Colored Noise Minimisation using machine learning algorithms

A report submitted in partial fulfilment of the requirements for the award of the degree of

Bachelor of Technology

In

Electrical Engineering

By

Annadanam Lakshmi Haritha 120EE1086

Kandimalla Vishwas Reddy 120EE0695

Akula Hemanth Kalyan Kumar 120EE0452

based on the research carried out under the supervision of

Prof. Shekha Rai

DEPARTMENT OF ELECTRICAL ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA

Colored Noise Minimisation using Machine Learning

Project Supervisor: Prof. Shekha Rai

Akula Hemanth Kalyan Kumar[120EE0452]

Kandimalla Vishwas Reddy[120EE0695]

Annadanam Lakshmi Haritha [120EE1086]

Department of Electrical Engineering

National Institute of Technology, Rourkela, India

120ee0452@nitrkl.ac.in, 120ee0695@nitrkl.ac.in, 120ee1086@nitrkl.ac.in

Abstract — This paper addresses the pervasive challenge of data quality degradation in on-site Phasor Measurement Units (PMUs) resulting from the application of anti-aliasing filters, leading to the introduction of colored noise. The interference caused by colored noise poses a significant hurdle in accurately predicting results, thereby affecting the overall reliability of PMU measurements. To mitigate this issue, we propose a novel approach that integrates Independent Component Analysis (ICA), Sparse Principal Component Analysis (SPCA), and Non-negative Matrix Factorization (NMF) within an ensemble stacking framework with the XgBoost algorithm.

Keywords— PMU, Anti-aliasing filter, ICA, SPCA, NMF, XgBoost, Ensemble stacking

1. INTRODUCTION

Modern power systems heavily rely on Phasor Measurement Units (PMUs) for real-time monitoring and control. These devices play a crucial role in capturing and analysing the dynamic behaviour of the power grid. However, the presence of colored noise, induced by anti-aliasing filters employed in the PMU sampling process, significantly compromises the accuracy of the collected data. The estimation bias introduced by colored noise hampers the precision of predicted modes, impacting the effectiveness of critical applications such as disturbance detection and system stability analysis.

To address these challenges, our research focuses on the development of an advanced methodology for colored noise minimization in on-site PMUs. By combining the strengths of Independent Component Analysis (ICA), Sparse Principal Component Analysis (SPCA), and Non-negative Matrix Factorization (NMF) in an ensemble stacking framework, complemented by the powerful XgBoost algorithm, we aim to enhance the accuracy of mode predictions. This ensemble approach is designed to provide a versatile solution, accommodating diverse input data types commonly encountered in real-world PMU applications.

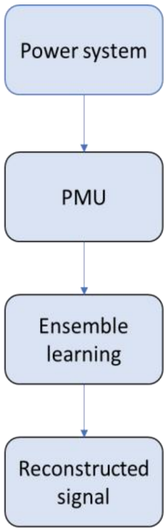

Fig 1: Block diagram of proposed method for coloured noise minimisation

In the subsequent sections of this paper, we delve into the theoretical foundations of the proposed method, detailing the integration of ICA, SPCA, and NMF within the ensemble stacking framework. We present experimental results, demonstrating the effectiveness of our approach in reducing colored noise and improving the precision of reconstructed signals. The compatibility of our method with various input data types further underscores its practical applicability in the field of power system monitoring and control.

2. LITERATURE REVIEW

Title of Paper	Name of the Journal/Publisher/Conference and Publication year	Review
A Modified TLS-ESPRIT Based Method for Low Frequency Mode Identification in Power Systems Utilizing Synchro phasor Measurement	IEEE Transactions on Power Systems, vol. 26, no. 2, pp. 719-727, May 2011	This report provides a review on how colored noise and white noise are being added to a clean signal and different algorithms to filter out high frequency components at the PMUs output.
Synchronized measurement based estimation of inter-area electromechanical modes using Ibrahim time domain method.	IEEE Transactions on Power Systems Res 118:85-95	This report provides how the Weiner filter plays a significant role in noise suppression and filtering of the clean signal.
An efficient K-SVD based Algorithm for detection of Oscillatory mode from ambient data for synchro phasor application	2021 IEEE 18th India Council International Conference (INDICON), 2021	Got a review on the effective application of K-SVD for the identifying and removing the oscillatory modes from ambient data and how ambient data is obtained from PMU.

Title of Paper	Year of Publication	Review
Using sparsity to estimate oscillatory mode from ambient data	Indian Academy of Sciences,2018	Got a review report on how it creates an effective dictionary from an input signal and then taking the representation coefficients that most accurately represent the noise-free signal from it.
Online Transfer Function Estimation and Control Design Using Ambient Synchro phasor Measurements	IEEE Transactions on Power Systems,2022	In this report, a methodology for identifying and designing adaptive controls for low-frequency oscillations in power systems is given and the design is based upon ambient synchro phasor measurements.
An XgBoost based method for identifying electromechanical oscillations from ambient measurements using WAMS	IEEE Transactions on Power Systems,2023	This report provides a review on how we are minimizing colored noise using XgBoost algorithm which enables more reliable and precise estimation.

3. Power System Dynamics

Power systems operate as intricate networks of generators, transformers, and transmission lines, where dynamic interactions between various components govern system behaviour. The stability and reliability of the power grid are paramount for ensuring a consistent and efficient supply of electricity.

4. Phasor Measurement Units (PMUs)

PMUs are advanced devices strategically placed within the power grid to capture precise information about voltage and current waveforms. Unlike conventional measurement devices, PMUs provide synchronized measurements at a high sampling rate, enabling the real-time monitoring of power system dynamics with unparalleled accuracy.

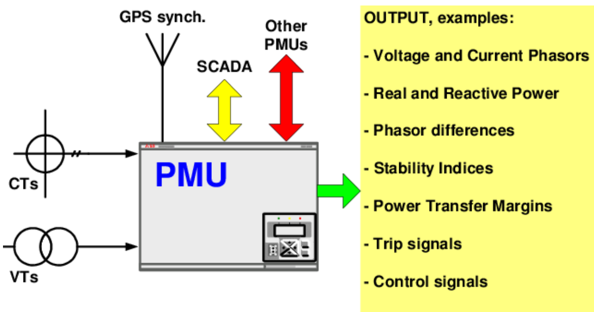

Figure 2: Phasor measurement units (PMU) terminal interfaces and outputs.

4.1 Data Flow from Power System to PMU:

4.1.1 Phasor Measurements:

Capturing Time-Synchronized Data:

PMUs capture voltage and current phasors, representing the amplitude and phase angle of sinusoidal waveforms. These measurements are timestamped with high precision, facilitating the synchronization of data across different locations within the power system.

4.1.2 Synchro phasors and Time Stamps:
Temporal Precision in Measurements:
Synchro phasors, the time-stamped phasor measurements, play a pivotal role in understanding the temporal dynamics of the power system. The precise time stamps associated with these measurements enable the reconstruction of events and disturbances with a high degree of accuracy.

4.1.3 Data Transmission to Centralized Systems:
Real-Time Communication:
The collected synchro phasor data is transmitted to centralized systems for real-time analysis and decision-making. This data flow is crucial for monitoring the health of the power system, detecting disturbances, and implementing corrective measures promptly.

5. Colored Noise Challenges

Anti-Aliasing Filters and Data Distortion:
Despite the advancements afforded by PMUs, the application of anti-aliasing filters introduces colored noise, complicating the accurate interpretation of synchro phasor data. This section highlights the challenges posed by colored noise and sets the stage for the proposed ensemble stacking approach to minimize its impact.

6. Proposed Methodology

Integrating ICA, SPCA, NMF, and XgBoost:
This section introduces the novel approach to colored noise minimization, incorporating Independent Component Analysis (ICA), Sparse Principal Component Analysis (SPCA), and Non-negative Matrix Factorization (NMF) within an ensemble stacking framework. The utilization of the XgBoost algorithm further enhances the effectiveness of the proposed methodology.

6.1 Independent Component Analysis:
Independent Component Analysis (ICA) is a computational technique used to separate a multivariate signal into additive, statistically independent components. The underlying assumption of ICA is that the observed signals are linear combinations of independent source signals. The goal of ICA is to identify these independent components and their mixing coefficients. The algorithm works iteratively to achieve this separation, often leveraging statistical properties of the data.

6.1.1 Data Preprocessing:
Input Data: Assume we have a dataset represented by a matrix X, where each column corresponds to a different observed signal (e.g., sensor readings, time series data).
Mean-Centring: The data is typically mean-centred to remove any DC offset.

6.1.2 Whitening the Data:
Covariance Matrix Calculation: Compute the covariance matrix of the mean-cantered data.
Eigenvalue Decomposition: Obtain the eigenvalues and eigenvectors of the covariance matrix.
Whitening Transformation: Use the eigenvectors to transform the data to a new space where the covariance matrix is diagonal.

6.1.3 Decorrelation (Rotation):

Orthogonal Transformation: Perform an orthogonal transformation on the whitened data to maximize statistical independence.

Fixed-Point Iteration: Iterate the fixed-point algorithm to find the rotation matrix that maximizes the non- Gaussianity of the components.

Negentropy Maximization: ICA often involves maximizing the negentropy (a measure of non-Gaussianity) of the estimated sources.

6.1.4 Component Extraction:

Final Independent Components: The independent components are obtained by applying the determined rotation matrix to the whitened data.

6.1.5 Postprocessing (Optional):

Scaling: The scaling of the independent components is often arbitrary. Postprocess the components to achieve desired scaling or unit variance.

6.1.6 Iterative Refinement (Optional):

Repeat Steps 2-5: In some cases, particularly when dealing with non-linear mixtures or complex source distributions, the whitening and decorrelation steps may need to be repeated to refine the separation.

6.1.7 Output:

Independent Components: The final output of the ICA algorithm is a set of statistically independent components that, when combined linearly, reconstruct the observed signals.

Independent Component Analysis works by transforming the observed signals into a space where the components are statistically independent and then extracting these independent components through a rotation process. The success of the algorithm relies on meeting the assumptions of non-Gaussianity and statistical independence of the sources.

6.2 Sparse Principal Component Analysis (Sparse PCA):

Sparse Principal Component Analysis (Sparse PCA) is an advanced data analysis technique that extends the classical Principal Component Analysis (PCA) by incorporating sparsity constraints into the decomposition process. This methodology is particularly beneficial when dealing with high-dimensional data and aiming for a more interpretable representation of underlying components.

6.2.1 Signal Generation and Colored Noise

The provided code begins by loading a true signal (x_sig) and introducing colored noise using a Butterworth filter. The incorporation of colored noise simulates real-world scenarios where signals are often contaminated by various forms of interference, making denoising techniques crucial for accurate analysis.

6.2.2 Applying Sparse PCA for Noise Reduction

The core of the code involves applying Sparse PCA to the signal corrupted with colored noise. The key parameter, alpha, controls the sparsity of the resulting components. A higher alpha promotes sparsity, emphasizing the extraction of essential features while mitigating the impact of noise.

6.2.3 Visualization of Results

The denoising process is visually represented through a series of plots. These plots compare the true signal, the signal with colored noise, and the reconstructed clear signal obtained through Sparse PCA. This visual exploration provides an intuitive understanding of how Sparse PCA effectively isolates meaningful components from noisy observations.

6.2.5 Quantifying Performance: Reconstruction Error

To quantitatively assess the performance of Sparse PCA, a reconstruction error is calculated. The reconstruction error is computed as the normalized Euclidean distance between the true signal and the denoised signal. This metric serves as a numerical indicator of the algorithm's ability to capture essential signal features while suppressing noise.

6.2.6 Implications

Sparse PCA emerges as a valuable tool for noise reduction in signal processing applications. Its application to the provided scenario demonstrates its effectiveness in enhancing signal quality and extracting meaningful information from noisy observations. The versatility of Sparse PCA makes it applicable across various domains, offering a powerful solution for data analysis challenges in fields such as image processing, finance, and biomedical signal analysis.

6.3 Non-Negative Matrix Factorization (NMF) :

Non-negative matrix factorization (NMF) is a technique commonly used for dimensionality reduction and feature extraction. In the context of signal processing, it can be applied to separate a signal into additive components.

6.3.1 Signal Representation:

Input Matrix with Colored Noise: Assume a non-negative data matrix V representing observed signals corrupted by colored noise. Each row corresponds to different observations, while each column represents features or variables with non-negative values.

6.3.2 NMF Decomposition for Noise Reduction:

Factorization Process: NMF decomposes the input matrix X into two lower-dimensional non-negative matrices, W (basis matrix) and H (coefficient matrix).

$$X \approx W \bullet H$$

Non-Negativity Constraints: The non-negativity constraints in both W and H ensure that the resulting factors represent only positive contributions, facilitating the extraction of meaningful components while mitigating the impact of colored noise.

6.3.3 Objective Function for Colored Noise Minimization:

Reconstruction Error Minimization: NMF aims to minimize the reconstruction error. This process is really good at reducing the impact of colored noise in the data. It helps make a more precise estimate of the actual signal that's hidden beneath the noise. So, by doing this, we get a clearer picture of the true information we're interested in. We need to choose rank of the matrix effectively. Rank in Non-negative Matrix Factorization (NMF) determines the number of components used to approximate data. Higher ranks allow capturing more complex patterns but may risk overfitting. Lower ranks simplify data representation, potentially losing details. Choosing an optimal rank balances model complexity with the ability to represent underlying structures effectively.

6.3.4 Interpretability and Noise Isolation:

Parts-Based Representation: NMF provides a parts-based representation of the signal, enhancing interpretability by isolating essential components while suppressing the effects of colored noise.

Colored Noise Reduction: The decomposition process inherently filters out undesirable noise components, allowing for a clearer identification and understanding of the intrinsic features in the data.

6.3.5 Applications in Colored Noise Scenarios:

Audio and Image Denoising: NMF can be applied in denoising scenarios, such as separating mixed audio sources or enhancing images corrupted by colored noise.

Signal Processing in Variable Environments: NMF proves valuable in scenarios where signals are influenced by varying environmental conditions, making it resilient to colored noise variations.

6.3.6 Output:

The output of the NMF algorithm is a factorized representation of the input matrix, consisting of the basis matrix (W) and coefficient matrix (H). This representation effectively minimizes the influence of colored noise, providing a clearer and interpretable depiction of the underlying signal.

7. Ensemble stacking

Ensemble stacking, also known as model stacking or stacking, is an ensemble learning technique that combines predictions from multiple diverse models to improve the overall predictive performance. It leverages the strengths of different algorithms and aims to mitigate their individual weaknesses.

7.1 Base Models (ICA, SPCA, NMF):

Ensemble stacking typically involves using diverse base models. In this case, the base models are derived from Independent Component Analysis (ICA), Sparse Principal Component Analysis (SPCA), and Non-negative Matrix Factorization (NMF), each contributing unique perspectives on the data.

7.2 Advantages:

Improved Robustness: Ensemble stacking helps mitigate overfitting and generalization issues by combining the strengths of different models.

Enhanced Predictive Power: The ensemble can achieve better predictive performance compared to individual models, especially when the base models capture different aspects of the underlying patterns in the data.

7.3 Ensemble Stacking with XgBoost:

Due to the limitations of input data types with XgBoost we have stacked it with different methods but there are few input datasets where XgBoost has its own advantages so along with ICA , SPCA, and NMF we have used XgBoost in final ensemble stacking to satisfy all the datasets.

Ensemble stacking combines the outputs of diverse models, including ICA, SPCA, NMF, and XgBoost. This integrated approach leverages the strengths of each component to achieve enhanced performance and robustness in handling complex data patterns.

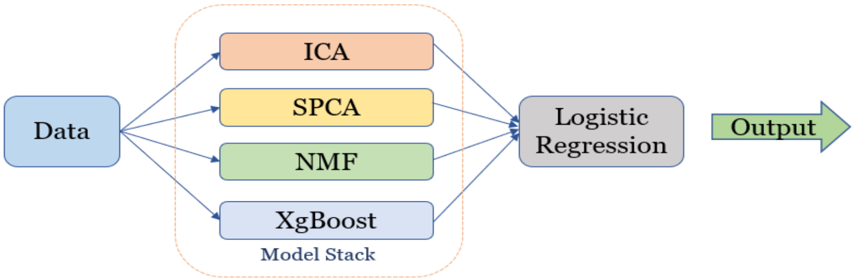

Figure 3 : Stacked model of Machine learning algorithms

8. Results and Discussion

This test signal corresponding to ensemble stacking was simulated with a sampling frequency of 10 Hz.

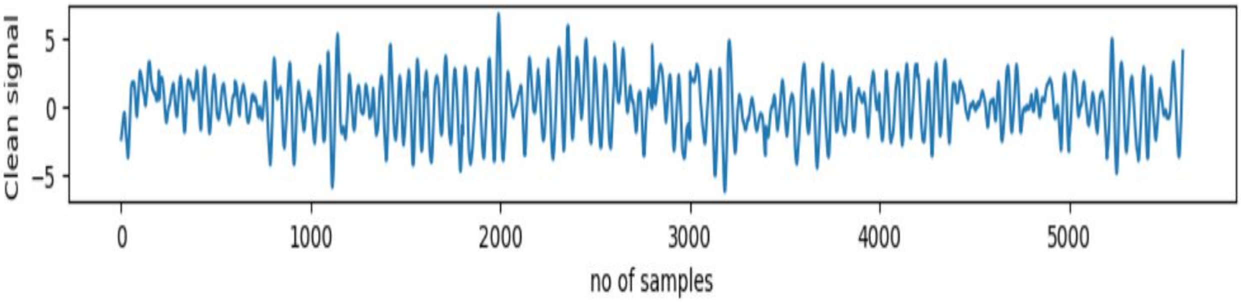

Figure 4 : Clean signal

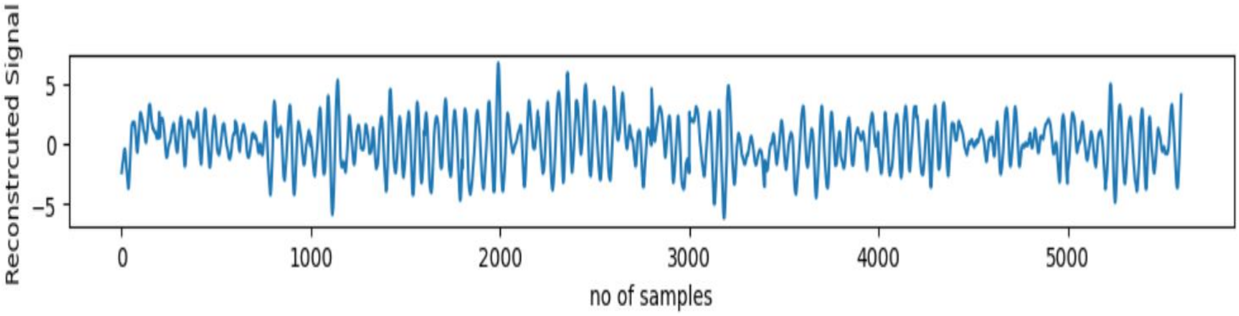

Figure 5 : Reconstructed signal with proposed method

Method	MSE
ICA	0.9999
SPCA	0.2161
NMF	0.7115
XgBoost	0.8457
Stacking	0.0017

9. Future agendas

The XgBoost algorithm was convenient but it was only allowed for a few data types. So the proposed ensemble stacking with ICA, SPCA, NMF and XgBoost will allow all the data types. In practical scenarios for testing 2 area networks and real time signals we may do the necessary modifications in future

10. Conclusion

In summary, our research introduced a comprehensive approach, employing a diverse suite of machine learning algorithms to enhance the versatility of our methodology in the context of Phasor Measurement Units (PMUs) and diverse data types. Significantly addressing the challenge of colored noise, our method showcased a remarkable reduction in its impact, facilitating more accurate and reliable mode estimations derived from ambient data. The comparative analysis of algorithms such as Independent Component Analysis (ICA), Sparse Principal Component Analysis (SPCA), Non-negative Matrix Factorization (NMF), and XgBoost unveiled their unique strengths and contributions. To optimize our solution, we incorporated ensemble stacking, a synergistic technique that effectively amalgamates the strengths of diverse sub-models, resulting in superior predictive accuracy. This approach not only mitigated the influence of colored noise but also preserved signal characteristics, yielding highly precise outcomes. In conclusion, the proposed ensemble stacking scheme emerged as a robust and efficient solution, affirming its effectiveness in the mitigation of colored noise for on-site PMUs.

11. Acknowledgement

We express our sincere appreciation for the invaluable direction provided by our supervisor, Prof. Shekha Rai ma'am, the cooperative endeavours of our research team, and the backing of our institution. We extend our heartfelt gratitude to the scholarly and scientific community for their valuable contributions. In combination, these constituents have enhanced our investigation on colored noise minimization using machine learning algorithms.

12. References:

1. Kundur P 1994 Power system stability and control. New York, McGraw-Hill
2. P. Tripathy, S. C. Srivastava and S. N. Singh, "A Modified TLS-ESPRIT-Based Method for Low-Frequency Mode Identification in Power Systems Utilizing Synchro phasor Measurements," in *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 719-727, May 2011, doi: 10.1109/TPWRS.2010.2055901.
3. S. Rai, P. Tripathy, and S. Nayak, “A robust TLS-ESPIRIT method using covariance approach for identification of low-frequency oscillatory mode in power systems,” in Power Systems Conference (NPSC), 2014 Eighteenth National, Dec 2014, pp. 1–6.
4. Zheng, Shiqiang & Shuangyi, Zhang & Song, Youyi & Lin, Zhizhe & Wang, Fei & Zhou, Teng. (2020). A Noise-eliminated Gradient Boosting Model for Short-term Traffic Flow Forecasting. 10.1109/ICDH51081.2020.00036.
5. S. Makeig, A.J. Bell, T.-P. Jung, and T.-J. Sejnowski. Independent component analysis of electro-encephalographic data. In Advances in Neural Information Processing Systems 8, pp. 145-151. MIT Press, 1996.
6. Yu-Xiong Wang ,Yu-Jin Zhang, “Nonnegative Matrix Factorization: A Comprehensive Review” in *IEEE Transactions on Knowledge and data Engineering (Volume : 25 , Issue : 6, June 2013). d*oi :10.1109/TKDE.2012.51
7. D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.