ENTROPY THEORY FOR STREAMFLOW FORECASTING

A Dissertation

by

HUIJUAN CUI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Vijay P. Singh
Committee Members,	Anthonny T. Cahill
	Ralph A. Wurbs
	Mohsen Pourahmadi
Intercollegiate Faculty Chair,	Ronald A. Kaiser

May 2015

Major Subject: Water Management and Hydrological Science

ABSTRACT

Entropy spectral analysis is developed for monthly streamflow forecasting, which contains the use of configurational entropy and relative entropy. Multi-channel entropy spectral analysis is developed for long-term drought forecasting with climate indicators.

The configurational entropy spectral analysis (CESA) is developed with both spectral power and frequency as random variables. With spectral power as a random variable, the configurational entropy spectral analysis (CESAS) identical to the original Burg entropy spectral analysis (BESA) when the underlying process is Gaussian. Through examination using monthly streamflow from the Mississippi Watershed, CESAS and BESA yield the same results and two methods are considered equivalent or as one method. With frequency as a random variable, the configurational entropy spectral analysis (CESAF) is developed and tested using monthly streamflow data from 19 river basins covering a broad range of physiographic characteristics. Testing shows that CESAF captures streamflow seasonality and satisfactorily forecasts both high and low flows. When relative drainage area is considered for analyzing streamflow characteristics and spectral patterns, it is found that upstream streamflow is forecasted more accurately than downstream streamflow.

Minimum relative entropy spectral analysis (MRESA) is developed under two conditions: spectral power as a random variable (RESAS) and frequency as a random variable (RESAF). The exponential distribution was chosen as a prior probability in the RESAS theory, and in the RESAF theory, the prior is chosen from the periodicity of

streamflow. Both MRESA theories were evaluated using monthly streamflow observed at 20 stations in the Mississippi River basin, where forecasted monthly streamflow shows higher reliability in the Upper Mississippi than in the Lower Mississippi.

The proposed univariate entropy spectral analyses are generally recommended over the classical autoregressive (AR) process for higher reliability and longer forecasting lead time. By comparing two MRESA theories with the two maximum entropy spectral analyses (MESA) (BESA and CESA), it is found that MRESA provided higher resolution in spectral estimation and more reliable streamflow forecasting, especially for multi-peak flow conditions. The MRESA theory is more accurate in forecasting streamflow for both peak and low flow values with longer lead time than MESA. Besides, choosing frequency as a random variable shows advantages over choosing spectral power. Spectral density estimated by the RESAF or CESAF theory shows higher resolution than the RESAS or BESA theory, respectively, and streamflow forecasted by RESAF or CESAF is more reliable than that by RESAS or BESA, respectively.

Finally, multi-channel entropy spectral analysis (MCESA) is developed for bi-variate or multi-variate time series forecasting. MCESA theory is verified by forecasting long-term standardized streamflow index with El Nino Southern Oscillation (ENSO) indicator. SSI was successfully forecasted using multi-channel spectral analysis with ENSO as an indicator. The monthly drought series is forecasted for lead times of 4-6 years by MCESA.

ACKNOWLEDGEMENT

This dissertation would not have been possible without the guidance from my committee, the help from my friends, and the support from my family.

I would like to express my most sincere gratitude to Dr. Vijay Singh, my committee chair, who led me to the world of hydrological science. His guidance, encouragement, support and patience has been indispensable to me throughout my graduate study, and the experience working with him will be a treasure for my future research life. I would like extend my appreciation to my committee members: Dr. Anthony Cahill, Dr. Ralph Wurbs, and Dr. Mohsen Pourahmadi, for helping me to develop a background in hydrology and statistics, and for their guidance during my research.

I would like to thank my friends and colleagues in our research group and department, Dr. Zengchao Hao, Dr. Di Long, Dr. Chao Li, Dr. Deepthi Rajsekhar, Dr. Clement Sohoulande, Gokhan Yildirim, Nandita Guar, Zhenlei Yang and Askarali Karimov for their help. Many thanks to my friends, Ying Liu, Yue Wang, Shuman Tan, Shanshui Yuan, Liang Chen, Si Gou, Shan Wang, Lin Jin and Xiaolei Xun for making my life in college station beautiful. Special thanks to Shanshui Yuan for helping to create maps.

Finally, I would like to thank my parents for their encouragement, and my husband Dr. Rui Wang for his understanding, support and love.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1 INTRODUCTION

Streamflow forecasting is the process of making estimates of streamflow volume or discharge in advance, based on available information. Streamflow forecasting plays an important role in flood control, reservoir operation and power generation, and provides useful information for water resources management. Timely streamflow forecasting enables civil protection authorities and public to make adequate preparation to reduce the impacts of floods or droughts. Based on the forecasting lead time, streamflow forecasting can be categorized as short-term forecasting for lead time less than 3 days, medium-term forecasting for 3 to 10 days, and long-term forecasting for longer than 10 days. Long term streamflow forecasting is important for reservoir storage and operation or water resources management and planning, while short-term forecasting is beneficial for flood warning and protection.

There are generally two classes of forecasting: deterministic methods and stochastic methods. Deterministic methods combine all hydrological processes to estimate future streamflow through predictor input, such as rainfall and snowmelt. However, as the forecasting lead time gets longer, the hydrological processes become more stochastic and more uncertainties are associated to predictors. In long-term streamflow forecasting, stochastic methods are therefore recommended, and time series analysis is often used (Hipel and McLeod, 1994; Noakes et al., 1985).

On the other hand, understanding time series characteristics of streamflow, such as seasonality and periodicity, is the key to improve the skill of streamflow forecasting. It

has been shown that spectral analysis is an efficient tool for characterizing the patterns of streamflow variation (Labat et al., 2005; Smith et al., 1998), identifying the periodicity of streamflow (Cengiz, 2011; Hameed, 1984; Sang et al., 2009; Sang et al., 2012), analyzing streamflow discontinuity (Adamowski and Prokoph, 2014), and examining the climatic influence on streamflow variability (Andreo et al., 2006; Kuhnel et al., 1990; Prokoph et al., 2012). It has been shown that spectral analysis is capable of extracting significant information for understanding the streamflow process and prediction thereof (Fleming et al., 2002; Ghil et al., 2002; Labat, 2005; Marques et al., 2006; Molenat et al., 1999). Thus, spectral analysis helps improve the reliability of forecasting.

Entropy, developed in thermodynamics, is a measure of system disorder. Since the development of Shannon (1948) entropy and the principle of maximum entropy (POME) (Jaynes, 1957a; b), the entropy theory has been widely applied in hydrology (Singh, 1997; 2011; Singh et al., 2007). However, it was not used for forecasting until Burg in 1960s. The development of Burg entropy or maximum entropy spectral analysis (MESA) (Burg, 1967; 1975) combines spectral analysis and time series analysis, and not only improves the resolution of the spectral density but also improves the reliability of prediction of streamflow. MESA has been applied to forecast hydrological series and it is recommended over classical methods (Krstanovic and Singh, 1989; 1991a; b; Singh, 2013a). Besides, there are other types of entropy, such as configurational entropy (Frieden, 1972; Gull and Daniell, 1978) and minimum relative entropy (Shore, 1979; 1981), that have been well known for signal processing and might be applicable to streamflow forecasting. This study

aims to develop methods based on these types of entropy to complete the entropy spectral theory for long-term streamflow forecasting.

Configurational entropy spectral analysis (CESA) is sometimes also referred to as maximum entropy method 2 (MEM2) or spectral MESA (SMESA) (Katsakos-Mavromichalis et al., 1985; Tzannes et al., 1985; Tzannes and Avgeris, 1981) since it is developed using the POME like Burg entropy spectral analysis (BESA). Superior to BESA, CESA is shown to not restrict to only AR process (Liefhebber and Boekee, 1987; Ortigueira et al., 1981). Configurational entropy has been later applied for spectral analysis and shown to have a better resolution than BESA for autoregressive moving average (ARMA) and moving average (MA) processes, and is comparable to BESA for the autoregressive (AR) process (Nadeu et al., 1981). On the contrary, Burg entropy appears to be better for white noise, as suggested by experiments on speech synthesis (Johnson and Shore, 1983; Katsakos-Mavromichalis et al., 1985). However, neither an explicit solution nor an equivalent extrapolation model had been developed until Wu (1983) used cepstrum analysis to derive an explicit solution for the configurational entropy.

Minimum relative entropy (MRE) theory, also called minimum cross-entropy (MCE), was introduced by Kullback (1959), which is an information-theoretic measure of the dissimilarity between two distributions. Two decades later, Shore (1979; 1981) developed the MRE based spectral analysis (MRESA) as an extension of Burg's maximum entropy spectral analysis, where the spectral power was considered as a random variable (Singh, 2013b). Later, another version of MRESA was developed by Tzaneess et al. (1985), considering frequency as a random variable. The MRE theory can determine the

spectra with the maximum value of kurtosis, which is the indicator reflecting the peskiness of the spectra (Endo and Randall, 2007). The MRE-based spectra are reported to have higher resolution and are more accurate in detecting peak location than other methods for spectral computation (Papademetriou, 1998). The theory refines the main frequencies and allows detection of very close peaks and does not create artificial peaks (Berger et al., 1990). When linking to time series analysis, the MRESA theory is equivalent to linear prediction with the smallest Itakura-Saito distortion (Gray et al., 1981; Schroeder, 1982; Shore, 1981; Shore and Gray, 1982). Beside, MRESA theory reduces the number of predictor coefficients by relying on the prior information (Schroeder, 1982). However, the MRE theory has only been applied for forward modeling and for solving inverse problems in groundwater (Woodbury and Ulrych, 1993; 1996; 1998), but has not been applied to streamflow forecasting yet.

1.1 Objectives

The main objective of this study is to improve monthly streamflow forecasting using entropy theory as well as long term drought forecasting. In general, two types of entropy spectral analysis, maximum entropy spectral analysis (MESA) and minimum relative entropy spectral analysis (MRESA), are developed by two different random variables as shown in Table 1-1. As shown in the table, the configurational entropy spectral analysis (CESA) to be developed and Burg entropy spectral analysis (BESA) developed by Burg (1967; 1975) form to maximum entropy spectral analysis (MESA).

Table 1-1 Entropy spectral analyses used in the study

Category	Entropy	Entropy spectral analysis	Random variable
Maximum entropy spectral analysis (MESA)	Burg entropy (BE)	BESA	Frequency
	Configurational entropy (CE)	CESA	Spectral power (CESAS) Frequency (CESAF)
Minimum relative entropy spectral analysis (MRESA)	Relative entropy (RE)	RESAS RESAF	Spectral power (RESAS) Frequency (RESAF)

The BESA was developed with frequency as a random variable, and will be revisited in the dissertation. Then the configurational entropy spectral analysis (CESA) will be first derived for streamflow forecasting in comparison with the Burg entropy using spectral power as a random variable (CESAS) and frequency as a random variable (CESAF). The minimum relative entropy spectral analysis (MRESA) will also be derived for two different procedures, one for considering spectral power as a random variable (RESAS) and the other for considering frequency as a random variable (RESAF). It is noted that when no prior information is given, RESAS and RESAF reduce to CESAS and CESAF respectively.

Finally, multichannel entropy spectral analysis (MCESA) will also be developed for long-term drought forecasting using climate indices.

Thus, the specific objectives of the study are stated as follows:

1. Develop configurational entropy spectral analysis with spectral power as a random variable for monthly streamflow forecasting
2. Develop configurational entropy spectral analysis with frequency as a random variable for monthly high-streamflow and low-streamflow forecasting.
3. Develop minimum relative entropy spectral analysis considering spectral power as a random variable for monthly streamflow forecasting.
4. Develop minimum relative entropy spectral analysis considering frequency as a random variable for monthly streamflow forecasting.
5. Compare the proposed entropy method with the classical AR method and Burg entropy spectral analysis, and compare maximum entropy spectral theories to minimum relative entropy spectral theories.
6. Develop multichannel-spectral analysis for long-term drought forecasting

1.2 Organization

The dissertation is organized in the following manner. A brief introduction about the study is provided in the first section, followed by literature review of the existing techniques. Mathematical preliminaries are provided in the third section, which introduces the theories of time series analysis, spectral analysis, cepstrum analysis and Burg entropy spectral analysis (BESA). Sections 4 and 5 develop the configurational entropy spectral analysis (CESA) for monthly streamflow forecasting with entropy in the form of Shannon

entropy. Spectral power is considered as a random variable in Section 4, which yields BESA in the end. On the contrary, frequency is considered as a random variable in Section 5. Sections 6 and 7develop two minimum relative entropy spectral analyses for monthly streamflow forecasting. In Section 6, spectral power is considered as a random variable, while in Section 7, frequency is considered as a random variable. Section 8 presents a comparison of the proposed four entropy spectral analysis methods. The multi-channel entropy spectral analysis is developed in Section 9 for long-term drought forecasting using ENSO teleconnections. The last section summarizes the dissertation.

2 LITERATURE REVIEW

There exists a multitude of methods for streamflow forecasting, which may roughly be divided into two categories: deterministic methods and stochastic methods. Deterministic methods are process-driven methods that consider streamflow as the output of the watershed system. Streamflow is forecasted through the underlying hydrological processes in the watershed, which involve rainfall, interception, evapotranspiration, infiltration, soil moisture or groundwater storage. Stochastic methods are data-driven methods. They identify mathematical connections between input and output without considering physical mechanisms of the streamflow process.

2.1 Deterministic forecasting methods

Deterministic methods can be roughly divided into rainfall-runoff models, watershed models, and hydrologic model-based ensemble streamflow prediction (ESP) methods. Watershed models or ESP are dependent on rainfall-runoff modeling. Hence, determination of the relationship between rainfall and runoff constitutes basis for ensemble streamflow prediction system (Wang et al., 2011).

Rainfall-runoff models can be categorized into lumped, semi-distributed, or distributed models (Sene, 2010; Singh, 1988). These models describe the rainfall and runoff relationship based on a combination of physically-based or empirical relations. A conceptual rainfall-runoff model can be used to forecast real-time streamflow, short-term streamflow and long-term streamflow. Thus, watershed models can be employed for forecasting. There are a large number of watershed models developed under different

conditions for different uses (Singh, 1995; Singh and Frevert, 2002a; b; 2006) For example, the simulation hydrology model (SIMHYD) (Chiew et al., 2002), the Systeme Hydrologique Europeen (SHE) (Abbott et al., 1986), TOPMODEL (Beven and Freer, 2001; Beven et al., 1984) can be used for streamflow forecasting. Because each model is developed for specific watersheds or topographies, its application is limited. Besides, the uncertainty of future rainfall is another drawback with watershed models.

ESP uses a conceptual hydrological model to forecast streamflow with weather information. A probabilistic forecast of future streamflow was provided by combining current snow, soil moisture, river and reservoir condition with historical meteorological data into model (Day, 1985). Wood et al. (2005) developed a climate model-based ESP approach for seasonal streamflow forecasting, of which precipitation and temperature produced from global spectral model were downscaled for use as forcing of the variable infiltration capacity hydrologic model. However, it was found that the accuracy was highly dependent on the determination of initial catchment state (Wood and Lettenmaier, 2008). Besides, applying ESP model required meteorological pre-processing, such as scale correction, under-dispersivity correction (Cloke and Pappenberger, 2009).

2.2 Stochastic forecasting methods

Stochastic methods have an advantage of using mathematical representation, as they avoid to represent the complex streamflow process under different conditions. Because of the difficulty of having an accurate knowledge of the streamflow process, stochastic methods became more popular in hydrology in the last two decades.

The stochastic methods fall into two categories, based on whether forecasting is done on other variables. One is the self-projecting approach, which entails univariate analysis and uses only past data to uncover its correlation to forecast future values. The other is cause-and-effect approach, which generates bi-variate or multivariate analysis by linking the series to be forecasted to one or more other series to which it is related.

2.2.1 Univariate forecasting

2.1.2.1 Time series model

Time series analysis is often used for forecasting monthly streamflow (Hipel and McLeod, 1994; Noakes et al., 1985).The most widely used time series models are autoregressive moving average (ARMA) and its derivatives, including autoregressive (AR), autoregressive integrated moving average (ARIMA), periodic ARMA (PARMA) or periodic AR (PAR), seasonal ARMA (SARMA) or seasonal AR (SAR), and fractionally integrated ARMA (ARFIMA).

AR and ARMA models are mathematically the simplest for time series forecasting, but their application is limited (Carlson et al., 1970; Haltiner and Salas, 1988; Jones and Brelsfor, 1967; Salas and Obeysekera, 1982). AR and ARMA were applied to forecast annual streamflow (Carlson et al., 1970; Hipel et al., 1977; Mcleod et al., 1977), but not directly applicable to streamflow with a time scale of less than 1 year. Because AR and ARMA were built under the assumption of stationary, though streamflow under 1 year usually exhibited strong seasonality or periodicity. Thus, ARIMA was developed to deal with the non-stationarity of the flow (Frausto-Solis et al., 2008). Instead, periodic models PAR or PARMA (Noakes et al., 1985; Salas and Obeysekera, 1992), SARMA (Salas et

al., 1982) were used for monthly streamflow forecasting. These models are sometimes even used for daily or shorter time scale streamflow forecasting (Abrahart and See, 2000; Toth et al., 2000). Later SARIMA was used to forecast monthly drought index with 2 or 3 month lead time (Durdu, 2010; Mishra and Desai, 2005; Modarres, 2007). ARFIMA was developed to deal with streamflow with long-memory condition, and applied to simulate monthly flow (Montanari et al., 2000; Ooms and Franses, 2001). However, the underlying linear assumption of above methods is the drawback and sometimes limits their application (Elshorbagy et al., 2002).

2.1.2.2 Kalman filter

Additional to the above time series methods, Kalman filter, also known as linear quadratic estimation, is an algorithm for forecasting future time series based on past observations developed by Kalman (1960). Kalman filter method operates recursively on noisy input data to produce a statistically optical estimate of underlying process. It was used for long-term periodic, daily, real-time forecasting, but all parameter matrices must be known (Bergman and Delleur, 1985a; b; Jimenez et al., 1989; Kitanidis and Bras, 1980a; b). Later, Ensemble transform Kalman filter (EnKF) approach is used to create initial uncertainties EPS forecast (Bishop et al., 2001; Cloke and Pappenberger, 2009; Reichle et al., 2002; Wei et al., 2006). However, standard implementation of EnKF is shown to be inappropriate because of the nonlinear states between the model and observed streamflow, and Clark et al.(2008) suggested transforming streamflow into log space.

2.1.2.3 Artificial neural networks

An artificial neural network (ANN) model is constructed from a flexible mathematical structure, thus it identifies the complex non-linear relationship between input and output data sets without understanding the nature of the process. Later, the accuracy in forecasting short-term streamflow had been increased by wavelet analysis in conjunction with ANN, though it is limited to a lead time less than a week (Adamowski, 2008; Kisi, 2009; 2010; Pramanik et al., 2011; Shiri and Kisi, 2010). ANN methods had an advantage in short-term forecasting (daily streamflow) over ARMA models, but they did not provide an explicit characterization and are unable to quantify physical conditions (Behzad et al., 2009; Frausto-Solis et al., 2008; Wu et al., 2009).

2.1.2.4 Entropy spectral method

The development of maximum entropy spectral analysis (MESA) by Burg (1967; 1975) entailed time series forecasting with entropy theory. It had been employed for long-term streamflow forecasting (Krstanovic and Singh, 1991a; b) and has been shown to have an advantage in long-term streamflow forecasting over traditional time series methods. Besides, the Burg-Levinson algorithm developed from MESA improved the prediction efficiency. However, due to the weakness in determining multi-peak spectral density for non-stationary conditions (Boshnakov and Lambert-Lacroix, 2012), MESA sometimes did not work well for monthly streamflow with strong seasonal and multi-periodic characteristics.

2.1.2.5 Nearest neighbor method

The nearest neighbor method (NNM) is a local approximation method. It divides the process into many subsets, if which the process has similar patterns, and then locally forecasts with a nonparametric or parametric model. It is good at approximation of nonlinear dynamics, and thus has an advantage in forecasting streamflow. The nonparametric NNM was applied to rainfall-runoff forecasting and shown comparable to ARMA with exogenous input (ARMAX) (Karlsson and Yakowitz, 1987a; b; Yakowitz, 1987). Later, it was found that the nonparametric NNM performed better than ARMA in one-step ahead daily discharge forecasting (Galeati, 1990) or is equivalent to ARMA for real-time flood forecasting (Toth et al., 2000). However, the nearest neighbor method is suited for large-sample time series and is limited to predict the values no higher than historic observations (Galeati, 1990; Karlsson and Yakowitz, 1987a; Toth et al., 2000).

2.2.2 Multivariate forecasting

Multivariate forecasting enables consideration of related variables, like rainfall, snowmelt, or other climate indices in streamflow forecasting. There are generally two categories that show how to use multi variables. One is by using a statistical relationship between climate series with streamflow (Chiew and McMahon, 2002; Sharma, 2000; Sharma et al., 2000; Westra et al., 2008), and the other by using a dynamic climate model (Wang et al., 2009). Involving climate indicators, a longer and more flexible range can be obtained for multi-variate forecasting, but the result is sensitive to the predictors so they should be carefully chosen.

Statistical relationships are widely used for multivariate forecasting. Regression analysis is one of the oldest methods for streamflow forecasting. Multiple regression models were developed to forecast monthly streamflow (Garen, 1992) or to forecast seasonal streamflow with snowpack, precipitation, and temperature(Pagano et al., 2009). With exogenous input, ARMA with exogenous input (ARMAX) model was used to forecast streamflow generated with additional rainfall input (Hannan and Kavalieris, 1984). MESA with multi-channel spectral analysis was applied to real-time streamflow forecasting by considering rainfall and streamflow as a bivariate vector (Krstanovic and Singh, 1993a; b).

A dynamic climate model was used to produce rainfall for forecasting seasonal streamflow by the Bayesian joint probability model (Wang and Robertson, 2011; Wang et al., 2009). It is a statistical approach for seasonal streamflow forecasting, where future streamflow was generated using Markov chain Monte Carlo samplings.

3 MATHEMATICAL PRELIMINARIES

This section introduces background theories which are fundamental to the dissertation study. In the first part, time series analysis is introduced with basic concepts of autocorrelation and linear prediction. Spectral analysis is presented in the second part with its definition and link to the time series variable. Cepstrum analysis introduced in the third part, has never been applied in hydrology, and will be used for developing configurational entropy spectral analysis (CESA) and minimum relative entropy spectral analysis (MRESA). The last part reviews Burg's maximum entropy spectral analysis (MESA).

3.1 Time series analysis

Consider observations of a random variable y over time t, and define y_t (y_1, y_2, …, y_T), a set of observations generated sequentially in time, t=1,… , T, where T is the total number of observations.

For a stationary process, it has a constant mean estimated by

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \tag{3-1}$$

and a constant variance, defined by

$$\sigma^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2 \text{ or } \sigma^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})^2 \tag{3-2}$$

For computational simplicity, data is often normalized as

$$y^* = (y - \bar{y})/\sigma \tag{3-3}$$

so that the mean of the y^* values is 0 and the variance is 1. For simplicity, from now onwards the normalized data will be represented by y instead of y^*.

3.1.1 Autocovariance and autocorrelation function

The autocovariance for a given lag n is defined as

$$\begin{aligned} R_n &= \operatorname{cov}(y_t, y_{t+n}) = \frac{1}{T-n}\sum_{t=1}^{T-n}(y_t - \bar{y})(y_{t+n} - \bar{y}) \\ &= \frac{1}{T-n}[(y_1 - \bar{y})(y_{1+n} - \bar{y}) + (y_2 - \bar{y})(y_{2+n} - \bar{y}) + \ldots + (y_{T-n} - \bar{y})(y_T - \bar{y}) \end{aligned} \tag{3-4}$$

when n=0, R_0 becomes variance:

$$\begin{aligned} R_0 &= \operatorname{cov}(y_t, y_t) = \frac{1}{T}[(y_1 - \bar{y})(y_1 - \bar{y}) + (y_2 - \bar{y})(y_2 - \bar{y}) + \ldots + (y_N - \bar{y})(y_N - \bar{y})] \\ &= \frac{1}{T}\sum_{t=1}^{T-n}(y_t - \bar{y})^2 = \sigma^2 \end{aligned} \tag{3-5}$$

Autocorrelation is obtained by dividing the autocovariance by variance, and varies from -1 to 1. The larger absolute value of autocorrelation means the values of variable are more correlated.

$$\rho_n = R_n / R_0 = \frac{\sum_{t=1}^{T-n}(y_t - \bar{y})(y_{t+n} - \bar{y})}{\sum_{t=1}^{T-n}(y_t - \bar{y})^2} \tag{3-6}$$

Specially, when n=0, autocorrelation yields

$$\rho_0 = R_0 / R_0 = 1 \tag{3-7}$$

Then the n by n autocorrelation matrix $\Re$ is defined as

$$\Re = E[Y^T Y] = \begin{bmatrix} R_0 & R_1 & \cdots & R_{n-1} & R_n \\ R_1 & R_0 & \cdots & R_{n-2} & R_{n-1} \\ \vdots & & & & \vdots \\ R_{n-1} & R_{n-2} & \cdots & R_0 & R_1 \\ R_n & R_{n-1} & \cdots & R_1 & R_0 \end{bmatrix} \tag{3-8}$$

It is seen from equation (3-8) that each element in the matrix is $\Re_{i,j} = \rho_{|i-j|}$. It is noted that the each descending diagonal from left to right is constant, thus the autocorrelation matrix $\Re$ is a Toeplitz matrix.

3.1.2 Spectral analysis

The streamflow time series y_t can represented with a wave-like function as the sum of sine and cosine functions as

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 \cos(2\pi f_1 t) + \beta_1 \sin(2\pi f_1 t) + \ldots + \alpha_q \cos(2\pi f_2 t) + \beta_q \sin(2\pi f_2 t) \\ &= \alpha_0 + \sum_{k=1}^{q} \alpha_k \cos(2\pi f_k t) + \beta_k \sin(2\pi f_k t) \end{aligned} \tag{3-9}$$

where frequency $f_k = k/T$. Equation (3-9) is also called Fourier series where α_k, β_k are the Fourier coefficients, which can be computed as

$$\alpha_0 = \bar{y} \tag{3-10a}$$

$$\alpha_k = \frac{2}{T} \sum_{t=1}^{T} y_t \cos(2\pi f_k t) \tag{3-10b}$$

$$\beta_k = \frac{2}{T} \sum_{t=1}^{T} y_t \sin(2\pi f_k t) \tag{3-10c}$$

It is noted that when considering normalized data, α_0=0 and hence it can be ignored.

Thus, spectral power at a given frequency is defined as

$$x_k = \frac{T}{2}(\alpha_k{}^2 + \beta_k{}^2) \tag{3-11}$$

For each frequency f_k, there is one corresponding spectral power. The highest frequency is 0.5 cycles per time interval since the smallest period is 2 intervals. For example, if the time interval is considered 1 month in the case of monthly streamflow, the smallest period of the cycle T is 2 months. Accordingly, $f=1/T$ giving the largest frequency as 0.5.

The frequencies $f_k = k/T$ are called harmonics of the fundamental frequency $1/N$, and allow f to vary continuously from 0 to 0.5. Thus, with continuous frequency, the power spectrum is defined as

$$G(f) = \frac{N}{2}(\alpha_f{}^2 + \beta_f{}^2) \tag{3-12}$$

where

$$\alpha_f = \frac{2}{N}\sum_{t=1}^{N} y_t \cos(2\pi ft) \tag{3-13a}$$

$$\beta_f = \frac{2}{N}\sum_{t=1}^{N} y_t \sin(2\pi ft) \tag{3-13b}$$

The spectral density is defined by dividing the power spectrum by variance, which is

$$p(f) = G(f)/\sigma^2 \tag{3-14}$$

where the variance σ^2 is computed from equation (3-2). Thus, spectral density can be considered as normalized power spectrum.

It is noted that the power spectrum or the spectral density can be linked with autocovariance/autocorrelation function, respectively, as follows:

$$G(f) = \sum_{k=-N+1}^{N-1} R_k \exp(-i2\pi fk) \quad f \in [0, 0.5] \tag{3-15a}$$

$$p(f) = \sum_{k=-N+1}^{N-1} \rho_k \exp(-i2\pi fk) \quad f \in [0, 0.5] \tag{3-15b}$$

It may note here that integration of *G(f)* over *f* between limits 0 to 0.5 is

$$\int_0^{0.5} G(f)df = \int_0^{0.5} \sum_{k=-N+1}^{N-1} R_k \exp(-i2\pi fk)df = \sum_{k=-N+1}^{N-1} R_k \int_0^{0.5} \exp(-i2\pi fk)df = R_0 \tag{3-16a}$$

and the integration of *p(f)* over *f* between limits 0 to 0.5 is

$$\int_0^{0.5} p(f)df = 1 \tag{3-16b}$$

It can be found from equations (3-16) that the power spectrum or the spectral density is in the form of the Fourier transform of the autocovariance or autocorrelation. Thus, inversely, autocovariance or autocorrelation can be written using the inverse Fourier transform as:

$$R_k = \int_{-W}^{W} G(f)e^{i2\pi fk\Delta t}df, -N \le k \le N \tag{3-17a}$$

$$\rho_k = \int_{-W}^{W} p(f)e^{i2\pi fk\Delta t}df, -N \le k \le N \tag{3-17b}$$

3.1.3 Linear prediction

Univariate time series forecasting predicts future events from past information. A linear prediction method assumes that the future series y_{T+t}, for $t>0$, $T>0$ can be forecasted using a linear combination of past values, which can be written as

$$\hat{y}_{T+t} = \sum_{j=1}^{m} a_j y_{T+t-j} \text{ , for } t>0 \tag{3-18}$$

where a_j is the prediction coefficient, and m is the model order. Equation (3-18) is also recognized as the autoregressive (AR) model.

The mean squared error of prediction can be denoted as

$$\upsilon = E(y_n - \hat{y}_n)^2 \tag{3-19}$$

Substituting $\hat{y}_n$ with equation (3-18) yields

$$\upsilon = E(y_n - \sum_{j=1}^{m} a_j y_{n-j})^2 = E(-\sum_{j=0}^{m} a_j y_{n-j})^2 = \sum_{k=0}^{m} a_k \sum_{j=0}^{m} a_j \rho_{n-j} \tag{3-20}$$

where $a_0 = -1$. It can be seen from equation (3-20) that to minimize the mean squared error, the right hand side of equation (3-20) must satisfy that

$$-\sum_{j=0}^{m} a_j \rho_{n-j} = \rho_n - \sum_{j=1}^{m} a_j \rho_{n-j} = 0 \tag{3-21a}$$

Thus,

$$\rho_n = \sum_{j=1}^{m} a_j \rho_{n-j} \tag{3-21b}$$

It is noted that equation (3-21b) is called the Yule-Walker equation (Yule, 1927).

The spectral density of the linear prediction is defined as

$$p(f) = \frac{\sigma^2}{\left|1 + \sum_{n=-N}^{N} a_n \exp(-i2\pi mfk)\right|^2} \tag{3-22}$$

3.1.4 Levinson algorithm

The prediction coefficient of Equation (3-21) can be computed from either Newton's method or Levinson algorithm (Levinson, 1946). The Levinson algorithm, also called the Durbin-Levinson algorithm, is a recursive algorithm and has an advantage over Newton's method as it involves the order of N^2 operations with memory storage on the order of N, and is faster than the general Gaussian elimination procedure.

The Levinson algorithm states that if y_t is a zero mean stationary process with autocovariance function R such that $R_0>0$ and $R_h \to 0$ as $h \to \infty$, then the forecasting coefficients $a_{n,j}$ (nth order jth coefficient) and mean squared errors υ_n defined by equations (3-18) and (3-19) satisfy $a_{1,1} = \rho_1$, $\upsilon_0 = R_0$,

$$a_{n,n} = \frac{R_n - \sum_{j=1}^{n-1} a_{n-1,j} R_{n-j}}{v_{n-1}} \tag{3-23a}$$

$$a_{n,j} = a_{n-1,j} - a_{n,n} a_{n-1,n-j}, \text{ for } j<n \tag{3-23b}$$

and $v_n = v_{n-1}(1 - a_{n,n}{}^2)$. (3-24)

By repeating equations (3-22), the coefficients are solved recursively from the values of previous order. The detailed proof can be found in most time series books, such as Box and Jenkins (1970) and Brockwell and Davis (1987).

3.2 Cepstrum analysis

Cepstrum, by its definition, is the inverse Fourier transform of the logarithm of spectrum. It is a measure of the rate of change in the spectrum bands, and is an efficient technique along with spectral analysis for separation and recovery of time series, as the homomorphic characteristics of time series are reminiscent of the cepstrum (Oppenheim and Schafer, 2004).

3.2.1 Definition

For a given streamflow time series $y(t)$, cepstrum can be computed using the following steps:

First, taking the Fourier transform of the original series $y(t)$, one obtains

$$Y(f) = FT[y(t)] = \sum_{n=-\infty}^{\infty} y(t) e^{-2\pi i f} \tag{3-25}$$

where $Y(f)$ is the Fourier transform of $y(t)$.

Second, taking the inverse Fourier transform of the log-magnitude of equation (3-25) one obtains the cepstrum of the Fourier transform as

$$C(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|Y(f)| e^{2\pi i f} df \tag{3-26}$$

It is stated in equation (3-15) that the spectral density is the Fourier transform of the autocorrelation, $p(f) = FT[\rho(n)]$. Thus, the cepstrum of autocorrelation can be defined by the inverse Fourier transform of the log-magnitude of $p(f)$, which yields

$$e(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|p(f)| e^{2\pi i f} df \tag{3-27}$$

However, it is known that the spectral density by definition can also be written as

$$p(f) = Y(f)^2 = \{FT[y(t)]\}^2 \tag{3-28}$$

Thus, the following relationship between the cepstrum of the series and the cepstrum of the autocorrelation can be obtained:

$$e(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|FT[y(t)]\right|^2 e^{2\pi nif}\, df = \frac{1}{2\pi}\int_{-\pi}^{\pi} 2\log\left|Y(f)\right| e^{2\pi nif}\, df = 2C(n) \tag{3-29}$$

3.2.2 Cepstrum analysis for finite data

Consider only the positive part of the autocorrelation function $\rho(n)$, for n>0 which is dependent on the one-sided autocorrelation instead of two-sided autocorrelation as

$$\rho^*(n) = \begin{cases} \rho(n), & , n > 0 \\ 1, & n = 0 \\ \rho(-n), & n < 0 \end{cases} \tag{3-30}$$

and let $e(n)$ be the cepstrum estimated from $\rho^*(n)$, which is

$$e(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|p^*(f)\right| e^{2\pi nif}\, df \tag{3-31}$$

where the spectral density $p^*(f)$ is obtained by Fourier transform from the positive half of $\rho(n)$, for n>0. It is noted that $p^*(f)$ is analytical. Equation (3-30) ensures that $\rho(n)$ is causal and stable. Thus, it is a minimum-phase function and for a minimum phase system the input and output are uniquely determined. It means if $e(n)$ is considered as the output from equation (3-30) and $\rho(n)$ as the input, $e(n)$ can be uniquely determined from $\rho(n)$. Let us define a two-sided output in the way such that

$$\hat{\rho}(n) = \begin{cases} 2e(n), & , n > 0 \\ e(0), & n = 0 \\ 2e(-n), & n < 0 \end{cases} \tag{3-32}$$

In such a way, $\hat{\rho}(n)$ can also be uniquely determined by $\rho(n)$ and vice versa.

Since $p^*(f)$ is analytical, $\log p^*(f)$ can also be considered as analytical. In such a case, following Oppenheim and Schafer (1975), there is the following relationship between the derivatives of z transformed $\hat{\rho}(n)$ and $\rho(n)$:

$$\hat{\rho}'(z)=\frac{\rho'(z)}{\rho(z)} \tag{3-33}$$

By multiplying z on both sizes of equations, Oppenheim and Schafer (1975) show equation (3-33) is equivalent to

$$z\hat{\rho}'(z)=\sum[-n\hat{\rho}(n)]z^{-n+1}=\frac{z\rho'(z)}{\rho(z)} \tag{3-34}$$

Then, the following equation can be obtained from equation (3-34):

$$z\rho'(z)=z\hat{\rho}'(z)\rho(z) \tag{3-35}$$

Taking the inverse z transform of equation (3-32), one obtains

$$n\rho(n)=\sum_{k=\infty}^{\infty}k\hat{\rho}(n)\rho(n-k) \tag{3-36}$$

Dividing equation (3-34) by n, the relationship between input and output becomes

$$\rho(n)=\sum_{k=-\infty}^{\infty}(\frac{k}{n})\hat{\rho}(n)\rho(n-k) \tag{3-37}$$

Transforming equation (3-37) with the use of equation (3-32), the autocorrelation function can be obtained from the following recursive formula (Oppenheim and Schafer, 1975):

$$\rho(n) = \begin{cases} \dfrac{\rho(0)}{2} e(n) + \sum_{k=1}^{n-1} \dfrac{k}{n} e(k)\rho(n-k)], & n > 0 \\ \exp e(n), & n = 0 \\ \rho(-n), & n < 0 \end{cases} \quad \text{(3-38)}$$

On the other hand, cepstrum $e(n)$ can be obtained from the reverse relation of equation (3-38) as:

$$e(n) = \begin{cases} \dfrac{2}{\rho(0)} [\rho(n) - \sum_{k=1}^{n-1} \dfrac{k}{n} e(k)\rho(n-k)], & n > 0 \\ \log \rho(n), & n = 0 \\ 0, & n < 0 \end{cases} \quad \text{(3-39)}$$

Thus, equations (3-38) and (3-39) explain the relationship between autocorrelation and cepstrum for finite time series.

3.3 Burg entropy spectral analysis

Burg (1967; 1975) defined entropy in the frequency domain and developed what is now called Burg entropy theory (BE). He developed "maximum entropy spectral analysis (MESA)" for time series forecasting, which is referred to as Burg entropy spectral analysis (BESA) in this study. BESA extends autocorrelation in a manner that maximizes the entropy of the underlying process. For a stationary random process BESA computed spectral power from autocorrelation of given lags, without assuming autocorrelation of unknown lags as zero (Edward and Fitelson, 1973). It had an advantage over classical methods in computational ease, the short and smooth spectra with a high degree of resolution, and the robustness of the estimates and their stability. As a result, it had been widely applied to spectral analyses of geomagnetic and climate indices, surface air

temperature, tide levels, precipitation, and runoff series (Currie, 1973; Dalezios and Tyraskis, 1989; Ghil et al., 2002; Hasanean, 2001; Padmanabhan and Rao, 1988; Pardo-Iguzquiza and Rodriquez-Tovar, 2006; Sang et al., 2009; Sang et al., 2012; Tosic and Unkasevic, 2005; Wang et al., 2004). Besides, Burg (1967; 1975) modified the Levinson algorithm provided a new recursive method to compute the AR parameters, which was noted as Burg-Levinson algorithm. The development of BESA and the Burg-Levinson algorithm is reviewed in this section.

3.3.1 Definition of Burg entropy

Burg (1967; 1975) defines entropy as

$$H(f) = \int_{-W}^{W} \ln[p(f)]df \tag{3-40}$$

where $W=1/(2\Delta t)$ is the Nyquist fold-over frequency and f is the frequency that varies from $-W$ to W, Δt is the sampling period, and $p(f)$ is the spectral density function. Here f is treated as a random variable. For example, for monthly streamflow, the sampling period is 1 month, thus Δt=1month, and $W=1/(2\Delta t)$=0.5, and the frequency f varies from -0.5 to 0.5.

3.3.2 Specification of constraints

From the definition of autocorrelation function

$$\rho_n = \int_{-W}^{W} p(f) e^{i2\pi f n\Delta t} df \ , \ -N \le n \le N \tag{3-41}$$

where N is usually taken from ¼ up to ½ of the total series. The basic assumption here is that the stationary time series being analyzed is a random or an unpredictable time series that is consistent with the measurement from equation (3-41).

3.3.3 Determination of spectral density

To obtain the least-biased spectral density, the entropy defined in equation (3-40) is maximized subject to equation (3-41) according to the principle maximum entropy theory (Jaynes, 1957a; b). The Lagrange function can be written as

$$L(f) = \int_{-W}^{W} \{\ln[p(f)] - \sum_{n=N}^{N} \lambda_n [p(f) e^{i2\pi fn\Delta t} - \rho_n]\} df \tag{3-42}$$

Taking the partial derivative of $L(f)$ with respect to $p(f)$, and equating the derivative to 0, one gets

$$\frac{\partial L(f)}{\partial p(f)} = \int_{-W}^{W} \{\frac{1}{p(f)} - \sum_{n=N}^{N} \lambda_n e^{i2\pi fn\Delta t}\} df = 0 \tag{3-43}$$

Thus, the spectral density becomes

$$p(f) = \frac{1}{\sum_{n=-N}^{N} \lambda_n e^{-i2\pi fn\Delta t}} \tag{3-44}$$

$$\text{Let } z = e^{-i2\pi f\Delta t}. \tag{3-45}$$

Then, $dz = -i2\pi\Delta t e^{-i2\pi f\Delta t} df = -i2\pi\Delta t z df$ and $df = \frac{dz}{-i2\pi\Delta t z}$. Equation (3-44) can be written by replacing $e^{-i2\pi f\Delta t}$ by z, which is called z transform, expressed as

$$p(f) = \frac{1}{\sum_{n=-N}^{N} \lambda_n z^n} \tag{3-46}$$

It is noted that the spectral density in equation (3-44) obtained from BESA is in the same form of linear prediction as shown in equation (3-22). Thus, by equating the two equations,

$$p(f) = \frac{\sigma^2}{\left|1 + \sum_{n=-N}^{N} a_n z^n\right|^2} = \frac{1}{\sum_{n=-N}^{N} \lambda_n z^n} \tag{3-47}$$

Thus, the following relationship between the prediction coefficient and Lagrange multipliers: can be found as

$$\begin{aligned} \sum_{s=-N}^{N} \lambda_s z^s &= \sigma^{-2}\left[1 + a_1 z + a_2 z^2 + \ldots + a_N z^Z\right] \quad \left[a + a_1^* z^{-1} + a_2^* z^{-2} + \ldots + a_N^* z^{-N}\right] \\ &= \sigma^{-2} \sum_{s=0}^{N} a_s z^s \sum_{s=0}^{N} a_s^* z^{-s} \end{aligned} \tag{3-48}$$

where a_s are the prediction coefficients and a_s* are their complex conjugates.

3.3.4 Estimation of Lagrange multipliers

The Lagrange multipliers can be computed by either inputting equation (3-46) into the constraint equation (3-41) or from equation (3-48) with the prediction coefficients first. Instead of using equation (3-46), substituting equation (3-47) into equation (3-41), one obtains

$$\rho_n = \int_{-W}^{W} \frac{\sigma^2}{\left|1 + \sum_{n=-N}^{N} a_n z^n\right|^2} z^n df = \frac{P_N}{2\pi i} \oint \frac{z^{-n-1} dz}{\sum_{s=0}^{N} a_s z^s \sum_{s=0}^{N} a_s^* z^{-s}}, \; -N < n < N \tag{3-49}$$

where P_N=$\sigma^2/\Delta t$. Multipling equation (3-46) by a_n^* and summing up for n from 0 to N,

$$\sum_{n=0}^{N} a_n^* \rho_{n-r} = \frac{P_N}{2\pi i} \oint \frac{z^{r-1} dz \sum_{n=0}^{N} a_n^* z^{-n}}{\sum_{s=0}^{N} a_s z^s \sum_{s=0}^{N} a_s^* z^{-s}} = \frac{P_N}{2\pi i} \oint \frac{z^{r-1} dz}{\sum_{s=0}^{N} a_s z^s}, \; r \geq 0 \tag{3-50}$$

Cauchy's integral formula gives

$$\oint \frac{z^{r-1} dz}{\sum_{s=0}^{N} a_s z^s} = \begin{cases} 2\pi i, & r = 0 \\ 0, & r \geq 1 \end{cases} \tag{3-51}$$

Thus, the integration of equation (3-50) yields

$$\sum_{n=0}^{N} a_n^* \rho_{n-r} = \frac{P_N}{2\pi i} \oint \frac{z^{r-1} dz}{\sum_{s=0}^{N} a_s z^s} = \frac{P_N}{2\pi i} 2\pi i = P_N, \quad r = 0 \tag{3-52a}$$

$$\sum_{n=0}^{N} a_n^* \rho_{n-r} = \frac{P_N}{2\pi i} \oint \frac{z^{r-1} dz}{\sum_{s=0}^{N} a_s z^s} = \frac{P_N}{2\pi i} 0 = 0, \quad r \geq 1 \tag{3-52b}$$

Since P_N is real, taking the conjugate of equations (3-52) yields

$$\sum_{n=0}^{N} \rho_{r-n} a_n = P_N, \quad r = 0 \tag{3-53a}$$

$$\sum_{n=0}^{N} \rho_{r-n} a_n = 0, \quad r \geq 1 \tag{3-53b}$$

which can be written in matrix form as:

$$\begin{bmatrix} \rho_0 & \rho_{-1} & \cdots & \rho_{-N} \\ \rho_1 & \rho_0 & & \rho_{1-N} \\ \vdots & \vdots & & \vdots \\ \rho_N & \rho_{N-1} & \cdots & \rho_0 \end{bmatrix} \begin{Bmatrix} 1 \\ a_1 \\ \vdots \\ a_N \end{Bmatrix} = \begin{Bmatrix} P_N \\ 0 \\ \vdots \\ 0 \end{Bmatrix} \tag{3-54}$$

Equation (3-54) is the matrix form of equation (3-21). It is to be noted that the autocorrelation matrix is positive definite only if P_N is positive, when P_N=0, the matrix is

singular. Again, like the linear prediction method, P_N and a_n can be solved for using Newton's method, the Durbin-Levinson algorithm or the Burg-Levinson algorithm developed by Burg (1967).

3.3.5 Levinson-Burg algorithm

It is shown in Section 3.1.4 that the Levinson algorithm [equation (3-23)] uses the coefficient and error computed in the previous order into the computation of next order. Burg improved the Levinson algorithm by computing forward and backward prediction error together to update the coefficient of next order(Collomb, 2009; Lin and Wong, 1990).

The forward prediction error can be defined as

$$f_m = a_{m-1} y_2 + \ldots + a_1 y_m + y_{m+1} \tag{3-55}$$

and backward prediction error as

$$b_m = y_1 + a_1 y_2 + \ldots + a_{m-1} y_m \tag{3-56}$$

Then, the mean squared error can be computed from weighing forward and backward errors by

$$MSE = \sum_{m=1}^{M} w_m \left| f_m + a_{n,n} b_m \right|^2 = \sum_{m=1}^{M} w_m \left| b_m + a_{nn} f_m \right|^2 \tag{3-57}$$

where w_m are weights and $\sum_{m=1}^{M} w_m = 1$.

Minimizing equation (3-57), one has

$$a_{n,n} = \frac{-2\sum_{m=1}^{M} w_m (b_m f_m)}{\sum_{m=1}^{M} w_m (b^2{}_m + f^2{}_m)} \tag{3-58}$$

Then Burg (1967) suggested to use equation (3-58) instead of (3-23a) to compute the prediction coefficients.

3.3.6 Extension of autocorrelation

Burg (1967; 1975) suggested the extension of autocorrelation under two circumstances. First, when matrix R(N) is singular, one can have

$$\sum_{j=0}^{m} \rho(N-j)a_j = 0 \tag{3-59}$$

Thus, autocorrelation can be extended as

$$\rho(N) = -\sum_{j=1}^{m} \rho(N-j)a_j \tag{3-60}$$

Otherwise, if matrix R(N) is non-singular, one can have

$$\sum_{j=0}^{m} \rho(N-j)a_j = -a_{n,n}\upsilon_{n-1} \tag{3-61}$$

In this case, autocorrelation is extended as

$$\rho(N) = -\sum_{j=1}^{m} \rho(N-j)a_j - a_{n,n}\upsilon_{n-1} \tag{3-62}$$

3.3.7 Time series forecasting

Time series can be forecasted using the coefficient of extended autocorrelation. To prove, let us assume a linear prediction equation as

$$\hat{y}_{T+t} = \sum_{j=1}^{m} b_j y_{T+t-j} \tag{3-63}$$

The mean square error of prediction then becomes

$$MSE = \overline{\sum_{m=0}^{N} b_m^* y_{T-m}^* \sum_{n=0}^{N} b_n y_{T-n}} = \sum_{m=0}^{N}\sum_{n=0}^{N} b_m^* \overline{y_{T-m}^* y_{T-n}} b_n = \sum_{m=0}^{N}\sum_{n=0}^{N} b_m^* \rho_{m-n} b_n \tag{3-64}$$

If autocorrelation matrix is positive definite, one has

$$\begin{bmatrix} \rho_0 & \rho_{-1} & \cdot & \rho_{1-N} & \rho_{-N} \\ \rho_1 & \rho_0 & \cdot & \rho_{2-N} & \rho_{1-N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{N-1} & \rho_{N-2} & \cdot & \rho_0 & \rho_{-1} \\ \rho_N & \rho_{N-1} & \cdot & \rho_{-1} & \rho_0 \end{bmatrix} \begin{Bmatrix} 1 \\ a_1 \\ \cdot \\ a_{N-1} \\ a_N \end{Bmatrix} = \begin{Bmatrix} \upsilon_N \\ 0 \\ \cdot \\ 0 \\ 0 \end{Bmatrix} \tag{3-65}$$

Thus, using equation (3-60),

$$\begin{aligned} &\sum_{m=0}^{N}\sum_{n=0}^{N} (a_m^* - b_m^*)\rho_{m-n}(a_n - b_n) \\ &= \sum_{m=0}^{N}\sum_{n=0}^{N} (a_m^*\rho_{m-n}a_n - a_m^*\rho_{m-n}b_n - b_m^*\rho_{m-n}a_n + b_m^*\rho_{m-n}b_n) \\ &= \sum_{m=0}^{N}\sum_{n=0}^{N} b_m^*\rho_{m-n}b_n - \upsilon_N \end{aligned} \tag{3-66}$$

Thus, by rearranging one obtains

$$MSE = \sum_{m=0}^{N}\sum_{n=0}^{N} b_m^*\rho_{m-n}b_n = \upsilon_N + \sum_{m=0}^{N}\sum_{n=0}^{N} (a_m^* - b_m^*)\rho_{m-n}(a_n - b_n) \tag{3-67}$$

It is clear that the minimum value occurs when b_n=a_n and *MSE* becomes υ_N, which means that the forecast coefficients need to be equal to the coefficients associated with extended with autocorrelation so that prediction satisfies the least square prediction.

3.4 Goodness of fit measurements

Goodness of fit measures are needed to evaluate the performance of a model and the are given in this section.

3.4.1 Itakura-Saito distortion

The performance of spectral estimation will be evaluated by Itakura-Saito distortion, which is a measure of the perceptual difference between an original spectrum and its estimate. The distortion is defined as

$$D_{I-S}(\hat{p}(f), p(f)) = \frac{1}{2\pi}\int [\frac{p(f)}{\hat{p}(f)} - \log(\frac{p(f)}{\hat{p}(f)}) - 1]df \tag{3-68}$$

where $p(f)$ represents the spectral density from FFT and $\hat{p}(f)$ is the estimated spectral density. The smaller value represents a better fit.

3.4.2 Relative error

The relative error, sometimes called absolute error, provides the average magnitude of differences between observed values and predicted values relative to observed values. It is computed as

$$RE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Q_f(i) - Q_o(i)}{Q_o(i)}\right| \tag{3-69}$$

where $Q_o(i)$ is the i-th observed streamflow; and $Q_f(i)$ is the i-th forecasted streamflow. It is noted that relative error is dimensionless.

3.4.3 Root mean square error

RMSE also represents the difference between observed and predicted values, however, it is scale-dependent.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Q_o(i) - Q_f(i))^2}{N-1}} \tag{3-70}$$

where $Q_o(i)$ is the i-th observed streamflow; and $Q_f(i)$ is the i-th forecasted streamflow.

3.4.4 Coefficient of determination

The coefficient of determination is defined as the square of the coefficient of correlation. It represent how well data fit the model. It ranges between 0 and 1, and its high values indicate better prediction.

$$r^2 = \left\{ \frac{\sum_{i=1}^{N} (Q_o(i) - \overline{Q_o})(Q_f(i) - \overline{Q_f})}{\left[\sum_{i=1}^{N} (Q_o(i) - \overline{Q_o})^2 \right]^{0.5} \left[\sum_{i=1}^{N} (Q_f(i) - \overline{Q_f})^2 \right]^{0.5}} \right\}^2 \tag{3-71}$$

where $Q_o(i)$ is the i-th observed streamflow, $Q_f(i)$ is the i-th forecasted streamflow, and $\overline{Q_o}$ and $\overline{Q_f}$ are the average values of observed and computed discharges, respectively.

3.4.5 Nash-Sutcliffe Efficiency

The Nash-Sutcliffe efficiency coefficient, defined by Nash and Sutcliffe [1970], ranges from negative infinity to 1. Higher values of *NSE* represent more agreement between model predictions and observations, and negative values indicate the model is worse than the mean value as a predictor (Nash and Sutcliffe, 1970; Schaefli and Gupta, 2007). However, it is known that r^2 and *NSE* are more sensitive to extreme values than the observations near the mean (Legates and Davis, 1997; Legates and McCabe, 1999). Therefore, *NSE* leads to overestimation of peak flows and underestimation of low flows. Thus, a modified *NSE* was suggested by Krause et al. (2005) and Legates and McCabe (1999) for a better overall evaluation, which is

$$NSE = 1 - \frac{\sum_{i=1}^{N} \left| Q_o(i) - Q_f(i) \right|^j}{\sum_{i=1}^{N} \left| Q_o(i) - \overline{Q_o} \right|^j} \tag{3-72}$$

where $Q_o(i)$ is the i-th observed streamflow, $Q_f(i)$ is the i-th forecasted streamflow, and $\overline{Q_o}$ is the average values of observed discharge.

4 CONFIGURATIONAL ENTROPY SPECTRAL ANALYSIS WITH SPECTRAL POWER AS A RANDOM VARIABLE

The maximum entropy spectral analysis was introduced in the previous section. One may noted that the definition of entropy Burg (1967) used was not the form often used. The Burg entropy is defined as the integration of the logarithm of the probability of frequency, however, the entropy often used is the expectation of the logarithm of the probability, which is in the form of Shannon (1948) entropy. Thus, one may wonder if it would be possible to use the form of Shannon entropy, which is called configurational entropy in spectral analysis introduced by Frieden (1972) and Gull and Daniell (1978), to derive the maximum entropy spectral analysis. Thus, in the following two sections, the maximum entropy spectral analysis will be developed using the configurational entropy with spectral power and frequency as a random variable in each section.

4.1 Development of configurational entropy spectral analysis

The configurational entropy spectral analysis (CESAS) is developed using spectral power as a random variable. Therefore, unlike the Burg entropy spectral analysis, the probability density function of spectral power is first obtained by maximizing entropy, and then the spectral density. It consists of the following steps: 1) define the entropy, 2) construct the constraints, 3) determine the probability density function, 4) determine the Lagrange multipliers, 5) determine the spectral power, 6) extend the autocorrelation or autocovariance, and 7) forecast streamflow.

4.1.1 Definition of entropy

Consider the spectral power x_k defined in equation (3-11) as a random variable. Let the streamflow time series be described by $\vec{x} = (x_1, x_2, ..., x_n)$ and the joint probability density by $p(\vec{x})$. In order to determine the probability density, each $p(x_k)$ is considered independent identically distributed. Thus, entropy is defined as

$$H(p) = -\int_D p(\vec{x}) \ln[p(\vec{x})] d\vec{x} \tag{4-1}$$

However, it was shown by Gray (1977) if x_k came from an N-dimensional Gaussian distribution, where the joint distribution is given by

$$p(\vec{x}) = (\frac{1}{2\pi})^{\frac{n}{2}} (\frac{1}{\det \Re})^{\frac{1}{2}} \exp(-\frac{1}{2} \vec{x}^t \Re^{-1} \vec{x}) \tag{4-2}$$

where $\Re$ is autocorrelation matrix defined in equation (3-8). Substituting equation (4-2) into equation (4-1) yields

$$H(p) = \log[(2\pi e)^{\frac{N}{2}} (\det \Re)^{\frac{1}{2}}] \tag{4-3}$$

It was shown in the section 3.1.2 that the autocorrelation is linked to the spectral density. Thus, replacing the autocorrelation in equation (4-3) with spectral density as shown in (3-12), the entropy will yield as $H(f) = \int_{-W}^{W} \ln[p(f)] df$, which is the Burg entropy. Nevertheless, how the configurational entropy spectral analysis yields Burg entropy spectral analysis will be shown through the derivation of maximum entropy theory.

4.1.2 Specification of constraint

First of all, the probability density function must satisfy that

$$\int_D p(\vec{x})d\vec{x} = 1 \tag{4-4}$$

The other constraints are constructed from the relationship between spectral power and autocorrelation. Let S_k denote the expected value of x_k, written as

$$S_k = \int_D x_k p(\vec{x})d\vec{x} = \int_D x_k p(x_k)dx_k \tag{4-5}$$

It is shown in Section 3.1.2 that autocovariance can be expressed as the inverse Fourier transform of S_k as

$$R_r = \sum_{k=1}^{n} S_k \exp(2\pi i r \Delta t f_k) = \sum_{k=-n}^{n} S_k c_{rk} \qquad -N \le r \le N \tag{4-6}$$

where $c_{rk} = \exp(2\pi i r \Delta t f_k)$, r is the lag, Δt is taken as 1 month for monthly streamflow, and N is the largest lag for given autocorrelation or autocovariance, usually taken as ¼ or ½ of streamflow length T. Substituting equation (4-5) into equation (4-6), the probability density function is linked to the autocovariance as

$$R_r = \int_D \sum_{k=1}^{n} x_k c_{rk} p(\vec{x})d\vec{x}, \; -N \le r \le N \tag{4-7}$$

Thus, the autocovariance function from lag $-N$ to N with equation (4-7) with equation (4-4) are considered as $2N+2$ constraints for applying the entropy theory.

4.1.3 Determination of distribution of spectral power

The probability density function is computed by maximizing entropy using the method of Lagrange multipliers. Using constraints in equations (4-4) and (4-7), the Lagrangian function can be written as

$$L(p) = -\int p(\vec{x}) \ln[p(\vec{x})] d\vec{x} + (\lambda_0 - 1)[\int_D p(\vec{x}) d\vec{x} - 1] + \sum_{r=-N}^{N} \lambda_r [\int_D \sum_{k=1}^{n} x_k c_{rk} p(\vec{x}) d\vec{x} - R_r] \quad (4\text{-}8)$$

where λr are Lagrange multipliers. Taking the partial derivative with respect to $p(\vec{x})$, and equating the derivative to zero, one gets

$$\frac{\partial L(p)}{\partial p(\vec{x})} = \ln[p(\vec{x})] + \lambda_0 + \sum_{r=-N}^{N} \lambda_r \sum_{k=1}^{n} x_k c_{rk} = 0 \quad (4\text{-}9)$$

Thus, the least-biased distribution of spectral power becomes

$$p(\vec{x}) = \exp[-\lambda_0 - \sum_{r=-N}^{N} \lambda_r \sum_{k=1}^{n} x_k c_{rk}] \quad (4\text{-}10)$$

4.1.4 Determination of Lagrange multipliers

It is noted from the equation (4-10) that the spectral power estimated by maximizing the entropy involves Lagrange multipliers. To compute the Lagrange multipliers, equation (4-11) is substituted into constraints equations (4-4) and (4-7). One obtains

$$\int_D \exp[-\lambda_0 - \sum_{r=-N}^{N} \lambda_r \sum_{k=1}^{n} x_k c_{rk}] d\vec{x} = 1 \quad (4\text{-}11)$$

and

$$R_r = \int_D \exp[-\lambda_0 - \sum_{r=-N}^{N} \lambda_r \sum_{k=1}^{n} x_k c_{rk}] \sum_{k=1}^{n} x_k c_{rk} d\vec{x} \quad (4\text{-}12)$$

It is shown that $2N+2$ nonlinear equations need to be solved for computing the Lagrange multipliers.

4.1.5 Determination of spectral power

The expected spectral power can be computed from

$$S_k = \int x_k p(x_k) dx_k = \int x_k \exp(-\frac{\lambda_0}{n} - \sum_{r=-N}^{N} \lambda_r x_k c_{rk}) dx_k \tag{4-13}$$

Let $A = \exp[-\lambda_0]$, then equation (4-10) can be written as

$$p(\vec{x}) = A \exp[-\sum_{r=-N}^{N} \lambda_r \sum_{k=-n}^{n} x_k c_{rk}] = A \exp[-\sum_{k=-n}^{n} \left(\sum_{r=-N}^{N} \lambda_r c_{rk} \right) x_k] \tag{4-14}$$

Integrating equation (4-14) for x_k over 0 to infinite yields

$$\begin{aligned} 1 &= \int_0^{\infty} p(\vec{x}) d\vec{x} = \iiint A \exp[-\sum_{k=-n}^{n} \left(\sum_{r=-N}^{N} \lambda_r c_{rk} \right) x_k] dx_k \\ &= A \prod_{k=1}^{n} \int \exp(-\left(\sum_{r=-N}^{N} \lambda_r c_{rk} \right) x_k) dx_k = A \prod_{k=1}^{n} \frac{1}{\left(\sum_{r=-N}^{N} \lambda_r c_{rk} \right)} \end{aligned} \tag{4-15}$$

$$\text{Thus, } A = \prod_{k} \left(\sum_{r=-N}^{N} \lambda_r c_{rk} \right) \tag{4-16}$$

Thus, inputting equation (4-16) into equation (4-14) becomes multi-variate exponential distribution as

$$p(\vec{x}) = \prod_{k=-n}^{n} \left(\sum_{r=-N}^{N} \lambda_r c_{rk} \right) \exp[-\left(\sum_{r=-N}^{N} \lambda_r c_{rk} \right) x_k] \tag{4-17}$$

Thus, the spectral power can be solved for by inputting (4-17) into equation (4-13), which yields

$$S_k = \frac{1}{\sum_{r=-N}^{N} \lambda_r c_{rk}} \tag{4-18}$$

It is noted that equation (4-18) is identical to that derived from the BESA in equation (3-47).

4.1.6 Extension of autocovariance

It is seen from equation (4-18) that the derived spectral power also follows the form of linear prediction defined in equation (3-22) as does BESA. Thus, equation (4-18) can be further expanded as

$$S_k = \frac{1}{\sum_{r=-N}^{N} \lambda_r c_{rk}} = \frac{1}{\sum_{r=-N}^{N} \lambda_r z^{-r}} = \frac{\sigma^2}{\left| 1 + \sum_{n=0}^{N} a_n z^{-n} \right|^2} \tag{4-19}$$

where $z = \exp(2\pi i \Delta f_k)$, and a_n are forecasting coefficients, which satisfy

$$\lambda_r = \frac{1}{\sigma^2} \sum_{k=0}^{N-r} a_{k+r} a_k^* \tag{4-20}$$

Furthermore, the extension of autocovariance follows a linear extension as

$$R_{N+1} = a_1 R_N + a_2 R_{N-1} + \ldots + a_m R_{N+m-1} \tag{4-21}$$

4.1.7 Forecast

It is shown in Section 3.3.7 that using the coefficient for extending the autocorrelation in equation (4-22) to weigh the time series yields the least squared prediction (Burg, 1975; Krstanovic and Singh 1991b). Thus, for stationary normalized time series, forecasting follows the way of extending the autocorrelation or the autocovariance function as:

$$y_{T+1} = a_1 y_T + a_2 y_{T-1} + \ldots + a_m y_{T+m-1} \tag{4-22}$$

where m is the order of forecasting model, which, based on the number of previous lags, will be used for forecasting the future, and is identified by the Akaike information criterion

(AIC) or Bayesian information criterion (BIC) (Box and Jenkins, 1970; Hipel and McLeod, 1994).

4.2 Application

4.2.1 Data description

The proposed derivation of Burg entropy spectral analysis was verified using observed streamflows from the U.S. Geological Survey (USGS) website. Five stations were selected from Minnesota River, Upper Mississippi River, Iowa River, Des Moines River and Illinois River, which were listed in Table 1 along with their drainage areas and locations.

Table 4-1 Selected stations from Mississippi River watershed

Location	Station	Area (km^2)	Latitude	Longitude
Minnesota River	05301000	10489	45°01'17"	95°52'05"
Upper Mississippi River	05420500	221704	41˚46'50"	90˚15'07"
Iowa River	05449500	1111	42°45'36"	93°37'18"
Des Moines River	05476000	3237	43°37'06"	94°59'05"
Illinois River	05543500	21391	41°19'37"	88°43'03"

The selected five stations are distributed over the whole Mississippi River watershed, and the drainage area varies from 1,111 km^2 to 221,704 km^2. Thus, streamflow characteristics are quite different from station to station. Basic statistics of streamflow of Mississippi River are listed in Table 2 that shows that averaged monthly streamflow discharge varies from 9 m^3/s to over 1700 m^3/s and the standard deviation varies from 4.8 m^3/s to 570 m^3/s. The average peak streamflow is 1.4 to 2.9 times the mean streamflow.

Table 4-2 Streamflow characteristics

Location	Mean (m^3/s)	Standard deviation (m^3/s)	Peak (m^3/s)	Peak/mean
Minnesota River	67.0	15.8	105.4	1.6
Upper Mississippi River	1693.8	566.1	2877.5	1.7
Iowa River	9.3	4.8	26.8	2.9
Des Moines River	13.7	9.1	37.0	2.7
Illinois River	375.1	86.2	531.4	1.4

4.2.2 Parameter estimation

The coefficients of prediction by CESAS and BESA theories are computed for the Minnesota River, and are listed in Table 4-3. The prediction coefficients of BESA were estimated from the Levinson-Burg algorithm, while the prediction coefficients of CESAS were obtained by solving equation (4-20) numerically. As shown in the table, the difference was minimal for two methods, which is less than 0.001. However, the computing speed was faster using the Levinson-Burg algorithm for BESA. The recursive Levinson-algorithm is more efficient, as it involves the order of N^2 operations with memory storage on the order of N comparing to the order of N^3 operations by Newton's method for solving nonlinear equations.

Table 4-3 Prediction coefficients estimated using BESA and CESA

Prediction coefficient	BESA	CESAS	Difference
a_0	1.000	1.000	0.000
a_1	0.462	0.462	0.000
a_2	0.182	0.181	-0.001
a_3	-0.043	-0.043	0.000
a_4	0.153	0.153	0.000
a_5	-0.238	-0.237	0.001

Table 4-3 Continued

Prediction coefficient	BESA	CESAS	Difference
a_6	0.039	0.039	0.000
a_7	0.068	0.068	0.000
a_8	-0.078	-0.078	0.000
a_9	-0.033	-0.034	-0.001
a_{10}	-0.055	-0.054	0.001
a_{11}	0.036	0.036	0.000
a_{12}	0.686	0.686	0.000

4.2.3 Results and comparison

Streamflow was forecasted using equation (4-22) by CESAS. Figure 4-1 plots the forecasted streamflow in the Mississippi watershed with 90% confidence intervals. The forecasting lead year varied from 1 year to 3 years based on the characteristic of streamflow. For rivers like the Upper Mississippi River with sharp repeated peaks every 12 months, CESAS was capable to forecast with high r^2 over 0.9 for up to 3 year lead time. But for the Minnesota River, where peak streamflow is less significant, CESAS only forecasted for 1 year lead time with r^2 of 0.766.

Though forecasted streamflow did not fit the observed values exactly, most of the observed values fell inside the 90% confidence intervals. For example, the mean of forecasted streamflow in Minnesota River turned out as 63.1 m^3/s, which was 5.7% less than the observed values. The peak flow in April was 93.4 m^3/s, which was 11.3% lower than the observed value. Nevertheless, all the observations fell between the upper and lower 90% confidence intervals, as shown in Figure 4-1. However, there's an exception for Iowa River. Streamflow in Iowa had an unexpected peak streamflow of 26.8 m^3/s during the second lead year, which exceeded by 38.5% over the past peak streamflow. In

this case, the forecasted streamflow was 12.75 m^3/s and the upper 90% was 17.1 m^3/s, still smaller than the observed value. It implies that forecasted streamflows with CESAS were not able to capture the irregular changes in the time series and may miss some unexpected large peaks.

On the other hand, CESAS was not good at forecasting streamflow in low flow season. It can be seen from Des Moines River plotted in Figure 4-1 that streamflow in this river does not monotonically decrease after the peak, but there is another small peak during the low flow season. In this case, CESAS forecasted streamflow higher than observation, and the differences between the observed values and the forecasted values became larger as lead time increased. As a result, the observed values fitted the lower 90% confidence intervals for the third lead year as shown in Figure 4-2.

Forecasted errors are computed and plotted versus the lead time in Figure 4-2. It is seen from the figure that errors for the Upper Mississippi River had the most random pattern, which suggested that forecasting by CESAS for this river was consistently well during the lead time of 3 years. However, forecasted errors for Iowa River and Des Moines River increase over time, which suggests that CESAS would not be valid for longer lead time forecasts for these rivers.

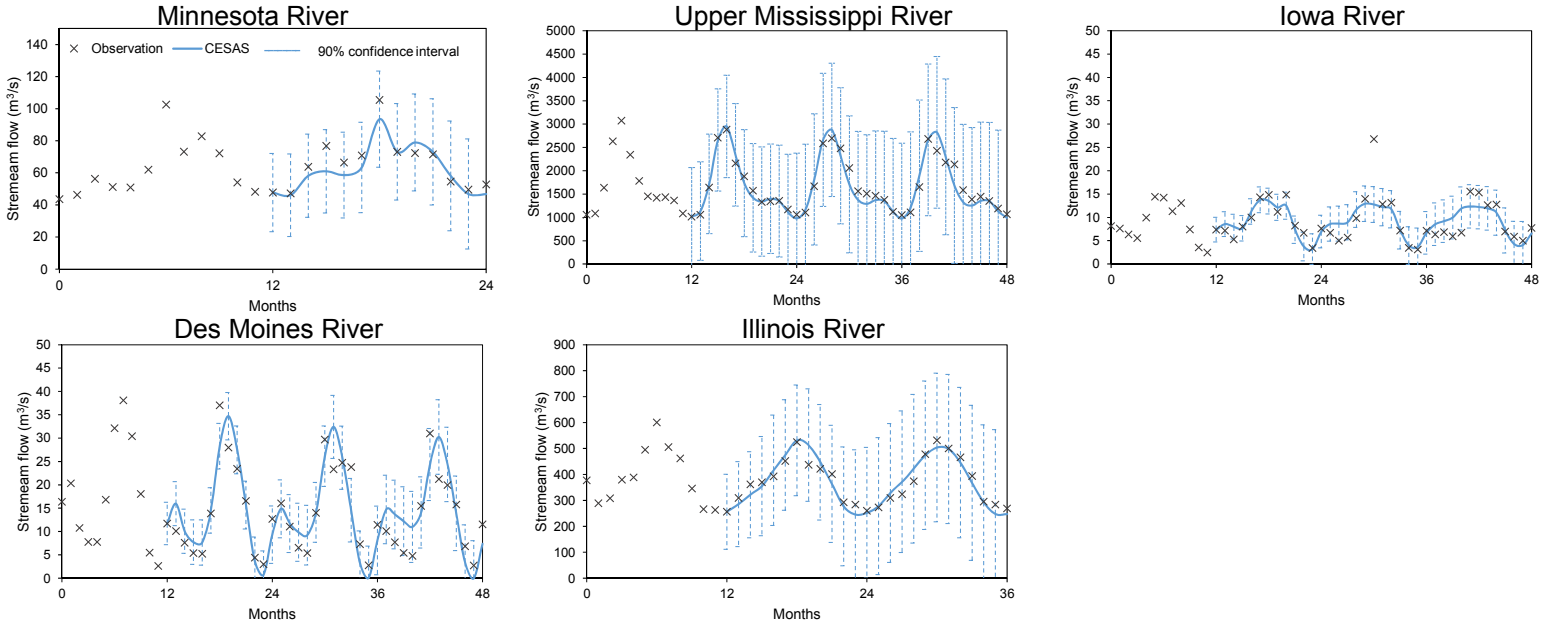


Figure 4-1 Forecasted streamflow with 90% confidence intervals

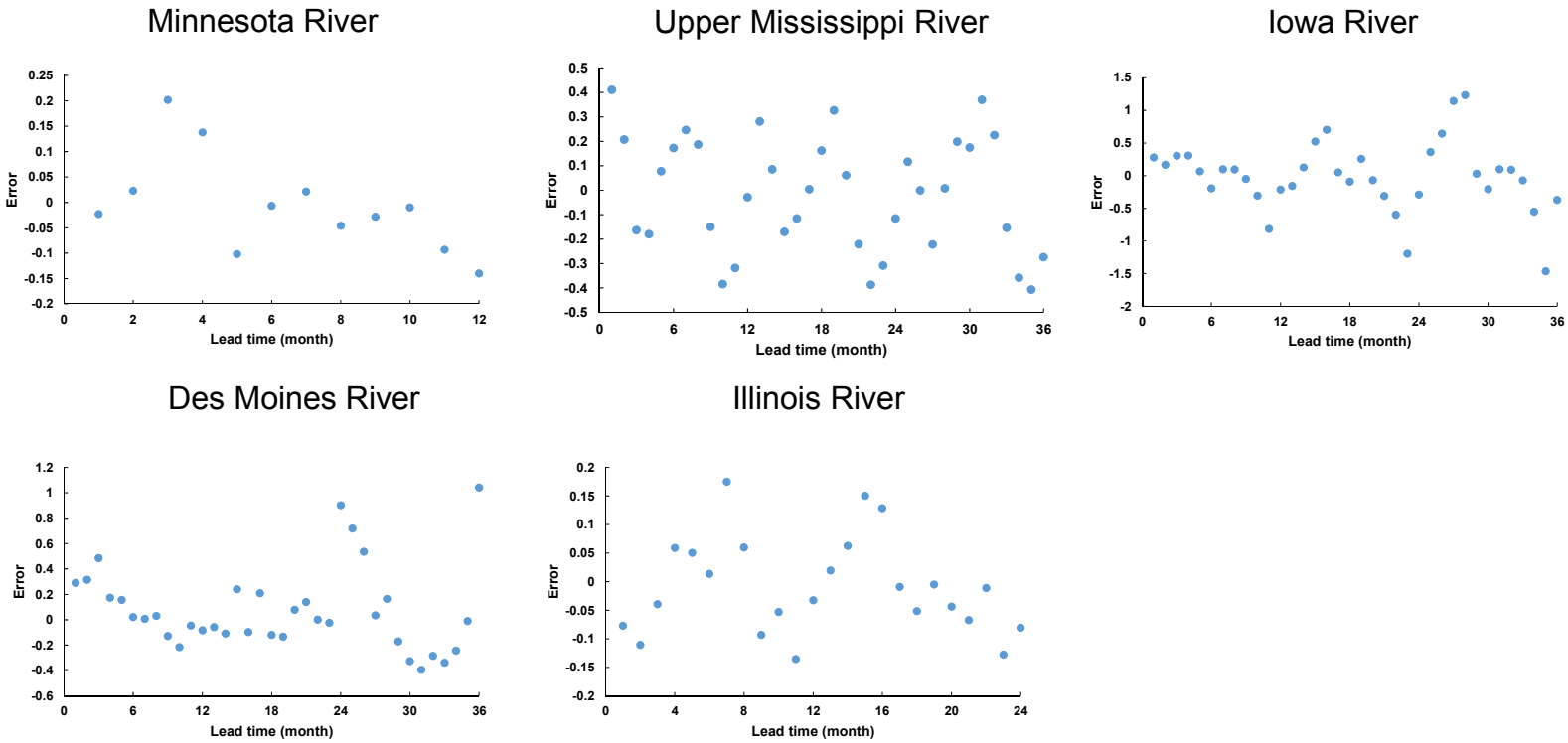


Figure 4-2 Forecasted errors using CESAS

The forecasted results for all five stations are summarized in Table 4-4. The results showed good forecasting by proposed CESAS. The RE values were around 0.1, which means that the forecasting error was around 10%. The r^2 values for all the rivers were above 0.7, which was even higher than 0.9 for Upper Mississippi River. Besides, the NSE values for all cases were higher than 0.4.

Table 4-4 Measures of forecasting results for five stations

Location	*RE*	*RMSE*	r^2	*NSE*
Minnesota River	0.083	7.320	0.766	0.484
Upper Mississippi River	0.072	163.578	0.914	0.737
Iowa River	0.129	3.091	0.872	0.675
Des Moines River	0.106	4.661	0.729	0.576
Illinois River	0.069	30.218	0.872	0.658

The prediction coefficients obtained by BESA and CESAS methods were about the same, with differences less than 0.001. As a result, streamflows forecasted by the BESA and CESAS theories coincided, as shown in Figure 4-3. No difference was found in forecasting streamflow for an accuracy of 0.001 m^3/s for rivers from the Mississippi watershed and BESA and CESAS methods shared the same confidence intervals as shown in Figure 4-3. The drawbacks of CESAS in forecasting irregular flow of dry season in Des Moines River and non-normal streamflow peak in Iowa River still occurred in the BESA forecasting. It suggested that the derived CESAS theory was completely identical to the BESA theory.

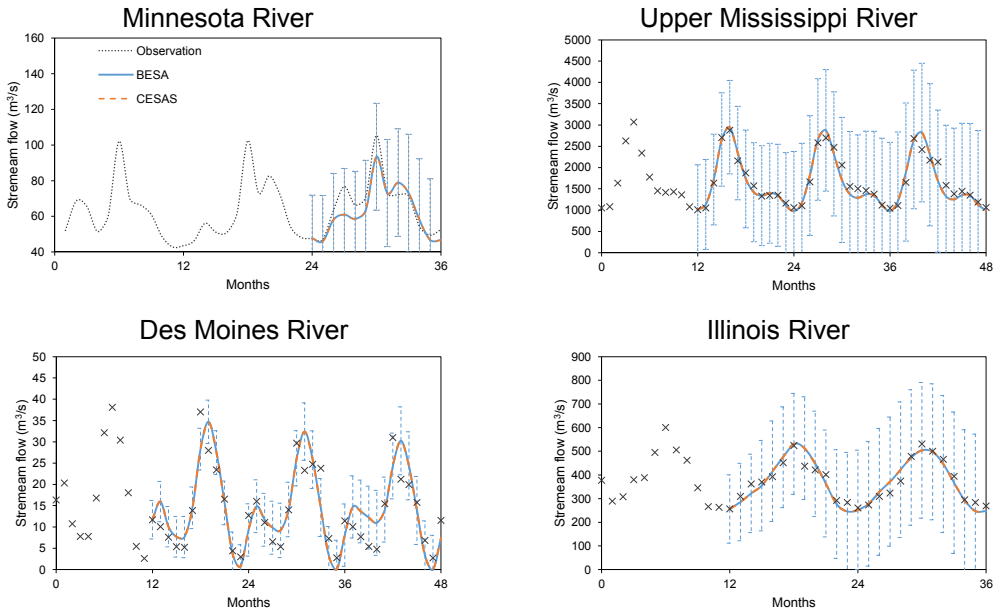


Figure 4-3 Streamflow forecasted using the CESAS and BESA methods

4.3 Summary

The configurational entropy spectral analysis was developed with spectral power as a random variable, which is identical to the Burg entropy spectral analysis that developed by Burg (1967; 1975). The derived CESAS was examined using streamflow data obtained from the Mississippi Watershed, which yielded the same results with the one estimated from BESA. Besides, CESAS had the drawbacks as with BESA in forecasting streamflow in low flow season and unexpected peak flows. However, the prediction coefficient was estimated from solving nonlinear equations, and the computing speed was slower than Levinson-Burg algorithm. To sum up, the CESAS and BESA methods can be seen as one method, and the Levinson-Burg algorithm for computing the parameters is recommended.

5 CONFIGURATIONAL ENTROPY SPECTRAL ANALYSIS WITH FREQUENCY AS A RANDOM VARIABLE*

In this section, configurational entropy spectral analysis (CESAF) is developed with frequency as a random variable. Development of CESAF does not need the spectral power to be Gaussian as Burg's (1967; 1975) BESA does. Liefhbber and Boekee (1987) showed that CESAF was not restricted to the AR process in contrast to BESA, which suggested that CESAF could be used for a wider range of streamflow conditions. It was found that the configurational entropy was shown to have better resolution than BESA for computing spectral density of autoregressive moving average (ARMA) and moving average (MA) processes, and comparable to BESA for that of the autoregressive (AR) process (Nadeu et al., 1981; Ortigueira et al., 1981). On the contrary, BESA appeared to be better for white noise, as suggested by experiments on speech synthesis (Johnson and Shore, 1983; Katsakos-Mavromichalis et al., 1985). However, due to the complexity in computing CESAF, it was not widely used until explicit solution was developed by Wu (1983), which requires the underlying process to be real, causal, and minimum phase. Since CESAS turned out to be as the same method as BESA in the previous section, CESAS will be referred as BESA, and CESAF will be referred as CESA henceforth. In this section, the CESA theory is developed for monthly streamflow forecasting, and compared to the BESA for forecasting.

* This section is reprinted with permission from "Configurational entropy theory for streamflow forecasting" by Cui, H. and Singh, V.P., 2015, *Journal of Hydrology*, 521(0):1-17, copyright 2014 by Elsevier B.V.

5.1 Development of configurational entropy spectral analysis

The development of configurational entropy spectral analysis (CESA) theory with frequency as a random variable and its application to streamflow forecasting involve following steps: (1) Define configurational entropy, (2) construct constraints, (3) derive maximum entropy-based spectral density, (4) compute the Lagrange multipliers, (5) extend the autocorrelation function, and (6) forecast streamflow time series. Cepstrum analysis is used to compute the Lagrange multipliers and relate them to forecasting coefficients.

5.1.1 Definition of configurational entropy

Let frequency f be considered a random variable, and the normalized spectral density $p(f)$ be considered its probability density function (pdf). Then, configurational entropy, $H(f)$, can be defined as (Frieden, 1972; Gull and Daniell, 1978):

$$H(f) = -\int_{-W}^{W} p(f) \ln[p(f)] df \tag{5-1}$$

where W is the Nyquist frequency. Equation (5-1) defines the entropy of the spectral density, which accounts for the uncertainty associated with streamflow in the frequency domain. In order to determine the least-biased $p(f)$, configurational entropy is maximized subject to specified constraints.

5.1.2 Construction of constraints

Recalling the relation between spectral density and autocorrelation function introduced in Section 3, constraints can be defined in terms of the autocorrelation function as

$$\rho_n = \int_{-W}^{W} p(f) e^{i2\pi fn\Delta t} df \qquad , -N \le n \le N \tag{5-2}$$

where Δt is the discretization or sampling interval, and $i = \sqrt{-1}$. If n=0, equation (5-2) reduces to

$$\int_{-W}^{W} p(f)\, df = 1 \tag{5-3}$$

It can be seen from equation (5-3) that $p(f)$ can be considered as a probability density function of frequency f.

5.1.3 Entropy maximizing

To obtain the least-biased $p(f)$, entropy is maximized using the method of Lagrange multipliers in which the Lagrangian function can be formulated as

$$L(f) = -\int_{-W}^{W} p(f) \ln[p(f)] df - \sum_{n=-N}^{N} \lambda_n [\int_{-W}^{W} p(f) \exp(i2\pi fn\Delta t) df - \rho_n] \tag{5-4}$$

where λ_n, n=0, 1, 2, …, N, are the Lagrange multipliers. Taking the partial derivative of equation (5-4) with respect to $p(f)$ and equating the derivative to zero, one obtains:

$$\frac{\partial L(f)}{\partial p(f)} = 0 = -\int_{-W}^{W} \{\ln[p(f)] + 1 + \sum_{n=-N}^{N} \lambda_n \exp(i2\pi fn\Delta t)\} df \tag{5-5}$$

Equation (5-5) yields the least-biased $p(f)$:

$$p(f) = \exp(-1 - \sum_{n=-N}^{N} \lambda_n e^{i2\pi fn\Delta t}) \tag{5-6}$$

Equation (5-6) expresses the spectral density with Lagrange multipliers λ_n obtained by maximizing the configurational entropy.

5.1.4 Computation of Lagrange multipliers

Unlike the spectral density of linear forecasting, either equation (3-21) of the AR process or equations (3-44) and (4-19) for the BESA theory, which were in the polynomial form, equation (5-6) is in the exponential form which is easier to manipulate. For computing the Lagrange multipliers efficiently, cepstrum analysis, which has not been used in hydrology, is used. Taking the logarithmic transform of equation (5-6) for computing the Lagrange multipliers, one obtains

$$1+\log[p(f)]=-\sum_{n=-N}^{N}\lambda_n e^{i2\pi fn\Delta t} \tag{5-7}$$

Taking the inverse Fourier transform of equation (5-7), it becomes

$$\int_{-W}^{W}\{1+\log[\ p(f)]\}\,e^{i2\pi fn\Delta t}df\ =\ \int_{-W}^{W}(-\sum_{n=-N}^{N}\lambda_n e^{i2\pi fn\Delta t})e^{i2\pi fn\Delta t}df \tag{5-8}$$

It can be seen that the second part of the left side is the autocepstrum defined as equation (3-25), and the first part of the left side yields the delta function as

$$\int_{-W}^{W}e^{i2\pi fn\Delta t}df\ =\ \int_{-W}^{W}\cos(2\pi fn\Delta t)df\ =\delta_n\ =\begin{cases}1, & n=0\\ 0, & n\neq 0\end{cases} \tag{5-9}$$

while the right side of equation (5-8) becomes

$$\int_{-W}^{W}(-\sum_{n=-N}^{N}\lambda_n e^{i2\pi fn\Delta t})e^{i2\pi fn\Delta t}df\ =-\sum_{s=-N}^{N}\lambda_s\int_{-W}^{W}e^{i2\pi f(n-s)\Delta t}df\ =\ \sum_{s=-N}^{N}\lambda_s\delta_{n-s} \tag{5-10}$$

Thus, equation (5-8) can be recast with equations (5-9) and (5-10) as

$$\delta_n+e(n)=-\sum_{s=-N}^{N}\lambda_s\delta_{n-s} \tag{5-11}$$

where $e(n)$ is the cepstrum of the autocorrelation and σ_n is the delta function.

Equation (5-11) then can be expanded as a set of N linear equations:

$$\begin{aligned} \lambda_0 &= -1 - e(0) \\ \lambda_1 &= -e(1) \\ &\vdots \\ \lambda_N &= -e(N) \end{aligned} \tag{5-12}$$

Equation (5-12) shows that the Lagrange multipliers can be determined from the values of cepstrum which entails the spectral density. However, as discussed in section 3.2.2 for a finite length of data the autocepstrum can be estimated by the following recursive relation, stated in equation (3-36), as

$$e(n) = 2[\rho(n) - \sum_{k=1}^{n-1} \frac{k}{n} e(k)\rho(n-k)], \quad n > 0 \tag{5-13}$$

In order to compute $e(k)$, one needs autocorrelation from lag 0 to k. Thus, for given N lag autocorrelations, the cepstrum of autocorrelation can be computed up to lag N, and beyond this lag, the cepstrum is defined as 0. Then, the calculated cepstrum from lag 0 to T can be used to compute the Lagrange multipliers using equation (5-9).

5.1.5 Extension of autocorrelation function

For streamflow forecasting the autocorrelation needs to be extended beyond the N lag. It may be noted from equation (5-13) that the first n lags of cepstrum can be completely determined from the first n lags of autocorrelation and vice versa. Thus, it is possible to write the inverse relation between the autocorrelation and the cepstrum as shown in equation (5-13), which is

$$\rho(n) = \frac{e(n)}{2} + \sum_{k=1}^{n-1} \frac{k}{n} e(k)\rho(n-k), \quad n > N \tag{5-14}$$

Therefore, with model order m determined, the autocorrelation function for $n>N$ can be estimated as

$$\rho_n = \sum_{k=1}^{m-1} \frac{k}{m} e(k)\rho(n-k) \tag{5-15}$$

Equation (5-15) extends the autocorrelation function with the configurational entropy maximized. Surprisingly, the autocorrelation extends with a linear combination of past lags, which is the same with the BESA or the AR method. Thus, equation (5-15) can be also written as

$$\rho_n = \sum_{k=1}^{m} \rho_{n-k} a_k , \quad n > N \tag{5-16}$$

with extension coefficients $a_k = \frac{k}{m} e(k)$, and m is the model order. A proper model order needs to be determined so that the autocorrelation function is extended to as close as the observed function at the lowest possible order.

5.1.6 Forecasting

Equation (5-16) suggests that through entropy maximizing, the extension of autocorrelation is as a linear combination. Thus, streamflow again can be forecasted by weighting time series with the extension coefficients as

$$y_t = \sum_{k=1}^{m} y_{t-k} a_k , \quad t > T \tag{5-17}$$

Equation (5-17) represents the forecast using the entropy-based extended autocorrelation, which satisfies least squared prediction as discussed in Section 3.3.7.

The computation procedure is shown in Figure 5-1 for forecasting y_{T+k} , k>0, from given series y_1, y_2,…, y_T. The computation steps are as follows: (1) For eliminating skewness and normalizing streamflow data, the Box-Cox transformation is recommended before applying the configurational entropy theory (Box and Cox, 1964; Jain and Singh, 1986). (2) By entropy maximizing, the maximum entropy-based spectral density $p(f)$ is obtained. (3) The cepstrum values e_1, e_2, …, e_N are determined for computing the Lagrange multipliers. (4) The autocorrelation function is extended after lag N from ρ_{N+1} onwards. (5) The extension coefficients a_n are determined and streamflow forecasting is done. (6) The forecast order m is identified by the Akaike information criterion (AIC) or Bayesian information criterion (BIC) (Box and Jenkins, 1970; Hipel and McLeod, 1994).

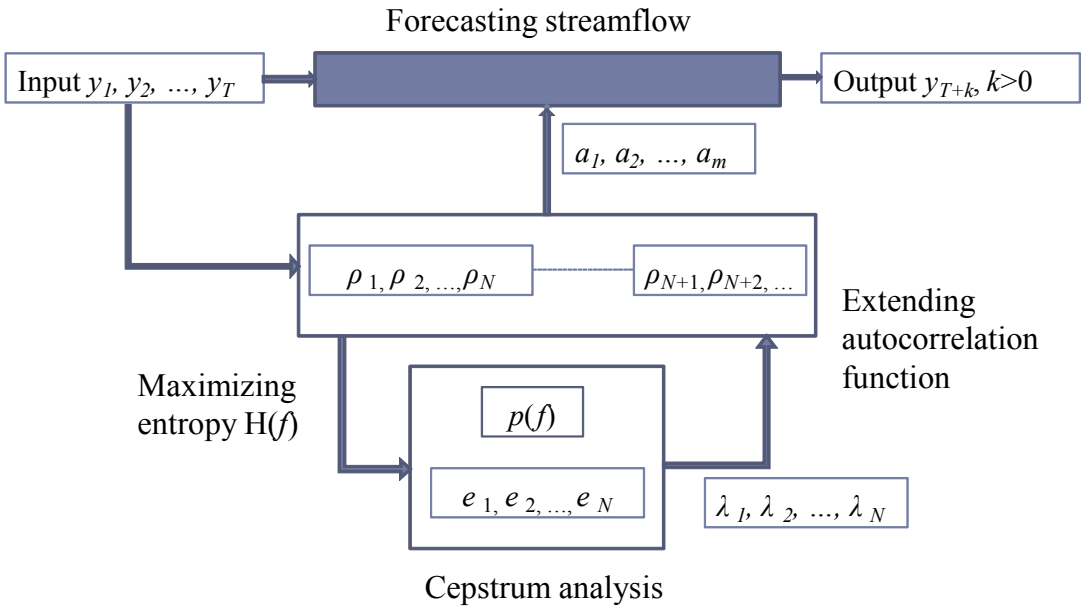


Figure 5-1 Computation procedure

5.2 Application

5.2.1 Observed data and characteristics

For evaluating the configurational entropy spectral analysis theory, monthly streamflow data from 19 river sites in the Mississippi River watershed, Tennessee River watershed, Missouri River watershed, Colorado River watershed and Texas-Gulf watershed, were obtained from the U.S. Geological Survey (USGS). Table 5-1 summarizes the basic information about the data used.

Greenbrier River is a tributary of New River in West Virginia, and belongs to the Mississippi River watershed. More than 100 years of streamflow data are available on Greenbrier River, where the average monthly streamflow during 1896-2012 is 57 m^3/s. Pigeon River and Piney River are tributaries of Tennessee River, with average monthly streamflow of 34 $m^3/$ and 9 m^3/s, respectively. Four sites in the Mississippi River watershed were chosen: the center of the Upper Mississippi River, Cedar River, Big Black River, and Ouachita River. The Upper Mississippi River has a drainage area of 221,704 km^2, and average monthly streamflow of 1,395 m^3/s; the Cedar flows into Upper Mississippi River and the others into Lower Mississippi River. Yellowstone River is a tributary of upper Missouri River, with an average monthly streamflow of 293 m^3/s for 1978-2013. Another station was chosen at lower Missouri River before it joins Mississippi River. It has a large drainage area of 814,814 km^2 with an average monthly streamflow of 839 m^3/s.

Table 5-1 Basic information of streamflow data used

No.	River	Station	Area (km^2)	Latitude	Longitude	Record length	Average (m^3/s)	Peak ratio*	Percentage**	Peak month
1	Greenbrier	03185000	136	37˚45'38"	81˚09'45"	1896-2012	57	2.77	45.4%	3
2	Pigeon	03461500	1725	35˚57'38"	83˚10'28"	1904-2012	34	2.17	31.9%	3
3	Piney	03602500	500	35˚52'16"	87˚30'05"	1926-2012	9	3.05	43.5%	3
4	Mississippi	05420500	221704	41˚46'50"	90˚15'07"	1875-2013	1395	2.17	38.0%	4
5	Cedar	05457700	2730	43˚03'44"	92˚40'25"	1964-2013	23	2.97	43.0%	4
6	Big Black	07290000	7283	32˚20'52"	90˚41'49"	1937-2013	109	3.41	49.8%	3
7	Ouachita	07359002	4015	34˚25'34"	92˚53'27"	1929-2013	70	2.72	34.3%	1
8	Yellowstone	06295000	103978	46˚16'00"	106˚41'29"	1978-2013	293	2.90	52.5%	6
9	Missouri	06486000	814814	42˚29'09"	96˚24'49"	1953-2013	839	1.50	31.5%	7
10	Upper Colorado	09095500	20684	39˚14'21"	108˚15'56"	1933-2012	8404	3.70	66.2%	6
11	Yampa	09251000	8762	40˚30'10"	108˚01'58"	1905-2012	2892	4.18	73.8%	5
12	Dolores	09180000	11862	38˚47'50"	109˚11'40"	1905-2013	1924	3.66	68.5%	5
13	Green	09315000	116162	38˚59'10"	110˚09'02"	1895-2012	4674	3.30	60.4%	6
14	San Juan	09355500	8443	36˚48'06"	107˚41'55"	1905-2012	2709	3.32	63.0%	5
15	Little Colorado	09402000	68529	35˚55'35"	111˚34'00"	1905-2012	421	4.89	43.0%	4
16	Lower Colorado	09427520	473193	34˚17'44"	114˚08'22"	1934-2012	38250	3.11	58.9%	6
17	Trinity	08062500	21101	32˚25'35"	96˚27'46"	1924-2013	89	3.49	41.2%	2
18	Brazos	08098290	78829	31˚08'02"	96˚49'29"	1957-2012	78	3.57	39.0%	5
19	Colorado (TX)	08136700	62660	31˚29'37"	99˚34'25"	1968-2012	4	4.27	36.4%	6

Peak ratio*=Peak/Average

Percentage**=streamflow during peak three months

Seven stations were chosen on Colorado River: Upper Colorado River is located at the very upstream near the source where the average monthly streamflow is 8,404 m^3/s. Green River is located near the junction with Colorado River with a drainage area of 116,162 km^2. Yampa, Dolores, and San Juan rivers are also tributaries of upper Colorado River with drainage areas of over 8000 km^2. Little Colorado River is the main tributary of lower Colorado River and possesses low streamflow volume. The Lower Colorado River is located near the mouth, before entering the Gulf of California with a drainage area of 473,193km^2. Three gaging stations were selected from the Texas-Gulf watershed: Trinity River, Brazos River, and Colorado (TX) River. These three rivers experience different drought conditions, with an average monthly streamflow of 89 m^3/s, 78 m^3/s and 4 m^3/s, respectively.

Yearly average monthly streamflow box-plots are presented in Figure 5-2. The central box represents the medium 50% of data, where upper and lower boundary lines are 75% and 25% percentile of data. The central line is the median of data and the solid line across the year is the mean of data. For the first nine rivers, except for Upper Mississippi River and Yellowstone River, (numbered 1, 2, 3, 5, 6, 7, 9), streamflow is distributed almost uniformly throughout the year, where streamflow occupies 30% to 40% of the yearly volume during the peak three months. The peak three months is referred to the three consecutive months with maximum streamflow. On the other hand, in rivers in the Colorado River basin (numbered 10, 11, 12, 13, 14, 16) streamflow is highly concentrated during the peak season (May-June), and after June or July, it drops down dramatically. For low flow rivers (numbered 17, 18, 19), the seasonal pattern is weak. Based on

consistency, seasonality, and peak characteristics the 19 stations (or rivers) can roughly be grouped into six categories as shown in Figure 5-3. The first category contains Greenbrier River (No. 1) with strong seasonality and 12-month periodicity, where the streamflow pattern is regular, and peak discharge and peak time are almost the same every year. The rivers (numbered 2, 3, 4, 5, 6, 7, 9) in the second category also exhibited strong seasonality and periodicity, but the peak time varied from year to year. The difference between the first two categories is the duration of peak season, which also can be seen from monthly hydrograph. Greenbrier River has a significantly large peak in March every year, reduces quickly from April. However, the duration of peak time in the second category may extend to 4 months and the peak may happen in any month within that period. For example, the annual peak of Missouri River in Figure 5-3 occurs in June, September, September, August, October, September, April in the row.

It is noted that the upper Mississippi River (numbered 4) was found to exhibit a bi-modal probability distribution as evidenced by the observed changes in the timing and amplitude of flow peaks (Baldwin and Lall, 1999). Rivers in the third (numbered 8, 10, 13, 14) and the fourth (numbered 11, 12, 16) categories have strong seasonality but the 12-month periodicity is weaker compared to the first two categories. For rivers in the third category, the variation of peak streamflow is less than 20%, whereas for those in the fourth category, the variation of peak streamflow is larger. Thus, streamflow in the fourth category may involve periodicity of other frequencies. Rivers 15 and 19 with weak seasonality and periodicity belong to the fifth category, and rivers 17 and 18 with an irregular pattern belong to the sixth category.

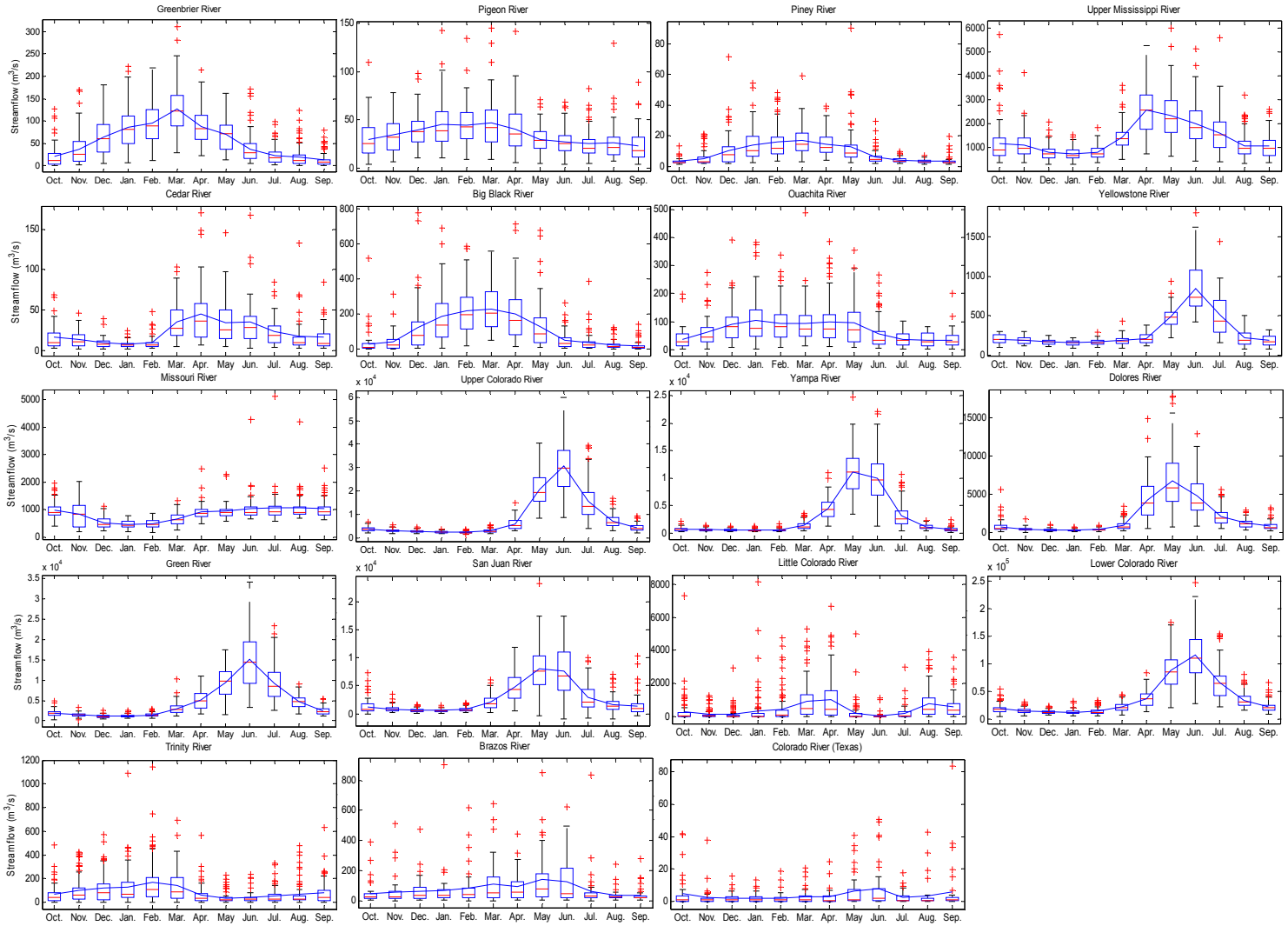


Figure 5-2 Monthly average streamflow for selected sites

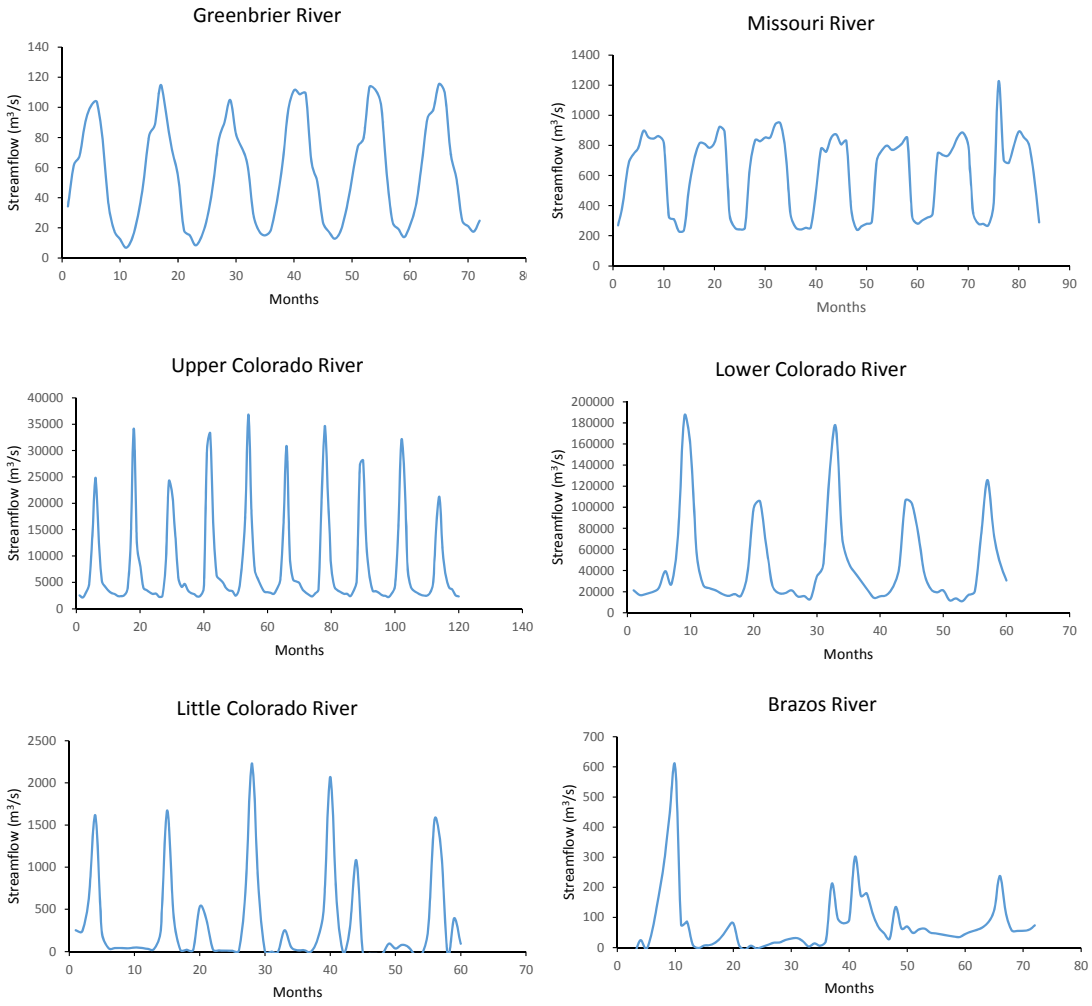


Figure 5-3 Streamflow time series of representative sites

Other than periodicity, autocorrelation is intuitively seen from autocorrelation (ACF) and partial autocorrelation (PACF) plots. The ACF plots of the first three categories are similar, thus, ACF and PACF are plotted in Figure 5-4 for Upper Colorado River as a representative. These are also plotted in Figure 5-4 for Lower Colorado River representing the fourth category. Generally, the correlation is not weak for the first four categories and a 12-month periodicity can be found in the ACF plot. The plot shows more significant values outside the confidence intervals of Upper Colorado River, which implies a stronger

autocorrelation in streamflow than for Lower Colorado River. For the last two category, the autocorrelation is not significant. Little Colorado River, representing the fifth category, has only significant autocorrelation at lags 1 and 12. Brazos River of the last category shows the weakest correlation, and it monotonically decreases to around zero with only first two lags outside the confidence bounds. As streamflow forecasting is based on the autocorrelation with the past series, the streamflow series with strong autocorrelation will be forecasted more reliably. Thus, one may expect that streamflow of the first four categories may be better forecasted, and one may be interested in how the configurational entropy theory performs for low streamflow forecasting with weak autocorrelation or seasonality.

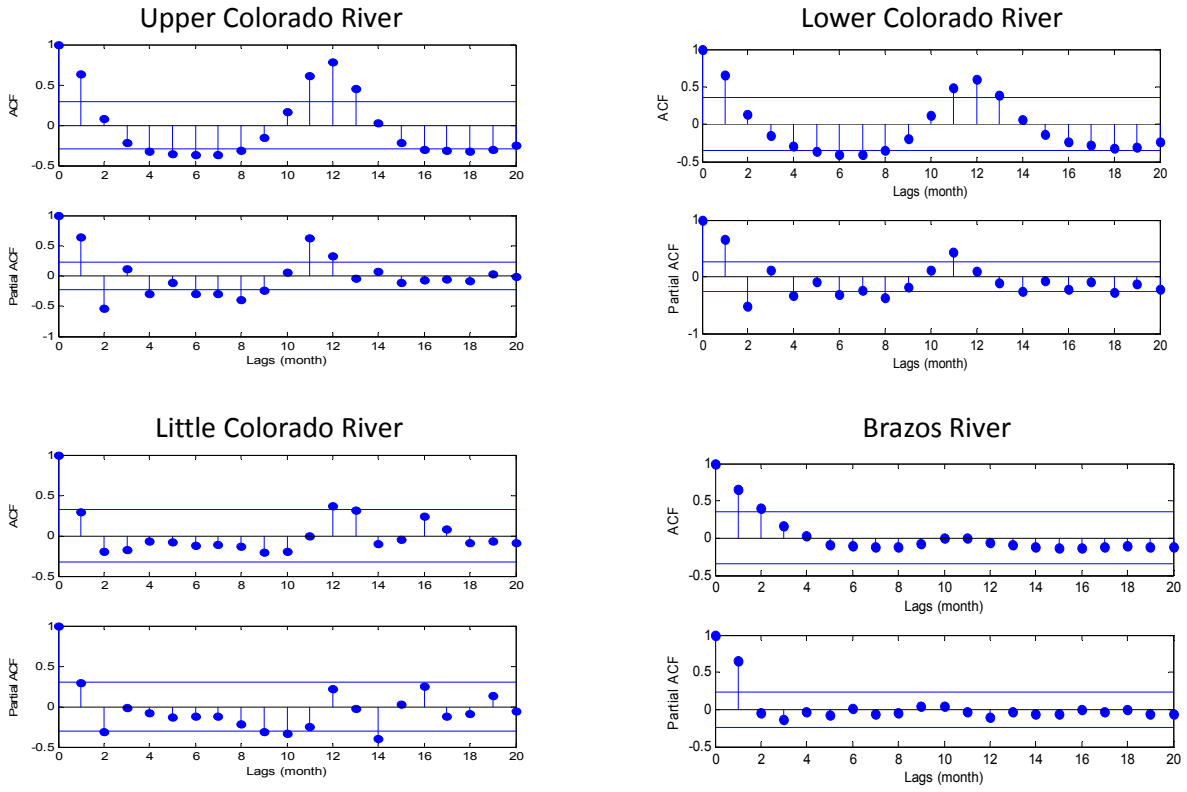


Figure 5-4 Autocorrelation and partial autocorrelation plot of representative sites

5.2.2 Spectral analysis

The spectral density is another method to check the periodicity by looking for the frequency of spectral peaks. Thus, the method, which can find the location of spectral peaks correctly, should be preferred for streamflow forecasting. Figure 5-5 shows spectral densities computed for streamflow using the configurational entropy spectral analysis (CESA) in comparison with the Burg entropy spectral analysis (BESA) and by fast Fourier transform (FFT). Six representative rivers were selected to show the ability of CE to estimate the spectral density; however, these six-categories are not the same as the previous six groups. The first two groups represent the uni-peak spectral density cases. It can be seen from the spectral density obtained by FFT of Upper Colorado River that there is a unique peak at frequency $1/12^{th}$. For this case, both CE and BE detected the same as by the FFT. On the other hand, though the spectral peak at frequency $1/12^{th}$ was significant, there was another suspected peak near frequency $1/4^{th}$ in the spectral density of Greenbrier River. Rivers 3 and 7 also belonged to this group. For this group, CESA still yielded the peak at the right frequency, but the one by BESA had shifted toward frequency $1/10^{th}$.

The third and fourth groups showed a multi-peak spectral case, but the major peak was still at frequency $1/12^{th}$. Rivers 2, 4, 6, 7, 10 belonged to the third group, where the second peak was not significant, and rivers 5, 8, 9, 14, and 15 belonged to the fourth group with more than three peaks. It implied that streamflow of the third group contained 12-month and 6-month periodicities, while streamflow of the fourth group exhibited 12-month, 6-month, 4-month and 3-month periodicities. However, the most significant period was still 12-month. Burg (1967) cautioned about the use of BESA for multi-peak spectral

cases. Thus, it would be interesting to determine how CESA would perform in this case. It is seen from the Mississippi River and Missouri River plots that CESA treated the spectral density as a uni-peak case and found the peak frequency at $1/12^{th}$, but for BESA the peak shifted dramatically. How it influenced streamflow forecasting would be discussed in the next section.

The last two groups showed significantly different spectral patterns. The spectral plot of Little Colorado River contained three peaks at frequencies $1/12^{th}$ and $1/6^{th}$ and $1/4^{th}$; however, the most significant peak was no longer at frequency $1/12^{th}$. Two peaks at frequencies $1/12^{th}$ and $1/6^{th}$ were detected by CESA, but with a wrong sequence. However, BESA yielded the spectral peak at frequency 0.15, which was shifted by 0.016. Colorado River (Texas) had a similar result. Referring to the ACF plot of the Brazos River streamflow, no significant periodicity was found, which implied that no significant large peak was expected in the spectral plot. As a result, the spectral density obtained by the FFT did not have one significantly large peak but consisted of generally decreasing multi-peaks-a pattern generally yielded by CESA, BESA led to uni-modal spectral density, with a wrong location for the spectral peak. To conclude, except for the fifth group, CESA performed better than BE. We now examine if this would lead to better streamflow forecasting.

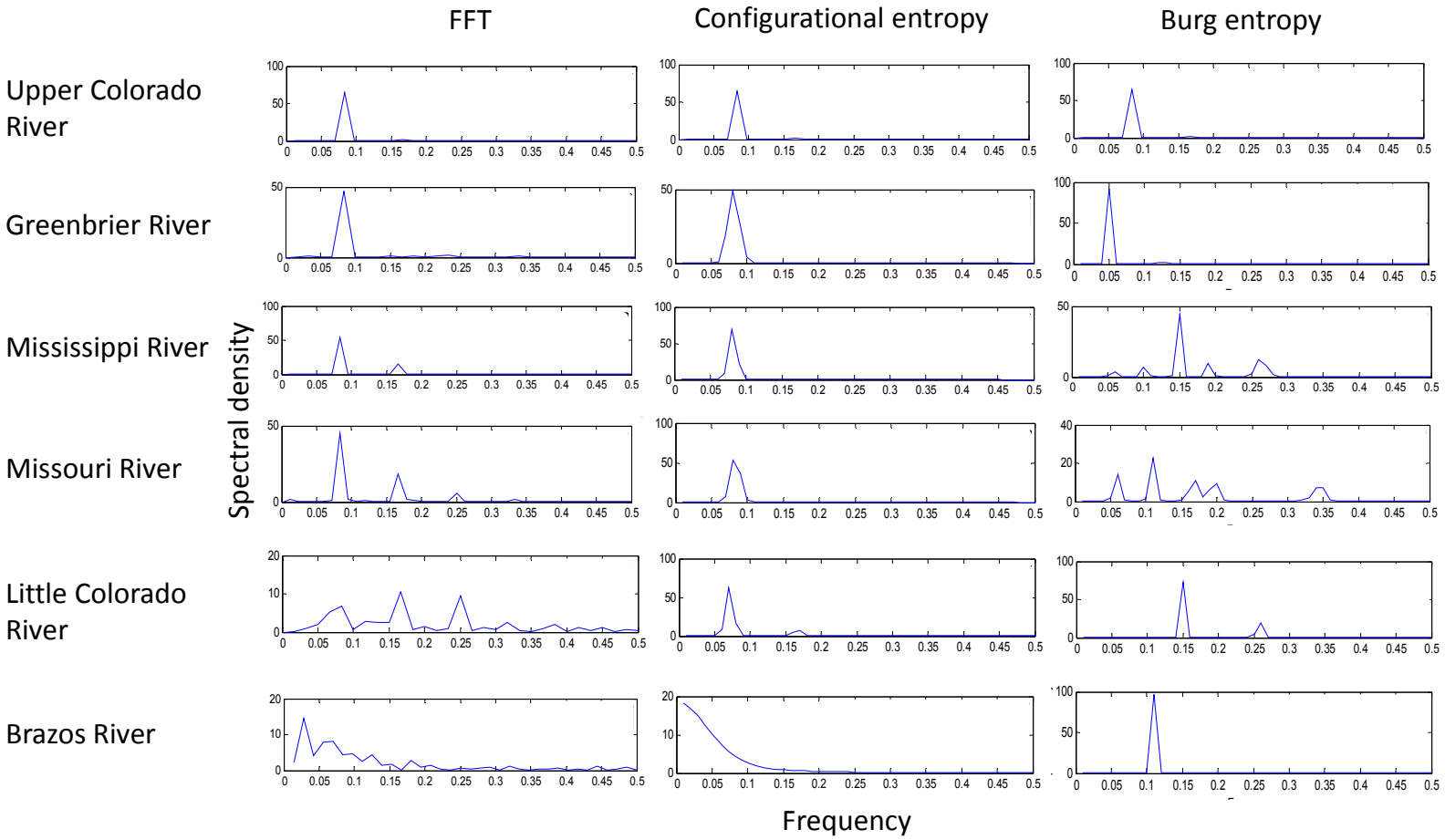


Figure 5-5 Spectral density of representative sites

5.2.3 Evaluation

The configurational entropy theory was applied to forecast streamflow as shown in Figure 5-6, with 90% confidence intervals, where the dotted line represents observed values and the solid line the forecasted values. The forecast statistics are given in Table 5-2 and the goodness of fit criteria in Table 5-3.

First, consider Greenbrier River from the first group. Since streamflow of this river exhibits strong seasonality, periodicity, consistency, and high autocorrelation, streamflow should be forecasted with high accuracy. CESA used a 12 order model to successfully forecast with a 48-month lead time, with an estimated average of 57 m^3/s compared to the observe average of 58 m^3/s. The forecasted hydrograph fitted observations well with r^2 as high as 0.929. The peak error of the first year (2009) was less than 1%, and from the second year afterward, the peaks forecasted subsequently were about 5%-10% less than observed peaks.

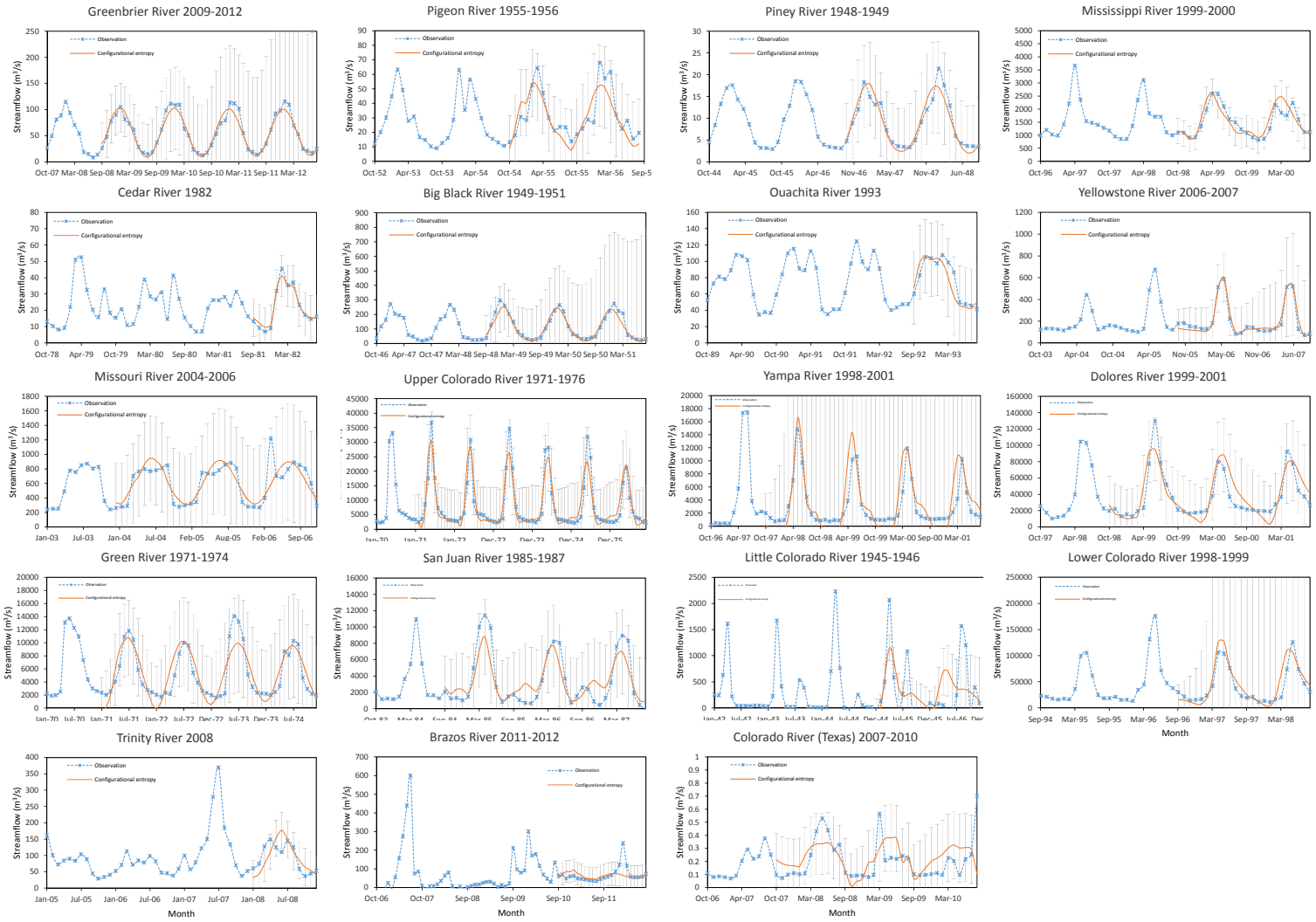


Figure 5-6 Forecasted streamflow with 90% confidence intervals

Table 5-2 Forecasting statistics for the entropy method

No.	River	Station	Model order	Leading time	Average		Standard deviation		Peak	
					Obs.	Est.	Obs.	Est.	Obs.	Est.
1	Greenbrier	03185000	12	48	58	57	35	18	116	97
2	Pigeon	03461500	13	24	1132	1047	593	541	2399	1906
3	Piney	03602500	13	24	326	319	197	199	758	632
4	Mississippi	05420500	13	36	52693	54109	19453	19714	91768	93330
5	Cedar	05457700	12	12	764	777	451	385	1604	1451
6	Big black	07290000	11	36	4235	3958	3179	2845	10497	8928
7	Ouachita	07359002	13	12	2732	2644	928	925	3806	3757
8	Yellowstone	06295000	10	24	7164	7268	5504	5652	20771	21266
9	Missouri	06486000	11	36	21458	21688	9250	7642	43241	33520
10	Upper Colorado	09095500	12	60	8349	8258	8899	8107	36822	30344
11	Dolores	09180000	12	36	37100	39347	27511	28143	129985	95156
12	Yampa	09251000	10	48	3207	3435	3503	4770	14836	16598
13	Green River	09315000	13	48	5194	5375	3711	3749	14091	11616
14	San Juan	09355500	12	36	3455	3390	3357	2138	11397	8782
15	Little Colorado	09402000	12	24	340	339	571	300	2069	1147
16	Lower Colorado	09427520	11	24	40701	45843	33889	40005	125822	126604
17	Trinity	08062500	11	12	3267	3232	1470	1719	5262	6253
18	Brazos	08098290	3	12	69	68	42	13	238	88
19	Colorado (Texas)	08136700	6	24	0.21	0.23	0.15	0.10	0.70	0.38

Table 5-3 Goodness of fit measures

No.	River	Configurational entropy spectral analysis				
		RE	*RMSE* (m^3/s)	r^2	*NSE*	*PE*
1	Greenbrier	0.137	9.3	0.929	0.773	-0.101
2	Pigeon	0.247	323.1	0.690	0.456	-0.206
3	Piney	0.177	65.3	0.886	0.683	-0.167
4	Mississippi	0.138	9725.1	0.739	0.540	0.017
5	Cedar	0.126	75.7	0.969	0.838	-0.095
6	Big black	0.205	924.8	0.913	0.755	-0.150
7	Ouachita	0.122	427.1	0.769	0.616	-0.013
8	Yellowoston	0.178	1233.7	0.948	0.760	0.024
9	Missouri	0.194	5356.1	0.655	0.539	-0.225
10	Upper Colorado	0.354	3319.8	0.859	0.749	-0.176
11	Dolores	0.347	13749.1	0.743	0.672	-0.268
12	Yampa	1.068	2329.8	0.548	0.310	0.119
13	Green River	0.343	1695.2	0.787	0.599	-0.176
14	San Juan	0.495	1633.2	0.757	0.545	-0.229
15	Little Colorado	-3.882	488.1	0.238	0.139	-0.446
16	Lower Colorado	0.400	16341.6	0.757	0.547	0.006
17	Trinity	0.346	1168.1	0.311	0.258	0.188
18	Brazos	0.437	42.5	0.011	0.016	-0.628
19	Colorado (Texas)	0.642	0.148	0.056	0.054	-0.457

Similar to the first category, forecasting in the third category was done well. Referring to Figure 5-5, the 12-month periodicity was observed for Upper Colorado River. Using CESA, streamflow was forecasted up to 60 months ahead (*RE* of less than 0.35) with r^2 of 0.859 and *NSE* of 0.749. For forecasting for the first 24 months, *RE* was less than 0.15, and r^2 was higher than 0.90. The forecast error in the average streamflow was only 1.2%, and the peak was forecasted about 10%-22.5% less. For the remaining rivers (4, 11 and 14) in the third category, streamflow was accurately forecasted with a 36-month lead time (*RE* of less than 0.36). The peak values were forecasted satisfactorily with no

shift in time, though Yampa River was an exception. Peak streamflow of Yampa River used to occur more often in May, but during 1999 and 2001, the observed peak shifted to June. CESA did not recognize this change, as it honored the 12-month periodicity.

Forecasting streamflow of rivers in the fourth category was similar to the second one. Thus, it was less accurate than in the third category due to the inconsistency from year to year. It can be seen for rivers 8, 12, 13 and 16 that peak streamflow varied more than 20% from year to year. CESA forecasted streamflow of Yellowstone River and Lower Colorado River with a 24-month lead time (*RE* of less than 0.40), of Dolores River with a 36-month lead time (*RE* of less than 0.35), and of Green River with a 48-month lead time (*RE* of less than 0.34). For forecasting one-year ahead of time, *RE* computed for above rivers was less than 0.17. The overall r^2 values were over 0.7 and *NSE* larger than 0.5.

Forecasting for the first four categories can be considered as satisfactory (*RE* of less than 0.30), although some drawbacks may be noted. The most challenging is to forecast low streamflows due to weak seasonality and periodicity, and low autocorrelation. Streamflow for rivers numbered 15 and 19 are forecasted with lead times of 24 months and 36 months. The *NSE* values were only 0.139 and 0.054. At least the average were forecasted accurately, where error was less than 2.5% and 9% for rivers 15 and 19, respectively. For Trinity and Brazos Rivers in the last category, forecasting was most difficult. These two rivers were irregular. Again, CESA was accurate in forecasting the average value, where the difference between the forecasted mean and the observed mean was less than 2%.

Analyzing the forecast residuals. Figure 5-7 shows a Q-Q plot of the forecast residuals versus the normal distribution. If the errors fitted the 45˚ line, it was safe to conclude that the residuals followed the normal distribution. The Q-Q plot shows that the residuals for rivers 3 and 18 are left skewed, and for rivers 5 and 9 they are slightly left skewed. Except these rivers, other residuals fitted the normal distribution well, and therefore CESA performed well.

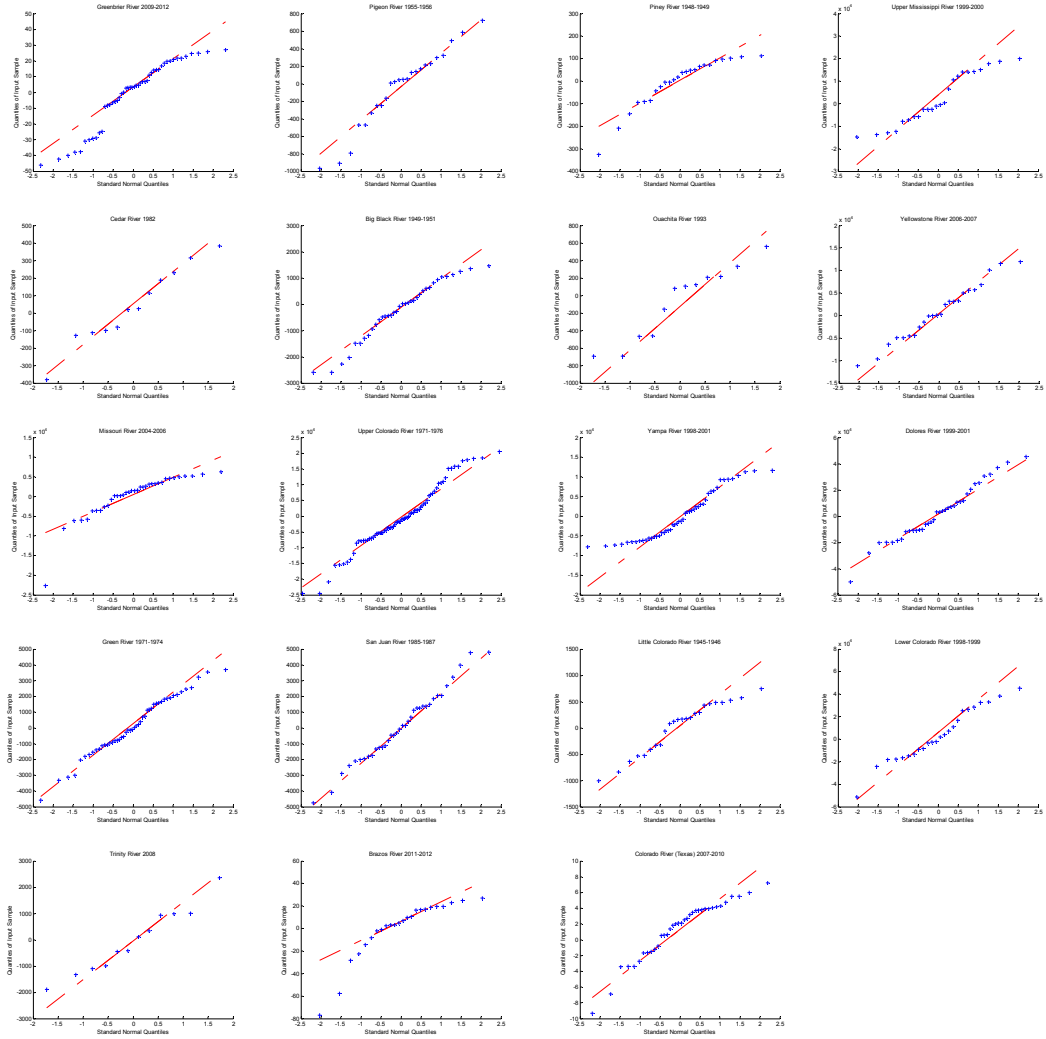


Figure 5-7 Q-Q plot of forecasting residuals

5.3 Comparison with BESA

The forecasted streamflows by two maximum entropy spectral analysis theories were plotted in Figure 5-8.

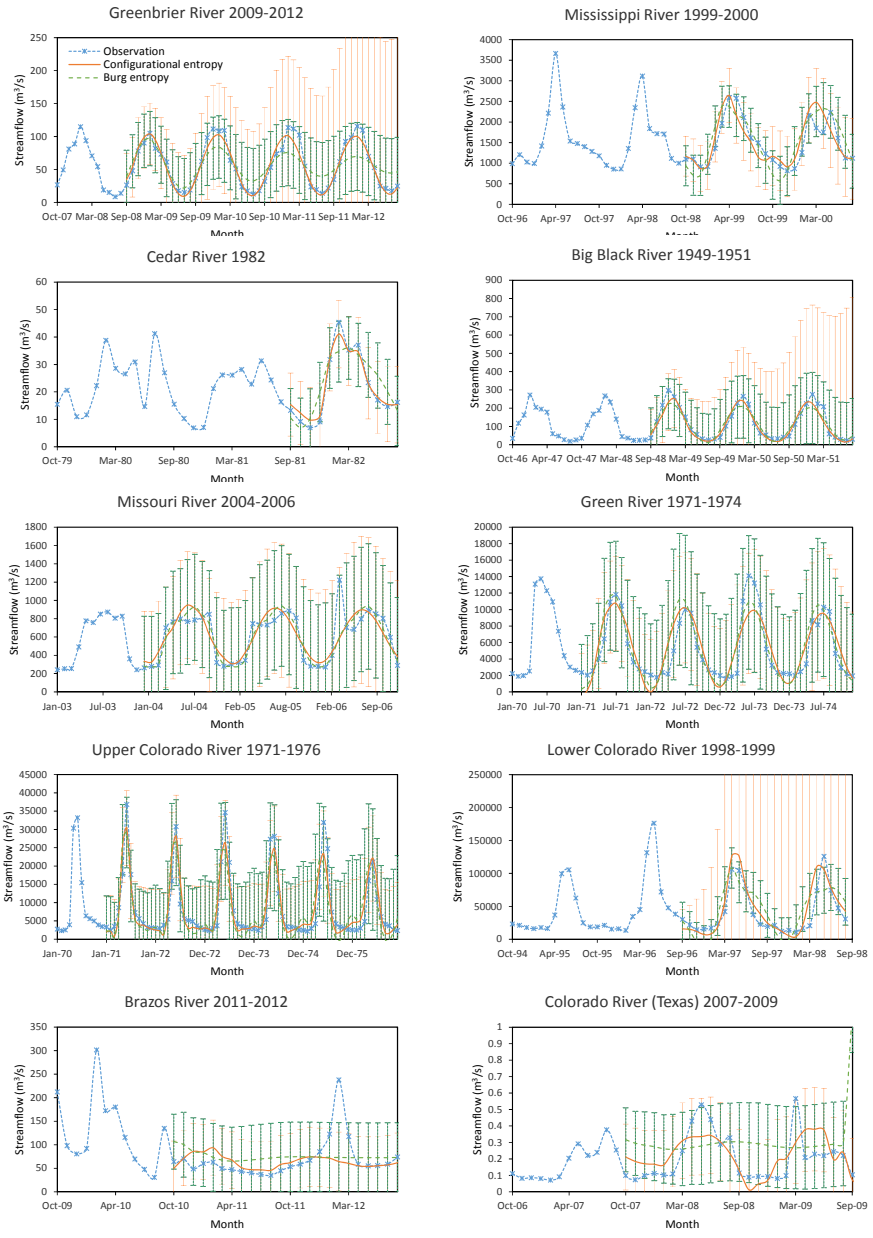


Figure 5-8 Comparison between CESA and BESA

For Greenbrier River the spectral peak shifted in the BE spectral estimation. As a result, the loss of detecting the periodicity led to less accurate forecasts. As shown in the first figure in Figure 5-8, BESA appropriately forecasted streamflow that matched the observed values during the first year (r^2= 0.89), but was in general not able to forecast after 24 months (r^2 lower than 0.50). It can be seen from the plot of variation of *NSE* over the forecast lead time in Figure 5-9 that the *NSE* value dropped faster in the beginning of the second year and became below 0.5 from the third year onwards for BESA, while that for CESA was consistently around 0.7 over the years. As a result, the goodness of fit measures for Greenbrier River were much lower for BESA.

Upper Colorado River and Green River had similar resolutions in estimating the spectral density for both BESA and CESA, as both were uni-peak spectral densities. BESA performed as well as CESA as shown in Figure 5-8, with *NSE* values of 0.523 and 0.687, respectively. However, as the lead time increased, *NSE* computed for BESA decreased faster than for CESA, as shown in Figure 5-9. Thus, for streamflows with unique periodicity, where there is uni-peak spectral density, both BESA and CESA forecasted streamflow accurately. But for longer lead-times, CESA was more consistent than BESA.

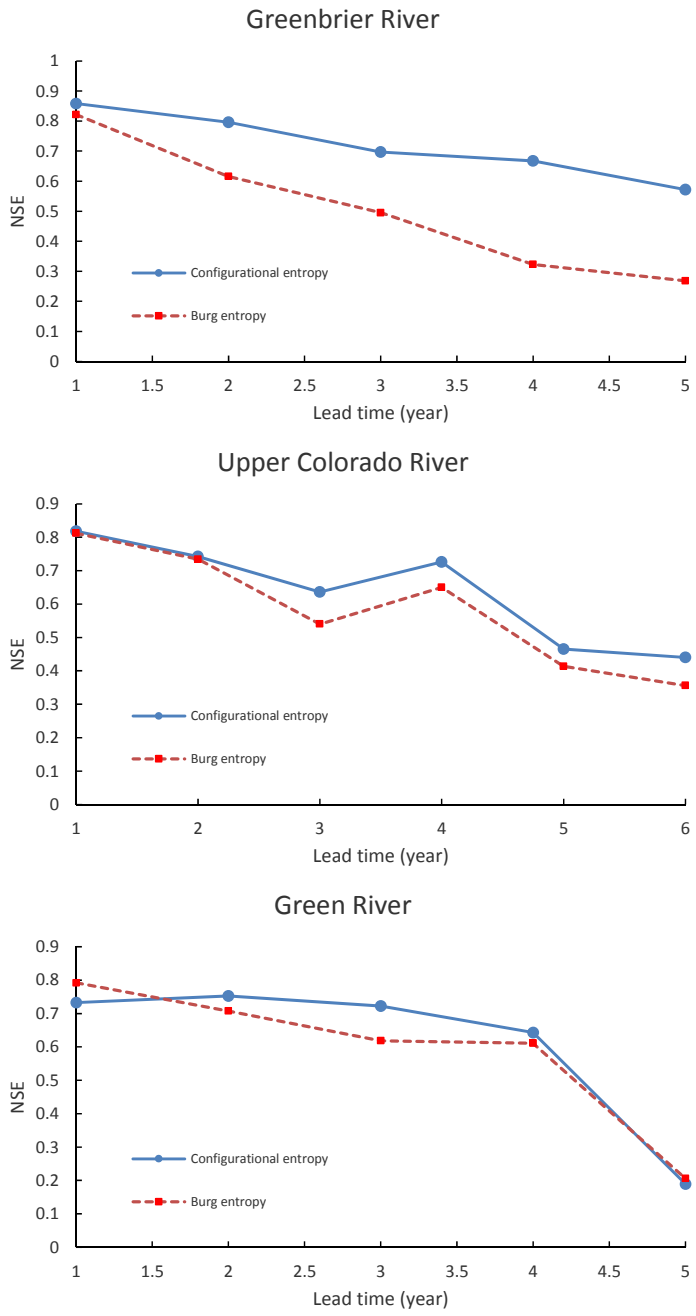


Figure 5-9 Nash-Sutcliffe coefficient against lead time for Greenbrier River, Upper Colorado River and Green River

It was shown in the previous section that streamflow of Mississippi River in the third category was correctly forecasted using CESA, but the spectral peak shifted by BESA. As a result, BESA forecasted peak streamflow during 1999 one month early but during 2000 one month late. We believe that it was due to the failure in detecting the location of spectral peak. The shift by BESA in forecast was also observed for Big Black River, Cedar River and Lower Colorado River, which belonged to the same spectral group as Mississippi River. For Mississippi River whose spectral density showed a multi-peak pattern, CESA is recommended over BESA. However, Missouri River was an exception to the above group. Though BESA failed to determine its spectral peak, streamflow was reasonably forecasted nonetheless. The *NSE* value for BESA was 0.511, which is comparable to 0.539 for CESA. The reason may be found from its "n"-shaped hydrograph. The peak streamflow of Missouri River lasted about 4 to 5 months, and the hydrograph was not as sharp as of other rivers. Thus, streamflow forecasting was less focused on peak time, which reduced the burden for BESA.

The performance of BESA in forecasting low streamflow was not satisfactory (*RE* higher than 0.7 and r^2 lower than 0.3). BESA was unable to forecast streamflow which remained flat after April, 2011, for Brazos River and showed small fluctuations for Colorado River (Texas). It showed that forecasted values were not correlated to the observed values and BESA was not valid for low streamflow forecasting. Though the predicted peaks for Brazos River were 53.5% lower than observed values in February, 2012, forecasts by CESA satisfactorily mimicked the streamflow pattern during the forecast period from October, 2010-December, 2011 with r^2 higher than 0.37, Besides, the

NSE values of CESA were positive and were larger than the negative values of BESA as shown in Tables 5-3 and 5-4. Thus, using CESA is suggested over BESA for low streamflow forecasting.

Table 5-4 Goodness of fit measures by Burg entropy spectral analysis

No.	River	Burg entropy spectral analysis				
		RE	*RMSE* (m^3/s)	r^2	*NSE*	*PE*
1	Greenbrier	0.458	20.7	0.646	0.457	-0.162
2	Pigeon	0.221	342.1	0.653	0.447	-0.264
3	Piney	0.333	106.9	0.693	0.523	-0.335
4	Mississippi	0.210	11430.0	0.640	0.408	-0.086
5	Cedar	0.321	211.7	0.759	0.548	-0.208
6	Big black	0.276	1115.8	0.873	0.688	-0.246
7	Ouachita	0.133	404.6	0.793	0.590	-0.039
8	Yellowoston	0.871	180.3	0.000	-0.264	-0.418
9	Missouri	0.190	5148.2	0.635	0.511	-0.261
10	Upper Colorado	0.593	4527.7	0.738	0.687	-0.209
11	Dolores	3.433	6070.2	0.000	-0.994	-0.097
12	Yampa	0.527	19233.0	0.497	0.664	-0.361
13	Green River	0.418	1850.8	0.746	0.523	-0.237
14	San Juan	0.255	2245.0	0.540	0.345	-0.307
15	Little Colorado	-8.583	474.9	0.279	0.017	-0.406
16	Lower Colorado	0.758	23308.1	0.506	0.300	-0.139
17	Trinity	0.536	1738.0	0.355	0.212	-0.397
18	Brazos	0.436	42.5	0.000	-0.196	-0.551
19	Colorado (Texas)	4.872	0.990	0.000	-3.956	0.037

5.4 Discussion

Based on consistency, seasonality and periodicity of observed data, 19 rivers were categorized into six different groups. The possible reasons for observing different characteristics are analyzed in this section.

It may be recalled that streamflow of rivers in the first two groups during the peak three months was between 30% and 50% of the annual volume, while that of the third and fourth groups was more than 60%. This pattern of streamflow may be influenced by geographical factors. In the first two groups, rivers are mainly from the watersheds of Mississippi River, Tennessee River and Missouri River, where precipitation over the year is sufficient and nearly uniformly distributed, and rivers are dry during low water periods. Another reason which makes Greenbrier River unique in the first group may be its non-artificiality. There is no dam or artificial hydraulic structure on this river, and thus the streamflow maintains a regular pattern, which makes it highly predicable. On the contrary, for Colorado River precipitation is concentrated during summer and it is unlikely to maintain high streamflow during the dry season. As a result, monthly hydrographs of rivers in the first two groups are fatter than that of Colorado River. These seasonal patterns are satisfactorily forecasted by CESA. The forecasted streamflow during March-May was 34%-45% for the rivers in first two groups, and forecasted streamflow during May to July was 53%-77% for the rivers in the Colorado River watershed.

In order to analyze the effect of drainage areas of the selected sites, the relative drainage area was computed by dividing the site drainage area by the total drainage area of the river. In general, the larger relative drainage area was, the closer to the mouth the selected site was. Thus, sites 1, 2, 3, 5, 6, 7, 10, 11, 14, and 18 were upstream near the source, and sites 4, 8, 9, 12, 13, 15, 16, 17, and 19 downstream near the mouth. It was found that the peak ratio was correlated to the relative drainage area, as shown in Figure 5-10. Though the data points were scattered, the peak/average ratio was larger for

downstream than upstream sites. It may be explained as follows. First, as the site went downstream, more and more water from different tributaries joined the mainstream, and as a result the flow variance increased. On the other hand, the water source for the upstream site was constant, which led to less flow variance. One may notice that river sites considered to be upstream fell into the first three groups, where streamflow was more consistent, had stronger seasonality and periodicity than other groups. The river sites in the fourth group, where the peak was inconsistent from year to year, were more likely to be downstream.

Furthermore, the uni-peak spectral density condition (river sites 1, 3, and 10) can only be satisfied at the upstream sites, where the source of the water was unique or simple. Thus, streamflow at the upstream sites was more likely to keep the uni-periodicity, say 12-month periodicity, than the downstream sites. Though river sites 2, 5, 6, 7 were upstream, their streamflows had multi-spectral peaks. However, the spectral peak of frequency other than $1/12^{th}$ was not comparable to the largest peak. Due to the complexity of tributaries providing water to the downstream stations, the river flows hardly kept uni-periodicity at the downstream sites. The spectral densities of streamflows of rivers 4, 8, 9, 12, 12, 13 that had multi-spectral peaks were categorized to be from downstream sites. For these rivers, 6-month, 4-month, and even 3-month as well as 12-month periodicities were detected from the spectral density plot. CESA neglected small peaks when estimating the spectral density, as a result, it ignored other monthly periodicities when forecasting streamflow. Thus, the forecasted streamflows for uni-periodicity rivers were more accurate with *NSE* higher 0.75 than multi-periodicity rivers with *NSE* higher 0.60. As a

result, the accuracy of forecasting streamflow was influenced by the relative drainage area. The upstream streamflow was forecasted with higher r^2 and *NSE*, as shown in Figure 5-10. The r^2 value for forecasting upstream was about 0.84, and was larger than 0.75 for forecasting downstream.

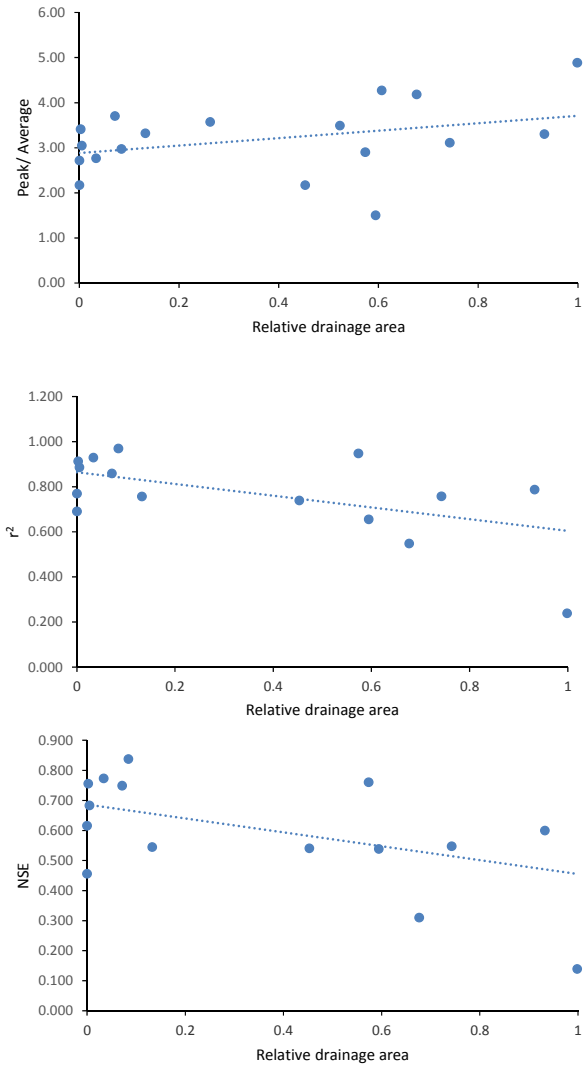


Figure 5-10 Forecasting characteristics related to relative drainage area

Consistency of climate condition may be another reason influencing the accuracy of forecast. The forecasted peak time was 1 month earlier for rivers 2, 7, and 9, and was 1 month late for rivers 3 and 6. The streamflow peaks observed for these rivers were not constantly the same over the years, but changed by a month or two due to the change in the distribution of precipitation or change of temperature. However, both CESA and BESA are based on the information from the past values, which implied a 12-month periodicity observed from data. Thus, streamflow forecasts using both these theories had peaks repeated every 12 months so that the location was consistent with the peak location that occurred in the previous year. The reason for the shift in the peak location was the change in input. It is known that the sources of tributaries of Upper Mississippi River and Upper Colorado River are snowmelt as well as rainfall. Thus, the occurrence of streamflow peak was not fixed but may shift due to different combinations of rainfall and snowmelt. The forecast theories discussed in this part did not consider rainfall or snowmelt as input, hence streamflow forecasting in the peak season affected by the change of climate condition might have bias as compared with observed values. However, the multi-channel spectral analysis with incorporation of climatic phenomena will be developed in Section IX to see how forecasting can be improved with additional climate information.

5.5 Summary

Configurational entropy spectral analysis was developed for monthly streamflow forecasting in this section. The development of CE theory linked time series, spectral analysis and cepstrum analysis, and provided a new approach for solving the autoregressive coefficient with autocepstrum.

The spectral density computed from the proposed CESA had higher resolution than from BESA. For uni-peak cases, both CESA and BESA determined the spectral density well and the spectral peak was obtained at frequency $1/12^{th}$. For multi-peak cases, the two theories produce different results. CESA ignored non-significant peaks and yielded uni-peak with $1/12^{th}$ frequency. On the contrary, BESA detected multi-spectral peaks but the whole spectral density is shifted. Thus, for streamflow forecasting, CESA was preferred, as the right peak location is an important criterion for periodicity. For computing spectral density for low streamflow, the spectral density obtained by CESA was consistent with the one obtained from Fourier transform, while BESA was unable to detect its pattern.

The streamflow forecasted by CESA was more accurate for upstream than downstream. CESA theory forecasts with less than 30% error for lead times up to 60 months for rivers located upstream and up to 48 months for downstream rivers. The streamflows located near upstream are found to possess stronger seasonality and periodicity and had more regular patterns than streamflow downstream. Thus, the streamflow forecasted for upstream locations can be more reliable for longer lead time.

Furthermore, CESA was comparable to BESA for forecasting streamflow with a strong unique periodicity, though CESA led to longer lead time forecasting than BESA. For streamflow with multi-periodicity, CESA had an advantage over BESA. The streamflow peaks forecasted by BESA were more often shifted than by CESA. For low streamflow forecasting, where BESA was not valid any more, CESA satisfactorily detected weak seasonality for low streamflow and the forecasted series fluctuates as do observations and hence is suggested over BESA.

However, if streamflow was uncorrelated to its past pattern, then both CESA and BESA were unable to forecast. With the change in the source of streamflow, the occurrence of streamflow peak may change. The two theories were not able to track the change but forecast the peak repeated with the periodicity detected from the spectral density. As a result, the peak location forecasted by both theories may be either earlier or later than observed. Both theories rely on the autocorrelation of data itself, and forecasting is based on its past series. Without additional input information, neither method was capable of detecting the sudden change.

6 MINIMUM RELATIVE ENTROPY SPECTRAL ANALYSIS WITH SPECTRAL POWER AS A RANDOM VARIABLE

The minimum relative entropy (MRE), also called minimum cross-entropy (MCE), was introduced by Kullback (1959), which is an information-theoretic measure of the dissimilarity between two probability distributions. Two decades later, Shore (1979; 1981) developed the minimum relative entropy spectral analysis (MRESA) as an extension of Burg's maximum entropy spectral analysis, where the spectral power was considered as a random variable. Later, another version of MRESA was developed by Tzaneess et al. (1985) considering frequency as a random variable.

The MRESA theory can determine the spectra with the maximum value of kurtosis, which is an indicator reflecting the peakedness of spectral density (Endo and Randall, 2007). The MRESA-based spectra are reported to have higher resolution and are more accurate in detecting peak locations than other methods for spectral computation (Papademetriou, 1998). The theory refines the main frequencies and allows a detection of very close peaks and does not create artificial peaks (Berger et al., 1990). When linking to time series analysis, the MRESA theory is equivalent to linear prediction with the smallest Itakura-Saito distortion (Gray et al., 1981; Schroeder, 1982; Shore, 1981; Shore and Gray, 1982). Beside, MRESA theory reduces the number of predictor coefficients by relying on the prior information (Schroeder, 1982). However, the MRESA theory has only been applied for forward modeling and for solving inverse problems in groundwater

(Woodbury and Ulrych, 1993; 1996; 1998), but has not been applied to streamflow forecasting yet.

In this and the following sections, the minimum relative entropy spectral analysis is developed for monthly streamflow forecasting. The MRESA is derived by assuming spectral power as a random variable in this section, and is abbreviated as RESAS.

6.1 Development of minimum relative entropy spectral analysis

Assuming each spectral power x_k, defined as equation (3-11), as a random variable, the development of RESAS theory consists of the following steps: 1) define the relative or cross entropy, 2) construct the constraints, 3) assume a prior distribution of spectral power, 4) determine the probability density function of spectral power, 5) determine the Lagrange multipliers, 6) determine the spectral power, 7) extend the autocorrelation or autocovariance function, and 8) forecast streamflow.

6.1.1 Definition of relative entropy

Let the streamflow time series be described by a single state $\vec{x}=(x_1,x_2,...,x_k,...,x_n)$ in domain *D*. If each x_k is considered to be a random variable, then the joint probability density function $q(\vec{x})$ can be used to describe the whole streamflow series. In order to determine the probability density function $q(\vec{x})$, each $q(x_k)$ is considered independent and identically distributed:

$$q(\vec{x}) = q(x_1)q(x_2)\cdots q(x_n) = \prod_{k=1}^{n} q(x_k) \tag{6-1}$$

Then, the relative entropy is defined as

$$H(q,p)=\int_D q(\vec{x})\ln[q(\vec{x})/p(\vec{x})]d\vec{x} \tag{6-2}$$

where $q(\vec{x})$ is the posterior probability density function, and $p(\vec{x})$ is the prior probability density function. It is noted that $q(x_k)\geq 0$ and $p(x_k)\geq 0$ for any $x_k \in D$. It is interesting to note that when there is no prior given or given prior is uniform, the minimization of equation (6-2) yields $H(q)_{\max}=-\int_D q(\vec{x})\ln[q(\vec{x})]d\vec{x}$, which is equivalent to the maximum entropy theory derived in equation (4-1). In order to obtain the least-biased posterior distribution $q(\vec{x})$, the relative entropy is minimized subject to given constraints.

6.1.2 Specification of constraints

First of all, the probability density function must satisfy that

$$\int_D q(\vec{x})d\vec{x}=1 \tag{6-3}$$

The other constraints are constructed from the relationship between spectral power and autocorrelation. Let T_k denote the expected value of x_k, written as

$$T_k=\int_D x_k q(\vec{x})d\vec{x} \tag{6-4a}$$

Since the probability of each frequency is independent of each other, the integration of equation (6-4a) can be done separately as

$$\begin{aligned} T_k &= \int_D x_k q(\vec{x})d\vec{x}=\iiint_j x_k\prod_{j=1}^{n} q(x_j)dx_1dx_2\cdots dx_n \\ &= \int x_k q(x_k)dx_k\iiint_{j\neq k} q(x_j)dx_j=\int x_k q(x_k)dx_k \end{aligned} \tag{6-4b}$$

where the second integration yields $\iiint_{j\neq k} q(x_j)dx_j=1$. It is known that the mean spectral

power T_k is the Fourier transform of autocovariance R_r. Vice versa, the autocovariance can be expressed as the inverse Fourier transform of T_k as

$$R_r = \sum_{k=1}^{n} T_k \exp(2\pi i r \Delta t f_k) = \sum_{k=-n}^{n} T_k c_{rk}\ ,\ -N \le r \le N \tag{6-5}$$

where $c_{rk} = \exp(2\pi i r \Delta t f_k)$, r is the lag, Δt is taken as 1 month for monthly streamflow, and N is the largest lag for given autocorrelation or autocovariance, usually taken ¼ or ½ of streamflow length T. Substituting equation (6-4a) into equation (6-5), the probability density function can be linked to the autocovariance as

$$R_r = \int_D \sum_{k=1}^{n} x_k c_{rk} q(\vec{x}) d\vec{x}\ ,\ -N \le r \le N \tag{6-6}$$

Thus, the autocovariance functions from lag $-N$ to N, along with equation (6-3), are considered as $2N+2$ constraints for applying the RESAS theory.

6.1.3 Hypothesis on prior distribution

Before minimizing the relative entropy, a distribution of the prior $p(\vec{x})$ should be hypothesized from prior information about data. Any kind of distribution can be taken as a prior distribution. The uniform, exponential, normal, and gamma distributions are taken as the prior distribution.

If the prior distribution is given as multivariate exponential distribution, then $p(x_k)$ can be written as exponential distribution with the expected spectral powers S_k at each frequency, which is

$$p(x_k) = \frac{1}{S_k} \exp(-\frac{x_k}{S_k}) \tag{6-7}$$

where the expected spectral powers $S_k = \int x_k p(x_k) dx_k$ is the prior information about the spectral power.

If the prior spectral powers identically follow the normal distribution, the prior distribution can be expressed as

$$p(x_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp[-\frac{(x_k - S_k)^2}{2\sigma_k^2}] \tag{6-8}$$

where S_k and σ_k are the mean and standard deviation of x_k, respectively.

If the prior spectral power is given to follow the gamma distribution, then prior distribution can be written as

$$p(x_k) = \frac{1}{\Gamma(\eta_k)\theta_k^{\eta_k}} x_k^{\eta_k - 1} e^{-x_k/\theta_k} \tag{6-9}$$

where θ_k and η_k are the given shape and scale parameters of the gamma distribution of the prior spectral power $S_k = \eta_k \theta_k$.

6.1.4 Determination of distribution of spectral power

The posterior probability density function is then derived by minimizing the cross entropy using the method of Lagrange multipliers. Using constraints in equations (6-3) and (6-5), the Lagrangian function can be written as

$$\begin{aligned} L(q, p) &= \int q(\vec{x}) \ln(q(\vec{x}) / p(\vec{x})) d\vec{x} + (\lambda_0 - 1)[\int_D q(\vec{x}) d\vec{x} - 1] \\ &+ \sum_{r=-N}^{N} \lambda_r [\int_D \sum_{k=1}^{n} x_k c_{rk} q(\vec{x}) d\vec{x} - R_r] \end{aligned} \tag{6-10}$$

where λ_r are Lagrange multipliers. Taking the partial derivative of equation (6-10) with respect to $q(\vec{x})$, and equating the derivative to zero, one obtains

$$\frac{\partial L(q,p)}{\partial q(\vec{x})} = \ln(q(\vec{x})/p(\vec{x})) + \lambda_0 + \sum_{r=-N}^{N} \lambda_r \sum_{k=1}^{n} x_k c_{rk} = 0 \tag{6-11}$$

Thus, the posterior distribution of spectral power becomes

$$q(\vec{x}) = p(\vec{x}) \exp[-\lambda_0 - \sum_{r=-N}^{N} \lambda_r \sum_{k=1}^{n} x_k c_{rk}] \tag{6-12a}$$

Equation (6-12a) is the joint posterior distribution that is determined from the prior and the Lagrange multipliers. If we assume the posterior probability distributions of spectral power x_k are identically independent of each other, the posterior probability distribution can be written by multiplying the probably of each spectral power as

$$\begin{aligned} q(\vec{x}) &= q(x_1)q(x_2)\ldots q(x_n) \\ &= \exp(-\lambda_0)[p(x_1)\exp(-\sum_{r=-N}^{N} \lambda_r x_1 c_{r,1})]\ldots[p(x_n)\exp(-\sum_{r=-N}^{N} \lambda_r x_n c_{r,n})] \end{aligned} \tag{6-12b}$$

Thus, dividing each probability distribution by $\exp(-\lambda_0)$, the probability distribution of the specific spectral power x_k can be written as

$$q(x_k) = p(x_k) \exp[-\frac{\lambda_0}{n} - \sum_{r=-N}^{N} \lambda_r x_k c_{rk}] \tag{6-13}$$

It can be seen from equation (6-13) that no matter what prior is chosen, the posterior distribution belongs to the exponential family but is sensitive to how the prior distribution is chosen.

6.1.5 Determination of Lagrange multipliers

The Lagrange multipliers can be numerically solved by inputting equations (6-12) into equations (6-3) and (6-6), respectively, as

$$\int_D p(\vec{x}) \exp[-\lambda_0 - \sum_{r=-N}^{N} \lambda_r \sum_{k=1}^{n} x_k c_{rk}] d\vec{x} = 1 \tag{6-14}$$

and

$$R_r = \int_D p(\vec{x}) \exp[-\lambda_0 - \sum_{r=-N}^{N} \lambda_r \sum_{k=1}^{n} x_k c_{rk}] \sum_{k=1}^{n} x_k c_{rk} d\vec{x}, \; -N \le r \le N \tag{6-15}$$

Since the probability distribution of each frequency is independent of the other, equation (6-14) becomes

$$\begin{aligned} 1 &= \exp(-\lambda_0) \iiiint \prod_k p(x_k) \exp(\sum_{r=-N}^{N} \lambda_r x_k c_{rk}) dx_k \\ &= \exp(-\lambda_0) \prod_k \int p(x_k) \exp(\sum_{r=-N}^{N} \lambda_r x_k c_{rk}) dx \end{aligned} \tag{6-16a}$$

which is equivalent to

$$\exp(\lambda_0) = \prod_k \int p(x_k) \exp(\sum_{r=-N}^{N} \lambda_r x_k c_{rk}) dx \tag{6-16b}$$

and equation (6-15) yields

$$\begin{aligned} R_r &= \exp(-\lambda_0) \iiiint \prod_k p(x_k) \exp(-\sum_{r=-N}^{N} \lambda_r x_k c_{rk}) \sum_{k=-n}^{n} x_k c_{rk} dx_k \\ &= \exp(-\lambda_0) \prod_k \left[\int p(x_k) \exp(-\sum_{r=-N}^{N} \lambda_r x_k c_{rk}) x_k c_{rk} dx_k + \int q(x_j) \sum_{j \ne k} x_j c_{rj} dx_j \right] \\ &= \exp(-\lambda_0) \prod_k \left[\int p(x_k) \exp(-\sum_{r=-N}^{N} \lambda_r x_k c_{rk}) x_k c_{rk} dx_k + \overline{\sum_{j \ne k} x_j c_{rj}} \right] \end{aligned} \tag{6-17}$$

Thus, the Lagrange multipliers are estimated by solving equations (6-16) and (6-17) with $2N+1$ nonlinear equations using the Newton-Raphson method. A FORTRAN code for solving these equations can be found in Woodbury (2004).

6.1.6 Determination of posterior spectral power

The posterior spectral power is needed to extend the autocovariance and can be obtained by inserting equation (6-13) in equation (6-4b) as

$$T_k = \int x_k q(x_k) dx_k = \int x_k p(x_k) \exp(-\frac{\lambda_0}{n} - \sum_{r=-N}^{N} \lambda_r x_k c_{rk}) dx_k \tag{6-18}$$

Equation (6-18) does not lend itself to an analytical solution for posterior spectral power, except for the prior using uniform and exponential distributions. When no specific prior distribution is assumed, the posterior expected spectral power from equation (6-18) is determined from steps explained in equation (4-14)-(4-19), which yields

$$T_k = \frac{1}{u_k} = \frac{1}{\sum_{r=-N}^{N} \lambda_r c_{rk}} \tag{6-19}$$

where $u_k = \sum_{r=-N}^{N} \lambda_r c_{rk}$. When the prior distribution is given as exponential for each frequency, with expected spectrum power S_k for each frequency, then the prior becomes

$$p(\vec{x}) = \prod_k \frac{1}{S_k} \exp(-\frac{x_k}{S_k}) \tag{6-20}$$

Inputting equation (6-20) into the posterior distribution equation (6-12), the result becomes

$$\begin{aligned} q(\vec{x}) &= p(\vec{x}) \exp[-\lambda_0 - \sum_{r=-N}^{N} \lambda_r \sum_{k=-n}^{n} x_k c_{rk}] \\ &= \exp(-\lambda_0) \prod_k \frac{1}{S_k} \exp(-\frac{x_k}{S_k}) \exp[-\sum_{r=-N}^{N} \lambda_r c_{rk} x_k] \\ &= \exp(-\lambda_0) \prod_k \frac{1}{S_k} \exp(-(u_k + \frac{1}{S_k}) x_k) \end{aligned} \tag{6-21}$$

where $u_k = \sum_{r=-N}^{N} \lambda_r c_{rk}$, the same as previously defined. It can be seen from equation (6-21) that the posterior distribution of spectral power is in the form that is transformable to the exponential distribution as $\eta \exp(-\eta x)$, where $u_k + \frac{1}{S_k}$ can be considered as the exponential parameter of each distribution. Thus, λ_0 needs to satisfy that $\exp(-\lambda_0)\prod_k \frac{1}{S_k} = \prod_k u_k + \frac{1}{S_k}$ so that equation (6-21) forms the multivariate exponential distribution as

$$q(\vec{x}) = \prod_{k=1}^{n} (u_k + \frac{1}{S_k}) \exp[-(u_k + \frac{1}{S_k}) x_k] \tag{6-22}$$

and the posterior distribution of each frequency is

$$q(x_k) = (u_k + \frac{1}{S_k}) \exp[-(u_k + \frac{1}{S_k}) x_k] \tag{6-23}$$

Thus, the expected posterior spectral power at each frequency becomes

$$T_k = \int x_k q(x_k) dx_k = \int x_k (u_k + \frac{1}{S_k}) \exp[-(u_k + \frac{1}{S_k}) x_k] dx_k = \frac{1}{\frac{1}{S_k} + \sum_{r=-N}^{N} \lambda_r c_{rk}} \tag{6-24}$$

6.1.7 Extension of autocovariance

The next question is how the autocorrelation or autocovariance can be extended by the RESAS theory. From the previous sections, spectral power is obtained by minimizing the relative entropy, subject to the constraints given as autocovariance functions of lag $r \leq N$. For lag r larger than N, the autocovariance function can be obtained from the spectral power according to the Wiener–Khinchin theorem, which says that the

autocovariance function satisfies the Fourier transform relation with the spectral power. Thus, using the spectral power obtained from equation (6-18), the autocovariance to be extended can be written as

$$R_{N+j} = \sum_{k=1}^{n} T_k c_{N+j,k} = \sum_{k=-n}^{n} c_{N+j,k} \int x_k p(x_k) \exp(-\frac{\lambda_0}{n} - \sum_{r=-N}^{N} \lambda_r x_k c_{rk}) dx_k , j>0 \qquad (6\text{-}25)$$

Equation (6-25) is the straight-forward mathematical expression of how the autocovariance can be extended under the RESAS theory. It can be seen from equation (6-25) that the extended autocovariance can be estimated from the Lagrange multipliers and the hypothesis on the prior distribution. However, the extension of autocovariance cannot be explicitly solved for.

On the contrary, the autocovariance or autocorrelation function can be extended in a more convenient way. Shore (1981) discussed that RESAS theory is equivalent to linear prediction with the smallest Itakura-Saito distortion with respect to the nontrivial prior used. It suggests that the autocovariance to be extended under the RESAS theory follows the linear form like BESA. It can be seen from comparing equation (6-18) to equation (6-19) that the only difference between the spectral power obtained by RESAS and BESA is on the first term of the denominator. Thus, by combining $1/S_k$ with λ_0, equation (6-20) can transferred to the form similar to that due to BESA, which is

$$T_k = \frac{1}{\sum_{r=-N}^{N} \lambda'_r \, c_{rk}} \qquad (6\text{-}26)$$

where λ'_r are modified Lagrange multipliers equaling

$$\lambda'_0 = \lambda_0 + \frac{1}{S_k} \tag{6-27a}$$

$$\lambda'_r = \lambda_r,\ r=1,\ \ldots,\ N \tag{6-27b}$$

Equation (6-26) shows that with the modified Lagrange multipliers, the spectral power is in the form of inverse polynomials similar to the BESA theory shown in equation (4-19). Thus, equation (6-27) can be further expanded as

$$T_k = \frac{1}{\sum_{r=-N}^{N} \lambda'_r\, c_{rk}} = \frac{1}{\sum_{r=-N}^{N} \lambda'_r\, z^{-r}} = \frac{\sigma^2}{\left|1 + \sum_{n=0}^{N} a_n z^{-n}\right|^2} \tag{6-28}$$

where $z = \exp(-2\pi i r \Delta t f_k)$, a_n are forecasting coefficients, and σ^2 is the gain satisfying the Yule-Walker equation (Yule, 1927). Equation (6-28) shows that the spectral power obtained by RESAS satisfies the form of linear prediction discussed in section 3.1.2 [equation (3-22)]. Thus, autocorrelation can be extended with linear combination of past lags using the prediction coefficients as

$$R_{N+1} = a_1 R_N + a_2 R_{N-1} + \ldots + a_m R_{N+m-1} \tag{6-29}$$

The prediction coefficient can be computed from the Lagrange multipliers. Equating the two equations on the right hand side of equation (6-28), the denominator becomes

$$\begin{aligned} &\sum_{r=-N}^{N} \lambda_r' z^{-r} = \sigma^{-2}\left[1 + a_1 z + a_2 z^2 + \ldots + a_N z^Z\right]\ \left[a + a_1^* z^{-1} + a_2^* z^{-2} + \ldots + a_N^* z^{-N}\right] \\ &= \sigma^{-2} \sum_{s=0}^{N} a_s z^s \sum_{s=0}^{N} a_s^* z^{-s} = \sigma^{-2} \sum_{r=-N}^{N} \sum_{k=0}^{N-r} a_{k+r} a_k^* z^r \end{aligned} \tag{6-30}$$

From equation (6-30), the modified Lagrange multipliers can be expressed by the convolution of forecasting coefficients as

$$\lambda_r' = \frac{1}{\sigma^2}\sum_{k=0}^{N-r} a_{k+r}a_k^* \tag{6-31}$$

Thus, prediction coefficients are solved from nonlinear equation (6-31).

6.1.8 Forecast

As discussed in sections 3.3.7 and IV.1.7, time series can be forecasted by weighting past series using coefficients computed from equation (6-31) as

$$y_{T+1} = a_1 y_T + a_2 y_{T-1} + \ldots + a_m y_{T+m-1} \tag{6-32}$$

where m is the order of forecasting model, which, based on the number of previous lags, will be used for forecasting the future, and is identified by the Akaike information criterion (AIC) or Bayesian information criterion (BIC) (Box and Jenkins, 1970; Hipel and McLeod, 1994).

6.2 Data description

As shown in Figure 6-1, 20 stations in the Mississippi River watershed were selected to apply the RESAS theory. Observed streamflows were downloaded from the U.S. Geological Survey (USGS) website. The stations are listed from upstream to downstream in Table 6-1 along with their drainage areas, locations and record lengths.

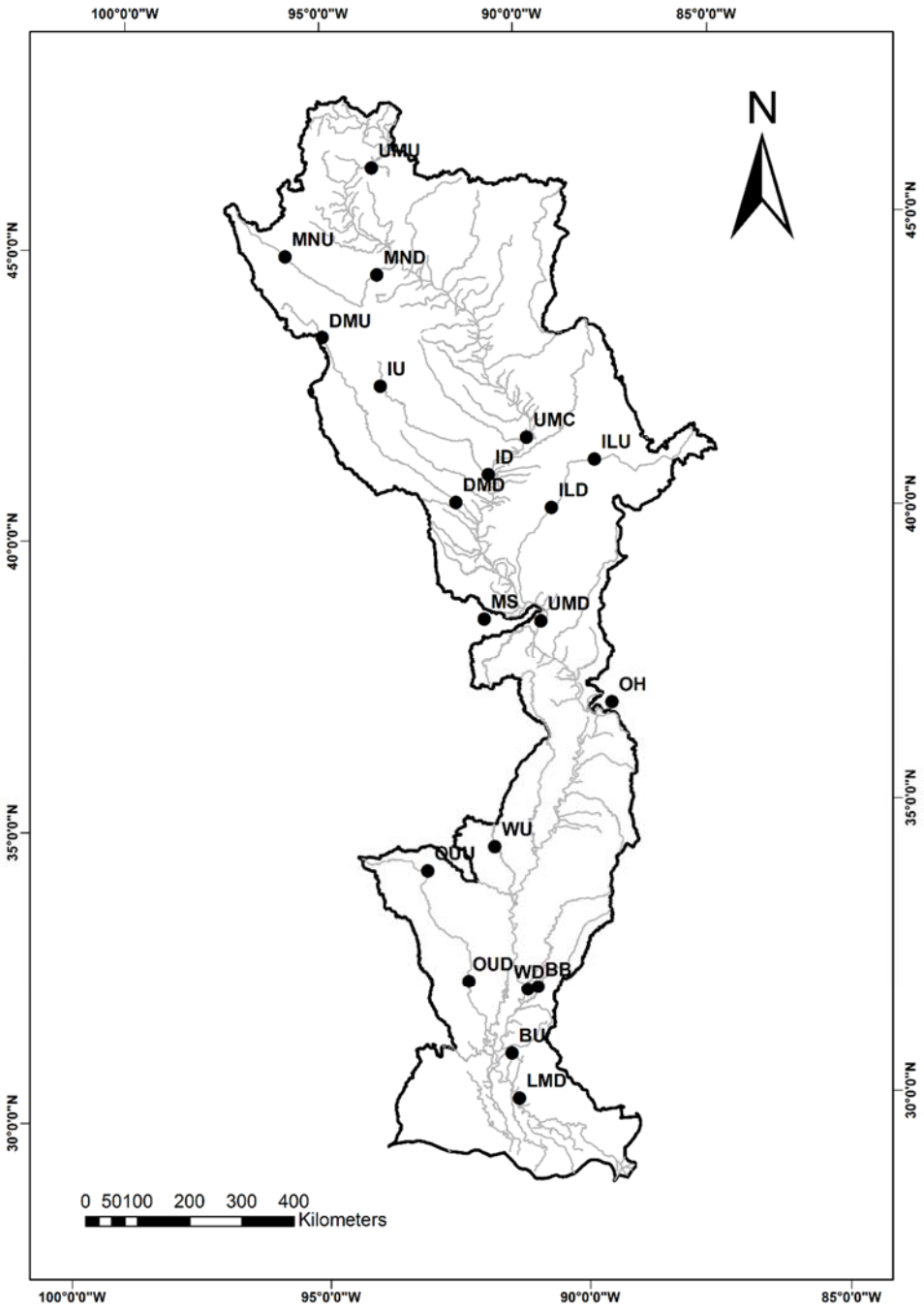


Figure 6-1 Selected stations on Mississippi River watershed

Table 6-1 Selected stations from Mississippi River watershed

Location	Name	Station	Area (km^2)	Latitude	Longitude	Record length	Tributary
Upper Mississippi River upstream	UMU	05227500	15903	46°32'26"	93°42'26"	1945-2013	
Minnesota River upstream	MNU	05301000	10489	45°01'17"	95°52'05"	1942-2013	Minnesota
Minnesota River downstream	MND	05330000	41958	44°41'35"	93°38'30"	1934-2013	Minnesota
Upper Mississippi River center	UMC	05420500	221704	41˚46'50"	90˚15'07"	1875-2013	
Iowa River upstream	IU	05449500	1111	42°45'36"	93°37'18"	1940-2013	Iowa
Iowa River downstream	ID	05465500	32375	41°10'41"	91°10'55"	1958-2013	Iowa
Des Moines River upstream	DMU	05476000	3237	43°37'06"	94°59'05"	1930-2013	Des Moines
Des Moines River downstream	DMD	05490500	36358	40°43'40"	91°57'34"	1969-2013	Des Moines
Illinois River upstream	ILU	05543500	21391	41°19'37"	88°43'03"	1919-2013	Illinois
Illinois River downstream	ILD	05568500	40968	40°33'11"	89°46'38"	1939-2013	Illinois
Upper Mississippi River downstream	UMD	07010000	1805222	38°37'44"	90°10'47"	1932-2013	
Missouri River	MS	06934500	1353269	38°42'35"	91°26'19"	1957-2013	Missouri
Ohio River	OH	03611500	525768	37°08'51"	88°44'27"	1928-2013	Ohio
White River upstream	WU	07077000	60606	34°47'40"	91°26'41"	1949-2013	White
White River downstream	WD	07289000	2964241	32°18'54"	90°54'21"	2008-2013	
Big Black River	BB	07290000	7283	32˚20'52"	90˚41'49"	1937-2013	Big black
Buffalo	BU	07295000	466	31°13'37"	91°17'44"	1942-2009	Buffalo
Ouachita River upstream	OUU	07359002	4015	34˚25'34"	92˚53'27"	1929-2013	Ouachita
Ouachita River downstream	OUD	07367005	39622	32°30'01"	92°07'11"	2007-2013	Ouachita
Lower Mississippi River downstream	LMD	07374000	2915848	30°26'44"	91°11'30"	2004-2013	

Mississippi River is the largest drainage system in North America, with many tributaries. The stations for Upper Mississippi River upstream (UMU), the Upper Mississippi River center (UMC), the Upper Mississippi River downstream (UMD) and the Lower Mississippi downstream (LMD) are on the main stem of the Mississippi River, and other stations are on its tributaries as noted in Table 6-1. For each tributary, one upstream station and one downstream station were chosen, if possible. The name of a station ending with the last letter *U* represents upstream, while letter *D* means downstream. Usually, Mississippi River can be divided into upper, middle and lower reaches. The Upper Mississippi (LM) is the part from head-waters to the confluence with the Missouri River, where the first 11 stations were selected. The middle Mississippi runs from downriver of Missouri (MS) to the Ohio River (OH), and the last 7 stations are from the Lower Mississippi (LM), which runs from the Ohio to the Gulf of Mexico. For simplicity, the MS and OH are categorized into the LM, thus, 11 UM stations and 9 LM stations are considered.

The selected 20 stations are distributed over the whole Mississippi River watershed, which is more than 3,000 km in length and covers a drainage area of over 2,981,000 km^2. As shown in Table 6-1, the drainage areas of selected stations vary from 466 km^2 to 2,915,858 km^2. Thus, streamflow characteristics are quite different from station to station. Basic statistics of streamflow of Mississippi River are listed in Table 6-2 that shows that average monthly streamflow discharge varies from 8 m^3/s to over 20,000 m^3/s over the whole watershed and the flow at the downstream reach is naturally larger than that from the upstream reach. The average peak streamflow is 1.3 to 7.9 times the mean streamflow,

and the peak over mean ratio is found to be generally larger for streamflow of the LM than of the UM. On the same tributary, the peak/mean ratio is also larger for a downstream station than an upstream station.

Table 6-2 Streamflow characteristics

Name	Mean (m^3/s)	Peak (m^3/s)	Peak/mean	Significant period*
UMU	77.1	133.6	1.7	12, 6, (4, 3)
MNU	67.0	105.4	1.6	12, 6, 4, 3
MND	149.8	388.4	2.6	12, 6, 4, 3
UMC	1693.8	2877.5	1.7	12, 6
IU	9.3	26.8	2.9	12, 6, (4, 3)
ID	304.8	513.6	1.7	12, (6, 4)
DMU	13.7	37.0	2.7	12, (6, 4, 3)
DMD	154.2	326.6	2.1	12, 6
ILU	375.1	531.4	1.4	12, 6, 3
ILD	638.5	929.2	1.5	12, 6, 3
UMD	6920.0	10408.4	1.5	12, (6)
MS	3376.4	4426.9	1.3	12, 6
OH	8723.4	18325.5	2.1	12
WU	6929.9	11992.4	1.7	12
WD	21596.4	41762.4	1.9	12, 6
BB	132.3	415.0	3.1	12, 6, 4
BU	8.6	17.4	2.0	12
OUU	90.9	145.5	1.6	12, 6, 4, 3
OUD	546.7	1079.3	2.0	12
LMD	616.4	4874.3	7.9	12

6.3 Periodicity analysis

Streamflow periodicity is a strong index for forecasting long-term streamflow. One of the efficient ways to detect periodicity is by transferring the time series data into frequency domain. The historical data 50 to 100 years long, before the year 2000, was

divided into 10-20 sub-sets so that one set of series contained at least five years of streamflow data. For each data sample of data, the spectral density was estimated by the Fast Fourier transform, which was plotted in Figure 6-2.

It can be seen from the figure that the 12 month periodicity was observed to be the most significant of all the stations in the Mississippi watershed. It was found that there were five types of spectral density obtained for the Mississippi River. The first type was uni-peak (f_k=1/12) spectral density observed at OH, White River upstream (WU), Buffalo River(BU), Ouachita River downstream (OUD), and LMD, which were all from the LM. The second was the double significant peak observed at f_k =1/12 and f_k =1/6, although the peak at 1/6 was not large compared to the peak at 1/12. The stations of UMC, Des Moines downstream (DMD), MS and White River downstream (WD) belonged to the second type, while UMD was somehow between the first and second types. The third type of spectral density was observed at Big Black River (BB) for having significant peaks at f_k =1/12 and f_k =1/4. The spectral density obtained for Illinois River belonged to the fourth type with significant peaks at f_k =1/12 and f_k =1/3. The last type of spectral density was significant at all four periodicities, which was the case with the Minnesota River having one large peak at 1/12 and small peaks at f_k =1/6, f_k =1/4 and f_k =1/3. The significant periodicity is tabulated in the last column of Table 6-2 that shows that in general the periodicity is almost consistent during the same stream, but the upstream reach may have more significant small peaks than the downstream reach.

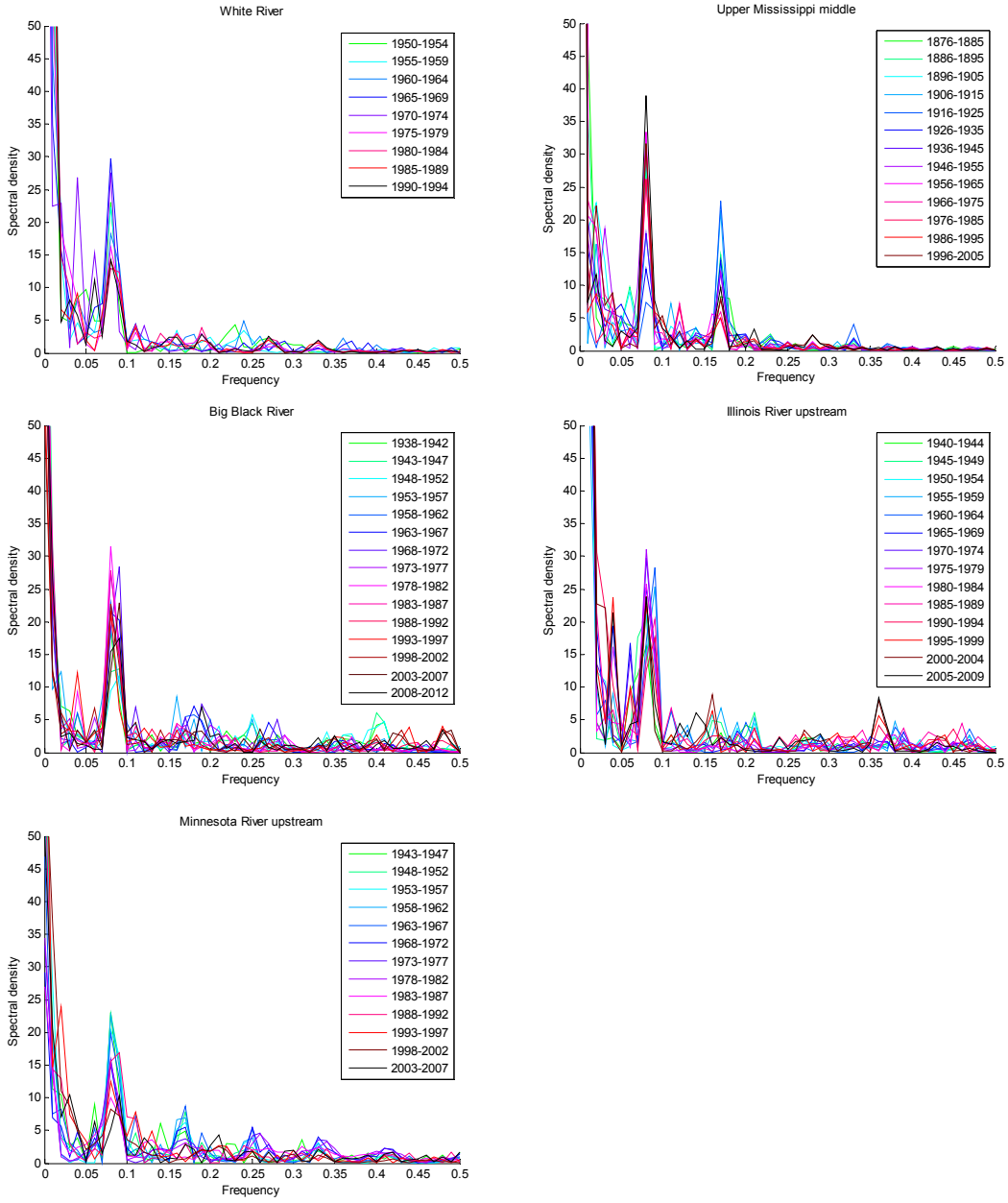


Figure 6-2 Spectral densities for different periodicities

6.4 Hypothesis on prior distribution

It was already shown in the previous section that the hypothesis on the prior distribution may considerably influence the final estimation of the spectral density and

hence further influence streamflow forecasting. Thus, it was necessary to determine which prior was proper for the observed series. To test the prior distribution hypothesis, one may impose the constraint from the expected spectral power using equation (6-3) and (6-4a). The Lagrangian function can be written as

$$L(q,p) = \int q(\vec{x}) \ln(q(\vec{x})/p(\vec{x}))d\vec{x} + (\gamma_0 - 1)[\int_D q(\vec{x})d\vec{x} - 1] + \sum_{k=1}^{n} \gamma_k [\int_D x_k q(\vec{x})d\vec{x} - T_k] \quad (6\text{-}33)$$

Taking the partial derivative with respect to $q(\vec{x})$,

$$\frac{\partial L(q,p)}{\partial q(\vec{x})} = \ln(q(\vec{x})/p(\vec{x})) + \gamma_0 + \sum_{k=1}^{n} \gamma_k x_k = 0 \quad (6\text{-}34)$$

Thus, the poster distribution $q(\vec{x})$ will have a form of

$$q(\vec{x}) = p(\vec{x}) \exp[-1 - \gamma_0 - \sum_{k=1}^{n} \gamma_k x_k] \quad (6\text{-}35)$$

Equation (6-35) is the posterior distribution obtained through the minimum relative entropy, and the Lagrange multipliers can be solved by inputting equation (6-35) into equations (6-3) and (6-4a). If the posterior probabilities of powers x_k are assumed to be identically independent to each other, the probability distribution of each power can be written as

$$q(x_k) = p(x_k) \exp[-\frac{\gamma_0}{n} - \gamma_r x_k] \quad (6\text{-}36)$$

The probability of the spectral power at different frequencies was estimated with the prior as uniform, exponential, normal or gamma distribution. The posterior probability

was evaluated with the Kolmogorov-Smirnov test using spectral values obtained from the historical data, and the result is shown in Table 6-3, where 1 represents the rejection of the probability distribution, otherwise 0. For all stations the probability density function estimated with the gamma prior probability was accepted by the test at all frequencies

Table 6-3 Hypothesis test on choosing the prior probability density for specific frequency (1=reject, 0=accept)

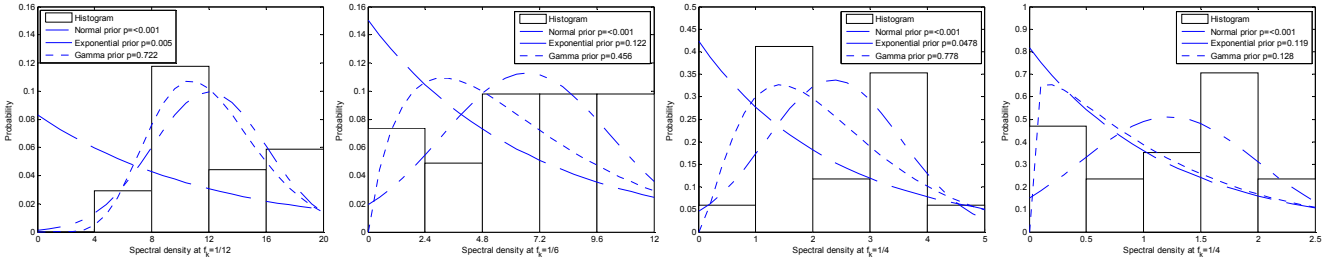
Station	Frequency	Uniform	Exponential	Normal	Gamma
UMU	1/12	1	1	1	0
	1/6	0	0	1	0
	1/4	1	1	1	0
	1/3	1	0	1	0
MNU	1/12	1	1	1	0
	1/6	0	0	1	0
	1/4	1	0	1	0
	1/3	1	1	1	0
UMC	1/12	1	1	1	0
	1/6	1	1	1	0
	1/4	1	0	1	0
	1/3	1	0	1	0
ILU	1/12	0	1	1	0
	1/6	1	1	1	0
	1/4	1	0	1	0
	1/3	1	0	1	0
UMD	1/12	1	1	1	0
	1/6	1	0	1	0
	1/4	1	0	1	0
	1/3	1	0	1	0
OH	1/12	1	1	1	0
	1/6	1	0	1	0
	1/4	1	0	1	0
	1/3	1	0	1	0

Table 6-3 Continued

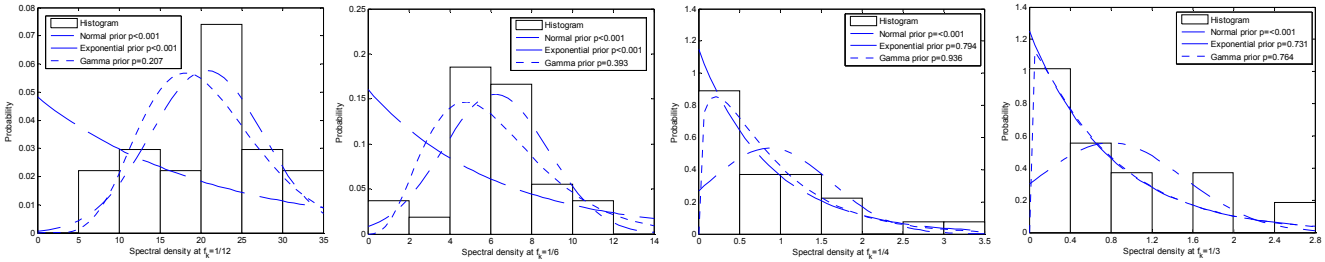
Station	Frequency	Uniform	Exponential	Normal	Gamma
BB	1/12	0	1	1	0
	1/6	0	1	1	0
	1/4	1	0	1	0
	1/3	0	0	1	0
BU	1/12	0	1	1	0
	1/6	0	0	1	0
	1/4	0	0	1	0
	1/3	0	0	1	0

On the other hand, the exponential probability prior was more possibly accepted for spectral powers at small frequencies, except for f_k=1/12. The uniform probability distribution as prior was rejected for almost all stations for the UM (Figure 6-3), but was accepted for the LM (Figure 6-4). The OH and BU in Figure 6-4 both showed that spectral density was only significant at f_k=1/12. Thus, it is reasonable to use uniform prior to determine the probability of those insignificant peaks. Though the posterior with normal prior seemed to fit the observed histogram at f_k=1/12 and f_k=1/6 in Figure 6-3 and of f_k =1/12 in Figure 6-4, the null hypothesis that the spectral power fitted the normal distribution, was rejected for p-values lower than 0.001 for all cases.

Upper Mississippi River upstream



Upper Mississippi River center



Upper Mississippi River downstream

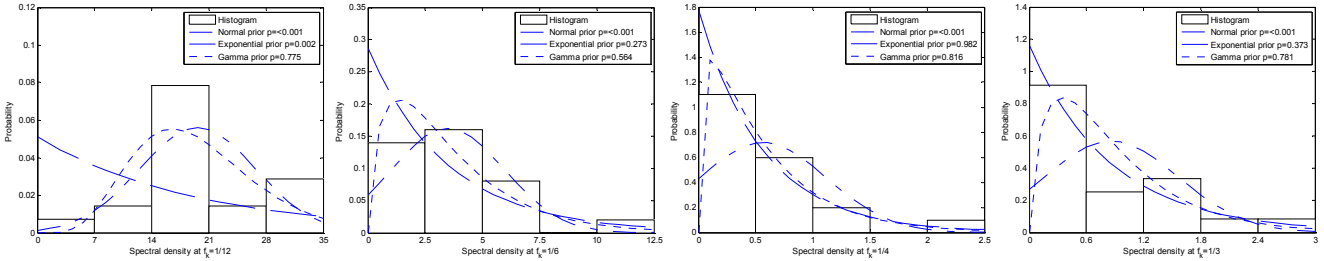


Figure 6-3 Posterior probability densities of spectral powers at frequencies 1/12, 1/6, 1/4 and 1/3 estimated with different priors for Upper Mississippi River

Ohio River

Buffalo River

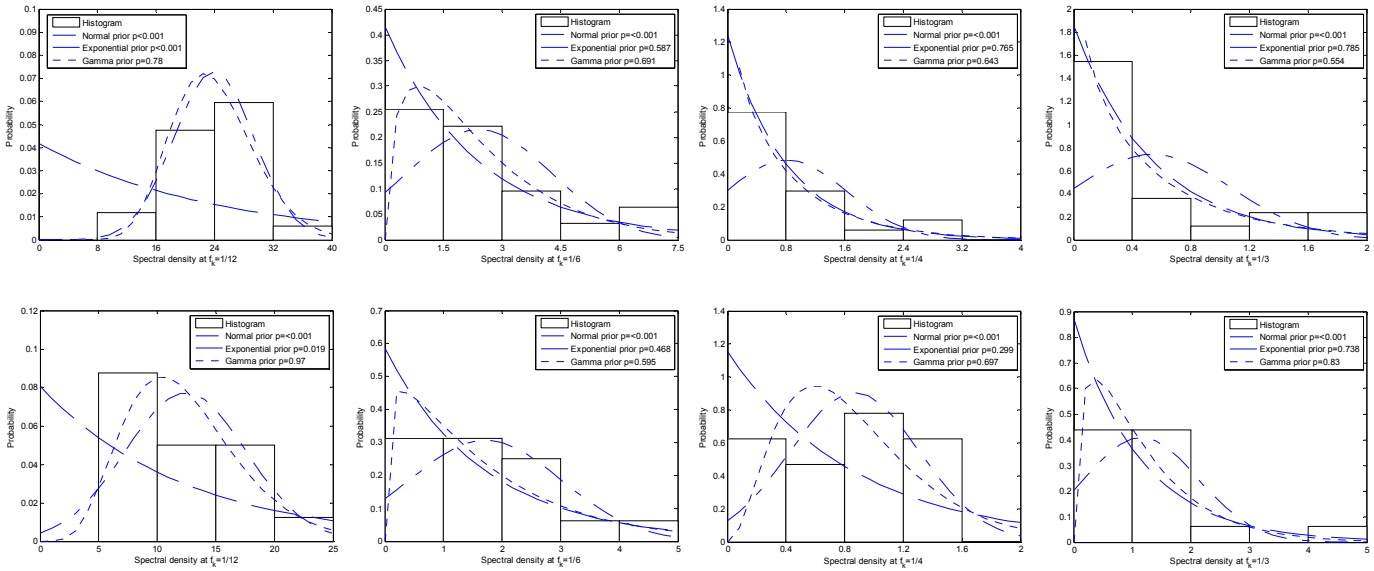


Figure 6-4 Posterior probability densities of spectral powers at frequencies 1/12, 1/6, 1/4 and 1/3 estimated with different priors for Ohio River and Buffalo River

When considering spectral powers of all frequencies as a vector $\vec{x}$, the final result of whether the multi-dimensional posterior probability density passes the Kolmogorov-Smirnov test was summarized in Table 6-4. Based on the test result, the exponential distribution and gamma distribution as priors passed the Kolmogorov-Smirnov test. However, the exponential distribution as a prior has an advantage over the gamma distribution in one less parameter and explicit derivation. In the case of streamflow, the exponential distribution has *n*-less parameters than the gamma distribution for multi-variate case. It is shown from the derivation of equations (6-20)-(6-24) that using the exponential distribution as a prior can be simply linked to the linear prediction. Thus, the exponential distribution was chosen for streamflow forecasting. It is noted that the result for WD, OUD, and LMD were left blank due to the lack of historical observations.

Table 6-4 Hypothesis test on choosing the prior probability density (1=reject, 0=accept)

Name	Uniform	Exponential	Normal	Gamma
UMU	1	0	1	0
MNU	1	1	1	0
MND	1	0	1	0
UMC	1	0	1	0
IU	1	0	1	0
ID	1	0	1	0
DMU	1	0	1	0
DMD	1	0	1	0
ILU	1	0	1	0
ILD	1	0	1	0
UMD	1	0	1	0
MS	1	0	1	0
OH	1	0	1	0
WU	0	0	1	0
WD	-	-	-	-

Table 6-4 Continued

Name	Uniform	Exponential	Normal	Gamma
BB	0	0	1	0
BU	0	0	1	0
OUU	0	0	1	0
OUD	-	-	-	-
LMD	-	-	-	-

6.5 Evaluation

6.5.1 Spectral density estimation

The spectral density for the forecasted series was estimated using RESAS and BESA and compared to FFT in Figure 6-5.

It was found that the spectral density estimated was not exactly the same as that estimated from historical data. For example, the one obtained from historical data of the Illinois River was found to be significant at f_k=1/12, 1/6 and 1/3, but here only peak at 1/12 was significant, even 1/6 was comparably small. They were not exactly the same, as the streamflow time series was influenced by other time associated factors, like climate, and may show different characteristics under different conditions.

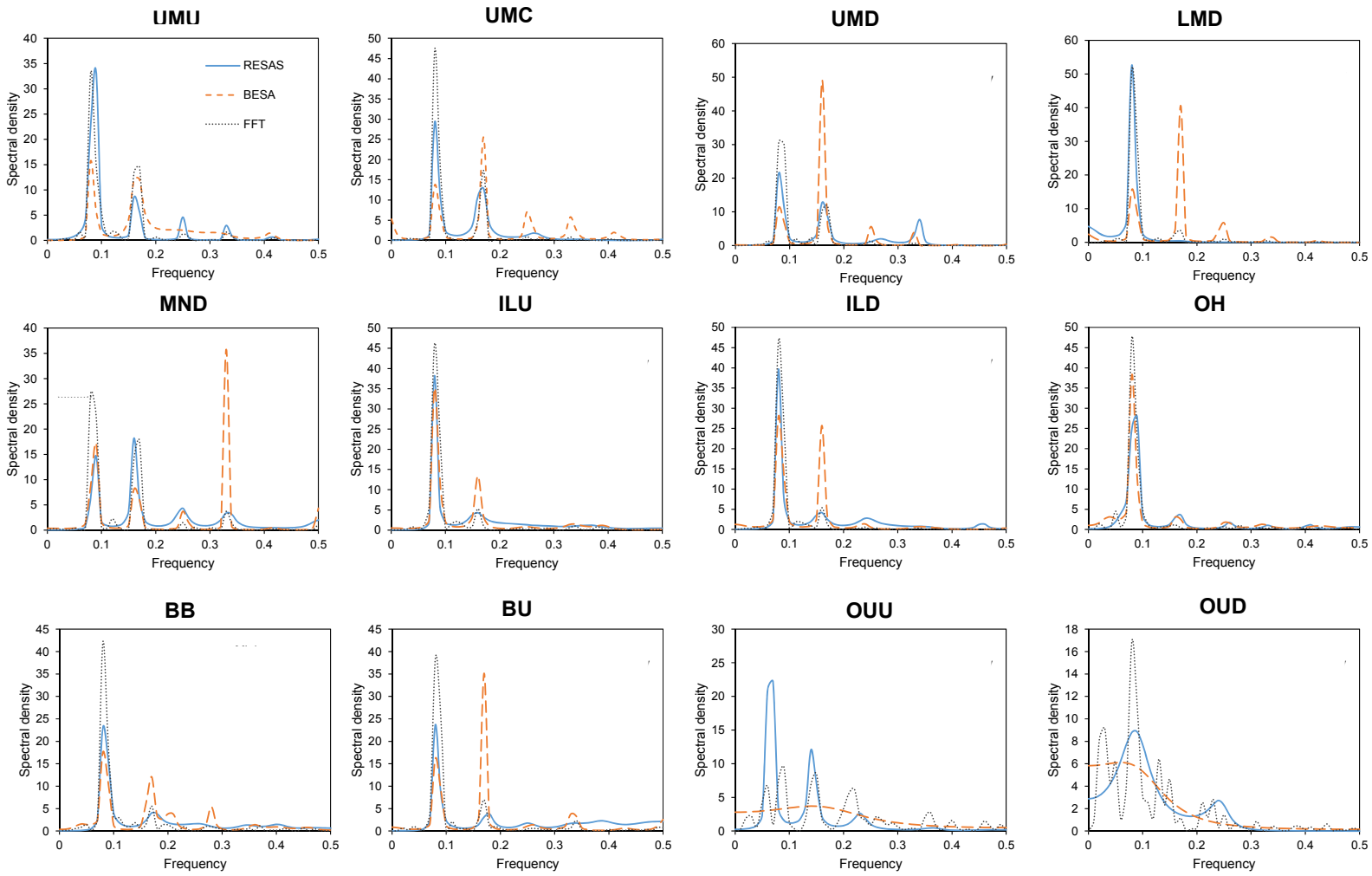


Figure 6-5 Estimated spectral densities by RESAS and BESA comparing to FFT

It can be seen from the results of Table 6-5 that the Itakura-Saito distortion values for RESAS are smaller for all cases than for BESA. The value of Itakura-Saito distortion by BESA was 1.3 to 75 times of that by RESAS. Thus, it suggested that RESAS theory provides higher resolution spectral density than BESA does. Furthermore, it was interesting to note that the Itakura-Saito distortion was larger for the upstream reach than for the downstream reach, which suggests that the difference in detecting the spectral density by two entropy methods for the upstream is larger than for the downstream. It agrees with the Kolomogorov-Smirnov test that shows the uniform distribution prior is more possibly accepted for the downstream reach than for the upstream reach.

Table 6-5 Itakura-Saito distortion obtained by RESAS and BESA theories

Name	RESAS	BESA
UMU	2.919	13.296
MNU	1.187	34.180
MND	1.606	9.475
UMC	0.942	13.356
IU	0.071	5.323
ID	2.510	10.754
DMU	1.995	25.579
DMD	2.904	8.570
ILU	0.505	20.933
ILD	1.356	33.325
UMD	1.917	49.393
MS	1.981	2.644
OH	0.235	2.783
WU	0.429	2.062
WD	0.631	563.501
BB	0.472	4.056
BU	1.250	33.181
OUU	0.651	16.724
OUD	0.873	1.958

Table 6-5 Continued

Name	RESAS	BESA
LMD	0.922	1.547

It was shown from the periodicity analysis that streamflow in the Mississippi River possesses strong yearly periodicity and the main frequency is f_k=1/12. It can be seen from the spectral density plot in Figure 6-5 that the RESAS theory did not miss the peak at f_k=1/12 but BESA sometimes did. For example, at UMC, UMD, LMD, MND, and BU, the largest peak determined by BESAS was at f_k=1/6 (1/3 for MND), which should stick to f_k=1/12. These stations were found in the downstream reach, suggesting that the RESAS theory was more reliable in detecting the peak location than BESA for estimating the spectral density for downstream reaches. For upstream stations, like UMU and ILU, both BESA and RESAS were both correct in the timing of spectral peaks. But the amplitude of the spectral density estimated by RESAS was closer to FFT than by BESA.

It was found that RESAS theory sometimes may miss small peaks when peaks are close to each other. For example, the spectral density of BU had an insignificant peak at f_k=1/3, but RESAS did not capture it at all. The spectral peaks for OUD from the largest to the smallest were at f_k=1/12, 1/24, 1/8, 1/16, 1/6, 1/4.5, but the spectral density obtained from RESAS combined small peaks and had two peaks at f_k =1/12 and ¼. Nevertheless, the RESAS spectral density more agreed with that observed from historical data and extracted and selected the most significant frequencies. On the contrary, for this case BESA did not capture any peaks at all.

The difference in the behavior of two entropy methods was due to the prior information imposed on the RESAS. For given limited autocorrelation series or not clear series, RESAS used a prior guess additional to the BESA. With the prior, the RESAS could select the important information from the given autocorrelation for detecting spectral peaks, but BESA could not. The prior information could be free from the limited information. Again, the RESAS was superior to BESA for imposing the prior information.

6.5.2 Streamflow forecasting result

The forecasting results for all stations are tabulated in Table 6-6. For all the stations, r^2 exceeded 0.7 and *NSE* was higher than 0.5. The *NSE* value was 0.737 for the UM and was 0.689 for the LM. This was due to different changes in the patterns of streamflow hydrograph, timing and amplitude of the peaks. Due to the stochastic streamflow behavior, though streamflow may not be forecasted exactly, all observations fell inside the 90% confidence intervals in Figure 6-6. As is shown in Figure 6-6, streamflow in the LM is less constant and regular than in the UM. Comparing the hydrograph between the UMC and the BB, it was found that the hydrograph pattern of streamflow varied from year to year in the BB, but remained similar in the UMC. The regular pattern of streamflow of the UM is clearly shown in the streamflow of Illinois River downstream (ILD) in Figure 6-6, which monotonically increases from September to March, then decreases. But the streamflow of the BB in the LM fluctuates during the low-flow season, though it still increases to peak in March. Besides, the forecasted errors plotted in Figure 6-7 showed more randomness in the UM than LM. Thus, the RESAS theory forecasts streamflow of the UM with higher reliability than that of the LM.

However, the UM is found to exhibit a bi-modal probability distribution, as evidenced by the observed changes in the timing and amplitude of flow peaks (Baldwin and Lall, 1999). Due to the stochastic nature of observations, no method is capable of forecasting the stochastic series exactly, and RESAS theory is no exception. For example, the peak timing at UMC was shifted from April to March in the last lead time in Figure 6-6. However, the RESAS forecast captured the first two peaks in April, but missed the last one that shifted to March. The streamflow of MND observed changes in the amplitude of peak flow. The relative error of the RESAS forecasted peaks and observed value were 15%, 1% and 11% for each lead time as shown in Figure 6-7, which was acceptable.

Table 6-6 Forecasting results of RESAS

Name	*RMSE* (m^3/s)	r^2	*NSE*
UMU	11.4	0.872	0.724
MNU	4.2	0.922	0.702
MND	31.6	0.891	0.671
UMC	108.2	0.962	0.816
IU	2.2	0.777	0.608
ID	26.6	0.964	0.829
DMU	3.1	0.883	0.674
DMD	32.4	0.913	0.700
ILU	21.0	0.938	0.788
ILD	38.5	0.942	0.783
UMD	353.1	0.963	0.814
MS	61.2	0.993	0.920
OH	1721.5	0.842	0.645
WU	1210.1	0.723	0.545
WD	3618.9	0.754	0.544
BB	45.7	0.822	0.624
BU	0.6	0.981	0.869
OUU	10.2	0.927	0.748
OUD	111.3	0.794	0.621
LMD	340.3	0.981	0.874

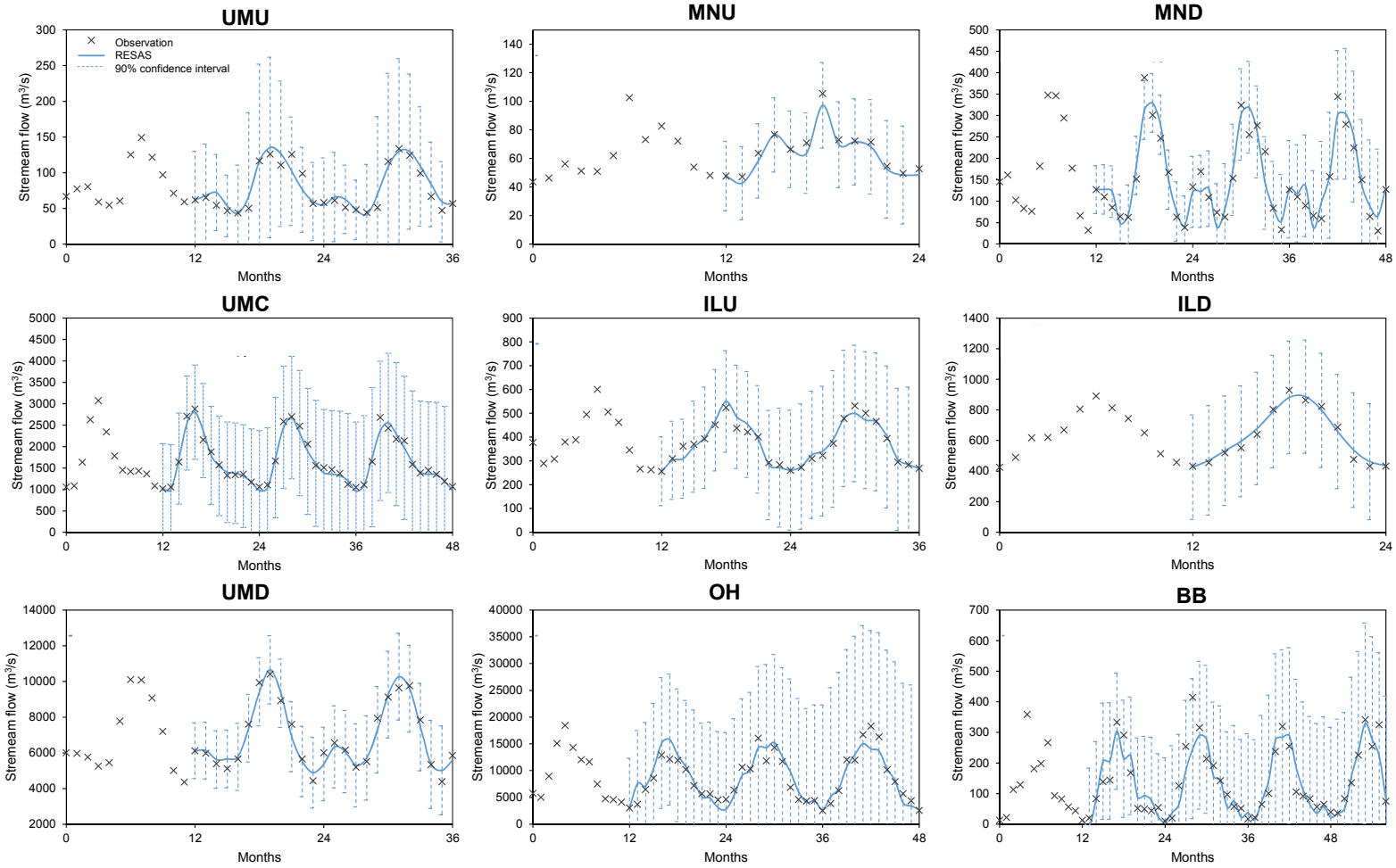


Figure 6-6 Forecasted streamflows using RESAS with 90% confidence interval

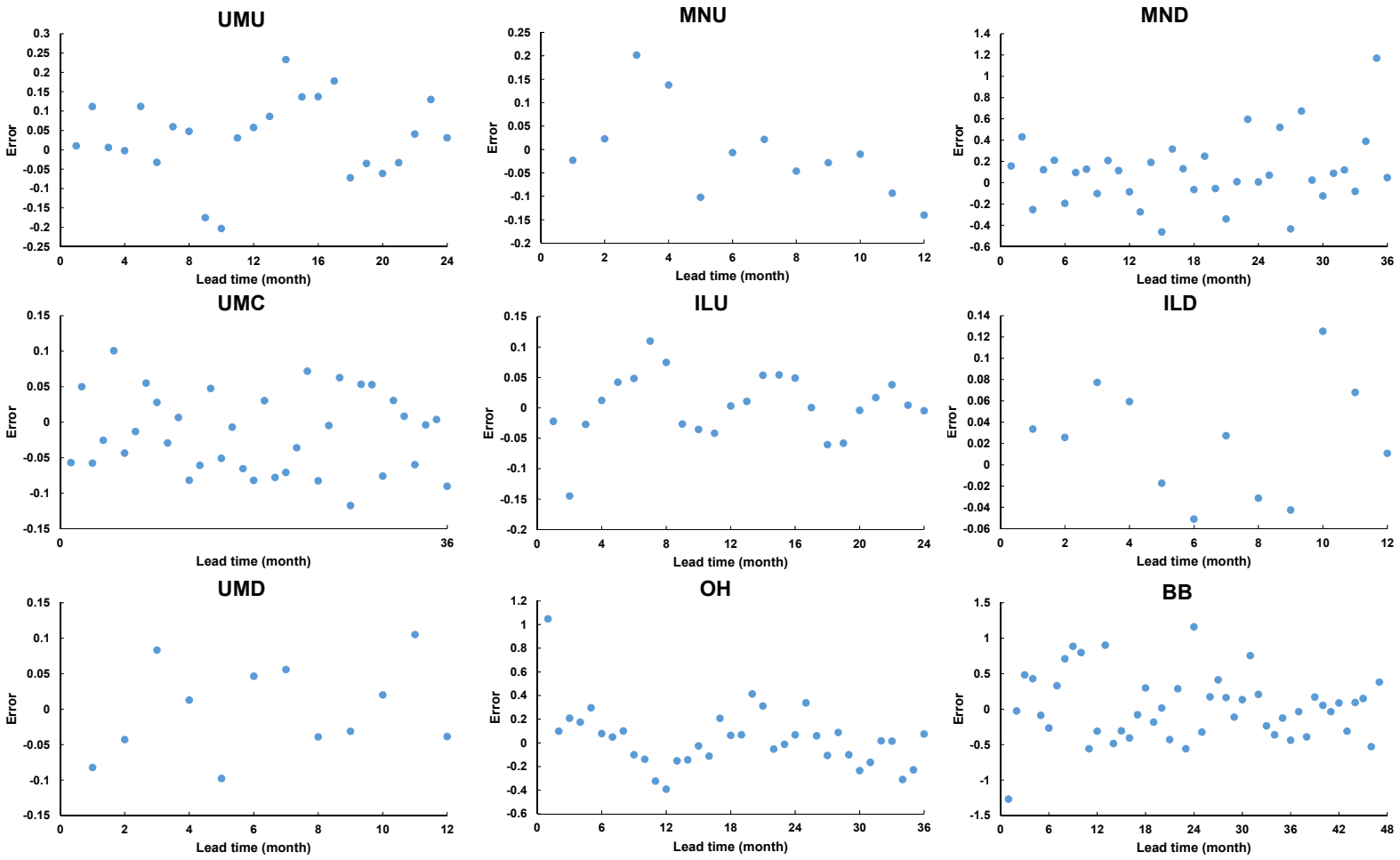


Figure 6-7 Forecasted errors by RESAS

6.6 Summary

This section developed the RESAS theory for monthly streamflow forecasting with spectral power as a random variable. The theory was applied to streamflow observed in the Mississippi River basin. Through analysis of the probability distribution of spectral power using historical series, the exponential distribution and gamma distribution prior fitted observations. However, using the exponential distribution prior had n-less parameters than using the gamma distribution. Furthermore, an explicit solution was derivable and the linear prediction was applicable when using the exponential distribution. Thus, the exponential distribution was chosen as a prior to apply the RESAS theory for forecasting streamflow.

Imposing the prior probability density on the spectral power, the RESAS theory had higher resolution in estimating the spectral density than BESA. The RESAS theory led to much smaller Itakura-Saito distortion than BESA, suggesting that the spectral density obtained by the RESAS theory was closer to FFT than BESA. However, the difference between the FFT and the two entropy methods is higher for streamflow at the upstream reach than at the downstream reach. The spectral density estimated by BESA sometimes shifted the peak, but the RESAS theory did not miss the largest peak at f_k=1/12, which is the main frequency of streamflow of the Mississippi River. The RESAS theory is also capable of estimating the spectral densities when peaks were close to each other while BESA cannot.

Applying RESAS theory, the Upper Mississippi streamflow was forecasted with higher accuracy than the Lower Mississippi streamflow. This was because the change of

hydrograph and timing and amplitudes of peaks were more constant than for the Lower Mississippi River.

7 MINIMUM RELATIVE ENTROPY SPECTRAL ANALYSIS WITH FREQUENCY AS A RANDOM VARIABLE

In the previous section on minimum relative entropy spectral analysis, there is a minor drawback that it suffers from restrictions on the nature of the process and dependence on the form of the assumed prior probability density function (Shore, 1981; Tzannes et al., 1985). Besides, an inadequate prior may lead to poor performance of RESAS, sometimes even worse than BESA without a prior assumption (Zhuang et al., 1993). The prior in Section 6 was selected by fitting 50-100 years of historical data, which may not always be available. When historic information is lacking, the selection of prior will be difficult. To overcome the restriction on the prior, Tzannes et al. (1985) developed a general method of minimum relative entropy spectral analysis with frequency as a random variable, thus abbreviated as RESAF. Later, it was found that RESAF is not restricted to the AR process; with different choices of prior, it can also be applied to a moving average (MA) or autoregressive and moving average (ARMA) series (Liefhebber and Boekee, 1987). When no specific prior is imposed, RESAF is equivalent to maximizing the configurational entropy spectral analysis (CESA) in Section 5. The RESAF theory is developed with frequency as a random variable in this section. The RESAF theory combines spectral analysis and cepstrum analysis with time series analysis for monthly streamflow forecasting.

7.1 Development of minimum relative entropy theory

Let monthly streamflow time series $y(t)$ be denoted as $y_1,\ldots, y_T$, where T is the total time period. Transferring to the frequency (f) domain, the information on streamflow is stored in the spectral density $q(f)$. Since the integration of spectral density over the limits of frequency equals 1, $q(f)$ can be treated as a probability density function of frequency (f) as a random variable. Thus, the development of minimum relative entropy theory contains the following steps: (1) define the relative entropy, (2) specify constraints, (3) derive minimum relative entropy-based spectral density, (4) compute the Lagrange multipliers, (5) extend the autocorrelation function, and (6) forecast streamflow.

7.1.1 Define the relative entropy

Let frequency f be the random variable, and the normalized spectral density be considered as its probability density function. The relative entropy, defined as in Tzannes et al. (1985), can be written as

$$H(q, p) = \int q(f) \ln[q(f) / p(f)] df \tag{7-1}$$

where $p(f)$ and $q(f)$ are normalized prior and posterior spectral density functions. The prior spectral density $p(f)$ is hypothesized from the observed periodicity of streamflow, and the least biased posterior spectral density $q(f)$ is estimated by minimizing equation (7-1). When no prior is given, $p(f)$ is taken as 1, and then the minimization of equation (7-1) is equivalent to maximizing the CE of $H(q)_{\max} = -\int q(f) \ln[q(f)] df$, the same as shown in equation (5-1).

7.1.2 Specification of constraints

Similar to MRES, the minimization of the relative entropy is again subjected to constraints defined by autocorrelation. However, in this case, the constraints are formed in a more straightforward manner. Using the Fourier transform relationship between the autocorrelation and the spectral density stated in Section 3.1.2, the constraints can be written as

$$\rho_r = \int_{-W}^{W} q(f) \exp(2\pi i r \Delta t f) df \quad , -N \leq r \leq N \tag{7-2}$$

where ρ_r is the autocorrelation of lag r, W is the Nyquist frequency, Δt is the sampling interval taken as 1 month (as monthly streamflow will be forecasted), and N is the largest lag where the autocorrelation can be correctly given from the observed time series data. It is noted that N is normally taken from 1/4 up to 1/2 of the length according to the periodicity of streamflow (Krstanovic and Singh, 1991b).

It is noted that when r=0, equation (7-2) reduces to

$$\int_{-W}^{W} q(f) = 1 \tag{7-3}$$

which satisfies the assumption of taking $q(f)$ as a probability density function of frequency f.

7.1.3 Estimation of spectral density

The least-biased estimate of the spectral density can be obtained by minimizing the relative entropy subject to equation (7-2). Using the Lagrange multipliers method, the Lagrangian function can be formulated as

$$L(f) = \int_{-W}^{W} q(f) \ln[q(f)/p(f)]df + \sum_{r=-N}^{N} \lambda_r [\int_{-W}^{W} q(f) \exp(i2\pi fr\Delta t)df - \rho_r] \qquad (7\text{-}4)$$

where λ_r, r=0, 1, 2, …, N, are the Lagrange multipliers. Taking the partial derivative of equation (7-4) with respect to $q(f)$ and equating the derivative to zero, one obtains:

$$\frac{\partial L(f)}{\partial q(f)} = 0 = \int_{-W}^{W} \{\ln[q(f)/p(f)] + 1 + \sum_{r=-N}^{N} \lambda_r \exp(i2\pi fr\Delta t)\}df \qquad (7\text{-}5)$$

Thus, by rearranging equation (7-5), the posterior distribution can formulated as

$$q(f) = p(f) \exp[-1 - \sum_{r=-N}^{N} \lambda_r e^{i2\pi fr\Delta t}] \qquad (7\text{-}6)$$

Equation (7-6) is the posterior spectral density obtained by minimizing the entropy based on prior information and given constraints. It is determined by solving the Lagrange multipliers and with the hypothesis on the prior. If the prior is white noise, the prior spectral density is constantly 1 for all frequencies, thus, the posterior density becomes

$$q(f) = \exp[-1 - \sum_{r=-N}^{N} \lambda_r e^{i2\pi fr\Delta t}] \qquad (7\text{-}7)$$

which is equivalent to the solution obtained by maximizing the CE.

7.1.4 Computation of Lagrange multipliers

The Lagrange multipliers can be solved for numerically by inputting equation (7-6) into equation (7-2), which yields

$$\rho_r = \int_{-W}^{W} p(f) \exp[-1 - \sum_{r=-m}^{m} \lambda_r e^{i2\pi fr\Delta t}] \exp(2\pi ir\Delta tf)df \text{ , } -N \le r \le N \qquad (7\text{-}8)$$

Another way to determine the Lagrange multipliers is by cepstrum analysis, like in CESA. Burr and Lytle (1986) and Wu (1983) showed how to combine MREF with

cepstrum analysis for estimating the explicit solution of Lagrange multipliers. Taking the inverse Fourier transformation of the log-magnitude of equation (7-6), one obtains

$$\int_{-W}^{W} \{1 + \log[\, q(f)] - \log[\, p(f)]\}\, e^{i2\pi fn\Delta t}\, df \;=\; \int_{-W}^{W} (-\sum_{n=-N}^{N} \lambda_n e^{i2\pi fn\Delta t}) e^{i2\pi fn\Delta t}\, df \qquad (7\text{-}9)$$

It can be seen from equation (7-9) that there are two terms relating to the spectral density that can turn to the cepstrum of autocorrelation, which is also called autocepstrum. Let the prior cepstrum of autocorrelation be denoted as $e_p(n)$, which is transferred from the prior spectral density as

$$e_p(n) = \int_{-W}^{W} \log p(f) e^{i2\pi fn\Delta t}\, df \qquad (7\text{-}10a)$$

and let $e_q(n)$ denote the posterior cepstrum of autocorrelation transferred from the posterior spectral density as

$$e_q(n) = \int_{-W}^{W} \log q(f) e^{i2\pi fn\Delta t}\, df \qquad (7\text{-}10b)$$

Then, doing the integration of both sides of equation (7-9), one gets

$$\delta_n + e_q(n) - e_p(n) = -\sum_{s=-N}^{N} \lambda_s \delta_{n-s} \qquad (7\text{-}11)$$

where δ_n is the delta function defined as:

$$\delta_n = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \qquad (7\text{-}12)$$

Equation (7-11) can be expanded as a set of N linear equations:

$$\begin{aligned}\lambda_0 &= -1 - e_q(0) + e_p(0)\\ \lambda_1 &= -e_q(1) + e_p(1)\\ &\vdots\\ \lambda_k &= -e_q(k) + e_p(k)\end{aligned} \tag{7-13}$$

Equation (7-13) enables to solve for the Lagrange multipliers in a more straight-forward manner than do nonlinear equation (7-8). Thus, the Lagrange multipliers can be estimated from the summation of two autocepstrums, the prior and posterior autocepstrums. The prior autocepstrum can be obtained from the observed periodicity of streamflow. When no prior is given, the cepstrum e_p equals 0 and diminishes, and the solution of equation (7-13) becomes the one derived using the CESA, equation (5-12). On the other hand, the posterior cepstrum can be obtained from the given lags of N autocorrelation as

$$e_q(n) = \int_{-W}^{W} \log q(f) e^{i2\pi fn\Delta t} df = \int_{-W}^{W} \log[\sum_{r=-N}^{N} \rho_r e^{-i2\pi ir\Delta tf}] e^{i2\pi fn\Delta t} df \tag{7-14}$$

However, based on the discussion of the relationship between autocepstrum and finite length of autocorrelation in Section 3.2.2, there is a recursive relation for autocepstrum that can be estimated from the autocorrelation defined as

$$e_q(n) = 2[\rho(n) - \sum_{k=1}^{n-1} \frac{k}{n} e_q(k)\rho(n-k)], \quad n > 0 \tag{7-15}$$

It is seen from equation (7-15) that the nth lag of cepstrum $e_q(n)$ is dependent on the previous n-1 lags of cepstrum and n-lag of autocorrelation. Thus, for given N lag autocorrelations, the cepstrum of autocorrelation can be computed up to lag N.

7.1.5 Extension of autocorrelation

The inverse relationship of equation (7-15) between autocorrelation and autocepstrum was stated in equation (3-37) as (Liefhebber and Boekee, 1987; Nadeu, 1992; Wu, 1983):

$$\rho_n = \frac{e(n)}{2} + \sum_{j=1}^{n-1} \frac{k}{n} e(j)\rho(n-j) \tag{7-16}$$

It is seen from equation (7-16) that the autocorrelation of *n*-th lag can be estimated from *n* cepstrums. Thus, for autocorrelations beyond lag *N*, it can be extended one by one using equation (7-16), in which manner ρ_{N+1} is extended from ρ_1, ρ_2, …, ρ_N using *N*+1 cepstrums and is used to extend to ρ_{N+2}, and so forth. Thus, the autocorrelation of *N*+*k* th lag can estimated from *N*+*k* cepstrums as

$$\rho_{N+k} = \frac{e_q(N+k)}{2} + \sum_{j=1}^{m} \frac{k}{N+k} e_q(j)\rho(N+k-j), \; k>0 \tag{7-17}$$

where *m* is the model order and $e_q(j)$ represents the posterior cepstrum obtained by MREF satisfying

$$e_q = \int_{-W}^{W} \log q(f) e^{i2\pi fj\Delta t} df = \int_{-W}^{W} \log\{p(f)\exp[-1-\sum_{r=-m}^{m} \lambda_r e^{i2\pi fr\Delta t}]\} e^{i2\pi fj\Delta t} df \tag{7-18}$$

It is noted from equation (7-18) that the posterior cepstrum is estimated from the prior and the Lagrange multipliers, which can be further equal to

$$e_q(j) = e_p(j) - \delta_{(j)} - \lambda_{(j)} \tag{7-19}$$

For different *j* values, the posterior cepstrum becomes

$$e_q = e_p(0) - 1 - \lambda_0, \text{ when } j=0 \tag{7-20a}$$

$$e_q = e_p(j) - \lambda_j, \text{ when } 1<j<N \tag{7-20b}$$

$$e'(j) = e_p(j), \text{ when } j>N \tag{7-20c}$$

since $\delta_j = 0$ when $j \neq 0$, and λ_j does not exist for $j>N$. Thus, replacing the cepstrum with equation (7-20), equation (7-17) can be written as

$$\rho_{N+k} = \frac{e_p(N+k)}{2} + \sum_{j=1}^{m} \frac{k}{N+k} e_q \rho(N+k-j) \tag{7-21}$$

where m is the order of model. Equation (7-20) is the extension of autocorrelation beyond the given lag N using RESAF. It is noted from equation (7-21) that when no prior is given, the first term diminishes and yields that derived from CESA.

7.1.6 Forecasting

Streamflow is forecasted in the manner that autocorrelation function is extended. It can be seen from equation (7-21) that autocorrelation of lag $N+k$ is extended from the previous m autocorrelations combined with m autocepstrums of autocorrelations. In the same manner, streamflow can be forecasted from previous m observations using m autocepstrums using equation (7-21) (Oppenheim and Schafer, 1975; Wu, 1983). Thus, to forecast streamflow, equation (7-21) can be written as a time series, where the input data changes to y_t and cepstrum $c(n)$ is used instead of $e(j)$:

$$y_{T+k} = \frac{c_p(T+k)}{2} + \sum_{j=1}^{m} \frac{k}{T+k} c_q(j) y(T+k-j) \tag{7-22}$$

where $c(j)$ is the cepstrum of the time series defined by

$$C(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left| \sum_{n=-\infty}^{\infty} y(t) e^{-2\pi nif} \right| e^{2\pi nif} df \tag{7-23}$$

It is noted that $e(n) = 2C(n)$ as shown in equation (3-28), and one can write equation (7-23) as

$$y_{T+k} = \frac{1}{4} e_p(N+k) + \frac{1}{2} \sum_{j=1}^{m} \frac{k}{T+k} e_q(j) y(T+k-j) \tag{7-24}$$

When no prior is given, e_p is 0, thus, equation (7-24) reduces to

$$y_{N+k} = \frac{1}{2} \sum_{j=1}^{m} \frac{k}{T+k} e(j) y(T+k-j) \tag{7-25}$$

Again, the order of forecasting model *m* is identified by the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) (Box and Jenkins, 1970; Hipel and McLeod, 1994).

The same 20 sets of streamflow data from the Mississippi River watershed used in the Section 6 were used to apply the MREF theory in this section.

7.2 Hypothesis on prior spectral density

To apply the RESAF theory, the prior hypothesis was constructed from periodicity discussed in Section 6.3. It was noted from the previous section that streamflow in the Mississippi River watershed has periodicities of 12 months, 6 months, 4 months, and 3 months with different degrees of significance. It is noted that the main periodicity of the Mississippi River is 12 month, and the periodicity is constant for the same river reach. The periodicity of small frequencies like 3 months or 4 months is less significant in streamflow for the downstream reach than for the upstream reach, and also less significant for the Lower Mississippi than for the Upper Mississippi. Thus, the prior spectral density was

assumed like a background noise for the six types of prior hypothesis, as stated in Table 7-1 and plotted in Figure 7-1.

Table 7-1 Hypothesis on the prior spectral density

No. of Prior	Prior spectral density
Prior 1 (CE)	p(f)=1
Prior 2	p(f)=0.01, p(1/12)=1
Prior 3	p(f)=0.01, p(1/12)=1, p(1/6)=0.6
Prior 4	p(f)=0.01, p(1/12)=1, p(1/6)=0.2, p(1/4)=0.2
Prior 5	p(f)=0.01, p(1/12)=1, p(1/6)=0.2, p(1/3)=0.2
Prior 6	p(f)=0.01, p(1/12)=1, p(1/6)=0.2, p(1/4)=0.2, p(1/3)=0.2

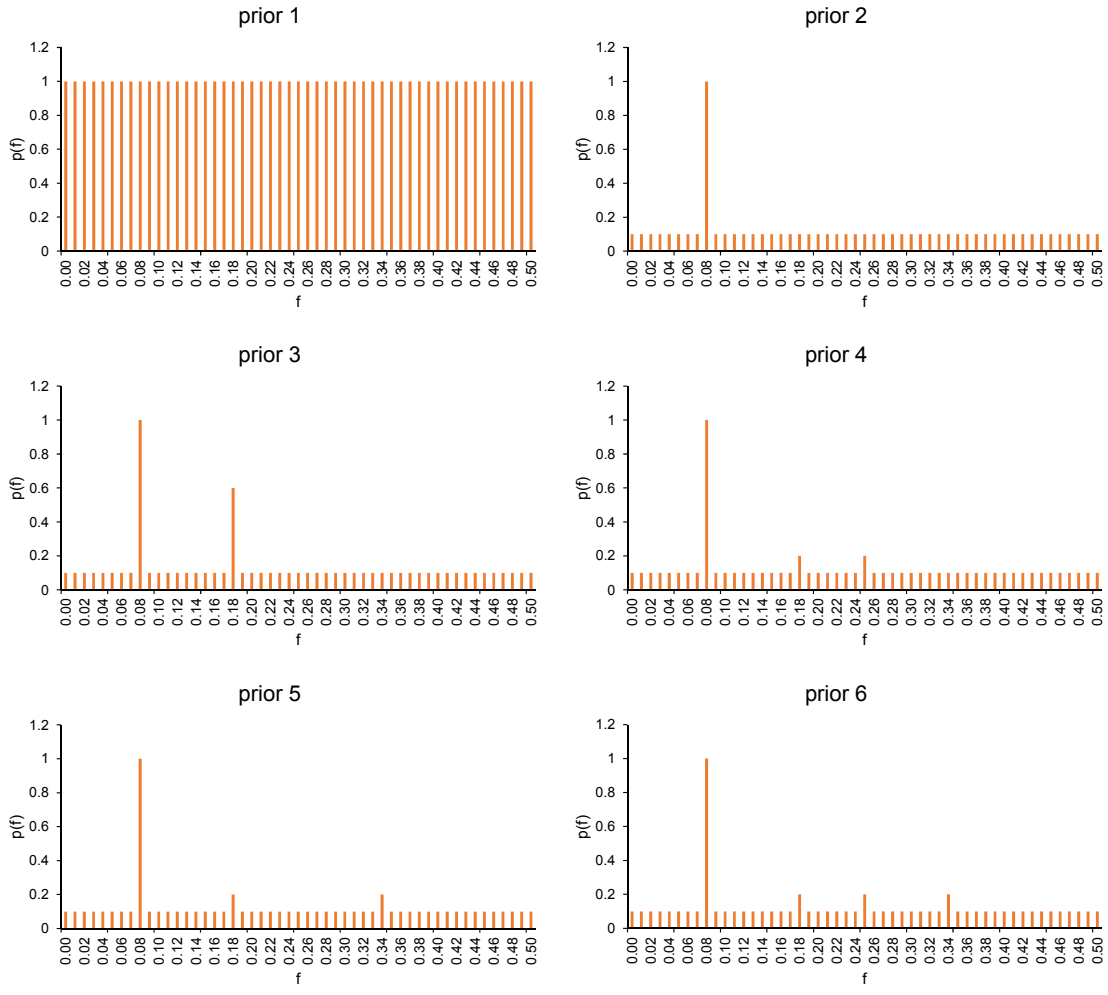


Figure 7-1 Six types of prior hypotheses

Prior 1 was defined by a condition when no prior information is given, thus, the spectral density was constantly 1. In this case, the RESAF was equivalent to the CESA theory. Priors 2-6 were constructed by combining the background noise $p(f)$=0.01 and the spectral peak at a specific frequency. Prior 2 was defined by the unique peak condition with assumed peak $p(f)$=1 at f=1/12, which may be applied at most of the stations in the Lower Mississippi. Prior 3 had two significant periodicities at frequencies 1/12 and 1/6,

however, the spectral peak at frequency 1/6 was assumed to be 0.6 not 1 with lower significance than at frequency 1/12. The streamflow observed in the main Mississippi River was prone to prior 3, prior 4 and prior 5 correspond to adding small spectral peaks at frequency of 1/3 or 1/4, which can be applied in the middle of the Mississippi River watershed. The last prior spectral density was composed of four peaks at all possible periodic frequencies, which was observed at Minnesota River (MN) and upstream Ouachita (OUU) River. The periodicity of streamflow is retabulated in Table 7-2, suggesting prior from the periodicity information.

Table 7-2 Periodicity of each station and suggested prior

Name	Location	Station	Periodicity	Suggesting prior
UMU	Upper Mississippi upstream	05227500	12, 6, (4, 3)	Prior 3
MNU	Minnesota upstream	05301000	12, 6, 4, 3	Prior 6
MND	Minnesota downstream	05330000	12, 6, 4, 3	Prior 6
UMC	Upper Mississippi center	05420500	12, 6	Prior 3
IU	Iowa upstream	05449500	12, 6, (4, 3)	Prior 3
ID	Iowa downstream	05465500	12, (6, 4)	Prior 2
DMU	Des Moines upstream	05476000	12, (6, 4, 3)	Prior 2
DMD	Des Moines downstream	05490500	12, 6	Prior 3
ILU	Illinois upstream	05543500	12, 6, 3	Prior 4
ILD	Illinois downstream	05568500	12, 6, 3	Prior 4
UMD	Upper Mississippi downstream	07010000	12, (6)	Prior 2
MS	Missouri	06934500	12, 6	Prior 3
OH	Ohio	03611500	12	Prior 2
WU	White upstream	07077000	12	Prior 2
WD	White downstream	07289000	12, 6	Prior 3
BB	Big Black	07290000	12, 6, 4	Prior 5
BU	Buffalo	07295000	12	Prior 2
OUU	Ouachita upstream	07359002	12, 6, 4, 3	Prior 6

Table 7-2 Continued

Name	Location	Station	Periodicity	Suggesting prior
OUD	Ouachita downstream	07367005	12	Prior 2
LMD	Lower Mississippi downstream	07374000	12	Prior 2

It is noted from equation (7-24) that the prior cepstrum is imposed as the background of streamflow fluctuation to maintain the periodicity assumed in the prior spectral density. Figure 7-2 plots the prior cepstrum corresponding to the six priors, which shows how spectral analysis impacts the streamflow time series. It is seen from the figure that the cepstrum fluctuated over time and the fluctuating frequency is different for each prior depending on which periodicity is assumed in the prior spectral density. When no prior information is given, the prior cepstrum constantly equals 0 and has no impact on streamflow forecasting.

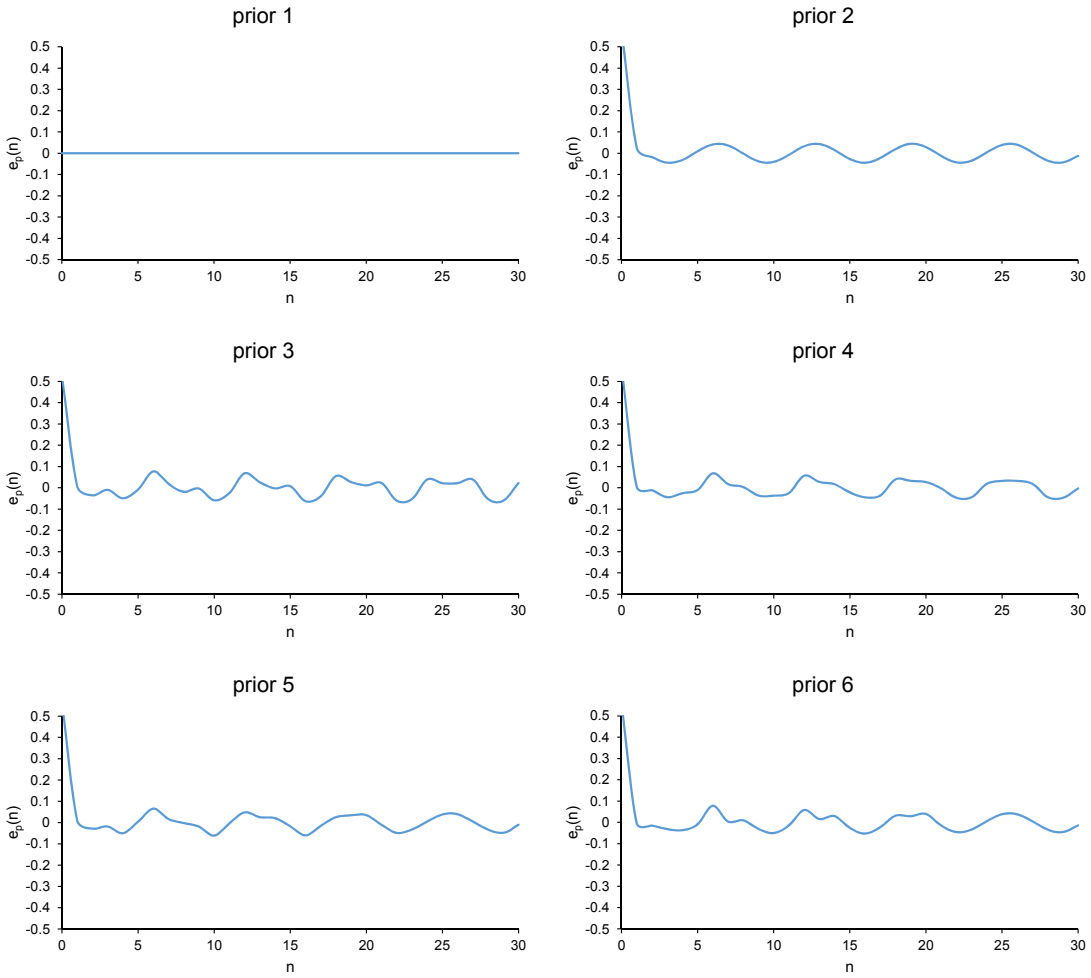


Figure 7-2 Prior cepstrum computed from six prior spectral densities

7.3 Evaluation

7.3.1 Spectral density estimation

For assumed six priors, the posterior spectral density was estimated by the RESAF theory. The estimated spectral density was verified with the one obtained from fast Fourier transform (FFT) through the Itakura-Saito (I-S) distortion, which is listed in Table 7-3. It was noted from the table that choosing priors 2-6 generally had higher resolution than

prior 1 except for the OUD and LMD. The I-S distortion estimated from prior 1 was 1.1 to 38.9 times that from prior 2, especially for White River downstream (WD) prior 1 became 180 times of prior 2. It suggested that prior information improves the resolution of spectral estimation.

Table 7-3 Computed Itakura-Saito distortion for each prior

Name	Prior 1	Prior 2	Prior 3	Prior 4	Prior 5	Prior 6	Choosing Prior
UMU	6.155	2.340	2.157	2.169	2.268	2.261	Prior 3
MNU	41.527	1.894	2.458	2.107	2.095	2.119	Prior 2
MND	3.688	1.409	1.288	1.171	1.091	1.194	Prior 5
UMC	6.243	1.276	0.916	0.918	0.930	0.977	Prior 3
IU	1.513	0.064	0.076	0.054	0.062	0.062	Prior 4
ID	5.504	1.997	2.275	2.039	2.150	2.180	Prior 2
DMU	19.464	2.218	2.060	1.911	1.856	1.908	Prior 5
DMD	8.956	2.454	2.632	2.247	2.468	2.350	Prior 4
ILU	26.459	0.692	1.018	0.945	0.925	1.080	Prior 2
ILD	27.722	1.065	1.368	1.302	1.295	1.450	Prior 2
UMD	16.128	2.862	2.801	2.825	2.928	2.946	Prior 3
MS	2.010	1.875	1.905	1.825	1.860	1.962	Prior 4
OH	2.799	0.274	0.614	0.520	0.534	0.665	Prior 2
WU	1.962	1.144	1.634	1.445	1.489	1.641	Prior 2
WD	158.595	0.881	1.165	1.041	1.025	1.225	Prior 2
BB	4.820	0.446	0.675	0.601	0.593	0.758	Prior 2
BU	26.210	1.481	1.659	1.569	1.559	1.705	Prior 2
OUU	39.996	1.612	2.396	2.135	2.150	2.419	Prior 2
OUD	1.793	1.839	2.505	2.261	2.311	2.525	Prior 1
LMD	0.896	1.439	1.615	1.559	1.569	1.687	Prior 1

Comparing the Itakura-Saito distortion of different priors, the one having the least distortion (prior*) was tabulated in the last column of Table 7-2. It was interesting to note that the prior with the least I-S distortion is not exactly the same as that suggested in Table

7-1. For example, the periodicity observed for the Minnesota River (MN) suggested prior 6 for four spectral peaks, while after verification, it reduced to prior 2 with uni-peak for the upstream reach and prior 5 with three peaks for the downstream reach. It implies that the small peaks observed at 3 month or 4 moth periodicity were not that significant for the MN. Besides, prior 6 was not suggested by any of the cases, which implied that none of the streamflow possessed equally strong periodicity for frequencies of 1/6, 1/4 and 1/3. On the contrary, for the Des Moines River upstream (DMU), where the only 12 month periodicity was observed to be significant, the spectral density estimated with prior 5 was suggested for additional peaks at 1/6 and 1/3 frequencies.

Though the prior selected from the observed periodicity may not possess the least I-S distortion, the difference between the values of I-S distortion for different priors was not significant. The variation of the I-S distortion was between 0.0005 and 0.10 from prior 2 to prior 6. Besides, the difference in the value of I-S distortion between prior 2 and prior* was not significant and varied from 0.01 to 0.36 as shown in Figure 7-3. Especially for the Lower Mississippi, prior 2 led to the least I-S distortion. It can also be seen from Figure 7-4 where the estimated spectral density with prior 1, prior 2 and prior* were plotted against the one from the FFT. For all the priors, the $1/12^{th}$ peak was clearly detected, though it is about 23% overestimated by prior 2 and 11.7% overestimated by prior*. The shape of the spectral peak by prior 2 and prior* were closer to the one from FFT than that by prior 1, as the peak by prior 1 is wider than observed. For example, the area under the $1/12^{th}$ peak of spectral density estimated for MND was 2 times the one estimated by FFT or other priors. Furthermore, it can be seen from the figure that the location of spectral

peak of the posterior spectral density was consistent with the one assumed from the prior. For the Upper Mississippi River (UMU, UMC, and UMD), spectral peak at frequencies 1/12 and 1/6 were estimated with prior 3, which were the hypothesized peaks by prior 3. For MND or MS, the additional peaks occurred at frequencies 1/3 or 1/4, where prior 5 or prior 4 was assumed. However, the estimated peaks at small frequencies sometimes were too large as compared with the one from FFT, and the use of prior* was doubtful. As shown in the figure, the spectral peak estimated was 2.8 times the observed peak at the $1/6^{th}$ frequency for UMD, more than 10 times the observed peak at the $1/4^{th}$ frequency for MS. Nevertheless, the prior* had estimated a slightly smaller I-S distortion, the estimate with prior 2 provided for a satisfactory resolution. Furthermore, when forecasting streamflow with the lack of periodicity information, prior 2 can be a proper choice as monthly streamflow possesses a strong 12 month periodicity.

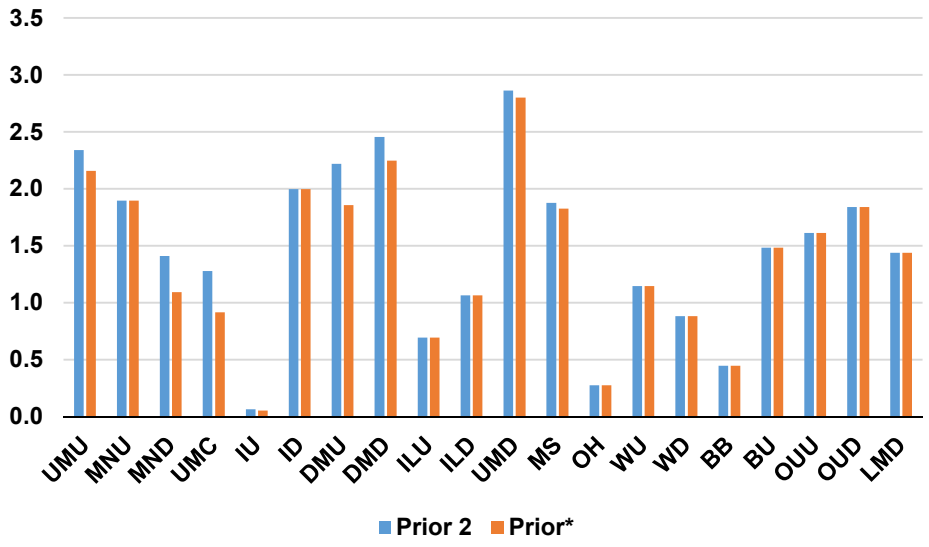


Figure 7-3 Comparison of the Itakura-Saito distortions for Prior 2 and Prior*

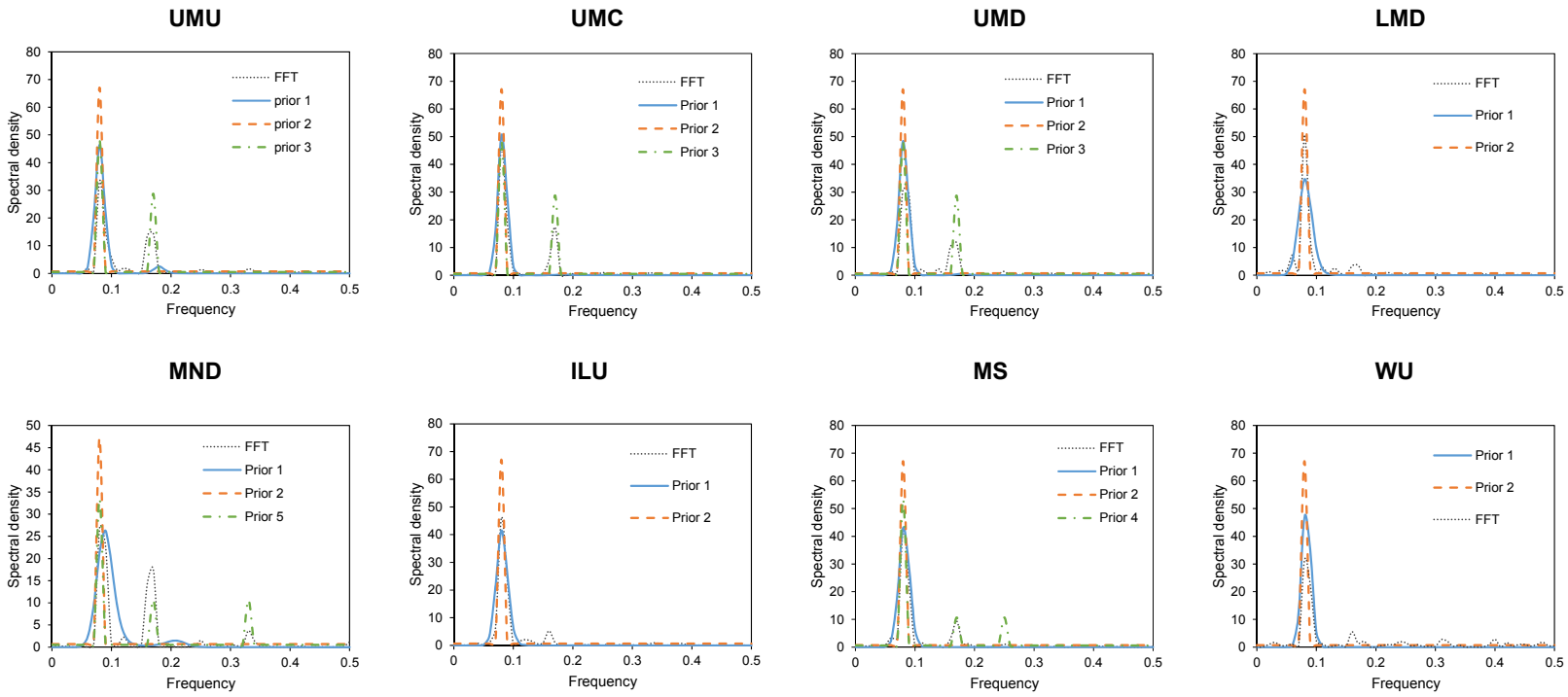


Figure 7-4 Estimated spectral densities with different priors

7.3.2 Streamflow forecasting

Streamflow was forecasted using prior 2 and prior*, the one with the least distortion, and were plotted in Figure 7-5 with forecasted errors plotted in Figure 7-6. As shown in the figure, streamflows forecasted with prior 2 and prior* were close to each other. For both priors, streamflow was forecasted properly, as the average values of *NSE* with the two priors were 0.77 and 0.78, respectively. Figure 7-7 plots values of *NSE* estimated with prior 2 and prior*, where it can be seen that the difference between the two priors was minimal. It was interesting to note that the I-S distortion for MNU was smaller for prior* than for prior 2, but *NSE* was higher for Prior 2. Besides, for Lower Mississippi, prior 2 was the only choice for forecasting. Thus, combining the discussion on the performance of spectral estimation by prior 2 and prior*, it suggested prior 2 be a proper assumption for estimating the spectral density and forecasting streamflow of the Mississippi River.

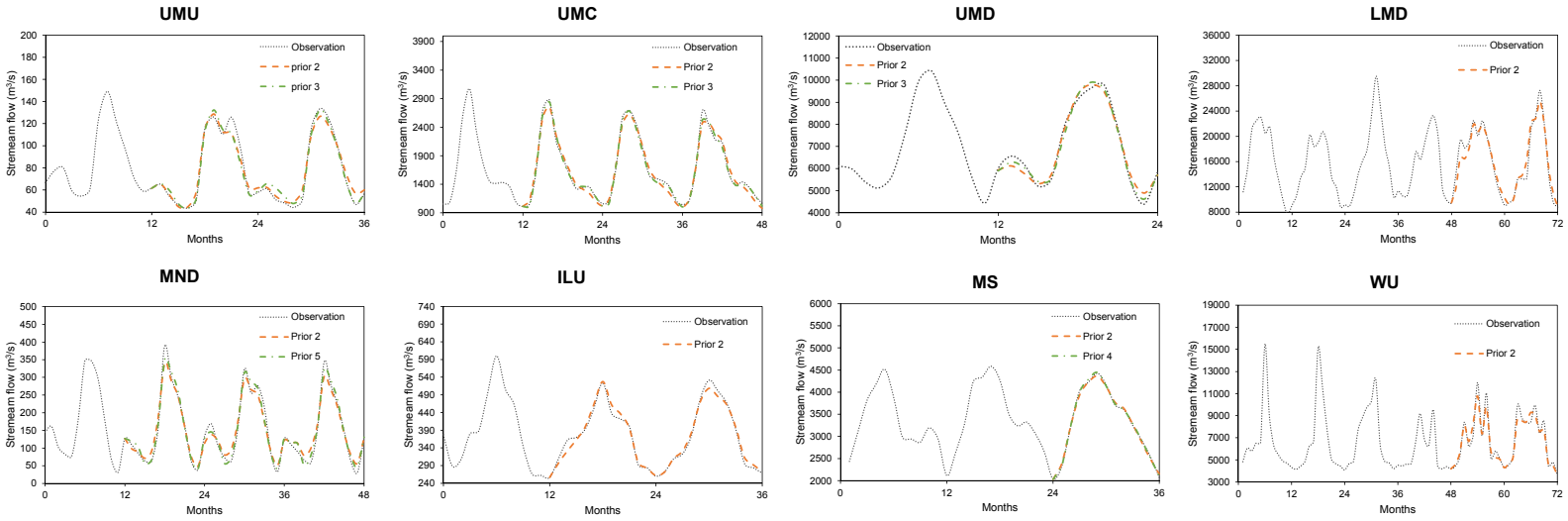


Figure 7-5 Forecasted streamflow with prior 2 and prior*

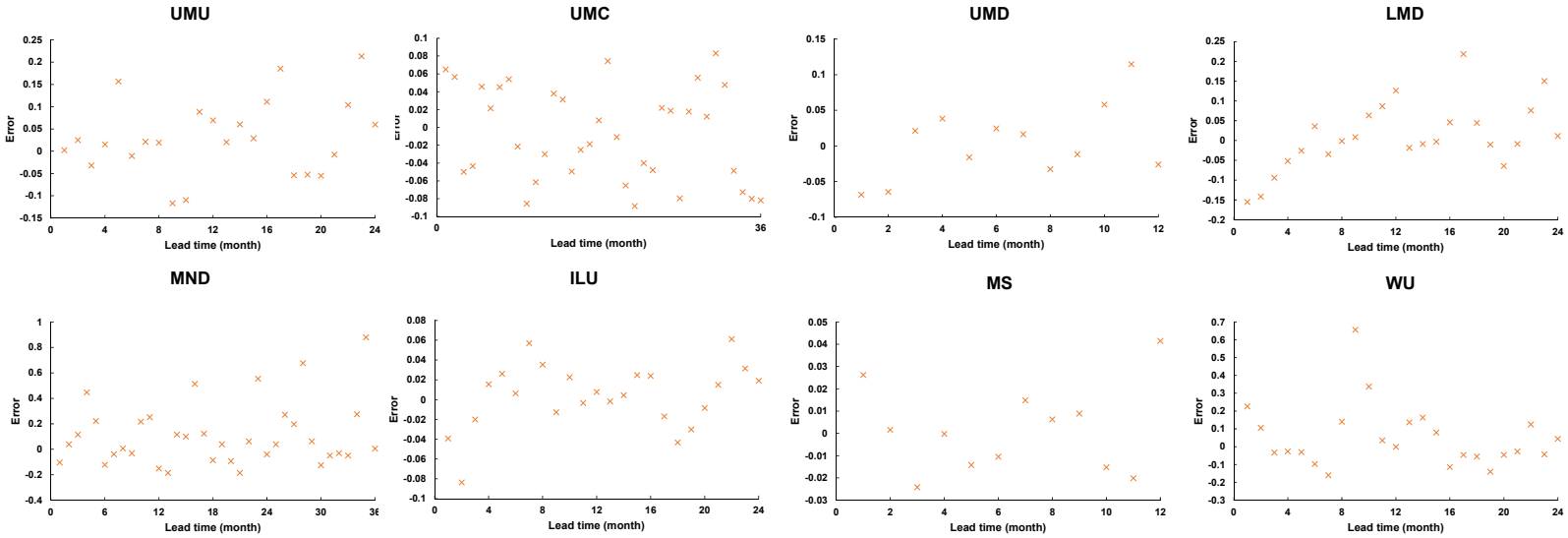


Figure 7-6 Forecasted errors by RESAF with prior 2

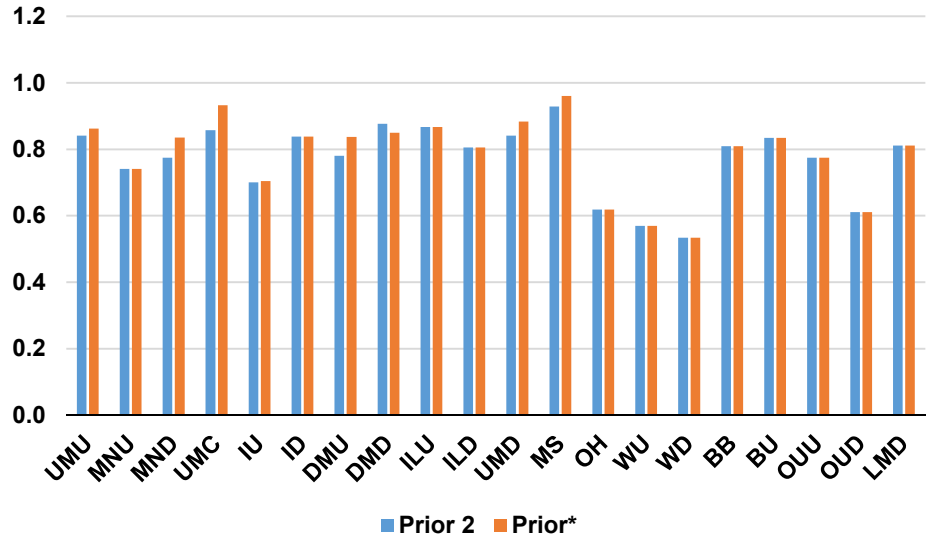


Figure 7-7 Comparison of NSE for prior 2 and prior*

The goodness of fit of forecasting results from prior 2 were evaluated by the root mean square error (*RMSE*), the coefficient of determination (r^2), the Nash-Sutcliffe efficiency coefficient (*NSE*), as tabulated in Table 7-4. It was noted that the RESAF theory was capable of forecasting streamflow with r^2 of 0.876; in particular it was 0.932 for the Upper Mississippi (UM) and 0.806 for the Lower Mississippi (LM). This shows that streamflow forecasting for the UM is more reliable than for the LM and is consistent with RESAS in Section 6, since streamflow hydrograph, timing and amplitude of the peaks were more regular and repeatable for the UM than LM. The plot of forecasted errors over lead time in Figure 7-6 also suggests that forecast for the UM was more consistent than for the LM, as errors were more randomly distributed for the UM. For example, it can be seen from streamflow of White River located in the LM in Figure 7-5 that the hygrograph, the timing of peak and peak volume are changing from year to year in a somewhat normal

fashion. For this case, the MREF was only able to provide the forecast with r^2 of 0.631 and *NSE* of 0.569 for a 2 year lead time, where the error in forecasted peak was about 0.7.

Table 7-4 Forecasted result of RESAF

Name	*RMSE* (m^2/s)	r^2	*NSE*
UMU	5.719	0.968	0.862
MNU	2.113	0.880	0.741
MND	15.816	0.973	0.835
UMC	46.866	0.934	0.932
IU	1.116	0.844	0.704
ID	13.316	0.891	0.838
DMU	1.531	0.971	0.837
DMD	16.215	0.978	0.850
ILU	10.488	0.946	0.867
ILD	16.236	0.891	0.805
UMD	209.338	0.986	0.884
MS	30.595	0.998	0.960
OH	860.748	0.660	0.618
WU	605.073	0.631	0.569
WD	1809.441	0.739	0.533
BB	22.956	0.855	0.809
BU	0.309	0.895	0.834
OUU	5.120	0.782	0.774
OUD	55.651	0.749	0.611
LMD	1179.229	0.950	0.811

7.4 Summary

The minimum relative entropy theory was developed by considering frequency as a random variable in this section. The development of RESAF theory entails the cepstrum analysis, as a result, an explicit solution is given for computing the Lagrange multipliers. The RESAF theory hypothesizes prior with information given from the periodicity of streamflow.

No matter which prior is chosen, even with no prior information, the MREF theory did not miss estimating the largest peak at frequency of 1/12, the main periodicity of the Mississippi River. However, with additional information as prior, the spectral density had higher resolution than that of the CESA theory. The difference in the estimation of spectral density with different priors was minimal, and prior 2 was applicable for monthly streamflow of any station on Mississippi River, as spectral peaks at other frequencies were not as significant as the one at frequency 1/12.

The MREF theory forecasted monthly streamflow from 20 stations in the Mississippi River watershed with a high value of *NSE* of 0.784. Similar to the MRES theory, streamflow forecasted for the Upper Mississippi River had a higher NSE than for the Lower Mississippi, but the difference was less than 0.1.

8 COMPARISON OF ENTROPY THEORIES FOR UNIVARIATE STREAMFLOW FORECASTING

Four entropy spectral analysis theories were developed using the configurational entropy and relative entropy for univariate streamflow time series forecasting. Both were developed using spectral power or frequency as a random variable. The configurational entropy spectral analysis (CESA) is obtained by maximizing entropy, and with Burg entropy spectral analysis (BESA) form the maximum entropy spectral analysis (MESA). The relative entropy spectral analysis is minimized, thus is called minimum relative entropy spectral analysis (MRESA) theory. However, the CESAS developed with spectral power as a random variable yielded the BESA, and was less convenient in estimating parameters than BESA. Thus, BESA is used to compare with other entropy based spectral analysis instead of CESAS, and CESA only represents the configurational spectral analysis developed with frequency as a random variable. This section compares the four entropy theory: BESA, CESA, RESAS and RESAF.

8.1 Entropy versus AR

8.1.1 Comparison between MESA and AR

First of all, the forecasted streamflows of Table 5-1 by two maximum entropy spectral analysis BESA and CESA were compared to the AR method, was shown in Figure 8-1.

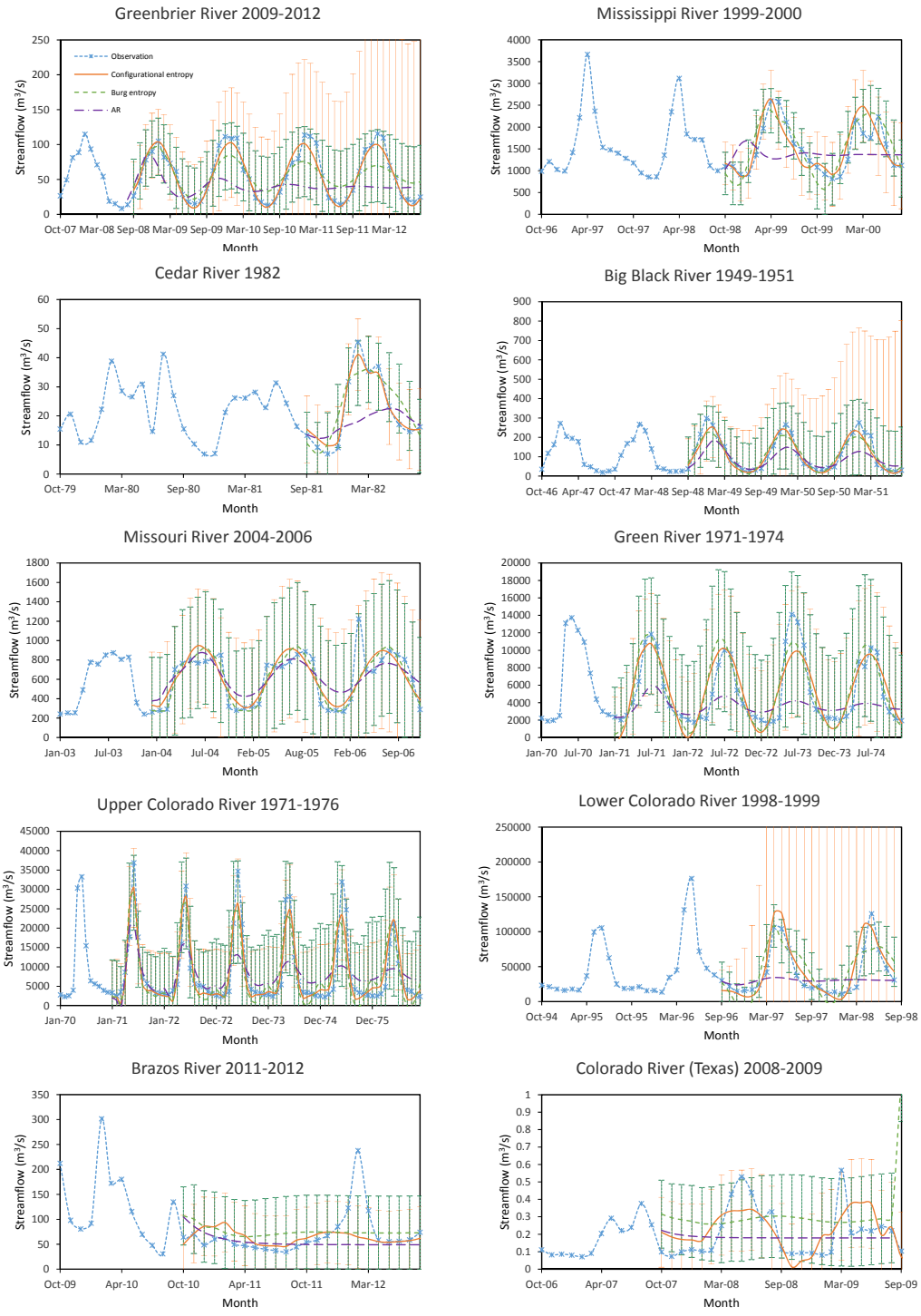


Figure 8-1 Comparison of two maximum entropy spectral analysis with the AR method

Using the Levinson algorithm, the AR method yielded a low order (3-5) of forecasting, while two entropy theories had an order of 10-13. The classical AR method was not comparable to the two MESA methods. Except for the forecasting for the Big Black River, Missouri River and Upper Colorado River, the forecasted streamflow using AR had r^2 lower than 0.5. The AR method requires longer past series so that it can provide the same accuracy as the entropy methods. For instance, both configurational entropy and the Burg entropy used 24 months streamflow series to forecast monthly streamflow of Greenbrier River up to 48 months lead time with r^2 higher than 0.8. The forecasted values by AR had r^2 lower than 0.7, and to increase the accuracy comparable to the entropy methods, past streamflow of at least 48 months needs to be given.

The other disadvantage of the AR method was its shorter lead time. The forecast of AR was not reliable over 48 months whereas two entropy methods can go as long as 60 months. The variation of *NSE* over the forecast lead time is shown in Figure 8-2. The *NSE* of AR dropped below 0.5 after the 1st year for Greenbrier River and after the 2nd year for others.

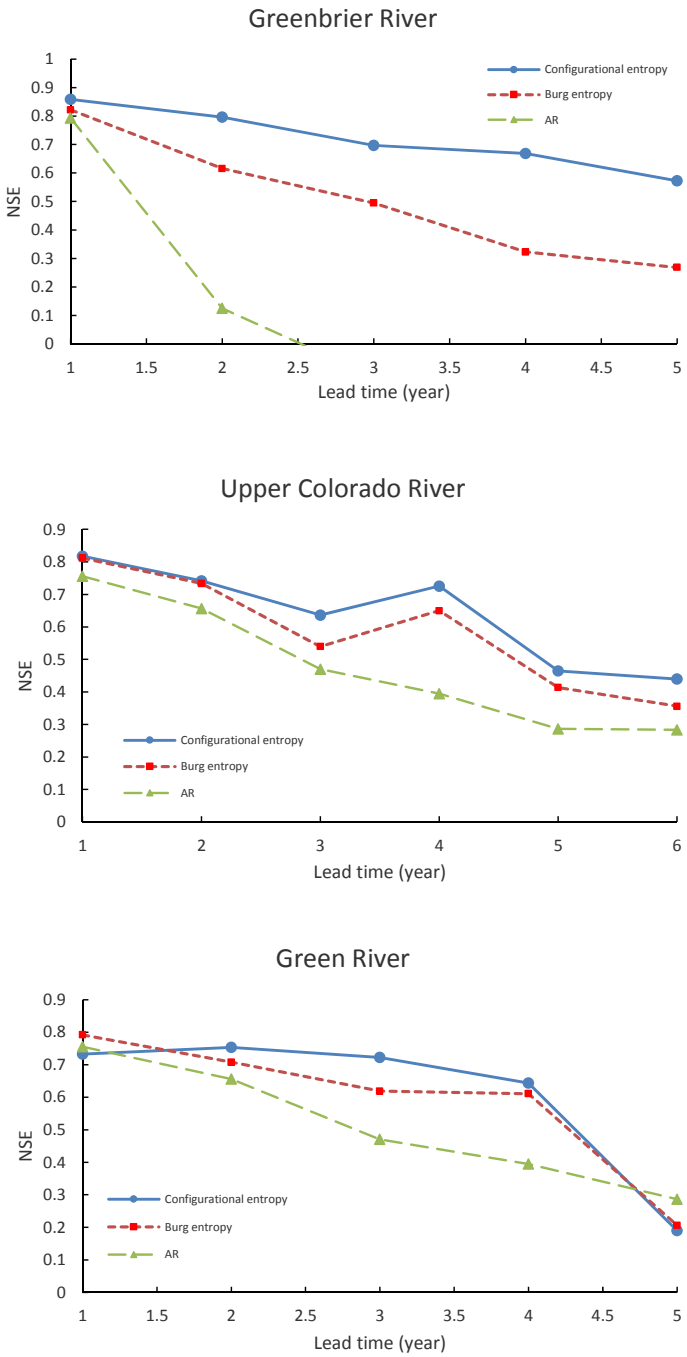


Figure 8-2 Nash-Sutcliffe coefficient against lead time for Greenbrier River, Upper Colorado River and Green River

8.1.2 Comparison of MRESA with AR

The orders of model used for forecasting by two MRESA (RESAS and RESAF) and AR are plotted in Figure 8-3. In almost all cases, the two entropy methods required higher order than did the AR method. The average order used by the MRES theory was around 13, by the MREF theory was around 14, but only 7 was for AR. Small order may cause poor prediction and incomplete peak fit in the spectrum. According to the *NSE* plot in Figure 8-4, the *NSE* values for all cases are higher for the entropy methods than for the AR method. Especially for WU, WD, OUU, OUD and LMD, where order of AR is less than 2, the NSE was less than 0.1. It showed that AR failed to capture the autocorrelation of these series and could not forecast monthly streamflow.

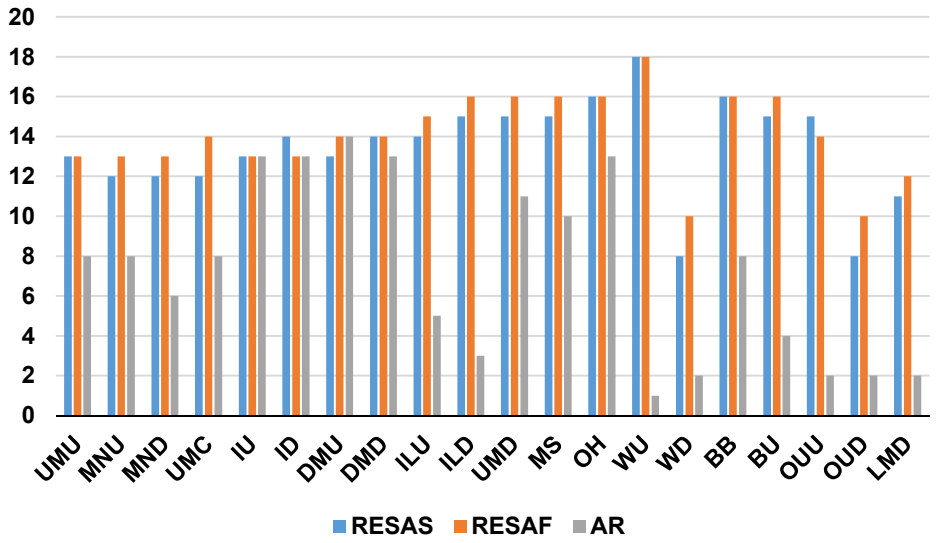


Figure 8-3 Model orders for RESAS, RESAF and AR

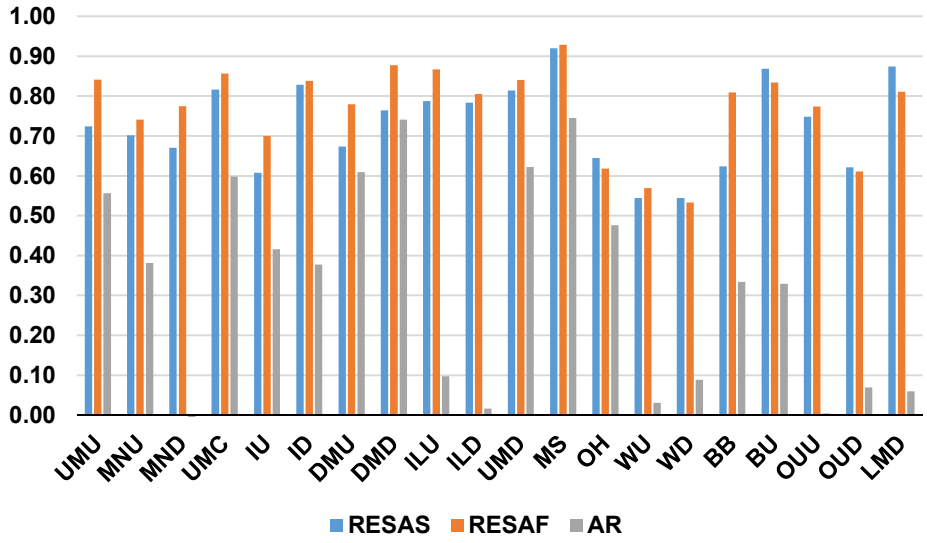


Figure 8-4 Comparison of *NSE* for RESAS, RESAF and AR

The poor performance of AR for these cases can be seen in Figure 8-5, where the AR forecast for LMD, MND, ILD, OUU and OUD roughly passed through the mean line, and did not fluctuate with observations. Thus, the relative errors, computed by $(Q_f - Q_o)/Q_o$ shown in Figure 8-6, had large absolute values and were dispersed far from 0 for the above stations. The range of relative errors estimated by AR was 3 to 4 times that by the minimum relative entropy spectral analsyis. It was a strong evidence that the AR method was not able to capture the autocorrelation of these streamflow series, thus, could not provide the fluctuating streamflow forecast.

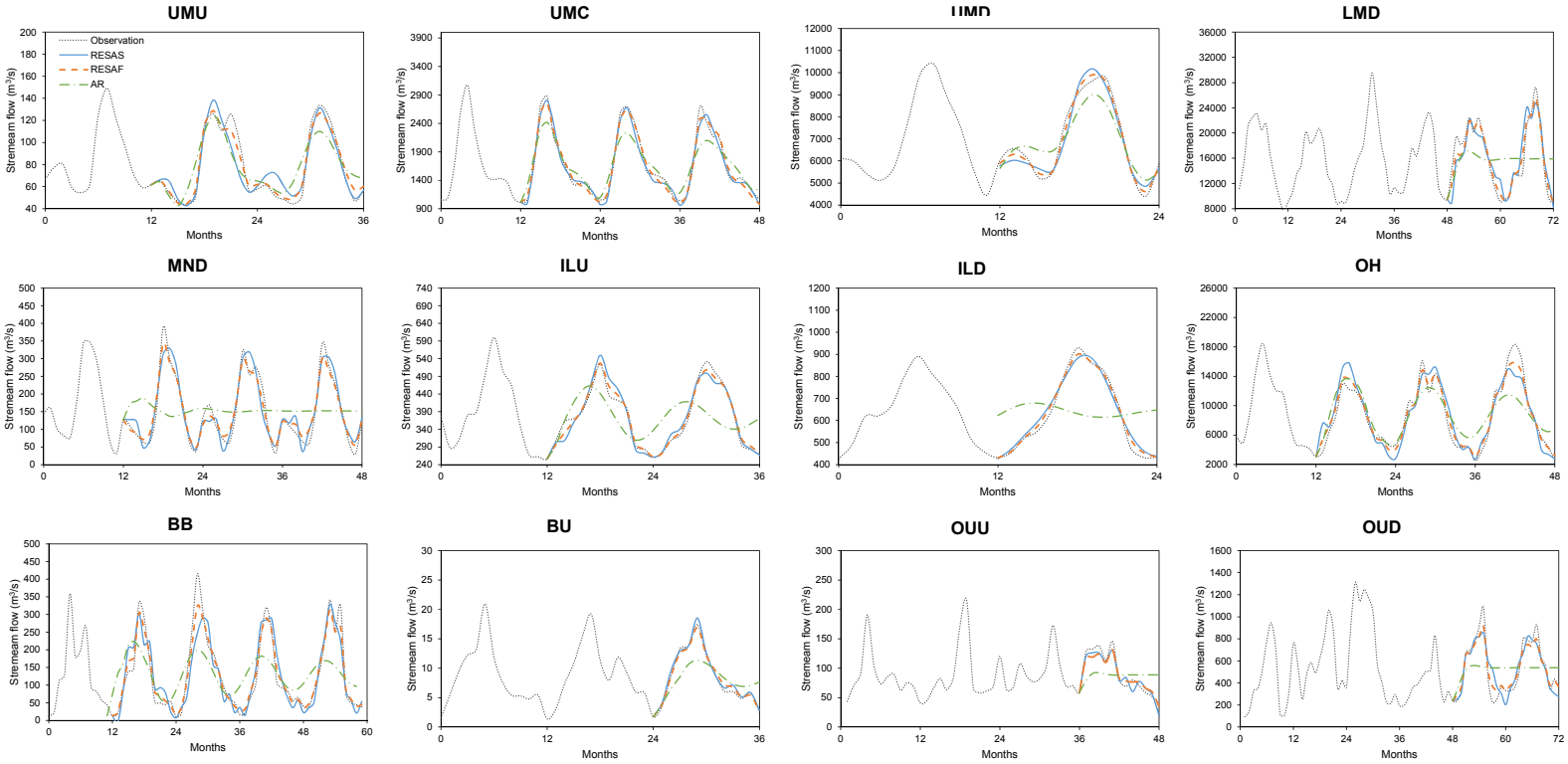


Figure 8-5 Streamflow forecasted by MRESA and AR

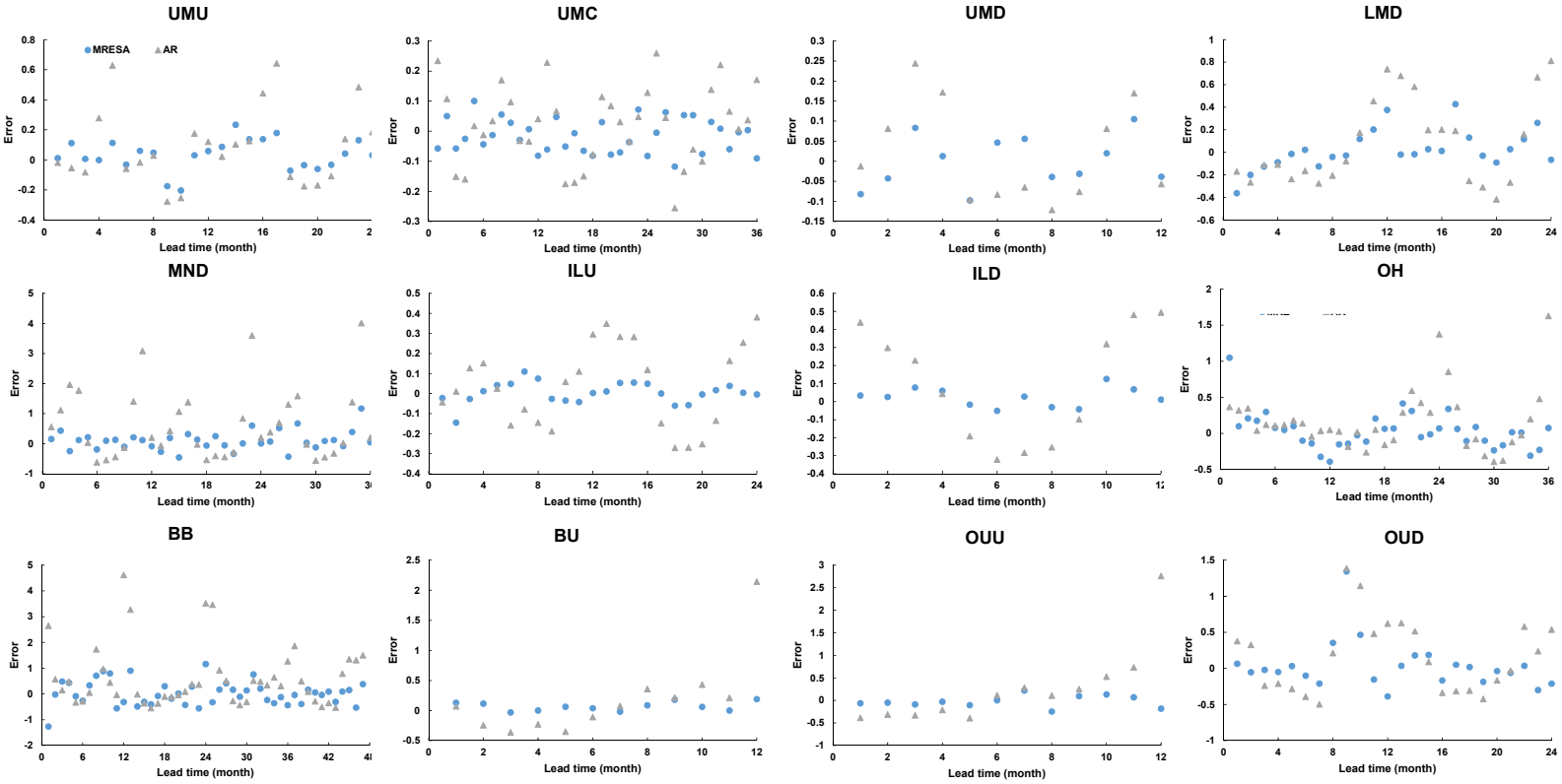


Figure 8-6 Relative errors against lead time by RESAS and AR

Though the average numbers of order used by the entropy methods are larger than AR, they enable the entropy methods to uncover longer autocorrelations of time series than would the AR method. The autocorrelation of the time series is the basis for forecasting univariate series from itself, thus the AR method may lose important information for forecasting, because the order represents how long the past data would be used to forecast the future. Thus, the autocorrelation, extended by either the maximum or minimum entropy, relied on longer past series than least squared extension, and forecasting was more reliable. Nevertheless, the higher order will make the computation of parameters more tedious, however, such tedium would be recommended for improving the accuracy.

Furthermore, the AR forecasting was weak to respond to sudden impulses of streamflow. For example, in the forecasting of streamflow at BB or BU, streamflow increased from December to January, paused in February then started to increase in March with a large slope. However, the AR method cannot forecast the change and fails to forecast the peak in March, while the entropy methods can. For other series, even though the forecast of AR fluctuated, it was still not comparable to that of the two entropy methods. For forecasted streamflow of UMU, UMC, UMD, OH, BB and BU, more than 20% of peak in March was lost by AR. It is noted that the relative errors forecasted for the March streamflow of the above stations periodically exceeded 0.65, 0.27, 0.25, 1.6, 4.8, and 2.2, respectively.

Furthermore, the AR forecast had a shorter lead time than the two entropy methods. For example, in the forecasting for UMC in Figure 8-5, the forecasts by the two entropy

methods were within 90% of the observed values, but AR dropped from 85% of the peak in the first lead year to 83% in the second lead year, and finally to 78% of the observed value in the third year. The average absolute relative error increased from 0.091 of the first lead year to 0.126 of the third lead year, while that by entropy methods stayed between 0.04 and 0.05 during the three-year lead time. To keep the error less than 0.2, the forecast by AR could not extend to longer than 24 months for all cases. In general, the AR method was not comparable to the two MRE-based forecasting methods.

8.2 Minimum relative entropy versus maximum entropy

The improvement in estimating spectral density with prior was discussed in Sections 6.5.1 and 7.3.1. This section compares the performance of forecasting streamflow by two MRESA theories and two MESA theories, and to see if imbedded prior information also improves time series forecasting.

8.2.1 Comparison between RESAS and BESA

As shown in Figure 8-7, the *NSE* values of RESAS forecasts were almost all higher than those of BESA, though for UMU, UMC, DMD, UMD and BB, the two entropy methods had similar results. It can be seen from Figure 8-8 for UMU and UMC that though the errors in peak forecasted using two entropy methods were both under 2.4%, the smallest values forecasted were much different. The mean value from September to February at UMU was about 52 m^3/s, while the mean value during this time forecasted by RESAS was 57 m^3/s and by BESA was 63 m^3/s. The BESA forecast exceeded the observations by 21%. Furthermore, it is seen from Figure 8-9 that the relative error forecasted by BESA was farther for this period, and the average of forecasted error for

this period by RESAS was 0.103 and by BE it was 0.173. In the previous section, it was shown that the two entropy methods had a longer lead time than AR. Here, RESAS showed even longer lead time than BESA. As shown in Figure 8-8 for BB, as time went far, the lower value forecasted by BESA was larger and the peak value forecasted by BESA was smaller. As a result, the peak forecasted decreased by 0.9% each year and the error in peak increased by 53% afterwards. Besides, the relative error exceeded 3 after forecasting 12 month ahead, and the average of relative error was larger than 0.5 after third lead year. On the contrary, the forecasting of RESAS was consistently good for up to 48 months lead time with average of relative error around 0.3-0.4. The peak in error was kept within 3.6% even for the fourth year's peak, and the forecast during low flow season still maintained a 93% accuracy.

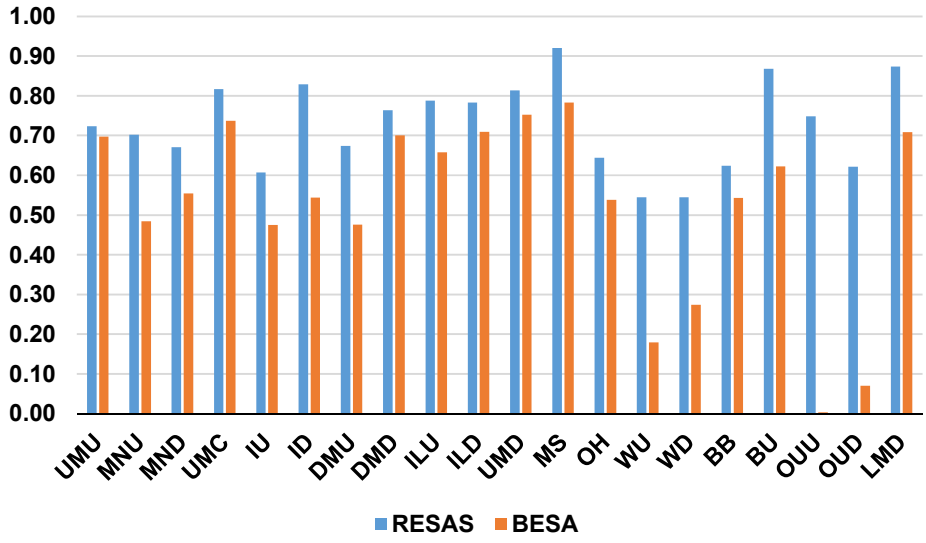


Figure 8-7 Forecasted NSE for RESAS and BESA

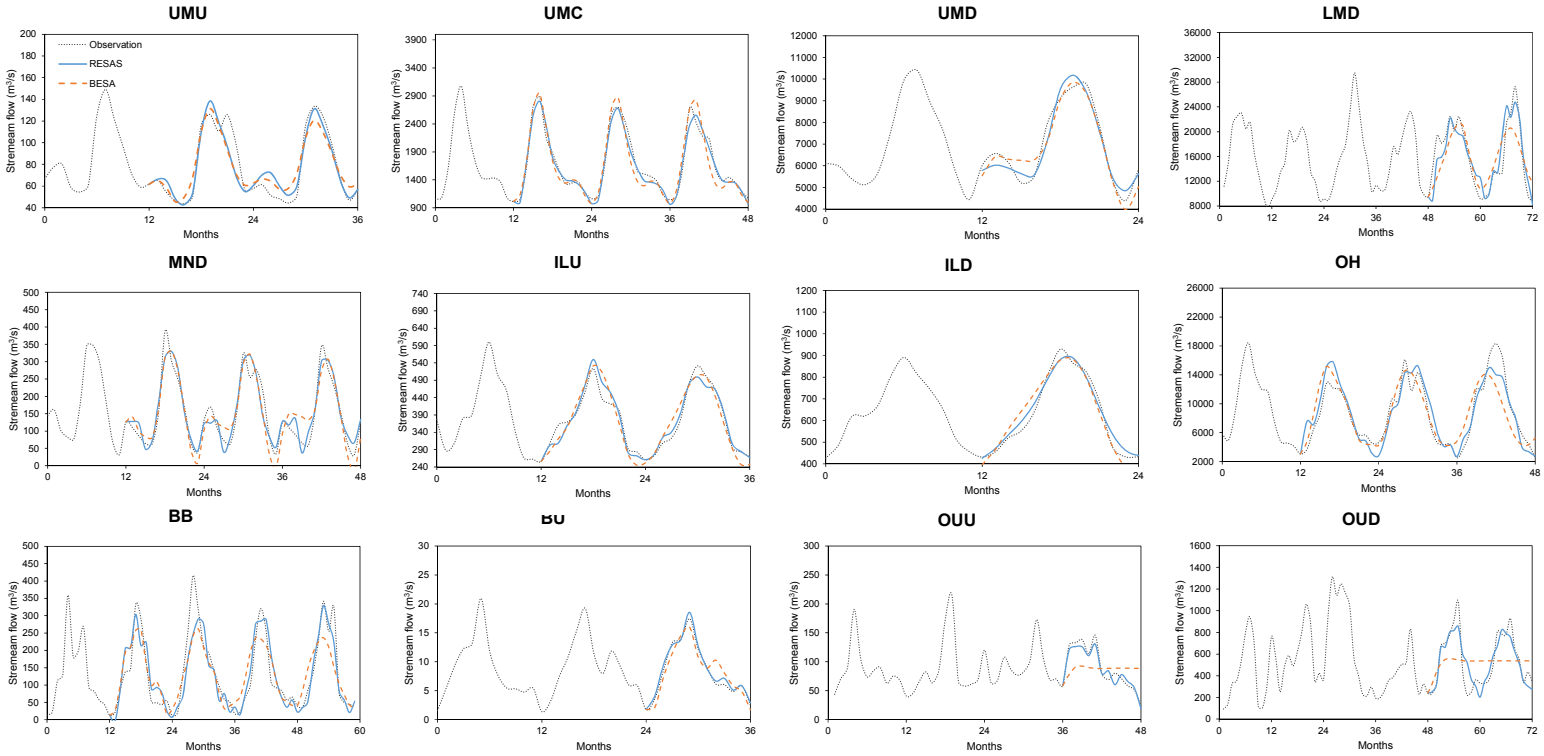


Figure 8-8 Forecasted streamflow by RESAS and BESA

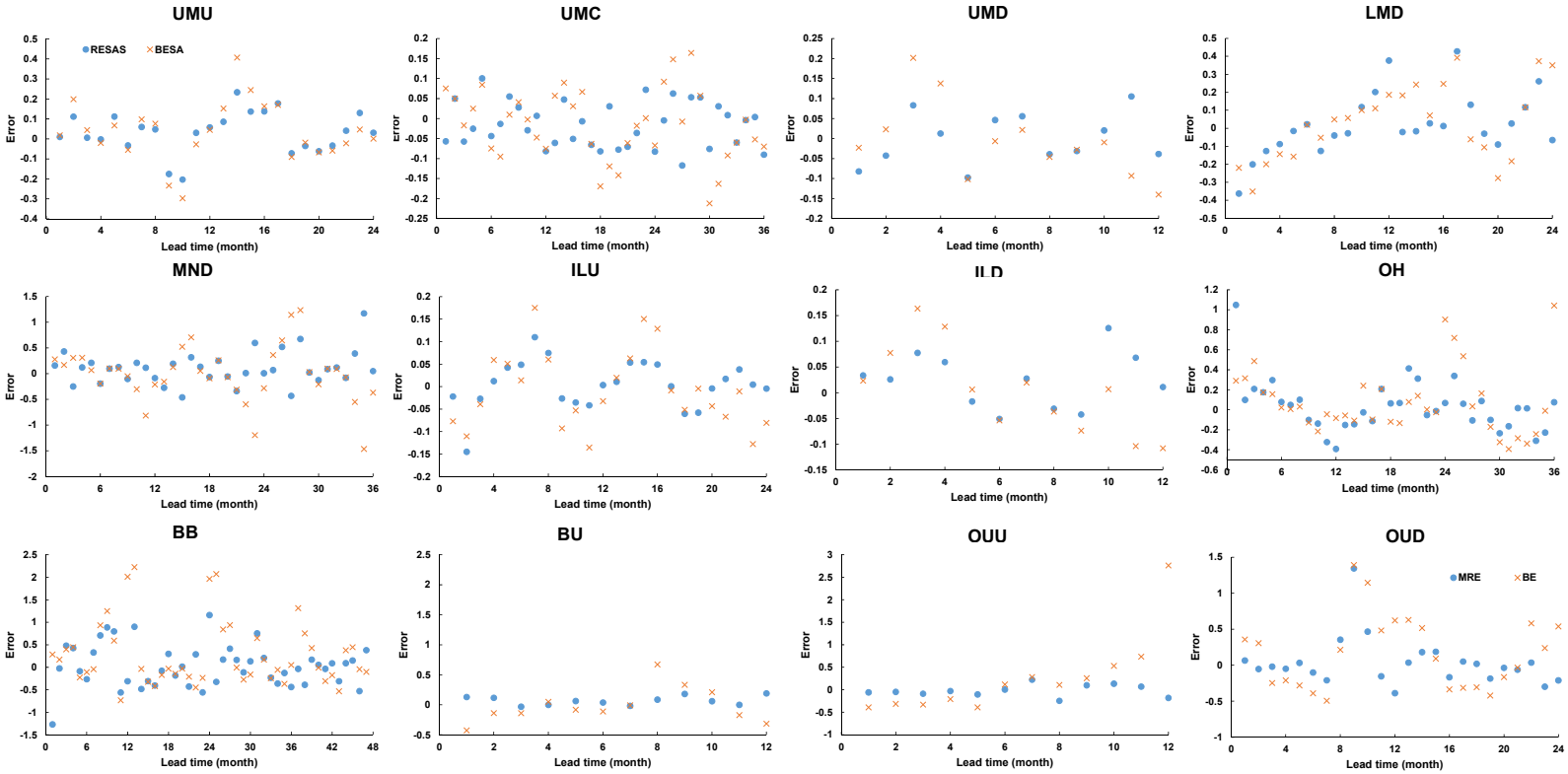


Figure 8-9 Forecasted errors by RESAS and BESA

However, the advantage of using RESAS over BESA was more significant in forecasting streamflow located downstream, where the determination of the locations of spectral peaks was more accurate. The *NSE* value of RESAS was about 1.2 to 1.5 times of that of BESA for ID, WD and OUD. While for MND, where BESA estimated the largest peak at f_k=1/3, the forecasted streamflow from July to February did not follow the observed periodicity pattern. It was interesting to find that during the test on the prior distribution, the probability of downstream spectral power was more prone to accepting the uniform prior than upstream. But when used for spectral estimation or streamflow forecasting, the advantage of choosing the exponential distribution over the uniform distribution as the prior was significant.

The streamflow of Ouachita possessed the least regular pattern and the least amount of information compared to other streamflow in the Mississippi River. BESA even failed to determine its spectral density as shown in Figure 6-5. Here, in Figure 8-8, the forecasting by BESA was as unsatisfactory as the one forecasted by AR. It failed to fluctuate with the observation but remained around the mean value, suggesting that the BESA method is limited to data with periodicity clearly given. It can be seen from the plotted errors that the errors from BESA coincided with those from AR and were all larger than RESAS. For limited data series or data with unclear periodicity, RESAS used prior information, additional to BESA, to clarify the periodicity, and further improved streamflow forecasting. In such cases, the advantage of imposing the prior distribution is significant. As a result, for forecasting streamflow of Ouachita River, the RESAS theory had an *NSE* of 0.75 and 0.62 for the upstream reach and the downstream reach respectively.

The cases where the BESA forecast was comparable to RESASor had *NSE* larger than 0.5 (for example UMU, ID, DMD or MS), had more possibility to be forecasted by the AR method than by others. This phenomenon showed an agreement with the result found by Liefhbber and Boekee (1987) that BESA was somehow limited to the AR series. It was reasonable, since the forecasting by BESA, though under entropy maximization, turned out to be a linear forecasting as the AR method. However, the RESAS theory was more general and not limited to the AR process. It makes the MRE more widely applicable than BE.

8.2.2 Comparison between RESAF and CESA

Streamflow forecasted by CESA theory and RESAF theory were plotted in Figure 8-10. Without any prior information, streamflow forecasted by CESA was smoother than with prior 2, and the shape of hydrograph kept the same from year to year. Thus, when observed streamflow is regular, repeatable, streamflow forecasted by CESA were as good as with other priors (see UMC, UMD and MS). However, streamflow forecasting by CESA was weak for an irregular change in hydrograph or a multi-peak flow. It can be seen from streamflow forecasted for LMD that CESA theory forecasted streamflow monotonically rising from October to May and then decreasing, while observed streamflow had fluctuating peaks. For UMU and MND Rivers, streamflow did not drop directly after the peak in March, however, there was another small rise in May, which was not forecasted using CESA. For the above cases, RESAF is recommended. It can be seen that RESAF with prior 2 was capable to forecast multi-peaks that fluctuated during the peak season of LMD and WU, and to forecast the small rise for UMU and MND.

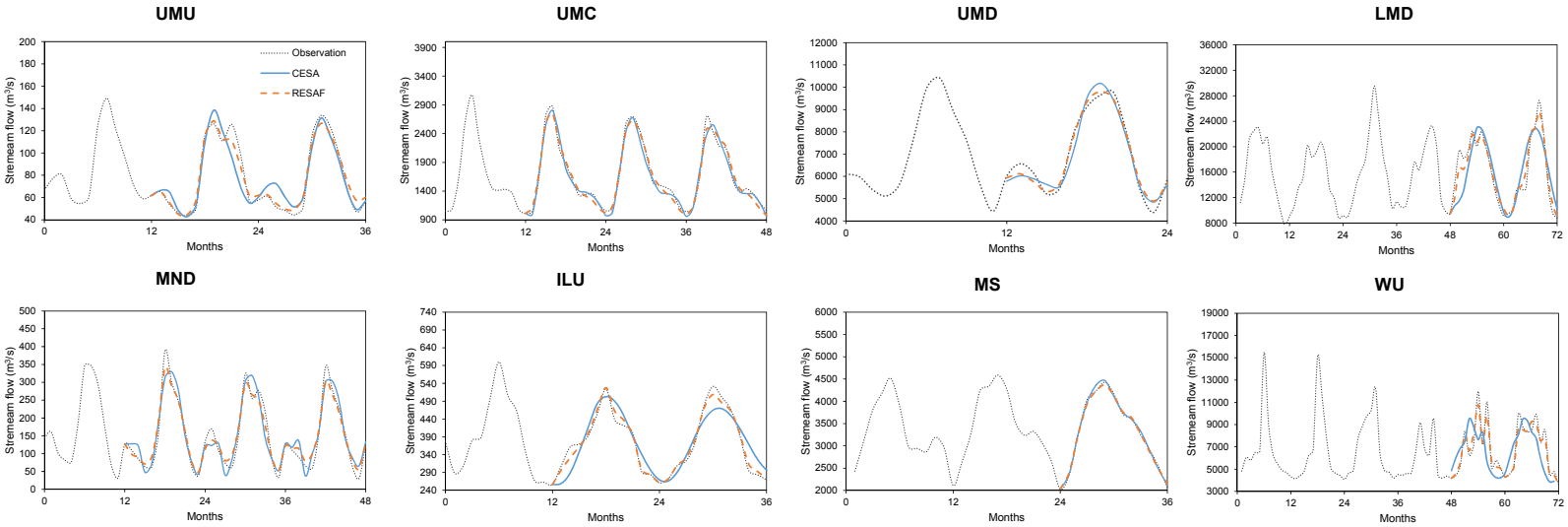


Figure 8-10 Streamflow forecasted by CESA and RESAF

Furthermore, streamflow forecasted during the low flow period was more reliable for choosing RESAF over CESA. It can be seen from the figure that streamflow forecasted with CESA during September to February was 2.7 m^3/s to 20.8 m^3/s higher than the observation, while for other prior the difference was less than 5.4 m^3/s. Besides, the peak flow forecasted by CESA dropped by 5.6% to 13.5% year by year. Due to the drop of forecasted peak of Illinois upstream (ILU) the forecasted NSE of the second lead year decreased from 0.769 of the first lead year to 0.547, when the forecasted peak from RESAF was within 4.3% variation.

8.3 Spectral power versus frequency as a random variable

The estimated spectral density and forecasted streamflow showed that adding prior information generally yielded higher resolution and accuracy. It suggested that no matter which variable is defined as a random variable, the prior information improves the MRESA theory over the MESA theory. Then comes the question of which way of defining entropy is better, spectral power as a random variable or frequency as a random variable.

When the random variable is defined by the spectral power, both the RESAS and BESA theories yielded the spectral density in the form of inverse polynomials, where the roots were the forecasting coefficients. In this case, the coefficients were solved from N nonlinear equations, and streamflow was forecasted by linear prediction. On the other hand, when the random variable is defined by frequency, the spectral density was in the form of exponential function for RESAF and CESA. Taking the inverse Fourier transform of the logarithm of the spectral density, which was the cepstrum of the autocorrelation, the Lagrange multipliers can be directly solved from the autocepstrum. The extension of

autocorrelation and forecasting of streamflow, in this case, were processed from the recursive function associated with the cepstrum. Furthermore, the prior is hypothesized for the distribution of spectral power for RESAS while the spectral density for RESAF. Comparison is made for different definition of entropy in this section.

8.3.1 Without prior information

Without considering the impact of prior information, the two maximum entropy spectral analysis theories, BESA and CESA, were first compared. It can be seen from Figure 8-11 that the spectral density estimated from BESA had multiple peaks but the order of significance was disordered. The main periodicity of the Mississippi River is 1/12, thus, the most significant peak should be found at frequency 1/12, which was estimated correctly by CESA. However, using BESA, the largest peak shifted, for example, the spectral peak estimated for UMC, UMD, and MS in the figure, shifted to frequency 1/6 but not at frequency 1/12. Besides, for the spectral density estimated for ILU, and ILD, though the largest peak was estimated at frequency 1/12 for BESA the same as for CESA, the second largest peak estimated at frequency 1/6 by BESA was 2.3 and 4.7 times larger than the one from FFT, respectively. However, CESA ignored small peaks to make sure the largest peak was correctly estimated, thus, the spectral density obtained by CESA was uni-peak with periodicity at frequency 1/12.

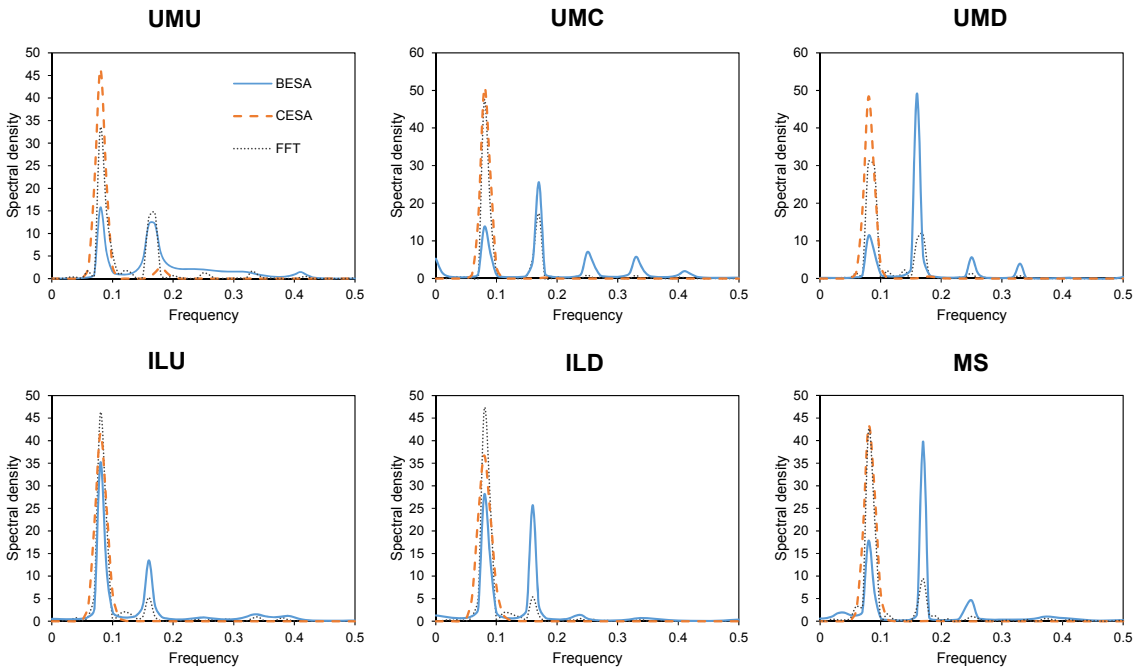


Figure 8-11 Computation of spectral densities estimated by BESA and CESA

Besides, the I-S distortion of spectral density estimated by two MESA theories in Table 8-1 shows that the I-S distortion is higher for BESA than CESA. It suggested that the resolution by CESA in estimating the spectral density was higher than by BESA.

Table 8-1 Computed Itakura-Saito distortion for CESA and BESA

Name	CESA	BESA
UMU	6.155	13.296
MNU	41.527	34.180
MND	3.688	9.475
UMC	6.243	13.356
IU	1.513	5.323
ID	5.504	10.754
DMU	19.464	25.579
DMD	8.956	8.570

Table 8-1 Continued

Name	CESA	BESA
ILU	26.459	20.933
ILD	27.722	33.325
UMD	16.128	49.393
MS	2.010	2.644
OH	2.799	2.783
WU	1.962	2.062
WD	158.595	563.501
BB	4.820	4.056
BU	26.210	33.181
OUU	39.996	16.724
OUD	1.793	1.958
LMD	0.896	1.547

The advantage of estimating the location of spectral peaks by CESA continued in streamflow forecasting. It can be seen from Figure 8-12 that graphs streamflow forecasted for the two MESA methods. Due to the shift in the estimation of spectral density for UMC, UMD and MS, the peak flow forecasted by BESA did not match the observation. The peak flow forecasted by BESA was 1 month late for the third lead year for UMC and 1 month early for UMD and MS. CESA can capture the peak flow correctly repeated every 12 months, as CESA strongly emphasized the 12 month periodicity for spectral estimation. Besides, the forecasting lead time for UMU and UMC is longer for CESA than BESA. As shown in Figure 8-12, the peak flow was forecasted with error less than 0.1 by CESA for two years lead time, but the error by BESA was larger than 0.3 for the forecasted peak second lead year.

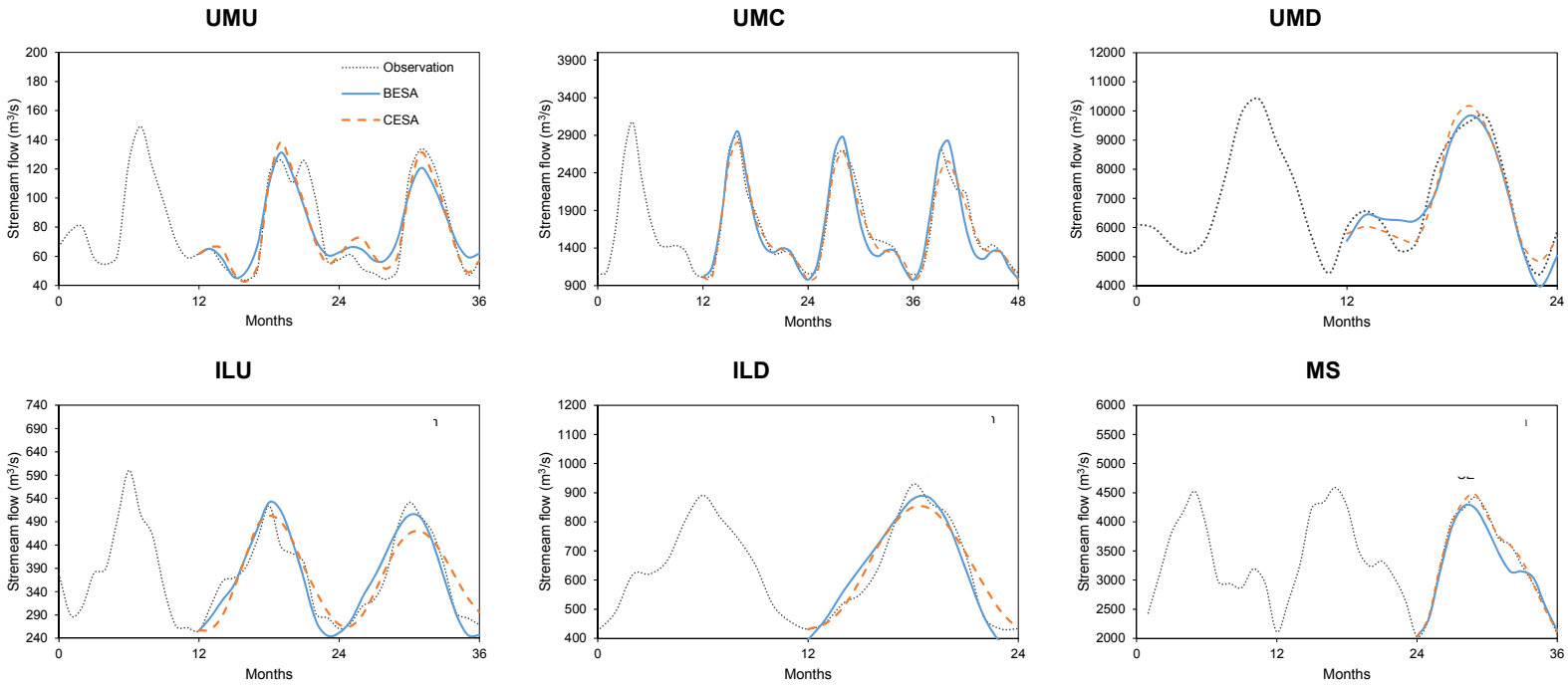


Figure 8-12 Comparison of streamflow forecasted by BESA and CESA

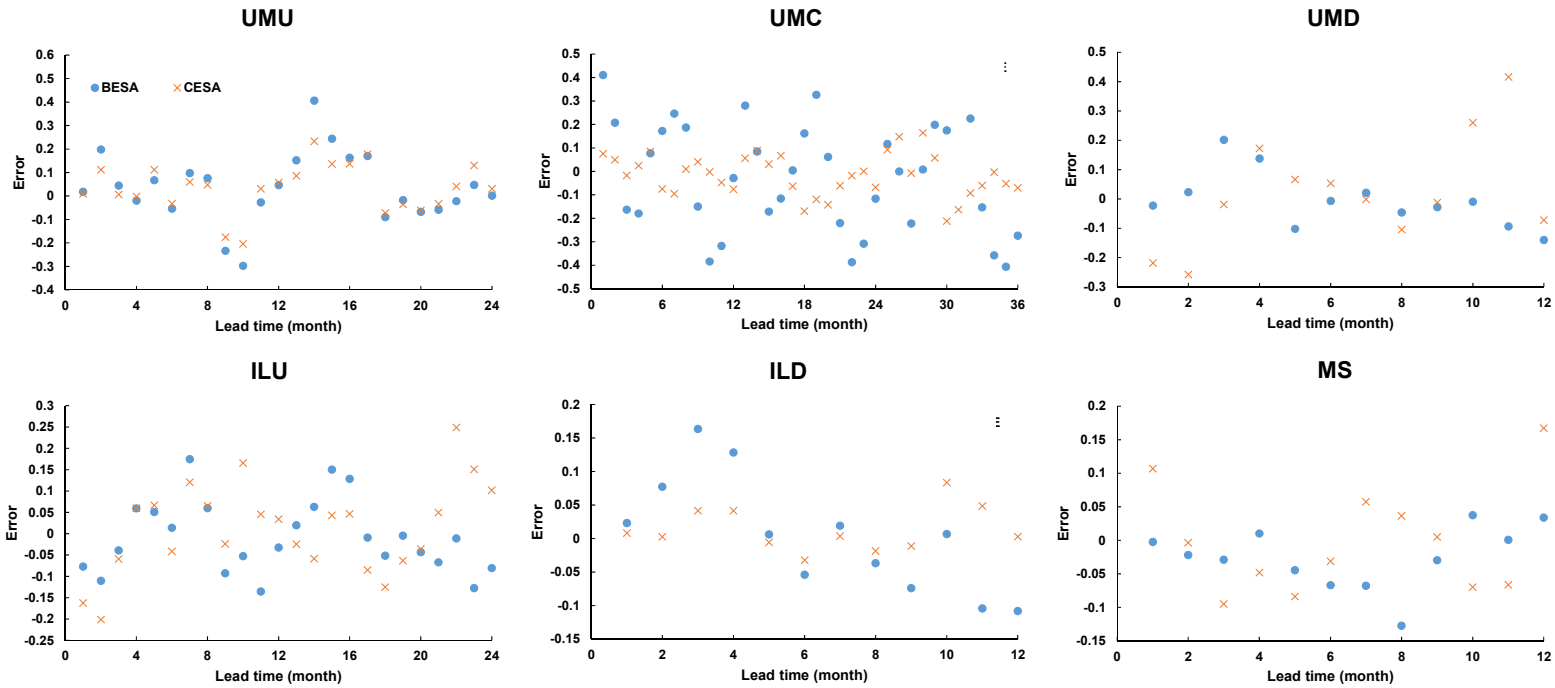


Figure 8-13 Forecasted errors by BESA and CESA

It can be seen from Figure 8-13, where the forecasted errors are plotted against lead time that forecasted error by BESA was earlier to exceed 0.2 than by CESA. However, both BESA and CESA were weak at forecasting streamflow during low flow season, which were from September to February. For example, during the 2nd lead year of UMU, the average streamflow forecasted by BESA was 21.1% higher than the observed values, while that of CESA was 19.3% higher than observation. It is noted from Figure 8-13 the absolute values of errors by BESA were larger than 0.3 during the low flow season comparing to those by CESA of around 0.2.

The order of forecasting model and NSE are given in Table 8-2, which shows that CESA used 1-2 more orders for streamflow forecasting than BESA, especially for Lower Mississippi, where BESA failed to provide reliable forecasting with model order of 2. CESA theory was more widely available to forecast streamflow than BESA. As a result, the goodness of forecasting measurement of CESA was higher than BESA. with *NSE* of 0.604 compared to that of 0.525 for BESA as shown in Table 7.

Table 8-2 Forecasting details of CESA and BESA

Name	Model order		NSE	
	CESA	BESA	CESA	BESA
UMU	13	13	0.725	0.697
MNU	13	12	0.492	0.484
MND	12	13	0.585	0.554
UMC	13	12	0.785	0.737
IU	12	13	0.498	0.475
ID	14	14	0.548	0.544
DMU	14	14	0.501	0.476
DMD	14	13	0.653	0.700
ILU	14	12	0.658	0.658

Table 8-2 Continued

Name	Model order		NSE	
	CESA	BESA	CESA	BESA
ILD	16	14	0.755	0.709
UMD	15	14	0.787	0.752
MS	15	15	0.790	0.783
OH	16	15	0.540	0.539
WU	18	6	0.432	0.179
WD	9	9	0.498	0.274
BB	15	16	0.481	0.543
BU	16	15	0.670	0.623
OUU	13	2	0.427	0.004
OUD	11	2	0.503	0.071
LMD	12	11	0.746	0.708
Average	13.75	11.75	0.604	0.525

8.3.2 With prior information

The difference in the spectral estimation between RESAS and RESAF was similar to that between BESA and CESA. It can be seen from Figure 8-14 that the spectral density estimated by RESAS is more likely to be multi-peak spectra compared to RESAF, though the location of small peaks may not exactly be consistent with the one from FFT. As shown in the figure, the spectral density estimated by RESAS had additional peaks at frequencies 1/4 and 1/3 for UMU, at frequency 1/3 for UMD and at frequency 1/4 for ILD. On the contrary, the spectral density estimated by RESAF had peaks at the specified frequency from the prior and seldom had peaks additional to the one from FFT. Like the maximum entropy case, RESAF was more likely to have the largest peak at the correct location.

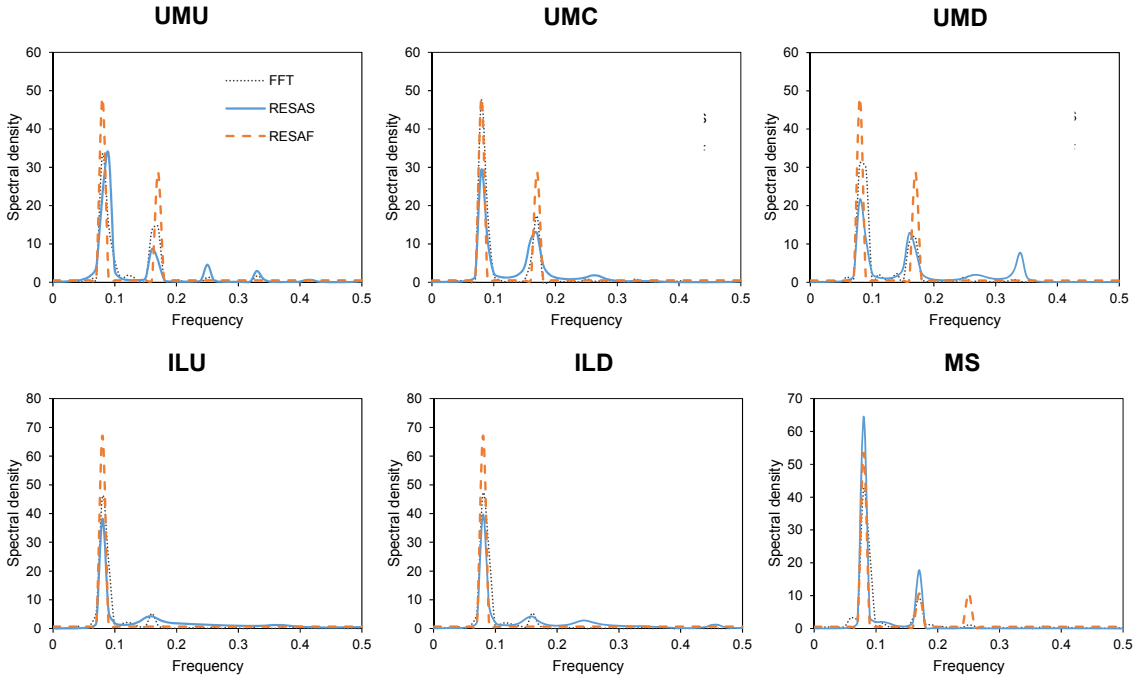


Figure 8-14 Comparison of spectral density by RESAS and RESAF

However, the capability of detecting multi-peaks of RESAS did not guarantee a smaller I-S distortion. As shown in Table 8-3 that the values of I-S distortion of two methods were similar, where the average value of I-S distortion was 1.386 for RESAS and 1.268 and RESAF. The number of counts that the I-S distortion was smaller for RESAF vs. RESAS was 10:10, as plotted in Figure 8-15. It was smaller than that by CESA vs. BESA of 14:6, which suggested that the advantage of estimating the spectral density using frequency as a random variable is more significant for applying MESA than MRESA.

Table 8-3 Computed Itakura-Saito distortion for RESAF and RESAS

Name	RESAF	RESAS
UMU	2.157	2.919

Table 8-3 Continued

Name	RESAF	RESAS
MNU	1.894	1.187
MND	1.091	1.606
UMC	0.916	0.942
IU	0.054	0.071
ID	1.997	2.510
DMU	1.856	1.995
DMD	2.247	2.904
ILU	0.692	0.505
ILD	1.065	1.356
UMD	2.801	1.917
MS	1.825	1.981
OH	0.274	0.235
WU	1.144	0.429
WD	0.881	0.631
BB	0.446	0.472
BU	1.481	1.250
OUU	1.612	0.651
OUD	1.839	0.873
LMD	1.439	0.922

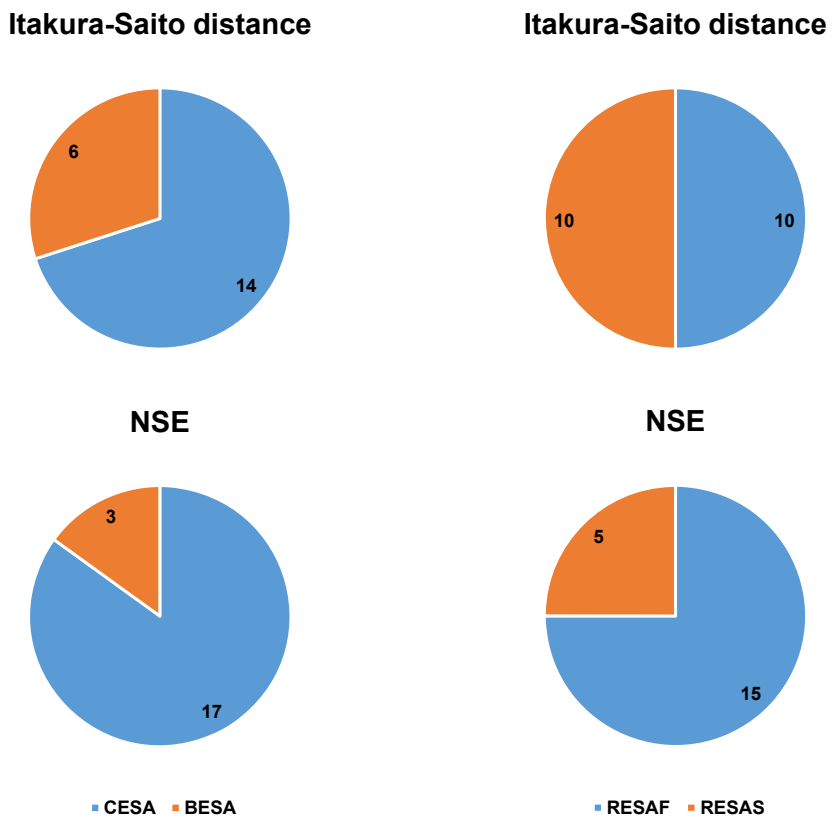


Figure 8-15 Preference of choosing CESA over BESA and RESAF over RESAS

The model orders used for streamflow forecasting by the two entropy methods were similar, as shown in Table 8-4, which suggested that the capability of possessing the past information was similar for the two methods. Thus, the difference was how the two methods dealt with that information to forecast streamflow. It was seen from the table that the average *NSE* obtained by RESAF was 0.784 slightly higher than 0.728 of RESAS. Based on the computed NSE, the preference for choosing RESAF over RESAS rose to 15:5 as shown in Figure 8-14. It suggested that though there were 5 stations that had higher

resolution spectral density using RESAS, streamflow forecasted by RESAF was more reliable.

Table 8-4 Forecasting details by RESAF and RESAS

Name	Model order		NSE	
	RESAF	RESAS	RESAF	RESAS
UMU	13	13	0.841	0.724
MNU	13	12	0.741	0.702
MND	13	12	0.774	0.671
UMC	14	12	0.857	0.816
IU	13	13	0.800	0.608
ID	13	14	0.888	0.829
DMU	14	13	0.780	0.674
DMD	14	14	0.877	0.764
ILU	15	14	0.867	0.788
ILD	16	15	0.905	0.783
UMD	16	15	0.841	0.814
MS	16	15	0.929	0.920
OH	16	16	0.618	0.645
WU	18	18	0.569	0.545
WD	10	8	0.533	0.544
BB	16	16	0.809	0.624
BU	16	15	0.834	0.869
OUU	14	15	0.774	0.748
OUD	10	8	0.611	0.621
LMD	12	11	0.811	0.874
Average	14.1	13.45	0.784	0.728

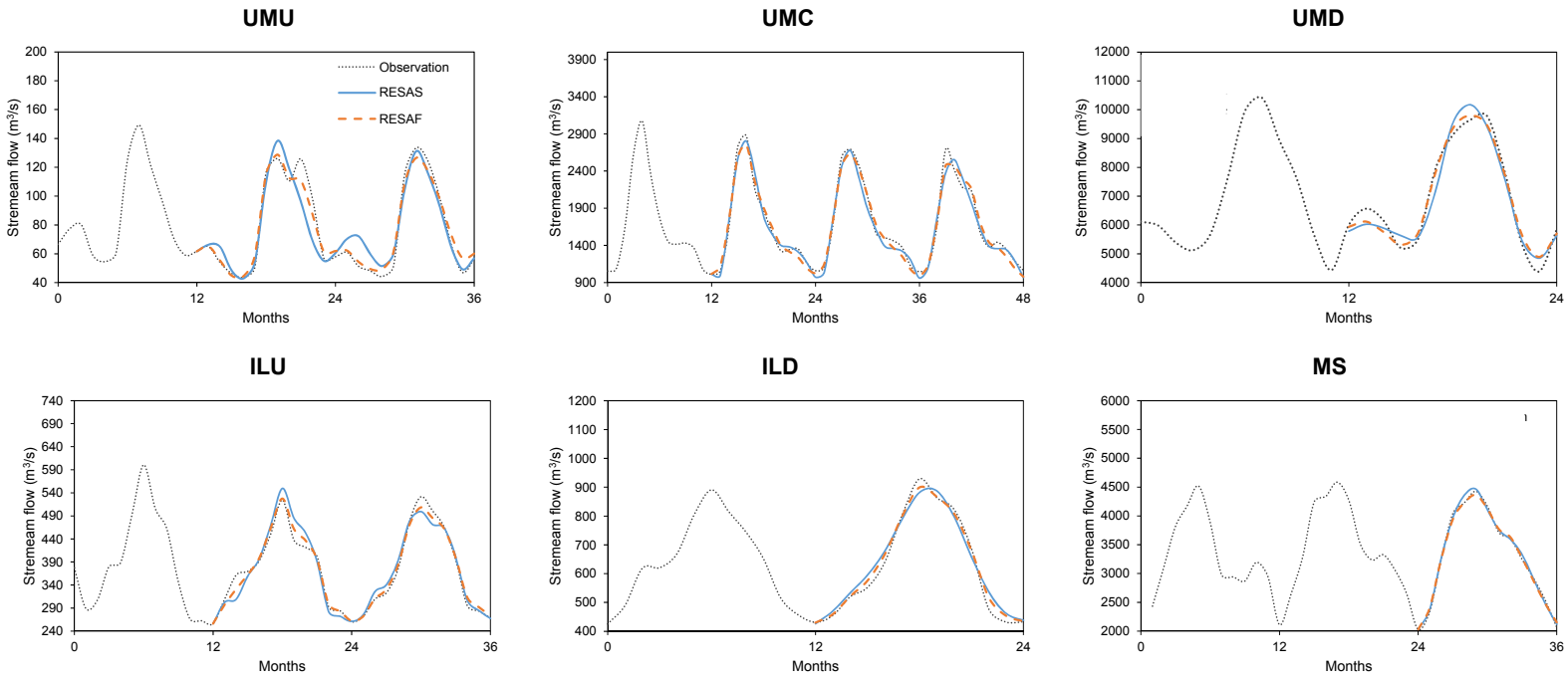


Figure 8-16 Comparison of streamflow by RESAS and RESAF

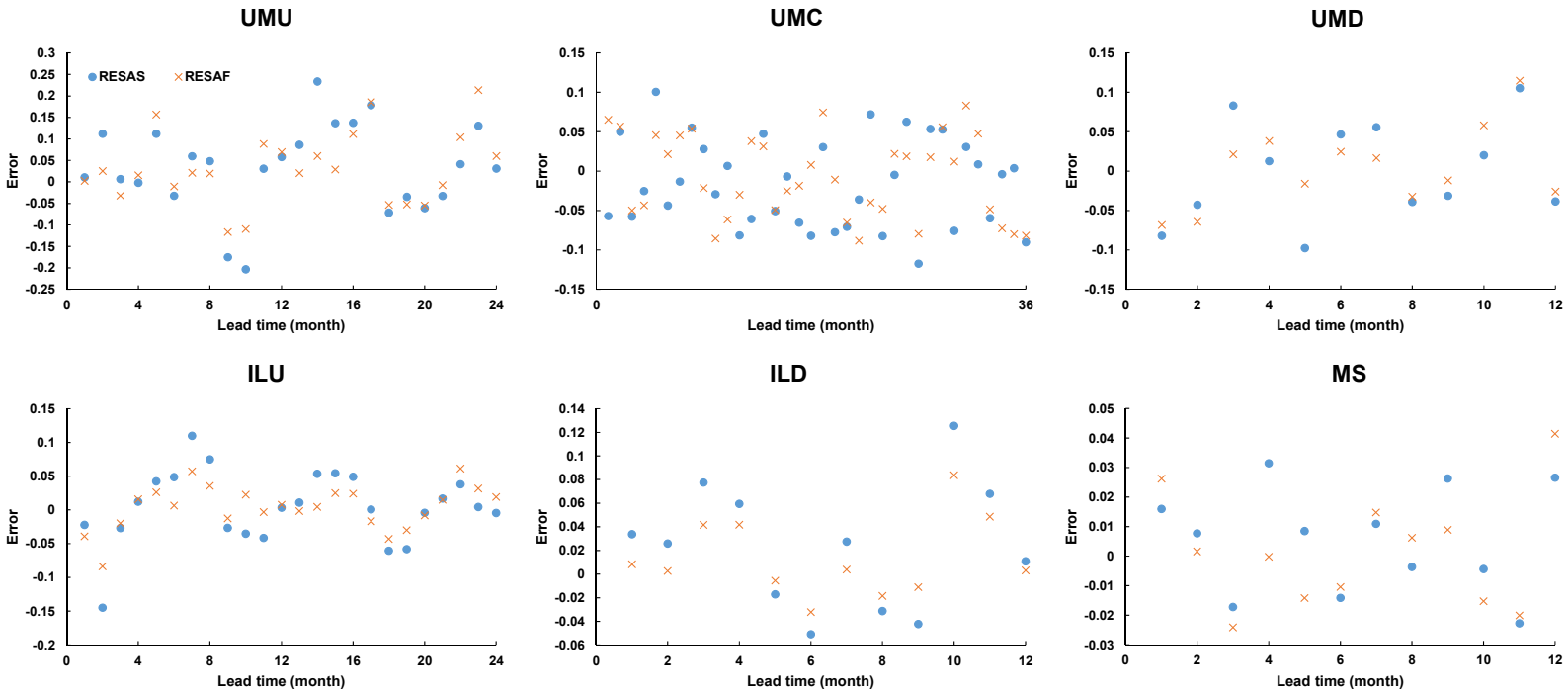


Figure 8-17 Forecasted errors by MRES and MREF

However, it was noticed that the difference between the values of NSE for RESAS and RESAF was less than 0.04. Besides, streamflows forecasted by the two MRESA theories were difficult to be distinguished from Figure 8-16, where both theories fitted observations satisfactorily. It can further be noted from forecasted errors in Figure 8-17 that two methods had similar accuracy and forecasting lead time for forecasting streamflow. But for UMU, MRES was slightly better than MREF for forecasting low flow. For low streamflow forecasting, RESAS yielded 11.7 m^3/s higher than observed flow, while that by RESAF was 1.1 m^3/s higher than observation. Though the forecasted errors were close and the differences in errors were less than 0.05 for almost all cases, but the difference for forecasted low flows of UMU was larger than 0.1.

8.3.3 Spectral power or frequency

With the spectral power as a random variable, the spectral density derived from RESAS or BESA was in the form of polynomial, while the other one, with frequency as a random variable, obtained spectral density in the form of exponential. With the exponential formula, the spectral density estimated by CESA and RESAF was no longer restricted to the AR process, but can also be applied to ARMA and MA (Liefhebber and Boekee, 1987) processes. However, RESAS or BESA was more likely to capture small peaks, while RESAF or CESA ignored small peaks and only focused on dominant peak.

Another difference was in the hypothesis on the way of imposing prior when applying MRESA. RESAS assumed a prior for the distribution of spectral power, while RESAF assumed one on the spectral density. The way of imposing prior by RESAF was more straight-forward than RESAS. To apply the RESAS, one need to collect long

historical data to obtain the past spectral power series and to analyze the distribution of spectral powers at each frequency, or at least the periodicity of streamflow need to be known, but RESAF did not. To apply RESAF, one need to add spectral peak at the significant period of the streamflow analyzed in addition to the background noise. In such a way RESAF can even be applied to data without periodicity information, since the 12 month periodicity is common for monthly streamflow. Thus, the prior 2 assumption is generally applicable for any monthly streamflow series. It suggested that to maintain the same resolution RESAS required more information to be collected a priori than RESAF.

The way for solving for the Lagrange multipliers and the forecasting coefficients was more straightforward for RESAF and CESA than for RESAS and BESA. RESAS or BESA estimated parameters by solving for N nonlinear equations, while RESAF or CESA used cepstrum analysis to estimate the Lagrange multipliers. From the given N autocorrelations, cepstrum can be obtained through the recursive function up to lag N. Though, BESA had an exception. Beyond solving the nonlinear equations by Newton's method, the Levinson-Burg algorithm can be used, which was more efficient than solving the nonlinear numerical equations. However, it was less convenient than using cepstrum analysis.

The autocorrelation was linearly extended by RESAS or BESA, while for RESAF and CESA it was extended from the N posterior cepstrum with or without prior cepstrum beyond the lag N for RESAF and CESA. Likewise, the forecasting by RESAS and RESAF yielded linear prediction, though the RESAF or CESA theory forecasted streamflow based on cepstrum analysis. Using the cepstrum analysis, small fluctuations transferred from the

spectral analysis can be loaded into the forecasting time series. It may be the reason why RESAF or CESA had the capability to forecast small peaks of streamflow, and the hydrograph is less smooth but closer to the observation than RESAS or BESA.

8.4 Summary

Through comparisons conducted in this section, following conclusions can be drawn.

Both MESA and MRESA were generally recommended over the classical AR process. Though the AR method yielded the lowest order of model to simplify computation, AR did not forecast well. It suggested that the AR method does not get enough information for forecasting. The AR process required longer memory to forecast than entropy theories but shorter lead time was reached. Though larger orders of model are used, the entropy methods extract information from much further part of the past and collect more information from the time series than the AR. As a result, forecasting by the entropy methods leads to higher accuracy

By comparing RESAS with BESA and RESAF with CESA, it was found that minimum relative entropy spectral analysis provided higher resolution in spectral estimation and more reliable streamflow forecasting, especially for multi-peak flow condition. In general, the MRESA theory was more accurate in forecasting streamflow for both peak and low flow values than MESA with longer lead time. By imposing the prior, MRESA theory can also be used to forecast streamflow where the periodicity information was limited, while BESA was not applicable and CESA was limited.

The spectral density estimated by the RESAF or CESA theory showed higher resolution than the RESAS or BESA theory, respectively, based on the I-S distortion, and streamflow forecasted by RESAF or CESA was more reliable than by RESAS or BESA. It suggested choosing frequency as a random variable. However, the advantage of choosing frequency as a random variable over spectral power was more significant between MESA theories than MRESA theories, since BESA shifted the peaks for multi-peak condition.

9 MULTI-CHANNEL ENTROPY SPECTRAL ANALYSIS FOR LONG-TERM DROUGHT FORECASTING

As indicated in Section 5 that low streamflow is difficult to forecast by itself as it is weakly autocorrelated. Thus, additional approaches may needed to increase the reliability of forecasting drought. One approach is to transfer streamflow to standardized streamflow index (SSI) to quantify drought. Similar to the widely used drought index, the standardized precipitation index (SPI), SSI incorporates hydrological processes and tells information about drought intensity and duration (Mishra and Singh, 2010). Besides, transferring streamflow observations to SSI is easier to identify the severity of the drought condition and turns highly skewed low streamflow normalized. The other approach is to combine related climatic indicators since climate indices are shown to be significant indicators of drought occurrences (Ozger et al., 2009). A strong relationship is noticed in the southern United States that dry conditions occur consistently during the La Nina events (McCabe et al., 2004; McKee et al., 1993; Piechota and Dracup, 1996). El Nino Southern Oscillation (ENSO) is shown to influence the hydrological processes such as precipitation, soil moisture, and streamflow (Barlow et al., 2001; Chiew and McMahon, 2002; Chiew et al., 1998; Kahya and Dracup, 1993; McCabe and Dettinger, 1999; Ning and Bradley, 2014; Redmond and Koch, 1991; Ropelewski and Halpert, 1986; Trenberth, 1997; Woolhiser et al., 1993), and a common three-month lag has been observed between and either precipitation or streamflow (Khedun et al., 2012; Woolhiser et al., 1993). ENSO-streamflow relationship can be suggested for use in streamflow or drought forecasting to

increase the forecasting reliability and lead time (Kalra et al., 2013; Maurer and Lettenmaier, 2003; Morid et al., 2007; Ruiz et al., 2007). Thus, multi-channel entropy spectral analysis (MCESA) is developed to forecast long-term drought represented by SSI with ENSO in this section.

9.1 Development of multi-channel entropy spectral analysis

Let us consider time series of k variable with each time series detonated as $y_l(t)$, for t=1, 2,…, n, and l=1, 2,…, k. The multi-variate time series can be written as vectors of $\vec{y}(t)=\left[y_1(t) \quad y_2(t) \quad \ldots \quad y_k(t)\right]^T$. The dependence of the variable is represented by autocorrelation and cross-correlation, which are defined as

$$\rho_l(n)=\frac{1}{T_l-n}\sum_{t=1}^{T_l-n} y_l(t)y_l(t+n) \tag{9-1}$$

$$\rho_{lj}(n)=\frac{1}{T-n}\sum_{t=1}^{T-n} y_l(t)y_j(t+n) \tag{9-2}$$

where T_l is the length of each series and T is the sum of length of each series. Equation (9-1) defines the autocorrelation of time series y_l and equation (9-2) defines the cross-correlation between time series y_l and y_j. Then the autocorrelation and cross-correlation of the same lag can be written in the matrix of

$$\rho(n)=\begin{vmatrix} \rho_1(n) & & \rho_{1k}(n) \\ & \ddots & \\ \rho_{k1}(n) & & \rho_k(n) \end{vmatrix} \tag{9-3}$$

It is noted that $\rho(n)$ is k by k matrix, and when k=1 it reduces to autocorrelation as defined in Section 3.

9.1.1 Definition of entropy

In spectral analysis, there marginal spectral density $p_l(f)$of each time series $y_l(t)$ and cross-spectral density $p_{lj}(f)$ between time series y_l and y_j. Then the entropy of each time series defined by configurational entropy can be written as

$$H_l(f) = -\int_{-W}^{W} p_l(f) \ln[p_l(f)]df \tag{9-4}$$

and similarly, the cross-entropy between two time series is defined from cross-spectral density as

$$H_{lj}(f) = -\int_{-W}^{W} p_{lj}(f) \ln[p_{lj}(f)]df \tag{9-5}$$

Likewise, the entropy of the multi-channel spectra can be written as a k by k matrix as

$$\mathrm{H}=\begin{vmatrix} H_1 & & H_{1k} \\ & \ddots & \\ H_{k1} & & H_k \end{vmatrix} \tag{9-6}$$

It can be seen from equation (9-6) that the diagonals in the matrix are marginal entropies of each time series. It is noted that when k=1, it reduces to the univariate case, which is equivalent to configurational entropy developed in Section 5. To obtain the least-biased estimation of the spectral density, the determinant of the entropy matrix needs to be maximized.

9.1.2 Constraints

The relationship between the spectral density and autocorrelation still holds in the multi-variate case. Thus

$$\rho_l(n) = \int_{-W}^{W} p_l(f) e^{i2\pi fn\Delta t} df \tag{9-7}$$

$$\rho_{lj}(n) = \int_{-W}^{W} p_{lj}(f) e^{i2\pi fn\Delta t} df \tag{9-8}$$

where Δt is the discretization or sampling interval, and $i = \sqrt{-1}$.

9.1.3 Entropy maximization

Maximization of the determination of maximum entropy can be done by the Lagrange multipliers method. With k by k Lagrange multipliers matrix $\Lambda(n)$, for n=1, 2, …, N, the Lagrange function matrix can be written as

$$L(f) = H - \sum_{n=-N}^{N} \Lambda(n)[\int_{-W}^{W} P(f)\exp(i2\pi fn\Delta t)df - \rho(n)] \tag{9-9}$$

where each element equals to

$$L_l(f) = -\int_{-W}^{W} p_l(f)\ln[p_l(f)]df - \sum_{n=-N}^{N} \Lambda_l(n)[\int_{-W}^{W} p_l(f)\exp(i2\pi fn\Delta t)df - \rho_l(n)] \tag{9-10}$$

$$L_{lj}(f) = -\int_{-W}^{W} p_{lj}(f)\ln[p_{lj}(f)]df - \sum_{n=-N}^{N} \Lambda_{ij}(n)[\int_{-W}^{W} p_{lj}(f)\exp(i2\pi fn\Delta t)df - \rho_{lj}(n)] \tag{9-11}$$

where Lagrange multipliers are a k by k matrix: $\Lambda(n) = \begin{bmatrix} \lambda_1(n) & & \lambda_{1k}(n) \\ & \ddots & \\ \lambda_{k1}(n) & & \lambda_k(n) \end{bmatrix}$.

It is noted that taking the derivative of equation (9-9) with respect to cross-spectral density $p_{lj}(f)$ is equivalent to taking derivative of the cross entropy term $L_{lj}(f)$ with respect to $p_{lj}(f)$, since $p_{lj}(f)$ is independent from other entropy elements. Thus, one has

$$\frac{\partial L(f)}{\partial p_{lj}(f)} = \frac{\partial L_{lj}(f)}{\partial p_{lj}(f)} = -\int_{-W}^{W} \{\ln[p_{lj}(f)] + 1 + \sum_{n=-N}^{N} \Lambda_{lj}(n)\exp(i2\pi fn\Delta t)\} df = 0 \tag{9-12}$$

Thus, equaling equation (9-12) to zero, the cross-spectral density is obtained as

$$P_{ij}(f) = \exp\left[-1 - \sum_{n=-N}^{N} \Lambda_{ij}(n) e^{i2\pi f n \Delta t}\right] \tag{9-13}$$

In the same manner, the each element of spectral density can be obtained and the spectral density can be written as

$$\begin{aligned} P(f) &= \begin{vmatrix} P_1(f) & & P_{1k}(f) \\ & \ddots & \\ P_{k1}(f) & & P_k(f) \end{vmatrix} \\ &= \begin{vmatrix} \exp\left[-1 - \sum_{n=-N}^{N} \Lambda_1(n) e^{i2\pi f n \Delta t}\right] & & \exp\left[-1 - \sum_{n=-N}^{N} \Lambda_{1k}(n) e^{i2\pi f n \Delta t}\right] \\ & \ddots & \\ \exp\left[-1 - \sum_{n=-N}^{N} \Lambda_{k1}(n) e^{i2\pi f n \Delta t}\right] & & \exp\left[-1 - \sum_{n=-N}^{N} \Lambda_k(n) e^{i2\pi f n \Delta t}\right] \end{vmatrix} \end{aligned} \tag{9-14}$$

9.1.4 Computation of Lagrange multipliers

Computation of the Lagrange multipliers follows the same method as in Section 5 using the cepstrum analysis. Let us define the cross-cepstrum from cross-spectral density as

$$e_{lj}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left|p_{lj}(f)\right| e^{2\pi n i f} \, df \tag{9-15}$$

For finite series of data, the k by k cross-cepstrum matrix can be obtained from cross-correlation using equation (3-39), which is

$$e(n) = 2[\rho(n) - \sum_{q=1}^{n-1} \frac{q}{n} e(q)\rho(n-q)], \quad n > 0 \tag{9-16}$$

Following the steps in equations (5-8) to (5-11), each element of equation (9-14) can be written as

$$\delta_n + e_{ij}(n) = -\sum_{s=-N}^{N} \Lambda_{ij}(s)\delta_{n-s} \tag{9-17}$$

where σ_n is the delta function. Equation (9-16) can be expanded in N linear equations as

$$\begin{aligned} \Lambda_{ij}(0) &= -1 - e_{ij}(0) \\ \Lambda_{ij}(1) &= -e_{ij}(1) \\ &\vdots \\ \Lambda_{ij}(N) &= -e_{ij}(N) \end{aligned} \tag{9-18}$$

It is seen from equation (9-18) that the Lagrange multipliers can be estimated from the cepstrum given in equation (9-16).

9.1.5 Extension of cross-correlation

In order to do time series forecasting, correlation is needed to be extended. Like the univariate condition, the correlation matrix can be extended by taking the inverse of equation (9-16), which is

$$\rho(n) = \sum_{q=1}^{m-1} \frac{q}{m} e(q)\rho(n-q) \tag{9-19}$$

It is noted that equation (9-19) extends the correlation matrix under the maximum entropy theory, which yields linear extension. Thus, the time series can be forecasted using a linear model with prediction coefficient matrix A equal to

$$A(q) = \frac{q}{m} e(q) = \frac{q}{m} \begin{vmatrix} e_1(q) & & e_{1k}(q) \\ & \ddots & \\ e_{k1}(q) & & e_k(q) \end{vmatrix} \tag{9-20}$$

9.1.6 Forecasting

As is shown in the Section 3, for linear forecasting the coefficient of extending autocorrelation function is the same as the prediction coefficient. Thus, multi-channel time series forecasting can be obtained by

$$y(t+T)=\sum_{q=1}^{m}A(q)y(t+T-q)\text{, for t>0} \tag{9-21}$$

It is noted that equation (9-21) is h-step forecasting, and all k time series are forecasted simultaneously. Thus, from the second step of forecasting, the previous forecasted values are combined. However, in this Section, the object is to forecast long-term drought using ENSO data as an indicator, and future ENSO data are supposed to be given, not to be forecasted. Let denote us $x(t)$ as ENSO series for t=1,…, T_x, and $y(t)$ for t=1, …, T_y is drought index series given and $\hat{y}$ are the forecasted values. The long-term drought is forecasted from

$$\begin{pmatrix}\hat{x}(T_y+t)\\ \hat{y}(T_y+t)\end{pmatrix}=\sum_{q=1}^{m}\begin{vmatrix}A_x & A_{xy}\\ A_{yx} & A_y\end{vmatrix}\begin{pmatrix}x(T_y+t-q)\\ y'(T_y+t-q)\end{pmatrix}\text{, for t>0} \tag{9-22}$$

where $y'(T_y+t-q)=\hat{y}(T_y+t-q)$, for t-q>0 and $y'(T_y+t-q)=y(T_y+t-q)$, for t-q<0. But no forecasted values will be used for ENSO. Equation (9-21) ensures drought forecasted from past drought information and future ENSO series.

9.2 Data description

9.2.1 ENSO

ENSO refers to the large-scale ocean-atmosphere climate phenomenon linked to periodic warm or cold in sea-surface temperature near the center and east center of

equatorial Pacific. The "Southern Oscillation" refers to variations in the temperature of the sea-surface, and El Nino represents warm phase in air surface pressure, while La Nina as cooling phase represents cool phase.

Monthly ENSO data estimated by Oceanic Nino Index (ONI) from 1953-2012 were obtained from the website of NOAA's National Weather Service (http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml). ENSO data are computed by 3 month running mean of ERSST.v3b (Smith et al., 2008) anomalies in the Niño 3.4 region (5°N-5°S, 120°-170°W). It is noted that El Nino and La Nina episodes are defined on a threshold of +/-0.5˚C. El Nino is evidenced when the 3-month average sea-surface temperature departure exceeds 0.5˚C, vice versa. Figure 9-1 plots the ENSO time series from January 1953 to December 2012. As seen from the figure, during the 60 year period, the total number of 205 months were under El Nino and 215 months were under La Nina.

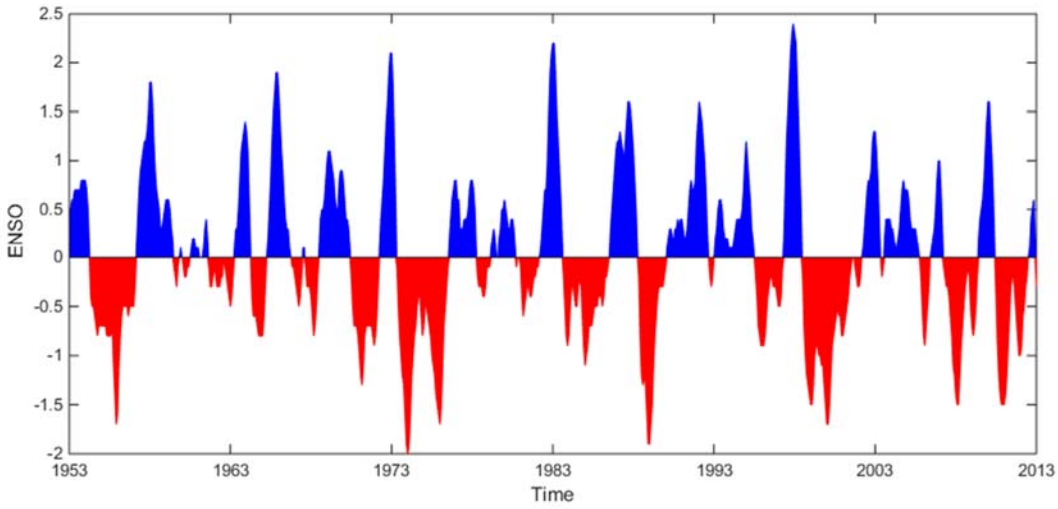


Figure 9-1 ENSO time series from 1953 to 2012

9.2.2 Streamflow data

Twelve stations in Texas were selected and downloaded from the USGS website for drought analysis, which are tabulated in Table 9-1. The average monthly streamflow for these stations were all less than 90 m^3/s and had large deviations. Except for the stations on Trinity River and Brazos River, the mean monthly streamflows from other rivers were smaller than 7.5 m^3/s. These rivers were under different degrees of drought, where the mode of flow was lower than 5 $l/sec/km^2$. The ratio of maximum monthly streamflow to mean streamflow was 10 to 136. Besides, streamflow had high values of skewness varying from 2.6 to 20. It can be seen from Figure 9-2, where histograms of observed streamflow are plotted, that the distributions were highly right skewed with long tails. More than 86% of data were observed in the first 10% quintiles.

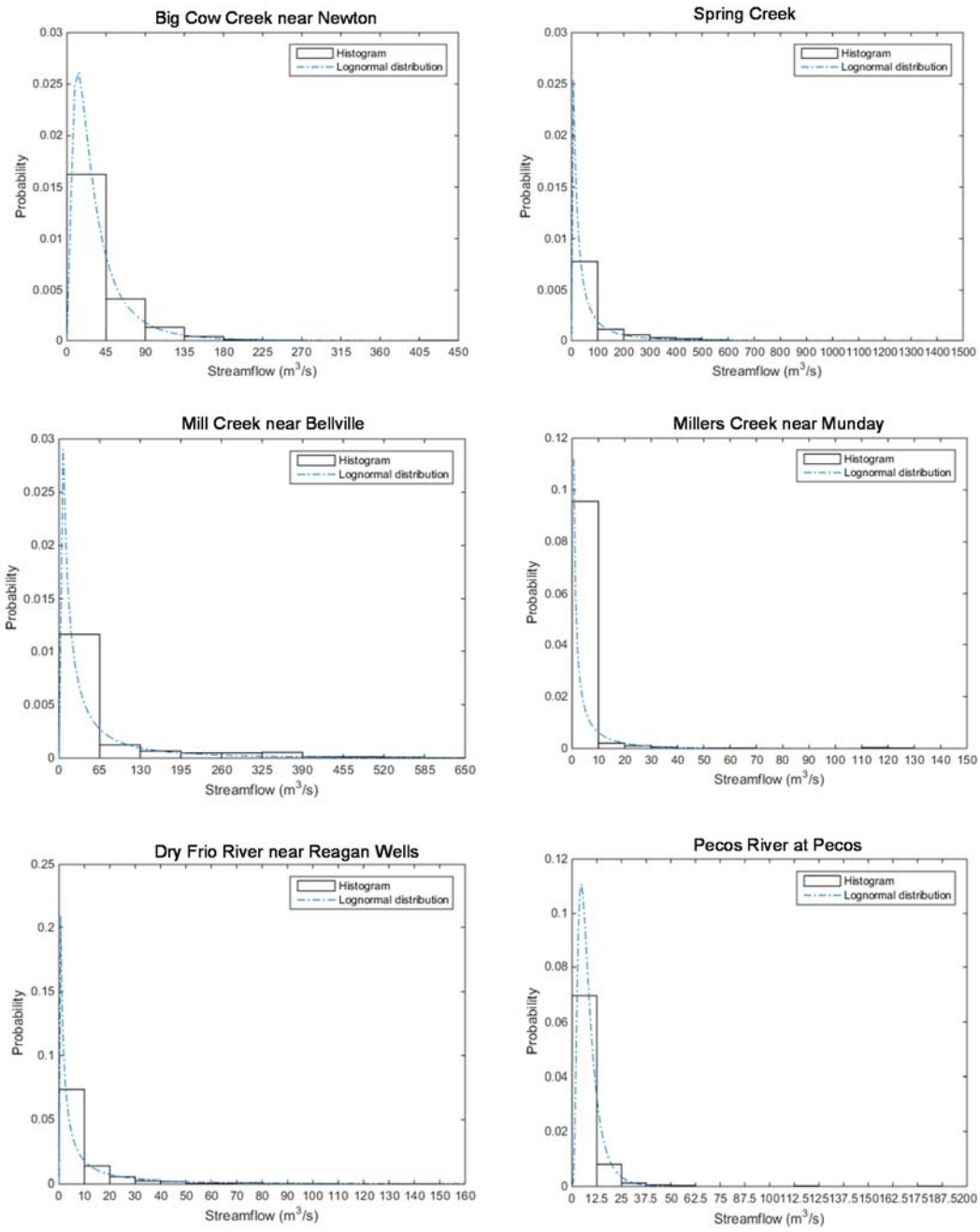


Figure 9-2 Histograms of observed streamflow in Texas

Table 9-1 Statistics of streamflow stations

Station name	Station no.	Latitude	Longitude	Drainage area (km^2)	Mean (m^3/s)	Max (m^3/s)	Min (m^3/s)	Std* (m^3/s)	Skewness
Canadian River near Amarillo	07227500	35°28'13"	101°52'45"	50363	6.90	226.9	0.00	18.20	6.73
Red River near Wayside	07297910	34°50'15"	101°24'49"	9723	0.74	39.9	0.00	2.22	11.57
Big Cow Creek near Newton	08029500	30°49'08"	93°47'08"	332	3.64	42.8	0.12	3.62	3.20
Trinity River near Rosser	08062500	32˚25'35"	96˚27'46"	21101	89.51	1143.3	0.93	120.47	3.20
Spring creek near Spring	08068500	30°06'37"	95°26'10"	1059	7.35	146.8	0.08	12.08	3.87
Brazos River near Highbank	08098290	31˚08'02"	96˚49'29"	78829	75.18	903.6	0.87	115.84	3.43
Millers Creek near Munday	08082700	33°19'45"	99°27'53"	269	0.16	10.0	0.00	0.79	9.57
Mill Creek near Bellville	08111700	29°52'51"	96°12'18"	974	5.88	61.9	0.00	10.41	2.61
Colorado River near Stacy	08136700	31˚29'37"	99˚34'25"	62660	3.71	83.6	0.00	8.10	4.49
Nueces River below Uvalde	08192000	29°07'25"	99°53'40"	4820	4.03	253.1	0.00	15.23	10.24
Dry Frio River near Reagan Wells	08196000	29°30'16"	99°46'52"	326	0.96	14.6	0.00	1.56	3.98
Pecos River at Pecos	08446500	31°06'47"	102°25'02"	76560	1.78	240.7	0.08	9.52	20.04

*Std=standard deviation

9.2.3 Standardized streamflow index

The standardized streamflow index (SSI), also called standardized runoff index (SRI), was derived by Shukla and Wood (2008) to determine the loss of streamflow due to climate change. Monthly SSI was shown to be a useful complement to standardized precipitation index for depicting hydrological aspects of drought (Mishra and Singh, 2010). One of the advantages standardizing streamflow is that it ensures the severity of drought in different locations to be compared independently of local climatic characteristics (Vicente-Serrano et al., 2012). Thus, there were common standards to quantify the drought severity, which are shown in Table 9-2.

Table 9-2 Weather classification based on SSI

SSI values	Weather class
SSI>2	Extremely wet
1.5<SSI<1.99	Very wet
1.0<SSI<1.49	Moderately wet
-0.99<SSI<0.99	Near normal
-1.49<SSI<-1	Moderately dry
-1.99<SSI<-1.5	Severely dry
SSI<-2	Extremely dry

According to discussion by Shukla and Wood (2008), the SSI can be computed from the concept for computing standardized precipitation index (SPI) developed by McKee et al. (1993). First of all, the probability distribution of streamflow is needed to calculate the SSI. Shukla and Wood (2008), Tabari (2013) and Zaidman et al. (2002), suggested fitting low streamflow with a lognormal distribution or a gamma distribution. The Kolgomorov-Smirnov test shows that streamflow in Texas follows the lognormal

distribution, which is plotted in Figure 9-2. Thus, the lognormal distribution was used for computing SSI.

Then, SSI was calculated as

$$SSI = +(W - \frac{C_0 + C_1 W + C_2 W^2)}{1 + d_1 W + d_2 W^2 + d_3 W^3}, \text{ for } F_y > 0.5 \quad (9\text{-}23a)$$

$$SSI = -(W - \frac{C_0 + C_1 W + C_2 W^2)}{1 + d_1 W + d_2 W^2 + d_3 W^3}, \text{ for } 0 < F_y \le 0.5 \quad (9\text{-}23b)$$

where $W = \sqrt{\ln\left(\frac{1}{F_y^2}\right)}$, for $0 < F_y \le 0.5$ (9-24a)

$$W = \sqrt{\ln\left[\frac{1}{(1 - F_y)^2}\right]}, \text{ for } F_y > 0.5 \quad (9\text{-}24b)$$

and F_y represents the fitted cumulative distribution for streamflow; C_0=2.515517, C_1=0.802853, C_2=0.010328, d_1=1.432788, d_2=0.189269, and d_3=0.001308. The time series of calculated SSI are shown in Figure 9-3. As shown in the figure, SSI shows stationarity over time, and has mean zero. Most of the SSI values fall between -2 and 2, but for the Canadian River and the Pecos River they drop below -2 more frequently than for other rivers.

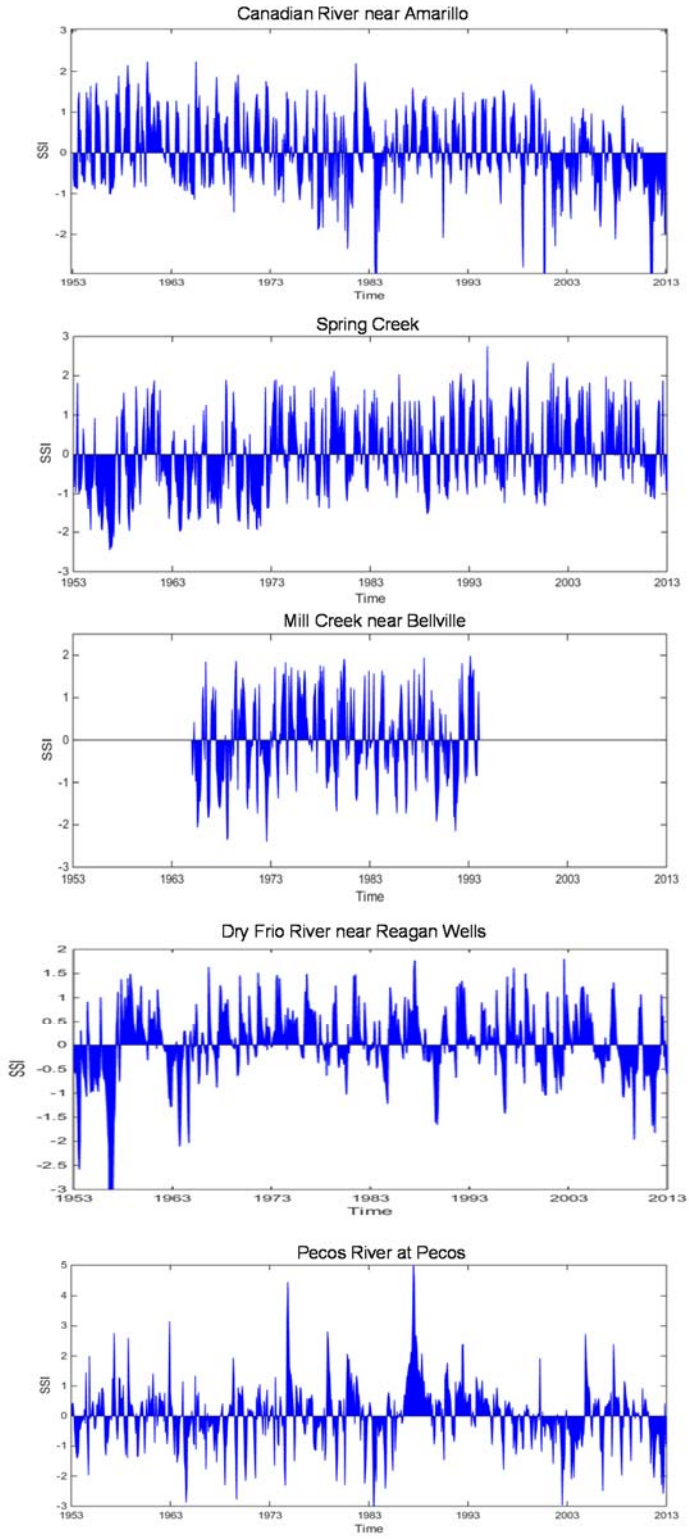


Figure 9-3 Computed SSI for rivers in Texas

Table 9-3 Flow conditions

River	Weather zone	Extremely wet	Very wet	Moderately wet	Near normal	Moderately dry	Severely dry	Extremely dry
Canadian River	Continental	0.6%	3.3%	11.7%	74.9%	5.3%	2.5%	1.8%
Red River		0.2%	1.6%	5.1%	83.9%	5.8%	2.0%	1.4%
Big Cow Creek	Subtropical humid	1.4%	6.3%	10.4%	65.3%	11.5%	4.0%	1.1%
Trinity River		1.9%	5.8%	11.1%	66.1%	8.8%	4.3%	1.9%
Spring Creek		1.1%	6.8%	11.0%	64.6%	11.4%	4.4%	0.7%
Brazos River		3.2%	5.7%	8.5%	69.1%	6.9%	4.6%	2.0%
Millers Creek	Subtropical semi-humid	0.3%	7.7%	8.3%	72.4%	6.6%	2.8%	1.9%
Mill Creek		0.0%	9.2%	9.2%	64.9%	10.3%	4.9%	1.4%
Colorado River		0.0%	1.5%	6.8%	89.8%	0.6%	0.2%	1.1%
Nueces River	Semi arid	0.0%	1.3%	5.7%	85.6%	4.6%	1.7%	1.1%
Dry Frio River		0.0%	1.0%	6.7%	85.7%	3.8%	1.4%	1.5%
Pecos River	Arid	3.1%	2.1%	6.1%	74.9%	7.9%	3.1%	2.9%

The percentage of SSI during 1953-2012 falling into different weather criteria in Table 9-2 are calculated and listed in Table 9-3. The weather zones where each station is located are stated in Table 9-3. The rivers under different climate zone show different distributions of the severity of drought as shown in Figure 9-4. It is seen that SSI of rivers under subtropical humid and subtropical semi-humid weather have the shortest duration of normal condition around 64%-72%, but more than 15% of time, they are under drought. Besides, for these river basins, the duration of wet condition is similar to that of dry condition. However, Pecos River from arid zone had the longest period (2.9%) during extreme drought, which is shown in Figure 9-4.

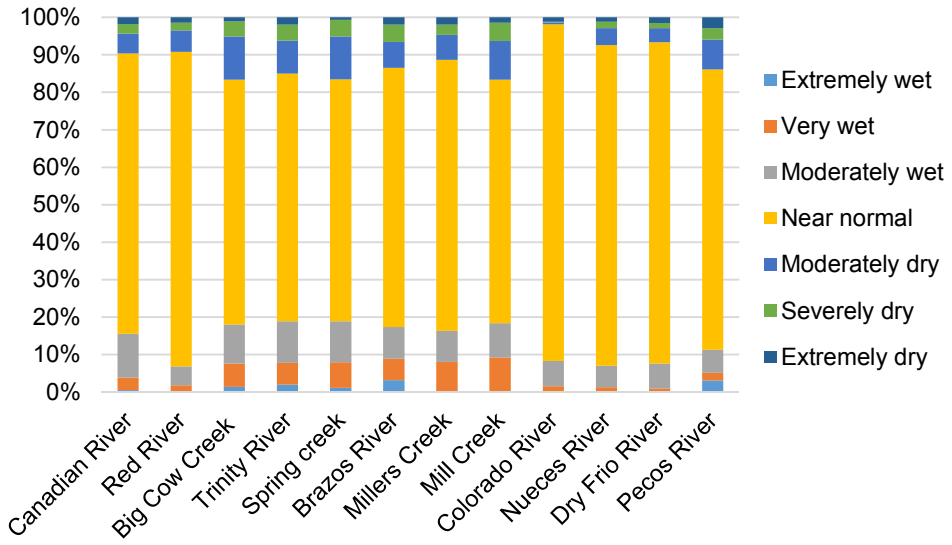


Figure 9-4 Percentage of weather condition based on SSI

9.3 Evaluation

9.3.1 Correlation

The autocorrelation of SSI and cross-correlation between SSI and ENSO are examined and are plotted in Figure 9-5. Recall the autocorrelation plot in Figure 5-4 in Section 5 that autocorrelation of low streamflow rarely showed the periodically repeating pattern or had significant values beyond the confidence bounds. As a result, the entropy approach for forecasting those low streamflow was not successful. However, when transferred to SSI, autocorrelation, shown in the first rows in Figure 9-5 shows a periodic pattern of 12 months, except for Dry Frio River. It reveals the periodicity which is contained but were too weak to be shown in the original series. Autocorrelation from Big Cow Creek under subtropical humid condition has the most significant values, as compared to rivers from other weather zones. Besides, more significant values are found in autocorrelation of SSI. For example, autocorrelation of SSI in Canadian River and Big Cow Creek is strong even after 4 years' lag time. Much stronger autocorrelation and periodicity of SSI than original streamflow made forecasting more reliable.

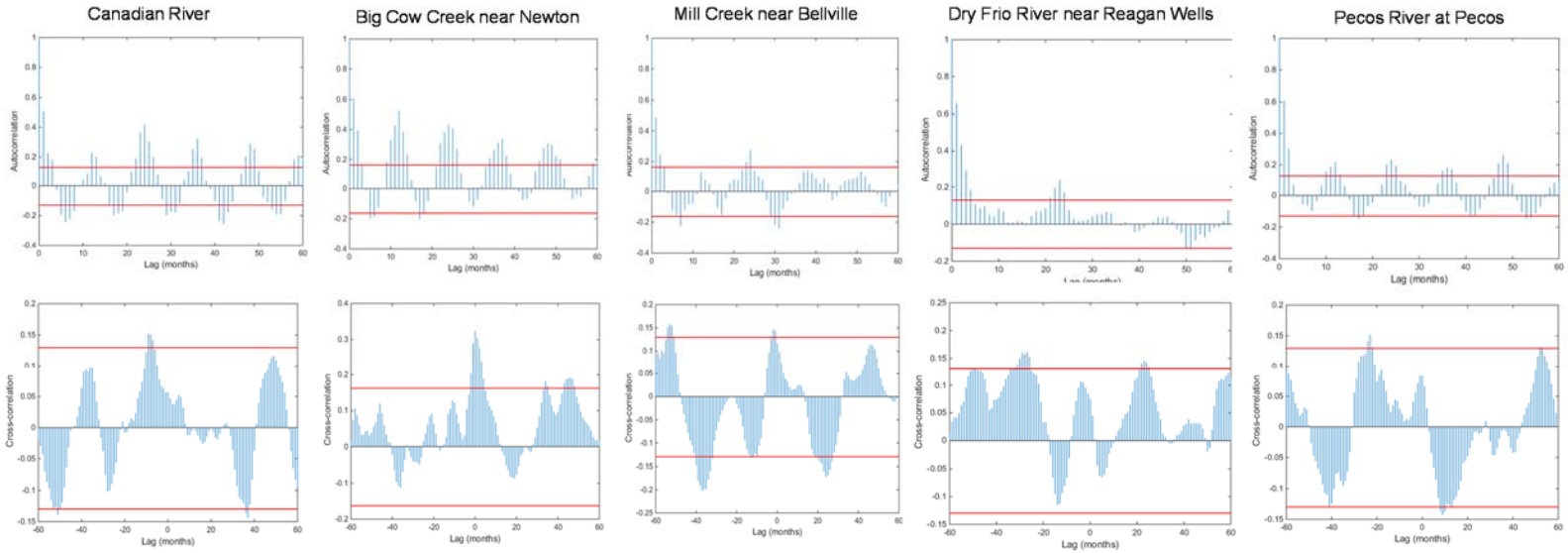


Figure 9-5 Autocorrelation and cross-correlation of SSI

The second row in Figure 9-5 plots the cross-correlation between ENSO and SSI. Many significant values are found for cross-correlation, which ensures the existence of link between ENSO and SSI. The most significant values are observed near lag 0 for Canadian River, Big Cow creek and Mill Creek, which suggests the synchronization between ENSO and SSI. Furthermore, it is seen from the figure that the cross-correlation is more significant for positive lags than negative lags. It suggests that the impact of ENSO may have some delay on SSI and the influence will maintain in the future series. Such correlation provided sufficient evidence to forecast SSI using ENSO. However, for rivers from semi-arid and arid region, the correlation between ENSO and SSI are weaker, and less significant values are observed.

9.3.2 Forecasted SSI

Based on equation (9-22), SSI was forecasted with ENSO for four to six years lead time, and the result can be found in Table 9-4. Though the mean values was forecasted with 8% to 21% error compared to observed values, forecasting SSI using multi-channel spectral analysis was more reliable than forecasting low streamflow using CESA. However, the forecasted SSI series had smaller deviations compared to observations. It was noted that r^2 using MCESA was larger than or around 0.2, which was less than 0.1 in Section 5. Besides, the NSE values were all positive and between 0.15 and 0.3, which was more than 10 times of that obtained for low streamflow. Especially, for Trinity River, Brazos River and Colorado River, forecasted low streamflow using CESA had r^2 of 0.258, 0.016 and 0.054, respectively, with 1 year lead time. But using MCESA, r^2 of forecasted SSI increased to 0.347, 0.373 and 0.153, and forecasting lead time increased to 6 years.

Though the forecasting for SSI was not as accurate as forecasting univariate high streamflow, the progress from forecasting low streamflow was evident.

Table 9-4 Forecasted result of SSI in Texas

Station name	Forecasting period	Obs. Mean	Est. Mean	Obs. Std	Est. Std	RMSE	r^2	NSE
Canadian River	1995-2000	0.056	0.033	1.579	0.426	1.267	0.347	0.175
Red River	2005-2010	0.103	0.085	1.134	0.887	0.145	0.329	0.188
Big Cow Creek	1989-1994	0.481	0.362	0.850	0.345	0.670	0.369	0.258
Trinity River	2007-2012	0.216	0.109	0.732	0.489	0.703	0.347	0.274
Spring creek	1995-2000	0.105	0.125	0.937	0.425	0.787	0.285	0.243
Brazos River	2006-2011	0.073	0.102	0.704	0.556	0.581	0.373	0.294
Millers Creek	2003-2006	0.044	0.006	0.062	0.041	0.065	0.219	0.173
Mill Creek	1983-1988	0.222	0.212	0.893	0.415	0.723	0.334	0.180
Colorado River	2007-2012	-0.645	-0.497	0.385	0.213	0.352	0.153	0.120
Nueces River	1997-2000	0.529	0.435	3.400	1.447	0.749	0.197	0.211
Dry Frio River	2007-2012	0.104	-0.004	0.677	0.293	0.575	0.267	0.124
Pecos River	1969-1974	0.071	-0.216	1.182	0.340	1.005	0.267	0.098

The forecasted SSI for the rivers under subtropical humid climate showed the highest reliability with r^2 over 0.35 and 6 years lead time in Table 9-4. As shown in Figure 9-6, the SSI of Big Cow Creek showed the most significant periodicity, where the dry and wet conditions occurred repeatedly. Thus, the forecasted SSI using MCESA showed periodically up and down. However, the forecasted severity of wet and drought conditions were different from observations. As a result, the residuals were large when the SSI observed was beyond normal condition (+/- 1) as shown in Figure 9-7.

For rivers under subtropical semi-humid, semi-arid or continental climate, the SSI was less regular and the periodicity was less significant compared to that under subtropical-humid condition, as shown in Figure 9-6. In these cases the MCESA was

capable to forecast SSI for 4-6 years lead time with r^2 around 0.2-0.3. However, the droughts of Mill Creek during 1963 and 1965 were underestimated, and the residuals during these periods were over 1.

The SSI of rivers under arid climate was much different from others, which was more frequently lower than 0. As shown in Figure 9-6, 87% of the forecasted SSI of Pecos River were negative. In this case, the significant residuals were mostly from the period of observed wet condition, for example, from late 1969 to 1970 and late 1974.

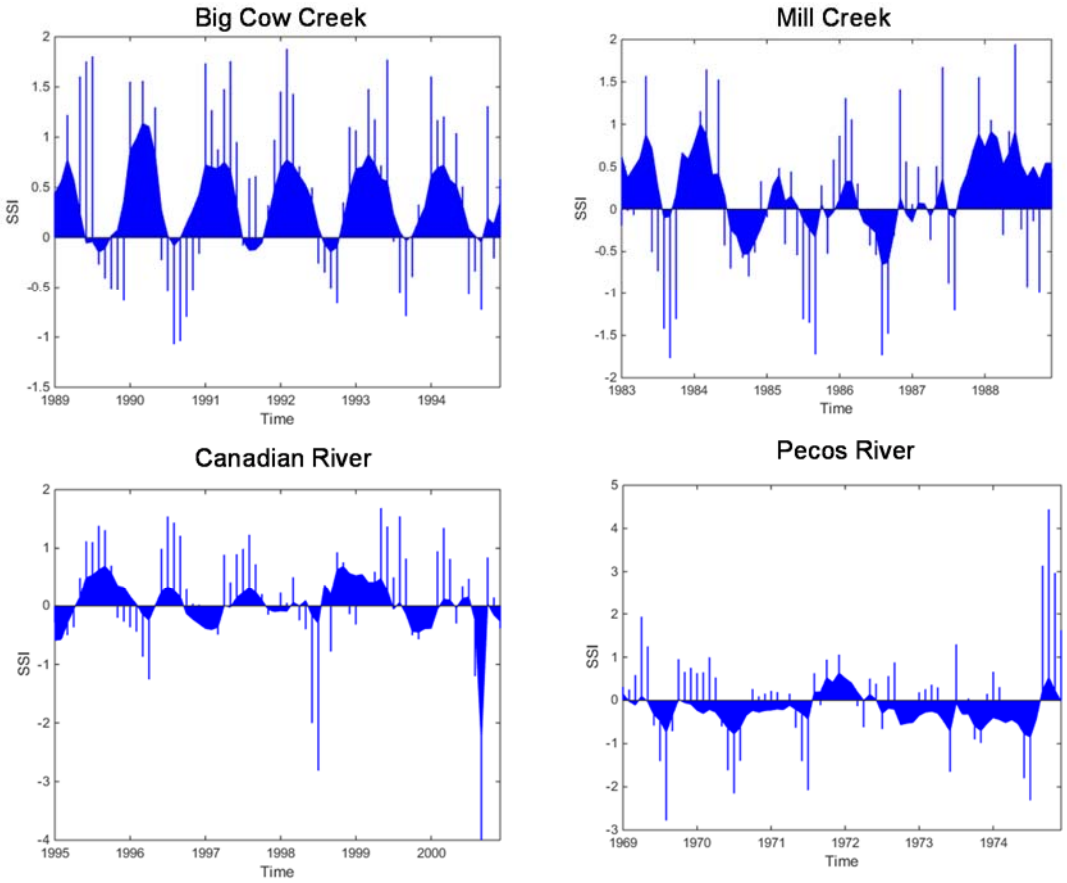


Figure 9-6 Forecasted SSI for Big Cow Creek, Mill Creek, Canadian River, and Pecos River

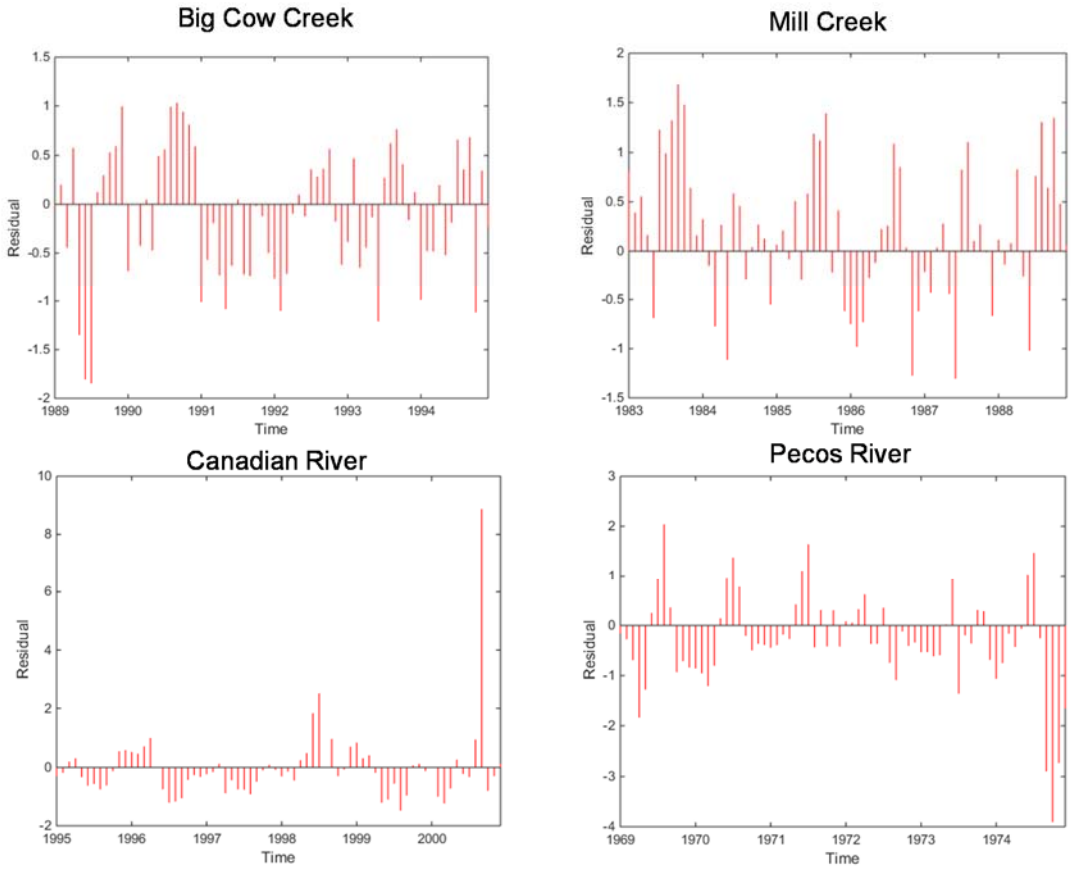


Figure 9-7 Residuals for Big Cow Creek, Mill Creek, Canadian River, and Pecos River

9.3.3 Comparison with univariate forecasting

To see how the proposed multichannel spectral analysis improves drought forecasting, forecasting of SSI with ENSO were compared to univariate forecasting without ENSO input. For comprehensive comparison, one El Nino condition and one La Nina condition were selected as shown in Figure 9-8.

The SSI forecasted for Big Cow Creek was under El Nino condition for 76% of lead time, which lasted from year 1990 to year 1994, as shown in Figure 9-8. Without

ENSO as an indicator, SSI was forecasted as a univariate series using the configurational entropy method developed in Section 5. The observed periodicity was correctly determined by the CESA method, and the forecasted SSI fluctuated with observations in the figure. However, the univariate forecasting yielded r^2 of 0.196 and NSE of 0.181. For ENSO, as an indicator incorporated in the model, the forecasted SSI had responded to observed El Nino phenomena. The drought forecasted by the proposed method was reduced as the ENSO indicated it was under El Nino condition. Thus, drought forecasted with ENSO were weaker than that without ENSO. Especially, during the summer of 1991, the negative SSI reached -0.5 by CE, which was corrected to near 0 with ENSO. As a result, the residuals were decreased by 60% during 1991, the r^2 increased to 0.369, and NSE to 0.258 with ENSO. However, the droughts during August to December in 1989, June to December in 1990, which showed less association with ENSO, were not forecasted by either of the methods.

On the contrary, the 91% of forecasting lead time was under La Nina in Spring Creek. It is seen in Figure 9-8, except for the first five months in 1995 and April 1997 to March 1998, other months were under the drought condition. Thus, forecasted SSI with ENSO enlarged the drought forecasted during the La Nina condition. As shown in Figure 9-8, the forecasted droughts during late 1999 and 2000 were increased with the ENSO indicator. On the other hand, for the El Nino condition during late 1997, the drought forecasted by CESA diminished after joining the ENSO indicator. Thus, forecasted residuals with ENSO were much smaller than without ENSO. As a result, r^2 increased from 0 to 0.285 and NSE was increased from -0.09 to 0.243 by multichannel option.

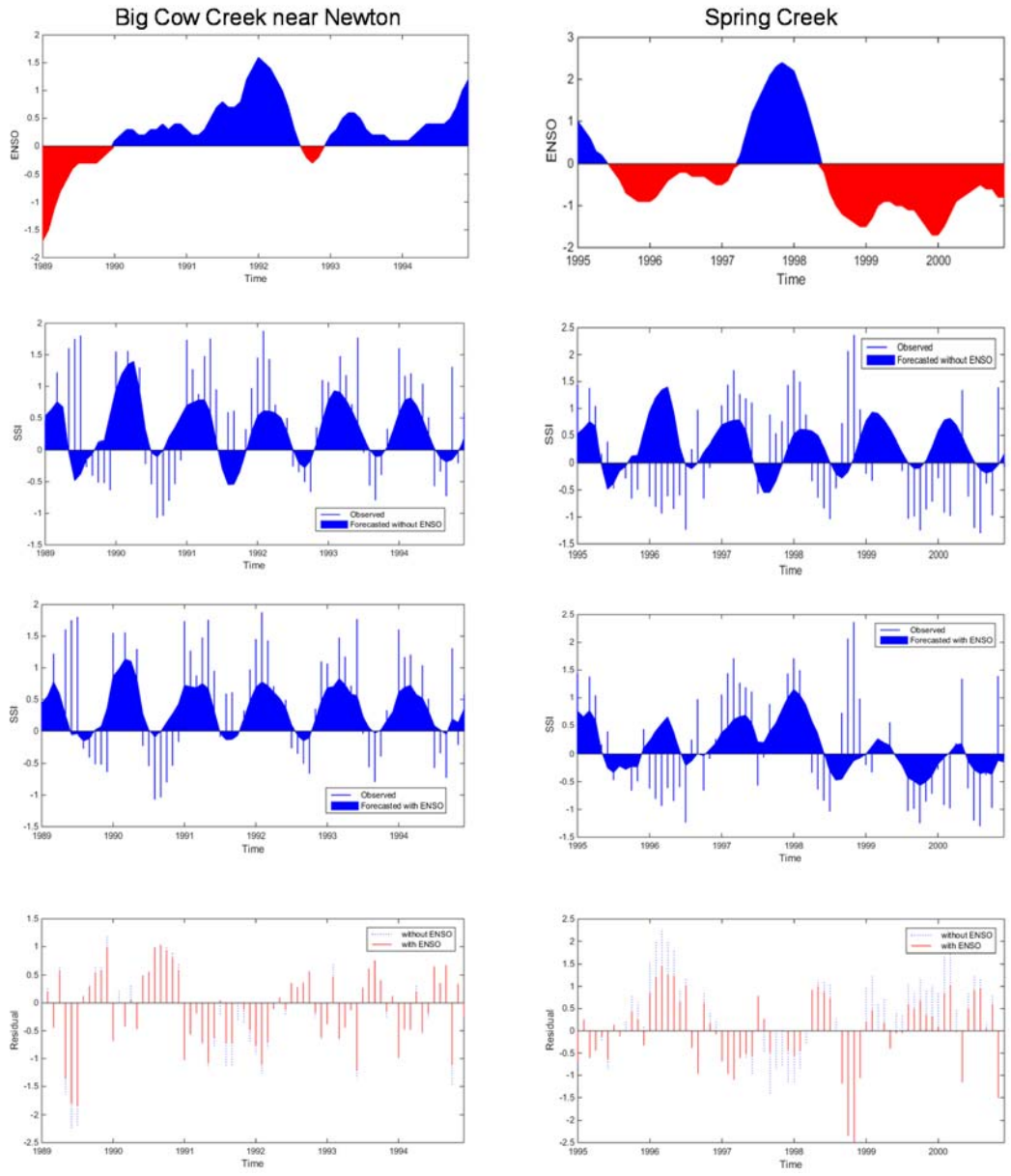


Figure 9-8 Forecasted SSI with and without ENSO

9.4 Summary

The multi-channel spectral analysis was developed based on the configurational entropy and applied to forecast long-term standardized streamflow index. When univariate

series is used, it is equivalent to the configurational entropy spectral analysis (CESA) developed in Section 5.

By transferring to SSI, the time series showed stationarity, and the 12 month periodicity appeared, which was weak in original streamflow data. Besides, the strong autocorrelation of SSI and strong cross-correlation with ENSO yield reliable forecasting of SSI.

Thus, SSI was successfully forecasted using multi-channel spectral analysis with ENSO as an indicator. The monthly drought was able to be forecasted for lead time of 4-6 years with NSE higher than 0.2. The forecasted SSI for rivers under subtropical humid condition were more reliable than other weather conditions for higher periodicity.

Comparison from the forecasting by CESA, it showed that proposed method did response to the climatic indicator. The forecasting under El Nino reduces the severity of drought, while that under La Nina increases the severity.

10 CONCLUSIONS

The study created in the dissertation builds a foundation for entropy theory for streamflow forecasting. CESAS and CESAF theory discussed in Sections 4 and 5 belong to the maximum entropy spectral analysis, while RESAS and RESAF developed in Sections 6 and 7 belong to the minimum relative entropy spectral analysis. The BESA, CESA, RESAS and RESAS theories, as a whole, complete the entropy theory for univariate time series forecasting. Besides, the cepstrum analysis applied in the study is first used in hydrology. The RESAF and CESA theory successfully link time series, spectral analysis and cepstrum analysis, and it provides a new approach for solving the autoregressive coefficient. These theories are verified by forecasting monthly streamflow and provide reliable results.

The conclusions drawn from the comparison of estimating spectral density and forecasting streamflow by different methods are summarized in Table 10-1. In general, MRESA theories, both RESAS and RESAF, require more information as the prior than the MESA theories including BESA and CESA. Besides, RESAS need longer historical data to assume the prior hypothesis than RESAF. The entropy theory of RESAS, RESAF or CESA need a larger model order than BESA, and the order used by all entropy theories are larger than AR. The spectral density estimation has a higher resolution for the MRESA theories than the MESA theories, and within two MESA theories CESA yields lower I-S distortion than BESA. Forecasted errors were smaller for MRESA than MESA, and using frequency as a random variable (RESAF or CESA) was smaller than that using spectral

power (RESAS or BESA). The forecasting lead time is found to be longer using MRESA than MESA, and longer using CESA than BESA. To sum up, the reliability of streamflow forecasting from high to low is RESAF, RESAS, CESA, BESA and then AR.

Table 10-1 Summary of different forecasting theories

Statistics	Order
Model order	AR< BESA< RESAS, CESA, RESAF
Itakaru-Saito distance	BESA> CESA> RESAS, RESAF
RMS	AR> BESA> CESA> RESAS > RESAF
NSE	AR< BESA< CESA< RESAS < RESAF
Lead time	AR< BESA< CESA< RESAS, RESAF
Information needed	BESA, CESA< RESAF< RESAS

The multichannel-spectral analysis is developed for long-term drought forecasting using configurational entropy, which allows entropy spectral analysis to be applied for multi-variate forecasting. The monthly SSI is successfully forecasted with ENSO as an indicator.

10.1 Recommendations

The proposed univariate entropy theories, though applied to monthly streamflow forecasting, are not restricted to monthly time scale, but may also be applied to daily, weekly and annual streamflows. They can also be applied to forecast other hydrological time series, such as rainfall, evapotranspiration and groundwater flow.

The multi-channel spectral analysis is developed based on the configurational entropy in this study. Other entropy theories for multi-channel can also be developed for multi-variate forecasting. The proposed multi-channel spectral analysis can be used to forecast streamflow with rainfall or flood with rainfall, and to see how it will work compared to traditional rainfall-runoff models. In such way, the failure to forecast streamflow due to sudden change may be mitigated.

Furthermore, the developed theories can be used for time series related to climate index both for univariate condition and multi-variate condition, and will be potentially useful to climate change analysis. ENSO is quantified by Nino 3.4 in the dissertation, however, there are SOI and MEI can also be used to represent ENSO. Except for ENSO, other climatic indicators, such as Pacific Decadal Oscillation, North Atlantic oscillation and North Atlantic Oscillation can be tested to forecast long-term drought or high streamflow for different regions. Besides, more than one indicator can be used together to forecast.

At last, the selection of model order is obtained based on AIC and BIC criteria in the study, while the uncertainty analysis can be conducted and be used to determine the model order.

NOTATIONS

a_j	prediction coefficient
b_m	backward prediction error
$C(n)$	cepstrum
$e(n)$	cepstrum of autocorrelation/ autocepstrum
$e_p(n)$	prior cepstrum of autocorrelation/ autocepstrum
$e_q(n)$	posterior cepstrum of autocorrelation/ autocepstrum
$G(f)$	power spectrum
H	entropy
N	length of given autocorrelation
$p(f)$	(prior) spectral density
$p(x_k)$	(prior) probability distribution of spectral power x_k
$p(\vec{x})$	(prior) joint probability distribution of spectral power x_k
$q(f)$	posterior spectral density
$q(x_k)$	posterior probability distribution of spectral power x_k
$q(\vec{x})$	posterior joint probability distribution of spectral power x_k
R	autocovariance
t	time interval
T	record length
W	Nyquist frequency
x_k	spectral power
$\vec{x}$	spectral power series

y	time series variable
$\bar{y}$	mean of time series observation
$\hat{y}$	forecasted value
$\Re$	autocorrelation matrix
σ	standard deviation of time series observation
υ_n	mean squared error
λ	Lagrange multiplier
ρ	autocorrelation

REFERENCES

Abbott, M.B., Bathurst, J.C., Cunge, J.A., Oconnell, P.E., Rasmussen, J., 1986. An introduction to the European Hydrological System - Systeme Hydrologique Europeen, SHE .2. Structure of a physically-based, distributed modeling system. J Hydrol, 87(1-2): 61-77.

Abrahart, R.J., See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrol Process, 14(11-12): 2157-2172.

Adamowski, J., Prokoph, A., 2014. Determining the amplitude and timing of streamflow discontinuities: A cross wavelet analysis approach. Hydrol Process, 28(5): 2782-2793.

Adamowski, J.F., 2008. River flow forecasting using wavelet and cross-wavelet transform models. Hydrol Process, 22(25): 4877-4891.

Andreo, B., Jimenez, P., Duran, J.J., Carrasco, F., Vadillo, I., Mangin, A., 2006. Climatic and hydrological variations during the last 117-166 years in the south of the Iberian Peninsula, from spectral and correlation analyses and continuous wavelet analyses. J Hydrol, 324(1-4): 24-39.

Baldwin, C.K., Lall, U., 1999. Seasonality of streamflow: The upper Mississippi River. Water Resour Res, 35(4): 1143-1154.

Barlow, M., Nigam, S., Berbery, E.H., 2001. ENSO, Pacific decadal variability, and US summertime precipitation, drought, and stream flow. Journal of Climate, 14(9): 2105-2128.

Behzad, M., Asghari, K., Eazi, M., Palhang, M., 2009. Generalization performance of support vector machines and neural networks in runoff modeling. Expert Syst Appl, 36(4): 7624-7629.

Berger, A., Melice, J.L., Van Der Mooren, I., Beran, A., Provenzale, A., Stanford, J.L., 1990. Evolutive Spectral Analysis of Sunspot Data over the past 300 Years [and Discussion]. Phil Trans R Soc A(330): 529-541.

Bergman, M.J., Delleur, J.W., 1985a. Kalman filter estimation and prediction of daily stream flows .1. review, algorithm, and simulation experiments. Water Resour Bull, 21(5): 815-825.

Bergman, M.J., Delleur, J.W., 1985b. Kalman filter estimation and prediction of daily stream flows .2. application to the Potomac River. Water Resour Bull, 21(5): 827-832.

Beven, K., Freer, J., 2001. A dynamic TOPMODEL. Hydrol Process, 15(10): 1993-2011.

Beven, K.J., Kirkby, M.J., Schofield, N., Tagg, A.F., 1984. Testing a physically-based flood forecasting-model (Topmodel) for 3 Uk catchments. J Hydrol, 69(1-4): 119-143.

Bishop, C.H., Etherton, B.J., Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. Mon Weather Rev, 129(3): 420-436.

Boshnakov, G.N., Lambert-Lacroix, S., 2012. A periodic Levinson-Durbin algorithm for entropy maximization. Comput Stat Data An, 56(1): 15-24.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J Roy Stat Soc B, 26(2): 211-252.

Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis; Forecasting and Control. Holden-Day series in time series analysis. Holden-Day, San Francisco,, 553 pp.

Brockwell, P.J., Davis, R.A., 1987. Time Series : Theory and Methods. Springer series in statistics. Springer-Verlag, New York, 519 pp.

Burg, J.P., 1967. Maximum entropy spectral analysis, Proceedings of 37th Meeting of Society Exploration Geophysics, pp. 34-41.

Burg, J.P., 1975. Maximum entropy spectral analysis. Ph.D. Thesis, Stanford University, 123 pp.

Burr, R.L., Lytle, D.W., 1986. A general-method of minimum cross-entropy spectral estimation - comments. Ieee T Acoust Speech, 34(5): 1324-1326.

Carlson, R.F., Maccormi.Aj, Watts, D.G., 1970. Application of linear random models to 4 annual streamflow series. Water Resour Res, 6(4): 1070-1078.

Cengiz, T.M., 2011. Periodic structures of Great Lakes levels using wavelet analysis. J Hydrol Hydromech, 59(1): 24-35.

Chiew, F.H.S., McMahon, T.A., 2002. Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability. Hydrolog Sci J, 47(3): 505-522.

Chiew, F.H.S., Peel, M.C., Western, A.W., 2002. Application and testing of the simple rainfall-runoff model SIMHYD. In: Singh, V.P., Frevert, D.K. (Eds.), Mathematical Models of Small Watershed Hydrlogy and Applications. Water Resour. Publ., Littleton, Colo.

Chiew, F.H.S., Piechota, T.C., Dracup, J.A., McMahon, T.A., 1998. El Nino Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. J Hydrol, 204(1-4): 138-149.

Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J., Uddstrom, M.J., 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Adv Water Resour, 31(10): 1309-1324.

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. J Hydrol, 375(3-4): 613-626.

Collomb, C., 2009. Burg's method, alogrithm and recursion.

Currie, R.G., 1973. Geomagnetic line spectra-2 to 70 years. Astrophysics and space science, 21: 14.

Dalezios, N.R., Tyraskis, P.A., 1989. Maximum-entropy spectra for regional precipitation analysis and forecasting. J Hydrol, 109(1-2): 25-42.

Day, G.N., 1985. Extended Streamflow Forecasting Using Nwsrfs. J Water Res Pl-Asce, 111(2): 157-170.

Durdu, O.F., 2010. Application of linear stochastic models for drought forecasting in the Buyuk Menderes river basin, western Turkey. Stoch Env Res Risk A, 24(8): 1145-1162.

Edward, J.A., Fitelson, M.M., 1973. Notes on maximum-entropy processing. Ieee T Inform Theory, 19(2): 3.

Elshorbagy, A., Simonovic, S.P., Panu, U.S., 2002. Estimation of missing streamflow data using principles of chaos theory. J Hydrol, 255(1-4): 123-133.

Endo, H., Randall, R.B., 2007. Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter. Mech Syst Signal Pr, 21(2): 906-919.

Fleming, S.W., Lavenue, A.M., Aly, A.H., Adams, A., 2002. Practical applications of spectral analysis to hydrologic time series. Hydrol Process, 16(2): 565-574.

Frausto-Solis, J., Pita, E., Lagunas, J., 2008. Short-term streamflow forecasting: ARIMA vs neural networks, American Conference on Applied Mathematics (MATH'08), Harvard, Massachusetts, USA., pp. 402-407.

Frieden, B.R., 1972. Restoring with maximum likelihood and maximum entropy. J Opt Soc Am, 62(4): 511-&.

Galeati, G., 1990. A comparison of parametric and non-parametric methods for runoff forecasting. Hydrolog Sci J, 35(1): 79-94.

Garen, D.C., 1992. Improved techniques in regression-based streamflow volume forecasting. J Water Res Pl-Asce, 118(6): 654-670.

Ghil, M., Allen, M.R., Dettinger, M.D., Ide, K., Kondrashov, D., Mann, M.E., Robertson, A.W., Saunders, A., Tian, Y., Varadi, F., Yiou, P., 2002. Advanced spectral methods for climatic time series. Rev Geophys, 40(1).

Gray, D.A., 1977. Maximum entropy spectrum analysis technique -A review of its theoretical properties, Weapons Research Establishment.

Gray, R.M., Gray, A.H., Rebolledo, G., Shore, J.E., 1981. Rate-distortion speech coding with a minimum discrimination information distortion measure. Ieee T Inform Theory, 27(6): 708-721.

Gull, S.F., Daniell, G.J., 1978. Image-reconstruction from incomplete and noisy data. Nature, 272(5655): 686-690.

Haltiner, J.P., Salas, J.D., 1988. Development and testing of a multivariate, seasonal ARMA(1,1) model. J Hydrol, 104(1-4): 247-272.

Hameed, S., 1984. Fourier-Analysis of Nile Flood Levels. Geophys Res Lett, 11(9): 843-845.

Hannan, E.J., Kavalieris, L., 1984. Multivariate linear time-series models. Adv Appl Probab, 16(3): 492-561.

Hasanean, H.M., 2001. Fluctuations of surface air temperature in the Eastern Mediterranean. Theor Appl Climatol, 68(1-2): 75-87.

Hipel, K.W., McLeod, A.I., 1994. Time series modelling of water resources and environmental systems, Developments in Water Science, v. 45. Elsevier, Burlington, pp. 1 online resource (1053 p.).

Hipel, K.W., Mcleod, A.I., Lennox, W.C., 1977. Advances in Box-Jenkins modeling .1. Model construction. Water Resour Res, 13(3): 567-575.

Jain, D., Singh, V.P., 1986. A comparison of transformation-methods for flood frequency-analysis. Water Resour Bull, 22(6): 903-912.

Jaynes, E.T., 1957a. Information theory and statistical mechanics .2. Phys Rev, 108(2): 171-190.

Jaynes, E.T., 1957b. Information theory and statistical mechanics.1. Phys Rev, 106(4): 620-630.

Jimenez, C., Mcleod, A.I., Hipel, K.W., 1989. Kalman filter estimation for periodic autoregressive-moving average models. Stoch Hydrol Hydraul, 3(3): 227-240.

Johnson, R.W., Shore, J.E., 1983. Which is better entropy expression for speech processing: -SLogS or LogS?, Naval Research laboratory, Washington, D. C.

Jones, R.H., Brelsfor, W.M., 1967. Time series with periodic structure. Technometrics, 9(1): 403-408.

Kahya, E., Dracup, J.A., 1993. United-States streamflow patterns in relation to the El-Nino Southern Oscillation. Water Resour Res, 29(8): 2491-2503.

Kalman, P.E., 1960. A new approach to linear filtering and prediction problems. ASME Transactions, Journal of Basic Engineering, 83D: 95-108.

Kalra, A., Miller, W.P., Lamb, K.W., Ahmad, S., Piechota, T., 2013. Using large-scale climatic patterns for improving long lead time streamflow forecasts for Gunnison and San Juan River Basins. Hydrol Process, 27(11): 1543-1559.

Karlsson, M., Yakowitz, S., 1987a. Nearest-neighbor methods for nonparametric rainfall-runoff forecasting. Water Resour Res, 23(7): 1300-1308.

Karlsson, M., Yakowitz, S., 1987b. Rainfall-runoff forecasting methods, old and new. Stoch Hydrol Hydraul, 1(4): 303-318.

Katsakos-Mavromichalis, N.A., Tzannes, M.A., Tzannes, N.S., 1985. Frequency resolution: A comparative study of four entropy methods. Kybernetes, 15(1): 25-32.

Khedun, C.P., Mishra, A.K., Bolten, J.D., Beaudoing, H.K., Kaiser, R.A., Giardino, J.R., Singh, V.P., 2012. Understanding changes in water availability in the Rio Grande/Rio Bravo del Norte basin under the influence of large-scale circulation indices using the Noah land surface model. J Geophys Res-Atmos, 117.

Kisi, O., 2009. Neural networks and wavelet conjunction model for intermittent streamflow forecasting. J Hydrol Eng, 14(8): 773-782.

Kisi, O., 2010. Wavelet regression model for short-term streamflow forecasting. J Hydrol, 389(3-4): 344-353.

Kitanidis, P.K., Bras, R.L., 1980a. Real-time forecasting with a conceptual hydrologic model .1. Analysis of uncertainty. Water Resour Res, 16(6): 1025-1033.

Kitanidis, P.K., Bras, R.L., 1980b. Real-time Forecasting with a conceptual hydrologic model .2. Applications and results. Water Resour Res, 16(6): 1034-1044.

Krause, P., Boyle, D.P., Base, F., 2005. Comparison of different efficiency criteria for hydrological model asssessment. Advances in Geosciences, 5: 89-97.

Krstanovic, P.F., Singh, V.P., 1989. An entropy based method for flood forecasting, New Directions for Surcace Water Modeling (Proceedings of the Baltimore Symposium, May 1989). IAHS Publ.

Krstanovic, P.F., Singh, V.P., 1991a. A univariate model for long-term streamflow forecasting .1. Development. Stoch Hydrol Hydraul, 5(3): 173-188.

Krstanovic, P.F., Singh, V.P., 1991b. A univariate model for long-term streamflow forecasting .2. Application. Stoch Hydrol Hydraul, 5(3): 189-205.

Krstanovic, P.F., Singh, V.P., 1993a. A real-time flood forecasting model based on maximum-entropy spectral analysis: I. Development. Water Resour Manag, 7(2): 109-129.

Krstanovic, P.F., Singh, V.P., 1993b. A real-time flood forecasting model based on maximum-entropy spectral analysis: II. Application. Water Resour Manag, 7(2): 131-151.

Kuhnel, I., Mcmahon, T.A., Finlayson, B.L., Haines, A., Whetton, P.H., Gibson, T.T., 1990. Climatic influences on streamflow variability - a comparison between Southeastern Australia and Southeastern United-States-of-America. Water Resour Res, 26(10): 2483-2496.

Kullback, S., 1959. Information Theory and Statistics. John Wiley and Sons, Inc.

Labat, D., 2005. Recent advances in wavelet analyses: Part I. A review of concepts. J Hydrol, 314(1-4): 275-288.

Labat, D., Ronchail, J., Guyot, J.L., 2005. Recent advances in wavelet analyses: Part 2 - Amazon, Parana, Orinoco and Congo discharges time scale variability. J Hydrol, 314(1-4): 289-311.

Legates, D.R., Davis, R.E., 1997. The continuing search for an anthropogenic climate change signal: Limitations of correlation-based approaches. Geophys Res Lett, 24(18): 2319-2322.

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res, 35(1): 233-241.

Levinson, N., 1946. The Wiener rms (root mean square) error criterion in filter design and prediction. J Math Phys Camb, 25(4): 261-278.

Liefhebber, F., Boekee, D.E., 1987. Minimum information spectral-analysis. Signal Process, 12(3): 243-255.

Lin, D.M., Wong, E.K., 1990. A survey on the maximum entropy method and parameter spectral estimation. Physics Reports, 193(2): 41-135.

Marques, C.A.F., Ferreira, J.A., Rocha, A., Castanheira, J.M., Melo-Goncalves, P., Vaz, N., Dias, J.M., 2006. Singular spectrum analysis and forecasting of hydrological time series. Phys Chem Earth, 31(18): 1172-1179.

Maurer, E.P., Lettenmaier, D.P., 2003. Predictability of seasonal runoff in the Mississippi River basin. J Geophys Res-Atmos, 108(D16).

McCabe, G.J., Dettinger, M.D., 1999. Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. Int J Climatol, 19(13): 1399-1410.

McCabe, G.J., Palecki, M.A., Betancourt, J.L., 2004. Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. P Natl Acad Sci USA, 101(12): 4136-4141.

McKee, T.B., Doesken, N., Kliest, J., 1993. The relationship of drought frequency and duration to time scales, Proceedings of the 8th conference on applied climatology, Anaheim, CA. American Meteorlogical Society, Boston, MA, pp. 179-184.

Mcleod, A.I., Hipel, K.W., Lennox, W.C., 1977. Advances in Box-Jenkins modeling .2. Applications. Water Resour Res, 13(3): 577-586.

Mishra, A.K., Desai, V., 2005. Drought forecasting using stochastic models. Stoch Env Res Risk A, 19(5): 326-339.

Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J Hydrol, 391(1-2): 204-216.

Modarres, R., 2007. Streamflow drought time series forecasting. Stoch Env Res Risk A, 21(3): 223-233.

Molenat, J., Davy, P., Gascuel-Odoux, C., Durand, P., 1999. Study of three subsurface hydrologic systems based on spectral and cross-spectral analysis of time series. J Hydrol, 222(1-4): 152-164.

Montanari, A., Rosso, R., Taqqu, M.S., 2000. A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. Water Resour Res, 36(5): 1249-1259.

Morid, S., Smakhtin, V., Bagherzadeh, K., 2007. Drought forecasting using artificial neural networks and time series of drought indices. Int J Climatol, 27(15): 2103-2111.

Nadeu, C., 1992. Finite length cepstrum modeling - a simple spectrum estimation technique. Signal Process, 26(1): 49-59.

Nadeu, C., Sanvicente, E., Bertran, M., 1981. A new algorthm for spectral esimation, International Conference on Digital Signal Processing, Florence, Italy, pp. 463-470.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through concptual models: Part1-A discussion of principles. J Hydrol, 10: 282-290.

Ning, L., Bradley, R.S., 2014. Winter precipitation variability and corresponding teleconnections over the northeastern United States. J Geophys Res-Atmos, 119(13).

Noakes, D.J., McLeod, A.I., Hipel, K.W., 1985. Forecasting monthly riverflow time series. International Journal of Forecasting, 1: 179-190.

Ooms, M., Franses, P.H., 2001. A seasonal periodic long memory model for monthly river flows. Environ Modell Softw, 16(6): 559-569.

Oppenheim, A.V., Schafer, R.W., 1975. Digital Signal Processing. Prentice-Hall, Englewood Cliffs, NJ.

Oppenheim, A.V., Schafer, R.W., 2004. From frequency to quefrency: A history of the Cepstrum. Ieee Signal Proc Mag, 21(5): 95-+.

Ortigueira, M.D., Garcia-Gomes, R., Tribolet, J.M., 1981. An iterative algorithm for maximum flatness spectral analysis, International Conference on Digital Signal Processing, Florence, Italy, pp. 810-818.

Ozger, M., Mishra, A.K., Singh, V.P., 2009. Low frequency drought variability associated with climate indices. J Hydrol, 364(1-2): 152-162.

Padmanabhan, G., Rao, A.R., 1988. Maximum-entropy spectral-analysis of hydrologic data. Water Resour Res, 24(9): 1519-1533.

Pagano, T.C., Garen, D.C., Perkins, T.R., Pasteris, P.A., 2009. Daily updating of operational statistical seasonal water supply forecasts for the western US. J Am Water Resour As, 45(3): 767-778.

Papademetriou, R.C., 1998. Experimental comparison of two information-theoretic spectral estimators, Signal Processing Proceedings, 1998. ICSP '98. 1998 Fourth International Conference, pp. 141-144 vol.1.

Pardo-Iguzquiza, E., Rodriquez-Tovar, F.J., 2006. Maximum entropy spectral analysis of climatic time series revisited: Assessing the statistical significance of estimated spectral peaks. J Geophys Res-Atmos, 111(D10).

Piechota, T.C., Dracup, J.A., 1996. Drought and regional hydrologic variation in the United States: Associations with the El Nino Southern Oscillation. Water Resour Res, 32(5): 1359-1373.

Pramanik, N., Panda, R.K., Singh, A., 2011. Daily river flow forecasting using wavelet ANN hybrid models. Journal of Hydroinformatics, 13(1): 49-63.

Prokoph, A., Adamowski, J., Adamowski, K., 2012. Influence of the 11 year solar cycle on annual streamflow maxima in Southern Canada. J Hydrol, 442: 55-62.

Redmond, K.T., Koch, R.W., 1991. Surface climate and streamflow variability in the Western United-States and their relationship to Large-Scale circulation Indexes. Water Resour Res, 27(9): 2381-2399.

Reichle, R.H., McLaughlin, D.B., Entekhabi, D., 2002. Hydrologic data assimilation with the ensemble Kalman filter. Mon Weather Rev, 130(1): 103-114.

Ropelewski, C.F., Halpert, M.S., 1986. North-American precipitation and temperature patterns associated with the Elnino Southern Oscillation (Enso). Mon Weather Rev, 114(12): 2352-2362.

Ruiz, J.E., Cordery, I., Sharma, A., 2007. Forecasting streamflows in Australia using the tropical Indo-Pacific thermocline as predictor. J Hydrol, 341(3-4): 156-164.

Salas, J.D., Boes, D.C., Smith, R.A., 1982. Estimation of ARMA models with seasonal parameters. Water Resour Res, 18(4): 1006-1010.

Salas, J.D., Obeysekera, J.T.B., 1982. ARMA model identification of hydrologic time-series. Water Resour Res, 18(4): 1011-1021.

Salas, J.D., Obeysekera, J.T.B., 1992. Conceptual basis of seasonal streamflow time-series models. J Hydraul Eng-Asce, 118(8): 1186-1194.

Sang, Y.F., Wang, D., Wu, J.C., Zhu, Q.P., Wang, L., 2009. The relation between periods' identification and noises in hydrologic series data. J Hydrol, 368(1-4): 165-177.

Sang, Y.F., Wang, Z.G., Liu, C.M., 2012. Period identification in hydrologic time series using empirical mode decomposition and maximum entropy spectral analysis. J Hydrol, 424: 154-164.

Schaefli, B., Gupta, H.V., 2007. Do Nash values have value? Hydrol Process, 21(15): 2075-2080.

Schroeder, M.R., 1982. Linear prediction, extermal entropy and prior information in speech signal analysis and synthesis. Speech Communication, 1(1): 9-20.

Sene, K., 2010. Hydrological forecasting, Hydrometeorology. Springer Netherlands, pp. 101-140.

Shannon, C.E., 1948. A mathematical theory of communication. At&T Tech J, 27(4): 623-656.

Sharma, A., 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 - A nonparametric probabilistic forecast model. J Hydrol, 239(1-4): 249-258.

Sharma, A., Luk, K.C., Cordery, I., Lall, U., 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 2 - Predictor identification of quarterly rainfall using ocean-atmosphere information. J Hydrol, 239(1-4): 240-248.

Shiri, J., Kisi, O., 2010. Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. J Hydrol, 394(3-4): 486-493.

Shore, J.E., 1979. Minimum cross-entropy spectral analysis, Naval Research laboratory, Washington, D. C.

Shore, J.E., 1981. Minimum cross-entropy spectral-analysis. Ieee T Acoust Speech, 29(2): 230-237.

Shore, J.E., Gray, R.M., 1982. Minimum cross-entropy pattern-classification and cluster-analysis. Ieee T Pattern Anal, 4(1): 11-17.

Shukla, S., Wood, A.W., 2008. Use of a standardized runoff index for characterizing hydrologic drought. Geophys Res Lett, 35(2).

Singh, V.P., 1988. Hydrologic Systems. Vol. 1: Rainfall-Runoff Modeling. Prentice Hall, Englewood Cliffs, New Jersey.

Singh, V.P., 1995. Computer Models of Watershed Hydrology. Water Resources Publications, Littleton, Colorado.

Singh, V.P., 1997. The use of entropy in hydrology and water resources. Hydrol Process, 11(6): 587-626.

Singh, V.P., 2011. Hydrologic synthesis using entropy theory: review. J Hydrol Eng, 16(5): 421-433.

Singh, V.P., 2013a. Entropy spectral analyses, Entropy Theory and its Application in Environmental and Water Engineering. John Wiley & Sons, Ltd, pp. 436-491.

Singh, V.P., 2013b. Minimum cross entropy spectral analysis, Entropy Theory and its Application in Environmental and Water Engineering. John Wiley & Sons, Ltd, pp. 492-516.

Singh, V.P., Frevert, D.K., 2002a. Mathmatical Modeling of Large Watershed Hydrology. Water Resources Publication, Littleton, Colorado, 891 pp.

Singh, V.P., Frevert, D.K., 2002b. Mathmatical Modeling of Small Watershed Hydrology and Applications. Water Resources Publication, Littleton, Colorado, 950 pp.

Singh, V.P., Frevert, D.K., 2006. Watershed Models. CRC Press, Boca Raton, Florida.

Singh, V.P., Jain, S.K., Tyagi, A., 2007. Entropy theory and its applications in risk analysis, Risk and Reliability Analysis, pp. 356-391.

Smith, L.C., Turcotte, D.L., Isacks, B.L., 1998. Stream flow characterization and feature detection using a discrete wavelet transform. Hydrol Process, 12(2): 233-249.

Smith, M.T., Reynolds, R.W., Peterson, T.C., Lawrimore, J., 2008. Improvement to NOAA's historical merged land-ocean surface temperature analysis (1980-2006). Journal of Climate, 21: 2283-2296.

Tabari, H., Nikbakht, J., Talaee, P.H., 2013. Hydrological drought assessment in northwestern Iran based on streamflow drought index (SDI). Water Resour Manag, 27(1): 137-151.

Tosic, I., Unkasevic, M., 2005. Analysis of precipitation series for Belgrade. Theor Appl Climatol, 80(1): 67-77.

Toth, E., Brath, A., Montanari, A., 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. J Hydrol, 239(1-4): 132-147.

Trenberth, K.E., 1997. The definition of El Nino. B Am Meteorol Soc, 78(12): 2771-2777.

Tzannes, M.A., Politis, D., Tzannes, N.S., 1985. A general method of minimum cross-entropy spectral estimation. Ieee T Acoust Speech, 33(3): 748-752.

Tzannes, N.S., Avgeris, T.G., 1981. A new approach to the estimation of continuous spectra. kybernetes, 10(2): 123-133.

Vicente-Serrano, S.M., Lopez-Moreno, J.I., Begueria, S., Lorenzo-Lacruz, J., Azorin-Molina, C., Moran-Tejeda, E., 2012. Accurate computation of a streamflow drought index. J Hydrol Eng, 17(2): 318-332.

Wang, D., Chen, Y.-F., Li, G.-F., Xu, Y.-H., 2004. Maximum entropy spectral analysis for annual maximum tide levels time series of the changjiang river estuary. Journal of Coastal Research(43): 101-108.

Wang, E.L., Zhang, Y.Q., Luo, J.M., Chiew, F.H.S., Wang, Q.J., 2011. Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data. Water Resour Res, 47.

Wang, Q.J., Robertson, D.E., 2011. Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. Water Resour Res, 47.

Wang, Q.J., Robertson, D.E., Chiew, F.H.S., 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. Water Resour Res, 45.

Wei, M.Z., Toth, Z., Wobus, R., Zhu, Y.J., Bishop, C.H., Wang, X.G., 2006. Ensemble transform Kalman filter-based ensemble perturbations in an operational global prediction system at NCEP. Tellus A, 58(1): 28-44.

Westra, S., Sharma, A., Brown, C., Lall, U., 2008. Multivariate streamflow forecasting using independent component analysis. Water Resour Res, 44(2).

Wood, A.W., Kumar, A., Lettenmaier, D.P., 2005. A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the western United States. J Geophys Res-Atmos, 110(D4).

Wood, A.W., Lettenmaier, D.P., 2008. An ensemble approach for attribution of hydrologic prediction uncertainty. Geophys Res Lett, 35(14).

Woodbury, A.D., 2004. A FORTRAN program to produce minimum relative entropy distributions. Comput Geosci-Uk, 30(1): 131-138.

Woodbury, A.D., Ulrych, T.J., 1993. Minimum relative entropy - forward probabilistic modeling. Water Resour Res, 29(8): 2847-2860.

Woodbury, A.D., Ulrych, T.J., 1996. Minimum relative entropy inversion: Theory and application to recovering the release history of a groundwater contaminant. Water Resour Res, 32(9): 2671-2681.

Woodbury, A.D., Ulrych, T.J., 1998. Minimum relative entropy and probabilistic inversion in groundwater hydrology. Stoch Hydrol Hydraul, 12(5): 317-358.

Woolhiser, D.A., Keefer, T.O., Redmond, K.T., 1993. Southern Oscillation effects on daily precipitation in the Southwestern United-States. Water Resour Res, 29(4): 1287-1295.

Wu, C.L., Chau, K.W., Li, Y.S., 2009. Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. Water Resour Res, 45.

Wu, N.L., 1983. An explicit solution and data extension in the maximum-entropy method. Ieee T Acoust Speech, 31(2): 486-491.

Yakowitz, S., 1987. Nearest neighbor method for time series analysis. J. Time Ser. Anal, 8(2): 235-247.

Yule, G.U., 1927. On a method of investigating periodicities in disturbed series, with special reference to Wofer's sunspot numbers. Philos T R Soc Lond, 226: 267-298.

Zaidman, M.D., Rees, H.G., Young, A.R., 2002. Spatio-temporal development of streamflow droughts in northwest Europe. Hydrol Earth Syst Sc, 6(4): 733-751.

Zhuang, X.H., Chen, L., Chen, S.S., 1993. An experimental comparison between maximum-entropy and minimum relative-entropy spectral-analysis. Ieee T Signal Proces, 41(4): 1730-1734.