CrossMark

Wavelet-based data compression for wide-area measurement data of oscillations

Lin CHENG[1], Xinchi JI[1], Fang ZHANG[1], He HUANG[2],
Song GAO[2]

MPCE

Abstract This paper proposes a wavelet-based compression method to compress the recorded data of oscillations in power systems for wide-area measurement systems. Actual recorded oscillations and simulated oscillations are compressed and reconstructed by the wavelet-based data compression method to select the best wavelet functions and decomposition scales according to the criterion of the minimum compression distortion composite index, for a balanced consideration of compression performance and reconstruction accuracy. Based on the selections, the relationship between the oscillation frequency and the corresponding optimal wavelet and scale is discussed, and a piecewise linear model of the base-2 logarithm of the frequency and the order of the wavelet is developed, in which different pieces represent different scales. As a result, the wavelet function and decomposition scale can be selected according to the oscillation frequency. Compared with the wavelet-based data compression method with a fixed wavelet scale for disturbance signals and the real-time data compression method based on exception compression and swing door trending for oscillations, the proposed method can provide high compression ratios and low distortion rates.

Keywords Data compression, Oscillation, Wavelet transform, Wide-area measurement system (WAMS)

✉ Fang ZHANG
  thu.zhangfang@gmail.com

  Lin CHENG
  chenglin@mail.tsinghua.edu.cn

  Xinchi JI
  jxc16@mails.tsinghua.edu.cn

  He HUANG
  huanghe.jiangsu@gmail.com

  Song GAO
  gaosongjiangsu@gmail.com

[1] State Key Laboratory of Control and Simulation of Power System and Generation Equipment, Tsinghua University, Beijing, China

[2] State Grid Jiangsu Electric Power Company, Nanjing, China

1 Introduction

Wide-area measurement systems (WAMSs) have been frequently used in power systems. The measured synchrophasor data can be used for both real-time applications such as dynamic monitoring, oscillation source location, smart control, and wide-area protection [1–5], and off-line applications such as post-event analysis [6]. With the rapid development of interconnected power systems, renewable energy sources, and high-voltage direct current (HVDC) transmission systems, oscillation problems such as low-frequency oscillations (LFOs) and subsynchronous oscillations (SSOs) are drawing more and more attention [7–10], which may threaten power system security and stability. WAMS can be used for solving these problems. However, large amounts of data in WAMS caused by high reporting rates and numerous data channels create a significant burden on communication and storage systems, which may cause data congestion in communication and a shortage of storage capacity.

For example, 34 phasor measurement units (PMUs) are implemented for WAMS in the Guizhou Power Grid in Southeast China. More than 3500 data channels are

STATE GRID
STATE GRID ELECTRIC POWER RESEARCH INSTITUTE

applied, and the reporting rate is 100 Hz. Over 120 GB of measurement data is transmitted to the phasor data concentrator (PDC) per day [11]. Moreover, the overall data size is expected to be larger as a result of more measurement devices and higher reporting rates in the future. Therefore, data compression techniques are desirable to mitigate this burden for communication and storage systems.

Various kinds of lossy and lossless compression methods have been developed for data compression in power systems. A type of lossy compression method achieves data compression by approximating the signal as segments, such as the exception compression (EC) and swing door trending (SDT) methods in [11]. These techniques are suitable for real-time applications, but the compression performance and reconstruction accuracy are heavily influenced by the parameters, which are set based on experience. The accuracy is not very high when a high compression ratio is achieved [11]. Another type of lossy compression method involves signal transformation algorithms based on frequency-domain analysis or time-frequency analysis, among which the most widely used are the wavelet transform (WT)-based data compression methods. The algorithms can achieve both high compression ratios and high accuracy, but they require data buffering and are thus not suitable for real-time applications. Rather than processing the signal itself, principal component analysis (PCA) can be used for data compression by reducing the redundancy between multidimensional data such as the measurement data of multiple PMUs [12] and multiple data channels [13]. The compression performance and reconstruction accuracy depend on the number of principal components selected. The compression ratio of PCA is not very high when high accuracy is achieved [12]. The PCA algorithm can be combined with other algorithms to achieve higher compression ratios. Lossless compression methods such as the Lempel-Ziv-Markov chain algorithm (LZMA) [14] and Golomb-Rice coding [15] are usually combined with other techniques because they cannot achieve high compression ratios by themselves. The above is summarized in Table 1.

This paper focuses on data compression in storage for off-line applications, which require high data accuracy but

Table 1 Comparison of different types of data compression methods

Performance	Lossy methods			Lossless methods
	Segments	WT-based	PCA	
CR	High	High	Low	Low
Accuracy	Low	High	High	Very high
Real-time	High	Low		

not much real-time capability [6]. Therefore, the widely used WT-based data compression technique is selected in this paper. In [16, 17], a fundamental multiresolution analysis (MRA) algorithm was utilized for data compression in power systems. The signals were decomposed into scaling coefficients (SCs) and layers of wavelet coefficients (WCs), from which the insignificant data points could be deleted through threshold methods. Many recent studies focused on selecting the best wavelet function and decomposition scale based on various criteria and improving the fundamental WT-based algorithms. In [18], based on the criterion of maximum wavelet energy, the db2 wavelet and scale 5 were chosen for the compression of disturbance signals generated by simulations. The fixed wavelet and scale may not be suitable for other signals. In [19, 20], a wavelet packet decomposition (WPD)-based data compression method, which is an expansion of wavelet decomposition (WD), was proposed for better accuracy. The best wavelet and scale were selected based on the maximum wavelet energy as well. However, the decomposition of high-frequency components in WPD may be meaningless for the compression of LFOs. In [21], an embedded zerotree wavelet transform (EZWT)-based data compression technique was proposed where the lossless encoding method was combined with the lossy WT-based compression method. Similarly, in [12], an efficient data compression method was introduced where several lossy and lossless compression methods such as PCA, WT, and LZMA were combined. In addition, many other WT-based data compression methods were developed in earlier years, including the lifting scheme [22], which is a fast algorithm of WD; the slantlet transform [23], with different wavelets for different scales; and the minimum description length (MDL) method [24], in which the abrupt change data points were extracted to achieve data compression. However, there are two main problems in the above WT-based methods: 1) it is difficult to select the optimal wavelet functions and decomposition scales with a balanced compression performance and reconstruction accuracy; 2) most of the above methods are only suitable for the compression of disturbance signals. For example, the MDL method is used for extracting abrupt changes from disturbance signals [24], and the criterion of maximum wavelet energy [18–20] means retaining the information of abrupt changes as much as possible. These approaches are not suitable for oscillations, such as LFOs and SSOs.

Our previous work [25] proposed a wavelet-based universal data compression method to solve these problems. The criterion of the minimum compression distortion composite index (CDCI) was proposed to select the best wavelet and decomposition scale. CDCI was proven to be suitable for the compression of different types of signals including oscillations and disturbance signals, and the

selected wavelet and scale can provide balanced compression performance and reconstruction accuracy. However, in [25], signals have to be compressed and reconstructed by all candidates of wavelets and scales. Only then can the CDCIs of all the candidates be calculated to select the best one. The entire algorithm can result in a significant amount of calculation.

Therefore, an improved wavelet-based data compression technique for oscillations in power systems is proposed in this paper based on [25]. The wavelet function and decomposition scale can be selected directly according to the oscillation frequency, which is the most significant characteristic of oscillations. The amount of calculation in the proposed method is much lower than that in [25]. Specifically,

1) The qualitative relationship between the oscillation frequency and the corresponding optimal wavelet function and decomposition scale is discussed. As samples, actual recorded oscillations and simulated oscillations are compressed and reconstructed by the wavelet-based data compression method to select the best wavelets and scales based on the criterion of the minimum CDCI in [25].

2) A quantitative relationship is established by linear regression. A piecewise linear model of the base-2 logarithm of the frequency ($\log_2 f$) and the order of the wavelet (N) is developed in which different pieces represent different scales.

3) To evaluate the performance of the proposed method, a comparison is made with the wavelet-based data compression method with a fixed wavelet and scale for disturbance signals in [18] and the EC and SDT data compression (ESDC) method in [11].

This paper is organized as follows. Section 2 introduces the wavelet-based universal data compression method and the criterion of CDCI for selecting the optimal wavelet function and decomposition scale. Section 3 presents the relationship between the oscillation frequency and the corresponding optimal wavelet and scale, and a piecewise linear model is developed. Section 4 carries out a comparison of other data compression methods. Section 5 presents the conclusion.

2 Wavelet-based universal data compression

2.1 Wavelet-based data compression

The procedure for the wavelet-based MRA is shown in Fig. 1. Through the low-pass filters g_i and high-pass filters h_i, a time series can be decomposed into SCs and WCs, corresponding to the approximations a_i and details d_i of the

(a) Decomposition
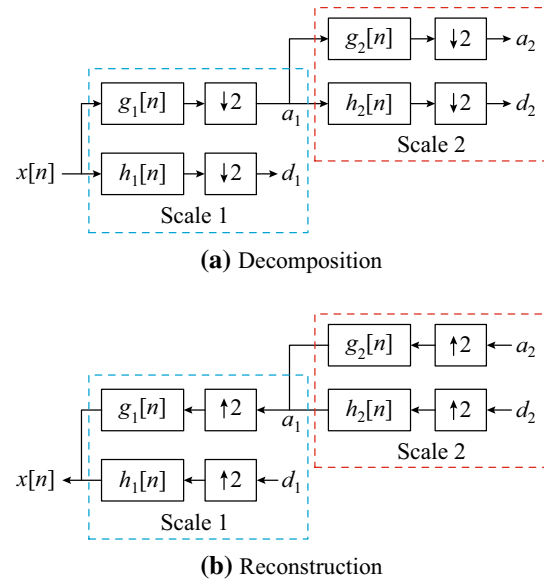

(b) Reconstruction

Fig. 1 Procedure of MRA

original signal, respectively. The filters, which are finite impulse response (FIR) filters, are constructed by a scaling function and a wavelet function that depend on the choice of wavelet function. MRA means decomposing the generated SC layer by layer to get SCs and WCs of different scales. SCs and WCs represent the low-frequency and high-frequency components of the signal, respectively. The reconstruction is the inverse procedure of the decomposition.

After the first decomposition, the lengths of the SCs and WCs will be $n_0 + K - 1$, in which n_0 is the number of original sampling points and K is the length of the filters. The total number will be approximately twice as many as before. Other layers are similar. Therefore, two-time down-sampling is necessary to avoid information redundancy in decomposition, as indicated by the "↓ 2" in Fig. 1a. Similarly, two-time up-sampling is obligatory for reconstruction, as indicated by the "↑ 2" in Fig. 1b. Therefore, assuming that the highest scale is I, the number of a_i and d_i at any scale i can be written as:

$$n_i = \frac{n_{i-1} + K - 1}{2} \approx \frac{n_{i-1}}{2} \quad i = 1, 2, \ldots, I \quad (1)$$

The general form is:

$$n_i \approx 2^{-i} n_0 \quad i = 0, 1, \ldots, I \quad (2)$$

The total number of sampling points before and after decomposition is:

$$n \approx 2^{-I} n_0 + \sum_{i=1}^{I} 2^{-i} n_0 = n_0 \quad (3)$$

As introduced above, SCs represent approximations of

the original signal, which is very important. By contrast, WCs represent the details in which the high-value data points experience abrupt changes and the low-value data points are mainly caused by noise. Therefore, a threshold can be applied to WCs to achieve data compression and retain important information of the signal. Many threshold methods have been developed, of which a selection deserves further study in the future. However, no matter what threshold method is chosen, the problem of selecting the best wavelet functions and decomposition scales with balanced compression performance and reconstruction accuracy always exists. Therefore, a common and widely used fixed-threshold and soft-thresholding method [26] is adopted in this paper. The method is simple but efficient, and the soft thresholding can avoid discontinuities in WCs, making the reconstructed signal smoother. The threshold λ_i and WCs after thresholding $\hat{d}_i$ at scale i can be calculated as:

$$\lambda_i = \sqrt{2\ln(n_i)} \quad i = 1, 2, \ldots, I \tag{4}$$

$$\hat{d}_i = \begin{cases} \operatorname{sign}(d_i)(|d_i| - \lambda_i) & |d_i| \geq \lambda_i \\ 0 & |d_i| < \lambda_i \end{cases} \tag{5}$$

The main computation of the wavelet-based data compression method is multiplication in convolutions with the wavelet filters. In the decomposition of scale i, the number of multiplications is about $2K(n_{i-1} + K - 1)$, so the total number of multiplications in the decomposition of MRA can be calculated as:

$$\begin{aligned} C(n_0, K, I) &= 2K \sum_{i=1}^{I} (n_{i-1} + K - 1) \\ &\approx 4Kn_0(1 - 2^{-I}) + 2KI(K - 1) \\ &\approx 4Kn_0(1 - 2^{-I}) \end{aligned} \tag{6}$$

Similarly, the number of multiplications in the reconstruction process is the same.

2.2 Selection of wavelet and decomposition scale

In our previous work [25], the optimal wavelet function and decomposition scale were selected based on the criterion of minimum CDCI for a balanced consideration of compression performance and reconstruction accuracy. The following is an introduction to CDCI.

The compression ratio can be used to evaluate compression performance. Assuming that each nonzero data point is of equal size and that zero points can be ignored, the compression ratio λ_{CR} can be calculated as:

$$\lambda_{CR} = \frac{\operatorname{len}(a_I) + \operatorname{len}\left(\sum_{i=1}^{I} \hat{d}_i\right)}{\operatorname{len}(x)} \tag{7}$$

where x is the original signal; function $\operatorname{len}(\cdot)$ means the size of data points.

The distortion rate can be used to evaluate reconstruction accuracy, which can be calculated as a root normalized mean square error. Assuming that the reconstructed signal is x'_n and L is the length of the signal, the distortion rate λ_{DR} can be expressed as:

$$\lambda_{DR} = \frac{||x'_n - x_n||}{||x_n||} = \frac{\sqrt{\sum_{n=1}^{L} (x'_n - x_n)^2}}{\sqrt{\sum_{n=1}^{L} x_n^2}} \tag{8}$$

Based on the compression ratio and the distortion rate, the CDCI is constructed as (9) to achieve a compromise between compression performance and reconstruction accuracy.

$$\begin{cases} \xi_{CDCI} = a\dfrac{1}{\lambda_{CR}} + b\lambda^*_{DR} = \dfrac{a}{\lambda_{CR}} + b\dfrac{\lambda_{DR}}{2 \times 10^{-3}} \\ a + b = 1 \end{cases} \tag{9}$$

where a and b are the weights of the compression performance and reconstruction accuracy, respectively. Owing to the significant difference in magnitude, both $1/\lambda_{CR}$ and λ_{DR} should be normalized. $1/\lambda_{CR}$ can be seen as a normalized value owing to its value range of (0, 1). The distortion rate should be normalized, and the base value is set as 2×10^{-3} in this paper according to the standard [27] in China.

The values of a and b should vary with the compression ratio and the distortion rate. If both the compression ratio and the distortion rate are small, the value of a should be higher because it is more important to increase the compression ratio than to decrease the distortion rate. Conversely, the value of b should be higher. For simplicity, assuming that the weights vary linearly with the distortion rate, the CDCI can be calculated as (10) in [25]. Obviously, a smaller CDCI means a better and more balanced performance.

$$\xi_{CDCI} = \frac{1 - \lambda^*_{DR}}{\lambda_{CR}} + (\lambda^*_{DR})^2 \tag{10}$$

According to [25], the wavelets of db2 to db10 and sym2 to sym10 can be chosen as the candidates of the wavelet functions for their properties of orthogonality, compact support, and good regularity. Since the filter length of wavelets db$N$ and sym$N$ is $2N$, the highest scale I can be calculated as (11) as demonstrated in [28] to guarantee the filter length is smaller than the length of WCs.

$$I = \text{floor}\left(\log_2 \frac{n_0}{2N-1}\right) \tag{11}$$

where function floor($\cdot$) stands for rounding down.

In [25], different types of actual recorded signals were compressed and reconstructed by the candidates of wavelets and scales, among which those corresponding to the smallest CDCI were selected as the best wavelets and scales. The results demonstrated that CDCI can be utilized to choose the best wavelet and scale for the compression of different types of signals, including oscillations and disturbances.

The main computation of the method in [25] consists of multiplications in the decomposition and reconstruction process, and multiplications in the calculation of the distortion rate. Since the candidates for the wavelet functions are db2 to db10 and sym2 to sym10, and the decomposition process can be reused when choosing the same wavelet and different scales, the total computation can be calculated as:

$$\begin{aligned} C_{\text{previous}} &\approx 2 \sum_{N=2}^{10} \left[C(n_0, 2N, I) \right. \\ &\quad \left. + \sum_{i=1}^{I} C(n_0, 2N, i) + 2In_0 \right] \tag{12} \\ &= 2 \sum_{N=2}^{10} (8N+2)n_0 I = 900 n_0 I \end{aligned}$$

Since the highest scale is 5 in [25], the amount of computation can be calculated as $4500n_0$.

3 Relationship between oscillation frequency and optimal wavelet and scale

As mentioned above, a signal can be decomposed into low-frequency components and layers of high-frequency components. Therefore, it can be speculated that the selection of the optimal wavelet function and decomposition scale depends on the oscillation frequency. In this section, oscillations of different frequencies are compressed and reconstructed by the wavelet-based method to select the best wavelets and scales based on the CDCI. According to the results, a piecewise linear model of the base-2 logarithm of the frequency ($\log_2 f$) and the order of the wavelet (N) is developed in which different pieces represent different scales. The decomposition scale and wavelet function can be selected based on the model.

The window length of data to be processed at a time deserves careful consideration. According to [25], a window length of 10 s is chosen for the compression of LFOs and SSOs to compromise between compression ratios and distortion rates. The sampling rate of the signals is 100 Hz,

so the number of data points to be processed at one time is 1000. According to (11), since the maximum order N is 10, as shown in Sect. 2.2, the highest scale can be calculated as 5.

3.1 Compression of actual recorded LFO signal

An actual recorded LFO signal of about 0.9 Hz, as shown in Fig. 2, is compressed and reconstructed by the wavelet-based method to select the best wavelet function and decomposition scale based on CDCI. The waveform is long enough to show the entire process of the LFO.

Two hundred pieces of waveforms are intercepted from the LFO signal randomly, and the length of each piece is 10 s. The pieces of waveforms are compressed and reconstructed by the candidates of wavelets (db2 to db10 and sym2 to sym10) and scales (1 to 5). The compression ratios and distortion rates of different wavelets and scales are then calculated to get the corresponding CDCIs. According to the criterion of the minimum CDCI, the best wavelet and scale of each waveform are selected. The results are listed in Table 2.

As shown in Table 2, the choices for different waveforms are close to each other. Specifically, the best decomposition scale is definitely 4, and the optimal N is about 5. It can be concluded that apart from the oscillation frequency, other characteristics such as the oscillation amplitude have a small effect on the selection.

To give a reasonable explanation of the above results, the compression and reconstruction of the waveform between 0 and 10 s are analyzed as representative of the LFO signal. The compression ratios and distortion rates for each scale and wavelet are shown in Fig. 3. The compression ratio limit of each scale can be calculated as $\lambda_{\text{CR,lim}}(i) = 2^i$ because all WCs are set to zero.

According to Fig. 3, some rules are summarized as follows:

1) The performance of db$N$ and sym$N$ are similar to each other on the same order and scale. The reason is that
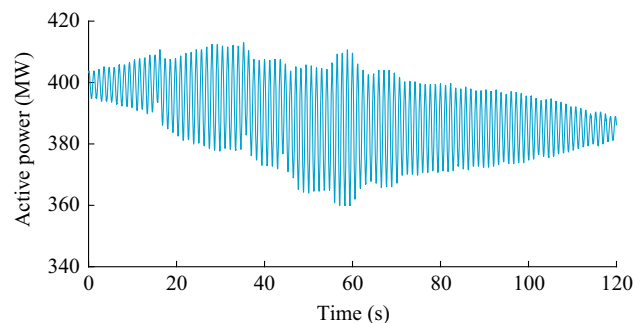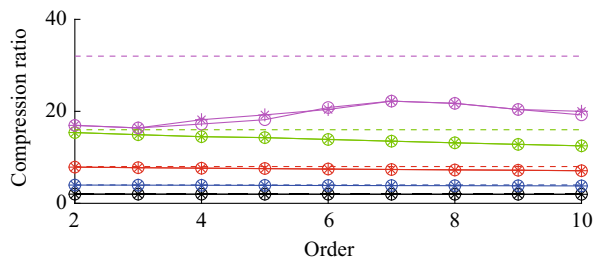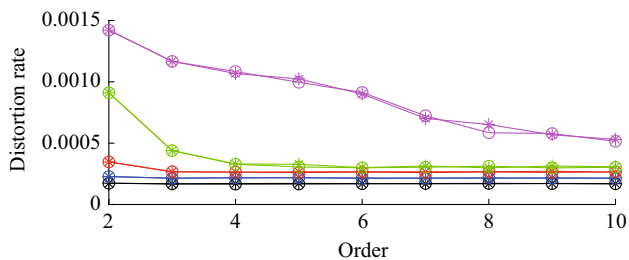

Fig. 2 Actual recorded LFO signal

Table 2 Number of each wavelet function and decomposition scale being selected in compression of LFO

Wavelet function	Decomposition scale	Number
db4	Scale 4	2
db5	Scale 4	61
db6	Scale 4	55
db7	Scale 4	3
sym5	Scale 4	52
sym6	Scale 4	21
sym7	Scale 4	6
Total		200

(a) Compression ratios

(b) Distortion rates

Fig. 3 Compression ratios and distortion rates of LFO between 0 and 10 s

sym$N$ is constructed based on db$N$ in the wavelet theory.

2) The selection of a decomposition scale is the more critical factor in determining the compression performance and reconstruction accuracy. On a scale of 1 to 4, the compression ratios increase by a factor of 2 per level and can almost reach the limits. The stable values of the distortion rates are quite low. On the contrary, the stable value of distortion rates at scale 5 is much higher, but the compression ratios are not much higher than those of scale 4. The result is consistent with the choice made by the CDCI. The explanation is as follows. On a scale of 1 to 4, WCs are considered to consist only of trivial information, which is eliminated by thresholding. By contrast, the valuable information of the oscillation is kept in SCs without being destroyed. However, at scale 5, the valuable information is mixed in WCs and then destroyed by thresholding, resulting in high distortion. Moreover, the amplitude of the WCs caused by the oscillation is too large to be zero. As a result, the distortion rates are quite low, and the compression ratios can reach the limits on a scale of 1 to 4, whereas the distortion rates are high and the compression ratios cannot reach the limits at scale 5.

3) The order of the wavelet has very little effect on the compression performance and reconstruction accuracy in most cases. With N increasing, the compression ratios have a tendency to decrease. The distortion rates decrease at first and then reach stable values. At scale 4, the distortion rate reaches the stable value at about order 5. Since the compression ratio at order 5 is slightly higher than that of order 6 and higher, the best order of the wavelet can be chosen as 5. However, the performance difference between order 6 and order 5 is small. The compression ratio at order 6 is lower than that at order 5, but the distortion rate at order 5 is higher than that at order 6. Therefore, the best order can be chosen as 6 as well. This result is consistent with the choice made by the CDCI. The explanation is as follows. As N increases, the amplitude-frequency response of the wavelet filter becomes more similar to a rectangular shape, making the low-frequency information more difficult to leak out to the WCs. Moreover, the regularity of the wavelet filter becomes better as N increases, bringing about smoother reconstruction waveforms. Therefore, the distortion rates decrease at first and then reach stable values with N increasing. As N increases, the number of sampling points after decomposition increases slightly according to (1), resulting in a slight decrease in the compression ratios.

3.2 Compression of actual recorded SSO signal

An actual recorded SSO signal of about 7 Hz is compressed and reconstructed by the wavelet-based method to select the best wavelet function and decomposition scale based on the CDCI, as shown in Fig. 4.

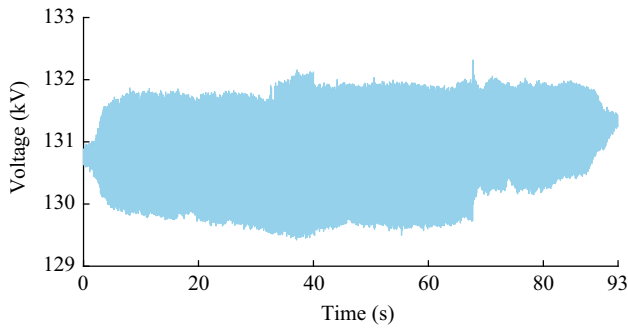Similar to the procedure in Sect. 3.1, 200 pieces of waveforms of 10 s are intercepted from the SSO signal

Fig. 4 Actual recorded SSO signal

randomly. The pieces of waveforms are compressed and reconstructed by the candidates of wavelets and scales, from which the best are selected by calculating the CDCIs. The results are shown in Table 3.

As shown in Table 3, the best decomposition scale is definitely 2 and the optimal N is about 9, which is different from the choice for the compression of LFO. Therefore, it can be concluded that the oscillation frequency has a significant effect on the selection.

The compression and reconstruction of waveforms between 50 and 60 s are analyzed, similar to Sect. 3.1. The compression ratios and distortion rates at each scale and wavelet are shown in Fig. 5.

The rules indicated in Fig. 5 are similar to those in Fig. 3. ① There is little difference between the performance of db$N$ and sym$N$. ② The selection of a decomposition scale is more important than that of the wavelet function. On a scale of 3 to 5, the distortion rates are too high to accept. By contrast, at scale 2, the distortion rates are low, and the compression ratios are high enough. This result is consistent with the choice made by the CDCI. ③ The order of the wavelet has very little effect on the performance if the distortion rate reaches the stable value. At scale 2, the distortion rate reaches the stable value at about order 8 or higher, whereas the performance difference

Table 3 Number of each decomposition scale and wavelet function being selected in compression of SSO

Wavelet function	Decomposition scale	Number
db7	Scale 2	1
db8	Scale 2	11
db9	Scale 2	87
db10	Scale 2	54
sym7	Scale 2	8
sym9	Scale 2	7
sym10	Scale 2	32
Total		200
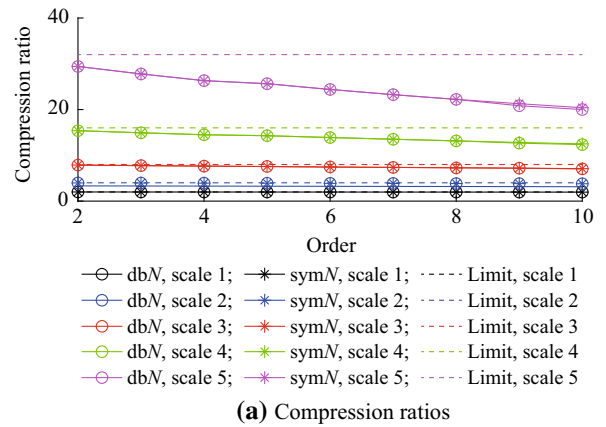
(a) Compression ratios
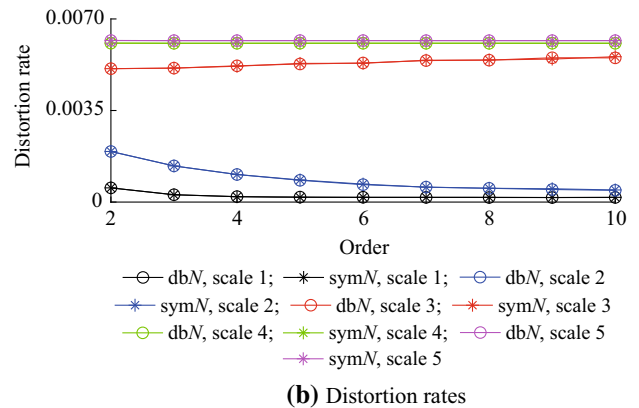
(b) Distortion rates

Fig. 5 Compression ratios and distortion rates of SSO between 50 and 60 s

between order 7 to order 10 is small. The result is consistent with the choice made by the CDCI.

3.3 Compression of simulated oscillations

As mentioned above, the oscillation frequency has a significant effect on the selection of the wavelet and decomposition scale, whereas other characteristics such as the oscillation amplitude have only a slight effect. In addition, the base value of the oscillations may affect the selection. In general, the base value is approximately constant in a time window. Therefore, actual recorded oscillations can be simplified as sinusoidal signals of different frequencies and amplitudes superimposed on different DC biases, as shown in Fig. 6. Furthermore, since
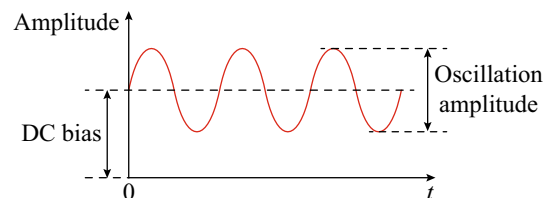

Fig. 6 Simulated oscillation

the wavelet filters are linear filters, and all WCs can be set to zero by thresholding in general when making the optimal selection, the compression process is approximately a linear transform. Therefore, it can be roughly proved that the optimal selection is the same if the ratio of the oscillation amplitude to the DC bias is constant. This proposition can be verified by numerical experiments as well. According to the proposition, the DC biases of the simulated signals can be set to a fixed value, and different oscillation amplitudes represent different signals.

As before, the sampling rate of the signals is 100 Hz, and the length of the signals is 10 s. Owing to the little difference between the performance of db$N$ and sym$N$, only the wavelets of db2 to db10 are chosen as candidates to simplify the algorithm. The candidates of the decomposition scales are on a scale of 1 to 5.

The frequency of the simulated oscillations is determined as follows. The frequency range of LFO is from 0.1 to 2.5 Hz in general [29], and the frequency of SSO is considered to be between the frequency of LFO and the power frequency. In this paper, the sampling rate of signals is 100 Hz. Therefore, the highest frequency of the signals is 50 Hz according to the sampling theorem. To prevent the valuable low-frequency information from being leaked into the WCs, the highest frequency of the signals to be compressed is 25 Hz. As a result, the frequency range of the oscillations to be compressed is 0.1 to 25 Hz. Since the upper frequency limit of SCs should be halved after each decomposition, the base-2 logarithm of the frequency ($\log_2 f$) should be approximately evenly distributed. The frequencies of the simulated oscillations are listed in Table 4.

According to the proposition above, the DC bias is set to a fixed value of 400 in this paper, and the amplitudes are set between 10 and 100 in steps of 10.

Similar to the procedures in Sects. 3.1 and 3.2, the simulated oscillations are compressed and reconstructed by the candidates of wavelets and decomposition scales, from which the best are selected by calculating the CDCIs. Part of the results are listed in Table 5.

Table 4 Frequencies of the oscillations to be compressed

Frequency range (Hz)	Step size (Hz)
[0.1, 0.25]	0.01
[0.26, 0.48]	0.02
[0.5, 0.95]	0.05
[1, 1.9]	0.10
[2, 4.8]	0.20
[5, 7.5]	0.50
[8, 24]	1.00

Table 5 Selected decomposition scales and wavelets for part of the simulated oscillations of different frequencies and amplitudes 10 and 100

Frequency (Hz)	10		100	
	Scale	Wavelet	Scale	Wavelet
0.1	5	db3	5	db3
0.2	5	db3	4	db5
0.5	5	db6	4	db6
1	4	db6	3	db6
1.5	4	db9	3	db6
2	3	db6	3	db9
3	3	db8	2	db6
5	2	db6	2	db10
7	2	db10	1	db7
10	1	db5	1	db10
24	0		0	

According to Table 5, some rules are summarized as follows:

1) For oscillations with the same amplitude, the selected decomposition scale decreases with an increase in the frequency. The reason is that if the frequency is higher, the valuable oscillation information will be leaked into the WCs on a lower scale.

2) For oscillations with the same amplitude, if the same scale is selected, the selected N will increase with an increase in the frequency. The reason is that if the oscillation frequency is higher, the valuable oscillation information will be leaked into the WCs more easily. As explained in Sect. 3.1, an increase in N can inhibit the leakage.

3) For oscillations with the same frequency, if the same scale is selected, the selected N will increase with an increase in the amplitude. The reason is that if the amplitude is higher, more valuable information will be leaked into the WCs. As explained in Sect. 3.1, an increase in N can inhibit the leakage.

3.4 Piecewise linear model

Since there is a positive correlation between the frequency and the selected scale and order, a linear fit can be applied to the data points in Table 5. Specifically, a linear fit of $\log_2 f$ and the selected N is carried out for oscillations with the same amplitude and the same selected scale. If the linear correlation coefficient r satisfies $r > 0.6$ and the P value satisfies $P < 0.01$, it can be considered that $\log_2 f$ and the selected N have a strong linear correlation. All lines of different amplitudes and decomposition scales are shown in Fig. 7. Different colors indicate different scales, and
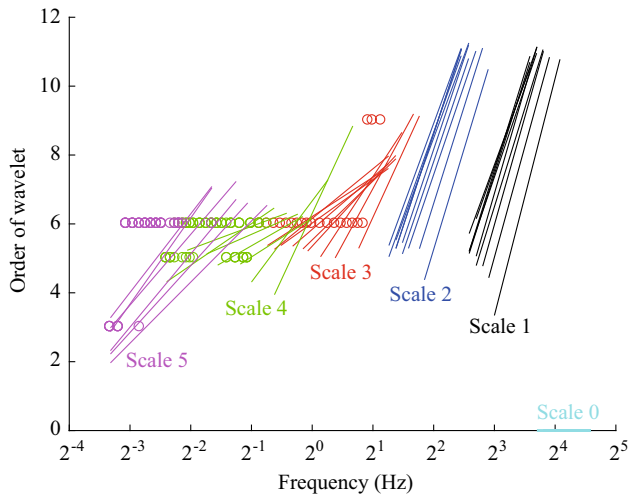
Fig. 7 Relationship between oscillation frequency and order of wavelet of different oscillation amplitudes and decomposition scales

scale 0 represents the corresponding oscillations that cannot be compressed by the compression method. The circles indicate data points that are not suitable for a linear fit.

According to Fig. 7, the rules are summarized as follows:

1) In most cases, there is a linear relationship between $\log_2 f$ and the selected N for oscillations with the same amplitude and the same selected scale, and there is still a positive correlation in other cases.

2) For oscillations with the same frequency and the same selected decomposition scale, the selected orders of the wavelet for oscillations of different amplitudes are close to each other. Since the order of the wavelet has very little effect on the compression performance and reconstruction accuracy, an averaged N is acceptable to represent all orders corresponding to different amplitudes.

According to the above two points, for oscillations with the same selected decomposition scale and different amplitudes, a linear fit of $\log_2 f$ and the selected N is acceptable. Although the best N may not be selected based on the linear model, the model is much more simplified, and the selected order is acceptable.

There are overlaps between different frequency ranges that correspond to different selected scales. In these areas, a higher scale means higher compression ratios and higher distortion rates, whereas a lower scale means lower compression ratios and lower distortion rates. It is difficult to decide which choice is better. Therefore, both linear models that correspond to both scales should be used to select the wavelets. The better choice can be selected based on the criterion of the minimum CDCI.

In conclusion, a linear fit of $\log_2 f$ and the selected N can be carried out for oscillations with the same selected decomposition scale and different amplitudes. As a result, a piecewise linear model with overlaps between different frequency ranges of different pieces is developed, in which different pieces represent different selected scales. All lines of different selected scales are shown in Fig. 8. The linear correlation coefficients r for a scale of 1 to 5 are about 0.79, 0.73, 0.65, 0.48, and 0.58 respectively, and all P values satisfy $P < 0.01$. Since the order of the wavelet has very little effect on the compression performance and reconstruction accuracy, the linear fit is considered to be acceptable.

The piecewise linear model can be expressed as:

$$N = \begin{cases} 1.357 \log_2 f + 8.081 & f \in [0.1, 0.6] & S5 \\ 0.538 \log_2 f + 6.288 & f \in [0.19, 1.6] & S4 \\ 1.330 \log_2 f + 5.944 & f \in [0.55, 3.4] & S3 \\ 4.113 \log_2 f - 0.257 & f \in [2.4, 7.5] & S2 \\ 4.437 \log_2 f - 6.401 & f \in [6, 17] & S1 \\ \text{Null} & f \in [13, 24] & S0 \end{cases}$$

(13)

where S represents the scale. Based on the equation, the decomposition scale and order of the wavelet for the compression of oscillations can be calculated according to the oscillation frequency. Specifically, the scale can be selected by the frequency ranges, and the order of the wavelet is the rounding of the calculation result of the corresponding equation. If the calculated N is greater than 10, the selected N should be 10. If the frequency is in different frequency ranges, all corresponding scales and calculated orders should be applied to the compression
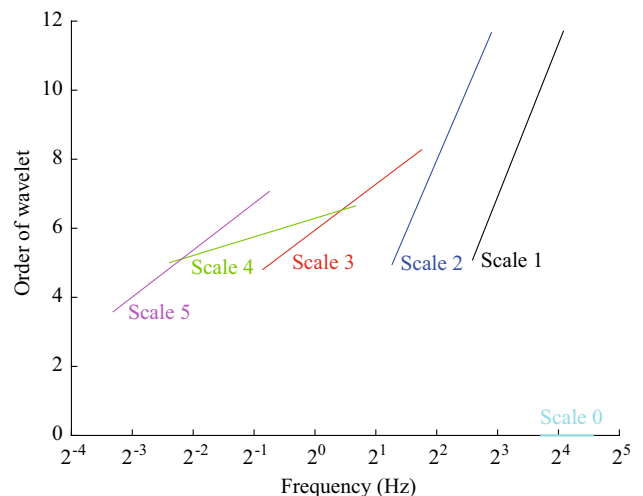

Fig. 8 Relationship between oscillation frequency and order of wavelet of different decomposition scales

STATE GRID
STATE GRID ELECTRIC POWER RESEARCH INSTITUTE

method, among which the best can be selected based on the CDCI.

According to (6), the amount of computation of the proposed method can be calculated as:

$$C_{\text{this}} = \begin{cases} 8n_0N(1 - 2^{-I}) & f \notin O \\ \sum[16n_0N_k(1 - 2^{-I_k}) + 2n_0] & f \in O \end{cases} \quad (14)$$

where O represents overlaps and k represents the corresponding frequency ranges. When $f \in [0.55, 0.6]$, scales 3 to 5 should be applied to the compression. Thus, the maximum amount of computation can be calculated as about $274.5n_0$. The computational burden is much lighter than that of [25].

4 Comparison

4.1 Proposed piecewise linear model-based data compression method

The actual recorded LFO and SSO signals in Sect. 3 are used in the proposed piecewise linear model-based method.

Since the frequency of the LFO signal is about 0.9 Hz, the optional decomposition scales and wavelet functions are calculated as db6–scale 4 and db6–scale 3. The signal is compressed and reconstructed with a window length of 10 s. Both choices are used in the compression method, among which db6–scale 4 is selected based on the criterion of minimum CDCI, for all data in different time windows. The choice is very close to the selection made in Sect. 3.1. The compression ratio is calculated as 13.62, and the distortion rate is 3.609×10^{-4}. The amount of computation is about $178n_0$ according to (14). The original and reconstructed LFO signals between 0 and 10 s are shown in Fig. 9.

Similarly, since the frequency of the SSO signal is about 7 Hz, the optional scales and wavelets are calculated as db10–scale 2 and db6–scale 1. The length of the time
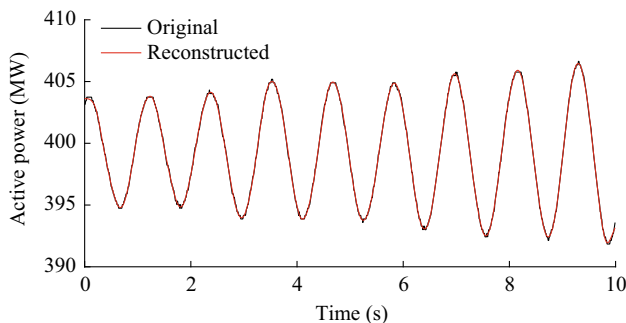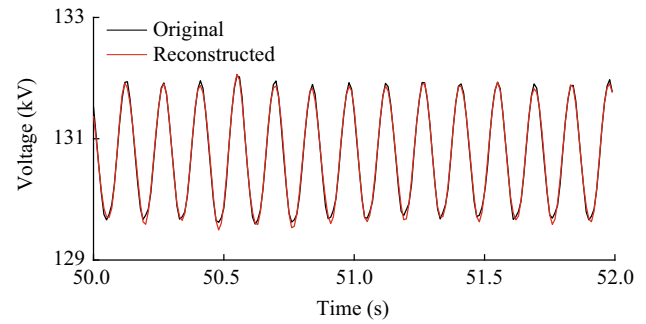

Fig. 10 Original and reconstructed SSO signals between 50 and 52 s by proposed method

window is 10 s as well. Both choices are used in the compression method, among which db10–scale 2 is selected based on the CDCI. The choice is very close to the selection made in Sect. 3.2. The compression ratio is calculated as 3.773, and the distortion rate is 3.728×10^{-4}. The amount of computation can be calculated as $172n_0$ according to (14). The original and reconstructed LFO signals between 50 and 52 s are shown in Fig. 10.

4.2 Wavelet-based data compression method with fixed wavelet and scale

Reference [18] selected the best wavelet and scale as db2–scale 5 for the data compression of disturbance signals, based on the criterion of maximum wavelet energy. To compare with the proposed piecewise linear model-based method, the wavelet of db2 and scale 5 is used in the compression of the LFO and SSO signals shown in Sect. 3. The amount of computation can be calculated as about $15.5n_0$ according to (14).

For compression of the LFO signal, the compression ratio is 11.09, and the distortion rate is 1.948×10^{-3}. The original and reconstructed LFO signals are shown in Fig. 11. According to the figure, there is a significant distortion of the reconstructed signal. Moreover, the
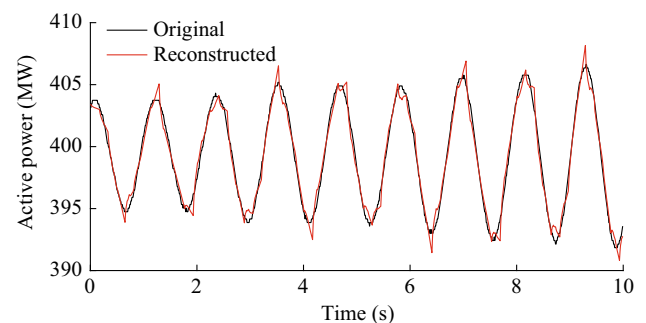

Fig. 9 Original and reconstructed LFO signals between 0 and 10 s by proposed method

Fig. 11 Original and reconstructed LFO signals between 0 and 10 s by db2 and scale 5

Springer

compression ratio is lower than that of the piecewise linear model-based method.

For compression of the SSO signal, the compression ratio is 28.62, and the distortion rate is 5.277×10^{-3}. The original and reconstructed LFO signals are shown in Fig. 12. According to the figure, the oscillation information is destroyed.

To sum up, the fixed wavelet and scale of db2–scale 5 are not suitable for the compression of oscillations owing to the severe distortion, despite the small amount of computation. The decomposition scale of 5 is too high for most oscillations, causing the valuable information of oscillation to be leaked out into WCs and then destroyed by the threshold method. As a result, there is a significant distortion of the reconstructed signal. At a wavelet order of 2, the oscillation information is more easily leaked out into WCs than other wavelets, causing more severe distortion. In addition, the regularity of db2 is worse than that of other wavelets, which means the reconstructed waveform is not smooth enough. Therefore, the wavelet of db2 may not be suitable for the compression of oscillations.

4.3 ESDC method

In [11], a real-time data compression method based on the EC and SDT methods was proposed where LFO signals were compressed with good compression performance. To compare this with the proposed piecewise linear model-based method, the LFO and SSO signals shown in Sect. 3 are compressed by the ESDC method. According to [11], the parameters are set as follows: $T_{max} = 0.2$ s, and $V_{ExcDev} = V_{CompDev} = 0.001 V_{base}$. The main computation of ESDC is multiplication and division to calculate the SDT criterion, and the number of calculations depends on the signal to be compressed. Therefore, the amount of computation can only be measured rather than calculated based on parameters.

For compression of the LFO signal, the base value of the active power is set as 686 MVA, which is the capacity of the connected generator. After the compression and reconstruction, the compression ratio is calculated as 16.949, and the distortion rate is 2.267×10^{-3}. The amount of computation is measured as about $0.87 n_0$ multiplications and divisions. The original and reconstructed LFO signals are shown in Fig. 13. Although the compression ratio is slightly higher than that of the piecewise linear model-based method, the distortion rate is much higher.

For compression of the SSO signal, the base value of the phase voltage is set as $230/\sqrt{3} \approx 133$ kV. After the compression and reconstruction, the compression ratio is calculated as 3.418, and the distortion rate is 7.318×10^{-4}. The amount of computation is measured as about $1.18 n_0$ multiplications and divisions. The original and reconstructed SSO signals are shown in Fig. 14. The compression ratio is lower and the distortion rate is higher than that of the piecewise linear model-based method.

To sum up, the proposed piecewise linear model-based method is more suitable for the compression of oscillations in storage than the ESDC method. Specifically, the amount of computation of the ESDC method is much less than that of the proposed method, whereas the distortion rate of the proposed method is lower than that of the ESDC method. The compression ratios of the two techniques are close to each other. The ESDC method was proposed for real-time applications, thus requiring a small amount of computation.
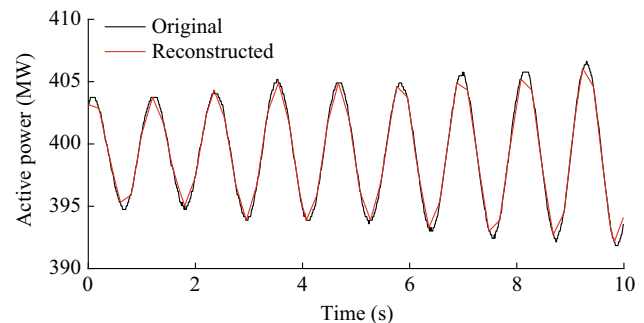

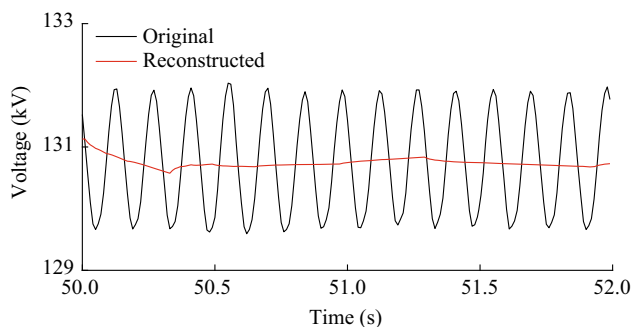**Fig. 13** Original and reconstructed LFO signals between 0 and 10 s by ESDC method

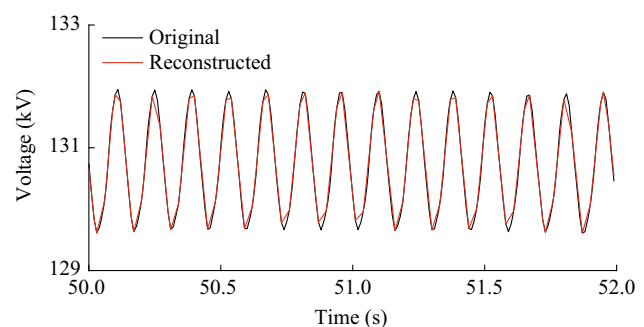**Fig. 12** Original and reconstructed SSO signals between 50 and 52 s by db2 and scale 5

Fig. 14 Original and reconstructed SSO signals between 50 and 52 s by ESDC method

By contrast, the proposed method is used for storage, and thus can afford a heavy computational burden. Therefore, the performance of the proposed method is better than that of the ESDC method in applications of data storage owing to the low distortion rate. Moreover, the parameters of the ESDC method are set based on experience. By contrast, the wavelet and scale can be directly selected based on the oscillation frequency in the proposed method.

5 Conclusion

Wavelet-based data compression techniques are widely used for data compression in power systems. However, most of the wavelet-based methods are only suitable for the compression of disturbance signals. In this paper, a wavelet-based data compression method is proposed for the compression of oscillations in power systems. Actual recorded oscillations and simulated oscillations are compressed and reconstructed by the wavelet-based data compression method to select the best wavelet functions and decomposition scales according to the criterion of the minimum CDCI for a balanced consideration of compression performance and reconstruction accuracy. Based on the selections, a piecewise linear model of the logarithm of the oscillation frequency and the order of the wavelet is developed, in which different pieces represent different scales. As a result, the decomposition scale and wavelet function can be selected according to the oscillation frequency. Compared with the wavelet-based data compression method with a fixed wavelet and decomposition scale for disturbance signals and the ESDC method for oscillations, the proposed method can provide high compression ratios and low distortion rates. Specifically, the compression ratio depends on the oscillation frequency, and can almost reach the compression ratio limit $\lambda_{CR,lim}(i) = 2^i$ of scale i. The distortion rate is on the order of 10^{-4} in general, and is always no more than 2×10^{-3}. The computational burden is not heavy for the compression in storage.

References

[1] Yang Q, Bi T, Wu J (2007) WAMS implementation in China and the challenges for bulk power system protection. In: Proceedings of the 2007 IEEE power engineering society general meeting, Tampa, USA, 24–28 June 2007, 6 pp

[2] Gupta AK, Verma K, Niazi KR (2017) Intelligent wide area monitoring of power system oscillatory dynamics in real time. In: Proceedings of the 2017 4th international conference on advanced computing and communication systems, Coimbatore, India, 6–7 January 2017, 6 pp

[3] Wang B, Sun K (2017) Location methods of oscillation sources in power systems: a survey. J Mod Power Syst Clean Energy 5(2):151–159

[4] Zacharia L, Hadjidemetriou L, Kyriakides E (2017) Cooperation of wide area control with renewable energy sources for robust power oscillation damping. In: Proceedings of the 2017 IEEE Manchester PowerTech, Manchester, UK, 18–22 June 2017, 6 pp

[5] Liu Z, Chen Z, Sun H et al (2015) Multiagent system-based wide-area protection and control scheme against cascading events. IEEE Trans Power Deliv 30(4):1651–1662

[6] Huang C, Li F, Zhou D et al (2016) Data quality issues for synchrophasor applications part I: a review. J Mod Power Syst Clean Energy 4(3):342–352

[7] Wang C, Lv Y, Huang H et al (2015) Low frequency oscillation characteristics of East China power grid after commissioning of Huai-Hu ultra-high voltage alternating current project. J Mod Power Syst Clean Energy 3(3):332–340

[8] Shen C, An Z, Dai X et al (2016) Measurement-based solution for low frequency oscillation analysis. J Mod Power Syst Clean Energy 4(3):406–413

[9] Liu H, Bi T, Chang X et al (2016) Impacts of subsynchronous and supersynchronous frequency components on synchrophasor measurements. J Mod Power Syst Clean Energy 4(3):362–369

[10] Wu M, Xie L, Cheng L et al (2016) A study on the impact of wind farm spatial distribution on power system sub-synchronous oscillations. IEEE Trans Power Syst 31(3):2154–2162

[11] Zhang F, Cheng L, Li X et al (2015) Application of a real-time data compression and adapted protocol technique for WAMS. IEEE Trans Power Syst 30(2):653–662

[12] Gadde PH, Biswal M, Brahma S et al (2016) Efficient compression of PMU data in WAMS. IEEE Trans Smart Grid 7(5):2406–2413

[13] Xie L, Chen Y, Kumar PR (2014) Dimensionality reduction of synchrophasor data for early event detection: linearized analysis. IEEE Trans Power Syst 29(6):2784–2794

[14] Kraus J, Štěpán P, Kukačka L (2012) Optimal data compression techniques for smart grid and power quality trend data. In: Proceedings of the 2012 IEEE 15th international conference on harmonics and quality of power, Hong Kong, China, 17–20 June 2012, pp 707–712

[15] Tate JE (2016) Preprocessing and Golomb–Rice encoding for lossless compression of phasor angle data. IEEE Trans Smart Grid 7(2):718–729

[16] Santoso S, Powers EJ, Grady WM (1997) Power quality disturbance data compression using wavelet transform methods. IEEE Trans Power Deliv 12(3):1250–1257

[17] Littler TB, Morrow DJ (1999) Wavelets for the analysis and compression of power system disturbances. IEEE Trans Power Deliv 14(2):358–364

[18] Ning J, Wang J, Gao W et al (2011) A wavelet-based data compression technique for smart grid. IEEE Trans Smart Grid 2(1):212–218

[19] Khan J, Bhuiyan S, Murphy G et al (2014) PMU data analysis in smart grid using WPD. In: Proceedings of the 2014 IEEE PES T & D conference and exposition, Chicago, USA, 14–17 April 2014, 5 pp

[20] Khan J, Bhuiyan S, Murphy G et al (2016) Data denoising and compression for smart grid communication. IEEE Trans Signal Inf Process Over Netw 2(2):200–214

[21] Khan J, Bhuiyan S, Murphy G et al (2015) Embedded-zerotree-wavelet-based data denoising and compression for smart grid. IEEE Trans Ind Appl 51(5):4190–4200

[22] Yan C, Liu J, Yang Q (2006) A real-time data compression and reconstruction method based on lifting scheme. In: Proceedings of the 2005/2006 IEEE/PES transmission and distribution conference and exhibition, Dallas, USA, 21–24 May 2006, pp 863–867

[23] Panda G, Dash PK, Pradhan AK et al (2002) Data compression of power quality events using the slantlet transform. IEEE Trans Power Deliv 17(2):662–667

[24] Hamid EY, Kawasaki ZI (2002) Wavelet-based data compression of power system disturbances using the minimum description length criterion. IEEE Trans Power Deliv 17(2):4190–4200

[25] Ji X, Zhang F, Cheng L et al (2017) A wavelet-based universal data compression method for different types of signals in power systems. In: Proceedings of the 2017 IEEE power and energy society general meeting, Chicago, USA, 16–20 July 2017, 5 pp

[26] Donoho DL (1995) De-noising by soft-thresholding. IEEE Trans Inf Theory 41(3):613–627

[27] State Grid Corporation of China (2015) Q/GDW 1131-2014: technology specifications of power system real time dynamic monitoring system. China Electric Power Press, Beijing

[28] Misiti M, Misiti Y, Oppenheim G et al (2016) Wavelet toolbox user's guide. https://www.mathworks.com/help/pdf_doc/wavelet/wavelet_ug.pdf. Accessed 6 September 2017

[29] Zhang L, Liu Y (2004) Bulk power system low frequency oscillation suppression by FACTS/ESS. In: Proceedings of the 2004 IEEE PES power systems conference and exposition, New York, USA, 10–13 October 2004, pp 219–226

Lin CHENG received the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2001. Currently, he is an associate professor with the State Key Laboratory of Control and Simulation of Power System and Generation Equipment, Department of Electrical Engineering, Tsinghua University. His research interests are reliability of power system, power system dynamics and control based on WAMS.

Xinchi JI received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2016. Currently, he is pursuing the M.S. degree with the State Key Laboratory of Control and Simulation of Power System and Generation Equipment, Department of Electrical Engineering, Tsinghua University. His research interests are in the areas of power system dynamic control based on WAMS.

Fang ZHANG received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2010 and 2015, respectively. Currently, he is a postdoctoral fellow with the State Key Laboratory of Control and Simulation of Power System and Generation Equipment, Department of Electrical Engineering, Tsinghua University. His research interests are in the areas of power system dynamic control based on WAMS.

He HUANG received the M.S. degree in electrical engineering from Southeast University, Nanjing, China, in 2001. Currently, he is a senior engineer in the Department of Development and Planning, State Grid Jiangsu Electric Power Company. His research interests are in the areas of distribution network planning.

Song GAO received the M.S. degree in electrical engineering from North China Electric Power University, Beijing, China, in 2006. Currently, he is a senior engineer in the Department of Development and Planning, State Grid Jiangsu Electric Power Company. His research interests are in the areas of distribution network planning.