Article

Entropy-Based Wavelet De-noising Method for Time Series Analysis

Yan-Fang Sang [1], Dong Wang [1,*], Ji-Chun Wu [1], Qing-Ping Zhu [2] and Ling Wang [3]

[1] State Key Laboratory of Pollution Control and Resource Reuse, Department of Hydrosciences, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210093, China; E-Mails: sunsangyf@gmail.com (Y.-F.S.); jcwu@nju.edu.cn (J-C.W.)

[2] China Water International Engineering Consulting Co. Ltd., Beijing 100053, China; E-Mail: qpzhu@mwr.gov.cn (Q.-P.Z.)

[3] Hydrology Bureau of Yellow River Conservancy Committee of Ministry of Water Resources, Zhengzhou 450001, China; E-Mail: wlyrcc@sina.com (L.W.)

* Author to whom correspondence should be addressed; E-Mail: wangdong@nju.edu.cn.

Abstract: The existence of noise has great influence on the real features of observed time series, thus noise reduction in time series data is a necessary and significant task in many practical applications. When using traditional de-noising methods, the results often cannot meet the practical needs due to their inherent shortcomings. In the present paper, first a set of key but difficult wavelet de-noising problems are discussed, and then by applying information entropy theories to the wavelet de-noising process, *i.e.*, using the principle of maximum entropy (POME) to describe the random character of the noise and using wavelet energy entropy to describe the degrees of complexity of the main series in original series data, a new entropy-based wavelet de-noising method is proposed. Analysis results of both several different synthetic series and typical observed time series data have verified the performance of the new method. A comprehensive discussion of the results indicates that compared with traditional wavelet de-noising methods, the new proposed method is more effective and universal. Furthermore, because it uses information entropy theories to describe the obviously different characteristics of noises and the main series in the series data is observed first and then de-noised, the analysis process has a more reliable physical basis, and the results of the new proposed method are more reasonable and are the global optimum. Besides, the analysis process of the new proposed method is simple and is easy

to implement, so it would be more applicable and useful in applied sciences and practical engineering works.

Keywords: time series analysis; de-noising; information entropy; wavelet transform; uncertainty

1. Introduction

Time series analysis is not only a very important technique for identifying different components and revealing variation characters of the variable studied, but also is the basis of many simulation and forecast works, thus it has been widely used in many different fields of applied researches and engineering practical works presently, such as electronics, business, medicine, physics, earth sciences, hydraulic engineering and among others [1-8]. However, due to the influence of many random and uncertain natural factors as well as the subjective factors, observed time series data always include many noises, which contaminate the real series data and cause many difficulties in the time series data analysis process, e.g., periods' identification, parameters' estimation, simulation and forecasting, *etc*. [9-14]. Due to the existence of noises, it is not an easy task to get accurate time series analysis results in practical works. As far as the noises are concerned, they are generally classified as either additive or dynamical, and additive noises are sometimes called measurement noises [15-16]. Comparatively, the dynamical noises are generated by certain physical mechanisms, so they usually show good correlations (or maybe constants sometimes) and can be identified and modified easily; but the additive noises often show random characters and are difficult to be analyzed and described accurately. In the present paper, on the basis of the different physical generation mechanisms between real series and noises in observed time series data [15], the components which have pure random characters and are generated by random and uncertain factors are defined as noises, and they are the main focus to be studied in this paper.

In order to obtain accurate and reliable time series data analysis results in practical works, noises reduction or removing should precede other tasks in the time series analysis process. At present, there have been a number of de-noising methods. Among them, one kind of de-noising methods is to establish suitable deterministic models to simulate the observed time series first, and then regard the difference between observed series and simulated series as noises [17-18]. However, many real natural evolution mechanisms cannot be understood completely now and sometimes even know nothing, so the real models are unknown in these cases and the de-noising results are unreliable. Another important kind of de-noising methods is based on spectral analysis [19]. Since most of time series in nature show many complex variation characters [20-22] (e.g., the hydrologic time series are the most representative because of usually showing extremely non-stationary and nonlinear characters which result from spatial and dynamical heterogeneities and also showing multi-temporal scale variation characters [9,23,24], so they will be the examples in "cases studies" in section 5), these traditional spectral analysis-based methods (Wiener filtering, Kalman filtering, Fourier transform and among others) have many disadvantages and also cannot meet the practical needs enough. For examples, Wiener filtering method and Kalman filtering method are only suitable for linear natural systems, and

the analysis results depend on the establishment of state space functions to a great extent; and Fourier transform method is just suitable for stationary and linear time series analysis. In recent years, wavelet analysis (WA) used widely is a new and powerful method of time series analysis in theory [25-30], by which noises in time series data can also be reduced or removed. However, when using the wavelet de-noising method, there are some key but difficult problems as discussed in section 2, and most of them have not been solved presently, so the de-noising results of it are also not as good as expected in many practical works.

To distinguish the new method which would be proposed in this paper, the de-noising methods mentioned above are called "traditional de-noising methods". The discussion results about these traditional de-noising methods indicate that: (1) most of the de-noising methods used presently have their own applicable conditions and also have many disadvantages, which would limit their uses and cause the difficulties in getting accurate analysis results in practice; (2) for certain time series data, the de-noising results vary with the methods used, sometimes analysis results of certain methods show unreasonable phenomena or even wrong completely, for examples, the separated noises show good auto-correlations, the de-noised series data losses some real components, *etc.*; (3) comparatively, the wavelet de-noising method is much more applicable and more powerful than others, since it can identify the variation characters of time series data both in temporal and frequency domains. However, several key problems impact its effectiveness and accuracies; and (4) for the de-noising methods used presently, they do not take the physical difference between the characters of real series data and noises into account effectively. However, the physical processes of the variable studied are always the most concerned in practical works, especially in hydraulic engineering and earth sciences. In this paper, the real series data in observed time series is called "main series", *i.e.*, the observed time series data is composed of the "main series" and "noises". If based on the different characters of noises and main series to de-noising, the analysis process would have more reliable physical basis and the results could become more accurate and reasonable.

Information entropy is a powerful and universal theoretical concept used for measures of disorder, uncertainty and complexity [31-35]. For a given system whose exact description is not precisely known, the entropy is defined as the amount of information needed to exactly specify the state of the system, given what we know about the system. Nowadays, information entropy theories have been applied across physics, mathematics, information theory and many other branches of applied sciences and engineering, and more and more applications have indicated the effectiveness and universality of it, for examples, the principle of maximum entropy (POME) [36] is widely used to estimate parameters and determine the probability distribution of the random variables studied [37-45], and maximum entropy spectral analysis (MESA) [46] has been a commonly used method for identifying the dominant periodicities of time series data [47-52]. In this paper, for the main objective of proposing a new wavelet de-noising method which is more applicable and effective in applied sciences and practical works, the information entropy theories are employed to the time series de-noising process mainly for describing the different characters of noises and main series and then providing reliable physical basis to the de-noising. To begin with, several key but difficult problems about wavelet de-noising are discussed, and the suggestions and approaches for solving them are put forward, then by using information entropy theories both to describe random characters of the noises and degrees of complexity of the main series in observed time series data, respectively, a new entropy-based wavelet

de-noising method is proposed. Finally, noises in some different synthetic series and typical observed time series data are separated by using the new proposed method and other traditional wavelet de-noising methods, and the results are compared and discussed in detail. The results indicate that better performances of the new method proposed in de-noising of time series data.

The paper is organized as follows. After the introduction, traditional wavelet de-noising methods are reviewed in Section 2, and then a set of key but difficult wavelet de-noising problems are discussed in detail in Section 3; in Section 4, the new entropy-based wavelet de-noising method is proposed by applying information entropy theories to de-noising process; some examples are analyzed by different methods for verifying the new method proposed in Section 5. Finally, a set of discussions about the new method conclude the paper.

2. Review of Traditional Wavelet De-noising Methods

2.1. Wavelet Transform (WT)

Just like other transform techniques (Fourier, Bessel, *etc.*), the wavelet transform also has its own base function, *i.e.*, the mother wavelet function. The mother wavelet functions must fulfill certain strict mathematical conditions called "admissibility conditions" as shown in Equation (1) in the temporal domain and Equation (2) in the frequency domain [24, 26]:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{1}$$

$$C_\psi = \int_{-\infty}^{+\infty} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} d\omega < \infty \tag{2}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of the mother wavelet function $\psi(t)$ at the frequency ω.

Defining $L^2(R)$ as a measurable square integral function space in real axis, the signal $f(t) \in L^2(R)$ can be analyzed by the continuous wavelet transform (CWT) as:

$$W_f(a,b) = \int_{-\infty}^{+\infty} f(t)\psi_{a,b}{}^*(t)dt = |a|^{-1/2} \int_{-\infty}^{+\infty} f(t)\psi^*(\frac{t-b}{a})dt \quad a,b \in R, a \neq 0 \tag{3}$$

where $\psi^*(t)$ is complex conjugate; $\psi_{a,b}(t)$ is the wavelet function gained by translating and expanding $\psi(t)$; a is temporal scale factor and b is time position factor; $W_f(a,b)$ are wavelet coefficients. Along with the varying values of parameters a and b, $W_f(a,b)$ can exhibit good localized characters both in the temporal and frequency domains [25-30]. Therefore, variation characters of signal $f(t)$ in multi-temporal scales can be understood by the CWT.

In practice, observed time series data are usually discrete signals like $f(k\Delta t)$ ($k = 1, 2, \ldots, N$; Δt is time interval), so a and b become discrete, and we then get the discrete wavelet transform (DWT) of the signal $f(k\Delta t)$ as:

$$W_{j,k} = a_0^{-j/2} \int_{-\infty}^{+\infty} f(t)\psi^*(a_0^{-j}t - kb_0)dt \tag{4}$$

where a_0 ($a_0>1$) and b_0 are constants. In practical works, the dyadic DWT is usually used by assigning $a_0 = 2$ and $b_0 = 1$. The integer j is the temporal scale factor which is analogous to the parameter a in Equation (3), and the integer k is the time position factor which is analogous to parameter b.

When the mother wavelet function satisfies more restrictive conductions called "regularity conditions" in Equation (5), it is said to have the regularity of order N [24,27]:

$$\int_{-\infty}^{+\infty} t^k \psi(t)dt = 0, \quad k = 1, \cdots, N-1 \tag{5}$$

If the wavelet function fulfills the condition in Equation (5), the signal can be reconstructed by using the wavelet coefficients $W_f(a,b)$ with Equation (6) or $W_{j,k}$ with Equation (7). Besides, different components in signal $f(t)$ can also be reconstructed by DWT:

$$f(t) = C_\psi^{-1} \int_{-\infty}^{+\infty} W_f(a,b)\psi_{a,b}(t)\frac{da}{a^2}db \text{ or } f(k\Delta t) = \sum_{j,k} W_{j,k}\psi^*(a_0^{-j}t - kb_0) \tag{6}$$

$$f(k\Delta t) = \sum_{j,k} W_{j,k}\psi^*(a_0^{-j}t - kb_0) \tag{7}$$

2.2. Traditional Wavelet De-noising Methods

The main series and noises in observed time series data are generated by different physical mechanisms and have obviously different variation characters, therefore, the values and variation characters of wavelet coefficients describing them are also different. Based on this difference, proper thresholds can be used to adjust the wavelet coefficients of DWT, and then the main series and noises can be separated by using the wavelet reconstruction method. This is the basic idea of wavelet threshold de-nosing [9,19,53,54]. There are four main and key problems in the wavelet de-noising process, namely: (1) choice of reasonable wavelet functions and (2) choice of proper time scale levels, both of which mainly determine the accuracy and reasonability of the DWT results; (3) determination of accurate thresholds under each time scale level by certain methods; and (4) choice of suitable thresholding rules. The analysis process of wavelet de-noising methods will be described in detail in Section 4, together with the information entropy analysis process.

Although being effective and powerful in theory, the wavelet threshold de-noising methods have several main defects when used in practice, for example, in runoff series data analysis as discussed in Section 5, and each of them is discussed in the following paragraphs.

The first is probability description of noises. When using traditional wavelet de-noising methods in practice, noises in observed time series data are generally thought of as following second-order stationary process, such as following normal probability distribution, so the standard deviation or variance of noises are mainly used to estimate the noise level and to estimate proper threshold values. However, in practice, the fact is that what type of probability distribution noises follow in observed time series data is usually unknown, e.g., noises in hydrologic series data generally do not follow second-order stationary process, but rather follow skewed probability distributions. Therefore, it is unreasonable and also not enough to only take the second-order stationary process (*i.e.*, standard deviation or variance) into account when de-noising time series data. In order to obtain accurate and reliable de-noising results, certain approaches (such as information entropy), which can conveniently

analyze and describe the random characters of noises following different probability distributions, should be given and introduced into the wavelet de-noising process.

The second is the accuracy of threshold estimation methods. Many traditional methods have their own disadvantages [9], so the analysis results obtained using them are not accurate and are also different to each other. Among the numerous threshold estimation methods, the universal threshold algorithm (UT) takes a prominent position, as it offers many optimality properties [55]. The basic idea of UT algorithm is estimating the thresholds values based on the series' length and noises' standard derivation. In practical applications, the UT algorithm is often found to be too conservative, *i.e.*, it removes too much of the underlying data, thereby causing blur in the output [55]. Another important method is the Stein unbiased risk estimation (SURE) algorithm, which is an unbiased estimator of the mean-squared error (MSE) of a given estimator [56], and to improve the performances of SURE algorithm, the heuristic SURE algorithm that was also proposed later. Besides, the minimax algorithm (MIN) is also a typical threshold estimation method which is a minimum estimator of the MSE of a given estimator by using regression models to describe the time series data analyzed [57]. The theories about these typical methods have been elaborated in many papers [19,56-60], from where it can be found that the basic idea of the SURE and MIN algorithm is to first establish an estimator (also called risk function) which is generally the MSE function describing the difference between the original series data and the de-noised series data, *i.e.*, the variance of the noises separated; and then the SURE algorithm or MIN algorithm can be used to estimate the minimum values of MSE function; finally the results corresponding to the minimum MSE are considered the best thresholds estimation results. But in practical works, the fact is that both the probability distribution and the amount of noise in observed time series data are unknown, while the methods above just take the second-order stationary process of noises into consideration, so their analysis results lack a physical basis and are unreasonable to a certain extent. Furthermore, just as Jansen and Bultheel pointed out [58]: "the main challenge with this MSE as an objective function is the fact that in real applications, it can never be computed exactly: its definition uses the value of the exact, unknown coefficients. In practical situations, this MSE has to be estimated". Although many other improved threshold estimation methods have been proposed later, in essence they are just the same as those three typical methods mentioned above, so their defects are not overcome effectively.

The third is the validity of thresholding rules. No matter whether the hard-thresholding rule or the soft-thresholding rule is used as described in Equation (8) and Equation (9), respectively, they each have their own disadvantages. Generally, the latter is better than the former because the dealt wavelet coefficients $W_{j,k}^{'}$ in Equation (8) is discontinuous at the points of both $-T_j$ and $+T_j$. However, the soft-thresholding rule also has its own defects, *i.e.*, there are constant deviations between $W_{j,k}$ and the real $W_{j,k}^{'}$, which will influence the precision of wavelet reconstruction results. In order to overcome the defects of hard- and soft-thresholding rules, many mid-thresholding rules have been put forward, mainly by combining the hard- and soft-thresholding rules together using different means [61-63]. In these improved mid-thresholding rules, some new parameters are generally used to coordinate the proportional relations of hard- and soft- thresholding rules, which would add to the difficulties of de-noising since these parameters are difficult to estimate and determine in practical works.

Hard-thresholding rule:

$$W'_{j,k} = \begin{cases} W_{j,k} & |W_{j,k}| > T_j \\ 0 & |W_{j,k}| < T_j \end{cases} \tag{8}$$

Soft-thresholding rule:

$$W'_{j,k} = \begin{cases} \text{sgn}(W_{j,k})(|W_{j,k}| - T_j) & |W_{j,k}| > T_j \\ 0 & |W_{j,k}| < T_j \end{cases} \tag{9}$$

where T_j is the wavelet coefficient threshold under time scale level j.

3. Discussions of Several Key Problems Concerning Wavelet De-noising

In order to propose a new effective wavelet de-noising method, by which more accurate and reliable time series data de-noising results can be obtained, especially in practical applied sciences and engineering applications, the following main and key problems of wavelet de-noising are discussed in detail, and several suggestions and approaches for solving them are also given, which are the choice of reasonable wavelet function, choice of proper time scale levels, determination of accurate thresholds and choice of suitable thresholding rules, respectively.

3.1. Choice of Reasonable Wavelet Function

According to the wavelet analysis theory, it is known that the first and key problem concerning WA is how to choose a reasonable mother wavelet function, since the analysis results of time series data vary with the wavelet function used. Many papers have discussed this problem [64-66]. In the authors' opinion, the mathematical properties of wavelets should be taken into account first when choosing a wavelet function, *i.e.*, it is preferable to first choose progressive, linear phase wavelets; secondly, the wavelet chosen should exhibit good localized properties in both the temporal and frequency domains; thirdly, the trade-off between time and scale resolutions of the chosen wavelet has to be adapted to the analysis process [67]; and fourthly, the chosen wavelet should meet the "regularity conditions" in Equation (5) for reconstructing different components in the original series data. The mathematical properties of commonly used wavelet functions are summarized in Table 1 [68].

Based on the mathematical properties of wavelet functions, a simple method of choice of reasonable wavelet function proposed in [69] can be used in this paper, whose basic idea is: first each of the wavelet functions is used to separate the main series and noises in the observed time series by DWT, and then the similarity degrees between original time series and the main series are compared, and we judge whether the characters of the separated noises are purely random or not; finally the most reasonable wavelet function can be chosen by comparing the analysis results of these wavelet functions.

Table 1. Mathematical properties of the mother wavelet functions used commonly.

Wavelet function	Abbreviation	Function number	Mathematical properties				
			Compactly supported	Symmetry	Vanishing moment	Orthogonality	Double-Orthogonality
Haar	haar	1	+	+	1	+	+
Daubechies	db$N$	10	+	−	N	+	+
Symlets	sym$N$	7	+	+*	N	+	+
Coiflets	coif$N$	5	+	+*	2N	+	+
Dmeyer	dmey	1	+	+	/	+	+
BiorSplines	bior$M.N$	15	+	+	M	−	+
ReverseBior	rbio$M.N$	15	+	+	M	−	+

Note: "+" means the wavelet function has the corresponding mathematical property, and "+*" means the wavelet function has the corresponding similar mathematical property, while "−"mean the wavelet function does not have the corresponding mathematical property. "/" means that this mathematical property need not be considered.

3.2. Choice of Proper Time Scale Levels

As is known to all, the noises and main series in observed time series are generated by different physical mechanisms, *i.e.*, the main series are generated by a deterministic physical mechanism, while noises are generated by many random and uncertain factors, so they have obviously different variation characters. Concretely, the noises show random characters and mainly reflect the inherent uncertainties in nature, while the main series are composed of deterministic components and mainly reflect the deterministic characters of the variable studied [15-16]. When applying DWT to analyze observed time series data, the components under different time scale levels obviously show different characteristics, *i.e.*, the main series usually locates in bigger time scale levels and shows low-frequency characteristics, while noises usually locate in small time scale levels and show high-frequency characters.

Based on the obviously different characters of the main series and noises, and in order to reduce noises in observed time series data effectively, it is suggested that the maximum time scale level be determined according to both the time series data analyzed and the scales (*i.e.*, resolution) concerned in practical works. For the hydrologic series data whose data points are just dozens or hundreds, the maximum time scale level generally can be valued at 2 or 3 in practical hydrologic de-noising works.

3.3. Determination of Accurate Thresholds

In order to reduce or remove noises in time series data accurately, by employing information entropy theories to the de-noising process, a method of determining accurate wavelet coefficients thresholds was proposed in [9] and is used in this paper. The theoretical and physical basis of this method is that both the values and variation characters of wavelet coefficients of the main series and noises in observed time series data are obviously different, *i.e.*, when applying DWT to the time series data analyzed, small wavelet coefficients are assumed to be dominated by noises and carry little useful information, but the main series carry all useful information and are concentrated in a limited number of big wavelet coefficients [58]. Moreover, from the energy point of view, the energies of the main series are concentrated on several time scale levels corresponding to the periods and trends of series

data, but the energies of noises scatter in the whole time scales and decrease rapidly as the time scale level increases.

Based on this difference, information entropy can be applied to the wavelet de-noising process. The main idea of the method proposed is to use entropy value H obtained by POME [36] to describe the random characters of the noises separated, and use wavelet energy entropy (WEE) [70] to describe the degrees of complexity of the main series reconstructed first, and then, according to the variations of noises' H and main series' WEE along with the increasing of wavelet coefficients thresholds, the separation process of noises in time series can be described and understood. After the noises are removed completely, values of H and WEE would become constants within a certain set of thresholds; and these thresholds can be regarded as the most reasonable final results:

$$H = -\int f(x) \ln(f(x)) dx \tag{10}$$

$$WEE = -\sum_{j=1}^{M} P_j \ln P_j \quad with \quad P_j = E_j / \sum_{j=1}^{M} E_j = \left(\sum_{k=1}^{K_j} w'^2_{j,k} \right) \Big/ \sum_{j=1}^{M} \left(\sum_{k=1}^{K_j} w'^2_{j,k} \right) \tag{11}$$

where *f(x)* is the probability density function used for describing random characters of noises. $W'_{j,k}$ are the wavelet coefficients of DWT adjusted by a certain thresholding rule. *M* is the maximum time scale level, and K_j is the number of wavelet coefficients in time scale level *j*. Besides, although the approach used for calculation of entropy value H by POME has been illustrated in many papers, it is described briefly again in Appendix A, mainly to help readers understand the new method proposed more clearly, and also for keeping the integrity of the contents about wavelet de-noising.

The information entropy theories are employed to determine wavelet coefficient thresholds for two main purposes. The first is to use information entropy theories to describe the obviously different characteristics of noises and main series in observed time series data, which can provide a more reliable physical basis for the process of threshold estimation and de-noising, so the analysis results can be more reasonable in practice. The other purpose is that no matter what probability distribution noises follow in the time series data analyzed, *i.e.*, more than just the noises following second-order stationary process, and no matter what amount of noises is included in time series data analyzed, they can be described and analyzed by POME quantitatively and accurately, so the analysis results can also be more reliable, and the new entropy-based wavelet de-noising method proposed in the following can become more effective and more universal.

3.4. Choice of Suitable Thresholding Rules

In order to overcome the disadvantages of hard- and soft-thresholding rules used commonly, and also to avoid the difficulties of parameters estimation in many improved mid-thresholding rules, by comprehensive analysis, the Equation (12), which does not include any new parameter, is chosen and used in this paper [71]. As shown in Equation (12), the wavelet coefficients are adjusted by using themselves and the thresholds, when $T_j = 0$, it is just the same as Equation (8); and when $T_j = 1$, it is the same as Equation (9). Therefore, the Equation (12) is the combination of both hard- and

soft-thresholding rules, and the analysis results by using Equation (12) are continuous and can reduce or even remove the constant deviations between $W_{j,k}$ and real $W'_{j,k}$:

$$W'_{j,k} = \begin{cases} \mathrm{sgn}(W_{j,k})(|W_{j,k}| - \dfrac{T_j}{\exp(\dfrac{|W_{j,k}| - T_j}{T_j})}) & |W_{j,k}| > T_j \\ 0 & |W_{j,k}| < T_j \end{cases} \tag{12}$$

4. New Entropy-Based Wavelet De-noising Method

Based on both the basic idea of wavelet threshold de-noising and the discussion results of four key problems about wavelet de-noising above, a new entropy-based wavelet de-noising method is proposed as follows:

(1) Firstly, we choose reasonable wavelet function and determine the proper time scale levels, then analyze the time series data by DWT in Equation (4) and obtain high frequency wavelet coefficients $W_{j,k}$ under time scale level j ($j = 1, 2, \ldots, M$).

(2) We set the same threshold T for different time scale level j [72] and use a certain small threshold T to adjust $W_{j,k}$ according to Equation (12). Then we use $W'_{j,k}$ to reconstruct the main series by Equation (7), and regard the difference between observed time series and reconstructed main series as noises.

(3) We determine the proper probability density function to describe the random characters of noises separated by using H in Equation (10), as described in appendix A, and describe the complexity degrees of the reconstructed main series by using WEE in Equation (11).

(4) The threshold value T in step (2) is increased gradually, and for each threshold value, we do the same analysis according to the steps (2) and (3), and then get two series of H and WEE values corresponding to a set of thresholds.

(5) After noises in the observed time series are removed completely, both the values of noises' H and the reconstructed main series' WEE would become constants, so the threshold T^* corresponding to the constants of H and WEE is the most reasonable threshold.

(6) We use threshold T^* to reduce noises in the observed time series data analyzed, and separate the main series and noises.

(7) We judge whether the de-noising results are reasonable or not by using the criterion proposed in [9] initially, and moreover, the prior information and experiences about the series data analyzed are used to judge the reasonability of the de-noising results further. If not, we do the same analysis according to steps (1)-(6), until accurate de-noising results are obtained.

Besides, the analysis process of time series data by using the new entropy-based wavelet de-noising method proposed is also depicted in Figure 1.

Figure 1. The analysis process of time series data by the new entropy-based wavelet de-noising method proposed (in the black pane, the analysis processes are information entropy theories based).
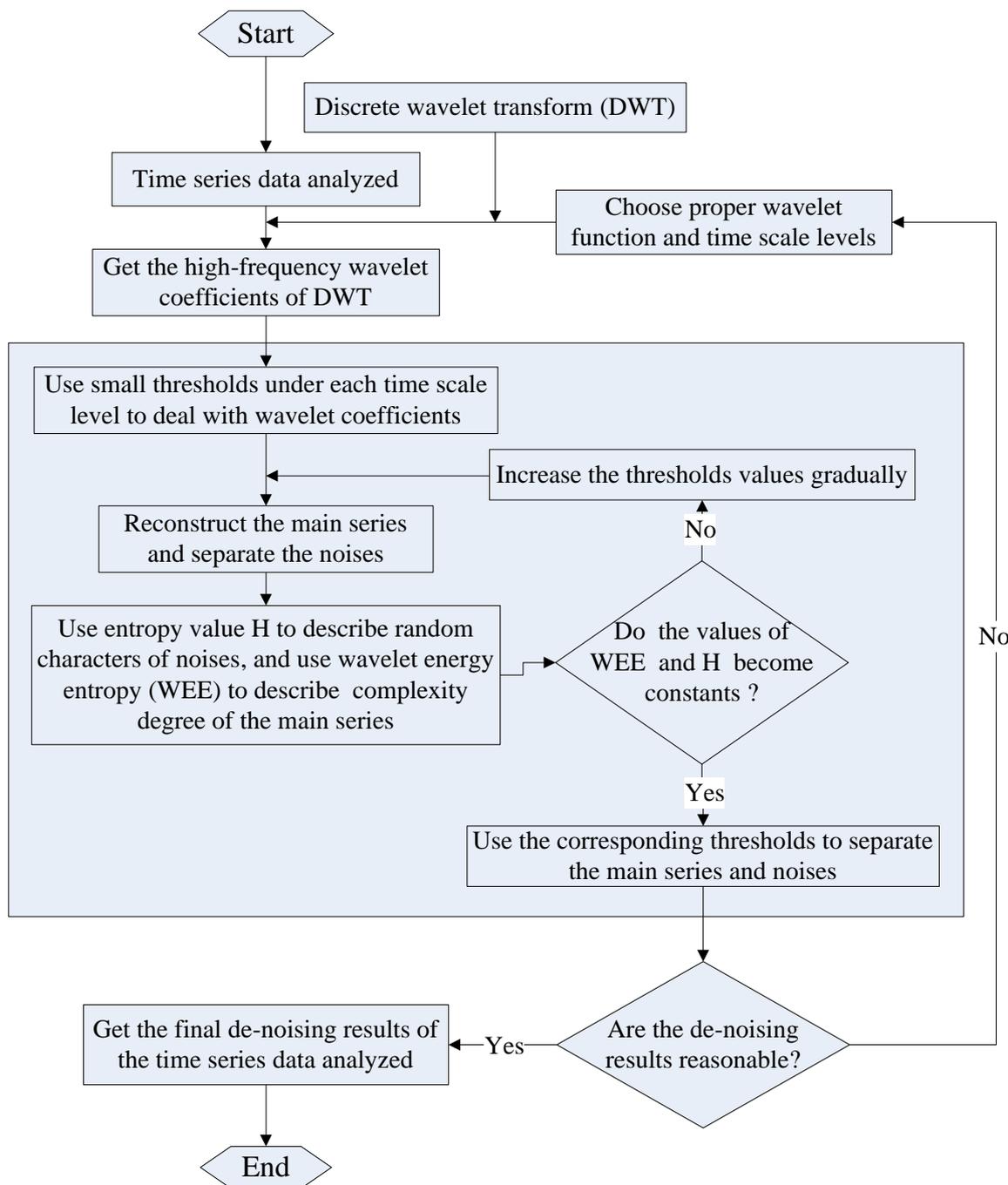
As described above, because information entropy theories are mainly used to describe the uncertainties of noises and the complexities of main series, and then based on the different characters of noises and main series in observed time series data to de-noising, it holds that the new method proposed has a reliable physical basis and the analysis results are reasonable and are the global optimum. Besides, the analysis process of the new method is simple and is easy to implement, so it is more applicable and useful in applied sciences and engineering practice. However, it should be pointed out that when using this new method, great attention should be paid to the choice of proper probability

distributions to describe noises. In practical works, in order to obtain accurate threshold estimation results, on one hand, as much prior information and experience as possible should be used to determine the proper probability distribution, which are also very important in the de-noising process by traditional wavelet de-noising methods; on the other hand, it is suggested that several probability distributions be used together, and by comprehensively comparing the analysis results of different distributions, the most reasonable results could be obtained finally.

Nevertheless, it should also be pointed out that since the basic idea of the new wavelet de-noising method proposed is based on the difference of wavelet coefficients' values and energies about the main series and noises in original time series data to de-noising, this new method has its own applicable condition: when there are too many noises in the time series data analyzed, *i.e.*, the wavelet coefficients of noises are close to or even much bigger than the coefficients of the main series, the energy of noises would be much bigger than that of the main series, and the main parts of time series data become noises, but not the main series. In these situations, the main series are submerged completely in the noises so cannot easily be identified by the new method proposed. Besides, because the amounts of real signals in different observed series data are unknown, it is difficult to determine the cutoff of SNR (signal to noises ratio) of the applicable condition. But from another point of view, in the authors' opinion, these series greatly contaminated by noises can be regarded as pure random series and then analyzed by proper statistical methods, and there is no need to reduce or remove noises again in practical works.

5. Case Studies

In order to verify the new entropy-based wavelet de-noising method proposed in this paper, both synthetic series and observed time series data are analyzed by different methods, and the results are compared and discussed in detail, all of which will be done in the following sections.

5.1. Synthetic Series Analysis

Two different synthetic series, S1 and S2 for short, were generated by Monte-Carlo method. Among them, noises in the S1 series follow a normal probability distribution of $N \sim (0, 5)$, while noises in the S2 series follow a Pearson-III (P-III for short) probability distribution of $P \sim (0, 8, 0.5)$, and their SNR are 9.51 and 6.27, respectively. Since the two synthetic series include different noises, they can be used to judge whether the new method proposed is suitable for analyzing different noises or not. Moreover, because the real series data in the two synthetic series are known clearly, the de-noised series (*i.e.*, the main series) can be compared with the real series data, and the performances of these de-noising methods used can be understood further.

Firstly, the "db4" mother wavelet function is chosen and the maximum time scale level 5 is used, then DWT is applied to the two synthetic series. Based on the DWT results, noises in both S1 series and S2 series are separated by using the new method proposed, and the results are shown in Figure 2 and Figure 3, respectively. Besides, the two synthetic series are also analyzed by using three other typical wavelet de-noising methods, namely the UT, HSURE (heuristic SURE) and MIN, and the statistical characteristic values, including mean ($\overline{X}$), standard deviation (σ), coefficient of skewness (C_s) and the first-order autocorrelation coefficient (r_1), of original synthetic series, main series and

noises obtained by different methods are calculated and summarized in Table 2 and Table 3, respectively. Furthermore, the de-noised series data obtained by different methods are also compared with the real series data by using the quantitative indicator of mean square error (MSE) in Equation (13), whose value can reflect the similar degrees of two series data to a certain extent:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x(i) - y(i))^2$$

(13)

where $x(i)$ and $y(i)$ are the two series data analyzed, and n is the length of series data $x(i)$.

Table 2. Statistical characteristics of the de-noising results of the S1 series obtained by different methods.

De-noising method	Series' type	Statistical characteristic values				
		$\overline{X}$	σ	C_s	r_1	MSE
	S1 series	75.05	20.46	−0.005	0.93	37.07
New method proposed	Main series	73.32	18.85	−0.003	0.98	8.21
	Noises	1.73	4.91	0.010	0.08	
UT	Main series	74.27	22.04	0.007	0.92	21.99
	Noises	0.78	6.65	−0.020	0.42	
HSURE	Main series	74.46	21.75	0.013	0.97	17.73
	Noises	0.59	4.01	−0.165	−0.26	
MIN	Main series	74.53	22.24	0.010	0.96	10.36
	Noises	0.52	5.44	−0.104	0.26	

Table 3. Statistical characteristics of the de-noising results of the S2 series obtained by different methods.

De-noising method	Series' type	Statistical characteristic values				
		$\overline{X}$	σ	C_s	r_1	MSE
	S2 series	174.77	45.60	−0.02	0.96	112.05
New method proposed	Main series	173.95	45.09	−0.02	0.98	22.35
	Noises	0.82	7.58	0.51	−0.03	
UT	Main series	174.77	44.11	−0.01	0.97	99.23
	Noises	−0.00	10.97	0.39	0.30	
HSURE	Main series	174.78	44.63	−0.03	0.99	26.09
	Noises	−0.01	8.28	0.35	−0.07	
MIN	Main series	174.78	44.22	−0.02	0.99	55.02
	Noises	−0.01	9.70	0.38	0.16	

Figure 2. The synthetic S1 series data (upper) and the de-noising results of S1 series (lower) by using the different methods (in synthetic series S1, noises follow a normal probability distribution).
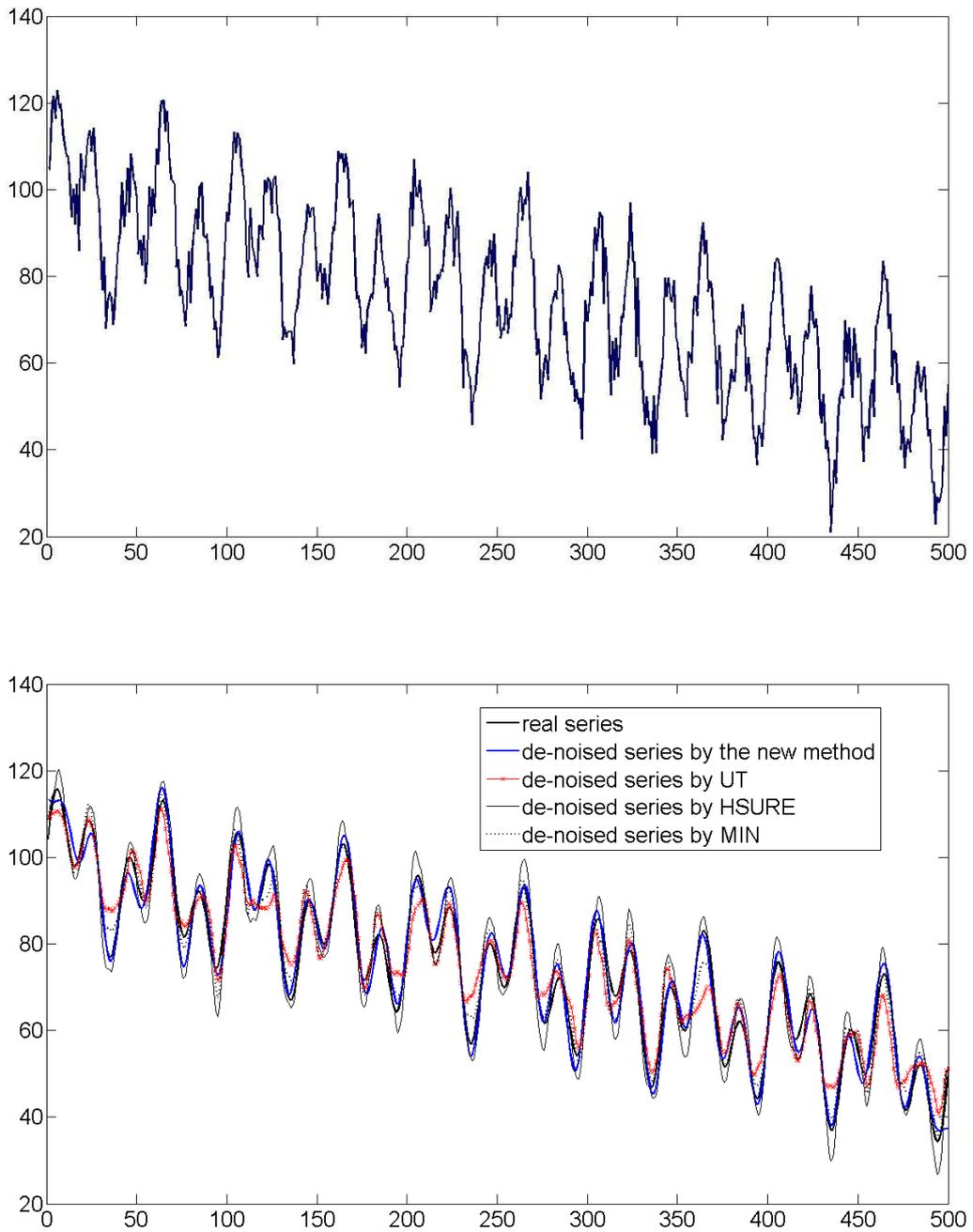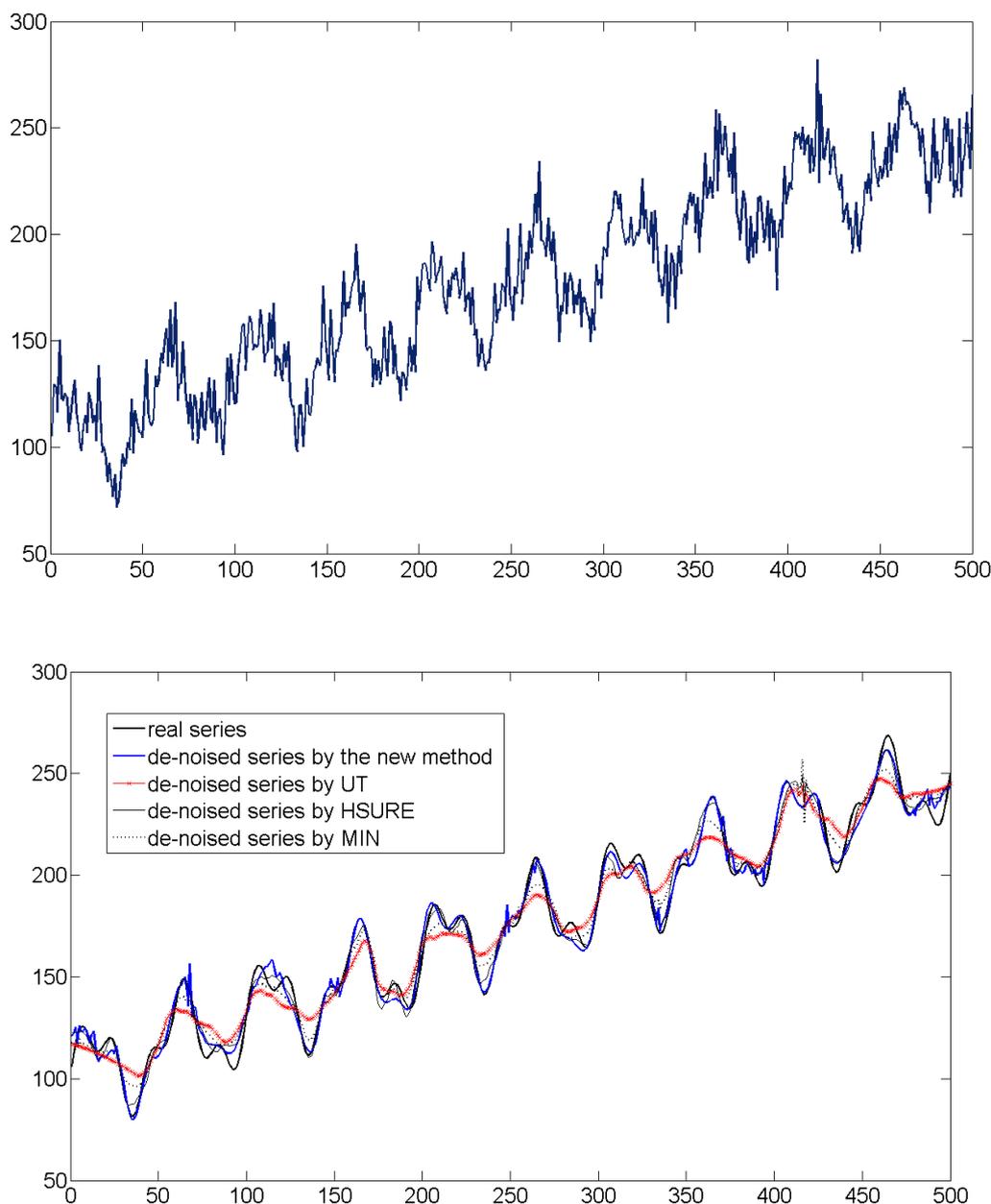
Figure 3. The synthetic S2 series data (upper) and the de-noising results of S2 series (lower) by using the different methods (In synthetic series S2, noises follow a P-III probability distribution).

By comparing and discussing the analysis results of the two synthetic series comprehensively, it can be found that: (1) when using the new method proposed to analyze both the S1 series and the S2 series, the statistical characteristic values of original synthetic series and the main series are very close, and the noises separated show pure random characters. Taking the S1 series for example, the r_1 values of S1 series and the main series are 0.93 and 0.98, respectively; while the r_1 value of the noises separated is 0.08. Besides, the analysis results plotted in Figure 2 and Figure 3 also show that the two de-noised series obtained by the new method are very similar to the corresponding real series data, respectively. Thus it is thought that the de-noising results are accurate, and which indicate the reliability of the new method proposed; (2) no matter whether we reduce normal noises in the S1 series or skewed noises in

the S2 series, the analysis results of the new entropy-based wavelet de-noising method are in good accord with the criterion proposed in reference [9]. Therefore, it can hold that the new method proposed not only has its own effectiveness but also has good universality; (3) noises separated from the synthetic series by traditional methods (UT, HSURE and MIN) are different and show good auto-correlations. For examples, r_1 values of noises separated from S1 series are 0.42, –0.26 and 0.26 corresponding to UT, HSURE and MIN, respectively. It means that the analysis results of traditional wavelet de-noising methods are not reasonable and these results should be viewed with caution when used; (4) for the de-noised results of the S1 series and the S2 series by the new method, the values of MSE are 8.21 and 22.35, respectively, which are the lowest in all the analysis results of the methods used. It means that the de-noised series obtained by the new method proposed are the most similar to the real series data, so the analysis results are the most accurate and reliable; (5) by comparing with the analysis results of different de-noising methods, it shows much better performances of the new method proposed in de-noising than other traditional wavelet de-noising methods.

5.2. Observed Time Series Analysis

Two hydrologic time series data, RS1 and RS2 for short, are also analyzed by different methods to further verify the performances of the new method proposed. The two hydrologic series data have complex non-stationary and multi-temporal scale characters and are the most representative observed time series data, so in the authors' opinion, it is deemed that if suitable for analyzing the two hydrologic series data here, the new method proposed can also be used to analyze other observed time series data accurately in practical works.

As illustrated in [9], RS1 presents 20 years (1978-1997) of monthly runoff series measured at the Dashankou hydrologic station at Kaidu River in Xinjiang province in the northwest of China. There are two recharge sources about Kaidu River, one is snowmelt from Tianshan Mountain, mainly happening from March to April every year, and the other is rainfall, mainly happening at August every year. Consequently two flood seasons happen in every year, and the RS1 series has two obvious periods: about 6 months and 12 months. RS2 presents 54 years (1950-2003) of annual runoff series measured at the Lijin hydrologic station at the estuary area of the Yellow River watershed in the north of China. The Yellow River, the second largest river in China, is an important water source in North China. After the 1970s, because of the great influence of human activities and climatic conditions changes in this area, runoff in the middle and lower Yellow River became seasonal and even presents a cutting-off trend, which causes serious sediment problems and eco-environment problems. Hydrologic regimes in the estuary area are controlled by the whole Yellow River watershed. In present studies, it is shown that the runoff in the Yellow River mainly has four dominant periods: 3, 7, 11 and 18 years.

Analysis of the variation characters (such as periods) of RS1 and RS2 series have great significance in understanding the physical hydrologic processes and for water resources management, as well as many other practical hydrologic works. However, due to the influence of noises, the periods of the two hydrologic series data cannot be identified accurately when analyzing the raw series data directly, especially when analyzing the RS2 series. If the raw series data is de-noised first by a certain method and then periods could be identified accurately, it can be deemed that the de-noising results are reliable and the corresponding de-noising method is effective.

Figure 4. The de-noising results of RS1 series and RS2 series by different methods (in Figure 4, "Method*" is the method proposed in reference [9]).
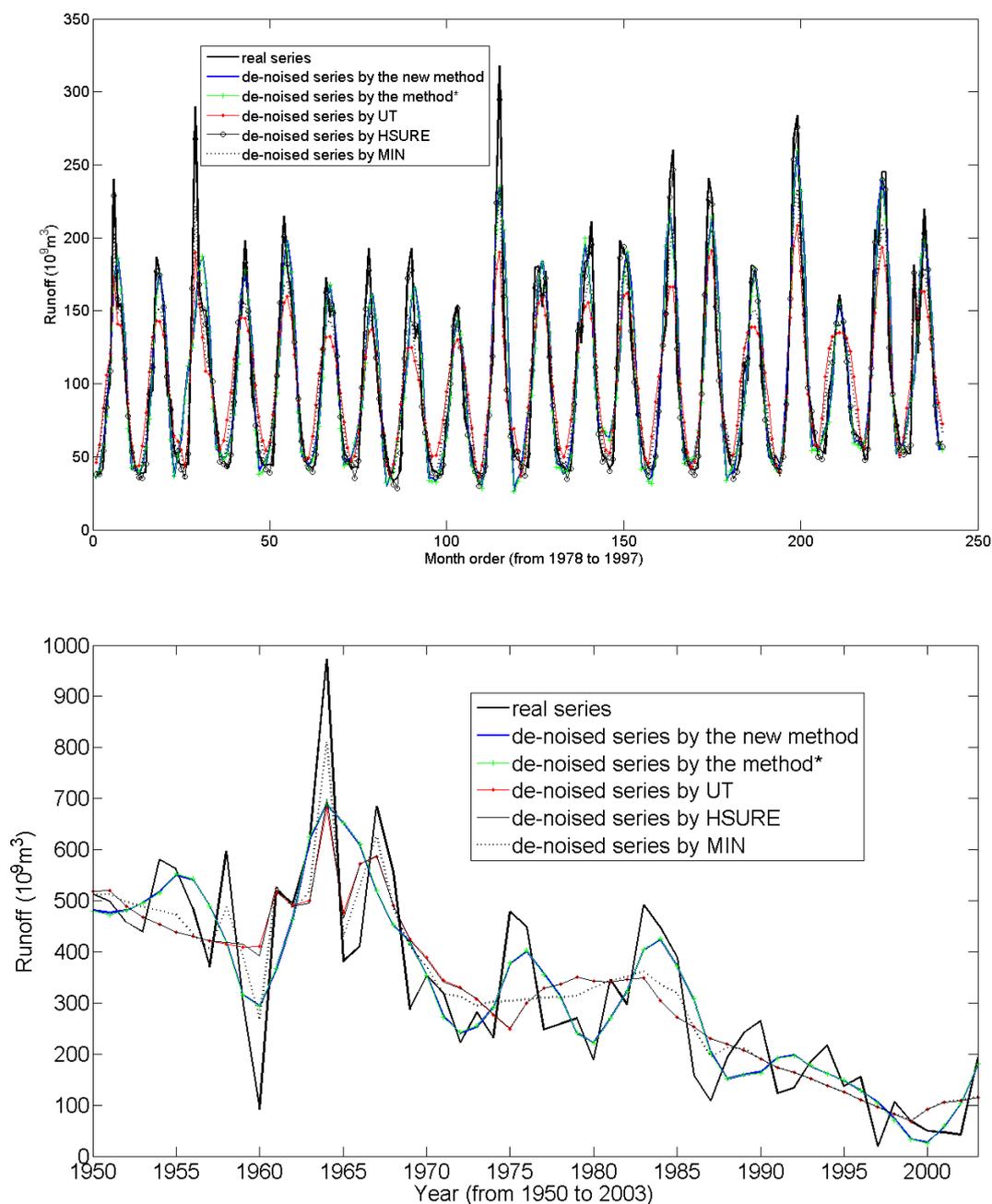
According to the analysis results in [9], here, the P-III probability distribution is used to describe the random characters of noises in the RS1 series, and the normal probability distribution is used to describe the random characters of noises in the RS2 series. Then the two hydrologic series data are analyzed by the new method proposed and three other typical methods (UT, HSURE and MIN). During the analysis process, the "dmey" wavelet is chosen and the maximum time scale level 3 is used to analyze the RS1 series; and the "db2" wavelet is chosen and the maximum time scale level 2 is used to analyze the RS2 series. Finally, the de-noising results of the two observed hydrologic series by different methods are depicted in Figure 4, and the characteristic values about each of these series data and calculated and summarized in Table 4 and Table 5, respectively.

Table 4. Statistical characteristics of the de-noising results of the RS1 series obtained by different methods.

De-noising method	Series types	Statistical characteristic values			
		$\overline{x}$	σ	r_1	C_s
	Original series (RS1)	101.55	61.24	0.73	0.99
New method proposed	Main series	101.53	57.26	0.84	0.82
	Noises	0.02	14.86	−0.11	0.36
Method*	Main series	101.53	58.32	0.82	0.77
	Noises	0.02	15.07	−0.16	0.33
UT	Main series	101.74	40.97	0.84	0.35
	Noises	−0.19	28.08	0.42	1.15
HSURE	Main series	101.57	59.14	0.79	0.83
	Noises	−0.02	7.00	−0.39	0.29
MIN	Main series	101.76	48.24	0.82	0.57
	Noises	−0.29	19.42	0.29	0.96

Table 5. Statistical characteristics of the de-noising results of the RS2 series obtained by different methods.

De-noising method	Series types	Statistical characteristic values			
		$\overline{x}$	σ	r_1	C_s
	Original series (RS2)	324.48	194.97	0.64	0.69
New method proposed	Main series	323.09	164.26	0.87	0.68
	Noises	1.39	65.37	−0.10	0.14
Method*	Main series	322.81	172.39	0.86	0.70
	Noises	1.67	63.31	−0.13	0.11
UT	Main series	328.75	150.43	0.91	0.32
	Noises	−4.27	88.34	0.34	0.30
HSURE	Main series	325.45	179.43	0.79	0.61
	Noises	−0.93	44.45	−0.44	0.04
MIN	Main series	327.57	165.06	0.81	0.60
	Noises	−3.09	61.28	−0.21	0.08

Note: in Table 4 and Table 5, "Method*" is the method proposed in reference [9].

Analysis results in Table 4 and Table 5 show that because of the use of Equation (12) to adjust the high frequency wavelet coefficients of DWT, the de-noising results of the new method proposed in the present paper are a little better than those obtained from [9] and much better than three other traditional methods. Besides, analysis results show that the noises separated from RS1 series follow a skew probability distribution since the value of C_s is bigger than 0.3. Furthermore, the statistical characteristic values of original observed series data, the main series and noises accord well with the criterion proposed in [9], so it can be deemed that the de-noising results of the two hydrologic time

series data are also reasonable and accurate, and the new method proposed is reliable and effective for de-noising. Finally, it can be found that although the real series data in the two observed hydrologic series data are unknown, Figure 4 shows that compared with the analysis results of other methods, trends of the de-noised series obtained by the new method proposed are more in accordance with the trends of the observed series data as a whole, which means that the new method proposed is comparatively more reliable, and moreover, because noises are reduced accurately and reliably, the periods of the two observed time series data can be identified accurately, as discussed in [9]. But for the de-noised series of other methods as shown in Figure 4, they have a little big difference with the de-noised series of the new method, which mean that they also include certain amount of noises or lose some real signals, so all the periods cannot be identified by using them. Since the issue of periods' identification is far beyond the scope of the present paper, more details about which can be found in detail in reference [9].

6. Summary and Conclusions

The authenticity and reliability of observed time series data are the very important basis of many applied research and engineering works. In practice, the existence of noises contaminates the real series data and causes many difficulties in time series analysis. When using traditional methods to reduce or remove noises in time series data, the results cannot meet the practical needs. In this paper, in order to overcome the disadvantages of traditional methods and to obtain accurate de-noising results of time series data, by employing information entropy theories to describe the obviously different characters of noises and main series, a new entropy-based wavelet de-noising method has been proposed. By analyzing both synthetic series and typical observed time series data, the performance of the new method proposed has been verified. By comprehensive analysis, the following conclusions about the new method proposed can be drawn: first, because of its basis on information entropy theories to describe the obvious difference of noises and main series in observed series data and then de-noising, the analysis process has a more reliable physical basis and the results of the new method are the global optimum in the whole aspect; secondly, compared with traditional methods, the de-noising results of the new method are more accurate and more reasonable; thirdly, since can be used to analyze both normal noises and skewed noises accurately, the new method shows good effectiveness and universality; and fourthly, the analysis process of the new method is simple and is easy to implement, so it is more applicable and useful in practical applied sciences and engineering works, and therefore, it can be used in future practical applications.

Nevertheless, great attention should be paid to several detailed problems when using the new method, such as determination of the proper probability distribution to describe noises, and choice of a reasonable wavelet function and time scale levels. Only by analyzing and solving these detailed problems accurately, reliable and reasonable de-noising results could be obtained finally.

Acknowledgements

References

1. Sassi, R.; Corino, V.D.A; Mainardi, L.T. Analysis of surface atrial signals: time series with missing data? *Ann. Biomed. Eng.* **2009**, *37*, 2082-2092.
2. Delgado, M.A. Recent advances in time series analysis: A volume honoring Peter M. Robinson introduction. *J. Econometrics* **2009**, *151*, 99-100.
3. Boesveldt, S.; Stam, C.J.; Knol, D.L.; Verbunt, J.P.A.; Berendse, H.W. Advanced time-series analysis of MEG data as a method to explore olfactory function in healthy controls and Parkinson's disease patients. *Hum. Brain Mapp.* **2009**, *30*, 3020-3030.
4. Warner, J.; Gmel, G.; Graham, K.; Erickson, B. A time-series analysis of war and levels of interpersonal violence in an English military town, 1700–1781. *Soc. Sci. Hist.* **2007**, *31*, 575-602.
5. Houle, T.T.; Lauzon, J.J.; Penzien, D.B.; Rains, J.C.; Harden, R.; Zinke, J. Time-series analysis: mathematical simulation demonstrates different data reduction/analytic strategies indicated for episodic *vs*. chronic headache. *Headache* **2005**, *45*, 804-804.
6. Navarro-Esbri, J.; Diamadopoulos, E.; Ginestar, D. Time series analysis and forecasting techniques for municipal solid waste management. *Resour. Conserv. Recycl.* **2002**, *35*, 201-214.
7. Berendsen, H.J.C.; Vanderspoel, D.; Vanderunen, R. Gromacs—a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43-56.
8. Dockery, D.W.; Pope, C.A.; Xu, X.P.; Spengler, I.D.; Ware, J.H.; Fay, M.E.; Ferris, B.G.; Speizer, F.E. An association between air-pollution and mortality in 6 United States cities. *N. Engl. J. Med.* **1993**, *329*, 1753-1759.
9. Sang, Y.F.; Wang, D.; Wu, J.C.; Zhu, Q.P.; Wang, L. The relation between periods' identification and noises in hydrologic series data. *J. Hydrol.* **2009**, *368*, 165-177.
10. Schreiber, T. Extremely simple nonlinear noise-reduction method. *Phys. Rev. E* **1993**, *47*, 2401-2404.
11. D'Astous, F.; Hipel, K.W. Analyzing environmental time series. *J. Environ. Engng. Div. ASCE* **1979**, *105*, 979-992.
12. Sang, Y.F.; Wang, D. A stochastic model for mid-to-long-term runoff forecast. *Proc. Int. Conf. Nat. Comput.* **2008**, *3*, 44-48.
13. Kipinski, L. Time series analysis of nonstationary data in encephalography and related noise modelling. In *Recent Advances in Mechatronics*; Springer: Berlin, Heidelberg, Germany, 2007; pp. 238-242.
14. Wang, B.; Yan, S.P.; Wu, X.Q. Effects of cross correlated noises on the mean first-passage time of optical bistable system. *Acta Phys. Sin.* **2009**, *58*, 5191-5195.

15. Yevjevich, V. *Stochastic Process in Hydrology*; Water Resources Publications: Highlands Ranch, CO, USA, 1972.

16. Ding, J.; Deng Y.R. *Stochastic Hydrology*; Science and Technology Press of the University of Chengdu: Chengdu, China, 1988. (in Chinese).

17. Elshorbagy, A.; Simonovic, S.P.; Panu, U.S. Noise reduction in chaotic hydrologic time series: facts and doubts. *J. Hydrol.* **2002**, *256*, 147-165.

18. Jayawardena, A.W.; Gurung, A.B. Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach. *J. Hydrol.* **2000**, *228*, 242-264.

19. Donoho, D.H. De-noising by soft-thresholding. *IEEE T. Inform. Theory* **1995**, *41*, 613-617.

20. Ahalpara, D.P.; Verma, A.; Parikh, J.C.; Panigrahi, P.K. Characterizing and modelling cyclic behavior in non-stationary time series through multi-resolution analysis. *Pramana-J. Phys.* **2008**, *71*, 459-485.

21. Ombao, H.; von Sachs, R.; Guo, W.S. SLEX analysis of multivariate nonstationary time series. *J. Am. Statist. Assoc.* **2005**, *100*, 519-531.

22. Pinnegar, C.R.; Mansinha, L. Time-local spectral analysis for non-stationary time series: The S-transform for noisy signals. *Fluct. Noise Lett.* **2003**, *3*, L357-L364.

23. Torrence, C.; Compo, G.P. A practical guide to wavelet analysis. *Bull. Amer. Meteorol. Soc.* **1998**, *79*, 61-78.

24. Labat, D. Recent advances in wavelet analyses: Part 1. A review of concepts. *J. Hydrol.* **2005**, *314*, 275-288.

25. Daubechies, I. The wavelet transform, time-frequency localization and signal analysis. *IEEE T. Inform. Theory* **1990**, *36*, 961-1005.

26. Daubechies, I. *Ten lectures on wavelets*; SIAM: Philadelphia, PA, USA, 1992; p. 357.

27. Kumar, P.; Foufoula-Georgiou, E. Wavelet analysis for geophysical applications. *Rev. Geophys.* **1997**, *35*, 385-412.

28. Vidakovic, B. *Statistical Modeling by Wavelet*; Wiley: New York, NY, USA, 1999.

29. Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2000.

30. Antoniadis, A.; Bigot, J.; Sapatinas, T. Wavelet estimators in nonparametric regression: A comparative simulation study. *J. Stat. Softw.* **2001**, *6*, 1-83.

31. Lindblad, G. Entropy, information and quantum measurements. *Commun. Math. Phys.* **1973**, *33*, 305-322.

32. Plastino, A.R.; Plastino, A. Information-theory, approximate time-dependent solutions of boltzmann-equation and tsallis entropy. *Phys. Lett. A* **1994**, *193*, 251-258.

33. Chatzisavvas, K.C.; Moustakidis, C.C.; Panos, C.P. Information entropy, information distances, and complexity in atoms. *J. Chem. Phys.* **2005**, *123*, 174111.

34. Harremoes, P.; Vajda, I. Efficiency of entropy testing. In *Proceedings of IEEE International Symposium on Information Theory 2008*, Toronto, Ontario, Canada, July 06–11, 2008; pp. 1-6, 2639-2643.

35. Wang D. Rethinking risk analysis: the risks of risk analysis in water issues as the case. *Hum. Ecol. Risk Assess.* **2009**, *15*, 1079-1083.

36. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620-630.

37. Singh, V.P. *Entropy-based parameter estimation in hydrology.* Kluwer Academic Publishers: Boston, MA, USA, 1998.

38. Harremoes, P. Maximum entropy on compact groups. *Entropy* **2009**, *11*, 222-237.

39. Hajicek, P. Quantum model of classical mechanics: Maximum entropy packets. *Found. Phys.* **2009**, *39*, 1072-1096.

40. Wang D.; Singh, V.P.; Zhu Y.S.; Wu, J.C. Stochastic observation error and uncertainty in water quality evaluation. *Adv. Water Resour.* **2009**, *32*, 1526-1534.

41. Boshnakov, G.N.; Lambert-Lacroix, S. Maximum entropy for periodically correlated processes from nonconsecutive autocovariance coefficients. *J. Time Ser. Anal.* **2009**, *30*, 467-486.

42. Wang D.; Singh, V.P.; Zhu Y.S. Hybrid fuzzy and optimal modeling for water quality evaluation. *Water Resour. Res.* **2007**, *43*, W05415.

43. Otwinowski, H. Maximum entropy method in comminution modeling. *Granul. Matter* **2006**, *8*, 239-249.

44. Mansour, Y.; Schain, M. Learning with maximum-entropy distributions. *Mach. Learn.* **2001**, *45*, 123-145.

45. Leon, Y.; Peeters, L.; Quinqu, M.; Surry, Y. The use of maximum entropy to estimate input-output coefficients from regional farm accounting data. *J. Agric. Econ.* **1999**, *50*, 425-439.

46. Burg. J.P.A. A new analysis technique for time series data. In *Proceedings of the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics*, Enschede, The Netherlands, August 12–23, 1986.

47. Sang, Y.F.; Wang, D. New method for estimating periods in hydrologic series data. *Proc. Int. Conf. Fuzzy Syst. Knowl. Discov.* **2008**, *5*, 645-649.

48. Pardo-Iguzquiza, E.; Rodriquez-Tovar, F.J. Maximum entropy spectral analysis of climatic time series revisited: Assessing the statistical significance of estimated spectral peaks. *J. Geophys. Res. Atmos.* **2006**, *111*, D10102.

49. Muto, S. Maximum-entropy spectral analysis of extended energy-loss fine structure and its application to time-resolved measurement. *Philos. Mag.* **2004**, *84*, 2793-2808.

50. Wang D.; Chen, Y.F.; Li, G.F.; Xu, Y.H. Maximum entropy spectral analysis for annual maximum tide levels time series of the Changjiang River Estuary. *J. Coast. Res.* **2004**, *SI43*, 101-108.

51. Pacey, M.N.; Wang, X.Z.; Haake, S.J.; Patterson, E.A. The application of evolutionary and maximum entropy algorithms to photoelastic spectral analysis. *Exp. Mech.* **1999**, *39*, 265-273.

52. Padmanabhan, G.; Rao A.R. Maximum entropy spectral analysis of hydrologic data. *Water Resour. Res.* **1988**, *24*, 1519-1533.

53. Bruni, V.; Vitulano, D. Wavelet-based signal de-noising via simple singularities approximation. *Signal Process.* **2006**, *86*, 859-876.

54. Chanerley, A.A.; Alexander, N.A. Correcting data from an unknown accelerometer using recursive least squares and wavelet de-noising. *Comput. Struct.* **2007**, *85*, 1679-1692.

55. Jansen, M. Minimum Risk Thresholds for data with heavy noise. *IEEE Signal Process. Lett.* **2006**, *13*, 296-299.

56. Donoho, D.L.; Johnstone, I.M. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Assoc.* **1995**, *90*, 1200-1244.

57. Zhao, M.; Kulasekera, K.B. Minimax estimation of linear functionals under squared error loss. *J. Stat. Plan. Infer.* **2009**, *139*, 3160-3176.

58. Jansen, M.; Bultheel, A. Asymptotic behavior of the minimum mean squared error threshold for noisy wavelet coefficients of piecewise smooth signals. *IEEE T. Signal Process.* **2001**, *49*, 1113-1118.

59. Jansen, M.; Malfait, M.; Bultheel, A. Generalized cross validation for wavelet thresholding. *Signal Process.* **1997**, *56*, 33-44.

60. Donoho, D.L.; Johnstone, I.M. Minimax estimation via wavelet shrinkage. *Ann. Stat.* **1998**, *26*, 879-921.

61. Bruce, A.G.; Gao H.Y. Waveshrink with firm shrinkage. *Stat. Sin.* **1997**, *4*, 855-874.

62. Gao, H.Y. Wavelet shrinkage denoising using the nonnegative garrote. *J. Comput. Graph. Stat.* **1998**, *7*, 469-488.

63. Johnstone, I.M.; Silverman, B.W. Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Stat.* **2004**, *32*, 1594-1649.

64. Labat, D.; Ababou, R.; Mangin, A. Rainfall-runoff relations for karstic springs. Part II: Continuous wavelet and discrete orthogonal multiresolution analyses. *J. Hydrol.* **2000**, *238*, 149-178.

65. Burrus, C.S.; Gopinath, R.A.; Guo, H.; Guo, H. *Introduction to wavelets and wavelet transforms: A primer*; Prentice-Hall: Upper Saddle River, NJ, USA, 1998.

66. Gaucherel, C. Use of wavelet transform for temporal characterization of remote watersheds. *J. Hydrol.* **2002**, *269*, 101-121.

67. Schaefli, B.; Maraun, D.; Holschneider, M. What drives high flow events in the Swiss Alps? Recent developments in wavelet spectral analysis and their application to hydrology. *Adv. Water Resour.* **2007**, *30*, 2511-2525.

68. Chui, C.K. *An introduction to wavelets*; Academic Press: San Diego, CA, USA, 1992; Vol. 1.

69. Sang, Y.F.; Wang, D. Wavelets selection method in hydrologic series wavelet analysis. *J. Hydraul. Eng.* **2008**, *39*, 295-300, 306. (in Chinese with English abstract)

70. Figliola, A.; Seerano, E. Analysis of physiological time series using wavelet transform. *IEEE Eng. Med. Biol. Mag.* **1997**, *16*, 74-79.

71. Ke, X.Z.; Wang, L.; Ni, G.R. Application of improved threshold de-noising based on wavelet transform to pulsar signal processing. *J. Xi'an Univ. Technol.* **2008**, *24*, 18-21. (in Chinese with English abstract)

72. Sang, Y.F.; Wang, D.; Wu, J.C. Analyses and description of the wavelet characters of white noises. *Water Syst. Water Resour. Manage.* 2009 (in press and in Chinese with English abstract).

Appendix A: Calculation of Entropy Value by using POME

The principle of maximum entropy (POME), proposed by Jaynes in 1957 [36], is mainly used to determine the least biased probability distribution of the random variable studied. By using POME, it is thought that the minimally prejudiced assignment of probabilities can be done since which maximizes the entropy subject to the given information.

Defining the probability density function (pdf) of random variable x as $f(x)$, based on the observed series sample X, the mathematical programming problem in Equation (A.1) can be established to determine the expression of $f(x)$, given that the necessary m linearly independent constraint conditions P_i have been obtained as shown in detail in reference [37]:

$$\text{max} \qquad H = -\int f(x)\ln[f(x)]dx$$

$$s.t. \quad \begin{cases} \int f(x)dx = 1 \\ \int f(x)p_1(x)dx = E[p_1(x)] = P_1 \\ \quad\vdots \\ \int f(x)p_j(x)dx = E[p_j(x)] = P_j \\ \quad\vdots \\ \int f(x)p_m(x)dx = E[p_m(x)] = P_m \end{cases} \tag{A.1}$$

We establish Lagrange function of the mathematical programming problem in Equation (A.1) as:

$$L(f) = -\int \left[\ln f(x) + (\lambda_0 - 1) + \sum_{j=1}^{m}\lambda_j p_j(x)\right]f(x)dx, \qquad j = 1,2,\cdots m \tag{A.2}$$

where $(\lambda_0 - 1)$, λ_1, …, λ_m are the Lagrange multipliers.

The variational derivative of the fonctionelle $L(f)$ is calculated as:

$$\partial L(f) = -\int \left[\ln f(x) + 1 + (\lambda_0 - 1) + \sum_{j=1}^{m}\lambda_j p_j(x)\right]\partial f(x) = 0 \tag{A.3}$$

According to the equations set in Equation (A.3), the final results of Equation (A.4) can be obtained, which is just the solution of the mathematical programming problem in Equation (A.1):

$$f(x) = \exp\left[-\lambda_0 - \sum_{j=1}^{m}\lambda_j p_j(x)\right] \tag{A.4}$$

The corresponding expression of entropy value H can be calculated by using Equation (A.5):

$$H = \lambda_0 + \sum_{j=1}^{m}\lambda_j p_j \tag{A.5}$$

By substituting the Equation (A.4) into Equation (A.1), we then get the relationship expression of the Lagrange multipliers as:

$$\lambda_0 = \ln \int \exp\left[-\sum_{j=1}^{m}\lambda_j p_j(x)\right] \tag{A.6}$$

The first-order partial derivatives of these Lagrange multipliers are calculated using Equation (A.7):

$$\frac{\partial \lambda_0}{\partial \lambda_j} = -E(p_j)$$
(A.7)

and the high-order partial derivatives can be calculated using Equation (A.8):

$$\begin{cases} \dfrac{\partial^2 \lambda_0}{\partial \lambda_j^2} = Var[p_j(x)] \\ \dfrac{\partial^2 \lambda_0}{\partial \lambda_j \lambda_r} = Cov[p_j(x)p_r(x)] \end{cases}$$
(A.8)

By solving Equation (A.7) and Equation (A.8) together, the values of Lagrange multipliers ($\lambda_0 - 1$), λ_1, …, λ_m can be estimated, and then the expression of $f(x)$ in Equation (A.4) can be determined, and also the entropy value H in Equation (A.5) can be calculated finally. This is the main analysis process of POME for determining the pdf of the random variable x studied. For the normal probability distribution and P-III distribution used in this paper, the brief processes of estimation of entropy values H are summarized in Table A1.

Table A1. Entropy functions of normal probability distribution and P-III probability distribution.

Analysis process	Probability distribution type	
	Normal distribution	**P-III distribution**
Probability density function	$f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp(-\dfrac{(x-\bar{x})^2}{2\sigma^2})$	$f(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)}(x-a_0)^{\alpha-1}.e^{-\beta(x-a_0)}$
Constraint conditions	$\begin{cases} \int_{-\infty}^{\infty} f(x)dx = 1 \\ \int_{-\infty}^{\infty} xf(x)dx = E[x] = \bar{x} \\ \int_{-\infty}^{\infty} x^2 f(x)dx = var[x] + (\bar{x})^2 = S_x^2 + (\bar{x})^2 \end{cases}$	$\begin{cases} \int_{c}^{\infty} f(x)dx = 1 \\ \int_{c}^{\infty} xf(x)dx = E(x) \\ \int_{c}^{\infty} \ln(x-c)f(x)dx = E[\ln(x-c)] \end{cases}$
Lagrange multipliers	$f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2)$	$f(x) = \exp[-\lambda_0 - \lambda_1 x - \lambda_2 \ln(x-c)]$
Equations of parameters' estimation	$\begin{cases} \bar{x} = E[x] \\ S_x^2 = var[x] \end{cases}$	$\begin{cases} \bar{x} = c + ab \\ S_x^2 = a^2 b \\ E[\ln(x-c)] = \ln a + \psi(b) \end{cases}$
Entropy function	$H(x) = \ln[S_x \sqrt{2\pi e}]$	$H(x) = \ln[a^b \Gamma(b)] + \left(\dfrac{E[x]-c}{a}\right) - (b-1)E[\ln(x-c)]$