

Article

Tsallis Wavelet Entropy and Its Application in Power Signal Analysis

Jikai Chen * and Guoqing Li

School of Electrical Engineering, Northeast Dianli University, Jilin 132012, China;

E-Mail: LGQ@mail.nedu.edu.cn

* Author to whom correspondence should be addressed; E-Mail: funcjk@163.com;
Tel.: +86-432-64806691; Fax: +86-432-64806691.

Received: 29 March 2014; in revised form: 29 March 2014 / Accepted: 20 May 2014 /

Published: 27 May 2014

Abstract: As a novel data mining approach, a wavelet entropy algorithm is used to perform entropy statistics on wavelet coefficients (or reconstructed signals) at various wavelet scales on the basis of wavelet decomposition and entropy statistic theory. Shannon wavelet energy entropy, one kind of wavelet entropy algorithm, has been taken into consideration and utilized in many areas since it came into being. However, as there is wavelet aliasing after the wavelet decomposition, and the information set of different-scale wavelet decomposition coefficients (or reconstructed signals) is non-additive to a certain extent, Shannon entropy, which is more adaptable to extensive systems, couldn't do accurate uncertainty statistics on the wavelet decomposition results. Therefore, the transient signal features are extracted incorrectly by using Shannon wavelet energy entropy. From the two aspects, the theoretical limitations and negative effects of wavelet aliasing on extraction accuracy, the problems which exist in the feature extraction process of transient signals by Shannon wavelet energy entropy, are discussed in depth. Considering the defects of Shannon wavelet energy entropy, a novel wavelet entropy named Tsallis wavelet energy entropy is proposed by using Tsallis entropy instead of Shannon entropy, and it is applied to the feature extraction of transient signals in power systems. Theoretical derivation and experimental result prove that compared with Shannon wavelet energy entropy, Tsallis wavelet energy entropy could reduce the negative effects of wavelet aliasing on accuracy of feature extraction and extract transient signal feature of power system accurately.

Keywords: Tsallis entropy; transient signal; wavelet entropy; data mining; feature extraction

1. Introduction

Transient feature extraction is an important part of signal analysis, and as a new feature extraction algorithm, wavelet entropy has attracted the attention of experts and scholars all over the world. Wavelet entropy is a combination of wavelet decomposition and entropy statistics theories, and it has the advantages of multi-resolution analysis and complexity evaluation for time-varying signals, which means that the macro and micro aspects of some special signals could be researched in the time-frequency domain. For the above reasons, wavelet entropy has gradually been used in engineering signal surveys such as electroencephalography (EEG) testing, machinery vibration detection, power system fault diagnosis and so on [1–3]. Various studies have also manifested the better performance of wavelet entropy in analyzing the variability and complexity of climate processes compared with traditional methods [4–6]. For estimating the cognitive workload of human brains, wavelet entropy is used to extract the variation features from EEG signals, and for a subject-independent multi-channel classification scheme, the entropy features achieved high accuracy, up to 98% for channels from the frontal lobes, in the delta frequency band [7]. Bing *et al.* [8] presented a sensor fault detection method based on wavelet entropy and applied it to fault diagnosis of micro-gas turbine engine sensors. Remarkably, Zhenyou *et al.* [9,10] introduced wavelet entropy to the signal analysis of power systems, which makes it possible for some special power signals, including voltage dips, voltage interruptions, voltage flicker, voltage pulse and so on, to be analyzed more extensively and deeply than ever.

However, most research on wavelet entropy involves engineering applications of Shannon wavelet energy entropy (SWEE), but its physical meanings, working mechanism, and application principle have not been well discussed yet. In fact, Shannon wavelet energy entropy has some disadvantages in processing non-stationary signals, which could result in inaccurate or wrong results.

In view of the abovementioned facts, in this paper, the theoretical basis of SWEE has been analyzed, and the emphasis of research was placed on the working mechanism for feature extraction of transient signals. Wavelet aliasing's negative effect on feature extraction, which uses Shannon wavelet entropy, is discussed to find the primary reason for the problem.

Because Tsallis entropy is good at expressing the uncertainty of generalized systems, with a combination of Tsallis entropy and wavelet decomposition, a wavelet entropy-Tsallis wavelet energy entropy (TWEE) is constructed and the physical meaning of its algorithm is also explained. Starting with an analysis of wavelet aliasing and feature extraction effects, this paper studied the working mechanism and the suitable scope of TWEE, analyzes the relationship and difference between TWEE and SWEE and provides the initial principle of non-extensivity index selection. Finally, for analyzing transient voltage fluctuations caused by non-faulty indirect lightning strikes, TWEE is applied to transient feature extraction of lightning strikes. The corresponding voltage data when a non-faulty indirect lightning strike took place on 110 kV overhead distribution line in Guangdong, China, is collected and analyzed using TWEE, and the comparison of feature extraction, based upon TWEE and SWEE, has been carried out. The experimental results show that TWEE is better than SWEE in analyzing transient signals.

2. SWEE and Its Theoretical Defects

2.1. The Definition of SWEE

Shannon wavelet entropy (SWE) is the cooperative product of wavelet decomposition and Shannon entropy theory. By changing the segment of wavelet coefficients at the corresponding scales, and for different statistic objects using Shannon entropy, several derived algorithms are proposed based on SWE theory. Among these, SWEE is one of the most widely used algorithms in all fields [10]. The definition of SWEE is given as follows: first, a sliding window (width $w \in N$) is defined upon wavelet coefficients $D = \{d_j(k), k = 1, \dots, L; j = 1, \dots, S\}$. Then, suppose that:

$$E(m) = \sum_{j=1}^S E_m(j)$$

represents the energy sum of D in sliding window, where:

$$E_m(j) = \sum_{k=1+m\delta}^{w+m\delta} (d_j(k))^2$$

represents the energy branch at scale j . At the moment $w/2 + m\delta$, SWEE is defined as:

$$W_{EE}(m) = -\sum_{j=1}^S p_m(j) \log_e(p_m(j)) \quad (1)$$

where L is the length of $d_j(k)$, δ is the sliding factor, S is the maximum of scale number, m is the number of sliding step, $p_m(j) = E_m(j) / E(m)$ represents the distribution of wavelet energy at different scales, with $\sum_{j=1}^S p_m(j) = 1$ [9].

2.2. The Theoretical Defects of SWEE

In the macro-perspective, SWEE is the data mining of wavelet coefficients (or reconstructed signals) base on Shannon entropy statistics, so it is inevitable that Shannon WE inherits the statistic properties of Shannon entropy. As well known, Shannon entropy is the extended application for B-G entropy in information theory. As an important index in thermodynamics, B-G entropy represents the measurement for the uncertainty of N-energy-level system, and is defined as:

$$S_{BG} = -k \sum_{i=1}^N p_i \log_a p_i \quad (2)$$

where p_i represents the probability of No. i system status and k is Boltzmann constant. Shannon was inspired to induce B-G entropy to measure the information loss in information theory, and given the definition of information entropy-Shannon entropy as following: suppose that a probability measure on X , as an X -tuple of numbers, is p_i ($i = 1, 2, \dots, N$) satisfying:

$$\sum_{i=1}^N p_i = 1$$

Shannon entropy is represented as:

$$H = - \sum_{i=1}^N p_i \log_a p_i \quad (3)$$

with the convention that $p_i \log_a p_i = 0$ as $p_i = 0$ [11,12]. In addition, when $a = 2$, e , or 10 , the unit of H is bit, nat or hart respectively.

From Equations (2) and (3), it is found that there is a strong connection between Shannon entropy and B-G entropy. In some ways, Shannon entropy naturally inherits the statistic property of B-G entropy such as convex linearity, extensivity, continuity, and so on, so Shannon entropy also satisfies the following Equation (4):

$$H(A+B) = H(A) + H(B) \quad (4)$$

where A and B are two subspaces of one information set. In other words, the statistic range of Shannon entropy is restricted to additive signals, which like B-G entropy only fits into extensive systems in thermodynamics [13].

It is well known that Boltzmann-Gibbs statistics (BGS) are inadequate for treating some complex systems [14]. These are systems with complex or long term interactions and correlations, systems showing often distribution laws are different from the usual ones (Gauss, Poisson), systems in chaotic or fractal states and often related to nonextensive phenomena in energy, entropy, heat, and other quantities [15]. In the process of wavelet decomposition, energy leakage exists at each scale, and the leakage of energy invades neighboring scales. At the same time, the extent of interaction, among neighbor scales, will change with a change in the component of signal. Obviously, it belongs to the complex system mentioned above. In other words, Equation (4) is not true when Shannon entropy is used to analyze wavelet decomposition statistically, which means SWEE is not entirely applicable to some transient signals with components change in the time-frequency domain, so it is necessary to discuss this issue in detail and solve it.

3. Research on Issue in SWEE

3.1. Energy Leakage and Wavelet Aliasing

The analyzed signal result using SWEE is undoubtedly connected to the consistence between the wavelet decomposition result and the corresponding component in the signal, so it is a key point for SWEE accuracy that the proper orthogonal wavelet basis be selected to obtain accurate coefficients or reconstructed signals. However, for most mother wavelets, wavelet aliasing still exists more or less at the neighbor scales, which means that the information and energy of signal are incomplete at the decomposition scales, so we can say that wavelet aliasing is the main reason for energy leakage. To better understand energy leakage, the mechanism of wavelet aliasing is researched as follows: wavelet aliasing is when frequency bands, corresponding with wavelet scales, overlap each other after wavelet decomposition. Here, the definition of continuous wavelet transform is given as below. For $\forall x(t) \in L^2(R)$, the continuous wavelet transform function is defined as:

$$W_{Tx}(a, b) = |a|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} x(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \quad (5)$$

where $\psi(t)$ represents the mother wavelet function, $\bar{\psi}(t)$ represents the complex conjugate of $\psi(t)$, a is the scale operator, and b is the shift operator.

From Equation (5), the derived frequency-domain expression of the correlation function is obtained as:

$$W_{Tx}(a,b) = \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{+\infty} X(\omega) \bar{\hat{\psi}}_a(\omega) \exp(j\omega b) d\omega \quad (6)$$

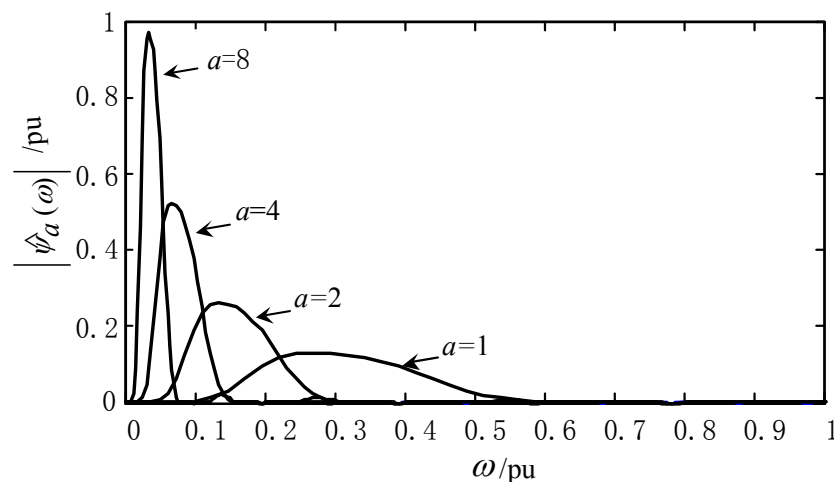
where $\bar{\hat{\psi}}_a(\omega)$ is the frequency-domain expression of $\bar{\psi}(\frac{t}{a})$, and $X(\omega)$ is the frequency-domain expression of $x(t)$. In addition, the center frequency is given as:

$$\omega_a = \frac{\omega_0}{a} \quad (7)$$

where ω_0 is the center frequency of $\hat{\psi}(\omega)$, and ω_a is the center frequency of $\hat{\psi}_a(\omega)$.

As the Db4 wavelet is taken as $\psi(t)$, according to Equation (6), the wavelet function curve, in the frequency domain, is drawn as in Figure 1. From Figure 1 and Equation (7), it is known that ω_a changes with the scale a , and wavelet coefficient $W_{Tx}(a,b)$ reaches its maximum value when the frequency of the signal is equal to ω_a . However, the curve of $\hat{\psi}_a(\omega)$ is not the rectangle, and even if Mallat algorithm is used in discrete wavelet decomposition, this means that the frequency overlap of $\hat{\psi}_a(\omega)$ still exists at the neighbor scales, so the energy of a signal in the frequency-domain window, corresponding to one scale, would leak into the neighbor scales. In conclusion, wavelet aliasing is the main reason of energy leakage, which is almost inevitable for most mother wavelets.

Figure 1. DB4 wavelet function curves with different scales in frequency domain.



3.2. The Negative Effect of Wavelet Aliasing on SWEE

Based on the above analysis, wavelet aliasing can have negative effects on the accuracy of SWEE, therefore we further research the relationship between SWEE and wavelet aliasing.

First, the time sequence of signal $x(t)$ is decomposed into N scales on a orthogonal wavelet, suppose that no wavelet aliasing takes place at neighbor scales, the mathematical expressions of expectation SWEE is derived as:

$$\begin{aligned}
 W_{EE} &= -\sum_{i=1}^N p_i \log_e p_i \\
 &= -\sum_{i=1}^{k-1} p_i \log_e p_i - p_k \log_e p_k - p_{k+1} \log_e p_{k+1} - \sum_{i=k+2}^N p_i \log_e p_i
 \end{aligned} \tag{8}$$

where:

$$\begin{aligned}
 p_i &= \frac{E_i}{E} \\
 E &= \sum_{i=1}^N E_i \quad (i=1, \dots, N)
 \end{aligned}$$

Similarly, suppose that wavelet aliasing takes place at neighbor scales k and $k+1$, and ε represents the energy loss from k to $k+1$, mathematical expressions of SWEE are derived as:

$$\begin{aligned}
 W'_{EE} &= -\sum_{i=1}^N p'_i \log_e p'_i \\
 &= -p'_k \log_e p'_k - p'_{k+1} \log_e p'_{k+1} - \sum_{i=1}^{k-1} p'_i \log_e p'_i - \sum_{i=k+2}^N p'_i \log_e p'_i \\
 &= -p'_k \log_e p'_k - p'_{k+1} \log_e p'_{k+1} - \sum_{i=1}^{k-1} p_i \log_e p_i - \sum_{i=k+2}^N p_i \log_e p_i
 \end{aligned} \tag{9}$$

where:

$$p'_i = \frac{E'_i}{E} \quad (i=1, \dots, N), \quad p'_k = \frac{E_k - \varepsilon}{E} = p_k - p_\varepsilon, \quad p'_{k+1} = \frac{E_{k+1} + \varepsilon}{E} = p_{k+1} + p_\varepsilon.$$

Then, constructing the difference function $\theta(\varepsilon)$ as:

$$\begin{aligned}
 \theta(\varepsilon) &= W'_{EE} - W_{EE} \\
 &= -p'_k \log_e p'_k - p'_{k+1} \log_e p'_{k+1} + p_k \log_e p_k + p_{k+1} \log_e p_{k+1} \\
 &= -(p_k - p_\varepsilon) \log_e (p_k - p_\varepsilon) - (p_{k+1} + p_\varepsilon) \log_e (p_{k+1} + p_\varepsilon) + p_k \log_e p_k + p_{k+1} \log_e p_{k+1} \\
 &= f(p_k - p_\varepsilon) - f(p_k)
 \end{aligned} \tag{10}$$

where:

$$\begin{aligned}
 f(p_k) &= -p_k \log_e p_k - p_{k+1} \log_e p_{k+1} \\
 &= -p_k \log_e p_k - (C - p_k) \log_e (C - p_k) \\
 f(p_k - p_\varepsilon) &= -(p_k - p_\varepsilon) \log_e (p_k - p_\varepsilon) - (p_{k+1} + p_\varepsilon) \log_e (p_{k+1} + p_\varepsilon) \\
 &= -(p_k - p_\varepsilon) \log_e (p_k - p_\varepsilon) - (C - p_k + p_\varepsilon) \log_e (C - p_k + p_\varepsilon)
 \end{aligned}$$

with $C = p_k + p_{k+1}$. Furthermore, for researching the analytic properties of $f(p_k)$, $f(p_k)$ is transformed as:

$$\begin{aligned}
f(p_k) &= -p_k \log_e p_k - p_{k+1} \log_e p_{k+1} \\
&= C \left(-\frac{p_k}{C} (\log_e \frac{p_k}{C} + \log_e C) - \frac{p_{k+1}}{C} (\log_e \frac{p_{k+1}}{C} + \log_e C) \right) \\
&= C \left(-\frac{p_k}{C} \log_e \frac{p_k}{C} - \frac{p_{k+1}}{C} \log_e \frac{p_{k+1}}{C} \right) - (p_k + p_{k+1}) \log_e C \\
&= CH(P) - C \log_e C
\end{aligned} \tag{11}$$

where:

$$H(P) = -\frac{p_k}{C} \log_e \frac{p_k}{C} - \frac{p_{k+1}}{C} \log_e \frac{p_{k+1}}{C}.$$

Obviously, $H(P)$ is the Shannon information entropy of a two-channel information source with the prior probability $(\frac{p_k}{C}, \frac{p_{k+1}}{C})$. According to the concavity of Shannon information entropy function, $H(P)$ meets the following conditions:

$$\begin{cases} H(\alpha P + (1-\alpha)Q) \geq \alpha H(P) + (1-\alpha)H(Q) \\ 0 \leq H(P) \leq \log_e 2 \\ 0 < \alpha < 1 \end{cases} \tag{12}$$

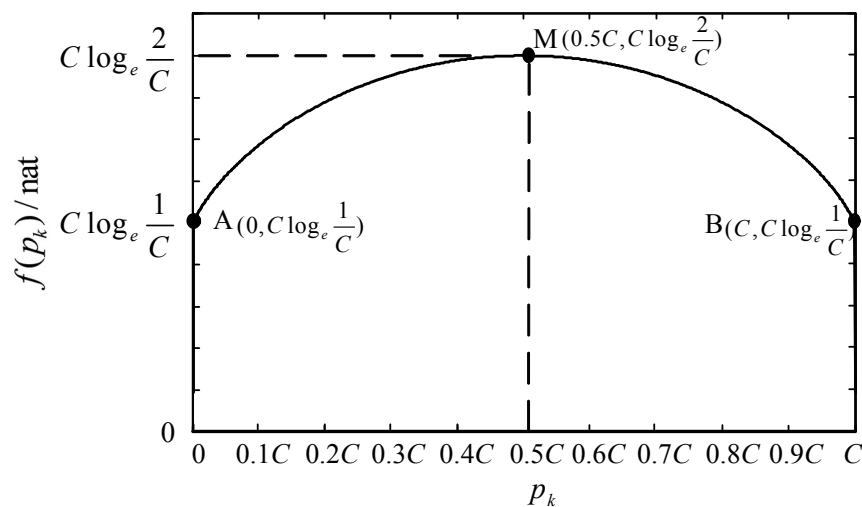
According to Equations (11) and (12), it can be deduced as:

$$\begin{cases} f(\alpha p_k + (1-\alpha)q_k) \geq \alpha f(p_k) + (1-\alpha)f(q_k) \\ C \log_e \frac{1}{C} \leq f(p_k) \leq C \log_e \frac{2}{C} \end{cases}$$

which means that $f(p_k)$ has concavity, maximum and minimum values.

According to the above analysis, the curve of $f(p_k)$ is drawn as shown in Figure 2.

Figure 2. Curve of $f(p_k)$.



Considering that wavelet aliasing and energy leakage exist at the neighbor scales, on the basic premise of $0 < p_\varepsilon < p_k$, $\theta(\varepsilon)$ is discussed as following:

(1) As $0 < p_k \leq p_{k+1}$,

The curve of $f(p_k)$ corresponds to $\widehat{AM}$ in Figure 2. With increasing value of p_ε , it is found that $\theta(\varepsilon) < 0$, $f(p_k)$ is greater than $f(p_k - p_\varepsilon)$, which means that W'_{EE} is less than W_{EE} .

(2) As $0 < p_{k+1} < p_k$,

(a) if $0 < p_\varepsilon < p_k - p_{k+1}$, with increasing value of p_ε , it is found that $\theta(\varepsilon) > 0$, $f(p_k - p_\varepsilon)$ is greater than $f(p_k)$, which means that W'_{EE} is greater than W_{EE} .

(b) if $p_k - p_{k+1} < p_\varepsilon$, with increasing value of p_ε , it is found that $\theta(\varepsilon) < 0$, $f(p_k - p_\varepsilon)$ is less than $f(p_k)$, which means that W'_{EE} is less than W_{EE} .

From the above, it is only when $p_\varepsilon = 0$ or $p_k - p_{k+1}$ that $\theta(\varepsilon) = 0$, which means W'_{EE} is equal to W_{EE} without wavelet aliasing at the neighbor scales. However, according to the analysis in Section 3.1, wave aliasing inevitably exists to varying degrees, so energy leakage caused by wavelet aliasing has detrimental effects on the accuracy of SWEE. Without considering the signal characteristics, using the SWEE algorithm directly, the accuracy of feature extraction is sometimes unbelievable. To prove the actual existence of a negative effect of wavelet aliasing on SWEE, we propose a numerical example as follows. For above all, a signal is constructed as :

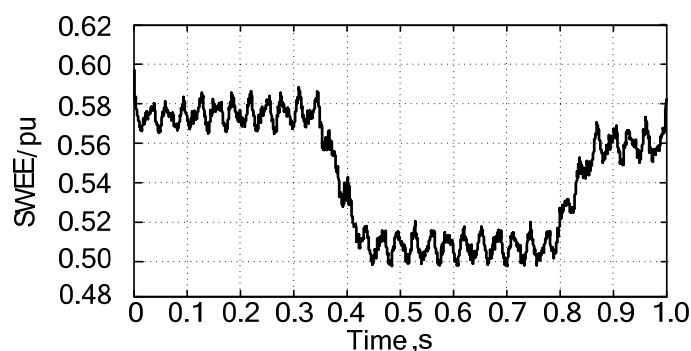
$$x(t) = 0.25 \sin(2\pi f_1 t) + 0.25 \sin(2\pi f_2 t) + 0.4 \sin(2\pi f_3 t) + x_1(t) + x_2(t) + W_n \quad (0 \leq t \leq 1.0 \text{ s})$$

where:

$$x_1(t) = \begin{cases} 0 & (0 \leq t \leq 0.4 \text{ s}) \\ 0.3 \sin(2\pi f_4 t) & (0.4 \leq t \leq 1.0 \text{ s}) \end{cases}, \quad x_2(t) = \begin{cases} 0 & (0 \leq t \leq 0.8 \text{ s}) \\ 0.1 \sin(2\pi f_5 t) & (0.8 \leq t \leq 1.0 \text{ s}) \end{cases}$$

$f_1 = 200 \text{ Hz}$, $f_2 = 390 \text{ Hz}$, $f_3 = 450 \text{ Hz}$, $f_4 = 625 \text{ Hz}$, $f_5 = 1,260 \text{ Hz}$. In addition, the sampling frequency $f_s = 5,000 \text{ Hz}$, and W_n represents white noise. Then, four-level wavelet decomposition on the basis of the Db4 wavelet is used for the signal $x(t)$. When $w = 30$ and $\delta = 1$, according to Equation (1), SWEE of $x(t)$ is obtained and the curve is drawn as in Figure 3.

Figure 3. SWEE under wavelet aliasing.



According to the definition of SWEE, the value of SWEE has to increase with the complexity of the signal in the time-frequency domain. For this reason, with the frequency component increasing, the value of SWEE, from 0 s to 0.4 s, should be minimum from 0 s to 1 s, and the value of 0.4–0.8 s

should be less than that of 0.8–1 s. However, from Figure 3, SWEE never represents the complexity change of the signal accurately, and the ripple disturbance is also serious, which does not correspond with the fact, so it is true that wavelet aliasing has a negative effect on the accuracy of SWEE, and this also verifies the correctness of the above theoretical deduction.

4. Tsallis Wavelet Energy Entropy—The Development of SWEE

4.1. Tsallis Entropy

Nonextensive statistical mechanics, pioneered by Tsallis, offer a consistent theoretical framework for the studies of complex systems with long-range interactions, long-time memories, multifractal and self-similar structures, or anomalous diffusion phenomena. As a nonextensive entropy, Tsallis entropy is the extension and development of extensive entropy (B-G entropy) in statistical physics [16]. It can explain some abnormal experimental phenomena, such as the complexity of non-additive systems, which cannot be explained by the theory of extensive entropy. Tsallis entropy in a discrete expression is defined as follows:

$$S_q = \frac{k}{q-1} \left(1 - \sum_{i=1}^n p(i)^q \right) \quad (q \in R) \quad (13)$$

where k is Boltzmann constant, $\{p(i)\}_{i=1}^n$ is a probability distribution set with i and $\sum_{i=1}^n p(i) = 1$.

Different from extensive entropy, q is introduced as a parameter, so it is called nonextensivity index, and it represents the extent of nonextensivity of the system in Tsallis entropy. The non-additivity of Tsallis entropy of a complex system, composed of A and B subsystems, is defined as follows:

$$\frac{S_q(A+B)}{k} = \frac{S_q(A)}{k} + \frac{S_q(B)}{k} + (1-q) \frac{S_q(A)S_q(B)}{k^2} \quad (14)$$

Note that $q < 1$ and $q > 1$ correspond respectively to the ultra-extensivity and sub-extensivity of the system. For researching the mathematical concave and convex nature of Tsallis entropy, let P and Q be probability variables, and satisfy:

$$P = (p(1), p(2), \dots, p(r)) \text{ and } Q = (q(1), q(2), \dots, q(r)),$$

where:

$$0 \leq p(i) \leq 1, 0 \leq q(i) \leq 1, i = 1, 2, \dots, r, \sum_{i=1}^r p(i) = 1 \text{ and } \sum_{i=1}^r q(i) = 1.$$

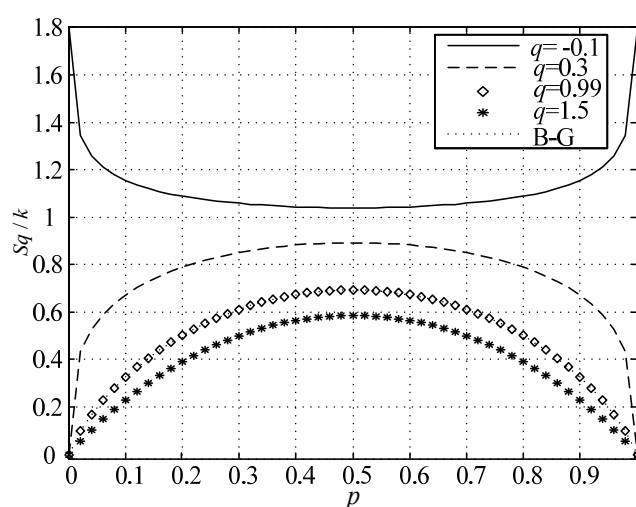
Suppose that $0 \leq \alpha \leq 1$, when S_q satisfies:

$$S_q(\alpha P + (1-\alpha)Q) \geq \alpha S_q(P) + (1-\alpha)S_q(Q) \quad (15)$$

From Equation (15), Tsallis entropy function represents the concave nature. At the same time, Tsallis entropy function has a definite concavity for q values (S_q is concave for $q > 0$ and convex for $q < 0$). Furthermore, we consider Tsallis entropy statistical characteristics as two independent subsystems. According to Equation (13), we have a curve trace like Figure 4 which indicates the change law of Tsallis entropy with the probability distribution under different q values.

From Figure 4, the variation of q values has considerable effects on the statistical characteristics of Tsallis entropy. When $q > 0$ and $q \neq 1$, the function curves take on the appearance of concavity, and it exists corresponding maximum values for all. When $q \rightarrow 1$, Tsallis entropy tends to B-G entropy and it can describe the complexity of additive systems when considering probabilistically independent events [17,18]. Moreover, when $q < 0$, the function curves are contrary to the former for $q > 0$, and there exists the corresponding minimum values. Based on the above analysis, Tsallis entropy with appropriate q value is not only more flexible in information measurement but more widespread in the statistical range of entropy.

Figure 4. Relations between entropy with q and probability distribution under two-level system.



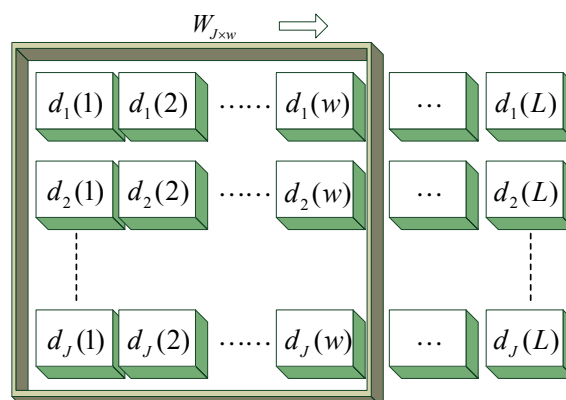
4.2. The Definition of TWEE

A sliding-window (w in width) is defined upon wavelet coefficients $D = \{d_j(k), k = 1, \dots, L\}$ with wavelet decomposition at j scales like Figure 5, here the sliding factor is $\delta \in N$. Then a sliding-window is defined as:

$$W(n; w, \delta) = \{d_j(k), k = 1 + n\delta, \dots, w + n\delta\}$$

where $n = 1, 2, \dots, N$, $N = (L - w) / \delta$.

Figure 5. Wavelet coefficients matrix in the sliding-window.



At the moment of $w/2 + n\delta$:

$$E(n) = \sum_{j=1}^J E_n(j)$$

is defined as the energy sum of high-frequency part of testing signal in $W(n; w, \delta)$ on j scales where:

$$E_n(j) = \sum_{k=1+n\delta}^{w+n\delta} (d_j(k))^2.$$

Suppose that $p_n(j) = E_n(j)/E(n)$ and:

$$\sum_{j=1}^J p_n(j) = 1,$$

then at the moment of $w/2 + n\delta$ TWEE is defined as:

$$W_{EE}(n) = \frac{1}{q-1} \left(1 - \sum_{j=1}^J (p_n(j))^q \right) \quad (16)$$

As the sliding-window slides, the trendline that TWEE changes with time can be obtained and plotted. In the above expression, the scale space corresponds to the frequency space, and TWEE indicates the energy distribution of signals. A wavelet functions does not have pulse selectivity in either the frequency or time domain, whereas it has a support region, so the partition of the original signal at scale space also indicates the signals' energy distribution at time-frequency domains. The more complex the analyzed signal is, more modes the energy congregates to and the larger the TWEE is. Therefore the TWEE is an index to evaluate the signal complexity or uncertainty.

4.3. The Inhibition on Wavelet Aliasing Using TWEE

According to Section 3, wavelet aliasing causes the energy loss at the neighbor scales, and the information set of wavelet coefficients (or reconstructed signals) at different frequency bands, is nonadditive, so SWEE is unsuited for extracting the features of complex signals, especially for transient signals. On the contrary, as the theoretical foundation of TWEE, Tsallis entropy is good at characterizing the uncertainty of nonextensive system because of its nonadditivity. The extent of energy loss has a close correlation with the degree of nonadditivity of information set, for this, if the appropriate value of q is chosen, using TWEE should be able to extract the features accurately without the wavelet aliasing effect.

To prove the above point, the example proposed in Section 3.2 is taken again as an analysis case. For the time sequence of $x(t)$, as $q = 0.1$, the feature of frequency complexity is extracted using TWEE as Figure 6.

Comparing Figure 6 with Figure 3, it is found that the value of TWEE from 0 s to 0.4 s, is less than that from 0.4 s to 0.8 s, and the value of 0.4–0.8 s is less than 0.8–1 s, which is coherent with the frequency variance of $x(t)$. The simulation results show TWEE can inhibit the negative effect of wavelet aliasing and extract the feature of transient signal accurately at the appropriate value of q .

Apart from the statistical characteristics of entropy, we drive the root reason why TWEE are not affected by wavelet aliasing and could characterize the accurate features of transient signals. Taking a

three-level system as the analysis object, according to Equation (13), the statistical results are calculated, and the corresponding relation between Tsallis entropy with q and probability distribution are shown in Figure 7.

Figure 6. TWEE under wavelet aliasing.

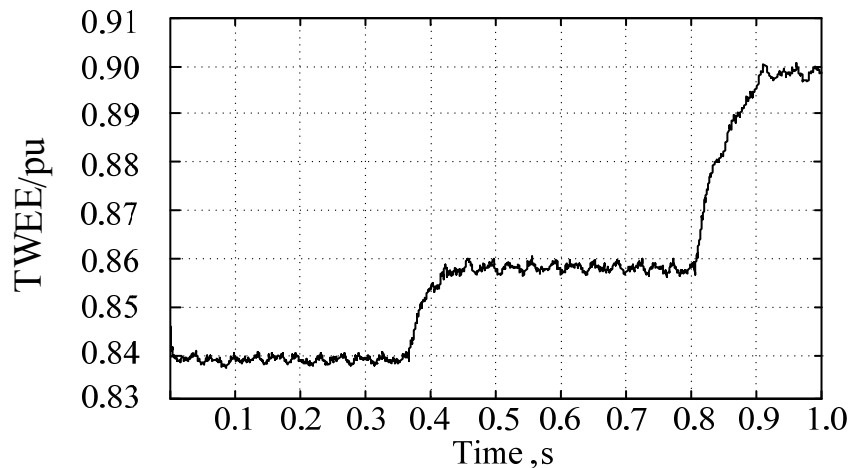
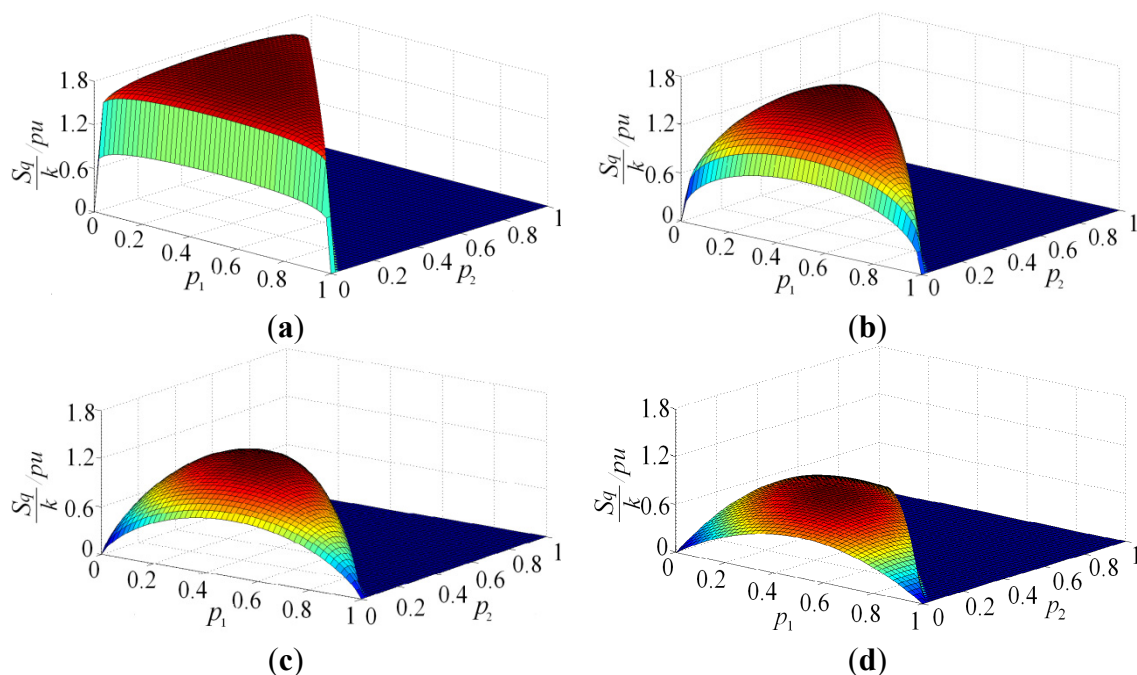


Figure 7. Relation between Tsallis entropy with (a) $q = 0.1$, (b) $q = 0.5$, (c) $q = 0.9$, (d) $q = 2$ and probability distribution under three-level system.



From Figure 7, as $q > 0$ and $q \neq 1$, for the small probability event, Tsallis entropy represents the concave nature, and extends the statistical range, but reduces sensitivity with the increasing value of q . On the contrary, with the decreasing value of q , the statistical range is narrowed down and sensitivity is enhanced. Therefore, when the energy loss at the neighbor scales is too serious to be ignored, it is suggested to reduce the value of q for TWEE, which could enhance the inhibition capability on wavelet aliasing. Of course, with the value of q decreasing, TWEE has less resolution to distinguish the changes of time-frequency complexity for the big probability event in transient signals.

5. TWEE's Application in Signal Analysis in Power System

5.1. Original Transient Power Signal

In the field of power quality, the research on transient signals such as transient overvoltages, voltage dips, voltage interruptions, voltage flicker and voltage pulses, have attracted great attention. As one of transient signals in power systems, the voltage fluctuations caused by indirect lightning strikes often happen in the overhead distribution lines in South China, where most indirect lightning strikes are low-energy and unlikely to cause faults directly. However, the safety of power transmission and the quality of power supply could be suffered from negative effects because of high incidence rate of lightning strikes. In addition to electromagnetic interference in PT secondary circuits, the short duration and low energy level of non-faulty indirect lightning strikes are another major reason which could result in failure in extracting the features of this kind of lightning strike, so the phase voltages, caused by indirect lightning strikes, are taken as analysis objects, and TWEE and SWEE are separately used to extract the transient features of lightning strikes. A set of phase voltage signals of a 110 kV transmission line, with lightning strike interference, are obtained from the Guangzhou EPRI (Figure 8). The details are as follows:

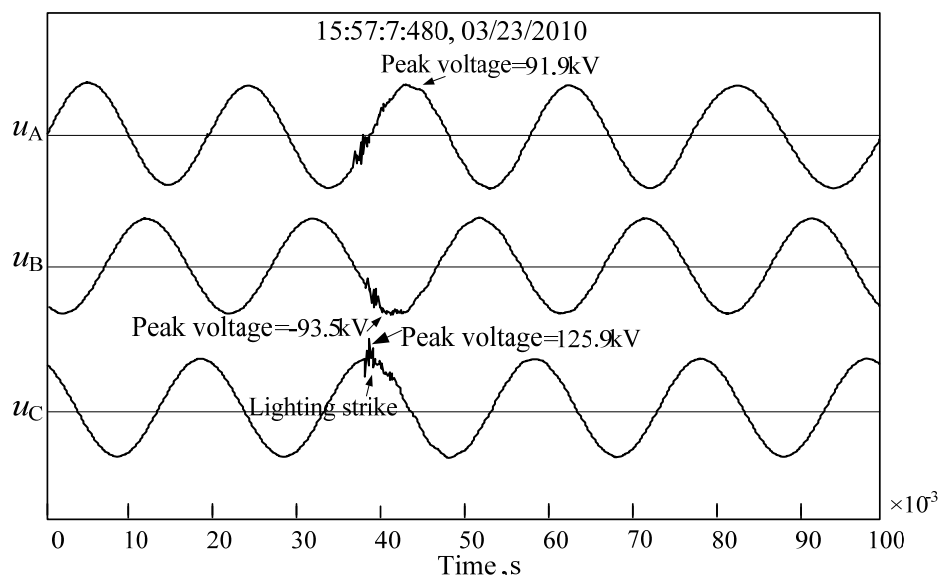
Fault line: the 110 kV Jiayi transmission line in Guangdong.

Fault type: lightning strike interference.

Recording site: the 110 kV side of main transformer of 220 kV Jiahe substation.

Occurrence time: 15:57:7:480, 23 March 2010.

Figure 8. Waveform of three-phase voltages with lightning strike.



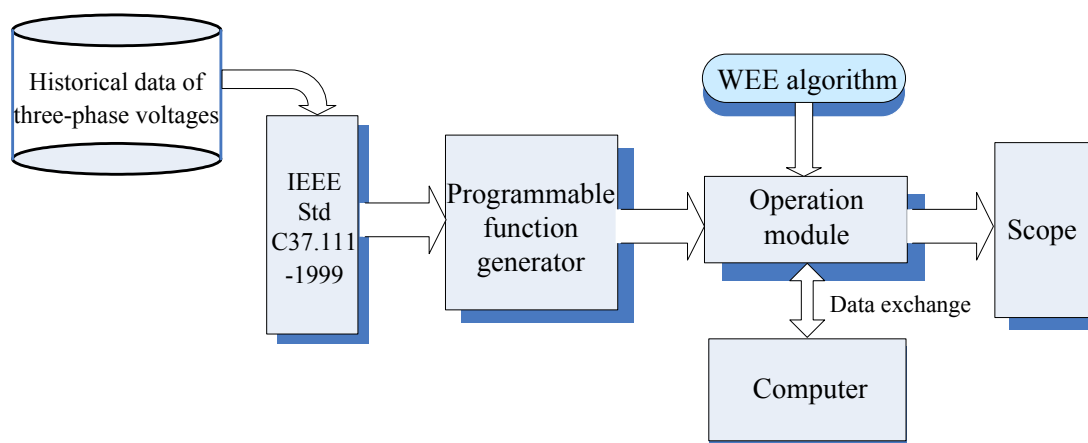
On the afternoon of 3 March 2010, lightning strikes sequentially occurred on the 110 kV Jiayi transmission line of Guangdong Province. The relay protection didn't operate because of the low lightning energy. The recorder located in the 220 kV Jiahe substation recorded the phase voltage waveforms when the lightning struck. Here is a set of collected lightning strike voltage waveforms is shown in Figure 8. In Figure 8, lightning strike interference occurs to the line's three phases in a

relative time of 38 ms. Because the traveling waves reciprocate in the transmission line, over-voltage and high-frequency components exist in the three-phase voltage signals. Meanwhile, with the energy loss in the transmission line, the high-frequency component of the voltage decreases and it almost disappears after about 50 ms. Finally, the three-phase voltage returns to normality.

5.2. The Results of Feature Extraction Using Two WEE Algorithm

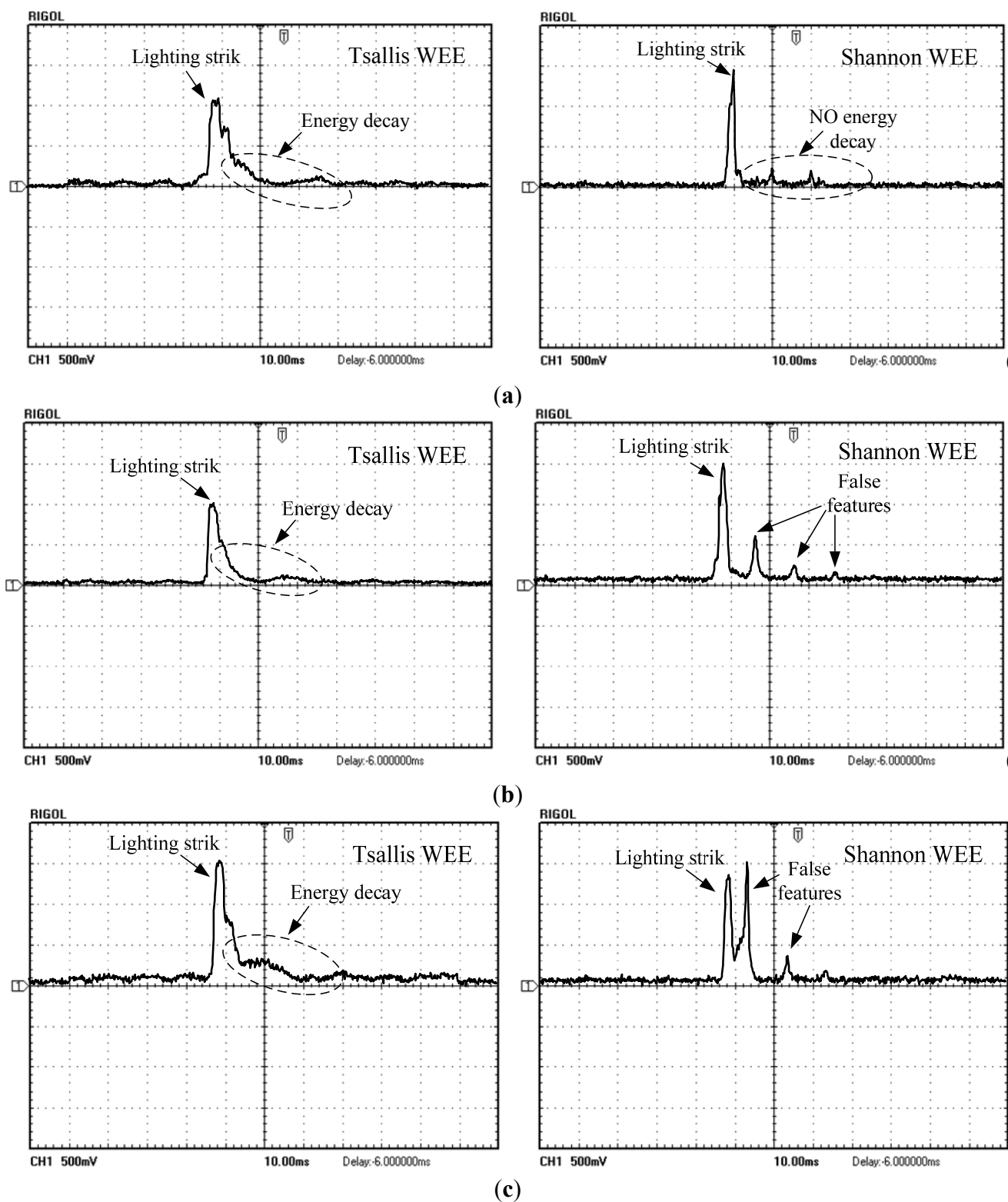
An experimental platform for signal processing is built to verify the practicality and validity of the proposed algorithm. Based on this platform, the historical data of three-phase voltages, which was recorded when a non-faulty lightning strike occurs, is converted to general data according to IEEE Std C37.111-1999. The transient lightning voltage signal is reconstructed by utilizing a programmable function generator, and then the operation module with DSP collects the real-time voltage signals from the programmable function generator through A/D channel. Furthermore, the TWEE algorithm and SWEE algorithm fixed in DSP are utilized separately to extract the transient features of the lightning signal. Finally, the results are output through the D/A channel and shown in the scope. The experiment flow is as shown as Figure 9.

Figure 9. Experiment flow chart.



Set sample frequency of A/D $f_s = 6400\text{Hz}$, width of sliding data window $w = 40$, noextensivity index $q = 0.7$. The results of feature extraction based on TWEE and SWEE are shown in Figure 10. In Figure 10, the feature curve, obtained by TWEE, not only has less fluctuation, but also indicates the occurrence time of the lightning strike and the exponential decay trend of the lightning energy accurately. It was introduced in [19] that lightning energy decays exponentially. Meanwhile, it shows that the peak of TWEE of C-phase is slightly larger than the A-phase and B-phase, which shows that the lightning interference of the C-phase is more serious than the other two phases, which accords with the facts. Although the lightning-strike feature curve, obtained by SWEE, shows the occurrence time of the lightning strike, there are false features in the B-phase and C-phase. The result proves that SWEE has poor anti-interference performance, and its feature curve has a significant fluctuation. Hence, it couldn't indicate the decay trend of the lightning energy accurately.

Figure 10. (a) Lightning-strike feature curves of A phase based on two WEE algorithm; (b) Lightning-strike feature curves of B phase based on two WEE algorithm; (c) Lightning-strike feature curves of C phase based on two WEE algorithm.



6. Conclusions

Entropy statistics theories and wavelet decomposition algorithm are the academic foundation of wavelet entropy, therefore the choice of entropy has a direct impact on the signal analysis results. Because wavelet aliasing exists at neighbor scales, the information set of wavelet coefficients (or reconstructed signals) is endowed with additivity property, and the extent of nonadditivity is much higher for information sets corresponding to transient signals with varying frequency. At this point, using the SWEE algorithm would result in a false analysis result against the fact. In this paper, considering the advantage of Tsallis entropy in measuring the uncertainty of generalized system, TWEE is proposed to extract the features of transient signals with changing the value of q . We apply the TWEE algorithm to the analysis of transient power signals and find that it can not only indicate the occurrence time of lightning strikes, but also characterize the decay trend of lightning energy without the effect of wavelet aliasing, which means the TWEE algorithm has stronger ability to analyze transient signals than SWEE. Of course, do more in-depth research on the wavelet entropy's theoretical basis and application range is needed, especially in the choice of nonextensivity index for the different signals.

Acknowledgments

The financial support received from the National Natural Science Foundation of China (Grant No. 51377016), 2012 Science and Technology Project of State Grid Corporation of China (Grant No.2012515) and the Scientific Research Foundation for the Doctoral Program of Northeast Dianli University (Grant No.BSJXM201335) is gratefully acknowledged.

Author Contributions

Jikai Chen mainly contributed to the Sections 2–4. Guoqing Li contributed to the Section 5.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Quiroga, R.Q; Rosso, O.A; Basar E.; Schürmann M. Wavelet entropy in event-related potentials: A new method shows ordering of EEG oscillations. *Biol. Cybern.* **2001**, *84*, 291–299.
2. AlNashash, H.A.; Thakor, N.V. Monitoring of global cerebral ischemia using wavelet entropy rate of change. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 2119–2122.
3. Cek, M.E.; Ozgoren, M.; Savaci, F.A. Continuous time wavelet entropy of auditory evoked potentials. *Comput. Biol. Med.* **2010**, *40*, 90–96.
4. Brunsell, N.A. A multiscale information theory approach to assess spatial-temporal variability of daily precipitation. *J. Hydrol.* **2010**, *385*, 165–172.
5. Chou, C.M. Wavelet-based multi-scale entropy analysis of complex rainfall time series. *Entropy* **2011**, *13*, 241–253.

6. Sang, Y.-F.; Wang, Z.-G.; Li, Z.-L. Discrete wavelet entropy aided detection of abrupt change: A case study in the Haihe river basin, China. *Entropy* **2012**, *14*, 1274–1284.
7. Zarjam, P.; Epps, J.; Chen, F.; Lovell, N.H. Estimating cognitive workload using wavelet entropy-based features during an arithmetic task. *Comput. Biol. Med.* **2013**, *43*, 2186–2195.
8. Yu, B.; Liu, D.; Zhang, T. Fault diagnosis for micro-Gas turbine engine sensors via wavelet entropy. *Sensors* **2011**, *11*, 9928–9941.
9. He, Z.Y.; Cai, Y.M.; Qian, Q.Q. A study of wavelet entropy theory and its application in electric power system fault detection. *Proc. Chin. Soc. Electr. Eng.* **2005**, *25*, 38–43.
10. He, Z.Y.; Chen, X.Q.; Luo, G.M. Wavelet entropy measure definition and its application for transmission line fault detection and identification. In Proceedings of the International Conference on Power System Technology, Chongqing, China, 22–26 October 2006.
11. Hnizdo, V.; Gilson, M.K. Thermodynamic and differential entropy under a change of variables. *Entropy* **2010**, *12*, 578–590.
12. Baez, J.C.; Fritz, T.; Leinster, T. A characterization of entropy in terms of information loss. *Entropy* **2011**, *13*, 1945–1957.
13. Wang, A. Nonextensive Statistics and Incomplete Information. *Eur. Phys. J. B* **2002**, *26*, 357–368.
14. Tsallis, C. Possible Generalization of Boltzmann-Gibbs Statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.
15. Huang, Z.F.; Ou, C.J.; Chen, J.C. Incomplete nonextensive statistics and the zeroth law of thermodynamics. *Chin. Phys. B* **2013**, *22*, 040501:1–040501:4.
16. Tsallis, C.; Mendes, R.S.; Plastino, A.R. The role of constraints within generalized nonextensive statistics. *Phys. A* **1998**, *261*, 534–554.
17. Martins, A.; Aguiar, P.; Figueiredo, M. Tsallis Kernels on Measures. In Proceedings of the IEEE Information Theory Workshop, Porto, Portugal, 5–9 May 2008.
18. Sneddon, R. The Tsallis entropy of natural information. *Phys. A* **2007**, *386*, 101–118.
19. Heidler, F.; Cvetic, J.M.; Stanic B.V. Calculation of lightning current parameters. *IEEE Trans. Power Deliv.* **1999**, *14*, 399–404.