

# Minimum relative entropy theory for streamflow forecasting with frequency as a random variable

Huijuan Cui<sup>1</sup>  · Vijay P. Singh<sup>2</sup>

Published online: 28 June 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** This paper develops a minimum relative entropy theory with frequency as a random variable, called MREF henceforth, for streamflow forecasting. The MREF theory consists of three main components: (1) determination of spectral density (2) determination of parameters by cepstrum analysis, and (3) extension of autocorrelation function. MREF is robust at determining the main periodicity, and provides higher resolution spectral density. The theory is evaluated using monthly streamflow observed at 20 stations in the Mississippi River basin, where forecasted monthly streamflows show the coefficient of determination ( $r^2$ ) of 0.876, which is slightly higher in the Upper Mississippi ( $r^2 = 0.932$ ) than in the Lower Mississippi ( $r^2 = 0.806$ ). Comparison of different priors shows that the prior with the background spectral density with a peak at 1/12 frequency provides satisfactory accuracy, and can be used to forecast monthly streamflow with limited information. Four different entropy theories are compared, and it is found that the minimum relative entropy theory has an advantage over maximum entropy (ME) for both spectral estimation and streamflow forecasting, if additional information as a prior is given. Besides, MREF is found to be more convenient to estimate parameters with cepstrum analysis than minimum relative entropy with spectral power as random variable

(MRES), and less information is needed to assume the prior. In general, the reliability of monthly streamflow forecasting from the highest to the lowest is for MREF, MRES, configuration entropy (CE), Burg entropy (BE), and then autoregressive method (AR), respectively.

**Keywords** Minimum relative entropy · Streamflow forecasting · Spectral density · Cepstrum analysis · Configurational entropy · Burg entropy

## 1 Introduction

The minimum relative entropy (MRE) spectral analysis, also called minimum cross-entropy spectral analysis, was developed by Shore (1979, 1981) with spectral power as a random variable. The MRE theory has shown higher resolution in spectral estimation, and is more accurate in detecting locations of spectral peaks than other spectral computation methods (Papademetriou 1998). Cui and Singh (2016a) applied MRE to monthly streamflow forecasting and showed its advantage over traditional autoregressive (AR) method or Burg entropy (BE) spectral analysis for both peak and low flow values with longer lead times. However, there is a minor drawback in the MRE theory 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). The selected exponential distribution as prior was tested by Cui and Singh (2016a) using 50–100 years of historical data, which may not always be available. To overcome the restriction on the prior, Tzannes et al. (1985) developed a general method of minimum relative entropy, where frequency was considered as a random variable. When a uniform prior is assumed, MRE reduces to the

✉ Huijuan Cui  
cuihj@igsnr.ac.cn

<sup>1</sup> Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11A Datun Road, Chaoyang District, Beijing 100101, People's Republic of China

<sup>2</sup> Department of Biological & Agricultural Engineering & Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843-2117, USA

configurational entropy (CE) theory (Frieden 1972; Gull and Daniell 1978). Thus, in this paper the general method of MRE theory will be developed for streamflow forecasting. To distinguish the two MRE theories in the series, the one used by Cui and Singh (2016a) will be referred to as MRES for S representing the spectral power, and the one in this paper developed with frequency as a random variable will be referred to as MREF henceforth.

It was found that MREF 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 (Girardin 2001; Liefhebber and Boeke 1987). Comparing CE to the Burg entropy (BE), the no prior case for MRES (Feng et al. 2013), it is shown that CE has higher resolution than BE for estimating the spectral density of ARMA and MA processes and two entropy theories are comparable for estimating the AR process (Nadeu et al. 1981). It was found that monthly streamflow forecasted by CE had higher reliability and longer lead time than BE (Cui and Singh 2015, 2016b). On the contrary, BE appears to be efficient for white noise, which is a random signal, as suggested by experiments on speech synthesis (Johnson and Shore 1983; Katsakos-Mavromichalis et al. 1985). One of the significant differences between MREF and MRES is the way of computing parameters. Combining MREF with cepstrum analysis, an explicit solution for estimating the Lagrange multipliers is developed (Burr and Lytle 1986; Wu 1983). Finite cepstrum analysis enables to construct a direct recursive relationship between the autocorrelation function and the cepstrum which is simpler than the Levinson algorithm used for the AR method (Nadeu 1992). However, there is no simpler way other than the Newton-Raphson method to solve the  $N$  nonlinear equations for computing the parameters for MRES. It is noted that cepstrum analysis is an efficient technique for separation and recovery of time series, as the homomorphic characteristics of time series are reminiscent of the cepstrum (Oppenheim and Schaffer 2004). Thus, it will be interesting to examine how it works for forecasting streamflow.

The objective of this paper therefore is to develop the MREF theory for streamflow forecasting with frequency as a random variable. The development of the theory contains four parts: (1) hypothesis on the prior spectral density (2) determination of the posterior spectral density, (3) estimation of the cepstrum, and (4) extension of the autocorrelation function. The paper is organized as follows. Providing a short introduction to the MREF theory in the first section, the second section introduces the development of MREF theory. Evaluation of the theory is discussed in the third section, followed by a discussion on MREF versus MRES in the fourth section. The paper is concluded in the

fifth section. Two appendices provide additional mathematical details of the study.

## 2 Development of minimum relative entropy theory

Let monthly streamflow time series  $y(t)$  be denoted as  $y_1, \dots, y_T$ , 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 limit 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. Each of these steps is now discussed.

### 2.1 Definition of 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 Kullback (1959), can be written as

$$H(q, p) = \int q(f) \ln[q(f)/p(f)] df \quad (1)$$

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 Eq. (1).

### 2.2 Specification of constraints

The minimization of the relative entropy needs to be subjected to constraints defined by autocorrelation, which can be formed in a straightforward manner. Using the Fourier transform relationship between the autocorrelation and the spectral density, the constraints can be written as

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

where  $\rho_r$  is the autocorrelation of lag  $r$ , which is estimated from observed time series,  $W$  is the Nyquist frequency,  $\Delta t$  is the sampling interval taken as 1 month in this paper, 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 1991).

It is noted that when  $r = 0$ , Eq. (2) reduces to

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

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

### 2.3 Estimation of spectral density

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

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

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

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

Thus, by rearranging Eq. (5), the posterior distribution can be formulated as

$$q(f) = p(f) \exp \left[ -1 - \sum_{r=-m}^m \lambda_r e^{i2\pi fr\Delta t} \right] \tag{6}$$

Equation (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

$$q(f) = \exp \left[ -1 - \sum_{r=-m}^m \lambda_r e^{i2\pi fr\Delta t} \right] \tag{7}$$

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

### 2.4 Computation of Lagrange multipliers

The Lagrange multipliers can be solved for numerically by inputting Eq. (6) into Eq. (2), which yields

$$\rho_r = \int_{-W}^W p(f) \exp \left[ -1 - \sum_{r=-m}^m \lambda_r e^{i2\pi fr\Delta t} \right] \exp(i2\pi ir\Delta t)df, \tag{8}$$

$-N \leq r \leq N$

Another way to determine the Lagrange multipliers is by cepstrum analysis (see Appendix 1 for details). Cepstrum, by definition, is the inverse Fourier transform of the logarithm of the estimated spectrum. It is a measure of the rate of change in the spectrum bands, and an efficient technique along with spectral analysis. Taking the inverse Fourier transform of the log-magnitude of Eq. (6), one obtains

$$\int_{-W}^W \{1 + \log[q(f)] - \log[p(f)]\} e^{i2\pi fn\Delta t} df = \int_{-W}^W \left( - \sum_{n=-N}^N \lambda_n e^{i2\pi fn\Delta t} \right) e^{i2\pi fn\Delta t} df \tag{9}$$

It can be seen from Eq. (9) that there are two terms relating to the spectral density that 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 defined as

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

and let  $e(n)$  denote the posterior cepstrum of autocorrelation as

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

Then, doing the integration of both sides of Eq. (9), one gets

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

where  $\delta_n$  is the delta function defined as:

$$\delta_n = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \tag{12}$$

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

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

Equation (13) enables to solve for the Lagrange multipliers in a more straight-forward manner than does non-linear Eq. (8). Thus, the Lagrange multipliers can be estimated from the summation of two cepstrums, the prior and posterior cepstrums. The prior cepstrum 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 Eq. (13) becomes the ones derived using the CE. On the other hand, the posterior cepstrum can be obtained from the given lags of  $N$  autocorrelation as

$$\begin{aligned} e(n) &= \int_{-W}^W \log q(f) e^{i2\pi fn\Delta t} df \\ &= \int_{-W}^W \log \left[ \sum_{r=-N}^N \rho_r e^{-i2\pi ir\Delta t} \right] e^{i2\pi fn\Delta t} df \end{aligned} \tag{14}$$

However, for a finite length of data, there is a recursive relation between the autocorrelation and the cepstrum developed by Nadeu (1992) to avoid the Fourier and inverse Fourier transforms. In his derivation, the causal part of autocorrelation, where  $0 < n \leq N$ , was used instead of  $-N \leq n \leq N$ . The details of derivation by Nadeu (1992) can be found in Appendix 2, which shows that the cepstrum can be estimated from the autocorrelation using

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

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

### 2.5 Extension of autocorrelation

It is shown in Eq. (15) that the autocorrelation and cepstrum are one to one related. On the contrary, the autocorrelation function can be estimated from the cepstrum using the inverse function of Eq. (15) as (Liefhebber and Boekee 1987; Nadeu 1992; Wu 1983):

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

It is seen from Eq. (16) that the autocorrelation of  $n$ -th lag can be estimated from  $n$  cepstrums. Thus, for the

autocorrelation beyond lag  $N$ , it can be extended one by one using Eq. (16), in which manner  $\rho_{N+1}$  is extended from  $\rho_1, \rho_2, \dots, \rho_N$  using  $N+1$  cepstrums and is used to extend to  $\rho_{N+2}$ , and so forth. Thus, the autocorrelation of  $N+k$  th lag can be estimated from  $N+k$  cepstrums as

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

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

$$\begin{aligned} e'(j) &= \int_{-W}^W \log q(f) e^{i2\pi fj\Delta t} df \\ &= \int_{-W}^W \log \left\{ p(f) \exp \left[ -1 - \sum_{r=-m}^m \lambda_r e^{i2\pi fr\Delta t} \right] \right\} e^{i2\pi fj\Delta t} df \end{aligned} \tag{18}$$

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

$$e'(j) = e_p(j) - \delta_{(j)} - \lambda_{(j)} \tag{19}$$

For different  $j$  values, the posterior cepstrum becomes

$$e'(j) = e_p(0) - 1 - \lambda_0, \quad \text{when } j = 0 \tag{20a}$$

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

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

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

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

where  $m$  is the order of model. Equation (20) is the extension of autocorrelation beyond given lag  $N$  using MREF. It is noted from Eq. (21) that when no prior is given, the first term diminishes and yields that derived from CE.

### 2.6 Forecasting

Streamflow is forecasted in the manner that autocorrelation function is extended. It can be seen from Eq. (21) that autocorrelation of lag  $N+k$  is extended from previous  $m$  autocorrelations combined with  $m$  cepstrums of autocorrelations. In the same manner, streamflow can be forecasted from previous  $m$  observations using  $m$  cepstrums using Eq. (21) (Oppenheim and Schaffer 1975; Wu 1983).

Thus, to forecast streamflow, Eq. (21) can be written in time series, where the input data changes to  $y_t$  and cepstrum  $c(n)$  is used instead of  $e(j)$ :

$$y_{T+k} = \frac{c_p(T+k)}{2} + \sum_{j=1}^m \frac{k}{T+k} c'(j)y(T+k-j) \tag{22}$$

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

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

It is noted that  $e(n) = 2C(n)$ (see Appendix 1), one can write Eq. (22) as

$$\hat{y}_{T+k} = \frac{1}{4}e_p(N+k) + \frac{1}{2} \sum_{j=1}^m \frac{k}{T+k} e'(j)y(T+k-j) \tag{24}$$

When a uniform prior is assumed,  $e_p$  is 0, thus, Eq. (24) reduces to

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

which is equivalent to streamflow forecasted by Cui and Singh (2015). 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).

### 3 Evaluation

The same 20 sets of streamflow data from the Mississippi River watershed, as used in the paper of Cui and Singh (2016a), were used in this paper. The Mississippi River covers the drainage area of over 2,981,000 km<sup>2</sup>, thus, it is divided as the Upper Mississippi (UM) River and the Lower Mississippi (LM) River. There are 11 stations chosen from the UM and 9 from the LM with the drainage area varying from 466 km<sup>2</sup> to 2,915,858 km<sup>2</sup>. As discussed in Cui and Singh (2016a), the most significant periodicity of the Mississippi River is 12 months, as well as less significant periodicities of 3, 4 and 6 months. The periodicity of each station was tabulated as shown in Table 1, which is the basis for choosing the prior spectral density. It is noted that 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. To apply the MREF theory, one needs to hypothesize a prior based on the streamflow periodicity first.

### 3.1 Hypothesis on prior spectral density

The prior hypothesis was formulated for the distribution of spectral power at each frequency when deriving MRES, and long records of historical data are needed for a proper hypothesis. However, in this paper, the periodicity of streamflow is the only information needed for the hypothesis on the spectral density. It is noted in Table 1 that streamflow in the Mississippi River watershed has periodicities of 12, 6, 4, and 3 months with different degrees of significance. Thus, the prior spectral density was assumed as background for the six types of prior hypothesis, as stated in Table 2 and plotted in Fig. 1.

Prior 1 is defined by a condition when no prior information is given, thus, the spectral density is constantly 1. In this case, the MREF is equivalent to the CE 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 is 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 has two significant periodicities at frequencies 1/12 and 1/6, however, the spectral peak at frequency 1/6 is assumed to be 0.6 not 1 for lower significance than at frequency 1/12. Based on observed streamflow periodicity, the suggesting priors are listed in the last column in Table 1. It is noted that streamflow observed in the main Mississippi River is prone to Prior 3. Prior 4 and Prior 5 correspond to adding small spectral peaks at the frequency of 1/3 or 1/4, which can be applied in the middle of the Mississippi River watershed. The last prior spectral density is composed of four peaks at all possible periodic frequencies, which is observed for Minnesota River (MN) and upstream Ouachita (OUU) River.

### 3.2 Spectral density

For assumed six priors, the posterior spectral density was estimated by the MREF theory. The estimated spectral density was verified with the one obtained from the fast Fourier transform (FFT) through the Itakura-Saito (I-S) distortion, which is a divergence measure between two spectral densities. The smaller value of Itakura-Saito distortion represents a higher agreement with FFT. It is noted in Table 3 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. This suggests that prior information improves the resolution of spectral estimation.

Comparing the Itakura-Saito distortion of different priors, the one having the least distortion (Prior\*) is tabulated

**Table 1** Periodicity of streamflow at each station and suggested prior

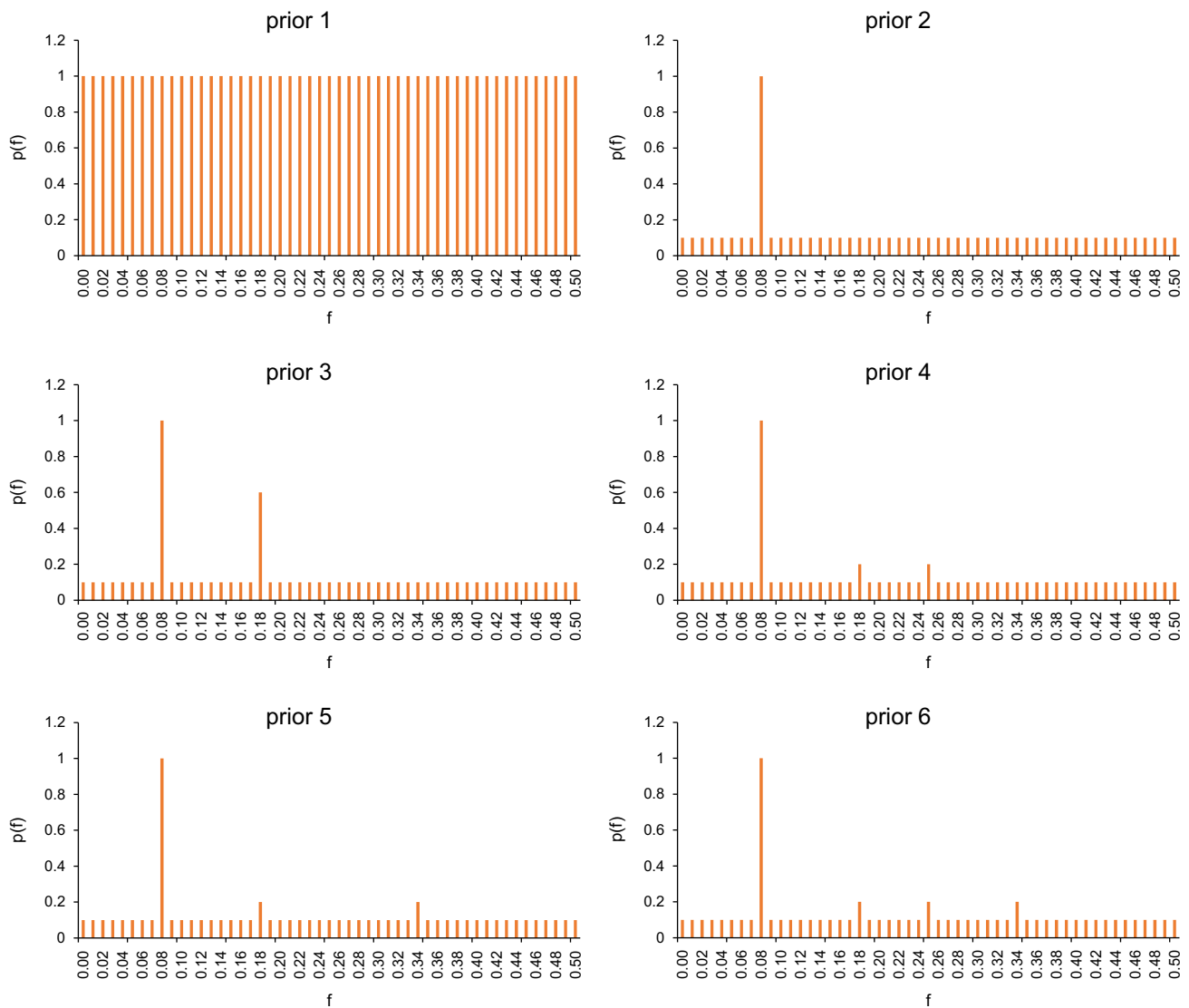
Name	Location	Station	Periodicity	Suggesting prior
UMU	Upper Mississippi upstream	05227500	12, 6, (4, 3)	Prior 3
MNU	Minnesota upstream	05301000	12, 6, 4, 3	Prior 6
MND	Minnesota downstream	05330000	12, 6, 4, 3	Prior 6
UMC	Upper Mississippi center	05420500	12, 6	Prior 3
IU	Iowa upstream	05449500	12, 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
OOU	Ouachita upstream	07359002	12, 6, 4, 3	Prior 6
OOD	Ouachita downstream	07367005	12	Prior 2
LMD	Lower Mississippi downstream	07374000	12	Prior 2

**Table 2** 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	$p(f) = 0.01, p(1/12) = 1, p(1/6) = 0.2, p(1/4) = 0.2, p(1/3) = 0.2$

in the last column of Table 3. It is interesting to note that the Prior\* with the least I-S distortion is not exactly the same as that suggested in Table 1 from observations. For example, the periodicity observed for the Minnesota River (MN) suggested Prior 6 for four spectral peaks at 1/12th, 1/6th, 1/4th and 1/3rd frequencies in Table 1. However, after verification, it reduced to Prior 2 with uni-peak for the upstream reach (MNU) and Prior 5 with three peaks for the downstream reach (MND). It implies that the small peaks observed at 3 or 4 months periodicity were not that significant for the MN. Besides, Prior 6 was not suggested by any of the cases, which implies that none of the streamflow series 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 months 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 Fig. 2. Especially for the Lower Mississippi, Prior 2 led to the least I-S distortion. It can also be seen from Fig. 3 where the estimated spectral density with Prior 1, Prior 2 and Prior\* are plotted against the one from the FFT. For all the priors, the 1/12th 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 was wider than observed. For example, the area under the 1/12th peak of spectral density estimated for MND was 2



**Fig. 1** Six types of prior hypotheses

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 1/6th frequency for UMD, more than 10 times the observed peak at the 1/4th frequency for MS. Nevertheless, the Prior\* had estimated a slightly smaller I-S distortion, the estimation

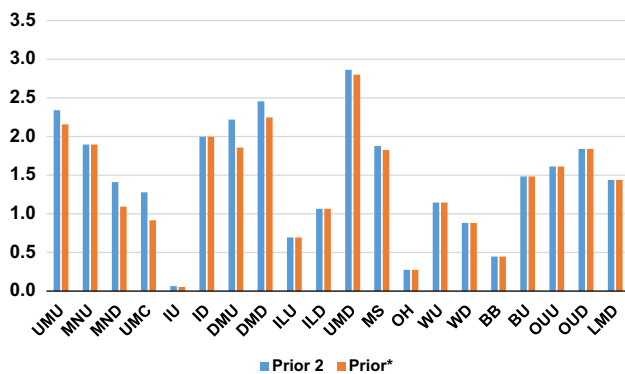
with Prior 2 provided for a satisfactory resolution. Furthermore, when forecasting streamflow with the lack of periodicity information, Prior 2 can be a proper choice as monthly streamflow possesses a strong 12 months periodicity.

### 3.3 Forecasting results

With the selected prior, streamflow was forecasted, as plotted in Fig. 4, while the goodness of fit was evaluated by the root mean square error (*RMSE*), the coefficient of determination ( $r^2$ ), and the Nash-Sutcliffe efficiency coefficient (*NSE*) as tabulated in Table 4. It is noted that the MREF theory was capable of forecasting streamflow with  $r^2$  of 0.876; in particular it is 0.932 for the Upper Mississippi (UM) and 0.806 for the Lower Mississippi (LM). The

**Table 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



**Fig. 2** Comparison of the Itakura-Saito distortions for Prior 2 and Prior\*

higher accuracy obtained for forecasting streamflow in the UM than that in the LM is related to the more regular and repeatable streamflow hydrograph, timing and amplitude of the peaks in the UM than in the LM. For example, it can be seen from the streamflow of White River located in the LM in Fig. 4 that the hydrograph, timing of the 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 years lead time.

Streamflow forecasted with Prior 1, Prior 2, and Prior\* are also plotted in Fig. 4. Without any prior information,

streamflow forecasted with Prior 1 is smoother than with other priors, and the shape of hydrograph kept the same from year to year. Thus, when observed streamflow is regular, repeatable, streamflow forecasted with Prior 1 was as good as with other priors (see UMC, UMD and MS). However, streamflow forecasting with Prior 1 is weak for an irregular change in hydrograph or a multi-peak flow. It can be seen from streamflow forecasted for LMD that Prior 1 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 Prior 1. For the above cases, the other prior information is recommended. It can be seen that Prior 2 was capable of forecasting multi-peaks that fluctuated during the peak season of LMD and WU, and to forecast the small rise for UMU and MND. Furthermore, streamflow forecasted during the low flow period was more reliable for choosing a prior other than Prior 1. It can be seen from the figure that streamflow forecasted with Prior 1 during September to February was  $2.7 \text{ m}^3/\text{s}$  to  $20.8 \text{ m}^3/\text{s}$  higher than the observation, while for other prior the difference was less than  $5.4 \text{ m}^3/\text{s}$ . Besides, the peak flow forecasted with Prior 1 dropped by 5.6–13.5 % year by year. Due to the drop in the forecasted peak of Illinois upstream (ILU), the forecasted  $NSE$  of the second lead year



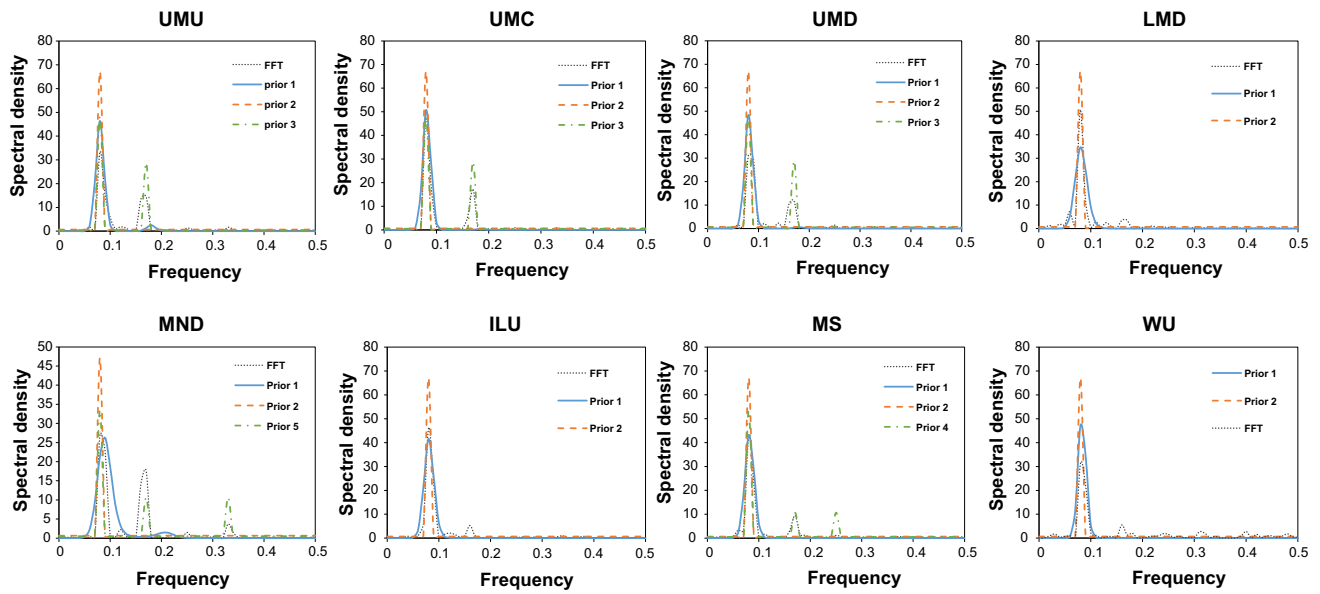


Fig. 3 Estimated spectral densities with different priors

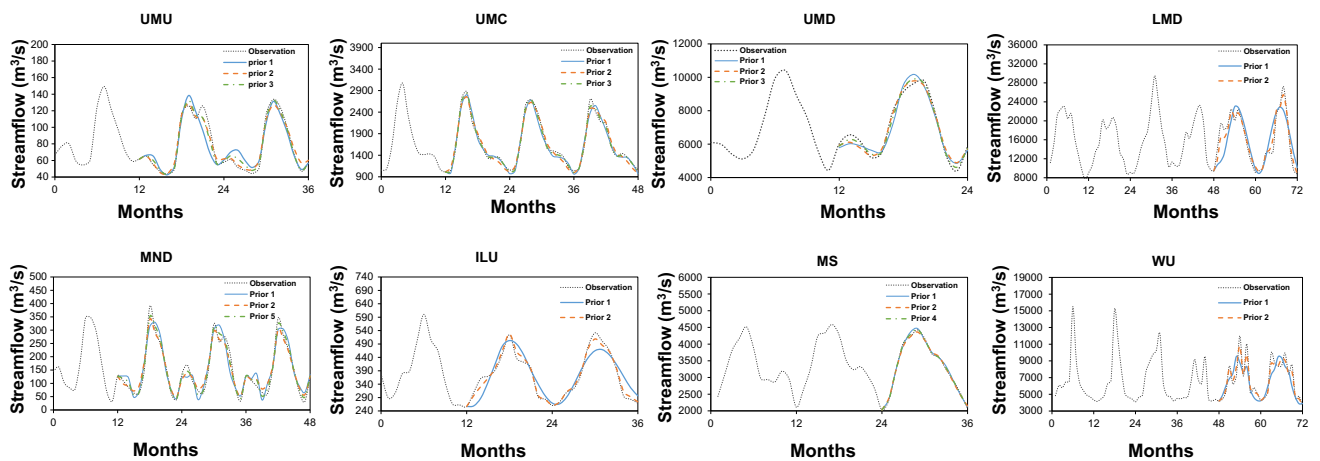


Fig. 4 Forecasted streamflow with Prior 1, Prior 2 and Prior\*

decreased from 0.769 of the first lead year to 0.547, when the forecasted peak from Prior 2 was within 4.3 % variation.

Thus, it can be noted that additional prior information improved forecasting, while which prior to be chosen is not that significant. Again, as shown in Fig. 4, streamflows forecasted by Prior 2 and Prior\* were close to each other, and the average values of *NSE* for the two priors were 0.77 and 0.78, respectively. It can be seen from Fig. 5, where the values of *NSE* estimated with Prior 2 and Prior\* are plotted, that the difference between the two priors is minimal. It is 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. This suggests that the minimal advantage of choosing Prior\* over Prior 2 in spectral estimation

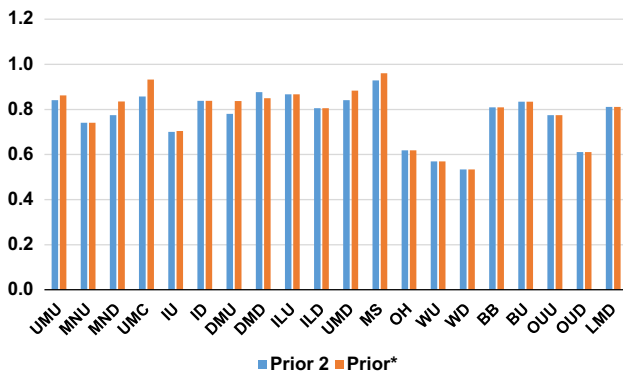
does not carry over to streamflow forecasting. Besides, for Lower Mississippi, Prior 2 is the only choice for forecasting. Thus, the main periodicity of 12 moth is reiterated and Prior 2 is a proper assumption for estimating the spectral density and forecasting streamflow of the Mississippi River.

### 4 Discussion

The estimated spectral density and forecasted streamflow showed that adding prior information generally yielded higher resolution and accuracy. It was shown by Cui and Singh (2016a) that the MRES theory had higher resolution in estimating the spectral density and forecasted with more

**Table 4** Forecasted result of MREF

Name	RMSE (m <sup>2</sup> /s)	r <sup>2</sup>	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
OOU	5.120	0.782	0.774
OOD	55.651	0.749	0.611
LMD	1179.229	0.950	0.811



**Fig. 5** Comparison of NSE for Prior 2 and Prior\*

accuracy with longer lead time than the BE theory. It can be summarized that no matter which variable is defined as a random variable, the prior information improves the MRE theory over the maximum entropy (ME) 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 MRES and BE theories yield the spectral density in the form of inverse polynomials, where the roots are the forecasting coefficients. In this case, the coefficients

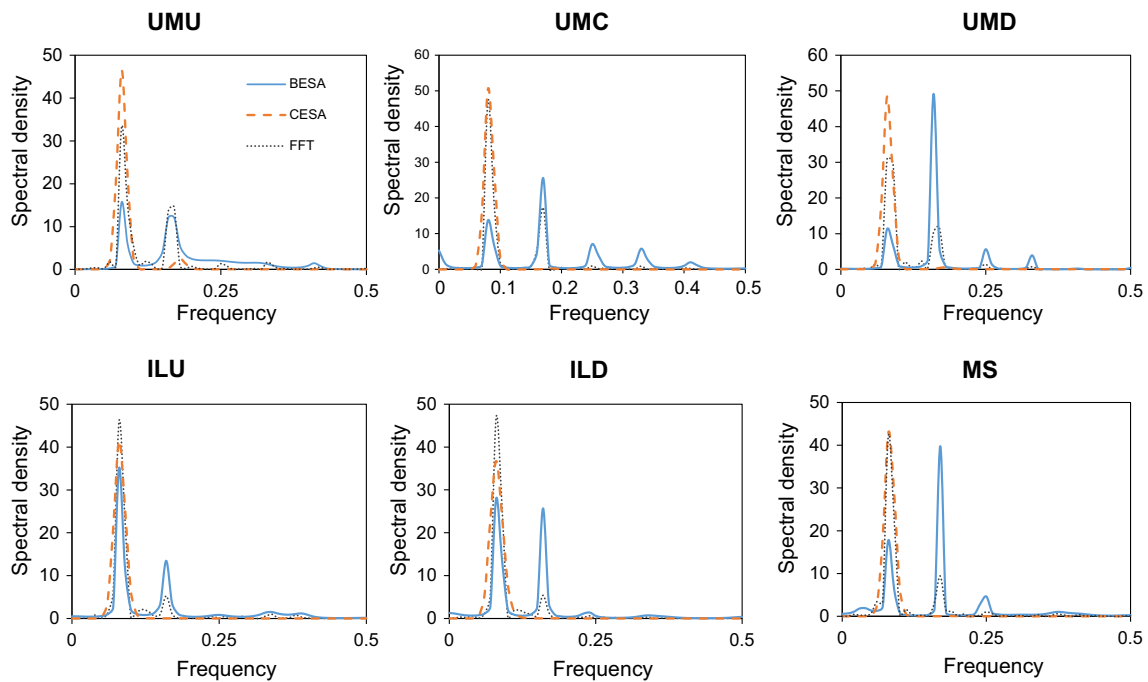
are solved from  $N$  nonlinear equations, and streamflow is forecasted by linear prediction. On the other hand, when the random variable is defined by frequency, the spectral density is in the form of exponential function for MREF and CE. Taking the inverse Fourier transform of the logarithm of spectral density, which is the cepstrum of the autocorrelation, the Lagrange multipliers can be directly solved from the cepstrum. The extension of autocorrelation and forecasting of streamflow, in this case, are processed from the recursive function associated with the cepstrum. Furthermore, the prior is hypothesized for the distribution of spectral power for MRES while for the spectral density for MREF. Let us now compare how much difference the two entropy definitions make.

**4.1 Without prior information**

Without considering the impact of prior information, we first compare BE and CE. It is found from the I-S distortion tabulated in Table 5 that the I-S distortion is higher for BE than CE, which suggests that the resolution in estimating the spectral density was higher by CE than by BE. It can be seen from Fig. 6 that the spectral density estimated from BE 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 CE.

**Table 5** Computed Itakura-Saito distortion for different methods

Name	MREF	MRES	CE	BE
UMU	2.157	2.919	6.155	13.296
MNU	1.894	1.187	41.527	34.180
MND	1.091	1.606	3.688	9.475
UMC	0.916	0.942	6.243	13.356
IU	0.054	0.071	1.513	5.323
ID	1.997	2.510	5.504	10.754
DMU	1.856	1.995	19.464	25.579
DMD	2.247	2.904	8.956	8.570
ILU	0.692	0.505	26.459	20.933
ILD	1.065	1.356	27.722	33.325
UMD	2.801	1.917	16.128	49.393
MS	1.825	1.981	2.010	2.644
OH	0.274	0.235	2.799	2.783
WU	1.144	0.429	1.962	2.062
WD	0.881	0.631	158.595	563.501
BB	0.446	0.472	4.820	4.056
BU	1.481	1.250	26.210	33.181
OOU	1.612	0.651	39.996	16.724
OOD	1.839	0.873	1.793	1.958
LMD	1.439	0.922	0.896	1.547



**Fig. 6** Comparison of spectral densities by BE and CE

However, using BE, the largest peak shifted, for example, the spectral peak estimated for UMC, UMD, and MS in Fig. 6, shifted to frequency 1/6 but was 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 BE the same as for CE, the second largest peak estimated at frequency 1/6 by BE was 2.3 and 4.7 times larger than the one from FFT, respectively. However, CE ignored small peaks to make sure the largest peak was correctly estimated, thus, the spectral density obtained by CE was uni-peak with periodicity at frequency 1/12.

The advantage of estimating the location of spectral peaks by CE continued in streamflow forecasting. It can be seen from Fig. 7 that graphs streamflow forecasted for the two ME methods. Due to the shift in the estimation of spectral density for UMC, UMD and MS, the peak flow forecasted by BE did not match the observations. The peak flow forecasted by BE was 1 month late for the third lead year for UMC and 1 month early for UMD and MS. CE can capture the peak flow correctly repeated every 12 months, as CE strongly emphasized the 12 months periodicity for spectral estimation. Besides, the forecasting lead time for UMU and UMC is longer for CE than BE. As shown in Fig. 7, the peak flow was forecasted with error less than 0.1 by CE for two year lead time, but the error by BE was larger than 0.3 for the forecasted peak second lead year. It can be seen from Fig. 8, where the forecasted errors were plotted against lead time, that forecasted error by BE was earlier to exceed 0.2 than by CE. However, both BE

and CE 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 BE was 21.1 % higher than the observation, while that of CE was 19.3 % higher than the observation. It is noted from Fig. 8 that the absolute values of errors by BE were larger than 0.3 during the low flow season compared to those of around 0.2 by CE. As listed in Table 6, CE uses orders of 1–2 more for streamflow forecasting than BE, especially for the Lower Mississippi, where BE failed to provide reliable forecasting with a model order of 2. CE was generally applicable to forecast streamflow for the whole watershed with *NSE* of 0.604 compared to that of 0.525 for BE as shown in Table 7.

#### 4.2 With prior information

The difference in the spectral estimation between MRES and MREF was similar to that between BE and CE. It can be seen from Fig. 9 that the spectral density estimated by MRES is more likely to be multi-peak spectra compared to MREF, though the location of small peaks may not exactly be consistent with the ones from FFT. As shown in the figure, the spectral density estimated by MRES 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 MREF had peaks at the specified frequency from the prior and seldom had peaks additional to the one from FFT. Like the

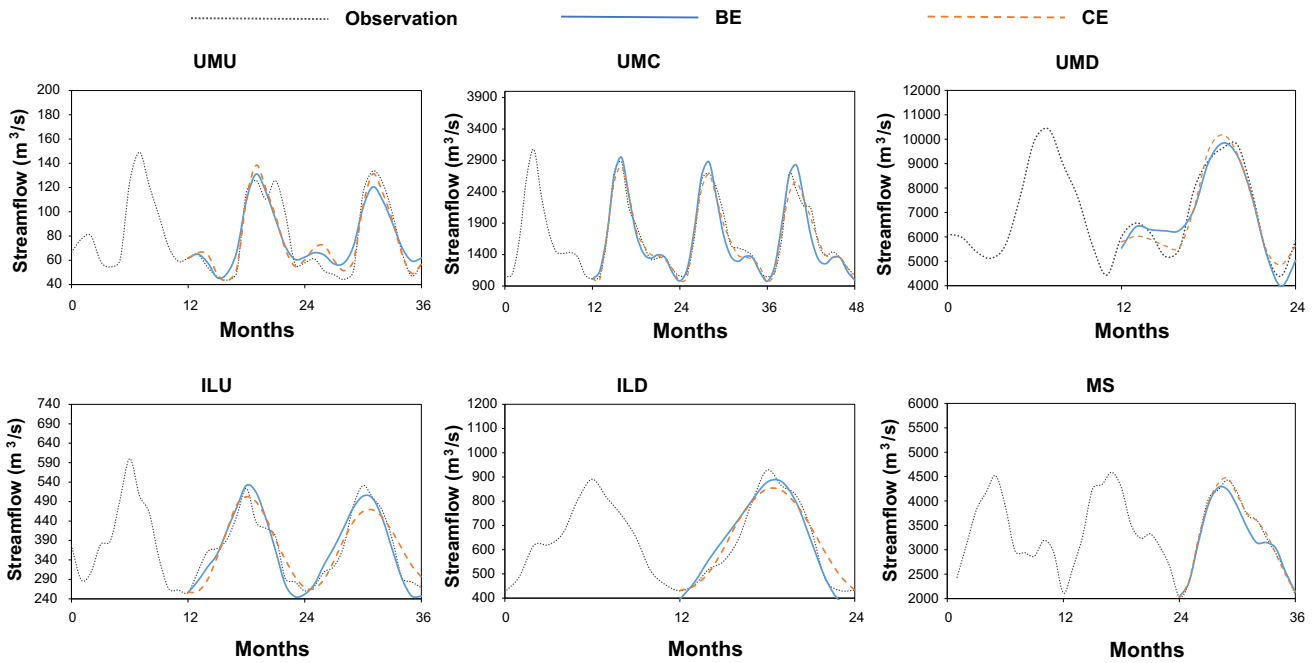


Fig. 7 Comparison of streamflow forecasting by BE and CE

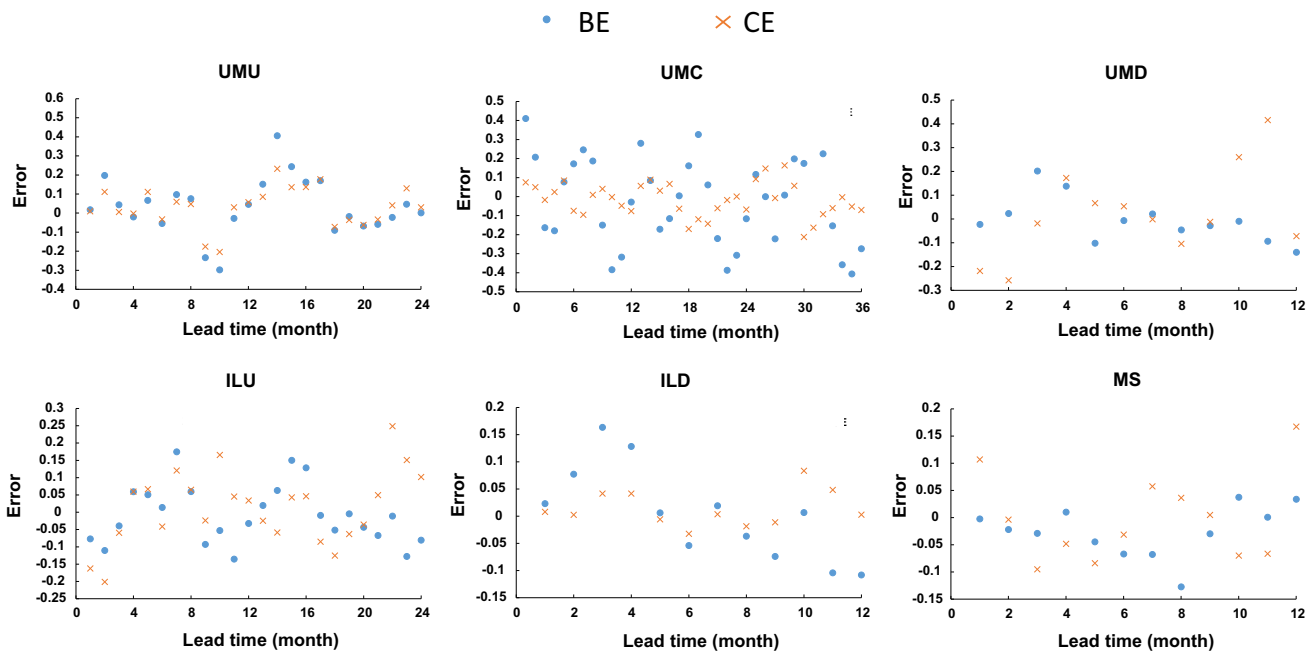


Fig. 8 Forecasted errors by BE and CE

maximum entropy case, MREF was more likely to have the largest peak at the correct location. However, the capability of detecting multi-peaks of MRES did not guarantee a smaller I-S distortion. As shown in Table 7, the values of I-S distortion of the two methods were similar, where the average value of I-S distortion was 1.386 for MRES and 1.268 and MREF. The number of counts that the I-S

distortion was smaller for MREF vs. MRES was 10:10, as plotted in Fig. 10. It was smaller than that by CE vs. BE of 14:6, which suggests that the advantage of estimating the spectral density using frequency as a random variable is more significant for applying ME than MRE.

The model orders used for streamflow forecasting by the two entropy methods were similar, as shown in Table 6,

**Table 6** Model orders for different methods

Name	MREF	CE	MRES	Burg	AR
UMU	13	13	13	13	8
MNU	13	13	12	12	8
MND	13	12	12	13	6
UMC	14	13	12	12	8
IU	13	12	13	13	13
ID	13	14	14	14	13
DMU	14	14	13	14	14
DMD	14	14	14	13	13
ILU	15	14	14	12	5
ILD	16	16	15	14	3
UMD	16	15	15	14	11
MS	16	15	15	15	10
OH	16	16	16	15	13
WU	18	18	18	6	1
WD	10	9	8	9	2
BB	16	15	16	16	8
BU	16	16	15	15	4
OUU	14	13	15	2	2
OOD	10	11	8	2	2
LMD	12	12	11	11	2
Average	14.1	13.75	13.45	11.75	7.3

**Table 7** Computed *NSE* for different methods

Name	MREF		MRES	CE	BE	AR
	Prior 2	Prior*				
UMU	0.841	0.862	0.724	0.725	0.697	0.556
MNU	0.741	0.741	0.702	0.492	0.484	0.381
MND	0.774	0.835	0.671	0.585	0.554	-0.062
UMC	0.857	0.932	0.816	0.785	0.737	0.598
IU	0.800	0.804	0.608	0.498	0.475	0.416
ID	0.888	0.888	0.829	0.548	0.544	0.377
DMU	0.780	0.837	0.674	0.501	0.476	0.609
DMD	0.877	0.850	0.764	0.653	0.700	0.741
ILU	0.867	0.867	0.788	0.658	0.658	0.097
ILD	0.905	0.905	0.783	0.755	0.709	0.016
UMD	0.841	0.884	0.814	0.787	0.752	0.622
MS	0.929	0.960	0.920	0.790	0.783	0.745
OH	0.618	0.618	0.645	0.540	0.539	0.476
WU	0.569	0.569	0.545	0.432	0.179	0.030
WD	0.533	0.533	0.544	0.498	0.274	0.088
BB	0.809	0.809	0.624	0.481	0.543	0.334
BU	0.834	0.834	0.869	0.670	0.623	0.329
OUU	0.774	0.774	0.748	0.427	0.004	0.004
OOD	0.611	0.611	0.621	0.503	0.071	0.069
LMD	0.811	0.811	0.874	0.746	0.708	0.060

which suggests that the capability of possessing the past information is similar for the two methods. Thus, the difference is how the two methods deal with that information to forecast streamflow. It is seen from the forecasted result that the *NSE* shown in Table 7 obtained by MREF was 0.784, slightly higher than 0.728 of MRES. Based on the computed *NSE*, the preference for choosing MREF over MRES rose to 15:5. This suggests that though there were 5 stations that had higher resolution spectral densities using MRES, streamflow forecasted by MREF was more reliable. It was noticed that for these cases the difference between the values of *NSE* for MRES and MREF was less than 0.04. However, streamflows forecasted by the two MRE theories were difficult to distinguish from Fig. 11, where both theories fitted observations satisfactorily. It can be noted from forecasted errors in Fig. 12 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. It is found that low flow forecasted by MRES was 11.7 m<sup>3</sup>/s higher than observed flow, while that by MREF was 1.1 m<sup>3</sup>/s higher than the observation. Though the forecasted errors plotted were close and the differences in errors were less than 0.05 for almost all cases, the difference in forecasted low flow of UMU was larger than 0.1.

### 4.3 Spectral power or frequency

No matter which entropy theory is used, both MRE theories captured the largest peak at the 1/12 frequency and emphasized the yearly periodicity. Now we return to the question of which way of defining entropy is better.

With the spectral power as a random variable, the spectral density derived from MRES is in the form of polynomial, while the other one, with frequency as a random variable, obtains spectral density in the form of exponential. With the exponential formula, the spectral density estimated by MREF is no longer restricted to the AR process, but can also be applied to the ARMA and MA (Liefhebber and Boekee 1987) processes. However, MRES was more likely to capture small peaks, while MREF ignored small peaks and only focused on the peaks through the hypothesized prior. The peak of the spectral power estimated by MREF is more impacted by the prior assumption.

Another difference is in the hypothesis on the prior. MRES assumes a prior for the distribution of the spectral power, while MREF assumes one on the spectral density. The way of imposing prior by MREF is more straightforward than MRES. To apply the MRES, one needs to have long historical data to obtain the past spectral power

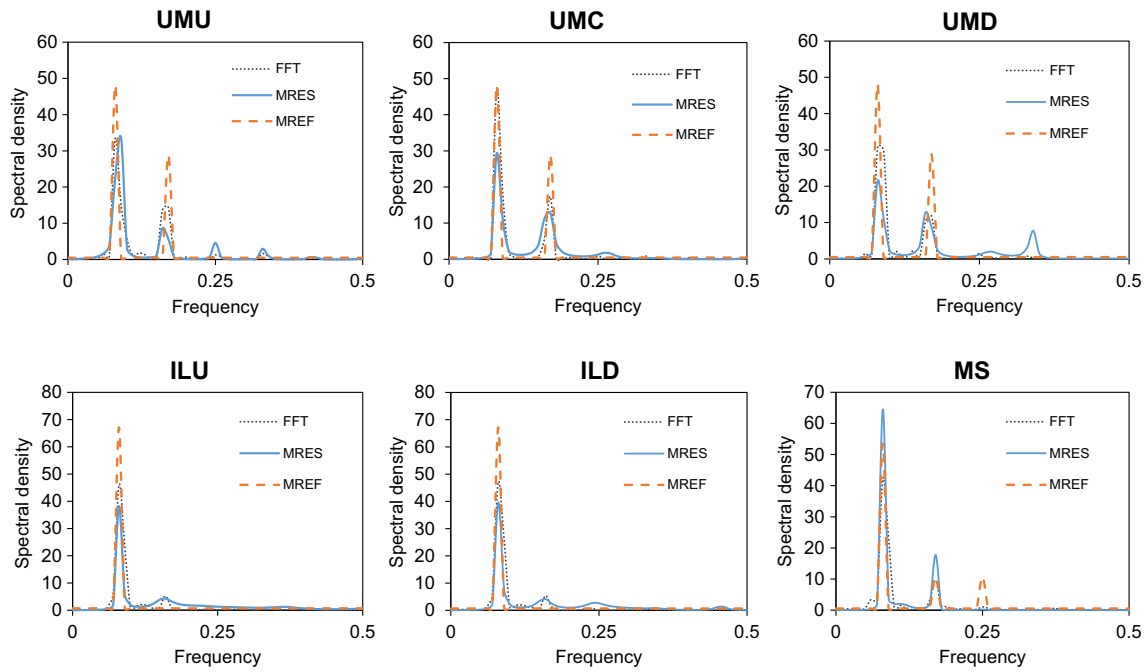


Fig. 9 Comparison of spectral densities by MRES and MREF

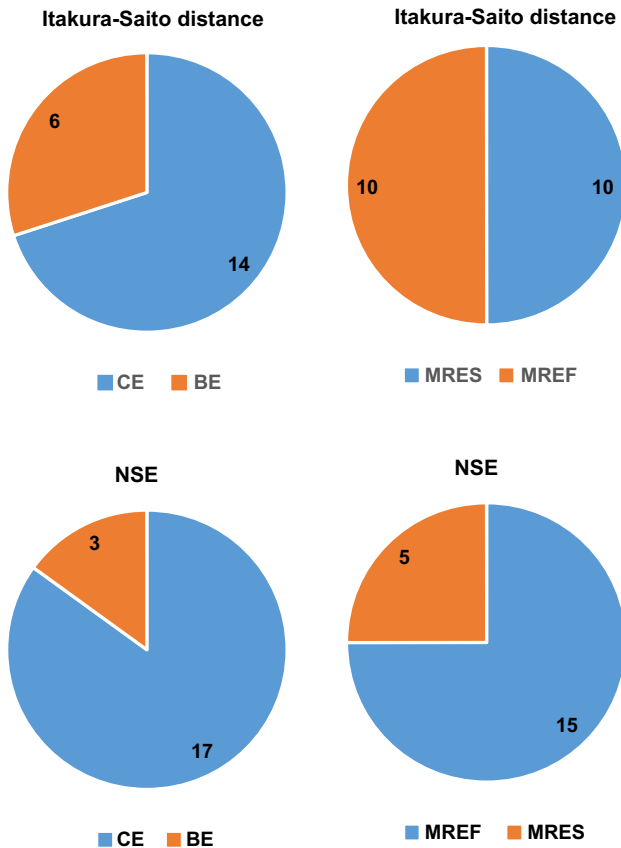


Fig. 10 Preference of choosing MREF over MRES and CE over BE

series and to analyze the distribution of spectral powers at each frequency, or at least the periodicity of streamflow needs to be known, but MREF does not. To apply MREF, one needs to add peak spectra at the significant period of the streamflow analyzed in addition to the background noise. In such a way MREF can even be applied to data without periodicity information, since the 12 months periodicity is common for monthly streamflow. Thus, the Prior 2 assumption is generally applicable for any monthly streamflow series. This suggests that to maintain the same resolution MRES needs more information to be collected a priori than MREF.

The way of solving for the Lagrange multipliers and the forecasting coefficients is more straightforward for MREF than for MRES. MRES estimates parameters by solving the  $N$  nonlinear equations (Alsaka et al. 1988), while MREF uses cepstrum analysis to estimate the Lagrange multipliers. From the given  $N$  autocorrelations, cepstrum can be obtained through the recursive function up to lag  $N$ . On the contrary, the autocorrelation is extended from the  $N$  posterior cepstrum and the prior cepstrum beyond the lag  $N$ . Furthermore, the MREF theory forecasts streamflow based on cepstrum analysis or specifically, say, the hypothesized prior cepstrum. It may be seen in the way that using the MREF method in Eq. (24), the prior cepstrum is imposed as the background of streamflow fluctuation to maintain the periodicity assumed in the prior spectral density. Figure 13 plots the prior cepstrum corresponding to the six priors used for MREF, which shows how

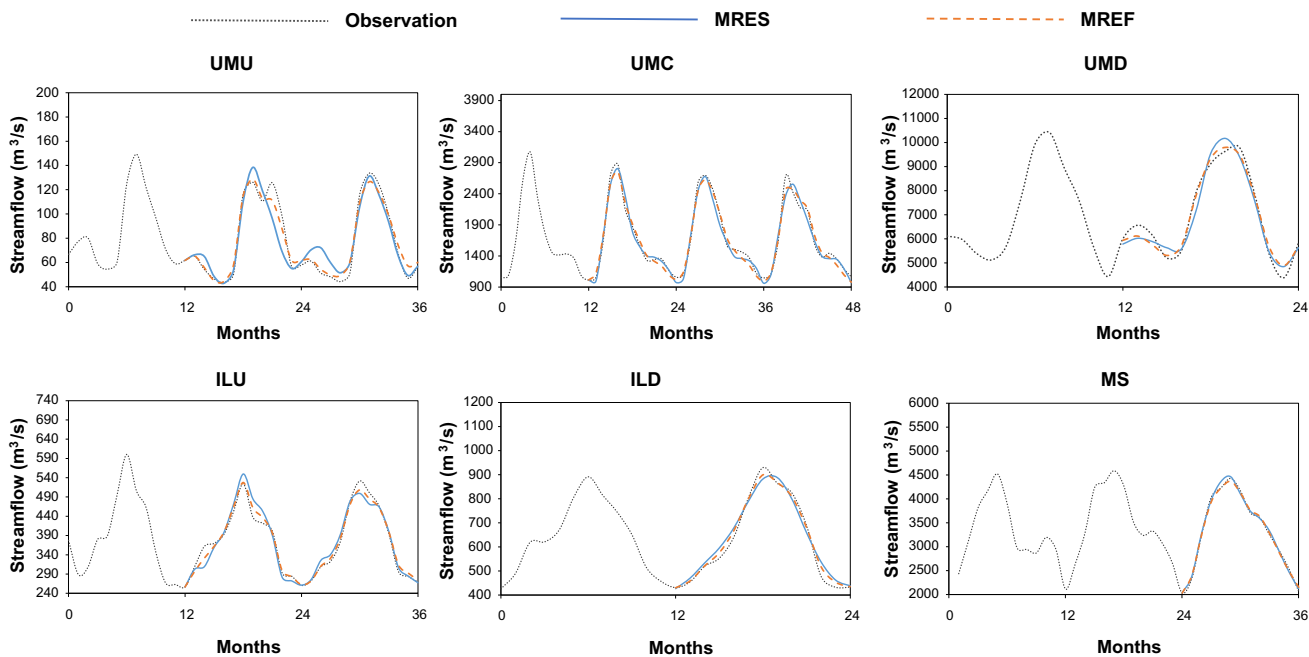


Fig. 11 Comparison of streamflow by MRES and MREF

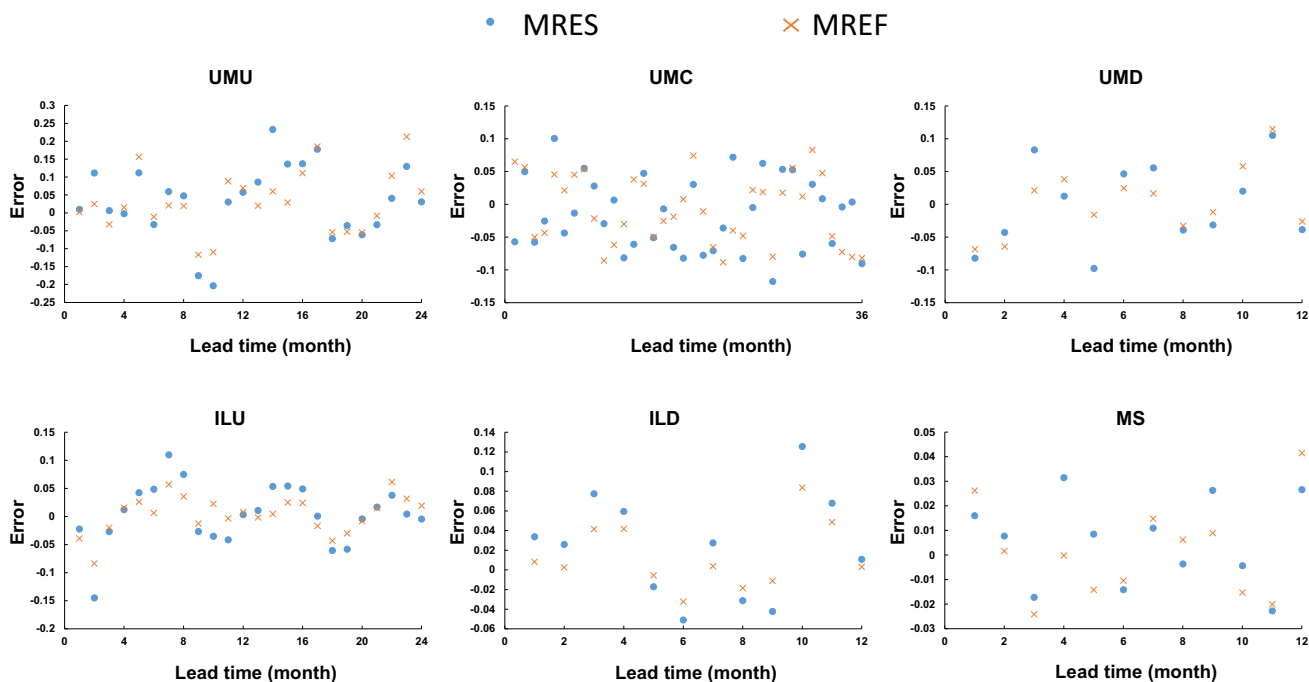
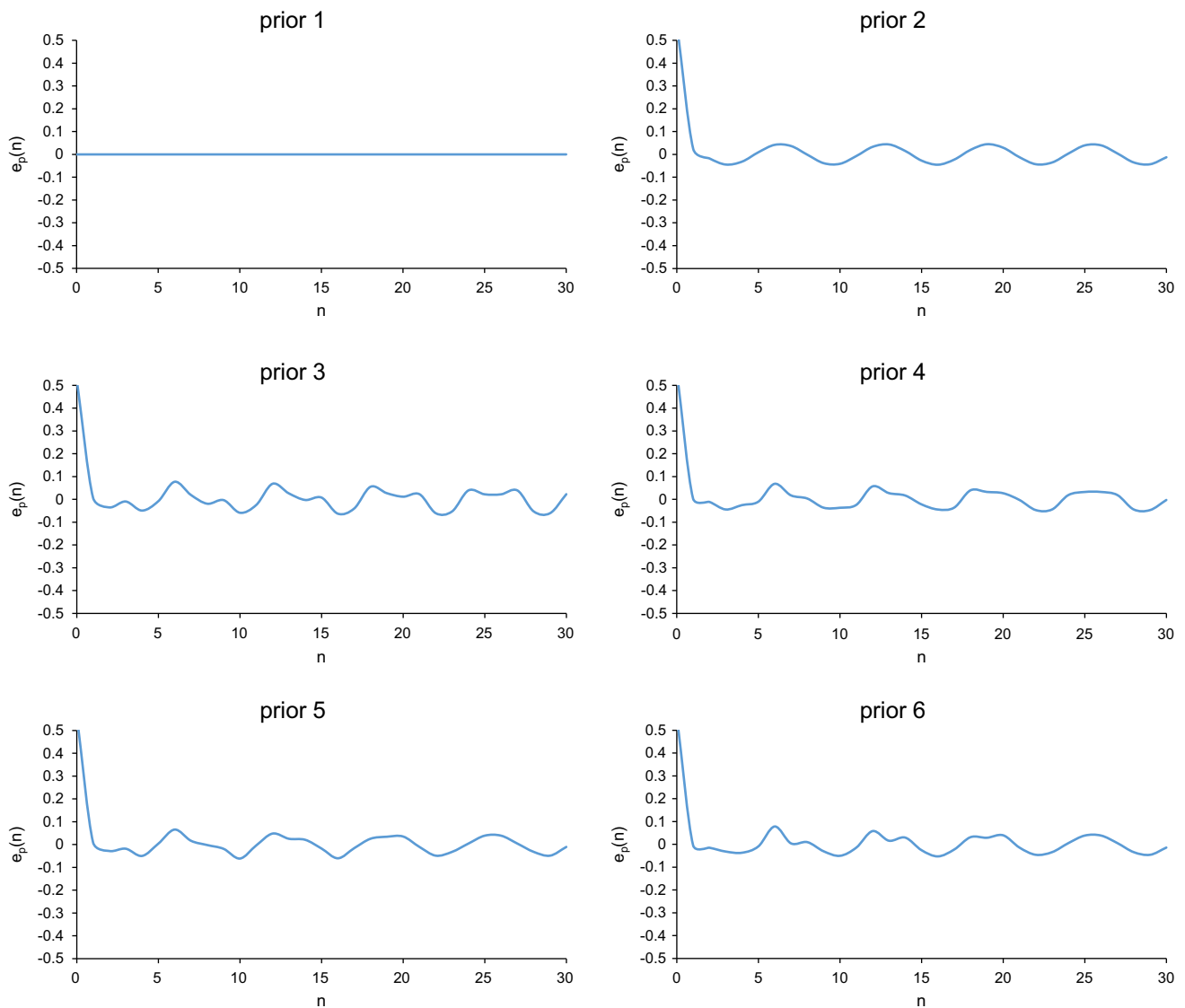


Fig. 12 Forecasted errors by MRES and MREF

spectral analysis impacts the streamflow time series. Cepstrum is inverse Fourier transform from spectral density, which represents the effect of different periodicities on the time series. As stated in Eq. (24), the cepstrum is not only used for estimating the prediction coefficients, but it directly influences the future prediction. 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. It may be the reason why MREF has the capability to forecast small peaks of streamflow, and the hydrograph is less smooth but closer to the observation than MRES. When no prior information is



**Fig. 13** Prior cepstrum computed from six prior spectral densities

**Table 8** Comparison for different methods

Statistics	Order
Model order	AR < BE < MRES, CE, MREF
Itakaru-Saito distance	BE > CE > MRES, MREF
RMS	AR > BE > CE > MRES > MREF
NSE	AR < BE < CE < MRES < MREF
Lead time	AR < BE < CE < MRES, MREF
Information needed	BE, CE < MREF < MRES

given, the prior cepstrum constantly equals 0 and has no impact on streamflow forecasting.

The above comparison of different methods is summarized in Table 8. In general, MRE theories, both MRES

and MREF, require more information as the prior than maximum entropy (ME) theories, including BE and CE, while MRES needs longer historical data to assume the prior hypothesis than does MREF. The entropy theories, MREF, MRES and CE, need a larger model order than BE, and all entropy theories have larger order than has AR. The spectral density estimation has a higher resolution for the MRE theories than the ME theories, and within two ME theories CE yields lower I-S distortion than BE. The forecasted errors are smaller for MRE than ME, and using frequency as a random variable (MREF or CE) is smaller than that using spectral power (MRES or BE). The forecasting lead time is obtained longer using MRE than ME, and longer using CE than BE. To sum up, the reliability of streamflow forecasting from high to low is for MREF, MRES, CE, BE and then AR.



### 5 Concluding remarks

The minimum relative entropy theory is developed by considering frequency as a random variable in this paper. The MREF theory forecasts 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 has a higher *NSE* than for the Lower Mississippi, but the difference is less than 0.1.

No matter which prior is chosen, the MREF theory does 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 has higher resolution than that of the CE theory. The difference in the estimation of spectral density with different priors is minimal, and Prior 2 is applicable for monthly streamflow of any stations on Mississippi River, as spectral peaks at other frequencies are not as significant as the one at frequency 1/12. Besides, forecasted streamflow with prior is more reliable than that from the CE theory, especially for multi-peak flow condition and low flows. Furthermore, the lead time by MREF is also longer than for CE. Thus, it can be summarized that the MRE theories are generally better than the ME theories for both higher resolution in the spectral estimation and higher reliability in streamflow forecasting.

The spectral density estimated by the MREF or CE theory has higher resolution than the MRES or BE theory based on the I-S distortion. However, the advantage of choosing frequency as a random variable over spectral power is more significant for ME than MRE, since BE shifts the peaks for multi-peak condition. Comparing to MRES and BE, MREF and CE emphasize the main periodicity of streamflow more, and sacrifice small peaks to maintain the largest peak at the correct location.

Streamflow forecasted by MREF or CE is more reliable than by MRES or BE. The advantage of using CE over BE is significant for streamflow forecasting. Though more model orders are required to apply CE, streamflow forecasted by CE is more consistent with periodicity and has a longer lead time. The difference between MREF and MRES is not as significant as that between BE and CE. Both MRE theories use a similar order of model and the goodness fit has less difference. However, *NSE* for MREF is higher than for MRES by about 0.56 for streamflow in the Mississippi River watershed. The main advantage of using MREF over MRES is in the straightforward prior assumption and in the parameter estimation with cepstrum analysis. The prior is needed to be assumed on the distribution of spectral power, thus more prior information is needed to construct the prior hypothesis for MRES than

for MREF. Though both MRE and MRES are capable of forecasting streamflow with limited streamflow information, MRES needs historical data to test the prior hypothesis, while MREF is even available for stations with lack of periodicity information with the Prior 2 option.

The MRE theory provides a foundation for streamflow forecasting using entropy theory. The total number of four entropy approaches are discussed in the paper, which are MRES, BE, MREF and CE. Though the proposed method is designed for monthly streamflow forecasting, it is not restricted to monthly time scale, as it may also be applied to other time scales, for example hourly, daily, weekly or annual time series. Furthermore, the entropy theory can be derived for multivariate condition by defining entropy in the form of vector so that streamflow can be forecasted linking to other climate indicators. It would be potentially useful for hydroclimate analysis.

### Appendix 1: Cepstrum analysis

Cepstrum is defined as the inverse Fourier transform of the log-magnitude of Fourier spectrum. For a given streamflow time series  $y(t)$ , the cepstrum can be computed using the following steps.

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

$$Y(f) = \sum_{n=-\infty}^{\infty} y(t)e^{-2\pi nif} \tag{26}$$

where  $Y(f)$  is the Fourier transform of  $y(t)$ . Taking the inverse Fourier transform of the log-magnitude of Eq. (26) one obtains the cepstrum of the Fourier transform as

$$C(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|Y(f)|e^{2\pi nif} df \tag{27}$$

It is known that the Fourier transform of autocorrelation leads to the spectral density, which is

$$q(f) = \sum_{n=-\infty}^{\infty} \rho(n)e^{-2\pi nif} \tag{28}$$

Thus, similar to Eq. (26), the cepstrum of autocorrelation can be defined by the inverse Fourier transform of the log-magnitude of  $P(f)$ , which yields

$$e(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|q(f)|e^{2\pi nif} df \tag{29}$$

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

$$q(f) = \{FT[y(t)]\}^2 \tag{30}$$

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

$$\begin{aligned} e(n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|FT[y(t)]|^2 e^{2\pi nif} df \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} 2 \log|Y(f)| e^{2\pi nif} df = 2C(n) \end{aligned} \tag{31}$$

### Appendix 2: Cepstrum of data of finite length

Consider only the positive part of the autocorrelation function  $\rho(n)$ , for  $n > 0$  and let  $e(n)$  be the cepstrum estimated from  $\rho(n)$ ,  $n > 0$ , which is

$$e(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|p^*(f)| e^{2\pi nif} df \tag{32}$$

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. Using only the positive part of  $\rho(n)$  ensures that  $\rho(n)$  is the minimum-phase function and for a minimum phase system the input and output are uniquely determined. This means  $e(n)$  can be uniquely determined from  $\rho(n)$ . Let us define a two-sided output in the way that

$$\hat{\rho}(n) = \begin{cases} 2e(n), & n > 0 \\ e(0), & n = 0 \\ 2e(-n), & n < 0 \end{cases} \tag{33}$$

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  $\rho(n)$ :

$$\hat{\rho}'(z) = \frac{\rho'(z)}{\rho(z)} \tag{34}$$

which is equivalent to

$$z\hat{\rho}'(z) = \sum [-n\hat{\rho}(n)]z^{-n+1} = \frac{z\rho'(z)}{\rho(z)} \tag{35}$$

The following difference equation can be obtained from Eq. (35):

$$z\rho'(z) = z\hat{\rho}'(z)\rho(z) \tag{36}$$

Taking the inverse  $z$  transform of Eq. (36), one obtains

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

Dividing Eq. (37) by  $n$ , the relationship between input and output becomes

$$\rho(n) = \sum_{k=-\infty}^{\infty} \left(\frac{k}{n}\right)\hat{\rho}(n)\rho(n-k) \tag{38}$$

Transforming Eq. (38) with the use of Eq. (33), the autocorrelation function can be obtained from the following recursive formula (Oppenheim and Schafer 1975):

$$\rho(n) = \begin{cases} \frac{\rho(0)}{2} e(n) + \sum_{k=1}^{n-1} \frac{k}{n} e(k)\rho(n-k), & n > 0 \\ \exp e(n), & n = 0 \\ \rho(-n), & n < 0 \end{cases} \tag{39}$$

On the other hand, the cepstrum  $e(n)$  can be obtained from the reverse relation of Eq. (39) as:

$$e(n) = \begin{cases} \frac{2}{\rho(0)} [\rho(n) - \sum_{k=1}^{n-1} \frac{k}{n} e(k)\rho(n-k)], & n > 0 \\ \log \rho(n), & n = 0 \\ 0, & n < 0 \end{cases} \tag{40}$$

### References

Alsaka YA, Tzannes NS, Marinelli WA (1988) An efficient algorithm for implementing the relative entropy method. *Acoust Speech Signal Process* 11–14(1988):2384–2387. doi:10.1109/ICASSP.1988.197120

Box GEP, Jenkins GM (1970) *Time series analysis: forecasting and control*. Holden-Day series in time series analysis. Holden-Day, San Francisco

Burr RL, Lytle DW (1986) A general-method of minimum cross-entropy spectral estimation—comments *Ieee T Acoust Speech* 34:1324–1326

Cui H, Singh VP (2015) Configurational entropy theory for streamflow forecasting. *J Hydrol* 521:1–17. doi:10.1016/j.jhydrol.2014.11.065

Cui H, Singh VP (2016a) Application of minimum relative entropy for streamflow forecasting *Stoch Env Res Risk A*, under review

Cui H, Singh VP (2016b) Maximum entropy spectral analysis for streamflow forecasting. *Phys A* 442:91–99. doi:10.1016/j.physa.2015.08.060

Feng X, Porporato A, Rodriguez-Iturbe I (2013) Changes in rainfall seasonality in the tropics. *Nat Clim Change* 3:811–815. doi:10.1038/Nclimate1907

Frieden BR (1972) Restoring with maximum likelihood and maximum entropy. *J Opt Soc Am* 62:511

Girardin V (2001) Relative entropy and covariance type constraints yielding ARMA models. In: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 20th International Workshop*, 2001. AIP Publishing, Melville, vol 1, pp 318–327

Gull SF, Daniell GJ (1978) Image-reconstruction from incomplete and noisy data. *Nature* 272:686–690

- Hipel KW, McLeod AI (1994) Time series modelling of water resources and environmental systems. *Develop Water Sci* 45:555–572
- Johnson RW, Shore JE (1983) Which is better entropy expression for speech processing: SLogS or LogS?. Naval Research Laboratory, Washington, DC
- Katsakos-Mavromichalis NA, Tzannes MA, Tzannes NS (1985) Frequency resolution: a comparative study of four entropy methods. *Kybernetes* 15:25–32
- Krstanovic PF, Singh VP (1991) A univariate model for long-term streamflow forecasting. 2. *Appl Stoch Hydrol Hydraul* 5: 189–205
- Kullback S (1959) *Information theory and statistics*. Wiley, New York
- Liefhebber F, Boekee DE (1987) Minimum information spectral-analysis. *Signal Process* 12:243–255
- Nadeu C (1992) Finite length cepstrum modeling - a simple spectrum estimation technique. *Signal Process* 26:49–59
- Nadeu C, Sanvicente E, Bertran MA (1981) new algorithm for spectral estimation. In: *International Conference on Digital Signal Processing*, Florence, Italy. pp 463–470
- Oppenheim AV, Schafer RW (1975) *Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ
- Oppenheim AV, Schafer RW (2004) From frequency to quefrequency: a history of the cepstrum. *IEEE Signal Proc Mag* 21:95
- Papademetriou RC (1998) Experimental comparison of two information-theoretic spectral estimators. In: *Signal Processing Proceedings, 1998. ICSP '98. Fourth International Conference*, pp 141–144. doi:[10.1109/ICOSP.1998.770170](https://doi.org/10.1109/ICOSP.1998.770170)
- Shore JE (1979) Minimum cross-entropy spectral analysis. Naval Research Laboratory, Washington DC
- Shore JE (1981) Minimum cross-entropy spectral-analysis. *IEEE Trans Acoust Speech* 29:230–237
- Tzannes MA, Politis D, Tzannes NS (1985) A general method of minimum cross-entropy spectral estimation. *IEEE Trans Acoust Speech* 33:748–752
- Wu NL (1983) An explicit solution and data extension in the maximum-entropy method. *IEEE Trans Acoust Speech* 31:486–491