Spatio-Temporal Data Recovery in Continuous Streaming of Synchro-waveforms: Combining Matrix Completion with Deep Learning

Sirui He[†], Zhirui Tian[†], Chenye Wu[†], Hamed Mohsenian-Rad[‡]
[†]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Guangdong, China
[‡]Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA
E-mails: {223015060, zhiruitian}@link.cuhk.edu.cn, chenyewu@yeah.net, hamed@ece.ucr.edu

***Abstract*—Synchro-waveforms pioneer power system monitoring, supporting critical smart grid applications such as event and fault detection, renewable resource integration, and network dynamics tracking. Data quality in synchro-waveform measurements plays a crucial role in ensuring reliable inference and situational awareness. However, two key challenges arise when working with continuous streaming of synchro-waveforms: 1) the extremely large volume and velocity of data transmission that often leads to frequent missing data points; and 2) high-precision time synchronization, which plays a critical role but encounters real-world issues such as temporary loss of the GPS signal. We propose an innovative hybrid approach that integrates matrix completion with a learning ensemble to address both challenges. Matrix completion leverages the low-rank structure of synchro-waveform data across sensor locations to capture spatial dependencies and basic recovery, while deep learning captures temporal dependencies to refine the recovery. Ensemble learning techniques combine deep learning methods to enhance data processing. Experimental results based on real-world synchro-waveform data demonstrate high data recovery accuracy and computational efficiency, thus advancing the emerging and future applications of synchro-waveforms in power systems.**

***Index Terms*—Learning ensemble, matrix completion, data recovery, synchro-waveforms, continuous measurement streaming, power systems monitoring.**

I. Introduction

Synchro-waveforms represent voltage or current waveforms measured from multiple locations on a power network with precise time-synchronization enabled by technologies such as the global positioning system (GPS) [1]. A waveform measurement unit (WMU) [2] serves as the sensor device for capturing synchro-waveforms. The fine time granularity and precise synchronization inherent in synchro-waveform technology enable various complex data-driven power system applications, including real-time monitoring, control, and protection [3], [4], event detection and characterization [5], integration of inverter-based resources [6]–[8], wildfire monitoring [9], and fault location identification [10]. These applications rely on accurate and continuous synchro-waveform data to ensure the optimal functioning of power networks.

This work was supported by the National Natural Science Foundation of China (Grant 72271213), the Shenzhen Science and Technology Program (Grants JCYJ20220530143800001 and RCYX20221008092927070), by the Guangdong Basic and Applied Basic Research Foundation (Grant 2024A1515240024), and by the University of California Riverside's Winston Chung Endowed Chair Professorship on Energy Innovation..

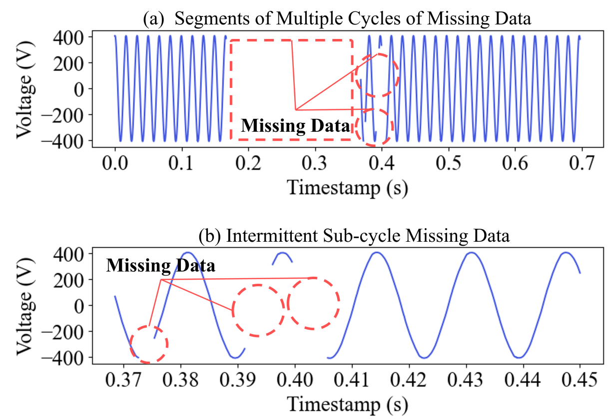

Fig. 1: Examples of Missing Data in Real-world Voltage Synchro-waveforms in a Pilot Testbed.

However, poor data quality poses a challenge in synchro-waveform data analysis. For example, Fig. 1 illustrates a real-world synchro-waveform testbed in Riverside, CA, where synchro-waveform data frequently gets lost due to limitations in the data communications infrastructure, depending on network traffic. Some segments experience extended periods of data loss spanning several cycles, while intermittent sub-cycle data loss also occurs. Although upgrading the communications infrastructure—such as adopting terabit Ethernet technologies—could significantly reduce data loss, this option involves high costs and may prove impractical in most real-world settings.

Therefore, we need reliable but also cost-efficient solutions to address missing data despite the extremely fast reporting rates of WMUs. We also need to address the common challenges of measurement noise, especially when we combine synchro-waveform measurements from multiple sensor locations. In this paper, we seek to address both challenges.

A. Related Work

Recent studies have provided valuable insights into challenges in data recovery for data-intensive power system monitoring, such as smart meters [11] and phasor measurement units (PMUs) [12]. Gao *et al.* [13] introduced a matrix decomposition approach that leverages the low-rank property of PMU data to recover missing entries, effectively addressing data gaps in large datasets. Similarly, Gao *et al.* [14] developed a robust recovery method by combining nuclear norm minimization with Bayesian estimation, exploiting the low-rank structure of state variables, and incorporating prior knowledge through second-order statistics. Wu *et al.* [15] applied network embedding techniques to recover missing PMU data, capturing complex interdependencies within the power grid via graph-based methods in voltage stability assessment. Zhang *et al.* [16] proposed a missing-data tolerant method for short-term voltage stability assessment, which strategically groups buses to maintain grid observability and employs a structure-adaptive ensemble learning model to adapt to available data, ensuring high assessment accuracy despite incomplete measurements.

The above studies underscore the importance of addressing missing data in power systems to ensure reliability and stability. However, these methods primarily focus on static or low-dimensional measurements and do not adequately tackle missing data recovery in waveforms or synchro-waveforms measured by WMUs. WMU data poses unique challenges due to its continuous streaming nature and intricate temporal dependencies, which must be preserved for accurate recovery. Moreover, WMU measurements represent high-dimensional signals that vary over both space and time [17] and are often contaminated by noise and outliers stemming from environmental disturbances and sensor limitations. These factors complicate the recovery process and necessitate specialized techniques tailored to the dynamic and complex characteristics of WMU data.

B. Technical Contributions

We address the above challenges by introducing a hybrid learning ensemble method that integrates different deep learning techniques. The main contributions of this paper can be summarized as follows:

- *Matrix Completion for Noise Containment*: We use Singular Value Decomposition (SVD) to exploit the low-rank structure of synchro-waveform data, recovering missing entries while preserving key features. SVD-based matrix completion approximates the measurement matrix by retaining the most significant singular values, ensuring robust recovery even under missing data and noise.
- *Deep Learning-based Robust Recovery*: After matrix completion, we apply Long Short-Term Memory (LSTM) networks and other deep learning methods to learn sequential patterns within the synchro-waveform data, smoothing out noise and enhancing temporal consistency, thereby improving the accuracy of data recovery.
- *Learning Ensemble for Boosted Performance*: To enhance robustness across different environments, these deep learning methods, including LSTM, GRU, and CNN, are integrated into an ensemble learning framework, crucial for effective synchro-waveform analysis.

II. METHODOLOGY

In this section, we describe the primary methods we employed in our data recovery research within synchro-waveforms. Then, we discuss the matrix completion for basic recovery. Last, we introduce the key components of learning ensemble and the application of learning ensemble.

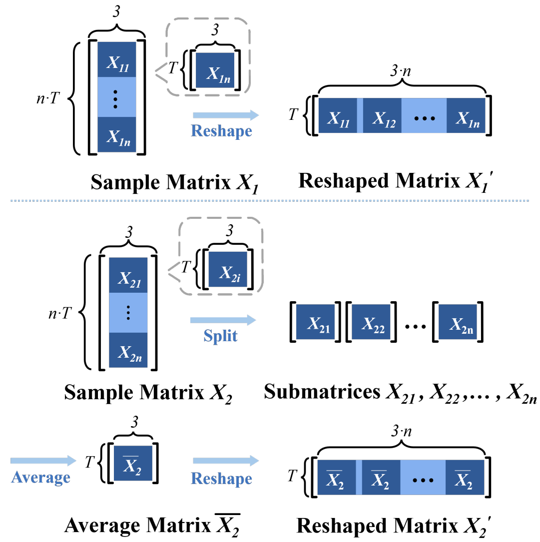

Fig. 2: Matrices X_1 and X_2 Reshaping Process Demonstration.

A. Overall Design Framework

Our approach comprises several key stages: initial data processing, matrix completion, advanced deep learning-based recovery, ensemble learning integration, and final evaluation. First, the data is organized according to different scenarios and split into training and testing subsets. In the initial recovery phase, we apply SVD-based matrix completion to perform denoising and fill in missing entries, providing a coarse but effective restoration. Subsequently, multiple deep learning models [18], [19] are employed in parallel to refine the recovery, leveraging their complementary strengths to further improve waveform reconstruction. This produces multiple recovered datasets corresponding to each deep learning method. To enhance robustness and accuracy, these diverse recovery results are combined using an ensemble learning strategy based on the Random Forest algorithm. This integration yields a unified, enhanced recovery outcome that better handles noise and missing data. Finally, the evaluation module rigorously assesses the performance of the overall recovery process using a suite of quantitative metrics.

B. Step 1: SVD-based Matrix Completion

In the context of synchro-waveforms, we use matrix completion to recover missing data from partial measurements. Matrix completion exploits the low-rank structure of the data to recover missing entries. The waveform data matrix (from WMUs) is assumed to be low-rank, meaning it can be approximated by a smaller set of dominant features.

We may have two closely spaced WMU devices located near each other, measuring the same physical quantity. The matrix X_1 (size $nT \times 3$) is obtained from the fully functional device, while the matrix X_2 (size $nT \times 3$) is obtained from the malfunctioning device. Here, T represents the period, and

n is the number of periods. We assume that the WMU with missing values contains n cycles, where n is typically a small number. Next, we divide the X_2 matrix into several $T \times 3$ submatrices, calculate the average of each submatrix, and then concatenate them to form X_2' (size $T \times 3n$) as shown in Fig. 2, which has the same size as X_1'. To facilitate the recovery process, we first reshape the X_1 matrix from an $nT \times 3$ matrix into a $T \times 3n$ matrix. The data is divided into several $T \times 3$ submatrices, which are concatenated horizontally as shown in Fig. 2, resulting in X_1' (size $T \times 3n$).

After data processing, we have two datasets: X_1' (complete data from one WMU) and X_2' (data from another WMU with missing values). Both datasets correspond to the same voltage measurements, but X_2' has missing data. The goal is to recover the missing data in X_2' using information from both datasets, ensuring the reconstructed data matches the actual voltage characteristics. The shared low-rank structure, temporal and spatial correlations between X_1' and X_2' aid in this recovery. We apply a customized SVD for matrix completion. First, we obtain the SVD of X_1', which are as follows:

$$X_1' = U_1 \Sigma_1 V_1^T, \Sigma_1 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_T). \tag{1}$$

Then, we select the top m singular values, representing the dominant features. These components are used to perform completion on X_2':

$$\Sigma_s = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_m), \quad X_2^{\text{recovery}} = U_2 \Sigma_s V_2^T. \tag{2}$$

The missing data is reconstructed using the low-rank approximation derived from X_1', resulting in the recovered matrix, X_2^{recovery}.

C. Step 2: Deep Learning Methods

After matrix completion for basic recovery, to further refine data recovery, we mainly utilize an LSTM-based architecture, which is well-suited for capturing sequential dependencies in time-series data. The LSTM model is governed by the following update equations, and the output h_t is fed into a fully connected layer to generate the recovery values for each time step:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{3}$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{5}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \tag{6}$$

$$h_t = o_t \cdot \tanh(c_t), \tag{7}$$

$$\hat{y}_t = W_{fc} \cdot h_t + b_{fc}, \tag{8}$$

where the parameters are defined as follows:

- i_t, f_t, o_t: Input, forget, and output gates at time step t.
- c_t, h_t: Cell state and hidden state at time step t.
- x_t: Input at time step t.
- $W_i, W_f, W_o, W_c, W_{fc}$: Weight matrices for input, forget, and output gate, candidate cell state, and fully connected layer.
- $b_i, b_f, b_o, b_c, b_{fc}$: Bias terms for input, forget, and output gates, candidate cell state, and fully connected layer.
- c_{t-1}: Cell state from the previous time step.
- $\hat{y}_t$: Predicted recovery value at time step t.

TABLE I: Parameters of Learning Methods

Model	Parameters	Value
LSTM	Feature Deminsion	6
	Bath Size	128
	Look-back Window	30
	Max Epoch	100
GRU	Batch Size	128
	Max Epoch	100
	Look-back Window	30
	GRU Layer Units	100
CNN	Batch Size	128
	Epochs	100
	Look-back Window	30
	Filters in Conv1D Layer	64
	Kernel Size in Conv1D Layer	3
	Pool Size in MaxPooling1D Layer	2
RF	Number of Trees	1,000

Following training, we evaluate the model's performance on the withheld test set to assess its generalization ability. During inference, gradient computation is disabled to speed up processing, ensuring faster recovery.

In addition to the LSTM, we also implement two alternative methods, Gated Recurrent Unit (GRU) and Convolutional Neural Networks (CNNs). GRUs, a simplified version of LSTMs, are used to test whether a more lightweight architecture can achieve similar results. Unlike LSTMs, GRUs have fewer parameters and are generally faster to train, but they may struggle with long-range dependencies. CNNs, although typically used for spatial data, are included to explore their effectiveness in capturing local temporal patterns in time-series data. Table I shows the specific settings of these parameters.

D. Step 3: Learning Ensemble

After multiple deep learning models generate diverse recovery results in parallel, these outputs are integrated using a Random Forest-based ensemble learning framework. This approach combines the different recovered datasets by learning from their complementary features, producing a unified reconstruction. The ensemble model is trained on representative samples and then applied to aggregate the results, forming a consolidated output for further processing.

In the learning ensemble model, we employ an RF framework to recover datasets, including three-phase voltage. Specifically, we integrate matrix completion and deep learning methods into the RF model, treating them as distinct datasets within the ensemble. Let D_1 and D_2 represent the two datasets, each containing three data streams (three-phase voltage), resulting in $D_{\text{total}} = D_1 \cup D_2$ consisting of twelve data streams for recovery. The RF algorithm uses Bagging, which combines multiple decision trees, where the number of trees T is specified in Table I. Each decision tree t is trained

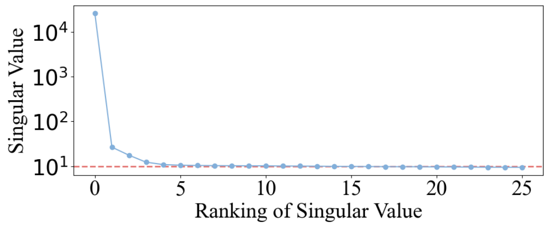

Fig. 3: Singular Value Example of Reshaped Voltage Matrix.

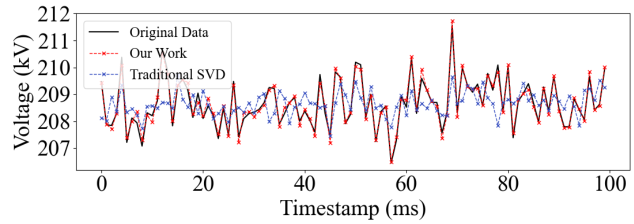

Fig. 4: Comparison of Waveform Completion Methods.

on a bootstrap sample $B_t \subset D_{\text{total}}$ sampled with replacement. The quality of splits is evaluated using the Gini index $G(t)$:

$$G(t) = 1 - \sum\nolimits_{k=1}^{K} p_k^2, \tag{9}$$

where p_k is the proportion of class k in the node. The tree is built recursively until stopping conditions are met, such as maximum depth $d_{\max}$ or minimum sample size $n_{\min}$.

Once the trees are built, results $\hat{y}_{\text{RF}}$ are aggregated: for regression tasks, the results are averaged, and for classification tasks, majority voting is applied:

$$\hat{y}_{\text{RF}} = \frac{1}{N_t} \sum\nolimits_{t=1}^{N_t} \hat{y}_t, \;\; \hat{y}_{\text{RF}} = \text{mode}(\{\hat{y}_t\}_{t=1}^{N_t}), \tag{10}$$

where N_t is the number of decision trees in the RF and $\hat{y}_t$ is the result from the t-th decision tree. This method enhances robustness and accuracy and reduces overfitting by leveraging the diversity of individual trees.

III. Case Studies and Results

In this section, we address the recovery of synchro-waveform data through the integration of different deep learning methods after SVD-based matrix completion, following the concept of ensemble learning. We also compare two types of data with varying levels of accuracy to validate the feasibility of our method.

A. *Data Selection*

Given the context of synchro-waveforms in power systems [20], we may have two closely spaced devices and measure the same physical quantity. The curves based on WMUs represent relatively clean data with minimal noise. These curves reflect typical characteristics observed in operational power systems, such as trim levels of noise and minor data disturbances. Such data supports our assumption and can serve as a reference to evaluate recovery performance under controlled conditions.

B. *Matrix Completion*

After data processing, we consider the two datasets from the two closely spaced devices: X_1' and X_2'.

The first step is to perform SVD on the matrix X_1'. This decomposes X_1' into three matrices: U_1, Σ_1, and V_1, where U_1 and V_1 are orthogonal matrices, and Σ_1 is a diagonal matrix containing the singular values. Next, the first m most significant singular values are selected from the diagonal matrix Σ_1. These values represent the most important components of the matrix. A new diagonal matrix Σ_m is then constructed, which contains only these m significant singular values. As shown in Fig. 3, especially the voltage part, we observe that the singular values decrease rapidly, with the value around the 20th singular value falling to around 10. Notably, these singular values are also smaller than 0.1% of the first singular value. This indicates that the majority of the information in the matrix is concentrated in the first few singular values, while the remaining ones contribute marginally to the overall structure of the data. As a result, for the subsequent recovery process, we select the first 20 singular values, as they capture the most significant components of the data. Consequently, for the subsequent recovery process, we select the first 20 singular values. After this, matrix completion is performed on X_2' using a low-rank approximation technique. By leveraging the significant singular values derived from X_1', the matrix completion process effectively recovers the primary structure of X_2'. Finally, the recovered data matrix X_2^{recovery} is obtained by performing matrix multiplication using the matrices from the matrix completion process. The resulting matrix provides an approximation of the original matrix X_2', which has been recovered through low-rank approximation.

Remark: Compared to the traditional matrix completion method (solely relying on the local measurement, without any help from the neighbor measurements), the presence of missing values introduces errors in the matrix completion process. As shown in Fig. 4, the traditional SVD method performs worse than our approach due to the presence of missing values, which adversely affect the accuracy of matrix completion. Missing data points cause the conventional SVD to produce biased singular vectors, leading to higher reconstruction errors. By leveraging data from two spatially close WMU devices, our method exploits the inherent spatial correlation to better estimate the missing entries, thereby significantly improving the completion performance.

C. *The Necessity of Learning Ensemble*

In this part, we investigate the performance of several learning models for synchro-waveform recovery. We evaluate the models using three standard metrics, Mean Absolute Error (MAE), Mean Square Error (MSE), and R-squared (R^2), across three scenarios: perfect, missing data, and noisy data. Then, after basic recovery with SVD-based matrix completion, we introduce the learning ensemble with the LSTM, GRU, and CNN models. The results are summarized in Table II.

1) Perfect Environment: The numerical evaluation under perfect environmental conditions demonstrates the superior performance of the proposed method compared to conventional learning ensemble techniques. As shown in Table II,

TABLE II: Performance Comparison of Different Learning Ensemble Methods

Matrix Completion	**Learning Method**	**Perfect Environment**				**Under Missing Data Conditions**				**Under Noisy Conditions**			
		MSE	MAE	R^2	T	MSE	MAE	R^2	T	MSE	MAE	R^2	T
None	GRU	8,831.39	43.51	0.96	6'47"	14,504.21	78.51	0.77	1'42"	10,221.19	58.35	0.90	6'26"
	CNN	14,457.31	71.96	0.74	**45"**	18,329.64	90.95	0.65	**12"**	15,047.90	86.73	0.67	**58"**
	LSTM	7,444.97	38.76	0.98	4'40"	12,925.40	61.22	0.91	1'09"	9,727.65	47.58	0.91	5'30"
	Our Work	539.78	10.61	0.99	30'09"	10,237.92	48.53	0.95	25'59"	8,049.91	40.74	0.93	33'30"
SVD	GRU	4,265.72	28.21	0.98	7'05"	8,375.99	45.19	0.88	1'59"	6,207.90	39.61	0.92	7'01"
	CNN	8,684.21	43.07	0.86	1'01"	14,547.34	70.73	0.81	27"	8,517.28	51.25	0.81	1'17"
	LSTM	3,266.61	20.81	0.99	4'45"	7,861.53	38.06	0.92	1'26"	5,744.61	28.74	0.93	5'54"
	Our Work	**422.73**	**9.50**	**0.99**	31'62"	**4,349.91**	**29.74**	**0.85**	27'28"	**2,913.19**	**20.27**	**0.96**	35'01"

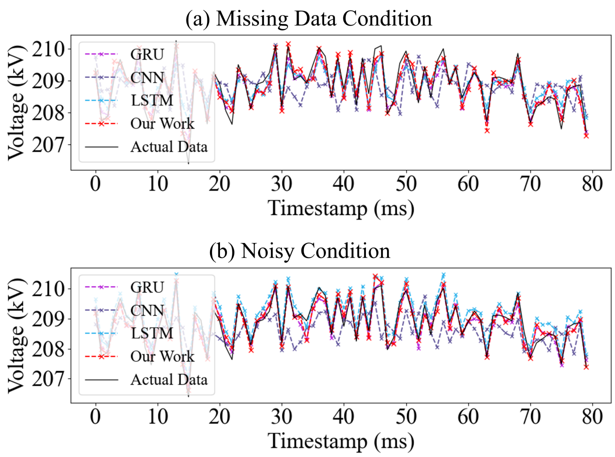

Fig. 5: Performance Comparison of Different Models under Missing Data and Noisy Conditions after Noisy Reduction.

the proposed approach achieves the lowest MSE and MAE across all noise reduction methods. Specifically, when no noise reduction is applied, the proposed method attains an MSE of 539.78 and an MAE of 10.61, significantly outperforming GRU, CNN, and LSTM models, which exhibit notably higher error values. Similarly, under the SVD noise reduction framework, the proposed method further reduces the MSE to 422.73 and MAE to 9.50, surpassing other recovery methods with clear margins. Additionally, the coefficient of determination values consistently approach 0.99 for the proposed method, indicating an excellent fit to the true waveform data and a highly reliable recovery performance.

2) Missing Data Conditions: To evaluate the performance of data recovery methods under missing data conditions, in WMU1, we also have a few cycles of data due to missing data. We observe significant differences in each model's ability to handle missing values. As shown in Table II and Fig. 5(a), without matrix completion, our method achieves an MSE of 10,237.92 and MAE of 48.53, outperforming GRU, CNN, and LSTM significantly under the same conditions, indicating its effectiveness in recovering missing data.

After applying matrix completion, recovery performance across all models improves significantly under missing data conditions. The proposed learning ensemble method demonstrates superior robustness and accuracy compared to individual models. Without matrix completion, the ensemble method already outperforms GRU, CNN, and LSTM, achieving an MSE of 10,237.92, an MAE of 48.53, and an R^2 of 0.95, indicating its effectiveness in handling incomplete waveform information. These results highlight the robustness of the proposed learning ensemble approach in recovering missing waveform data under challenging conditions, surpassing both standalone models and conventional recovery methods enhanced with matrix completion. The visual comparisons in Fig. 5(a) further confirm that the ensemble method combined with SVD achieves the most accurate reconstruction, outperforming individual models in scenarios with missing data.

3) Noisy Conditions: Under noisy conditions, where Gaussian noise was artificially added to the original data, the proposed learning ensemble method is designed to effectively address the challenges posed by such complex noise scenarios. Unlike the SVD noise reduction technique, which primarily targets inherent noise in the data, our approach combines three deep learning models to enhance robustness against both intrinsic and externally introduced noise. According to Table II and Fig. 5(b), without noise reduction, our ensemble method achieves an MSE of 8,049.91 and an MAE of 40.74, significantly outperforming GRU, CNN, and LSTM, which show higher error metrics under the same noisy conditions.

After matrix completion, the ensemble's performance improves further, reducing the MSE to 2,913.19 and the MAE to 20.27. This indicates that the complementary use of SVD for completion and the ensemble's integrated deep learning models effectively mitigate the impact of both intrinsic and externally added noise, resulting in more accurate waveform recovery. In contrast, individual deep learning models show higher error rates and less consistent recovery performance under the same noisy conditions, demonstrating the advantage of the learning ensemble approach in handling noise scenarios.

4) Summary of Learning Ensemble Model Performance: Above all, we have evaluated the performance of various deep learning models and learning ensemble methods in data recovery tasks under different conditions. The results show that traditional models struggle significantly when faced with missing data or noisy environments. However, the integration of learning ensembles demonstrated superior adaptability and recovery capabilities in these challenging scenarios. Similarly, the SVD-based matrix completion as a denoising method also significantly improved the experimental results. Deep learning methods like GRU, CNN, and LSTM, as well as learning ensemble methods, further enhanced recovery accuracy.

Overall, the learning ensemble after matrix completion stands out as a highly reliable and resilient approach for data recovery, especially in missing data and noisy environments.

D. Time Efficiency of Various Data Recovery Methods

In studying data recovery under different scenarios, both time efficiency and accuracy are crucial. Table II evaluates several machine learning methods, including GRU, CNN, LSTM, and our work with matrix completion, comparing their time consumption for data recovery in different environments. Integrating matrix completion with these methods results in a slight increase in computational time, which is a small trade-off considering the improvements in recovery accuracy. The learning ensemble method with matrix completion, while offering superior recovery accuracy, requires more computational time than others. Although not the fastest, the modest increase in time is justified by the significant improvements in recovery accuracy, particularly when handling missing and noisy data. In conclusion, while other methods may be more time-efficient, our work stands out as the best choice, offering enhanced recovery accuracy while maintaining computational efficiency. This approach is especially beneficial for handling noisy or incomplete data, where the slight increase in time consumption is well justified by the improvement in the quality of the recovered data.

IV. Conclusions

This study investigates synchro-waveform data recovery using a learning ensemble methodology that combines the strengths of three different deep learning networks after matrix completion by SVD. The proposed method helps to address the inherent challenges of recovering time-series data, especially in the presence of noise and missing values, commonly encountered in real-world power system monitoring scenarios. Specifically, the ensemble model demonstrated a notable reduction in MSE and MAE while maintaining an R^2 value, suggesting an effective recovery of the missing data under different conditions. When evaluated under more challenging conditions, such as the introduction of Gaussian noise and random data removal, our approach continued to demonstrate remarkable resilience. The findings underscore the potential of integrating multiple learning techniques to enhance the reliability and accuracy of synchro-waveform data recovery. The more accurate and complete synchro-waveform could enable demand-side management via customized pricing [21], false data injection attack detection [22], and distributed user profiling [23].

References

[1] H. Mohsenian-Rad and W. Xu. Synchro-waveforms: A window to the future of power systems data analytics. *in the IEEE Power and Energy Magazine*, 21(5), Sept./Oct. 2023.

[2] C. Halliday. Visualization of real-time system dynamics using enhanced monitoring (VISOR). *SP Energy Networks, Final Report*, May 2018.

[3] A. Shirsat, H. Sun, K. J. Kim, J. Guo, and D. Nikovski. 1 convednet: A convolutional energy disaggregation network using continuous point-on-wave measurements. In *in Proc. of the 2022 IEEE Power & Energy Society General Meeting (PESGM)*, Denver, CO, USA, Jul. 2022. IEEE.

[4] H. Mohsenian-Rad, Kezunovic M, and F. Rahmatian. Synchro-waveforms in wide-area monitoring, control, and protection: Real-world examples and future opportunities. *in the IEEE Power and Energy Magazine*, 23(1), January 2025.

[5] N. Ehsani, F. Ahmadi-Gorjayi, Z.-J. Ye, A. McEachern, and H. Mohsenian-Rad. Sub-cycle Event Detection and Characterization in Continuous Streaming of Synchro-waveforms: An Experiment Based on GridSweep Measurements. In *in the Proc. of the IEEE North American Power Symposium*, Asheville, NC, USA, Qct. 2023.

[6] F. A. Gorjayi and H. Mohsenian-Rad. Data-Driven Models for Sub-Cycle Dynamic Response of Inverter-Based Resources Using WMU Measurements. *in the IEEE Trans. on Smart Grid*, 14(5), Sep. 2023.

[7] H. Mohsenzadeh-Yazdi, C. Li, and H. Mohsenian-Rad. Ibr responses during a real-world system-wide disturbance: Synchro-waveform data analysis, pattern classification, and engineering implications. *IEEE Trans. on Smart Grid*, pages 1–4, April 2025.

[8] H. Mohsenzadeh-Yazdi, F. Ahmadi-Gorjayi, and H. Mohsenian-Rad. Data-driven modeling of sub-cycle dynamics of inverter-based resources using real-world synchro-waveform measurements. *IEEE Trans. on Power Delivery*, pages 1–13, May 2025.

[9] H. Mohsenian-Rad, A. Shahsavari, and M. Majidi. Analysis of power quality events for wildfire monitoring: Lessons learned from a california wildfire. In *in the Proc. of the 2023 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT-LA)*. IEEE, Nov. 2023.

[10] M. Izadi and H. Mohsenian-Rad. Synchronous waveform measurements to locate transient events and incipient faults in power distribution networks. *in the IEEE Trans. on Smart Grid*, 12(5), Jan./Feb. 2021.

[11] J. Zheng, D. W. Gao, and L. Lin. Smart meters in smart grid: An overview. In *in the Proc. of the 2013 IEEE green technologies conference (GreenTech)*, Denver, CO, USA, Apr. 2013. IEEE.

[12] R. E. Wilson. PMUs [phasor measurement units]. *in the IEEE Potentials*, 13(2), Apr. 1994.

[13] P. Gao, M. Wang, S. G. Ghiocel, J. Chow, B. Fardanesh, and G. Stefopoulos. Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements. *in the IEEE Trans. on Power Systems*, 31(2), Mar. 2016.

[14] P. Gao, M. Wang, J. Chow, M. Berger, and L. Seversky. Missing data recovery for high-dimensional signals with nonlinear low-dimensional structures. *in the IEEE Trans. on Signal Processing*, 65(20), Oct. 2017.

[15] T. Wu, Y.-J. A. Zhang, Y. Liu, W. C. Lau, and H. Xu. Missing data recovery in large power systems using network embedding. *in the IEEE Trans. on Smart Grid*, 12(1), Jan. 2021.

[16] Y. Zhang, Y. Xu, R. Zhang, and Z. Y. Dong. A missing-data tolerant method for data-driven short-term voltage stability assessment of power systems. *in the IEEE Trans. on Smart Grid*, 10(5), Sept. 2019.

[17] H. Mohsenian-Rad. *Smart Grid Sensors: Principles and Applications*. Cambridge University Press, 2022.

[18] W. Sun, Z. Tian, and C. Wu. Enhancing probabilistic peak load forecasting with fuzzy information granulation and deep learning. In *2024 3rd International Conference on Power Systems and Electrical Technology (PSET)*, pages 810–815. IEEE, Aug. 2024.

[19] X. Wang, Y. Qiao, D. Wu, C. Wu, and F. Wang. Cluster based heterogeneous federated foundation model adaptation and fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 21269–21277, Feb. 2025.

[20] Hussain M Mustafa, Vasavi Sivaramakrishnan, Vignesh VG Krishnan, and Anurag Srivastava. Realistic synchrophasor data generation for anomaly detection using cyber-power testbed. In *2024 56th North American Power Symposium (NAPS)*, pages 1–6, 2024.

[21] J. Cui and C. Wu. Robust long-term rate design via matrix completion. In *2021 5th International Conference on Smart Grid and Smart Cities (ICSGSC)*, pages 141–145. IEEE, Jun. 2021.

[22] Q. Huang and C. Wu. Boosting false data injection attack detection with structural knowledge. In *2022 American Control Conference (ACC)*, pages 4595–4600. IEEE, Jun. 2022.

[23] Q. Huang, W. Jiang, J. Shi, C. Wu, D. Wang, and Z. Han. Federated shift-invariant dictionary learning enabled distributed user profiling. *IEEE Transactions on Power Systems*, 39(2):4164–4178, 2023.