Curvature Quantified Douglas-Peucker-based Phasor Measurement Unit Data Compression Method for Power System Situational Awareness

Weitao Tan, *Student Member*, *IEEE*, Tianhan Zhang, *Student Member*, *IEEE*, Yuanqian Ma, Shengyuan Liu, Li Yang, *Member*, *IEEE*, and Zhenzhi Lin, *Senior Member*, *IEEE*

***Abstract*—Facing constraints imposed by storage and bandwidth limitations, the vast volume of phasor measurement unit (PMU) data collected by the wide-area measurement system (WAMS) for power systems cannot be fully utilized. This limitation significantly hinders the effective deployment of situational awareness technologies for systematic applications. In this work, an effective curvature quantified Douglas-Peucker (CQDP)-based PMU data compression method is proposed for situational awareness of power systems. First, a curvature integrated distance (CID) for measuring the local flection and fluctuation of PMU signals is developed. The Douglas-Peucker (DP) algorithm integrated with a quantile-based parameter adaptation scheme is then proposed to extract feature points for profiling the trends within the PMU signals. This allows adaptive adjustment of the algorithm parameters, so as to maintain the desired compression ratio and reconstruction accuracy as much as possible, irrespective of the power system dynamics. Finally, case studies on the Western Electricity Coordinating Council (WECC) 179-bus system and the actual Guangdong power system are performed to verify the effectiveness of the proposed method. The simulation results show that the proposed method achieves stably higher compression ratio and reconstruction accuracy in both steady state and in transients of the power system, and alleviates the compression performance degradation problem faced by existing compression methods.**

***Index Terms*—Curvature quantified Douglas-Peucker, data compression, phasor measurement unit, power system situational awareness.**

I. Introduction

Because of the increasing deployment of phasor measurement units (PMUs) and phasor data concentrators (PDCs), vast volumes of PMU data are being collected at a fairly high reporting rate [1]. Limited by communication and storage capacities, these vast volumes may cause severe communication congestion and discontinuous storage. This greatly restricts systematic applications of enhanced situational awareness techniques (e.g., power system event detection [2] identification [3], and location [4]). Also, time-latency in the range of 100 ms to 5 s is required by most WAMS applications [5]. To address these problems, compressing PMU data in WAMS substations is an effective approach. Once the compressed PMU data are transmitted through the communication network, they can be reconstructed at WAMS master stations to restore their original resolution for high-accuracy applications.

Compression techniques for PMU data include lossless and lossy categories. Lossless compression techniques tend to reduce the scale of PMU data without any loss, i.e., the compressed data can be completely reconstructed into the original PMU data. In [6], a preprocessing method based on frequency-compensated difference encoding is presented to reduce the complexity of PMU data, and the Golomb-Rice codes-based entropy encoder is then exploited to compress the preprocessed data. In [7], a slack-referenced encoding (SRE)-based technique is used for the compression of PMU data from different sources. Nevertheless, these lossless compression techniques are adapted from other fields for general applications, prioritizing high-accuracy reconstruction at the expense of achieving low compression ratios. Hence, as mentioned in [8], lossless compression methods cannot achieve the high compression ratio required for big data applications, such as wide-area synchrophasors networks where the data of thousands of PMUs are shared among multiple entities. In pursuit of a balanced approach, the lossy compression technique that achieves higher compression ratios, albeit with a degree of information loss, is preferred and adopted for the compression of PMU data in this study.

Lossy compression techniques can significantly reduce the scale of PMU data, though at the cost of introducing limited errors between the reconstructed data and the originals. Signal feature analysis is a

Received: February 28, 2023
Accepted: October 8, 2023
Published online: March 1, 2024
Zhenzhi Lin is with College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China (linzhenzhi@zju.edu.cn).
DOI: 10.23919/PCMP.2023.000148

well-studied lossy compression technique for PMU data, among which the most widely used include the principal component analysis (PCA)-based [9] and wavelet transform (WT)-based [10] methods. However, data buffering is required for PCA-based and WT-based methods, resulting in insufficient timeliness for real-time applications. Another type of lossy compression technique is based on segment approximation algorithms. They approximate the PMU signals as segments and select many fewer data points to represent their trends. For real-time event detection, the ordinary least squares (OLS) [11] and swinging door trending (SDT)-based compression algorithms [12], [13] are developed to generate multiple compression segments of the PMU data. However, the selection of algorithm parameters is a serious challenge for OLS and SDT-based methods. More specifically, the tunable parameters of OLS and SDT-based algorithms are empirically set and remain constant, but the compression performance primarily depends on these algorithm parameters. While a group of algorithm parameters may achieve relatively high compression ratios with minimal distortion in the steady state of the power system, they may not be suitable for compressing rapidly changing PMU data under severe disturbance. This can lead to significant fluctuations in compression ratios. Similarly, the algorithm parameters for disturbed PMU data are also unsuitable for PMU data in the steady state of the power system. This phenomenon that the compression ratios undergo significant and unpredictable changes because of the mismatch of algorithm parameters for varying power system dynamics is referred to as compression performance degradation. The DP algorithm is also a well-studied lossy compression method for vector data based on segment approximation [14], [15]. The fundamental concept behind the DP algorithm is to progressively extract feature points of the data curve, transitioning from coarse to fine levels of detail, and thus the compressed data can retain the overall trend of signals with minimal distortion. However, a notable limitation of the DP algorithm is its inability to adequately capture the local characteristics of the PMU signals. This hinders its capacity to mitigate the issue of compression performance degradation, because of the challenge persisting in the selection of the algorithm parameters for different system dynamics.

Fundamentally, there is no absolute "steady state" for power systems [16]. Power systems typically operate in quasi-steady state, where the system operating conditions change primarily because of stochastic variations in demand/generation, while they can also experience transient states resulting from severe disturbances [17], [18]. Hence, it is of great significance for the compression method to tune the algorithm parameters for varying system dynamics, so as to maintain the desired compression ratio and reconstruction accuracy as much as possible, irrespective of the power system dynamics.

To address the compression performance degradation problem, an effective CQDP-based lossy compression method is proposed. Compared to previously used compression methods, the advantage of the proposed approach lies in its use of the curvature integrated distance (CID) and parameter adaptation scheme. This strategy effectively mitigates the degradation of compression performance caused by inappropriate parameters during significant shifts in power system dynamics caused by severe disturbances. As a result, it consistently delivers superior compression performance. First, CID is presented to measure the local flection and fluctuation trends within the PMU signals. Then, a DP-based algorithm is combined with a quantile-based parameter adaptation scheme to extract feature points that profile the contour of the PMU signals. Finally, the compressed data is reconstructed by a linear interpolation model to restore to the original resolution, for systematic applications in power systems. The contributions of this work are:

1) A CQDP-based PMU data compression method is proposed for power system situational awareness. By using the CID that measures local flection and fluctuation of PMU signals, and incorporating the DP algorithm that extracts feature points, the scale of PMU data can be compressed for easy communication and storage.

2) A quantile-based parameter adaptation scheme is embedded in the CQDP to effectively alleviate compression performance degradation for varying power system dynamics. Compared with previous lossy compression methods, the proposed method achieves a higher compression distortion composite index when the operating status of the power system changes, thus verifying its higher compression performance than other algorithms.

II. CQDP-based Data Compression Method for PMU Measurement

A noteworthy characteristic of the PMU data is that it keeps almost constant in the steady state of the power system, but fluctuates significantly under disturbances. Thus, as mentioned in Section I, robust compression methods for PMU data are required for varying power system dynamics. In the proposed method, the DP algorithm is improved by combining it with the CID and quantile-based parameter adaptation schemes, so as to alleviate the compression performance degradation problem caused by inappropriate algorithm parameters, particularly during significant changes in power system dynamics.

A. CQDP-based PMU Data Compression Method

Assume M PMUs are installed at major buses of the power system, and for each PMU, the reporting rate f_r and window length L are given. Then, the general form

of the N-point data series of the ith PMU $(i=1,2,\cdots,M)$ is denoted as:

$$V_i=[V_{i,1},V_{i,2},\cdots,V_{i,N}] \tag{1}$$

where $V_{i,j}$ ($j=1,2,\cdots,N$) is the PMU measurement (e.g., voltage amplitude data) at time t_j; and N is determined by L and f_r (i.e., $N=L\times f_r$). Then, N-1 equal intervals are segregated by N-2 interior points (i.e., $V_{i,2}$, $V_{i,3}$, $\cdots$, $V_{i,N-1}$) in the middle. Considering that the abrupt fluctuations and sharp variations in the PMU signals contain valuable insights into the power system dynamics, the motivation of the CQDP-based method is to extract data points that can profile the trends within the PMU signals.

CID is used to quantify the local flection and fluctuation of PMU signals, denoted as the product of curvature and Euclidean distance as:

$$\mu_{i,j}^{\mathrm{CID}}=\delta_{i,j}d_{i,j}=\delta_{i,j}\frac{\left|kt_j+b-V_{i,j}\right|}{\sqrt{1+k^2}},1\leqslant i\leqslant M,1<j<N \tag{2}$$

where $d_{i,j}$ is the point-to-edge distance from the interior point $V_{i,j}$ to the straight line concatenated by the start and end points; k and b are the respective slope and intercept coefficients of the straight line; and $\delta_{i,j}$ is the curvature at point $V_{i,j}$, represented as:

$$\delta_{i,j}=\frac{\left|\arctan\left(\frac{V_{i,j+1}-V_{i,j}}{t_{j+1}-t_j}\right)-\arctan\left(\frac{V_{i,j}-V_{i,j-1}}{t_j-t_{j-1}}\right)\right|}{\sqrt{(t_{j+1}-t_{j-1})^2+(V_{j+1}-V_{j-1})^2}}=\frac{\left|\arctan(f_r(V_{i,j+1}-V_{i,j}))-\arctan(f_r(V_{i,j}-V_{i,j-1}))\right|}{\sqrt{4(V_{i,j+1}-V_{i,j-1})^2/f_r^2}} \tag{3}$$

A higher CID indicates more pronounced and significant changes in the PMU signals and vice versa. Therefore the feature point of the PMU signals is defined as the data point with the largest CID in each iteration of the DP algorithm. Note that for PMU data with small fluctuation when there is no large disturbance in an actual power system, the CIDs of the interior data points are all significantly low (close to 0). This indicates that these data points are approximately linear, and can be highly compressed. On the contrary, a much higher CID will be obtained if a severe disturbance occurs on the corresponding data point. Hence, CID can be used for the compression of PMU data in both steady state and disturbance of the power system.

With the local flection and fluctuation measured by CID, the DP algorithm is adopted to extract feature points. The diagram of the CQDP-based compression method is presented in Fig. 1, where the red points denote extracted feature points, while the black and red lines denote raw PMU data and compressed PMU data, respectively. The CID threshold $\varepsilon_{\mathrm{CID}}$ is the only parameter in the CQDP-based compression algorithm, and it controls the error limit of the compression process. In each iteration, the maximum CID will be compared with $\varepsilon_{\mathrm{CID}}$, and only if the maximum CID $>\varepsilon_{\mathrm{CID}}$ will the data point be extracted as the feature point.

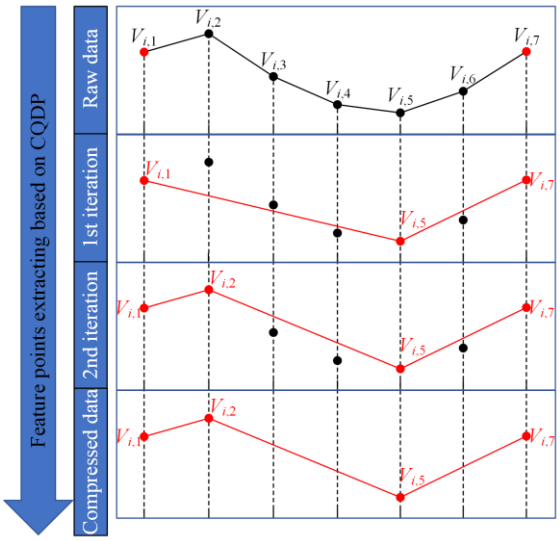


Fig. 1. Diagram of the CQDP-based algorithm for PMU data compression.

B. Quantile-based Parameter Adaptation Scheme for Varying Power System Dynamics

For the CQDP-based data compression algorithm, the CID threshold $\varepsilon_{\mathrm{CID}}$ is a critical parameter for compression and reconstruction performance. Generally, the value of $\varepsilon_{\mathrm{CID}}$ is set empirically to minimize the compression ratio within a limited error of the reconstructed data. However, the optimal selection of $\varepsilon_{\mathrm{CID}}$ may fluctuate greatly for different dynamics of the power system, and thus a constant $\varepsilon_{\mathrm{CID}}$ is not favorable for real-time applications. Therefore a quantile-based parameter adaptation scheme is presented here to adaptively select the value of $\varepsilon_{\mathrm{CID}}$.

With all CIDs calculated by (3), the CID vector is formed as:

$$\boldsymbol{\zeta}_i=[\mu_{i,1}^*,\mu_{i,1}^*,\cdots,\mu_{i,N-2}^*] \tag{4}$$

where $\mu_{i,1}^*>\mu_{i,2}^*>\cdots>\mu_{i,N-2}^*$. The CID threshold $\varepsilon_{\mathrm{CID}}$ is set as the p quantile value of $\boldsymbol{\zeta}_i$, expressed as:

$$\varepsilon_{\mathrm{CID}}=\mu_{i,\lfloor p^*\rfloor}^*+(\mu_{i,\lfloor p^*\rfloor+1}^*-\mu_{i,\lfloor p^*\rfloor}^*)(p^*-\lfloor p^*\rfloor) \tag{5}$$

$$p^*=1+(N-3)\times p \tag{6}$$

where p^* is the index of the p quantile value of $\boldsymbol{\zeta}_i$; and the symbol $\lfloor\cdot\rfloor$ is the floor function that calculates the maximum integer not greater than the input data. As the CIDs measure the local flection and fluctuation of PMU signal segments between two adjacent feature points, it can be inferred that the smaller the CIDs of interior

points between two adjacent feature points are, the higher compressibility these interior points have (i.e., the distribution of these points is approximately a line). Therefore, the p quantile can indicate the compressibility of PMU data points, and the data points with high compressibility will be removed in iterations of the DP algorithm, so as to accelerate the extraction process of feature points. The quantile of $p = 0.75$, also known as the quartile, is commonly used to reflect the outlier of a distribution in statistics [19]–[21]. This is also the setting in this study.

Since the p quantile defines the relative value of CIDs in a non-biased way, ε_{CID} can be selected adaptively irrespective of power system dynamics. To clarify further, the pseudocode of the CQDP-based compression method integrated with the quantile-based parameter adaptation scheme is presented in Algorithm 1.

Algorithm 1 CQDP-based data compression with parameter adaption

Input: PMU data series $\boldsymbol{V}_i = [V_{i,1}, V_{i,2}, \cdots, V_{i,N}]$
Set the number of reserved points N_i^{res}
Initialize feature points set as $\boldsymbol{F}_i = [V_{i,1}, V_{i,N}]$
While the length of $\boldsymbol{F}_i < N_i^{\text{res}}$:
  For adjacent feature points $V_{i,m}, V_{i,n} (m < n)$, in $\boldsymbol{F}_i$:
    If $(n-m) < 2$:
      Continue
    Else:
      Calculate CIDs from interior points $V_{i,j} (m < j < n)$ to the straight line $\overline{V_{i,m} V_{i,n}}$ via (2)
    End if
  End for
  Adjust the CID threshold ε_{CID} by (5)
  For adjacent feature points $V_{i,m}, V_{i,n} (m < n)$ in $\boldsymbol{F}_i$:
    If max $(\mu_{i,j}^{\text{CID}}) < \varepsilon_{\text{CID}}$:
      Remove the interior points $V_{i,j} (m < j < n)$
      $N \leftarrow N - n + m + 1$
    End if
  End for
  Append V_{i,j^*} to $\boldsymbol{F}_i$, where $j^* = \arg\max_{1<j<N} \mu_{i,j}^{\text{CID}}$
  $N \leftarrow N$ - 1, update the interior points and their CIDs
Output: feature point set $\boldsymbol{F}_i$, i.e., the compressed PMU data

Note that the PMU data point is a phasor including amplitude and phase angle components, which are treated as two independent data series to be compressed separately in this study. Even though the compression is processed separately, the amplitude and phase angle data will be reconstructed in the WAMS master station to restore the original resolution. This ensures the simultaneous and accurate representation of the phasor data. The complexity of the DP algorithm (i.e., extracting process of feature points), which primarily determines the execution time of the proposed method, is proved to be $O(mn)$ [22], where m is the iteration number of the DP algorithm, and n is the number of raw data points. For the compression of PMU data, the value of m is related to the status of the power system and the desired compression scale. The more PMU data can be compressed, or the more stable the data are, the smaller the value of m is. The value of n directly depends on the size of the time window given the fixed reporting rate of PMU. Thus, to avoid a high complexity, it is recommended to set a small window and a high compression ratio whenever feasible.

C. Reconstruction of the Compressed PMU Data

The compressed PMU data are formed by feature points that are sparsely extracted from the raw data, and the reconstruction is required first before further analysis and application. In order to achieve high fidelity of the raw PMU data, the linear interpolation model is used for reconstruction, and the reconstructed PMU data can be represented as:

$$\tilde{V}_{i,j} = \begin{cases} V_{i,j} & \text{if } V_{i,j} \neq \text{NULL} \\ V_{i,j-\Delta j_1} + \dfrac{V_{i,j+\Delta j_2} - V_{i,j-\Delta j_1}}{\Delta j_1 + \Delta j_2} \Delta j_1 & \text{if } V_{i,j} = \text{NULL} \end{cases} \quad (7)$$

where $\tilde{V}_{i,j}$ is the reconstructed data at time t_j; Δj_1 and Δj_2 are the minimum previous and forward intervals that make $V_{i,j-\Delta j_1} \neq \text{NULL}$ and $V_{i,j+\Delta j_2} \neq \text{NULL}$, respectively.

D. Performance Evaluation of the Compression Method

The performance of PMU data compression methods can be evaluated from two aspects: the compression scale and the accuracy of the reconstructed data.

Compression ratio (CR) is the most widely used indicator to measure the compression scale, shown as:

$$\eta_{\text{CR}} = (N_i^{\text{raw}} - N_i^{\text{res}}) / N_i^{\text{raw}} \quad (8)$$

where N_i^{raw} and N_i^{res} are the scales of the raw and the reserved PMU data, respectively.

Normalized mean square error (NMSE) is adopted to evaluate the accuracy of reconstructed data, denoted as:

$$\eta_{\text{NMSE}} = \frac{\left\| \boldsymbol{V}_i - \tilde{\boldsymbol{V}}_i \right\|_2}{\left\| \boldsymbol{V}_i \right\|_2} = \sqrt{\sum_{j=1}^{N} (V_{i,j} - \tilde{V}_{i,j})^2 \Big/ \sum_{j=1}^{N} V_{i,j}^2} \quad (9)$$

It should be noted that CR is usually associated with NMSE for the same compression method, and a higher CR will lead to a higher NMSE. As a compromise between CR and NMSE, the compression distortion composite index (CDCI) can be used to represent the comprehensive performance of the compression method, calculated as:

$$\eta_{\text{CDCI}} = a\eta_{\text{CR}}^* + b\eta_{\text{NMSE}}^* = a\frac{1}{\eta_{\text{CR}}} + b\frac{\eta_{\text{NMSE}}}{2\times 10^{-3}} \quad (10)$$

where η_{CR}^* is the normalized value of CR, denoted as $1/\eta_{\text{CR}}$ due to its range of (0,1]; whereas η_{NMSE}^* is the normalized value of NMSE with its base value set as 2×10^{-3} from [10]; a and b are the weights of CR and

NMSE, respectively. The smaller the CDCI is, the better performance of the compression method will be. Considering that the CR rather than the NMSE determines the communication time delay, a higher CR is more important for real-time applications. Thus, a and b are set as 0.7 and 0.3 respectively in this study.

III. Numerical Results and Comparisons

A. Compression Results with Simulated PMU Data

The proposed method is evaluated with the simulated PMU data on the western electricity coordinating council (WECC) 179-bus system, whose diagram can be seen in [23]. A three-phase short circuit fault is artificially set at Bus 8 and is cleared in 0.16 s. The simulated data is of 5 s duration, including the 1 s pre-event and the 4 s post-event.

The CQDP-based method is applied to the compression of PMU data, with the window length of 1 s in this case. Note that for the same compression method, the compression scale is contradicted with the reconstruction fidelity (i.e., the higher CR, the worse fidelity, and vice versa). To reach a trade-off, η_{NMSE} is required to be kept below 10^{-2}, which is a strict constrain for ensuring the accuracy of PMU data according to the IEEE standard for synchrophasor measurements [24]. Take the voltage amplitude data of Bus 1 as an example, the CRs, NMSEs and CDCIs over different time ranges are presented in Table I.

TABLE I
Compression Performances for Simulated Voltage Amplitude Data of Bus 1

Window range (s)	CR (%)	NMSE ($\times 10^{-3}$)	CDCI	Average CDCI
0 –1	98.0	0.065	0.724	
1 –2	91.0	0.970	0.923	
2 –3	90.0	0.424	0.841	0.831
3 –4	91.0	0.450	0.830	
4 –5	91.0	0.408	0.839	

It can be seen from Table I that, for the pre-event stage (0–1 s), the voltage amplitude of Bus 1 is almost constant, and thus only the start point and the end point are reserved, i.e., $\eta_{\mathrm{CR}} = 98\%$. Significantly low NMSE of 0.0065% and CDCI of 0.724 are also observed for these stable data. For the post-event stages (1–5 s), the CR slightly decreases but is still higher than 90%, with a NMSE of no higher than 0.097%. The reconstruction results of the compressed data are obtained with the linear interpolation, as shown in Fig. 2, where the blue lines denote the original PMU data, the black circles denote the compressed PMU data, and the red lines denote the reconstructed data. As can be seen in Fig. 2, the reconstructed data closely proximate the original one (η_{NMSE} ranges from 0.0065% to 0.097%) despite an average CR of 92.2%. Moreover, owing to CID, the compressed data (i.e., extracted feature points) are densely distributed in segments with abrupt changes and bends on the PMU signals. This aligns with the fact that more data are required for preserving critical changes in PMU signals as the disturbance information is more important to power system operators. Thus, the proposed method is effective to compress the simulated PMU data.

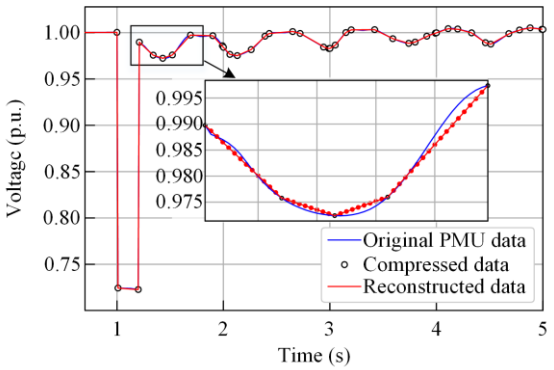


Fig. 2. Reconstruction result of the voltage amplitude data of Bus 1.

B. Compression Results with Actual Recorded PMU Data of the Guangdong Power System

The proposed method is further verified with actual PMU data of the Guangdong power system (GDPS) in this subsection. GDPS is a large interconnected power grid in southern China, with PMUs installed on all power plants and critical substations. Time-stamped streaming PMU data including voltage, frequency, and real/reactive power are recorded at 50 frames per second (FPS). Note that the actual recorded PMU data, which comes from our industrial partners in GDPS, have been filtered before archiving, so as to reduce the impact of noise.

On March 2, 2017, an actual generator ramping event was recorded by the PMUs. However, because of the limited storage and transmission capabilities, only partial recorded data in 31 PMUs were actually transmitted and saved. The recorded PMU data are shown in Fig. 3, where the PMU record is 56 s long from 11:50:00–11:50:56, including the whole duration of the event.

The compression performances for the CQDP-based method are presented in Fig. 4. It can be seen that:

1) With the proposed CQDP-based method, most of the PMU data are highly compressed yet the reconstruction error is still relatively low for all the data types in the steady state of the GDPS, i.e., the minimum CR is 75.4% for real power data, and the maximum NMSE is 0.153×10^{-3} for reactive power.

2) There are only slight changes on the CDCI during the generator ramping event, which means that the compression performance of the proposed method is largely unaffected by this disturbance.

3) The compression performances are different for different data types. Specifically, the CDCIs of the real and reactive power data are higher than those of the frequency and voltage data, but are still lower than 0.943 even in the worst case scenario.

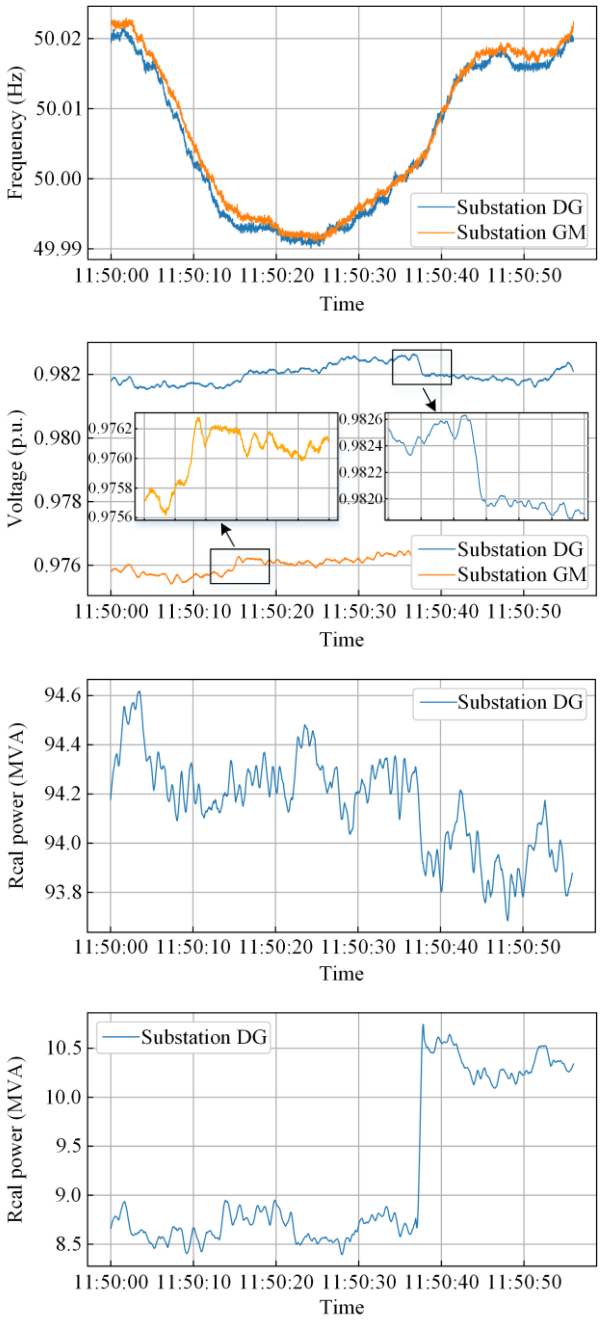
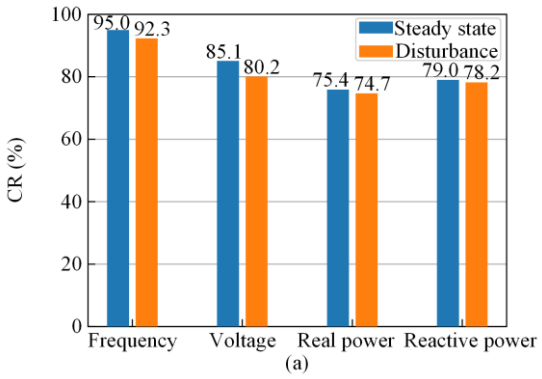


Fig. 3. Actual recorded PMU data of frequency, voltage, real/reactive power in GDPS.



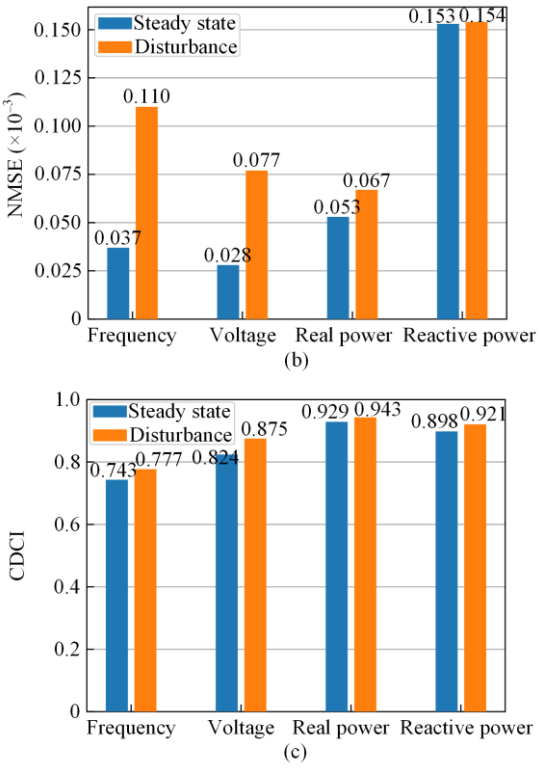


Fig. 4. Compression performance of the actual recorded PMU data in GDPS. (a) Compression ratio (CR). (b) Normalized mean square error (NMSE). (c) Compression distortion composite index (CDCI).

It can be concluded that the proposed method is effective in compressing PMU data while ensuring reconstruction accuracy for different data types and power system dynamics.

C. Sensitivity Analysis of the Window Length on Compression Performance

Because of the limited communication bandwidth, the PMU data are compressed and transmitted only when they accumulate to a certain size for achieving a high CR. Therefore the window length (i.e., the data buffering) is of great significance to the time delay. In this subsection, the compression performance with different window lengths are discussed. It should be noted that the average communication delay of PMU data is around 100 ms, and the data buffering for the WAMS master station is generally more than 1 s according to [25]. Thus, the window lengths of 1 s, 2.5 s, and 5 s are discussed here.

The test cases are performed on the actual recorded PMU data of the GDPS, and the PCA-based method [9] is used for comparison. Taking the voltage data of substation DG as an example, the reconstruction results with a 1 s window length are presented in Fig. 5, where the CRs for these two methods are both 90%. It can be seen that there are obvious distortions in the reconstructed data for the PCA-based method since crucial information during the severe disturbance is lost, while

the reconstructed data for the proposed method fit the original PMU data well, achieving much higher fidelity.

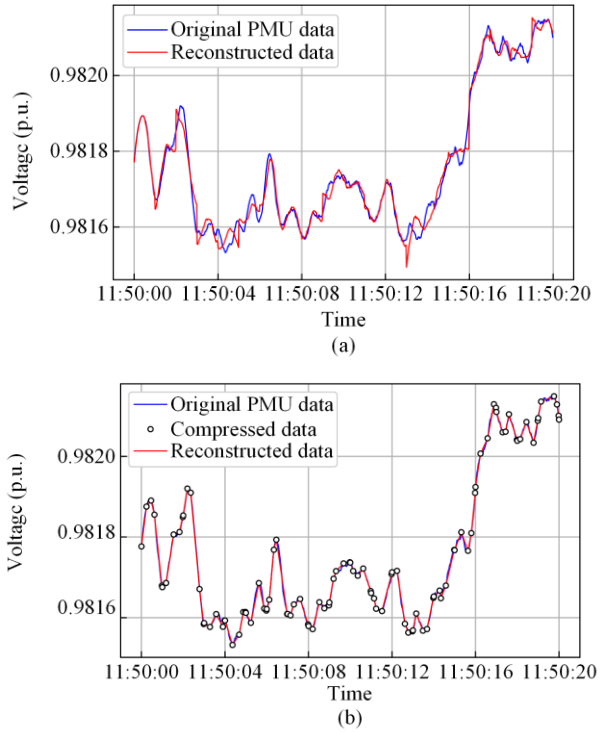


Fig. 5. Reconstruction results of actual recorded PMU data in GDPS with the 1s-length window. (a) PCA-based compression. (b) CQDP-based compression.

The overall compression performance with different window lengths are demonstrated in Table II.

TABLE II
COMPRESSION PERFORMANCE FOR VOLTAGE DATA WITH WINDOWS OF VARIOUS LENGTHS

Method	Window length (s)	CDCI	Execution time (s)
PCA [9]	1.0	1.106	0.005
	2.5	0.935	0.038
	5.0	0.905	0.181
CQDP	1.0	0.875	0.005
	2.5	0.874	0.031
	5.0	0.872	0.155

It can be seen that the CDCI of PCA changes from 0.905 to 1.106 as the window length decreases from 5 s to 1 s, which means that the compression performance is sensitive to the window length in case of disturbance, while a bigger window length is preferred. Hence the PCA-based method is more suitable for situations with low timeliness such as the data archival in PDCs integrated with multiple PMUs. For the CQDP-based method, the CDCI is largely unaffected by the window length, i.e., less data buffering is required. In practice, the selection of window length depends on the actual requirements of communication time delays (e.g., if a 2 s time delay is required, the window length can be set as 1.5 s, and the remaining time of 0.5 s is used for transmission and reconstruction). In addition, considering that a single PMU also has a certain data storage and processing capability, the CQDP-based compression method can be executed on a single PMU instead of the PDC, which further improves the near real-time processing capability.

D. Computational Complexity Analysis

To illustrate the practical computational complexity of the CQDP algorithm, case studies are conducted to calculate the time delay on a typical personal computer. The total time delay can be denoted as [26]:

$$\tau_{\text{tot}} = \tau_{\text{comp}} + \tau_{\text{trans}} + \tau_{\text{prop}} + \tau_{\text{queu}} + \tau_{\text{PMU}} \tag{11}$$

where τ_{tot} is the total delay; τ_{comp} is the execution time of the compression algorithm; τ_{trans} is the time delay of transmission, which approximately takes 2 ms for 10 phasors with a 10 Mbps bandwidth; and τ_{prop} is the time required for sending an electric signal, this is about 0.3 ms for a 300 km transmission line, and thus can be ignored; τ_{queu} is the queuing time delay related to the bandwidth of the communication network, and is insignificant until the network bandwidth downgrades to a certain threshold, and so can also be ignored as a high compression ratio is applied; τ_{PMU} is the PMU reporting delay, and its maximum value is $2/f_r$, where f_r is the reporting rate of PMU.

The average execution times of multiple tests for the proposed CQDP method are shown in Table III, where the time windows are all 100 samples-length.

TABLE III
RESULTS OF THE AVERAGE EXECUTION TIME FOR THE PROPOSED METHOD

Tested PMU data	Power system status	Average execution time (s)
Simulated data	Steady state	3.9
	Disturbance	15.6
Actual data	Steady state	6.1
	Disturbance	17.3

It can be seen that the maximum average execution time of the proposed CQDP method is 17.3 ms. As a result, in a real power grid, the proposed method could deliver compression results to PDC in less than 0.1 s after the PMU data is collected. Hence, the total time delay of the proposed method can satisfy the time-latency requirement (i.e., from 100 ms to 5 s) in [5], and can be applied for real-time applications in WAMS.

E. Comparisons with Other PMU Data Compression Methods

To demonstrate the superiority of the proposed method, the other two segment approximation-based methods, SDT [12] and DP [14] are employed for comparison. The comparison is first performed on the WECC 179-bus system with the simulated voltage data, and an ablation test is designed as follows. The PMU data in a steady

state of the power system is compressed with SDT, DP and CQDP methods, and the algorithm parameters of the compression methods are fine-tuned to achieve the same CDCI of 0.787 (i.e., the CRs are all 90%, and NMSEs are all around 3.6×10^{-5}). A three-phase short circuit fault is then applied, and the compression methods are used to compress the PMU data in the disturbance, without changing the algorithm parameters. The average CRs, NMSEs, and CDCIs of the above compression methods are calculated, as shown in Table IV.

TABLE IV
COMPRESSION PERFORMANCE FOR SIMULATED PMU DATA UNDER DISTURBANCE USING DIFFERENT METHODS

Method	CR (%)	NMSE ($\times10^{-3}$)	CDCI
SDT [12]	42.5	0.155	1.667
DP [14]	27.7	0.074	2.424
Proposed CQDP	79.0	0.523	0.839

It can be seen from Table IV that CDCIs of SDT and DP reach 1.667 and 2.424 respectively after the three-phase short circuit fault. The main reason for the increases in CDCI is the rapid decrease in CR, while the compensated reconstruction accuracy does not increase significantly. This indicates that the compressed data obtained by SDT and DP still retain high compressibility potential. Thus, inappropriate algorithm parameters may lead to the degradation of compression performance when there are large changes in power system dynamics, even though these parameters have been proved to be effective for the steady state of the power system. The benefits of the CID and parameter adaptation scheme in the CQDP method are fully demonstrated when compared with the basic DP algorithm. The CDCI of the proposed method only increases slightly from 0.787 to 0.839 after the severe disturbance. Thus, the results of the ablation test show that the merit of the proposed method is to alleviate the compression performance degradation when the power system dynamics change.

Then, the SDT, DP and CQDP compression methods are performed on the actual recorded PMU data of GDPS for further comparisons, and the results are shown in Table V.

TABLE V
COMPRESSION PERFORMANCE FOR VOLTAGE DATA WITH WINDOWS OF VARIOUS LENGTHS

Method	Data type	CR (%)	CDCI	Average CDCI
SDT [12]	Frequency	86.0	0.813	1.222
	Voltage	87.1	0.804	
	Real power	62.9	1.122	
	Reactive power	32.6	2.149	
DP [14]	Frequency	75.9	0.923	1.093
	Voltage	81.5	0.860	
	Real power	73.2	0.956	
	Reactive power	42.9	1.631	

As can be seen there, for SDT and DP-based methods, there is significant degradation in compression performance of the reactive power data, i.e., only 32.6% of the reactive power data is compressed for SDT, and 42.9% for DP. The reason is that the parameters of SDT and DP are set empirically based on the base value and error tolerance, but the optimal parameters are different for different data types and power system dynamics. As a result, the CRs of SDT and DP greatly reduce since there are more drastic fluctuations in reactive power. For the proposed method, it can adjust the algorithm parameter adaptively based on the trends within the PMU data, and thus outperforms SDT and DP in terms of CR for reactive power data (the CR is 78.2% as shown in Fig. 4). Also, the average CDCI for CQDP is 0.871, which is also much better than the 1.222 for SDT and 1.093 for DP.

IV. CONCLUSIONS

To address the challenges of insufficient communication and storage capabilities caused by large volumes of PMU data, an effective CQDP-based PMU data compression method is proposed for power system situational awareness. Case studies on simulated and actual PMU data demonstrate that the proposed method is capable of highly compressing PMU data in the presence of steady state and dynamic transient states of the power system, and achieves higher compression ratios and less data buffering for near real-time situational awareness.

ACKNOWLEDGMENT

Not applicable.

AUTHORS' CONTRIBUTIONS

Weitao Tan: complete full-text writing and the construction of the paper framework. Tianhan Zhang: interpret the simulation results and the drafting of the manuscript. Shengyuan Liu and Yuanqian Ma: provide simulation data and suggestions for improvement of the proposed method. Li Yang and Zhenzhi Lin have been the organizer and technical adviser for the total work. All authors read and approved the final manuscript.

FUNDING

This work is supported by the National Natural Science Foundation of China (No. 52077195).

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

DECLARATIONS

Competing interests: The authors declare that they

have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

AUTHORS' INFORMATION

Weitao Tan received the B.E. degree in the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2021. He is currently pursuing his M.E. degree at the College of Electrical Engineering, Zhejiang University, Hangzhou, China. His research interests include power system situation awareness and data-driven approaches in power systems.

Tianhan Zhang received the B.E. degree in the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2019. He is currently pursuing his Ph.D. degree at the College of Electrical Engineering, Zhejiang University, Hangzhou, China. His research interests include power system situation awareness and application of blockchain in power system.

Yuanqian Ma received the B.S. and Ph.D. degrees in the Electrical Engineering and its Automation from Sichuan University, Chengdu, China, in 2014 and 2019, respectively. She was a visiting student in Electrical Power Engineering with the University of Alberta for two years. She was a visiting scholar with the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2021. She is currently a lecturer with Zhejiang Sci-Tech University, Hangzhou, China. Her main research interests include demand response, electricity retail plan design and recommendation, and electricity market.

Shengyuan Liu received the B.E. degree in the Electrical Engineering from Shandong University, Jinan, China, in 2017, and the Ph.D. degree from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2022. From 2019 to 2020, he is also a visiting Ph.D. student for one year with The University of Tennessee, Knoxville, TN, USA. His research interests include wide-area monitoring and control of power systems, situation awareness of power systems, and data-driven approaches in power systems.

Li Yang received the Ph.D. degree in the Electrical Engineering from Zhejiang University, Hangzhou, China, in 2004. From 2007 to 2008, she was a postdoctoral with the Department of Electrical Engineering, Turin Polytechnic University, Tashkent, Uzbekistan. She is currently an associate professor with the College of Electrical Engineering, Zhejiang University. Her research interests include power market, power system economics, and distribution network planning.

Zhenzhi Lin received the Ph.D. degree in the Electrical Engineering from the South China University of Technology, Guangzhou, China, in 2008. From 2007 to 2008, he was a research assistant with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong, China, a research scholar with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA, from 2010 to 2011, and Research Associate with the College of Engineering and Computing Sciences, Durham University, Durham, U.K., from 2013 to 2014. He is currently a professor with the College of Electrical Engineering, Zhejiang University, Hangzhou, China. His research interests include power system situational awareness, power system restoration, power system economics, data mining, and artificial intelligence applications in power systems. He is also an associate editor for IEEE Transactions on Power Systems and IEEE Power Engineering Letters, subject editor of Energy Conversion and Economics (IET), and the editor of Protection and Control of Modern Power Systems (PCMP), Automation of Electric Power Systems and Journal of Modern Power Systems and Clean Energy (MPCE).

REFERENCES

[1] B. Appasani, A. V. Jha, and S. K. Mishra *et al*., "Communication infrastructure for situational awareness enhancement in WAMS with optimal PMU placement," *Protection and Control of Modern Power Systems*, vol. 6, no. 1, pp. 124-135, Jan. 2021.

[2] Y. Wang, C. Wang, and M. Chao *et al*., "Sub-and super-synchronous oscillation mode identification based on amplitude and frequency characteristics of synchronous phasor data," *Power System Protection and Control*, vol. 51, no. 19, pp. 1-11, Oct. 2023. (in Chinese)

[3] S. Liu, S. You, and Z. Lin *et al*., "Data-driven event identification in the U.S. power systems based on 2D-OLPP and RUSBoosted trees," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 94-105, Jan. 2022.

[4] S. Liu, K. Sun, and C. Zeng *et al*., "Practical event location estimation algorithm for power transmission system based on triangulation and oscillation intensity," *IEEE Transactions on Power Delivery*, vol. 37, no. 6, pp. 5190-5202, Dec. 2022.

[5] P. Kansal and A. Bose, "Bandwidth and latency requirements for smart transmission grid applications," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1344-1352, Sep. 2012.

[6] J. E. Tate, "Preprocessing and Golomb–Rice encoding for lossless compression of phasor angle data," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 718-729, Jun. 2016.

[7] R. Klump, P. Agarwal, and J. E. Tate *et al*., "Lossless compression of synchronized phasor measurements," in *IEEE PES General Meeting,* Minneapolis, MN, USA, July 25-29, 2010, pp. 1-7.

[8] R. Pourramezan, R. Hassani, and H. Karimi, "A real-time synchrophasor data compression method using singular value decomposition," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 564-575, Jan. 2022.

[9] P. H. Gadde, M. Biswal, and S. Brahma *et al*., "Efficient compression of PMU data in WAMS," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2406-2413, Sep. 2016.

[10] L. Cheng, X. Ji, and F. Zhang *et al*., "Wavelet-based data compression for wide-area measurement data of oscillations," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 6, pp. 1128-1140, Nov. 2018.

[11] Y. Ge, A. J. Flueck, and D. K. Kim *et al*., "Power system real-time event detection and associated data archival reduction based on synchrophasors," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 2088-2097, Jul. 2015.

[12] M. Cui, J. Wang, and J. Tan *et al*., "A novel event detection method using PMU data with high precision," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 454-466, Jan. 2019.

[13] F. Zhang, L. Cheng, and X. Li *et al*., "Application of a real-time data compression and adapted protocol technique for WAMS," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 653-662, Mar. 2015.

[14] Y. Pan, R. An, and D. Fu *et al*., "Unsupervised fault detection with a decision fusion method based on Bayesian in the pumping unit," *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21829-21838, Oct. 2021.

[15] C. Liu, J. Wang, and A. Liu, *et al*., "An asynchronous trajectory matching method based on piecewise space-time constraints," *IEEE Access*, vol. 8, pp. 224712-224728, Dec. 2020.

[16] J. Zhao, A. Gómez-Expósito, and M. Netto *et al*., "Power system dynamic state estimation: motivations, definitions, methodologies, and future work," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3188-3198, Jul. 2019.

[17] G. Zhang, Q. Zong, and X. Ke *et al*., "Online warning of transient power angle stability of power systems with wind power based on critical inertia and anticipated faults," *Power System Protection and Control,* vol. 51, no. 16, pp. 72-83, Aug. 2023. (in Chinese)

[18] C. Ren, Y. Xu, and Y. Zhang, "Post-disturbance transient stability assessment of power systems towards optimal accuracy-speed tradeoff," *Protection and Control of Modern Power Systems*, vol. 3, no. 2, pp. 194-203, Apr. 2018.

[19] A. Gholami, A. Vosughi, and A. K. Srivastava, "Denoising and detection of bad data in distribution phasor measurements using filtering, clustering, and Koopman mode analysis," *IEEE Transactions on Industry Applications*, vol. 58, no. 2, pp. 1602-1610, Mar. 2022.

[20] A. M. Shah, B. R. Bhalja, and R. M. Patel *et al*., "Quartile based differential protection of power transformer," *IEEE Transactions on Power Delivery*, vol. 35, no. 5, pp. 2447-2458, Oct. 2020.

[21] G. Li, H. Xiang, and Z. Zhao, "Fast fractal image encoding algorithm based on quartiles feature," *Computer Engineering and Applications*, vol. 47, no. 22, pp. 145-148, Nov. 2011.

[22] J. Hershberger and J. Snoeyink, "An $O(n\log n)$ implementation of the Douglas-Peucker algorithm for line simplification," in *Proceedings of the 10th Annual Symposium Computational Geometry*, pp. 383-384, 1994.

[23] S. Maslennikov, B. Wang, and Q. Zhang *et al*., "A test cases library for methods locating the sources of sustained oscillations," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, Boston, MA, USA, Jul. 17-21, 2016, pp. 1-5.

[24] IEEE standard for synchrophasor measurements for power systems. IEEE Std C37.118.1-2011, Dec. 2011.

[25] J. Zhao, Y. Zhang, and P. Zhang, *et al*., "Development of a WAMS based test platform for power system real time transient stability detection and control," *Protection and Control of Modern Power Systems*, vol. 1, no. 1, pp. 37-47, Jul. 2016.

[26] F. Ye, and A. Bose, "Multiple communication topologies for PMU-based applications: introduction, analysis and simulation," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5051-5061, Jul. 2020.