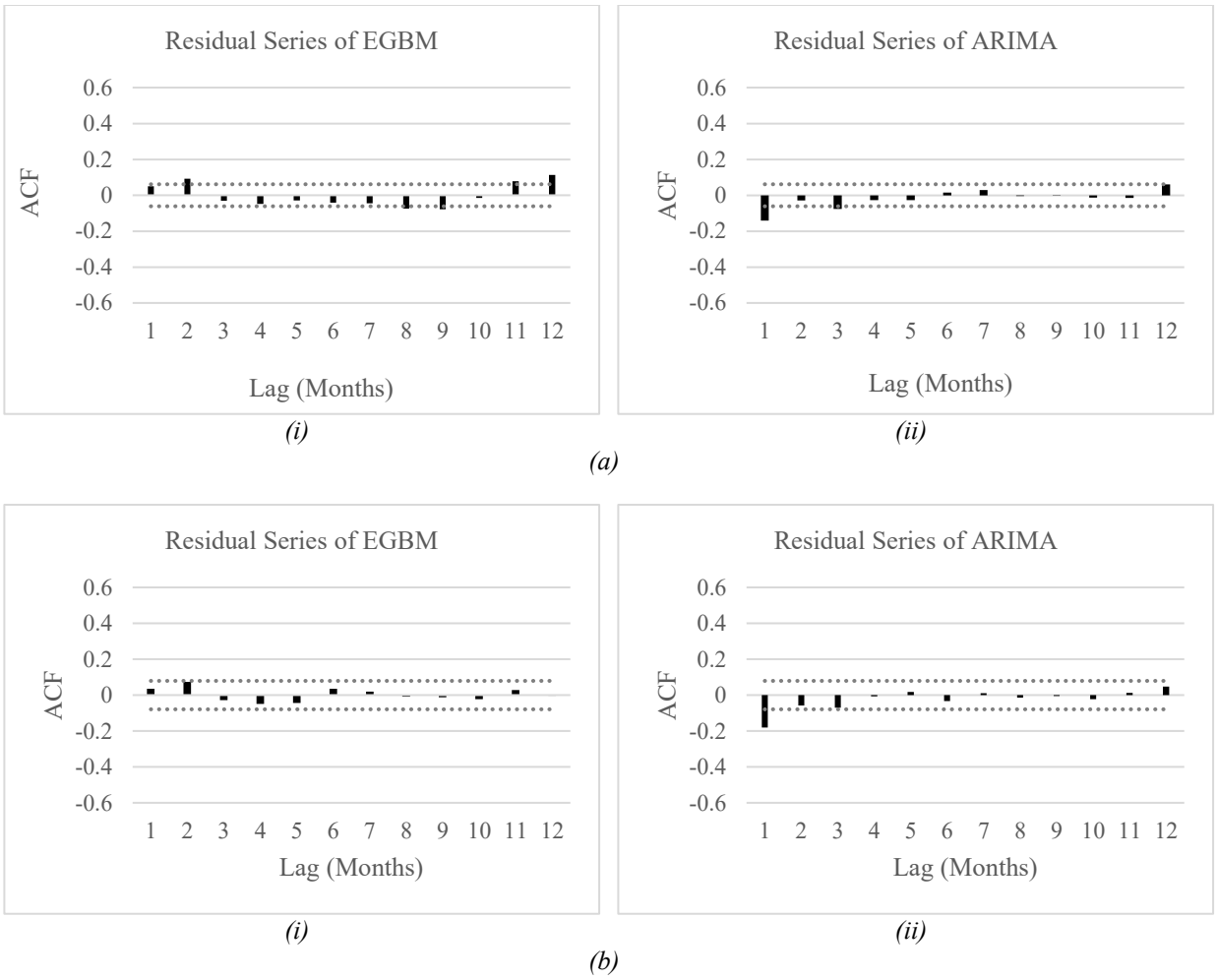



*Figure 4.8. Correlograms of Residual Series of (a) Basswood River and (b) Black Sturgeon River using (i) EGBM and (ii) ARIMA models.*

### 4.3.2 Streamflow Synthesis for Pink and Black-Noise Watersheds: Textural Pattern Recognition System

From the 102 watersheds characterized by black noise and the 39 watersheds characterized by pink noise, three watersheds from each group were randomly selected and are listed in Table 4.8.

All stations are located within Ontario. Black-noise-dominated stations are primarily concentrated in Southern and Northwestern Ontario, whereas pink-noise-dominated stations are more widely distributed, from Central Ontario extending into Southern Ontario (see Figure 4.1).

A study by Wang et al. (2016), which investigated the response of Ontario to global warming, aligns with the findings of this research. In regions of Southern and Northwestern Ontario (found

in this study to be dominated by black noise), the common features are mainly similar rainfall trends, total precipitation trends, and temperature trends.

In regions of Central and Southern Ontario (found in this study to be dominated by pink noise), the common features between Southern and Central Ontario are mostly related to snow trends, and the cross-correlation between temperature and precipitation is generally weak, except during summer (Wang et al., 2016). Therefore, it is reasonable to expect similar hydrological behaviour within each noise-dominated group.

The three stations selected for further analysis were randomly chosen from each noise category, with the additional criterion that their λ values span the full range observed for that noise type (Table 4.8). For example, pink-noise stations have λ values ranging from 0 to 2; if two randomly selected stations had nearly identical λ values (e.g., 1.1 and 1.112), one was replaced by another station to better capture the range. This method ensures that the selected stations are representative of the variability within each noise-dominated group.

The first step in the TPRS methodology is data segmentation. The monthly streamflow data of the six above-noted watersheds (Table 4.8) were segmented into two seasons ($k$ = 1 and $k$ = 2) following the approach of Panu et al. (1978), and then encoded into textural features, as shown in Figure 4.9.

*Table 4.8. Characteristics of the selected pink- and black-noise-dominated watersheds used for streamflow synthesis.*

| **Hydrometric Gauging Station** | **Hurst coefficient (H)** | **scaling exponent (λ)** | **Dominant Noise Type** |
|---|---|---|---|
| Little White River near Bellingham (02CC005) | 0.5087 | 1.777 | Pink |
| Mattagami River near Timmins (04LA002) | 0.5122 | 1.129 | |
| Root River at Sault Ste. Marie (02CA002) | 0.488 | 1.648 | |
| Collins Creek near Kingston (02HM005) | 0.4922 | 2.484 | Black |
| Kaministiquia River at Kaministiquia (02AB006) | 0.4819 | 2.389 | |
| North Magnetawan River near Burk's Falls (02EA005) | 0.5157 | 3.302 | |

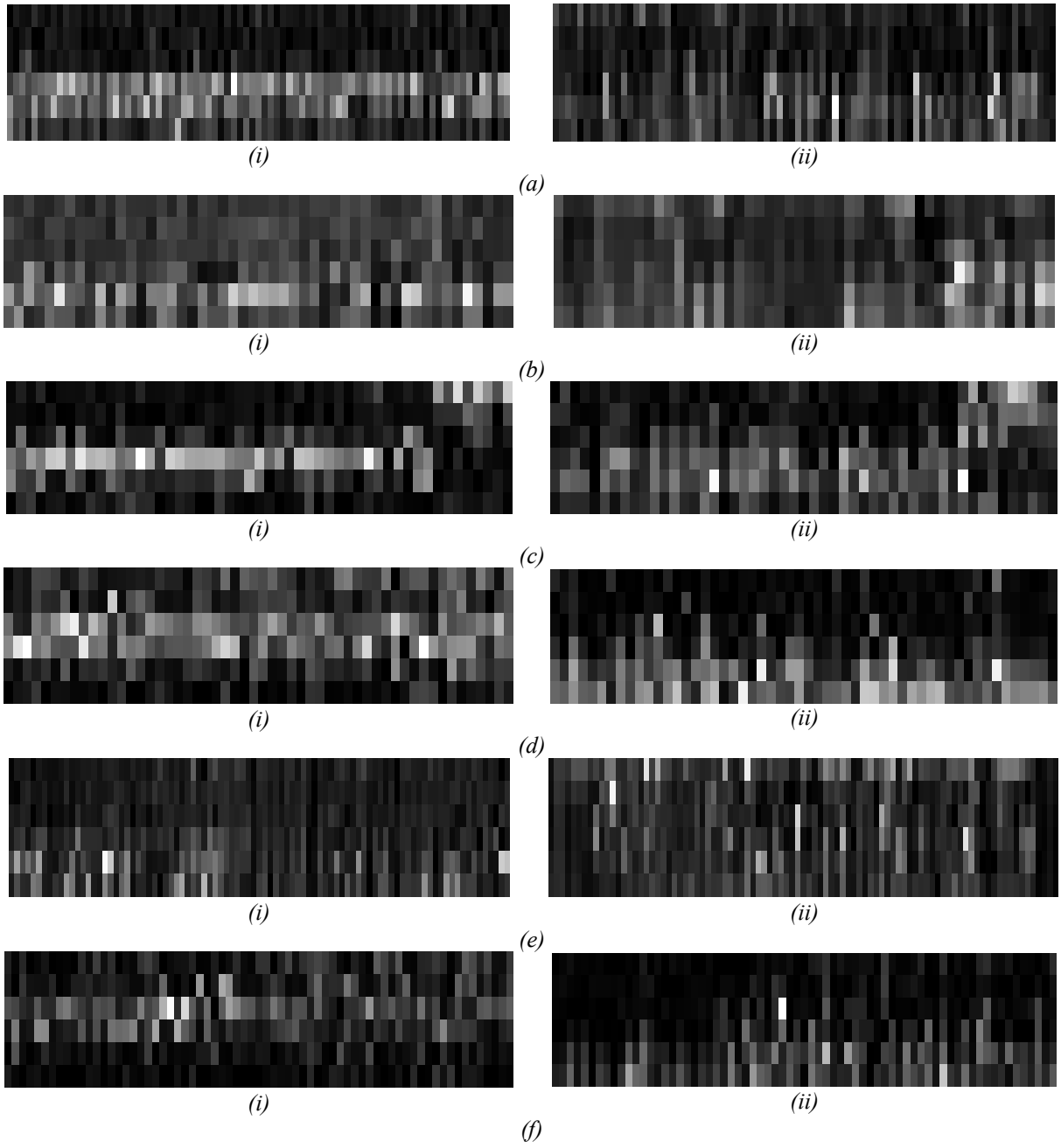



*Figure 4.9. Encoded seasonal grayscale images for the season (i) k =1 and (ii) k =2 of (a) Little White River near Bellingham, (b) Mattagami River, (c) Root River, (d) Collins Creek, (e) Kaministiquia River and (f) North Magnetawan River.*

The process for extracting textural image features (described in section 3.2.2.2) was employed. To measure the dependency in the intensity of a pixel in each grayscale textural image, a lag-ζ (=1) approach was utilized. The lagging process involved in assessing the Lag-ζ (=1) correlation in the horizontal direction, the first row needs to be excluded due to shifting (i.e., lagging) in the

computations of the $SACF^k$ for the season k. In the vertical direction, a lag-ζ (=1) approach indicates that the synthesis of 30 years requires the consideration of 30 years plus one additional year from the immediate past of streamflow data (for additional detail, refer to section 3.2.2.2). Consequently, the matrices, $SACF^{k=1}$ and $SACF^{k=2}$, respectively for the season (k=1 and k=2) were separately developed to represent the intrarelationship of adjacent pixels within each of the seasonal grayscale images.

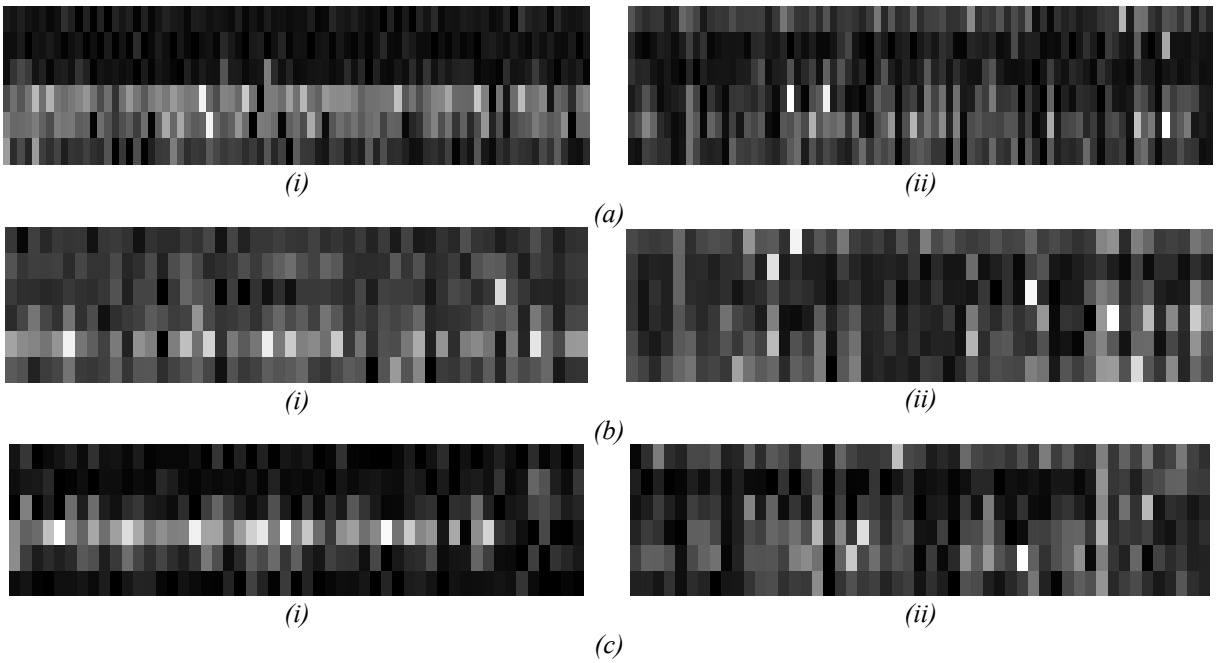
Upon the transformation of these matrices into the frequency domain using a DFT, it is feasible to compute the power spectrum matrices for each season (k = 1 and k = 2), hereafter referred to as the seasonal textural features. The DFT is employed because the autocorrelation function of an image is directly related to the power spectrum obtained via the Fourier transform.

Once the power spectrum is obtained, it is divided into high-, and low-power components using a cut-off point (such as the mean of the power spectrum). The approach of separating the power spectrum into high- and low-power components is effective. However, using the mean as the threshold may not be suitable for all cases, particularly for rivers with highly skewed flow distributions. Exploring alternative thresholding methods could improve the flexibility and generalizability of the model.

An approximation of the underlying Gaussian distribution of the power spectrum enables the generation (or synthesis) of values for the power spectrum using the standard Gaussian probability function. The process involves the generation of random numbers in the range of 0 and 1 and a decision about the mean and standard deviation. If this random number is lower than the probability of occurrence of a low-power spectrum group for each month, the first random generator generates a random value belonging to the Gaussian distribution approximated to the low-power spectrum, and the second random generator generates a random value belonging to the Gaussian distribution approximated to the high-power spectrum. For each pixel of $(i, j)$, the generator power spectrum must be paired with the reference vector of $R_{PH}^k \pm C \times St_{PH}^k$ and $R_{PL}^k \pm C \times St_{PL}^k$ in the low-power group. If yes, the second random generator is activated to generate the phase component. Such synthesized power spectrum values were used as amplitudes (from the first generator), and newly generated phase information (from the second generator) was added to form the Fourier coefficient matrices for each season across the six rivers.

Once the Fourier coefficients were generated for each season, image intensity was computed using a quadratic equation with two positive roots. To ensure adherence to the original image properties, a specific constraint was necessary. The selection of the root directly impacted the pixel intensity of the synthesized image. Hence, an alternative method was needed to determine the root selection based on the original pixel intensity of a seasonal image. If the original pixel intensity of a seasonal image being synthesized exceeded the mean intensity of all seasonal pixels, the positive sign was chosen $(-b + \sqrt{b^2 - 4ac}/2a)$. Conversely, the negative sign was employed otherwise $(-b - \sqrt{b^2 - 4ac}/2a)$.

Figure 4.10 displays the outcomes of textural features extracted from two separately synthesized seasonal images. A comparison of the original encoded seasonal grayscale images and the corresponding synthesized ones shows similar textural patterns reconstructed in both seasons for all Rivers. While visual inspection provides an initial assessment, a standardized criterion (or method) is needed for a more objective comparison between the images. Table 4.9 summarizes SSIM for all rivers.



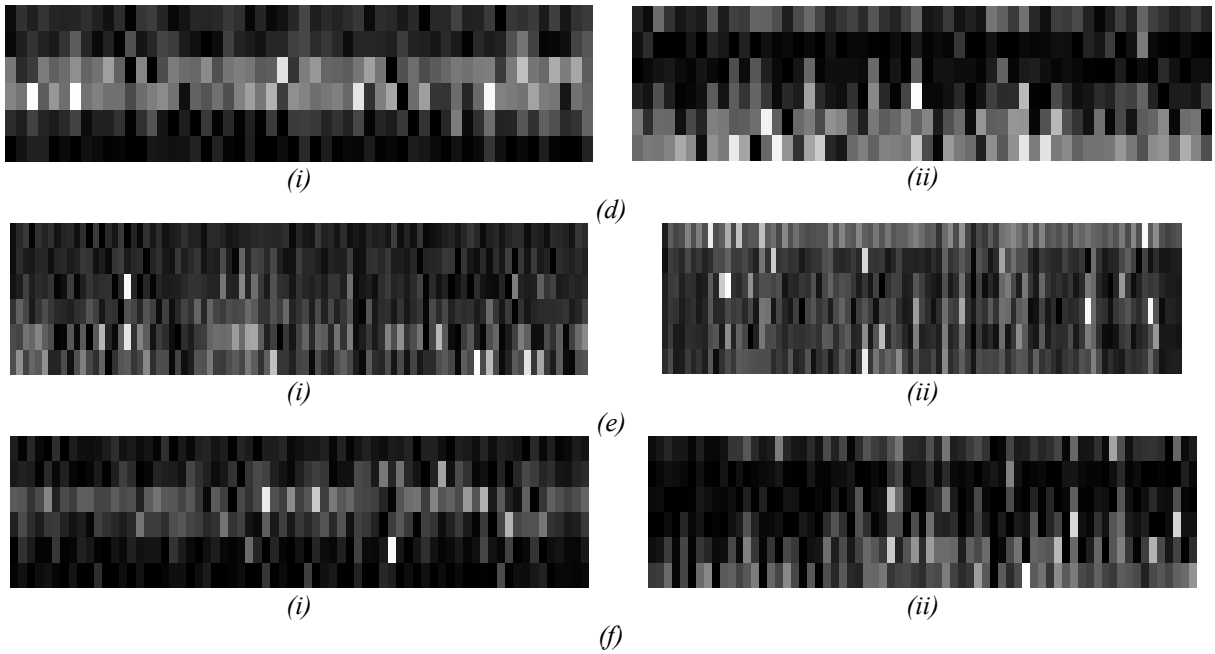



*Figure 4.10. Synthesized encoded seasonal grayscale images for the season (i) k=1 and (ii) k=2 of (a) Little White River near Bellingham, (b) Mattagami River, (c) Root River, (d) Collins Creek, (e) Kaministiquia River, and (f) North Magnetawan River.*

The SSIM values for the six Canadian rivers, summarized in Table 4.9, indicate varying levels of similarity between observed and modelled streamflow patterns. Among the rivers analyzed, the Root River at Sault Ste. Marie (02CA002) exhibits the highest similarity, with SSIM values of 0.768 and 0.835 for the two seasonal segments, thereby suggesting that the proposed modelling and segmentation approach effectively captures both seasonal dynamics and overall temporal variability of patterns. Collins Creek near Kingston (02HM005) also shows relatively high SSIM values (0.745 and 0.809), indicating a strong agreement between the observed monthly streamflow data and its correspondingly transformed textural images.

In contrast, the North Magnetawan River near Burk's Falls (02EA005) shows the lowest similarity with SSIM values of 0.535 and 0.520 for season k = 1 and k = 2, suggesting that the TPRS model may not fully be able to capture the temporal variability or seasonal characteristics of this river. Other rivers, including the Little White River near Bellingham (02CC005), Mattagami River near Timmins (04LA002), and Kaministiquia River at Kaministiquia (02AB006), fall between these two extremes. Overall, the results highlight that while the proposed approach generally performs well across diverse river systems, its accuracy may vary depending on river-specific hydrological characteristics and seasonal flow patterns.

*Table 4.9. Summary of SSIM values for the original and synthesized images.*

| Hydrometric Gauging Station (Watershed) | SSIM Values | |
|---|---|---|
| | k = 1 | k = 2 |
| Little White River near Bellingham (02CC005) | 0.667 | 0.583 |
| Mattagami River near Timmins (04LA002) | 0.601 | 0.742 |
| Root River at Sault Ste. Marie (02CA002) | 0.768 | 0.835 |
| Collins Creek near Kingston (02HM005) | 0.745 | 0.809 |
| Kaministiquia River at Kaministiquia (02AB006) | 0.652 | 0.617 |
| North Magnetawan River near Burk's Falls (02EA005) | 0.535 | 0.520 |

To enable a direct comparison with the ARIMA model, the synthesized textural images were converted back into streamflow time series format. This transformation allows the reconstructed time series to be evaluated using conventional time-domain performance metrics and facilitates a consistent assessment between the proposed image-based approach and the ARIMA-based time series modelling framework.

The fitted ARIMA/SARIMA models, shown in Table 4.10, indicate that monthly streamflow at all six hydrometric stations (or watersheds) is governed by a combination of short-term autocorrelation and pronounced seasonal dependence. For all six rivers, first- and second-order autoregressive [AR (1) and AR (2)] terms are statistically significant, demonstrating that current streamflow is strongly influenced by flows in the previous one to two months. The negative AR coefficients observed for pink noise-dominated watersheds (Little White River, Mattagami River, and Root River) suggest an oscillatory and mean-reverting behaviour, whereas the positive and relatively large AR coefficients for the Black noise dominated watersheds (Collins Creek, Kaministiquia River, and North Magnetawan River) indicate strong persistence and longer memory in these systems. In particular, the AR (1) coefficients are approaching unity for the Collins Creek ($\phi_1 = 0.991$) and the North Magnetawan River ($\phi_1 = 0.908$), which reflect highly persistent dynamics, consistent with catchments that exhibit sustained hydrologic storage and delayed runoff responses.

Seasonal moving average components further highlight the dominant role of annual periodicity in monthly streamflow variability. All watersheds include a seasonal MA term at lag 12 months, confirming the importance of yearly hydrologic cycles associated with snow accumulation and melt, precipitation seasonality, and evapotranspiration patterns. The Mattagami River additionally requires a second seasonal MA component at lag 24 months, suggesting a more complex multi-

year seasonal structure. The existence of generally small standard errors associated with both AR and seasonal MA parameters indicates robust parameter estimation and good model stability. Overall, the fitted models effectively capture both short-term dependence and strong seasonal controls, supporting their suitability for representing the temporal structure of monthly streamflow and for use as benchmarks in comparison with synthesized or alternative modelling approaches.

*Table 4.10. A summary of the optimized parametric estimates of ARIMA for six Ontario Stations.*

| Hydrometric Station | Model* | Parameter | SE Coefficient |
|---|---|---|---|
| Little White River near Bellingham (02CC005) | AR (1) | $\phi_1$= -0.691 | 0.029 |
| | AR (2) | $\phi_2$= -0.394 | 0.029 |
| | Seasonal MA (12) | $\Theta_1$= 0.394 | 0.021 |
| Mattagami River near Timmins (04LA002) | AR (1) | $\phi_1$= -0.652 | 0.039 |
| | AR (2) | $\phi_2$= -0.269 | 0.039 |
| | Seasonal MA (12) | $\Theta_1$= 0.524 | 0.001 |
| | Seasonal MA (24) | $\Theta_2$= 0.469 | 0.004 |
| Root River at Sault Ste. Marie (02CA002) | AR (1) | $\phi_1$= -0.453 | 0.040 |
| | AR (2) | $\phi_2$= -0.216 | 0.039 |
| | Seasonal MA (12) | $\Theta_1$= - 0.086 | 0.041 |
| Collins Creek near Kingston (02HM005) | AR (1) | $\phi_1$= 0.991 | 0.128 |
| | AR (2) | $\phi_2$= 0.394 | 0.082 |
| | MA (1) | $\Phi_1$= 0.524 | 0.140 |
| | Seasonal MA (12) | $\Theta_1$= 0.469 | 0.0374 |
| Kaministiquia River at Kaministiquia (02AB006) | AR (1) | $\phi_1$= 0.844 | 0.0275 |
| | AR (2) | $\phi_2$= 0.394 | 0.0275 |
| | Seasonal MA (12) | $\Theta_1$= - 0.219 | 0.0299 |
| North Magnetawan River near Burk's Falls (02EA005) | AR (1) | $\phi_1$= 0.908 | 0.0308 |
| | AR (2) | $\phi_2$= 0.467 | 0.0308 |
| | Seasonal MA (12) | $\Theta_1$= -0.290 | 0.001 |

A comparative analysis was conducted between the TPRS and the AR-family of models, summarized in Table 4.11. This table presents a comparison of key statistical properties between the historical records and the synthetic time series generated by the TPRS and the ARIMA models. Overall, the TPRS demonstrates a strong ability to reproduce the first four moments of the observed streamflow distributions, including mean, standard deviation, skewness, and kurtosis, across all six watersheds analyzed in this thesis. In most cases, the TPRS-synthesized time series closely match the historical means and standard deviations, indicating that the TPRS effectively preserves both central tendency and overall variability. For example, for the Little White River and Collins Creek, the results obtained by TPRS show near-identical means and variances relative to the historical data, highlighting the robustness of the texture-based approach in maintaining basic statistical structure.

In contrast, while the ARIMA models generally reproduce the historical means with reasonable accuracy, they show notable discrepancies in higher-order moments and variability for several watersheds. In particular, the ARIMA family of models substantially underestimates the standard deviation for Collins Creek, Kaministiquia River, and North Magnetawan River, indicating an inability to fully capture flow variability and extreme events. This underestimation is also reflected in reduced skewness and kurtosis for some watersheds, suggesting that the ARIMA family of models tends to smooth the time series and dampen heavy-tailed behaviour. Conversely, in some cases, such as the Kaministiquia River, ARIMA models overestimate skewness and kurtosis, implying excessive representation of extremes. These results suggest that in comparison to ARIMA, the TPRS provides a more consistent reproduction of both low-order and higher-order statistical properties, supporting its suitability for synthetic streamflow generation and texture-based temporal pattern modelling.

The results suggest that for the first three watersheds (Table 4.10), which are characterized by dominantly pink-noise behaviour, the ARIMA and the TPRS approaches exhibit comparable performance in reproducing the key statistical properties. In contrast, for the remaining watersheds, which exhibit Black noise dominance, the ARIMA models show reduced capability in capturing variability and higher-order moments, whereas the TPRS maintains stronger agreement with the historical statistics. These findings indicate that for streamflow time series with dominant pink-noise characteristics, ARIMA modelling may be sufficient, while for more complex hydrologic regimes with dominance of Black noise, the texture-based TPRS provides clear advantages.

At this stage, the TPRS is compared with traditional time series modelling approaches to further confirm its effectiveness and robustness. For pink noise, to clarify, the $\varphi_1$ values remain relatively consistent across AR-family models, suggesting that a simple AR-type model is sufficient to capture the temporal structure. In contrast, the presence of black noise exhibits stronger temporal correlations, and while a textural pattern could have been used to describe its behaviour, such an option was not exercised. The rationale for such an option is that if AR-family models can adequately represent the observed dynamics, there is no need to introduce additional complexity. For brown noise, the process can be represented using a Brownian model; although geometric Brownian motion or a textural pattern could capture more complex behaviour, the AR-family still provides a reasonable description, and the added complexity of textural patterns is unnecessary.

*Table 4.11. Comparison of statistical properties between historical and synthetic streamflow series generated by the textural pattern recognition system (TPRS) and ARIMA models.*

| River | Sample | Statistical Properties | | | |
|---|---|---|---|---|---|
| | | **Mean ($m^3/s$)** | **Std. Dev. ($m^3/s$)** | **Skewness** | **Kurtoses** |
| Little White River near Bellingham (02CC005) | Historical | 27.12 | 24.50 | 1.96 | 3.96 |
| | Synthesized by TPRS | 26.51 | 24.10 | 1.60 | 3.65 |
| | Synthesized by ARIMA | 28.28 | 28.90 | 1.89 | 3.67 |
| Mattagami River near Timmins (04LA002) | Historical | 65.65 | 41.30 | 2.00 | 5.88 |
| | Synthesized by TPRS | 66.24 | 42.37 | 1.93 | 6.08 |
| | Synthesized by ARIMA | 65.50 | 40.17 | 2.04 | 6.14 |
| Root River at Sault Ste. Marie (02CA002) | Historical | 2.37 | 1.20 | 2.09 | 6.68 |
| | Synthesized by TPRS | 1.90 | 1.96 | 2.03 | 5.26 |
| | Synthesized by ARIMA | 1.98 | 1.35 | 2.01 | 6.49 |
| Collins Creek near Kingston (02HM005) | Historical | 2.59 | 2.74 | 1.51 | 2.59 |
| | Synthetic by TPRS | 2.53 | 2.54 | 1.56 | 2.55 |
| | Synthetic by ARIMA | 2.59 | 1.28 | 0.80 | 1.23 |
| Kaministiquia River at Kaministiquia (02AB006) | Historical | 59.01 | 36.80 | 2.86 | 12.37 |
| | Synthesized by TPRS | 58.41 | 38.52 | 2.14 | 7.71 |
| | Synthesized by ARIMA | 59.22 | 17.57 | 3.14 | 14.75 |
| North Magnetawan River near Burk's Falls (02EA005) | Historical | 4.51 | 5.72 | 2.18 | 6.63 |
| | Synthesized by TPRS | 4.45 | 5.09 | 2.23 | 8.42 |
| | Synthesized by ARIMA | 4.50 | 1.86 | 1.50 | 4.71 |

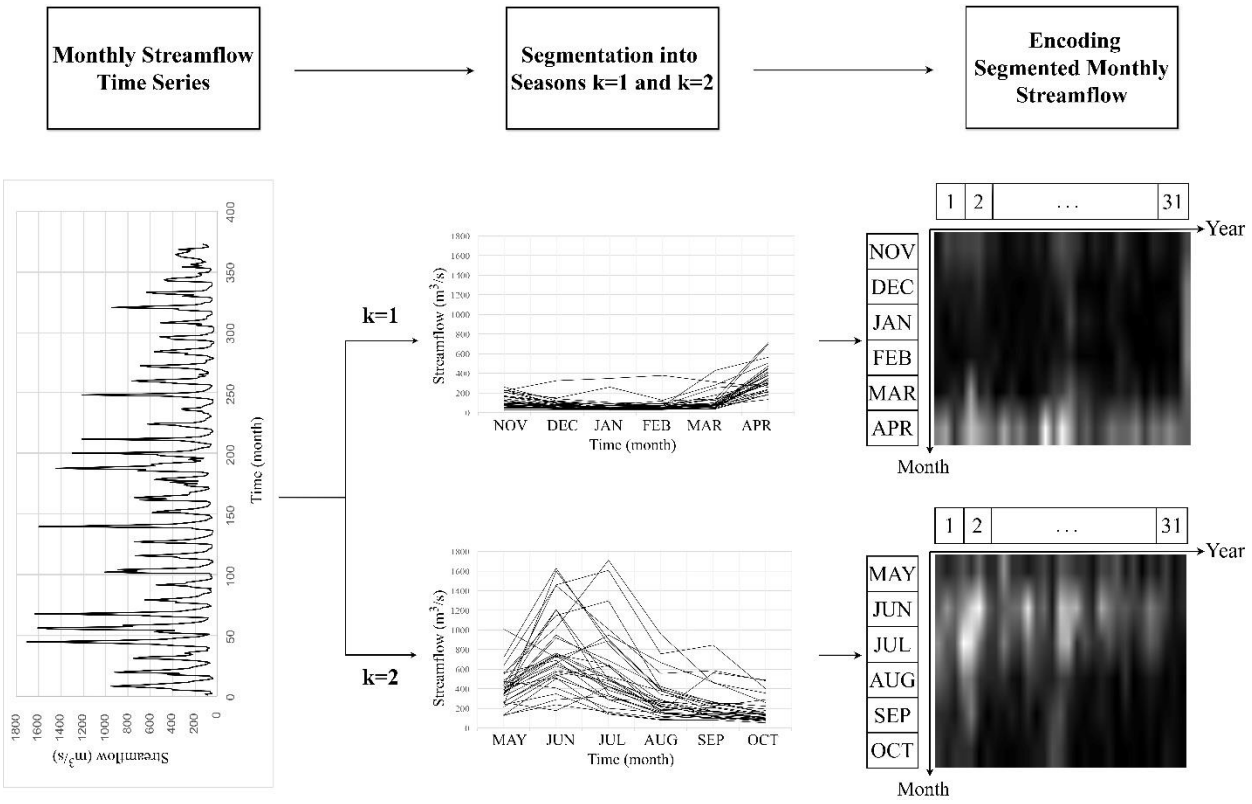
### 4.3.3 Comparison between the Textural Pattern Recognition System and Traditional Pattern Recognition Systems

This section aims to compare the proposed TPRS model with the Traditional Pattern Recognition System. For clarity, the detailed results of the TPRS are presented first, followed by a comparative evaluation of the two approaches. In addition, Appendix 4C presents a manual calculation of the textural image synthesis using TPRS for the Black River over a five-year period.

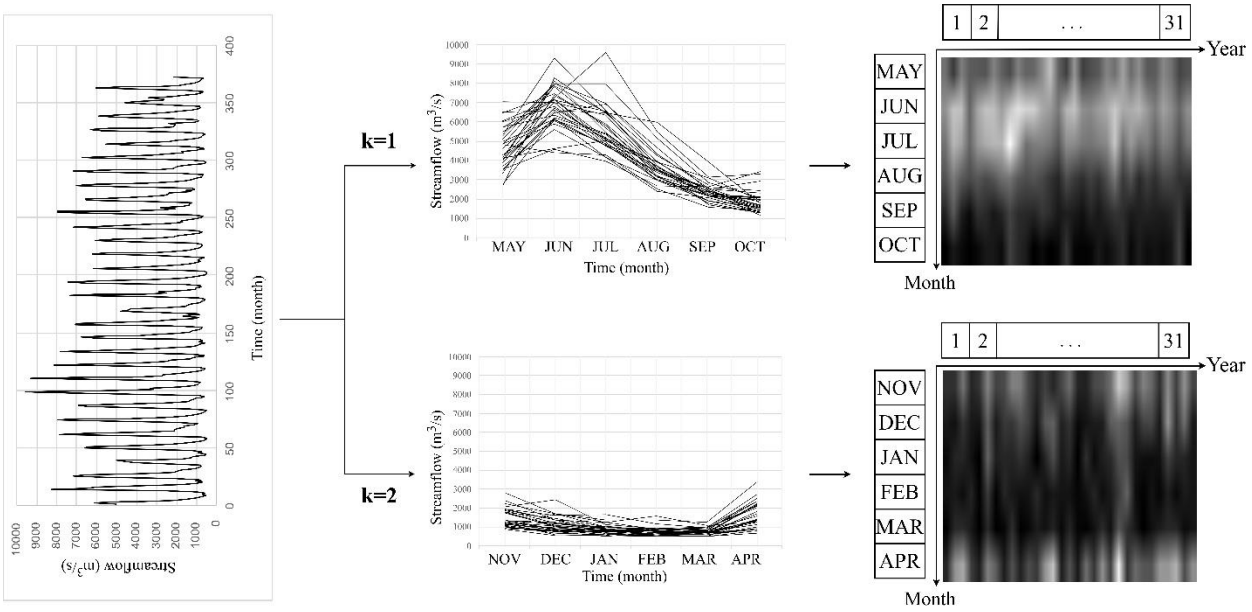
The first step in the proposed methodology is data segmentation. Monthly streamflow data from six Canadian rivers (South Saskatchewan, Black, Fraser, N.E.M, St. John, and Grand) were divided into two seasonal segments (k = 1 and k = 2) following Panu et al. (1978). Each seasonal segment, consisting of six-monthly values, was encoded into 8-bit grayscale textural images (Figure 4.11), with each image comprising 186 pixels arranged in 6 rows (months) and 31 columns (years), including one additional year for accommodating the requirement of the shifting and/or lagging operations.

Figure 4.11 illustrates distinct streamflow texture patterns among the rivers. The St. John River exhibits consistent patterns across years, whereas the N.E.M and Grand Rivers show greater variability, reflected by disrupted grayscale textures in season k = 1 and more stable patterns in season k = 2, including a bright row in May associated with peak flows. Recurring low-intensity patterns suggest persistent dry periods in the South Saskatchewan and Black Rivers, while the Fraser River exhibits cyclical behaviour and localized high-intensity regions indicative of anomalous flash flood events.

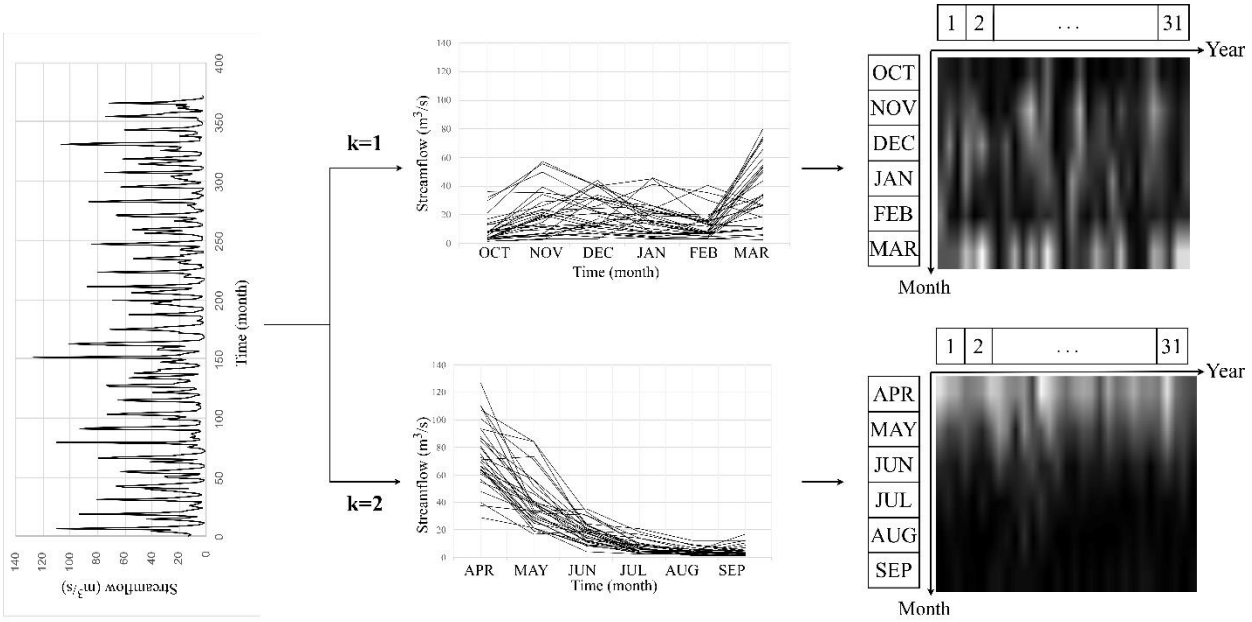
Textural features were extracted using a semi-automated approach based on lag-$\zeta$ (= 1) dependency analysis. Horizontal and vertical lagging were applied to compute the $SACF^k$ for each season. Due to lagging, the first row was excluded in the horizontal direction, and an additional year of data was included for vertical analysis. As a result, $SACF^k$ Matrices for k = 1 and k = 2 were constructed with dimensions of 5 × 30, representing the spatial dependence among adjacent pixels within each seasonal grayscale image.



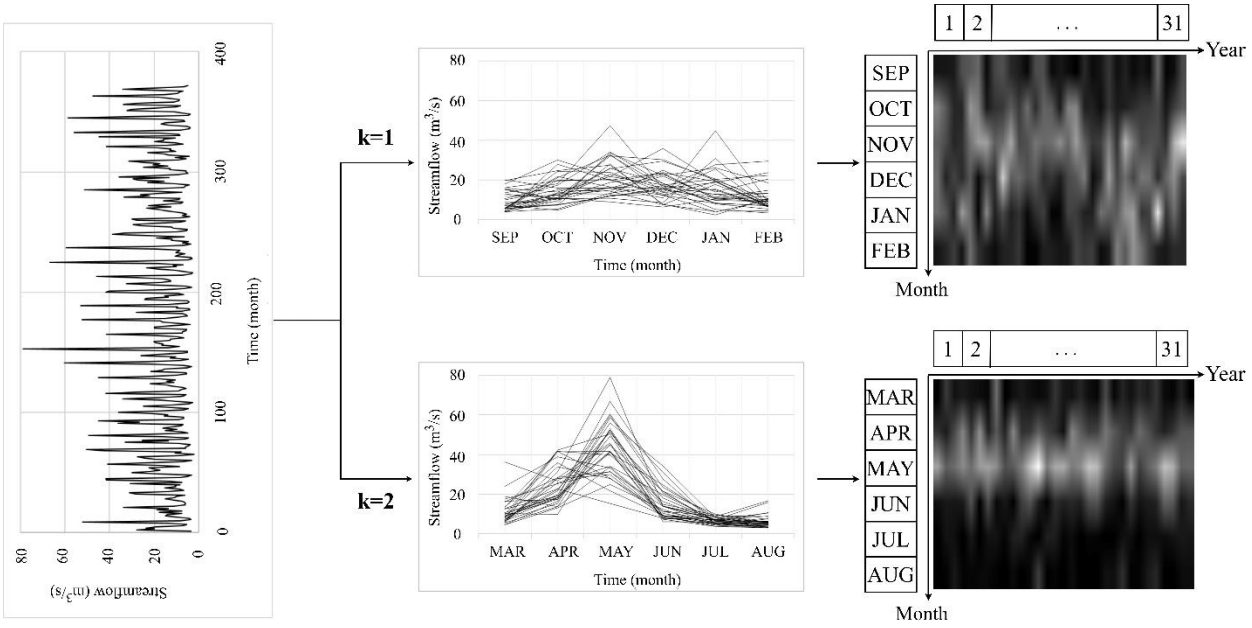

*(a) S. Saskatoon River (Nov 1912 - Oct 1942)*



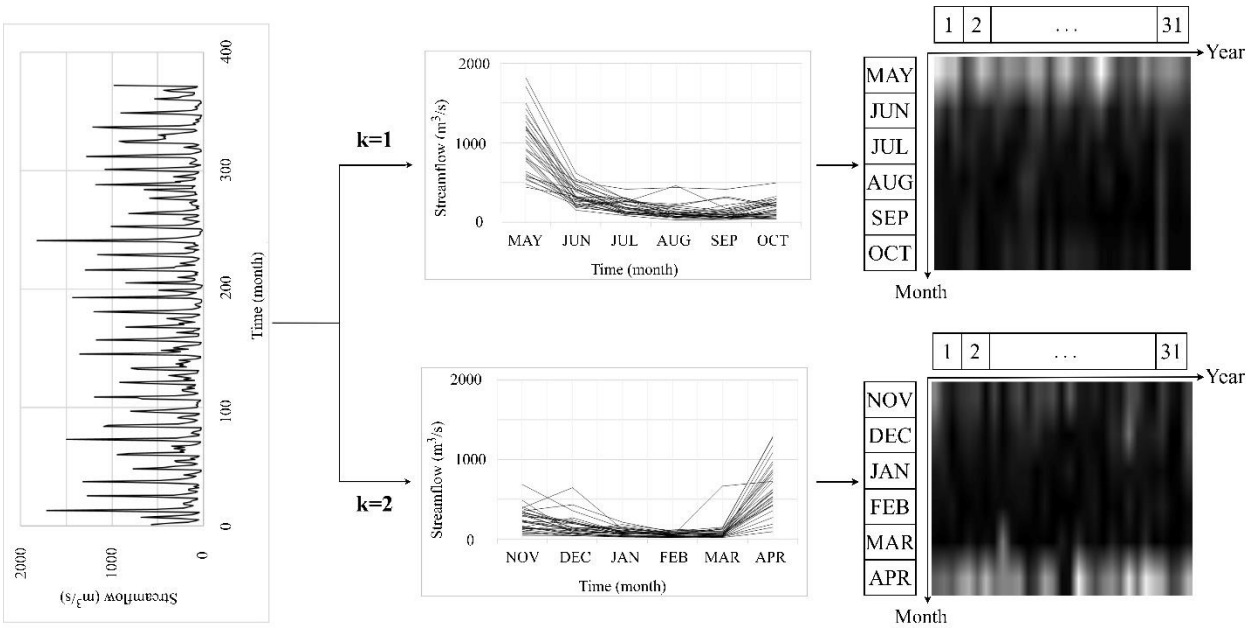

*(b) Fraser River (May 1912-Apr 1943)*



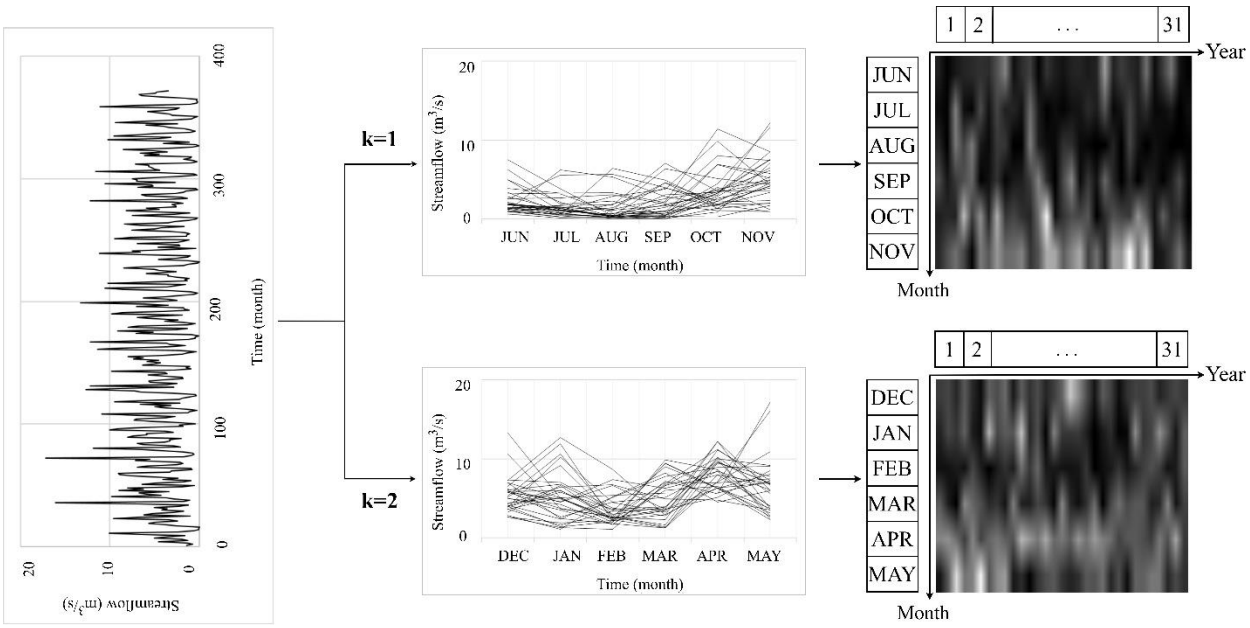

*(c) Black River (Oct 1915- Sep 1946)*




*(d) N.E.M (Sep. 1929-Aug 1960)*




*(e) St. John (May 1926-April 1957)*




*(f) Grand River (June 1920-May 1951)*

*Figure 4.11. Time Series data representation for season k = 1 and k = 2 of (a) S. Saskatchewan River (Nov 1911-Oct 1942), (b) Fraser River (May 1912-Apr 1943), (c) Black River (Oct 1915- Sep 1946), (d) N.E.M (Sep. 1929-Aug 1960), (e) St. John (May 1926-April 1957) and (f) Grand River (June 1920-May 1951).*

Once the intra-pattern analysis was accomplished, a Markovian transition probability was computed and $SACF^k$ becomes a matrix with 6×30 dimensions. Upon the transformation of these matrices into the frequency domain using a DFT, it is feasible to compute the power spectrum matrices for each season (k = 1 and k = 2), hereafter referred to as the seasonal textural features. Once the power spectrum is obtained, it is divided into high- and low-power components using a cut-off point (such as the mean of the power spectrum). The number of times the power spectrum falls in the high- or low-power spectrum group (Table 4.12). These numbers are later used to get the probability of occurrence. The statistical properties of seasonal textural image features of the N.E.M. River are also summarized in Table 4.13.

*Table 4.12. The frequency count indicates the occurrences of the power spectrum within the High- or Low-power spectrum groups of the N.E.M River.*

| Power Spectrum Groups | Season k=1 | | | | | | Season k=2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Seasonal Pattern Elements | | | | | | Seasonal Pattern Elements | | | | | |
| | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apl | May | Jun | Jul | Aug |
| High | 20* | 18 | 20 | 21 | 20 | 18 | 22 | 21 | 22 | 19 | 22 | 21 |
| Low | 11** | 13 | 11 | 10 | 11 | 13 | 9 | 10 | 9 | 12 | 9 | 10 |

Note: where * and ** respectively represent the number of occurrences in high and low groups.

*Table 4.13. Statistical properties of the extracted textural features of the N.E.M River.*

| Statistics | Season k = 1 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Low-Power Spectrum ($R_{PL}^{k=1}$ and $St_{PL}^{k=1}$) | | | | | | High-Power Spectrum ($R_{PH}^{k=1}$ and $St_{PH}^{k=1}$) | | | | | |
| | Seasonal Pattern Elements | | | | | | Seasonal Pattern Elements | | | | | |
| | Sep | Oct | Nov | Dec | Jan | Feb | Sep | Oct | Nov | Dec | Jan | Feb |
| Mean ($m^3/s$) | 23.4 | 22.0 | 22.2 | 24.1 | 26.2 | 24.2 | 112.6 | 136.3 | 108.4 | 143.7 | 108.4 | 119.3 |
| Std. Dev. ($m^3/s$) | 17.5 | 15.4 | 15.1 | 23.0 | 13.8 | 14.5 | 40.9 | 39.6 | 46.6 | 57.6 | 46.6 | 46.1 |
| **Statistics** | **Season k = 2** | | | | | | | | | | | |
| | Low-Power Spectrum ($R_{PL}^{k=2}$ and $St_{PL}^{k=2}$) | | | | | | High-Power Spectrum ($R_{PH}^{k=2}$ and $St_{PH}^{k=2}$) | | | | | |
| | Seasonal Pattern Elements | | | | | | Seasonal Pattern Elements | | | | | |
| | Mar | Apl | May | Jun | Jul | Aug | Mar | Apl | May | Jun | Jul | Aug |
| Mean ($m^3/s$) | 32.0 | 36.2 | 33.9 | 36.8 | 34.7 | 38.0 | 169.5 | 139.8 | 141.2 | 156.7 | 142.3 | 139.8 |
| Std. Dev. ($m^3/s$) | 25.0 | 22.1 | 17.5 | 16.5 | 17.5 | 19.7 | 83.1 | 45.3 | 51.8 | 58.5 | 46.8 | 45.3 |

The likelihood of higher-power spectrum values corresponding to major patterns is greater than that of lower-power spectrum values corresponding to minor patterns, as displayed in Table 4.12. Such results suggest that the power spectra of signals in both seasons have a greater concentration of energy in the high-power range. It is also known that the presence of a higher power spectrum

indicates that there is more energy in the frequencies of signals normally corresponding to the more significantly prominent patterns within the dataset.

Further insights into the power spectrum matrix are depicted in *Figure 4.12*, illustrating a normalized histogram when comparing the low-power-spectrum and high-power-spectrum groups. It shows the relative likelihood of occurrence of the power spectrum falling within each interval. For all months in both seasons (k=1 and k=2), the high-power spectrum has a lower probability density due to its spread over a wider range of values along the x-axis and thus results in a flatter graph corresponding to the total area under the graph being equal to 1. On the other hand, the presence of a sharper peak of the PDF in the low-power spectrum suggests that the data values are more concentrated around specific values in a narrower range.

An approximation of the underlying Gaussian distribution of the power spectrum enables the generation (or synthesis) of values for the power spectrum using the standard Gaussian probability function. The process involves the generation of random numbers in the range of 0 and 1 and a decision about the mean and standard deviation. If this random number is lower than the probability of occurrence of a low-power spectrum group for each month obtained from Eq. 3.22), the first random generator generates a random value belonging to the Gaussian distribution approximated to the low-power spectrum (for N.E.M River, it is shown in *Figure 4.12 a-i* and *Figure 4.12 b-i*), otherwise the generator generates a random value belonging to the Gaussian distribution approximated to the high-power spectrum (for N.E.M River, it is shown in *Figure 4.12 a-ii* and *Figure 4.12 b-ii*). For each pixel $(\mathrm{i},\mathrm{j})$, the generator power spectrum must be paired with the reference vector of $\mathrm{R}_{\mathrm{PH}}^{\mathrm{k}} \pm \mathrm{C} \times \mathrm{St}_{\mathrm{PH}}^{\mathrm{k}}$ and $\mathrm{R}_{\mathrm{PL}}^{\mathrm{k}} \pm \mathrm{C} \times \mathrm{St}_{\mathrm{PL}}^{\mathrm{k}}$ in the low-power group (obtained using the results of Table 4.13). If yes, the second random generator is activated to generate the phase component. These synthesized power spectrum values were used as amplitudes (from the first generator), and newly generated phase information (from the second generator) was added to form the Fourier coefficient matrices for each season across the three rivers.

Once the Fourier coefficients were generated for each season, image intensity was computed using a quadratic equation with two positive roots. To ensure adherence to the original image properties, a specific constraint was necessary. The selection of the root directly impacted the pixel intensity of the synthesized image. Hence, an alternative method was needed to determine the root selection based on the original pixel intensity of a seasonal image. If the original pixel intensity of a seasonal

image being synthesized exceeded the mean intensity of all seasonal pixels, the positive sign was chosen ($-b + \sqrt{b^2 - 4ac}/2a$). Conversely, the negative sign was employed ($-b - \sqrt{b^2 - 4ac}/2a$).

*Figure 4.13* displays the outcomes of textural features extracted from two separately synthesized seasonal images. A comparison of the original encoded seasonal grayscale images and the corresponding synthesized ones shows similar textural patterns reconstructed in both seasons for the N.E.M. River and the Grand River. While visual inspection provides an initial assessment, a standardized criterion (or method) is needed for a more objective comparison between the images. Table 4.14 summarizes the SSIM for all six rivers. This table reveals that in both seasons, for the N.E.M River, there is notable similarity, with season k = 2 showing a higher SSIM index, indicating a higher resemblance. On the other hand, the St. John River exhibits a lower similarity in season k = 1 and notably higher similarity in season k = 2. The Grand River shows moderate similarity, with slightly higher values of SSIM in season k = 2 compared to season k = 1. In general, the synthesized images in season k = 2 exhibit greater structural similarity to the original images. across all rivers.

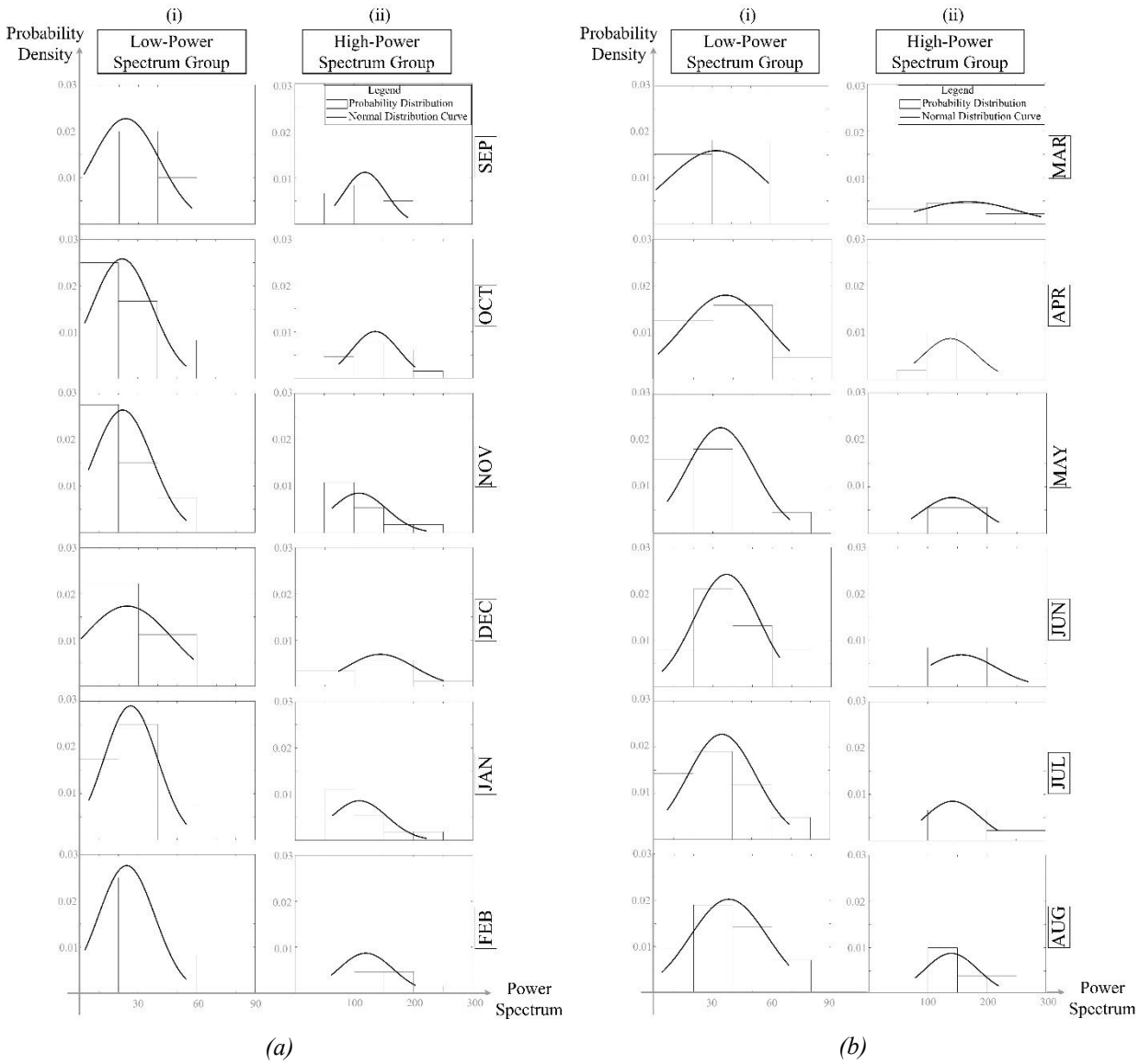
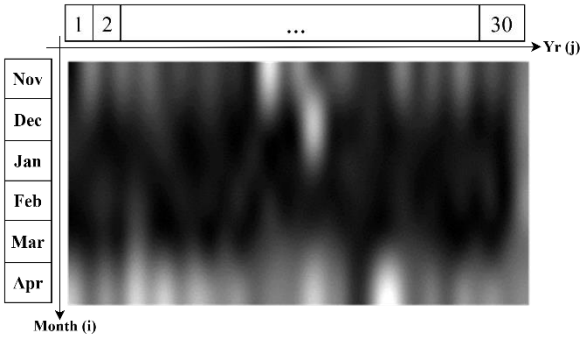
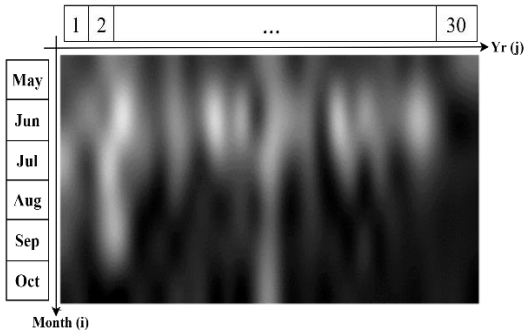



*Figure 4.12. Probability Distribution Functions for the N.E.M River in various seasons (a) Season k = 1 (Sep-Feb) and (b) Season k=2: Mar-Aug for **(i)** Low-power-spectrum group and **(ii)** High-power-spectrum group.*

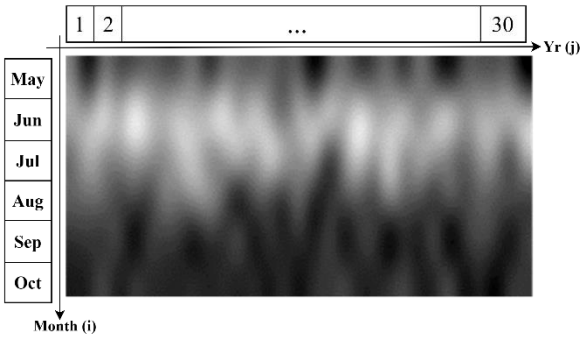



*(i) Season k=1*

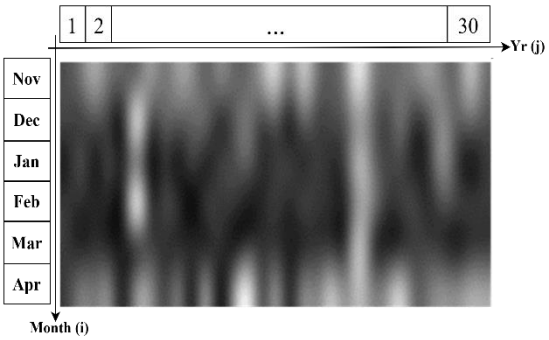



*(ii) Season k=2*

*(a) Saskatchewan River (Nov 1911- Oct 1941)*

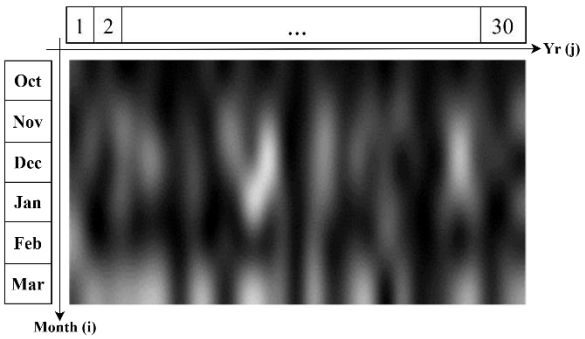



*(i) Season k=1*

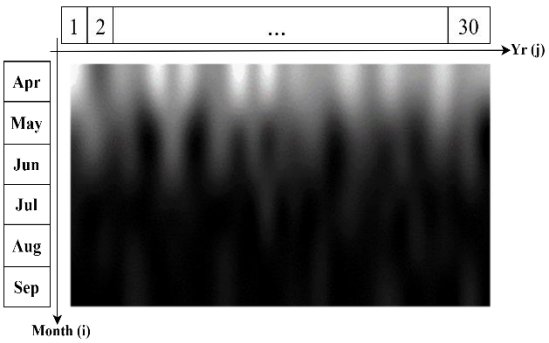



*(ii) Season k=2*

*b) Fraser River (May 1913-Apr 1943)*

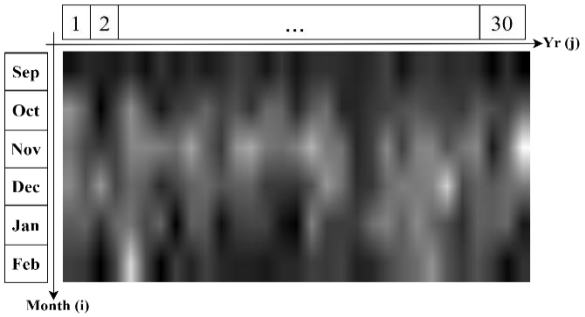



*(i) Season k=1*

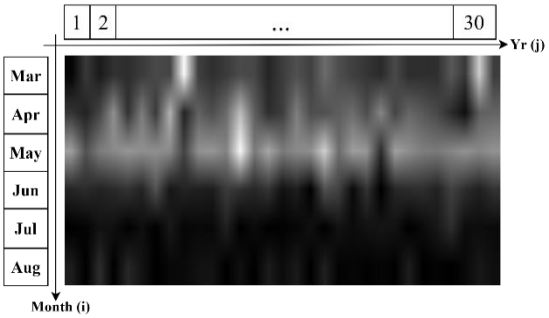



*(ii) Season k=2*

*c) Black River (Oct 1916-Sep 1946)*

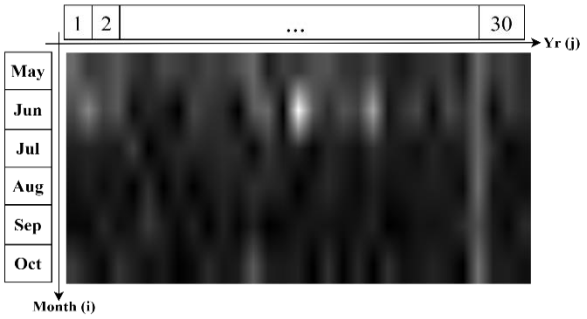



*(i) Season k=1*

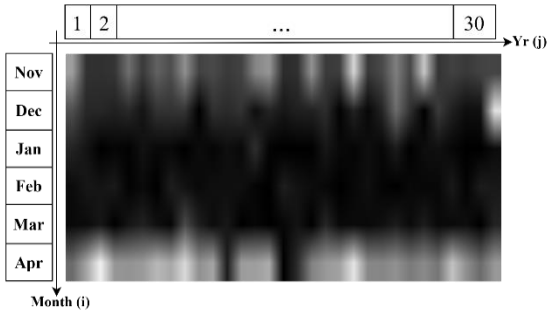



*(ii) Season k=2*

*d) NEM River (Sep 1930-Aug 1960)*

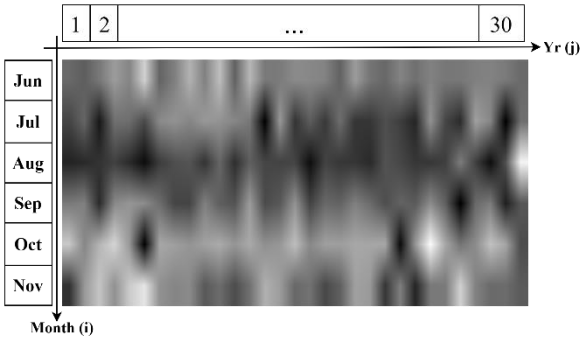



*(i) Season, k=1*

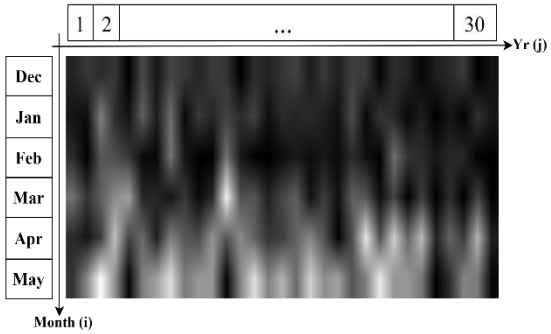



*(ii) Season, k=2*

*e) St. John River (May 1927-Apr 1957)*




*(i) Season, k=1*




*(ii) Season, k=2*

*f) Grand River (June 1921-May 1951)*

*Figure 4.13. Synthesized encoded seasonal grayscale images for the season i) k=1 and ii) k=2 of (a) S. Saskatchewan River (Nov 1912-Oct 1942), (b) Fraser River (May 1913-Apr 1943), (c) Black River (Oct 1916- Sep 1946), (d) N.E.M (Sep. 1930-Aug1960), (e) St. John (May 1927-April 1957), and (f) Grand River (June 1921-May 1951) rivers.*

*Table 4.14. Summary of SSIM values for the original and synthesized images.*

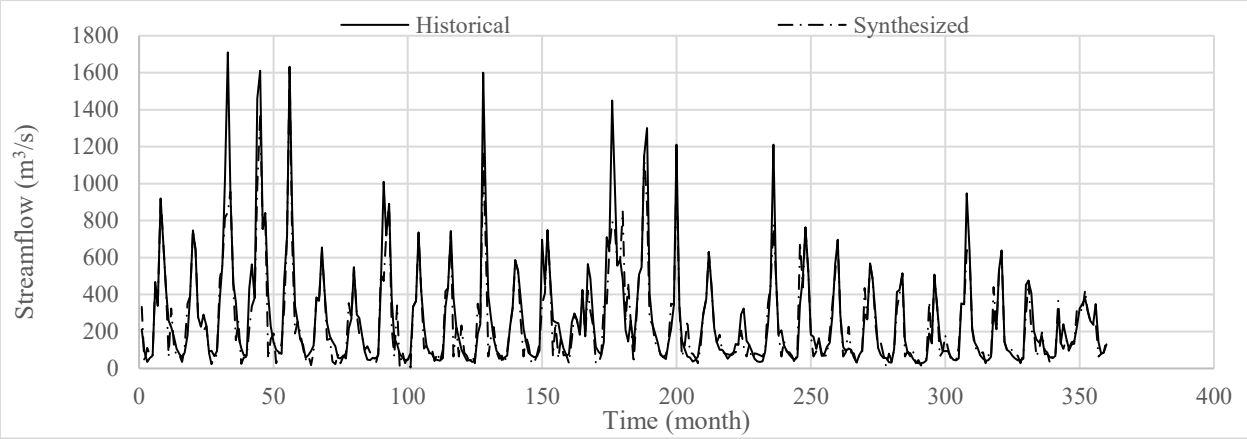
| River | SSIM Values | |
|---|---|---|
| | k = 1 | k = 2 |
| South Saskatchewan | 0.698 | 0.720 |
| Fraser | 0.726 | 0.654 |
| Black | 0.701 | 0.800 |
| N.E.M | 0.64 | 0.73 |
| St. John | 0.40 | 0.62 |
| Grand | 0.55 | 0.65 |

The final stage of the TPRS involves transforming the synthesized images back into streamflow time series. This transformation occurs by reversing the computed seasonal matrices of $SACF^{k}$ into seasonal segments for both seasons and then reorganizing these segments into successive geophysical years, forming the desired synthesized streamflow realization. For assessing the

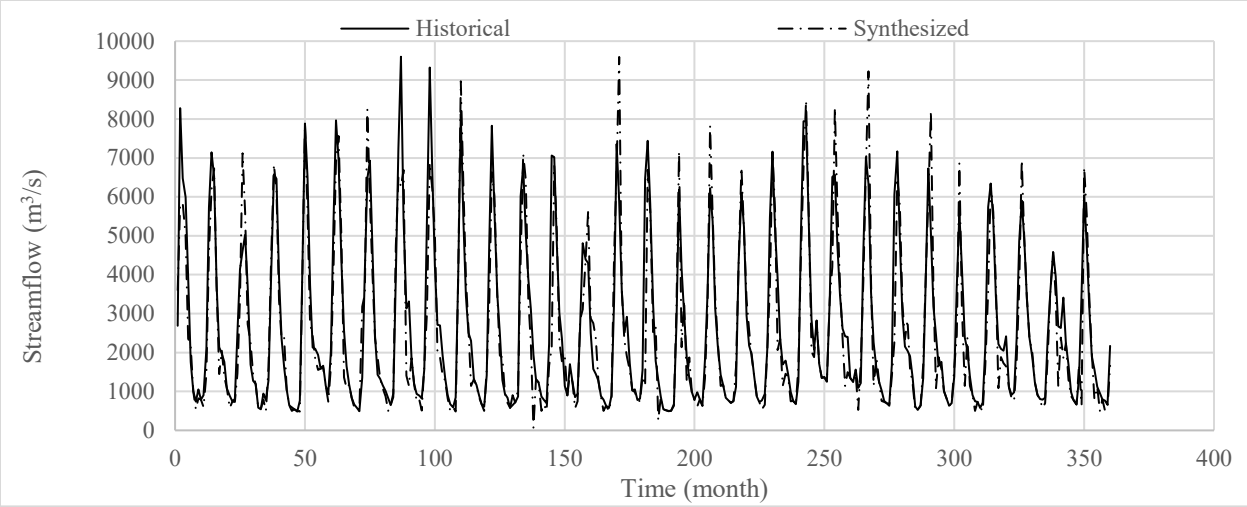
effectiveness of the TPRS for streamflow synthesis, the synthesized streamflow time series can be statistically and/or graphically compared to the original historical streamflow data. *Figure 4.14* illustrates a comparison between the historical streamflow time series and the average of 30 synthesized realizations using the TPRS model.

Upon further analysis of the statistical properties of both historical and synthesized samples, a resemblance is evident based on the summary statistics provided (Table 4.15). Similar values of the Permutation Entropy across all rivers between synthetic realizations and historical streamflow, especially in the N.E.M River and Grand River, endorse the assertion that the TPRS model adeptly captures the inherent patterns and dynamics of the historical streamflow time series. However, the results also indicate that the TPRS model experienced some challenges in effectively replicating the characteristics of the St. John River. For example, the comparison between the historical and synthetic data for the St. John River showcases noticeable differences in various statistical properties.

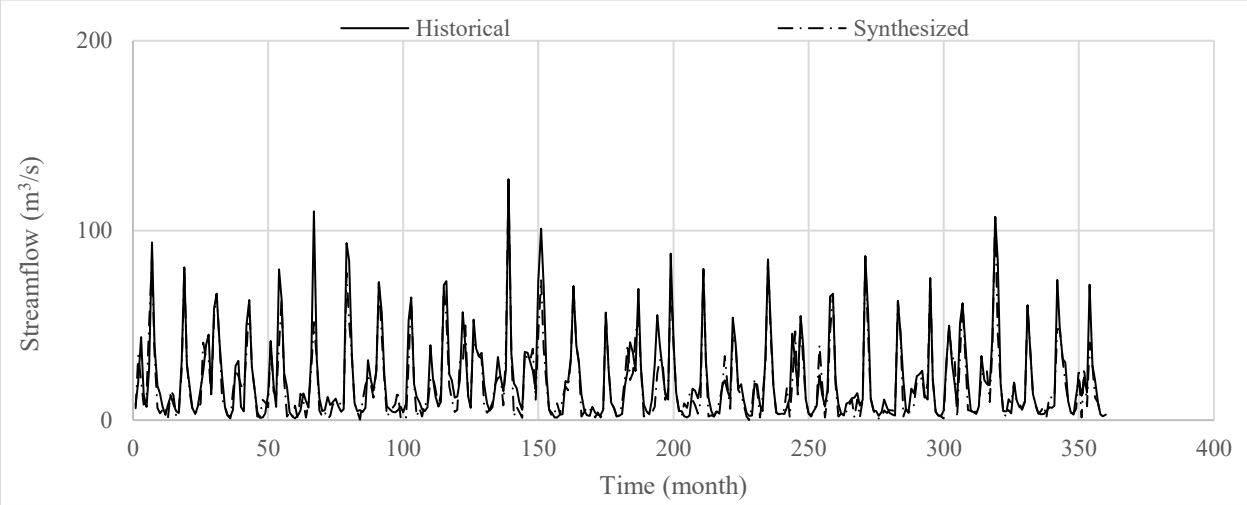
To ascertain whether the synthesized realization and historical data belong to the same population, a null hypothesis test was conducted, and the corresponding p-values have been presented in Table 4.16. The null hypothesis suggests that the statistical properties, including mean, standard deviation, skewness, and kurtosis, between the historical (Hist.) and synthesized (Syn.) datasets for each river are equal or indistinguishable. The reported p-values, all higher than the conventional significance level of 0.05, support the null hypothesis, suggesting that there exists no significant difference in the statistical properties between the historical and synthesized datasets for all six rivers.



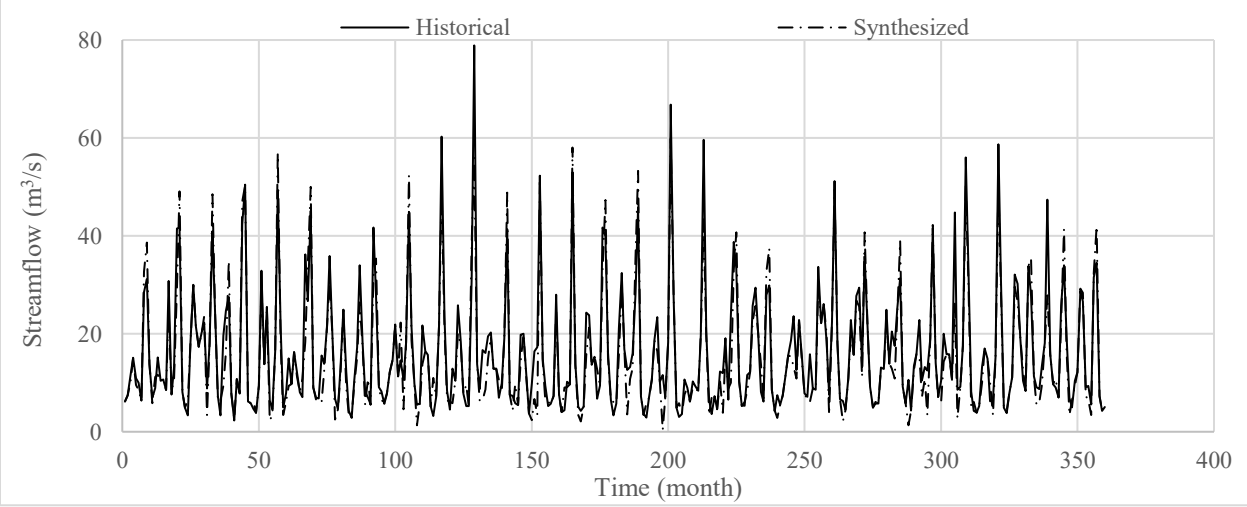

*a) S. Saskatchewan River (Nov 1912 - Oct 1942).*



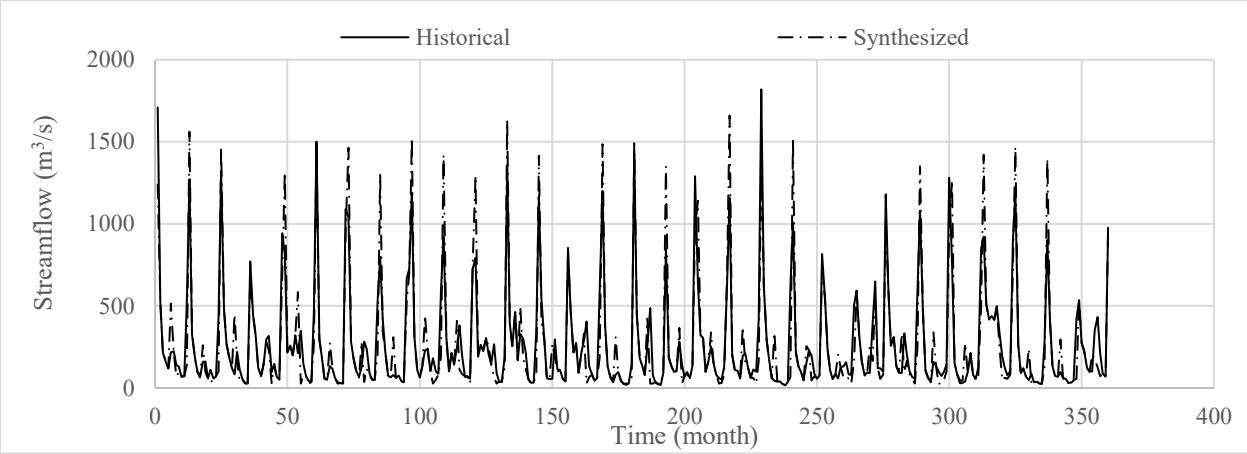

*b) Fraser River (May 1913 - Apr 1943)*



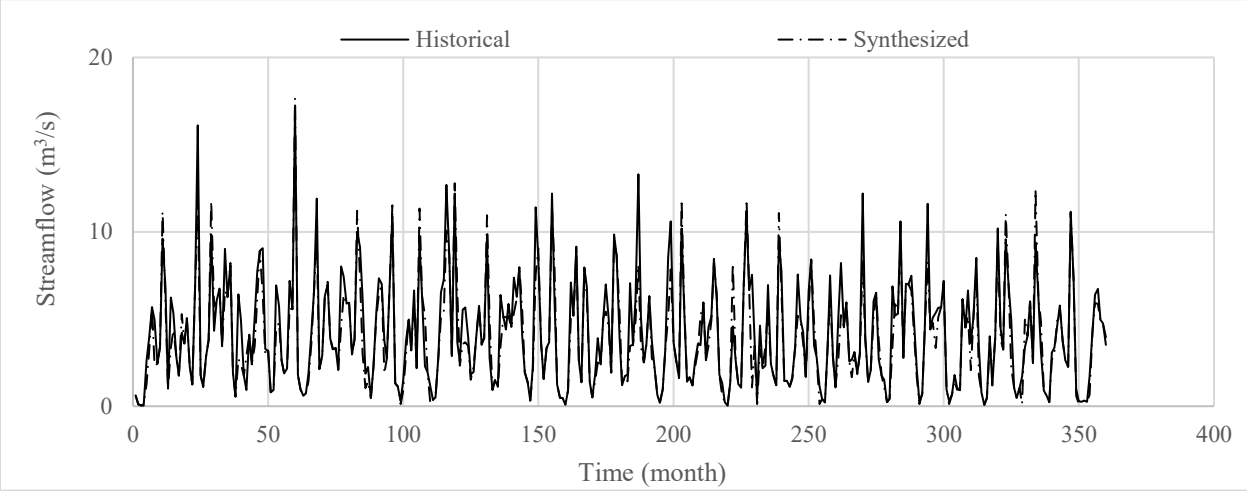

*c) Black River (Oct. 1916 – Sep. 1946)*




*(d) NEM River (Sep. 1930 - Aug. 1960).*




*(e) St. John River (May 1927 - Apr 1957).*




*(f) Grand River (June 1921 - May 1951).*

*Figure 4.14. Historical streamflow and the average of 30 synthesized realizations of (a) S. Saskatchewan River (Nov 1912-Oct 1942), (b) Fraser River (May 1913 - Apr 1943), (c) Black River (Oct 1916 - Sep 1946), (d) N.E.M River (Sep. 1930 – Aug 1960), (e) St. John River (May 1927- April 1957), and (f) Grand River (June 1921-May 1951).*

*Table 4.15. A Summary of Statistical Parameters of the Historical and Synthesized Streamflow.*

| River | Sample | Statistical Properties | | | | |
|---|---|---|---|---|---|---|
| | | Mean ($m^3/s$) | Std. Dev. ($m^3/s$) | Skewness | Kurtoses | Permutation Entropy |
| South Saskatchewan | Historical | 272.9 | 292.61 | 2.29 | 8.31 | 1.020 |
| | Synthetic* | 281.43 | 300.56 | 1.79 | 7.41 | 1.083 |
| Fraser | Historical | 2644.13 | 2101.49 | 1.09 | 3.17 | 0.926 |
| | Synthetic* | 2523.2 | 2075.2 | 1.23 | 3.52 | 1.011 |
| Black | Historical | 21.79 | 22.92 | 1.73 | 5.24 | 1.039 |
| | Synthetic* | 21.06 | 23.71 | 1.84 | 5.39 | 1.078 |
| N.E.M | Historical | 16.224 | 12.370 | 1.717 | 6.432 | 1.082 |
| | Synthetic* | 16.955 | 13.184 | 1.561 | 6.032 | 1.087 |
| St. John | Historical | 268.303 | 316.149 | 2.201 | 7.914 | 1.032 |
| | Synthetic* | 216.114 | 221.242 | 1.796 | 6.463 | 1.011 |
| Grand | Historical | 4.188 | 3.094 | 0.945 | 3.852 | 1.081 |
| | Synthetic* | 4.178 | 2.879 | 0.439 | 2.478 | 1.086 |

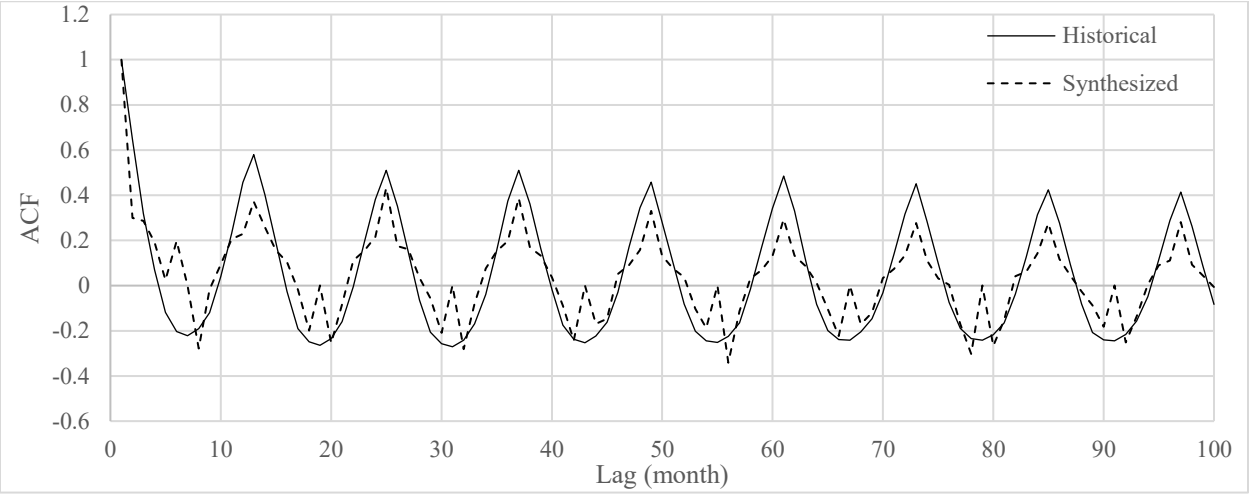
Note: *Synthetic implies the average of 30 synthesized realizations.

*Table 4.16. A Summary of Results of the Null Hypothesis for Various Statistical Parameters.*

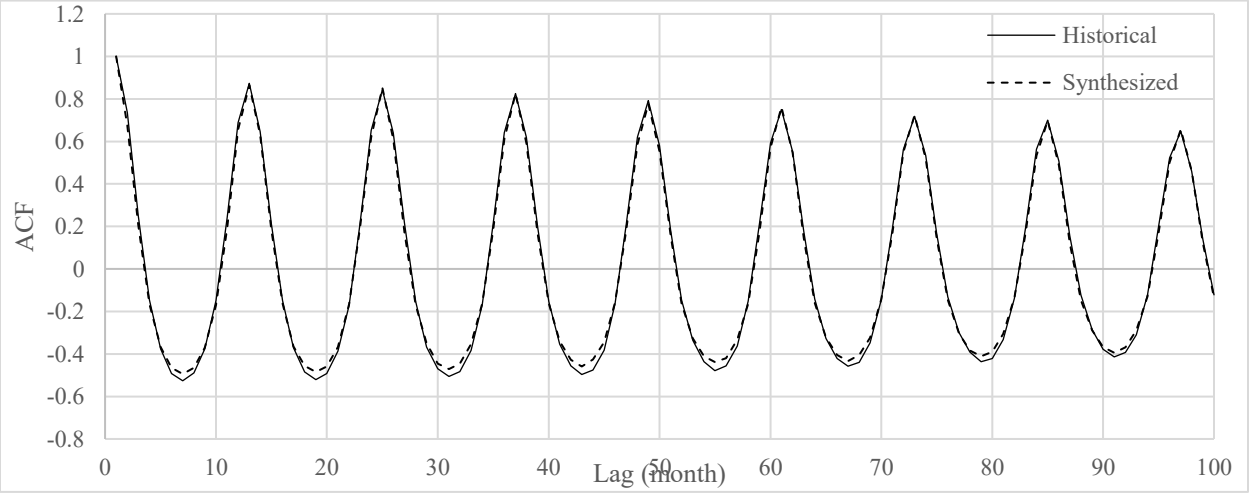
| River | Hypothesis Testing (Ho) | | |
|---|---|---|---|
| | $\bar{X}_{Hist} = \bar{X}_{Syn}$ | $\mathbf{Std.Dev}_{Hist} = \mathbf{Std.Dev}_{Syn}$ | $\mathbf{Skewness}_{Hist} = \mathbf{Skewness}_{Syn}$ |
| South Saskatchewan | 0.699 | 0.970 | 0.746 |
| Fraser | 0.437 | 0.839 | 0.987 |
| Black | 0.67 | 0.937 | 0.986 |
| N.E.M | 0.443 | 0.228 | 0.495 |
| St. John | 0.125 | 0.248 | 0.138 |
| Grand | 0.963 | 0.571 | 0.753 |

To explore further the similarity between the historical and synthesized streamflow, the autocorrelation function (ACF) of the synthesized streamflow realizations and the historical monthly streamflow is presented in Table 4.15. In all six rivers, the ACF of synthesized streamflow is found to closely resemble the historical streamflow up to higher lags of 100, and thus the synthesized streamflow can reproduce the temporal dependence and correlation structure of the historical streamflow until higher lags. The ability of the textural image feature prediction model to capture the temporal dependence and correlation structure of the observed streamflow data up

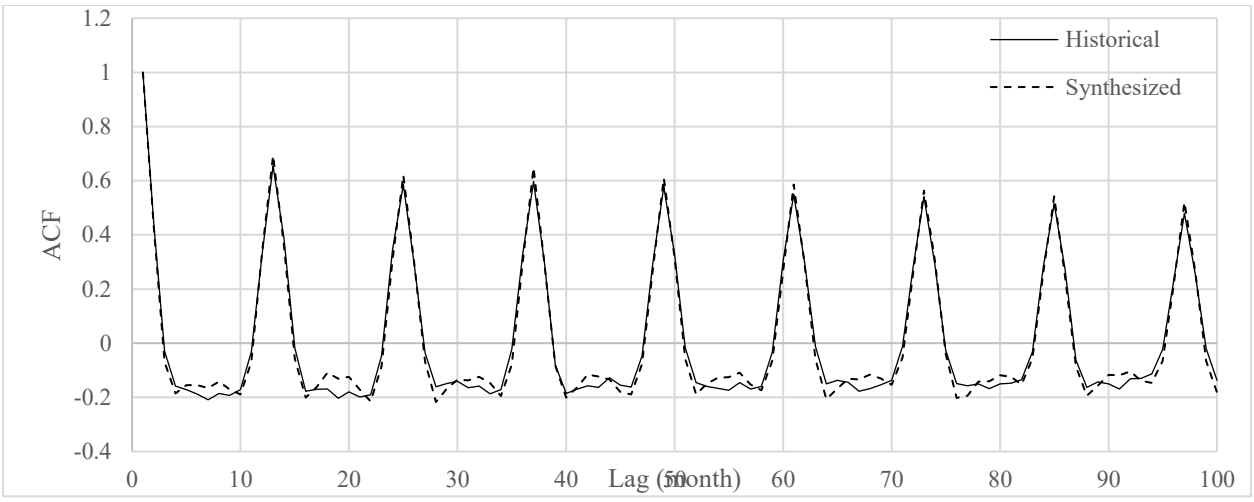
to higher lags is particularly important, as it indicates that the model can capture the long-term dynamics and memory of the system.



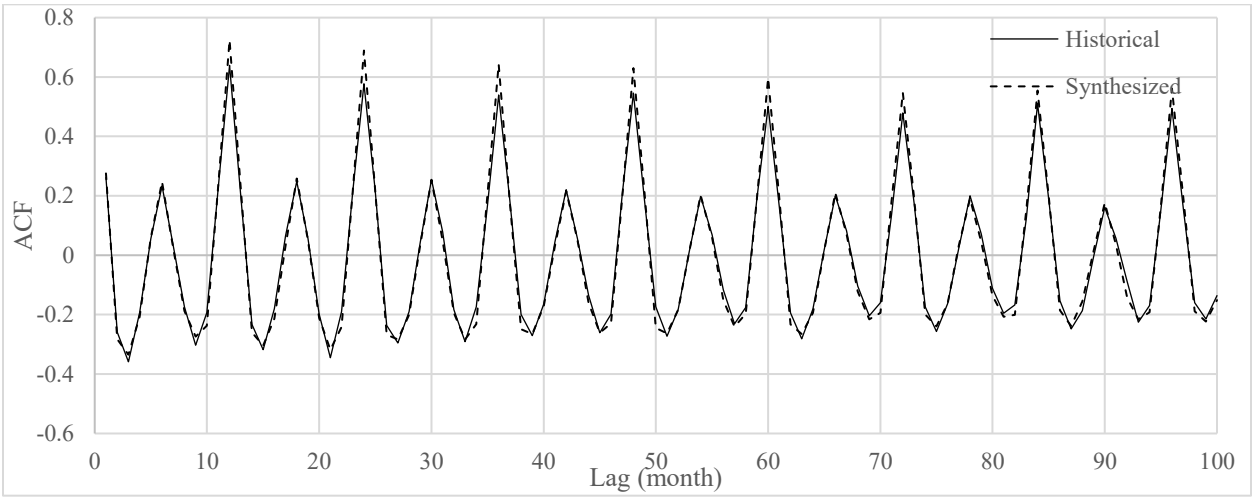

*(a) South Saskatchewan River.*



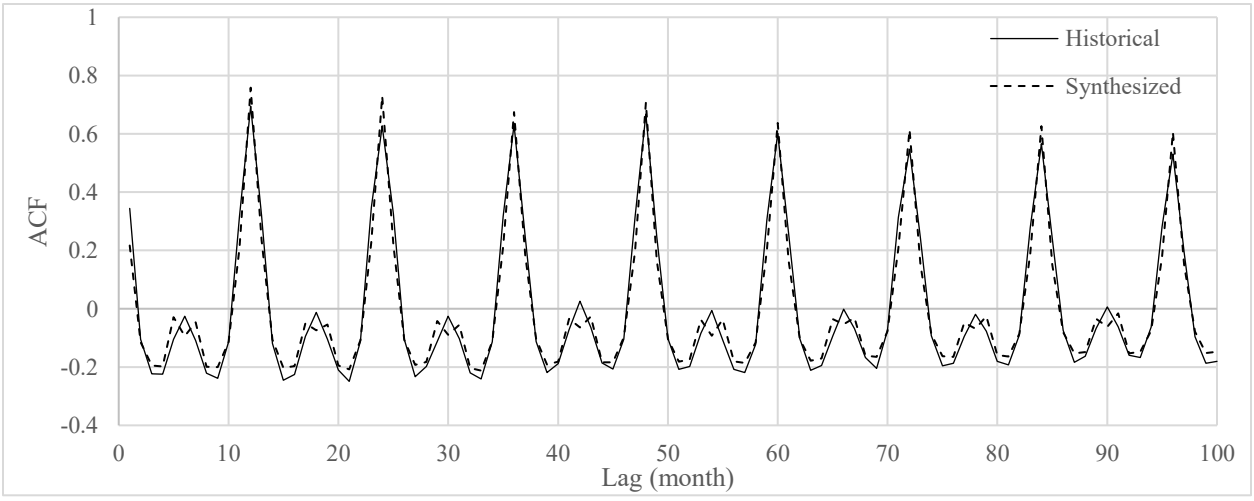

*(b) Fraser River.*



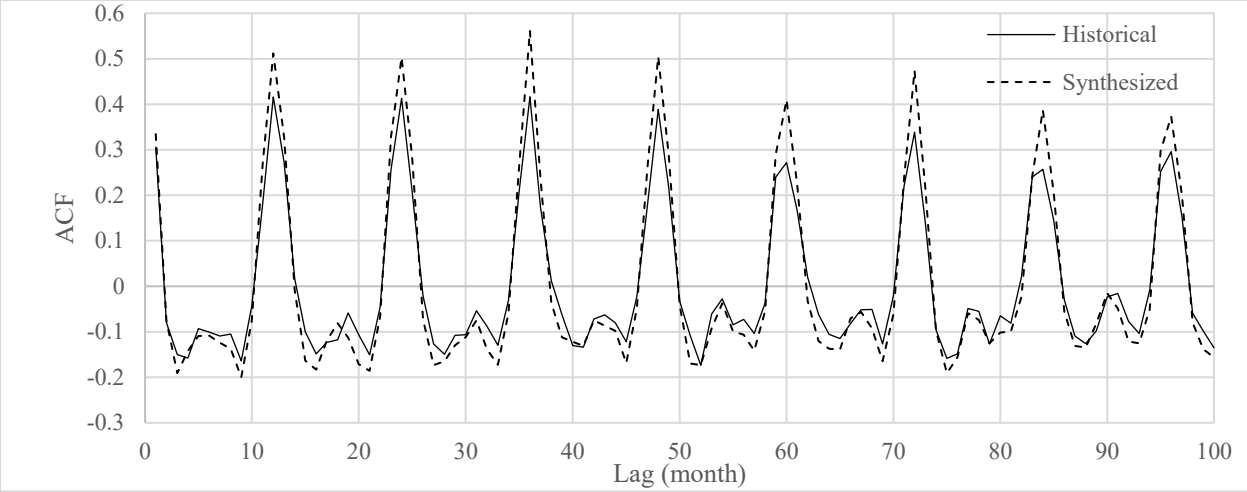

*(c) Black River.*




*(d) N.E.M River.*



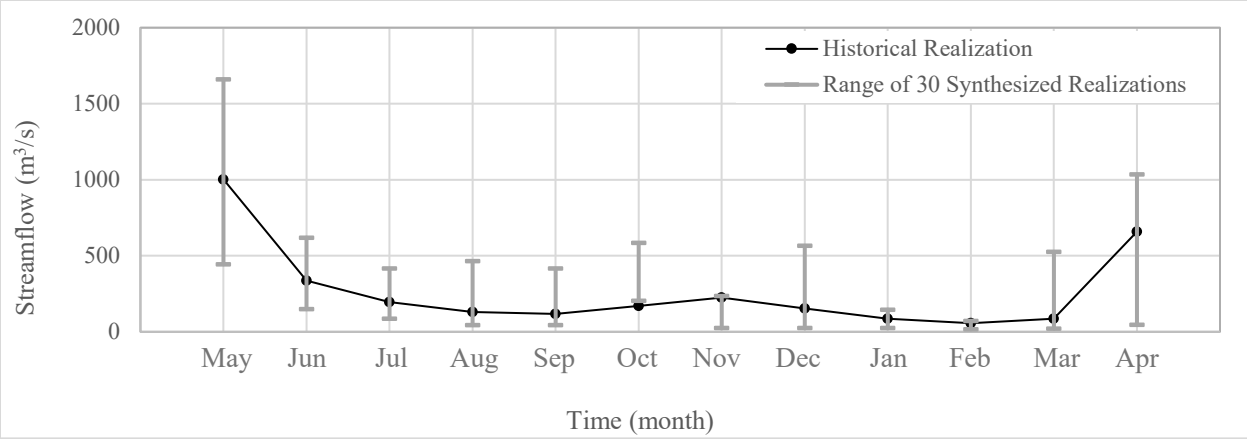

*(e) St. John River.*



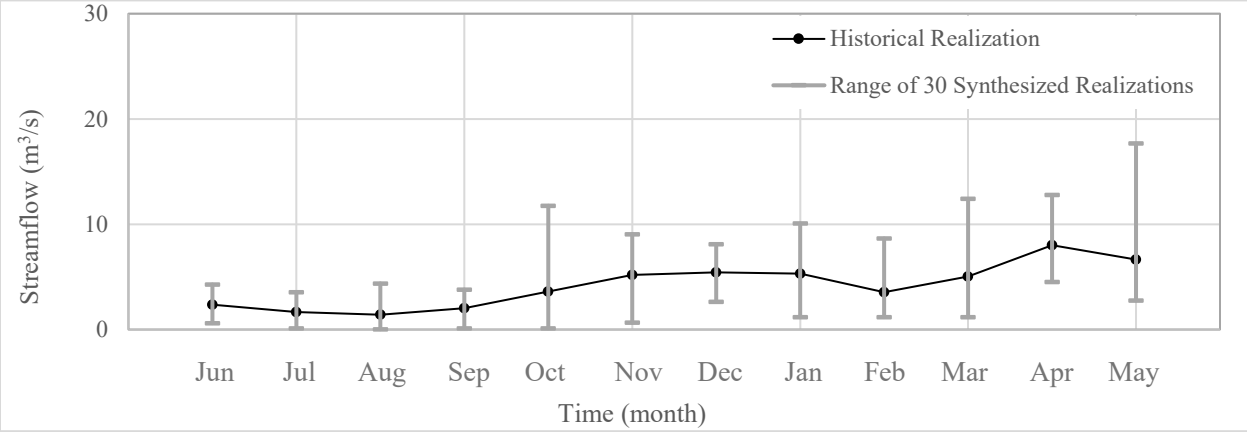

*(f) Grand River.*

*Figure 4.15. A comparison of the ACF of Historical and Synthesized Streamflow of (a) South Saskatchewan River, (b) Fraser River, (c) Black River, (d) N.E.M. River, (e) St. John River, and (f) Grand River.*

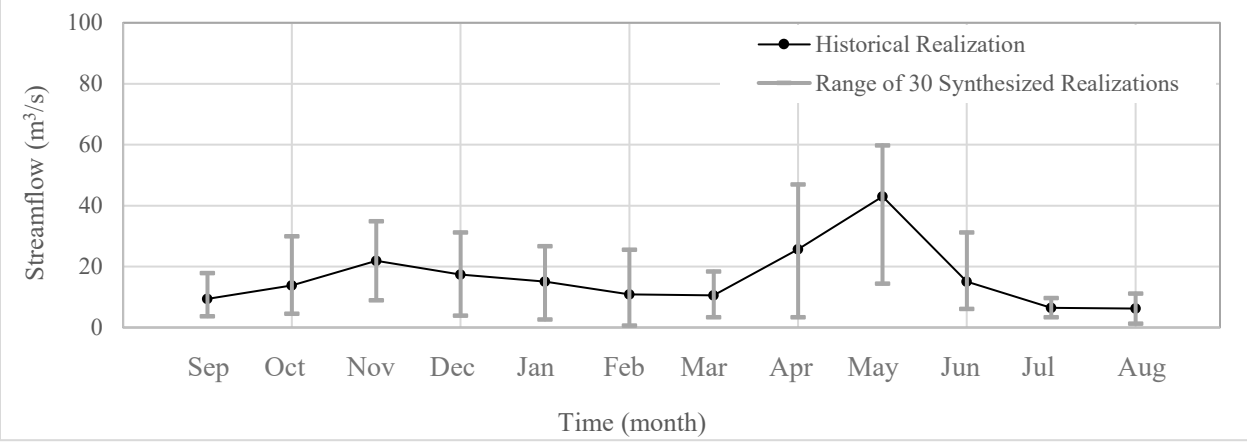
Additional investigations were accomplished through monthly comparisons for the Fraser, Black, and N.E.M. rivers. Figure 4.16**Error! Reference source not found.** illustrates the monthly mean streamflow values of historical data along with the monthly range values obtained from the 30 synthesized streamflow realizations. A review of this figure indicates that the historical mean monthly values, in general, lie within the established range boundaries for all rivers. Further, the historical mean monthly data for the St. John River falls within the range for every month except October. Moreover, the model shows a wider range of values, especially when streamflows are higher. For example, in the St. John River, the month of May has the highest mean streamflow of 1002.63 $m^3/s$, with a range of 443 –1660 $m^3/s$, while February has the lowest mean of 56.77 $m^3/s$, with a range of 16 –71 $m^3/s$. Similarly, in the Grand River, the peak streamflow (42.99 $m^3/s$) occurs in January, whereas the model exhibits a wider range of 14.46 –59.85 $m^3/s$.




*(a)*




*(b)*




*(c)*

*Figure 4.16. Monthly mean streamflow values for historical data, along with the range of the 30 synthesized streamflow realizations (represented by vertical lines) for a) St. John b) Grand and c) N.E.M Rivers.*

The textural feature recognition system shows strong capability in replicating observed streamflow patterns with high precision. It excels in capturing both short-term and long-term persistence and periodicities, as evidenced by a strong match between the observed and synthetic data in terms of key statistical parameters, such as monthly means, monthly standard deviations, ACF and the Hurst coefficient.

Notably, the proposed model offers an advantage over the previous Pattern Recognition System. It slightly improves the Hurst coefficient, thereby better representing long-term persistence, and significantly extends the ability to capture the autocorrelation structure from 24 lags to 100 lags. This enhancement is particularly critical, as streamflow patterns typically exhibit long-range temporal dependencies due to factors like seasonality and watershed dynamics.

However, the proposed model currently utilizes a single lag ($\zeta$=1) in developing the simultaneous autocorrelation function (SACF). Future research should explore higher lags or the implementation of optimization algorithms in the identification of optimal lags for further improvement in the performance of the model. Additionally, alternative time series imaging techniques — such as Gramian Angular Summation/Difference Field (GASF/GADF), Markov Transition Field (MTF), and Recurrence Plot (RP) — could also be investigated to further enhance the efficacy of the proposed model.

The proposed model, based on pattern recognition, is flexible and can be applied at finer (daily or weekly) or coarser (seasonal) temporal scales. For finer scales, the 2D correlation captures the immediate past (e.g., yesterday for daily data) and long-term dependencies (e.g., 7 days ago). For monthly data, segmentation would involve 12 months per year, while for weekly or daily scales, the segmentation process would need to reflect the appropriate time frame. The initial stage of data pre-processing should also consider these adjustments to ensure proper alignment with the selected temporal scale.

It is also important to acknowledge that this study segmented the data into two seasons. While this approach was effective for these specific case studies, it may not be universally applicable. Further research into the impact of varying seasonal lengths on streamflow synthesis is necessary to broaden the applicability of the model across different geographic and hydrological study areas. Local stationarity within each segment is assumed in the proposed methodology to enable effective

pattern recognition and synthesis. In this sense, first-order and second-order stationarity within each segment is invoked.

It is acknowledged that higher-order non-stationarity, such as evolving distributional shapes or more complex temporal dependencies, is not explicitly addressed in the current approach.

While this study focused on streamflow synthesis, it should be noted that the proposed model can also be adapted for forecasting purposes. In such applications, care should be taken to define an appropriate starting point, typically using observed data from one year prior to the current time where forecasting is initiated.

## 4.4 Summary Remarks

This chapter first distinguished the noise characteristics of streamflow in Ontario. For watersheds exhibiting predominantly brown noise, a geometric Brownian motion extension was successfully developed and applied. For pink- and brown-dominated watersheds, further analysis was conducted on the relationship between the Hurst exponent and the scaling exponent, followed by an examination of the first-order autoregressive parameter (AR (1)).

Three watersheds were randomly selected for both pink- and black-dominated categories, and synthetic series were generated using the TPRS and fitted ARIMA models. The comparison showed that for pink-noise-dominated watersheds, although the TPRS performs better, there is no compelling reason to use the more complex TPRS, where a high accuracy is not warranted.

Finally, TPRS was compared with traditional pattern recognition systems using six stations from this study. The results indicate that TPRS achieves higher performance in capturing the autocorrelation function (ACF) at multiple lags, previously limited to 24 lags by traditional methods. This demonstrates the superiority of TPRS in representing temporal dependencies over longer horizons.

# Supplementary Material for Chapter 4:

## Appendix 4A: Hydrometric Stations used for Noise Recognition Analysis.

*Table 4A.17. Summary of relevant information on hydrometric stations used for noise recognition analysis.*

| No | Station Name | Station Number | Location (Lat. & Long.) | Drainage Area ($km^2$) | Years of Records (Jan-Dec) | Hurst Coefficient (H) | Scaling Exponent (λ) |
|---|---|---|---|---|---|---|---|
| 1 | Alder Creek near New Dundee | 02GA030 | 43° 22' 20" N 80° 33' 03" W | 47.4 | 1970-2023 | 0.471 | 2.146 |
| 2 | Ausable River near Springbank | 02FF002 | 43° 04' 18" N 81° 39' 35" W | 865 | 1948-2023 | 0.507 | 2.307 |
| 3 | Aux Sables River at Massey | 02CE002 | 46° 12' 53" N 82° 04' 14" W | 1,340 | 1921-2023 | 0.495 | 1.916 |
| 4 | Avon River below Stratford | 02GD018 | 43° 20' 41" N 81° 06' 59" W | 140 | 1965-2023 | 0.569 | 2.596 |
| 5 | Basswood River near Winton | 05PA012 | 48° 04' 57" N 91° 39' 04" W | 4,510 | 1930-2015 | 0.722 | 2.008 |
| 6 | Batchawana River near Batchawana | 02BF001 | 47° 00' 12" N 84° 30' 56" W | 1,230 | 1968-2023 | 0.557 | 1.321 |
| 7 | Bayfield River near Varna | 02FF007 | 43° 33' 04" N 81° 35' 22" W | 460 | 1967-2023 | 0.500 | 2.141 |
| 8 | Bear Creek near Petrolia | 02GG006 | 42° 54' 21" N 82° 07' 08" W | 249 | 1967-2023 | 0.536 | 3.000 |
| 9 | Beaver River near Clarksburg | 02FB009 | 44° 31' 11" N 80° 28' 04" W | 587 | 1961-2023 | 0.635 | 2.419 |
| 10 | Big Creek near Delhi | 02GC006 | 42° 50' 15" N 80° 30' 35" W | 370 | 1956-2023 | 0.675 | 2.641 |
| 11 | Big Creek near Walsingham | 02GC007 | 42° 41' 08" N 80° 32' 18" W | 567 | 1956-2023 | 0.683 | 3.209 |
| 12 | Big East River near Huntsville | 02EB013 | 45° 23' 33" N 79° 09' 35" W | 610 | 1974-2023 | 0.639 | 2.600 |
| 13 | Bighead River near Meaford | 02FB010 | 44° 34' 12" N 80° 38' 54" W | 298 | 1958-2023 | 0.627 | 2.609 |
| 14 | Black Creek near Weston | 02HC027 | 43° 40' 27" N 79° 30' 15" W | 58 | 1967-2023 | 0.575 | 2.483 |
| 15 | Black River near Actinolite | 02HL003 | 44° 32' 22" N 77° 22' 10" W | 429 | 1956-2023 | 0.613 | 1.958 |
| 16 | Black River near Washago | 02EC002 | 44° 42' 49" N 79° 16' 53" W | 1,510 | 1917-2023 | 0.555 | 1.980 |
| 17 | Black Sturgeon River at Highway No. 17 | 02AC002 | 48° 54' 15" N 88° 22' 36" W | 2,980 | 1972-2023 | 0.509 | 2.048 |
| 18 | Blackwater River at Beardmore | 02AD010 | 49° 35' 51" N 87° 57' 55" W | 652 | 1972-2023 | 0.504 | 1.210 |
| 19 | Blanche River above Englehart | 02JC008 | 47° 53' 20" N 79° 52' 45" W | 1,780 | 1973-2023 | 0.477 | 2.202 |
| 20 | Blue Springs Creek near Eden Mills | 02GA031 | 43° 34' 34" N 80° 06' 32" W | 41.5 | 1966-2023 | 0.595 | 2.511 |
| 21 | Bonnechere River at Castleford | 02KC009 | 45° 29' 47" N 76° 33' 52" W | 2,380 | 1963-2023 | 0.603 | 3.077 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 22 | Burnt River near Burnt River | 02HF003 | 44° 42' 35" N 78° 40' 39" W | 1,250 | 1963-2023 | 0.575 | 1.585 |
| 23 | Canagagigue Creek near Elmira | 02GA023 | 43° 34' 47" N 80° 30' 33" W | 114 | 1965-2023 | 0.499 | 2.329 |
| 24 | Carp River near Kinburn | 02KF011 | 45° 25' 03" N 76° 11' 55" W | 258 | 1972-2023 | 0.484 | 1.643 |
| 25 | Cat River below Wesleyan Lake | 04GA002 | 51° 10' 25" N 91° 35' 40" W | 5,390 | 1971-2023 | 0.629 | 3.147 |
| 26 | Catfish Creek near Sparta | 02GC018 | 42° 44' 45" N 81° 03' 25" W | 295 | 1965-2023 | 0.595 | 2.478 |
| 27 | Cedar River below Wabaskang Lake | 05QE008 | 50° 30' 27" N 93° 15' 30" W | 1,690 | 1970-2023 | 0.638 | 2.770 |
| 28 | Chukuni River near Ear Falls | 05QC001 | 50° 52' 25" N 93° 29' 05" W | 4,920 | 1963-2023 | 0.704 | 2.880 |
| 29 | Clyde River at Gordon Rapids | 02KF013 | 45° 08' 02" N 76° 37' 47" W | 291 | 1972-2023 | 0.557 | 1.593 |
| 30 | Clyde River near Lanark | 02KF010 | 45° 02' 45" N 76° 24' 03" W | 618 | 1971-2023 | 0.478 | 1.907 |
| 31 | Coldwater River at Coldwater | 02ED007 | 44° 42' 25" N 79° 38' 37" W | 168 | 1966-2023 | 0.479 | 2.958 |
| 32 | Collins Creek near Kingston | 02HM005 | 44° 15' 23" N 76° 36' 45" W | 160 | 1970-2023 | 0.492 | 2.484 |
| 33 | Conestogo River at Glen Allan | 02GA028 | 43° 39' 17" N 80° 42' 07" W | 571 | 1960-2023 | 0.488 | 1.426 |
| 34 | Consecon Creek at Allisonville | 02HE002 | 44° 01' 38" N 77° 22' 00" W | 119 | 1970-2023 | 0.499 | 2.622 |
| 35 | Credit River near Cataract | 02HB001 | 43° 50' 09" N 80° 01' 22" W | 209 | 1916-2023 | 0.574 | 2.548 |
| 36 | Credit River near Orangeville | 02HB013 | 43° 53' 28" N 43° 53' 28" N | 60.6 | 1968-2023 | 0.693 | 2.652 |
| 37 | Credit River West Branch at Norval | 02HB008 | 43° 38' 47" N 79° 51' 58" W | 131 | 1964-2023 | 0.658 | 3.104 |
| 38 | Crowe River at Marmora | 02HK003 | 44° 28' 53" N 77° 41' 05" W | 1,930 | 1960-2023 | 0.515 | 2.553 |
| 39 | Depot Creek at Bellrock | 02HM002 | 44° 28' 18" N 76° 45' 44" W | 181 | 1957-2023 | 0.487 | 1.876 |
| 40 | Don River at Todmorden | 02HC024 | 43° 41' 09" N 79° 21' 41" W | 319 | 1963-2022 | 0.730 | 2.609 |
| 41 | East Humber River near Pine Grove | 02HC009 | 43° 47' 24" N 79° 35' 03" W | 191 | 1954-2023 | 0.724 | 2.519 |
| 42 | East Sixteen Mile Creek near Omagh | 02HB004 | 43° 29' 57" N 79° 46' 36" W | 193 | 1958-2023 | 0.619 | 2.694 |
| 43 | English River at Umfreville | 05QA002 | 49° 52' 24" N 91° 27' 35" W | 6,230 | 1922-2023 | 0.707 | 1.918 |
| 44 | Eramosa River above Guelph | 02GA029 | 43° 32' 52" N 80° 10' 55" W | 231 | 1963-2023 | 0.597 | 2.963 |
| 45 | Etobicoke Creek below Queen Elizabeth Highway | 02HC030 | 43° 36' 06" N 79° 33' 22" W | 205 | 1967-2023 | 0.562 | 2.580 |
| 46 | Fairchild Creek near Brantford | 02GB007 | 43° 08' 50" N 80° 09' 16" W | 389 | 1965-2015 | 0.528 | 2.615 |
| 47 | French River at Dry Pine Bay | 02DD010 | 46° 04' 06" N 80° 36' 41" W | 13,900 | 1962-2022 | 0.570 | 2.498 |
| 48 | Goulais River near Searchmont | 02BF002 | 46° 51' 39" N 83° 58' 18" W | 1,140 | 1968-2023 | 0.565 | 1.326 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 49 | Grand River at Brantford | 02GB001 | 43° 07' 57" N 80° 16' 02" W | 5,200 | 1948-2023 | 0.602 | 2.678 |
| 50 | Gull River at Norland | 02HF002 | 44° 43' 54" N 78° 49' 05" W | 1,280 | 1963-2023 | 0.681 | 2.664 |
| 51 | Holland River East Branch at Holland Landing | 02EC009 | 44° 05' 41" N 79° 29' 22" W | 176 | 1966-2023 | 0.662 | 3.043 |
| 52 | Humber River at Weston | 02HC003 | 43° 41' 56" N 79° 31' 13" W | 802 | 1949-2023 | 0.720 | 3.166 |
| 53 | Indian River near Blakeney | 02KF012 | 45° 14' 43" N 76° 15' 37" W | 212 | 1972-2023 | 0.483 | 1.843 |
| 54 | Jackson Creek at Peterborough | 02HJ001 | 44° 18' 10" N 78° 19' 17" W | 116 | 1963-2023 | 0.499 | 2.601 |
| 55 | Jock River near Richmond | 02LA007 | 45° 14' 57" N 75° 47' 26" W | 526 | 1970-2023 | 0.522 | 1.574 |
| 56 | Kaministiquia River at Kaministiquia | 02AB006 | 48° 31' 56" N 89° 35' 42" W | 6,470 | 1935-2023 | 0.482 | 2.389 |
| 57 | Kemptville Creek near Kemptville | 02LA006 | 44° 59' 39" N 75° 39' 44" W | 411 | 1970-2023 | 0.499 | 1.736 |
| 58 | Kettle Creek at St. Thomas | 02GC002 | 42° 46' 39" N 81° 12' 50" W | 331 | 1968-2023 | 0.520 | 2.639 |
| 59 | Kwetabohigan River near the Mouth | 04KA001 | 51° 09' 39" N 80° 51' 50" W | 4,250 | 1973-2023 | 0.477 | 1.711 |
| 60 | Little Pic River near Coldwell | 02BA003 | 48° 50' 56" N 86° 36' 25" W | 1,320 | 1973-2023 | 0.495 | 2.423 |
| 61 | Little Rouge Creek near Locust Hill | 02HC028 | 43° 54' 28" N 79° 12' 58" W | 83.6 | 1964-2022 | 0.665 | 2.357 |
| 62 | Little White River near Bellingham | 02CC005 | 46° 23' 38" N 83° 16' 58" W | 1,970 | 1943-2023 | 0.509 | 1.777 |
| 63 | Lynde Creek near Whitby | 02HC018 | 43° 52' 31" N 78° 57' 37" W | 100 | 1976-2020 | 0.645 | 3.185 |
| 64 | Lynn River at Simcoe | 02GC008 | 42° 49' 24" N 80° 17' 22" W | 144 | 1958-2023 | 0.744 | 3.371 |
| 65 | Madawaska River at Palmer Rapids | 02KD004 | 45° 19' 41" N 77° 30' 55" W | 5,800 | 1930-2023 | 0.529 | 2.584 |
| 66 | Maitland River above Wingham | 02FE005 | 43° 54' 54" N 81° 15' 51" W | 527 | 1954-2023 | 0.487 | 2.303 |
| 67 | Mattagami River near Timmins | 04LA002 | 48° 24' 15" N 81° 26' 54" W | 5,570 | 1974-2023 | 0.512 | 1.129 |
| 68 | McKenzie Creek near Caledonia | 02GB010 | 43° 02' 02" N 79° 56' 59" W | 173 | 1962-2023 | 0.692 | 3.250 |
| 69 | Medway River in London | 02GD008 | 43° 00' 49" N 81° 16' 49" W | 203 | 1971-2023 | 0.498 | 1.910 |
| 70 | Michipicoten River at Scott Falls | 02BD002 | 47° 54' 38" N 84° 44' 35" W | 5,310 | 1924-2020 | 0.599 | 1.732 |
| 71 | Middle Maitland River near Belgrave | 02FE008 | 43° 48' 46" N 81° 18' 24" W | 645 | 1968-2023 | 0.471 | 2.565 |
| 72 | Middle Maitland River near Listowel | 02FE003 | 43° 43' 38" N 80° 58' 21" W | 73.4 | 1954-2023 | 0.531 | 2.621 |
| 73 | Middle Thames River at Thamesford | 02GD004 | 43° 03' 32" N 80° 59' 41" W | 306 | 1949-2023 | 0.554 | 2.595 |
| 74 | Millhaven Creek near Millhaven | 02HM006 | 44° 13' 35" N 76° 45' 32" W | 144 | 1971-2023 | 0.486 | 2.738 |
| 75 | Missinaibi River at Mattice | 04LJ001 | 49° 36' 39" N 83° 16' 04" W | 8,570 | 1921-2023 | 0.521 | 1.509 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 76 | Missinaibi River below Waboose River | 04LM001 | 50° 35' 07" N 82° 05' 27" W | 22,900 | 1976-2023 | 0.489 | 1.526 |
| 77 | Mississippi River at Appleton | 02KF006 | 45° 10' 34" N 76° 07' 24" W | 2,940 | 1919-2023 | 0.537 | 2.241 |
| 78 | Moira River near Deloro | 02HL005 | 44° 29' 59" N 77° 37' 06" W | 297 | 1966-2023 | 0.493 | 1.983 |
| 79 | Moira River near Foxboro | 02HL001 | 44° 15' 13" N 77° 25' 07" W | 2,590 | 1916-2023 | 0.491 | 2.075 |
| 80 | Montreal River near Montreal River Harbour | 02BE002 | 47° 12' 50" N 84° 37' 10" W | 2,880 | 1936-2020 | 0.705 | 1.458 |
| 81 | Moon River at Highway No. 400 | 02EB011 | 45° 03' 54" N 79° 47' 24" W | 4,790 | 1966-2023 | 0.640 | 2.656 |
| 82 | Musquash River at Highway No. 400 | 02EB012 | 45° 01' 21" N 79° 46' 36" W | N/A | 1966-2023 | 0.544 | 2.374 |
| 83 | Nagagami River at Highway No. 11 | 04JC002 | 49° 46' 22" N 84° 32' 13" W | 2,180 | 1951-2023 | 0.488 | 1.963 |
| 84 | Namakan River at the Outlet of Lac La Croix | 05PA006 | 48° 22' 57" N 92° 10' 34" W | 13,400 | 1923-2023 | 0.668 | 2.682 |
| 85 | Nanticoke Creek at Nanticoke | 02GC022 | 42° 48' 35" N 80° 04' 34" W | 177 | 1970-2023 | 0.591 | 3.064 |
| 86 | Napanee River at Camden East | 02HM007 | 44° 20' 04" N 76° 50' 20" W | 700 | 1974-2023 | 0.475 | 2.823 |
| 87 | Neebing River near Thunder Bay | 02AB008 | 48° 23' 00" N 89° 18' 23" W | 187 | 1954-2022 | 0.601 | 1.936 |
| 88 | Niagara River at Queenston | 02HA003 | 43° 09' 25" N 79° 02' 50" W | 686,000 | 1860-2021 | 0.906 | 3.342 |
| 89 | Nith River above Nithburg | 02GA038 | 43° 29' 02" N 80° 50' 06" W | 326 | 1973-2023 | 0.490 | 2.633 |
| 90 | Nith River near Canning | 02GA010 | 43° 11' 23" N 80° 27' 18" W | 1,030 | 1914-2023 | 0.496 | 2.326 |
| 91 | North Branch Muskoka River at Port Sydney | 02EB004 | 45° 12' 46" N 79° 16' 31" W | 1,410 | 1916-2023 | 0.580 | 2.074 |
| 92 | North French River near the Mouth | 04MF001 | 51° 04' 36" N 80° 45' 50" W | 6,680 | 1967-2023 | 0.506 | 1.450 |
| 93 | North Magnetawan River above Pickerel Lake | 02EA010 | 45° 42' 13" N 79° 18' 31" W | 155 | 1969-2022 | 0.635 | 1.598 |
| 94 | North Magnetawan River near Burk's Falls | 02EA005 | 45° 40' 10" N 79° 22' 45" W | 329 | 1916-2023 | 0.516 | 1.302 |
| 95 | North Thames River at St. Mary’s | 02GD005 | 43° 15' 20" N 81° 08' 44" W | 1,080 | 1952-2023 | 0.525 | 2.408 |
| 96 | North Thames River near Mitchell | 02GD014 | 43° 27' 01" N 81° 12' 24" W | 315 | 1914-2023 | 0.533 | 2.630 |
| 97 | North Thames River near Thorndale | 02GD015 | 43° 08' 57" N 81° 11' 31" W | 1,320 | 1954-2023 | 0.500 | 2.514 |
| 98 | Nottawasaga River near Baxter | 02ED003 | 44° 14' 59" N 79° 49' 17" W | 1,230 | 1950-2023 | 0.668 | 3.260 |
| 99 | Ogoki River above Whiteclay Lake | 04GB004 | 50° 52' 06" N 88° 55' 53" W | 11,200 | 1972-2023 | 0.683 | 3.217 |
| 100 | Oshawa Creek at Oshawa | 02HD008 | 43° 55' 49" N 78° 53' 29" W | 95.8 | 1960-2022 | 0.654 | 3.068 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 101 | Pagwachuan River at Highway No. 11 | 04JD005 | 49° 45' 51" N 85° 13' 34" W | 2,120 | 1968-2022 | 0.487 | 1.872 |
| 102 | Parkhill Creek above Parkhill Reservoir | 02FF008 | 43° 09' 50" N 81° 37' 53" W | 113 | 1974-2023 | 0.519 | 2.347 |
| 103 | Petawawa River near Petawawa | 02KB001 | 45° 53' 10" N 77° 18' 55" W | 4,120 | 1916-2022 | 0.534 | 2.773 |
| 104 | Pic River near Marathon | 02BB003 | 48° 46' 26" N 86° 17' 47" W | 4,220 | 1971-2022 | 0.488 | 2.387 |
| 105 | Pipestone River at Karl Lake | 04DA001 | 52° 34' 50" N 90° 11' 12" W | 5,960 | 1967-2022 | 0.590 | 1.693 |
| 106 | Rainy River at Fort Frances | 05PC019 | 48° 36' 30" N 93° 24' 12" W | 38,600 | 1906-2022 | 0.688 | 2.704 |
| 107 | Raisin River near Williamstown | 02MC001 | 45° 09' 19" N 74° 38' 17" W | 404 | 1961-2023 | 0.488 | 1.741 |
| 108 | Rideau River at Ottawa | 02LA004 | 45° 22' 53" N 75° 41' 49" W | 3,810 | 1949-2023 | 0.657 | 1.642 |
| 109 | Root River at Sault Ste. Marie | 02CA002 | 46° 33' 46" N 84° 16' 54" W | 109 | 1972-2023 | 0.488 | 1.648 |
| 110 | Rouge River near Markham | 02HC022 | 43° 51' 30" N 79° 14' 00" W | 181 | 1962-2022 | 0.550 | 2.615 |
| 111 | Ruscom River near Ruscom Station | 02GH002 | 42° 12' 41" N 82° 37' 44" W | 125 | 1972-2023 | 0.489 | 2.739 |
| 112 | Salmon River near Shannonville | 02HM003 | 44° 12' 26" N 77° 12' 33" W | 907 | 1959-2023 | 0.494 | 2.563 |
| 113 | Sauble River at Sauble Falls | 02FA001 | 44° 40' 39" N 81° 15' 21" W | 913 | 1957-2023 | 0.653 | 3.160 |
| 114 | Saugeen River near Port Elgin | 02FC001 | 44° 27' 23" N 81° 19' 35" W | 3,950 | 1915-2023 | 0.528 | 2.421 |
| 115 | Saugeen River near Walkerton | 02FC002 | 44° 07' 13" N 81° 06' 55" W | 2,140 | 1915-2023 | 0.548 | 3.107 |
| 116 | Seine River at Sturgeon Falls Generating Station | 05PB009 | 48° 44' 40" N 92° 17' 05" W | 5,880 | 1964-2021 | 0.681 | 2.401 |
| 117 | Serpent River at Highway No. 17 | 02CD001 | 46° 12' 38" N 82° 30' 44" W | 1,350 | 1967-2023 | 0.530 | 1.910 |
| 118 | Severn River at Swift Rapids | 02EC003 | 44° 51' 25" N 79° 32' 30" W | 5,850 | 1954-2021 | 0.606 | 2.878 |
| 119 | Shelter Valley Brook near Grafton | 02HD010 | 43° 59' 30" N 78° 00' 04" W | 63.8 | 1966-2023 | 0.572 | 2.595 |
| 120 | Sixteen Mile Creek at Milton | 02HB005 | 43° 30' 50" N 79° 52' 47" W | 101 | 1959-2023 | 0.582 | 2.836 |
| 121 | Skootamatta River near Actinolite | 02HL004 | 44° 32' 58" N 77° 19' 41" W | 678 | 1959-2023 | 0.496 | 2.118 |
| 122 | South Maitland River at Summerhill | 02FE009 | 43° 41' 03" N 81° 32' 28" W | 371 | 1968-2023 | 0.493 | 2.837 |
| 123 | South Nation River near Plantagenet Springs | 02LB005 | 45° 31' 01" N 74° 58' 41" W | 3,810 | 1949-2023 | 0.686 | 1.838 |
| 124 | Speed River below Guelph | 02GA015 | 43° 32' 01" N 80° 15' 08" W | 568 | 1951-2023 | 0.594 | 1.933 |
| 125 | Speed River near Armstrong Mills | 02GA040 | 43° 38' 19" N 80° 16' 12" W | 167 | 1974-2023 | 0.602 | 2.662 |
| 126 | Sturgeon River at McDougall Mills | 05QA004 | 50° 10' 02" N 91° 32' 26" W | 4,440 | 1962-2023 | 0.654 | 2.849 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 127 | Sturgeon River at the Outlet of Salvesen Lake | 05QE009 | 50° 21' 08" N<br>94° 27' 59" W | 1,530 | 1965-2023 | 0.584 | 2.193 |
| 128 | Sydenham River near Alvinston | 02GG002 | 42° 49' 50" N<br>81° 51' 06" W | 701 | 1949-2023 | 0.537 | 2.196 |
| 129 | Sydenham River near Owen Sound | 02FB007 | 44° 31' 20" N<br>80° 55' 48" W | 183 | 1949-2023 | 0.683 | 2.875 |
| 130 | Thames River at Byron | 02GE002 | 42° 57' 45" N<br>81° 19' 54" W | 3,080 | 1956-2023 | 0.543 | 3.033 |
| 131 | Thames River at Ingersoll | 02GD016 | 43° 02' 28" N<br>80° 53' 10" W | 510 | 1958-2023 | 0.564 | 2.678 |
| 132 | Thames River at Thamesville | 02GE003 | 42° 32' 41" N<br>81° 58' 02" W | 4,370 | 1956-2023 | 0.619 | 2.659 |
| 133 | Thames River near Ealing | 02GD001 | 42° 58' 24" N<br>81° 12' 30" W | 1,340 | 1916-2023 | 0.501 | 2.547 |
| 134 | Turtle River near Mine Centre | 05PB014 | 48° 51' 00" N<br>92° 43' 25" W | 4,770 | 1921-2023 | 0.713 | 3.019 |
| 135 | Twenty Mile Creek at Balls Falls | 02HA006 | 43° 08' 00" N<br>79° 22' 59" W | 292 | 1958-2023 | 0.569 | 2.657 |
| 136 | Wabigoon River at Dryden | 05QD016 | 49° 46' 55" N<br>92° 50' 34" W | 2,340 | 1971-2021 | 0.669 | 2.252 |
| 137 | Wabigoon River near Quibell | 05QD006 | 49° 57' 28" N<br>93° 24' 01" W | 6,490 | 1954-2023 | 0.621 | 2.575 |
| 138 | Wanapitei River near Wanup | 02DB005 | 46° 20' 44" N<br>80° 50' 22" W | 3,150 | 1953-2023 | 0.543 | 2.637 |
| 139 | Welland River below Caistor Corners | 02HA007 | 43° 01' 18" N<br>79° 37' 04" W | 223 | 1958-2023 | 0.648 | 2.869 |
| 140 | West Humber River at Highway No. 7 | 02HC031 | 43° 45' 30" N<br>79° 40' 44" W | 142 | 1966-2023 | 0.510 | 2.292 |
| 141 | Whitemans Creek near Mount Vernon | 02GB008 | 43° 07' 34" N<br>80° 23' 01" W | 386 | 1962-2023 | 0.615 | 2.877 |
| 142 | Whitson River at Chelmsford | 02CF007 | 46° 35' 00" N<br>81° 11' 56" W | 277 | 1961-2023 | 0.489 | 1.776 |
| 143 | Wilmot Creek near Newcastle | 02HD009 | 43° 55' 48" N<br>78° 37' 07" W | 80.7 | 1966-2022 | 0.634 | 2.853 |

## Appendix 4B: ARIMA Model Selection Analysis

This appendix provides further details on the stations where ARMA models were applied, including Basswood River and Black Sturgeon River, which were used for comparison with the EGBM model, as well as Little White River, Mattagami River, Root River, Collins Creek, Kaministiquia River, and North Magnetawan River, which were used for comparison with the TPRS model. All relevant information is summarized in Table 4B-1.

*Table 0B.1. Summary of ARIMA model selection analysis for hydrometric stations based on log-likelihood and information criteria (AIC, AICc, and BIC).*

| Hydrometric Station | Model (d = 2, D = 1) | Log Likelihood | AICc | AIC | BIC |
|---|---|---|---|---|---|
| Basswood River near Winton (05PA012) | p = 2, q = 0, P = 0, Q = 1* | -773.47 | 1554.98 | 1554.94 | 1574.65 |
| | p = 2, q = 0, P = 1, Q = 1 | -773.66 | 1557.38 | 1557.32 | 1581.95 |
| | p = 2, q = 0, P = 0, Q = 2 | -776.90 | 1563.87 | 1563.81 | 1588.44 |
| | p = 2, q = 0, P = 2, Q = 1 | -779.41 | 1570.91 | 1570.83 | 1600.38 |
| | p = 2, q = 0, P = 2, Q = 2 | -839.74 | 1693.58 | 1693.47 | 1727.95 |
| | p = 1, q = 0, P = 0, Q = 1 | -863.07 | 1732.17 | 1732.14 | 1746.92 |
| | p = 1, q = 0, P = 2, Q = 1 | -861.07 | 1732.20 | 1732.14 | 1756.76 |
| | p = 1, q = 0, P = 1, Q = 1 | -862.75 | 1733.53 | 1733.49 | 1753.19 |
| | p = 1, q = 0, P = 0, Q = 2 | -865.33 | 1738.69 | 1738.65 | 1758.36 |
| | p = 1, q = 0, P = 2, Q = 2 | -930.40 | 1872.88 | 1872.80 | 1902.35 |
| | p = 1, q = 0, P = 1, Q = 2 | -991.31 | 1992.68 | 1992.62 | 2017.25 |
| | p = 2, q = 0, P = 1, Q = 2 | -1046.13 | 2104.35 | 2104.26 | 2133.82 |
| | p = 2, q = 0, P = 0, Q = 0 | -1081.44 | 2168.90 | 2168.88 | 2183.65 |
| | p = 1, q = 0, P = 2, Q = 0 | -1134.20 | 2276.45 | 2276.41 | 2296.11 |
| | p = 1, q = 0, P = 1, Q = 0 | -1144.55 | 2295.13 | 2295.10 | 2309.88 |
| | p = 2, q = 0, P = 1, Q = 0 | -1149.10 | 2306.24 | 2306.20 | 2325.90 |
| | p = 1, q = 0, P = 0, Q = 0 | -1171.61 | 2347.23 | 2347.22 | 2357.07 |
| | p = 2, q = 0, P = 2, Q = 0 | -1231.23 | 2472.53 | 2472.47 | 2497.10 |
| Black Sturgeon River at Highway No. 17 (02AC002) | p = 2, q = 0, P = 1, Q = 1* | -504.11 | 1018.32 | 1018.22 | 1040.29 |
| | p = 2, q = 0, P = 0, Q = 1 | -505.84 | 1019.74 | 1019.68 | 1037.33 |
| | p = 2, q = 0, P = 0, Q = 2 | -504.83 | 1019.75 | 1019.65 | 1041.72 |
| | p = 2, q = 0, P = 2, Q = 1 | -506.46 | 1025.06 | 1024.92 | 1051.40 |
| | p = 2, q = 0, P = 2, Q = 2 | -544.74 | 1103.66 | 1103.48 | 1134.37 |
| | p = 1, q = 0, P = 1, Q = 1 | -558.37 | 1124.81 | 1124.75 | 1142.40 |
| | p = 1, q = 0, P = 0, Q = 1 | -561.02 | 1128.07 | 1128.03 | 1141.27 |
| | p = 1, q = 0, P = 0, Q = 2 | -560.04 | 1128.15 | 1128.08 | 1145.74 |
| | p = 1, q = 0, P = 2, Q = 1 | -559.17 | 1128.44 | 1128.34 | 1150.41 |
| | p = 1, q = 0, P = 2, Q = 2 | -589.28 | 1190.70 | 1190.56 | 1217.04 |
| | p = 1, q = 0, P = 1, Q = 2 | -592.99 | 1196.08 | 1195.98 | 1218.05 |
| | p = 2, q = 0, P = 0, Q = 0 | -675.67 | 1357.37 | 1357.33 | 1370.57 |
| | p = 1, q = 0, P = 0, Q = 0 | -727.40 | 1458.82 | 1458.80 | 1467.63 |
| | p = 1, q = 0, P = 1, Q = 0 | -734.18 | 1474.40 | 1474.36 | 1487.60 |

| | | | | | |
|---|---|---|---|---|---|
| | p = 2, q = 0, P = 1, Q = 0 | -757.54 | 1523.14 | 1523.07 | 1540.73 |
| | p = 2, q = 0, P = 1, Q = 2 | -758.56 | 1529.26 | 1529.12 | 1555.61 |
| | p = 1, q = 0, P = 2, Q = 0 | -767.33 | 1542.72 | 1542.66 | 1560.31 |
| | p = 2, q = 0, P = 2, Q = 0 | -885.82 | 1781.73 | 1781.63 | 1803.70 |
| | p = 2, q = 1, P = 2, Q = 1 | -1318.81 | 2651.81 | 2651.63 | 2682.52 |
| Little White River near Bellingham (02CC005) | p = 2, q = 0, P = 0, Q = 1* | -4394.94 | 8795.90 | 8795.87 | 8810.51 |
| | p = 1, q = 0, P = 0, Q = 1 | -4394.32 | 8796.69 | 8796.65 | 8816.16 |
| | p = 0, q = 0, P = 0, Q = 1 | -4632.18 | 9268.38 | 9268.37 | 9278.13 |
| Mattagami River near Timmins (04LA002) | p = 2, q = 0, P = 0, Q = 2* | -3028.27 | 6066.64 | 6066.53 | 6088.52 |
| | p = 1, q = 0, P = 0, Q = 2 | -3031.55 | 6071.18 | 6071.11 | 6088.70 |
| | p = 0, q = 0, P = 0, Q = 2 | -3211.05 | 6428.15 | 6428.11 | 6441.30 |
| Root River at Sault Ste. Marie (02CA002) | p = 2, q = 0, P = 0, Q = 1* | -4670.67 | 9349.4 | 9349.3 | 9369.0 |
| | p = 1, q = 0, P = 0, Q = 1 | -4674.88 | 9355.8 | 9355.8 | 9370.5 |
| | p = 0, q = 0, P = 0, Q = 1 | -5095.61 | 10195.2 | 10195.2 | 10205.1 |
| Collins Creek near Kingston (02HM005) | p = 2, q = 1, P = 0, Q = 1* | -1383.39 | 2774.85 | 2774.78 | 2792.60 |
| | p = 1, q = 0, P = 0, Q = 1 | -1419.04 | 2844.12 | 2844.08 | 2857.44 |
| | p = 0, q = 0, P = 0, Q = 1 | -1503.53 | 3011.07 | 3011.06 | 3019.96 |
| | p = 2, q = 2, P = 0, Q = 1 | -1891.26 | 3794.66 | 3794.53 | 3821.25 |
| Kaministiquia River at Kaministiquia (02AB006) | *p = 2, q = 0, P = 0, Q = 1** | *-5422.77* | *10853.6* | *10853.5* | *10873.5* |
| | *p = 1, q = 0, P = 0, Q = 1* | *-5436.20* | *10878.4* | *10878.4* | *10893.4* |
| | *p = 0, q = 0, P = 0, Q = 1* | *-5892.46* | *11788.9* | *11788.9* | *11798.9* |
| North Magnetawan River near Burk's Falls (02EA005) | *p = 2, q = 0, P = 0, Q = 1** | *-1519.68* | *3047.41* | *3047.35* | *3065.25* |
| | *p = 1, q = 0, P = 0, Q = 1* | *-1526.83* | *3059.70* | *3059.66* | *3073.08* |
| | *p = 0, q = 0, P = 0, Q = 1* | *-1630.01* | *3264.03* | *3264.01* | *3272.96* |

***** *Best model with minimum AICc.*

## Appendix 4C: Manual Computations of Synthesized Streamflow

Considering two seasons $k = 1$ and $k = 2$, below is a sample computation for a 5-year period in the Black River for the second season. It should be noted that to synthesize 4 years of streamflow, it is necessary to consider data spanning 5 years.

$G^{k=1}=$

| - | 1915 | 1916 | 1917 | 1918 | 1919 |
|---|---|---|---|---|---|
| Oct | 39 | 22 | 22 | 49 | 13 |
| Nov | 33 | 62 | 37 | 79 | 89 |
| Dec | 42 | 138 | 23 | 125 | 98 |
| Jan | 73 | 40 | 11 | 143 | 18 |
| Feb | 128 | 18 | 8 | 40 | 11 |
| Mar | 82 | 84 | 99 | 188 | 157 |

$G^{k=2}=$

| - | 1915 | 1916 | 1917 | 1918 | 1919 |
|---|---|---|---|---|---|
| Apr | 221 | 187 | 161 | 133 | 126 |
| May | 103 | 77 | 58 | 81 | 55 |
| Jun | 40 | 35 | 33 | 45 | 30 |
| Jul | 12 | 26 | 11 | 12 | 7 |
| Aug | 3 | 13 | 5 | 3 | 1 |
| Sep | 1 | 4 | 12 | 2 | 1 |

For each raw (month): $\mu_i^{k=2} = [165.6 \quad 74.8 \quad 36.60 \quad 13.60 \quad 5.00 \quad 4.00\,]^T$

$$(S_i^{k=2}) = [39.30 \quad 19.45 \quad 5.94 \quad 7.23 \quad 4.69 \quad 4.64\,]^T$$

For each column (year): $\mu_j^{k=2} = [63.34 \quad 57.00 \quad 46.67 \quad 46.00 \quad 36.67]$

$$(S_j^{k=2}) = [86.18 \quad 68.54 \quad 59.32 \quad 52.41 \quad 48.55]$$

$SACF^{k=2}$

$$= \begin{bmatrix} - & - & - & - & - \\ - & (77-74.8)(103-74.8)(77-57)(187-57)/_{19.45^2 \times 68.54^2 = 0.091} & -0.036 & -0.305 & -0.225 \\ - & (35-57)(77-57)(35-36.3)(40-36.3)/_{68.54^2 \times 5.94^2 = 0.014} & -0.007 & 0.011 & 0.081 \\ - & -0.055 & -0.085 & 0.001 & 0.0169 \\ - & -0.2111 & 0 & 0 & 0.163 \\ - & 0 & 0 & -0.51 & 0.151 \end{bmatrix}$$

To compute the values for the first missed row (due to shifting and/or lagging process), $SMC^{k=2}$ needs to be computed. Therefore, the first row of $G_{k=2}$ and the last raw of $I_{k=1}$ must be considered and compared to their means ($\mu_{i=6}^{k=1} = 122$ and $\mu_{i=1}^{k=2} = 165.6$).

Therefore, the probability of transition from state $i$ in the season $k = 1$ to state $j$ in the season $k = 2$ is obtained using a comparison between $G^{k=1}(6,j)$ and $G^{k=2}(1,j)$

| | | $G_{k=2}(1,j)$ | |
|---|---|---|---|
| | | $G^{k=2}(1,j) < \mu_{i=1}^{k=2}$ | $G^{k=2}(1,j) \geq \mu_{i=1}^{k=2}$ |
| $G^{k=1}(6,j)$ | $G^{k=1}(6,j) < \mu_{i=6}^{k=1}$ | 1 | 2 |
| | $G^{k=1}(6,j) \geq \mu_{i=6}^{k=1}$ | 2 | 0 |

$$\longrightarrow \quad P = \begin{bmatrix} 0.33 & 0.67 \\ 1 & 0 \end{bmatrix}$$

Moreover, to determine the probability of transition from pixel $j-1$ to the next one in the same row $j$ within the same season $k=2$, $\hat{P}$ is obtained as below:

| | | $G^{k=2}(1,j)$ | |
|---|---|---|---|
| | | $G^{k=2}(1,j) < \mu_{i=1}^{k=2}$ | $G^{k=2}(1,j) \geq \mu_{i=1}^{k=2}$ |
| $G^{k=2}(1,j-1)$ | $G^{k=2}(1,j-1) < \mu_{i=1}^{k=2}$ | 2 | 0 |
| | $G^{k=2}(1,j-1) \geq \mu_{i=1}^{k=2}$ | 1 | 1 |

$$\longrightarrow \quad \hat{P} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$$

$SMC^{k=2}(i,j)$ for the first raw of $SACF^{k=2}$ will be

$SMC^{k=2} = [- \quad \frac{2}{3} \times 0.5 \quad \frac{2}{3} \times 0.5 \quad 1 \times 1 \quad 1 \times 1]$

Finally, $SACF^{k=2}=$

$$\begin{bmatrix} - & 0.33 & 0.33 & 1 & 1 \\ - & 0.091 & -0.036 & -0.305 & -0.255 \\ - & 0.014 & -0.007 & 0.011 & 0.081 \\ - & -0.055 & -0.085 & 0.001 & 0.0169 \\ - & -0.211 & 0 & 0 & 0.163 \\ - & 0 & 0 & -0.51 & 0.151 \end{bmatrix}$$

To transform $SACF^{k'}$ to synthesized grayscale image two possible roots are obtained using the solution of quadratic equation as below:

$$\frac{\left(\mu_j^k+\mu_i^k\right)\pm\sqrt{\left(\mu_j^k+\mu_i^k\right)^2-4\times\left[\mu_j^k\times\mu_i^k-\frac{SACF^{k\prime}(i,j)\times(S_j^k)^2\times(S_i^k)^2}{\left[G'(i-1,j)-\mu_j^k\right]\left[G'(i,j-1)-\mu_i^k\right]}\right]}}{2}$$

The outcomes of selecting either a positive or negative root are displayed in Table 4B-1. As observed, the model can capture the pattern and periodicity of the image well, regardless of the sign chosen. However, Table 4B-1 demonstrates that selecting a positive sign tends to overestimate the synthesized streamflow, while choosing the negative sign underestimates it. Hence, there is a need for a threshold to select the appropriate sign.

This cut-off point is defined based on the mean value of the original grayscale image. This means that each pixel is assigned to either the group with values higher than the mean or the group with values lower than the mean. If a pixel belongs to the former group, the positive sign is selected; otherwise, it belongs to the latter group, and a negative sign is selected.

*Table 0C.1. A summary of the model performance to synthesize the grayscale image using two different roots.*

| $G'(i,j)$ | Statistical properties | Historical | Synthesized |
|---|---|---|---|
| $\frac{\left(\mu_j^k+\mu_i^k\right)-\sqrt{\left(\mu_j^k+\mu_i^k\right)^2-4\times\left[\mu_j^k\times\mu_i^k-\frac{SACF^k_{1\theta_H,\zeta_1}{}'(i,j)\times(S_j^k)^2\times(S_i^k)^2}{\left[G'(i-1\eta,j)-\mu_j^k\right]\left[G'(i,j-1\theta_V)-\mu_i^k\right]}\right]}}{2}$ | Mean $m^3/s$ | 272.91 | 174.49 |
| | Std. Dev. $m^3/s$ | 292.61 | 154.48 |
| $\frac{\left(\mu_j^k+\mu_i^k\right)+\sqrt{\left(\mu_j^k+\mu_i^k\right)^2-4\times\left[\mu_j^k\times\mu_i^k-\frac{SACF^k_{1\theta_H,\zeta_1}{}'(i,j)\times(S_j^k)^2\times(S_i^k)^2}{\left[G'(i-1\eta,j)-\mu_j^k\right]\left[G'(i,j-1\theta_V)-\mu_i^k\right]}\right]}}{2}$ | Mean $m^3/s$ | 272.91 | 388.58 |
| | Std. Dev. $m^3/s$ | 292.61 | 286.68 |

# 5. Conclusion and Recommendations

This dissertation has focused on the classification and stochastic synthesis of monthly streamflow across Canadian rivers, emphasizing the role of coloured noise in model selection. The study distinguished watersheds based on dominant noise characteristics—brown, pink, or black—and developed tailored synthesis approaches for each category.

For brown-noise-dominated watersheds, an extended geometric Brownian motion model was found to more effectively capture long-term persistence and variability. In contrast, pink- and black-noise-dominated watersheds exhibited more complex temporal patterns that were successfully synthesized using the TPRS. Comparative analysis demonstrated that TPRS outperformed multiplicative ARIMA and traditional pattern recognition systems in reproducing autocorrelation structures, seasonal patterns, and higher-order statistical properties, highlighting its robustness for complex streamflow time series.

**The findings provide several practical insight**s:

**Noise-informed model selection**: The classification of streamflow stations (or watersheds) by noise type can guide the choice of an appropriate synthesis model, ensuring efficiency and reliability.

**Effectiveness of TPRS**: Textural approaches offer significant advantages over conventional methods for pink- and black-noise-dominated stream time series, particularly in capturing extended temporal dependencies and higher order statistics. TPRS is a pattern-based model that extracts strong temporal features and can, in principle, be applied to any watershed. In pink-noise-dominated watersheds, where long-term correlations are weaker and variability is more persistent across scales, simpler models may already provide sufficient accuracy for decision-making and policy purposes. This does not imply that TPRS cannot be applied; rather, its use depends on the level of accuracy required for the specific project. If the decision makers are satisfied with the performance of a simpler model, such as a regression approach, it may be more practical to use that. However, for projects requiring finer resolution of temporal dynamics or for more complex systems, TPRS offers clear advantages.

**Streamflow Data Synthesis:** The framework developed in this thesis enables reliable synthesis of streamflow sequences for water resource planning, modelling, and risk assessment.

## 5.1 Comments on the Textural Pattern Recognition System

A semi-textural image feature extraction system aiming to consider both the short- and long-term persistence for identifying patterns in monthly time series data has been represented using 8-bit grayscale images, which aids in efficiently storing, managing, and analyzing high-resolution data. Unlike conventional ML techniques that require large quality datasets for training, this approach effectively utilizes a single historical image representation to capture key features of the time series. In contrast, though the TPRS pertains to monthly hydrologic time series, it has the capacity to be easily extended to weekly and daily hydrological and meteorological time series. Moreover, segmentation adaptability is emphasized based on the characteristics and time intervals of the data set, utilizing correlograms, periodograms, and the Lyapunov time horizon. In the TPRS approach developed in this thesis, recognition of dual temporal patterns in textural image time series is restricted to considering only one lag (in both directions); however, future investigations may easily explore the incorporation of additional lags based on correlation properties and partial autocorrelation function (PACF) or the option for optimization algorithms to determine optimal lag selections for pattern identification. The primary focus of the TPRS approach is to extract features for synthesizing monthly streamflow, but it can be applied to data exhibiting trends and heteroskedasticity, provided that the first step involves the traditional pre-processing of the data set involving such basic functions as detrending, outlier treatment, etc. Additionally, addressing heteroskedasticity through variance stabilization techniques, such as the Box-Cox transformation, can further enhance the robustness and applicability of the TPRS approach.

The TPRS approach developed for the univariate case can also be easily adapted to multivariate cases, including multi-channel input involving such variables as precipitation, temperature, and land use changes, and by fusing features extracted from various images or merging images

corresponding to each of the multivariate cases into one using time-series imaging methods, such as Principal Component Analysis (PCA) Fusion.

## 5.2 Future Research Directions

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 methods, approaches and/or models. Accordingly, this dissertation presents the development of a semi-automated TPRS approach.

In the future, this approach can be extended in several directions. First, it can incorporate additional hydrological variables, such as precipitation, temperature, and land use, for multi-variable synthesis and forecasting. Second, it can be applied to disaggregation tasks, generating finer temporal resolution data from monthly or seasonal records. Third, the framework can inspire the development and testing of other semi-automated approaches and/or models, combining the efficiency of AI/ML with critical human oversight. Such extensions will enhance the robustness, versatility, and practical utility of semi-automated approaches in hydrological modelling and water resource management.

# References


Abdelaziz, S., Mahmoud Ahmed, A. M., Eltahan, A. M., & Abd Elhamid, A. M. I. (2023). Long-Term Stochastic Modelling of Monthly Streamflow in the River Nile. *Sustainability*, *15*(3), 2170.

Ahmed, J. A., & Sarma, A. K. (2007). Artificial neural network model for synthetic streamflow generation. *Water Resources Management*, *21*, 1015-1029.

Akaike, H. (1975). Markovian representation of stochastic processes by canonical variables. *SIAM journal on control*, *13*(1), 162-173.

Arselan, C. A. (2012). Stream flow simulation and synthetic flow calculation by the modified Thomas Fiering model. *Al-Rafidain Engineering Journal (AREJ)*, *20*(4), 118-127.

Awchi, T. A., & Srivastava, D. (2004. ). 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, . https://doi.org/https://doi.org/10.13140/2.1.1818.8164.

Bandt, C., & Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Physical Review Letters*, *88*(17), 174102.

Bayazit, M., Önöz, B., & Aksoy, H. (2001). Nonparametric streamflow simulation by wavelet or Fourier analysis. *Hydrological Sciences Journal*, *46*(4), 623-634.

Benson, M. A., & Carter, R. W. (1973). *A national study of the streamflow data-collection program*. US Government Printing Office.

Bergman, M., & Delleur, J. (1985). KALMAN FILTER ESTIMATION AND PREDICTION OF DAILY STREAM FLOWS: I. REVIEW, ALGORITHM, AND SIMULATION EXPERIMENTS 1. *JAWRA Journal of the American Water Resources Association*, *21*(5), 815-825.

Boughton, W., & McKerchar, A. (1968). Generating synthetic stream-flow records for New Zealand Rivers. *Journal of Hydrology (New Zealand)*, 112-123.

Bourdin, D. R., Fleming, S. W., & Stull, R. B. (2012). Streamflow modelling: a primer on applications, approaches and challenges. *Atmosphere-Ocean*, *50*(4), 507-536.

Bowles, D. S., James, W. R., & Kottegoda, N. T. (1987). Initial model choice: An operational comparison of stochastic streamflow models for drought. *Water Resources Management*, *1*, 3-15.

Brunner, M. I., Bárdossy, A., & Furrer, R. (2019). Stochastic simulation of streamflow time series using phase randomization. *Hydrology and Earth System Sciences*, *23*(8), 3175-3187.

Brunner, M. I., & Gilleland, E. (2020). Stochastic simulation of streamflow and spatial extremes: a continuous, wavelet-based approach. *Hydrology and Earth System Sciences*, *24*(8), 3967-3982.

Chong, K. L., Lai, S. H., & El-Shafie, A. (2019). Wavelet transform-based method for river stream flow time series frequency analysis and assessment in a tropical environment. *Water Resources Management*, *33*, 2015-2032.

Cover, K. A., & Unny, T. (1986). APPLICATION OF COMPUTER INTENSIVE STATISTICS TO PARAMETER UNCERTAINTY IN STREAMFLOW SYNTHESIS 1. *JAWRA Journal of the American Water Resources Association*, *22*(3), 495-507.

Dastour, H., & Hassan, Q. K. (2023). A machine-learning framework for modelling and predicting monthly streamflow time series. *Hydrology*, *10*(4), 95.

Deka, P. C. (2014). Support vector machine applications in the field of hydrology: a review. *Applied soft computing*, *19*, 372-386.

Deka, P. C. (2019). *A primer on machine learning applications in civil engineering*. CRC Press: Boca Raton, FL, USA.

Dettinger, M. D., & Diaz, H. F. (2000). Global characteristics of stream flow seasonality and variability. *Journal of Hydrometeorology*, *1*(4), 289-310.

Dooley, K. J., & Van de Ven, A. H. (1997). *A primer on diagnosing dynamic organizational processes*. Strategic Management Research Center, University of Minnesota, MN, USA. Available online:

https://scholar.google.com/scholar_lookup?title=A+Primer+on+Diagnosing+Dynamic+Organizational+Processes&author=Dooley,+K.J.&author=Van+de+Ven,+A.H.&publication_year=1997.
Elshorbagy, A., Panu, U., & Simonovic, S. (2001). Analysis of cross-correlated chaotic streamflows. *Hydrological Sciences Journal*, *46*(5), 781-793.
Ersoy, O. K. (1994). A comparative review of real and complex Fourier-related transforms. *Proceedings of the IEEE*, *82*(3), 429-447.
Fan, M., Liu, S., Lu, D., Gangrade, S., & Kao, S.-C. (2023). Explainable machine learning model for multi-step forecasting of reservoir inflow with uncertainty quantification. *Environmental Modelling & Software*, *170*, 105849.
Fernandez, B., & Salas, J. D. (1990). Gamma-autoregressive models for stream-flow simulation. *Journal of Hydraulic Engineering*, *116*(11), 1403-1414.
Fiering, M., & Jackson, B. (1971). Synthetic Streamflows, Water Resour. Monogr. Ser.
Fiering, M. B., & Bund, B. (1971). *Synthetic streamflows* (Vol. 1). American Geophysical Union, Washington, DC, USA.
Ghimire, G. R., & Krajewski, W. F. (2020). Exploring persistence in streamflow forecasting. *JAWRA Journal of the American Water Resources Association*, *56*(3), 542-550.
Girihagama, L., Naveed Khaliq, M., Lamontagne, P., Perdikaris, J., Roy, R., Sushama, L., & Elshorbagy, A. (2022). Streamflow modelling and forecasting for Canadian watersheds using LSTM networks with an attention mechanism. *Neural Computing and Applications*, *34*(22), 19995-20015.
Gonzalez, R. C. (2009). *Digital Image Processing*. Pearson Education India.
Hao, Z., & Singh, V. P. (2013). Modelling multisite streamflow dependence with maximum entropy copula. *Water Resources Research*, *49*(10), 7139-7143.
Heilbronner, R. P. (1992). The autocorrelation function: an image processing tool for fabric analysis. *Tectonophysics*, *212*(3-4), 351-370.
Hipel, K. W., McLeod, A. I., & Lennox, W. C. (1977). Advances in Box-Jenkins Modelling: 1. Model Construction. *Water Resources Research*, *13*(3), 567-575.
Hirsch, R. M. (1979). Synthetic hydrology and water supply reliability. *Water Resources Research*, *15*(6), 1603-1615.
Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American society of civil engineers*, *116*(1), 770-799.
Jardim, D., Maceira, M., & Falcao, D. (2001). Stochastic streamflow model for hydroelectric systems using clustering techniques. 2001 IEEE Porto Power Tech Proceedings (Cat. No. 01EX502),
Jia, Y., & Culver, T. B. (2006). Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. *Journal of Hydrology*, *331*(3-4), 580-590.
Kendall, D. R., & Dracup, J. A. (1991). A comparison of index-sequential and AR (1) generated hydrologic sequences. *Journal of Hydrology*, *122*(1-4), 335-352.
Keskin, M. E., Taylan, D., & Terzi, O. (2006). Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series. *Hydrological Sciences Journal*, *51*(4), 588-598.
Khare, S., & Gajbhiye, A. (1943). Literature Review on Application of Artificial Neural Network (ANN) in the Operation of Reservoirs. *International Journal of computational Engineering research (IJCER) IJCER| June 2013| VOL 3 ISSUE 6*, 63.
Kim, D. H., Rao, P. S. C., Kim, D., & Park, J. (2016). 1/f noise analyses of urbanization effects on streamflow characteristics. *Hydrological Processes*, *30*(11), 1651-1664.
Kirsch, B. R., Characklis, G. W., & Zeff, H. B. (2013). Evaluating the impact of alternative hydro-climate scenarios on transfer agreements: Practical improvement for generating synthetic streamflows. *Journal of Water Resources Planning and Management*, *139*(4), 396-406.
Klemeš, V. (1974). The Hurst phenomenon: A puzzle? *Water Resources Research*, *10*(4), 675-688.
Kojiri, T., Unny, T., & Panu, U. (1994). Cluster-Based Pattern Recognition and Analysis of Streamflows. In *Stochastic and Statistical Methods in Hydrology and Environmental Engineering: Time Series Analysis in Hydrology and Environmental Engineering* (pp. 363-380). Springer.

Kottegoda, N., Natale, L., & Raiteri, E. (2000). Daily streamflow simulation using recession characteristics. *Journal of Hydrologic Engineering*, *5*(1), 17-24.
Koutsoyiannis, D. (2002). The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrological Sciences Journal*, *47*(4), 573-595.
Lawrance, A., & Kottegoda, N. (1977). Stochastic modelling of riverflow time series. *Journal of the Royal Statistical Society: Series A (General)*, *140*(1), 1-31.
Lettenmaier, D. P., & Burges, S. J. (1977). Operational assessment of hydrologic models of long-term persistence. *Water Resources Research*, *13*(1), 113-124.
Li, C., & Singh, V. P. (2014). A multimodel regression-sampling algorithm for generating rich monthly streamflow scenarios. *Water Resources Research*, *50*(7), 5958-5979.
Li, F.-F., Cao, H., Hao, C.-F., & Qiu, J. (2021a). Daily streamflow forecasting based on flow pattern recognition. *Water Resources Management*, *35*, 4601-4620.
Li, F.-F., Cao, H., Hao, C.-F., & Qiu, J. (2021b). Daily streamflow forecasting based on flow pattern recognition. *Water Resources Management*, *35*(13), 4601-4620.
Liu, Y., Li, H., Yang, Y., Pang, X., & Niu, L. (2025). Enhancing machine learning runoff simulation via wavelet-based abnormality pattern recognition. *Journal of Hydrology*, *661*, 133729.
Ma, Y., Zhong, P.-a., Wang, G., & Xiao, Y. (2024). Performance of multisite streamflow stochastic generation approaches for a multi-reservoir system. *Stochastic environmental research and risk assessment*, *38*(6), 2135-2155.
Marković, Đ., Plavšić, J., Ilich, N., & Ilić, S. (2015). Non-parametric stochastic generation of streamflow series at multiple locations. *Water Resources Management*, *29*, 4787-4801.
Mehdiyev, N., Majlatow, M., & Fettke, P. (2025). Quantifying and explaining machine learning uncertainty in predictive process monitoring: an operations research perspective. *Annals of Operations Research*, *347*(2), 991-1030.
Mihailović, D. T., Malinović-Milićević, S., Han, J., & Singh, V. P. (2023). Complexity and chaotic behaviour of the US Rivers and estimation of their prediction horizon. *Journal of Hydrology*, *622*, 129730.
Molina, A. A. R., Frame, J. M., Halgren, J., & Gong, J. (2024). A Proof of Concept for Improving Estimates of Ungauged Basin Streamflow Via an LSTM-Based Synthetic Network Simulation Approach. *Authorea Preprints*.
Mujumdar, P., & Kumar, D. N. (1990). Stochastic models of streamflow: some case studies. *Hydrological Sciences Journal*, *35*(4), 395-410.
Muzik, I. (1980). Analysis of capacity requirements for storage reservoirs: a case study. *Canadian Journal of Civil Engineering*, *7*(2), 388-392.
Niu, J., & Sivakumar, B. (2013). Scale-dependent synthetic streamflow generation using a continuous wavelet transform. *Journal of Hydrology*, *496*, 71-78.
O'Connell, P. E. (1974). Stochastic modelling of long-term persistence in streamflow sequences. *Doctoral thesis, University of London*.
Ochoa-Rivera, J., García-Bartual, R., & Andreu, J. (2002). Multivariate synthetic streamflow generation using a hybrid model based on artificial neural networks. *Hydrology and Earth System Sciences*, *6*(4), 641-654.
Panu, U. S., & Unny, T. (1980a). Extension and application of the feature prediction model for the synthesis of hydrologic records. *Water Resources Research*, *16*(1), 77-96.
Panu, U. S., & Unny, T. (1980b). Stochastic synthesis of hydrologic data based on concepts of pattern recognition: II. Application of natural watersheds. *Journal of Hydrology*, *46*(3-4), 197-217.
Panu, U. S., Unny, T., & Ragade, R. (1978). A feature prediction model in synthetic hydrology based on concepts of pattern recognition. *Water Resources Research*, *14*(2), 335-344.
Parasuraman, K. (2007). *Hydrologic prediction using pattern recognition and soft-computing techniques,* University of Saskatchewan. Saskatoon, Saskatchewan, Canada.
Partington, D., Brunner, P., Frei, S., Simmons, C. T., Werner, A. D., Therrien, R., Maier, H. R., Dandy, G. C., & Fleckenstein, J. (2013). Interpreting streamflow generation mechanisms from integrated

surface-subsurface flow models of a riparian wetland and catchment. *Water Resources Research*, *49*(9), 5501-5519.
Patskoski, J., & Sankarasubramanian, A. (2015). Improved reservoir sizing utilizing observed and reconstructed streamflows within a Bayesian combination framework. *Water Resources Research*, *51*(7), 5677-5697.
Pei, S.-C., & Lin, C.-N. (1995). Image normalization for pattern recognition. *Image and Vision Computing*, *13*(10), 711-723.
Pender, D., Patidar, S., Pender, G., & Haynes, H. (2016). Stochastic simulation of daily streamflow sequences using a hidden Markov model. *Hydrology Research*, *47*(1), 75-88.
Pereira, G., & Veiga, A. (2018). PAR (p)-vine copula-based model for stochastic streamflow scenario generation. *Stochastic environmental research and risk assessment*, *32*, 833-842.
Petters, A. O., & Dong, X. (2016). Stochastic Calculus and Geometric Brownian Motion Model. In *An Introduction to Mathematical Finance with Applications: Understanding and Building Financial Intuition* (pp. 253-327). Springer New York. https://doi.org/10.1007/978-1-4939-3783-7_6
Phien, H. N., & Ruksasilp, W. (1981). A review of single-site models for monthly streamflow generation. *Journal of Hydrology*, *52*(1-2), 1-12.
Porto, V. C., de Souza Filho, F. d. A., Carvalho, T. M. N., de Carvalho Studart, T. M., & Portela, M. M. (2021). A GLM copula approach for multisite annual streamflow generation. *Journal of Hydrology*, *598*, 126226.
Prairie, J. R., Rajagopalan, B., Fulp, T. J., & Zagona, E. A. (2006). Modified K-NN model for stochastic streamflow simulation. *Journal of Hydrologic Engineering*, *11*(4), 371-378.
Raman, H., & Sunilkumar, N. (1995). Multivariate modelling of water resources time series using artificial neural networks. *Hydrological Sciences Journal*, *40*(2), 145-163.
Rasmussen, P. F., Salas, J. D., Fagherazzi, L., Rassam, J. C., & Bobée, B. (1996). Estimation and validation of contemporaneous PARMA models for streamflow simulation. *Water Resources Research*, *32*(10), 3151-3160.
Riggs, H. C. (1985). *Streamflow characteristics*. Elsevier.
Rippl, W. (1883). The Capacity Of Storage-Reservoirs For Water Supply. (Including Plate). Minutes of the Proceedings of the Institution of Civil Engineers,
Rodriguez-Iturbe, I., & Rinaldo, A. (1997). *Fractal river basins: chance and self-organization*. Cambridge University Press: Cambridge, UK.
Salas, J. D., Delleur, J. W., Yevjevich, V., & Lane, W. L. (1980). *Applied modelling of hydrologic time series*. Water Resources Publication, Littleton, Colorado. USA
Santos, E. G., & Salas, J. D. (1992). Stepwise disaggregation scheme for synthetic hydrology. *Journal of Hydraulic Engineering*, *118*(5), 765-784.
Sivakumar, B. (2016). *Chaos in hydrology: bridging determinism and stochasticity*. Springer: Berlin/Heidelberg, Germany.
Srinivas, V., & Srinivasan, K. (2001). A hybrid stochastic model for multiseason streamflow simulation. *Water Resources Research*, *37*(10), 2537-2549.
Srivastav, R. K., & Simonovic, S. P. (2014). An analytical procedure for multi-site, multi-season streamflow generation using maximum entropy bootstrapping. *Environmental Modelling & Software*, *59*, 59-75.
Stagge, J., & Moglen, G. (2013). A nonparametric stochastic method for generating daily climate-adjusted streamflows. *Water Resources Research*, *49*(10), 6179-6193.
Stedinger, J. R., Lettenmaier, D. P., & Vogel, R. M. (1985). Multisite ARMA (1, 1) and disaggregation models for annual streamflow generation. *Water Resources Research*, *21*(4), 497-509.
Stedinger, J. R., Pei, D., & Cohn, T. A. (1985). A condensed disaggregation model for incorporating parameter uncertainty into monthly reservoir simulations. *Water Resources Research*, *21*(5), 665-675.
Stedinger, J. R., & Taylor, M. R. (1982). Synthetic streamflow generation: 1. Model verification and validation. *Water Resources Research*, *18*(4), 909-918.

Stolte, W. (1980). The limitations and usefulness of streamflow generation methods: a case study. *Canadian Journal of Civil Engineering*, *7*(1), 185-191.

Stoyanov, M., Gunzburger, M., & Burkardt, J. (2011). Pink noise, 1/f α noise, and their effect on solutions of differential equations. *International Journal for Uncertainty Quantification*, *1*(3).

Studnicka, S., & Panu, U. (2025a). Techniques and Developments in Stochastic Streamflow Synthesis—A Comprehensive Review. *Encyclopedia*, *5*(4).

Studnicka, S., & Panu, U. (2025b). Streamflow Synthesis Using an Encoded Textural Pattern Recognition System. I: Model Development. *Journal of Hydrologic Engineering*, *30*(6), 04025039.

Studnicka, S., & Panu, U. (2025c). Streamflow Synthesis Using an Encoded Textural Pattern Recognition System. II: Model Applications. *Journal of Hydrologic Engineering*, *30*(6), 04025040.

Studnicka, S., & Panu, U. (2025d). Exploring the Role of Coloured Noise in Modelling Short- and Long-term Dependence Structures in Monthly Streamflows Across Ontario. *Journal of European Water 89*, 80–99.

Sudheer, K., Srinivasan, K., Neelakantan, T., & Srinivas, V. (2008). A nonlinear data-driven model for synthetic generation of annual streamflows. *Hydrological Processes: An International Journal*, *22*(12), 1831-1845.

Tang, Y., Sun, Y., Han, Z., Wu, Q., Tan, B., & Hu, C. (2023). Flood forecasting based on machine learning pattern recognition and dynamic migration of parameters. *Journal of Hydrology: Regional Studies*, *47*, 101406.

Tasker, G. D., & Dunne, P. (1997). Bootstrap position analysis for forecasting low flow frequency. *Journal of Water Resources Planning and Management*, *123*(6), 359-367.

Telesca, L., Lovallo, M., Lopez-Moreno, I., & Vicente-Serrano, S. (2012). Investigation of scaling properties in monthly streamflow and Standardized Streamflow Index (SSI) time series in the Ebro basin (Spain). *Physica A: Statistical Mechanics and its Applications*, *391*(4), 1662-1678.

Thomas, H., & Burden, R. P. (1963). Operations research in water quality management. (Harvard Wa•er Resources Group, Harvard University, Cambridge, Mass.)

Thomas, J., & Fiering, M. (1962). 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* (pp. 459-493). Harvard University Press.

Thompson, S. E., & Katul, G. G. (2012). Multiple mechanisms generate Lorentzian and 1/fα power spectra in daily stream-flow time series. *Advances in Water Resources*, *37*, 94-103.

Treistman, F., Maceira, M. E. P., Penna, D. D. J., Damázio, J. M., & Rotunno Filho, O. C. (2020). Synthetic scenario generation of monthly streamflows conditioned to the El Niño–Southern Oscillation: application to operation planning of hydrothermal systems. *Stochastic environmental research and risk assessment*, *34*(2), 331-353.

Tuceryan, M., & Jain, A. K. (1993). Texture analysis. *Handbook of pattern recognition and computer vision*, 235-276.

Unny, T. E., Panu, U. S., Macinnes, C. D., & Wong, A. K. (1981). Pattern analysis and synthesis of time-dependent hydrologic data. In *Advances in HYDROSCIENCE* (Vol. 12, pp. 195-295). Elsevier.

Vogel, R. M., & Stedinger, J. R. (1987). Generalized storage-reliability-yield relationships. *Journal of Hydrology*, *89*(3-4), 303-327.

Vogel, R. M., & Stedinger, J. R. (1988). The value of stochastic streamflow models in overyear reservoir design applications. *Water Resources Research*, *24*(9), 1483-1490.

Wang, W., Hu, S., & Li, Y. (2011). Wavelet transform method for synthetic generation of daily streamflow. *Water Resources Management*, *25*, 41-57.

Wang, X., Huang, G., & Liu, J. (2016). Observed regional climatic changes over Ontario, Canada, in response to global warming. *Meteorological Applications*, *23*(1), 140-149.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600-612.

Wen, H., & Liu, Z. (2016). Separating fractal and oscillatory components in the power spectrum of a neurophysiological signal. *Brain topography*, *29*, 13-26.

Wijayaratne, L. H., & Chan, P. C. (1987). Synthetic flow generation with stochastic models. Flood Hydrology: Proceedings of the International Symposium on Flood Frequency and Risk Analyses, 14–17 May 1986, Louisiana State University, Baton Rouge, USA,

Woods, R., & Sivapalan, M. (1999). A synthesis of space-time variability in storm response: Rainfall, runoff generation, and routing. *Water Resources Research*, *35*(8), 2469-2485.

Yaseen, Z. M. (2023). A new benchmark on machine learning methodologies for hydrological processes modelling: a comprehensive review for limitations and future research directions. *Knowledge-Based Engineering and Sciences*, *4*(3), 65-103.

Yevjevich, V. (1972). Stochastic processes in hydrology. *Water Resources Publication, Fort Collins, Colorado, USA*.

You, G. J.-Y., Thum, B.-H., & Lin, F.-H. (2014). The examination of reproducibility in hydro-ecological characteristics by daily synthetic flow models. *Journal of Hydrology*, *511*, 904-919.

Yue, S., & Wang, C. Y. (2002). Regional streamflow trend detection with consideration of both temporal and spatial correlation. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *22*(8), 933-946.

Zhang, H., Silva, F. H., Ohata, E. F., Medeiros, A. G., & Rebouças Filho, P. P. (2020). A two-dimensional approach based on transfer learning for alcoholism predisposition classification via EEG signals. *Frontiers in Human Neuroscience*, *14*, 365.