Enhancing Smart Grid Security with SHA-SARIMAX: Identifying and Restoring Corrupted Files from FDIA

Anand Srivatsa, T Ananthapadmanabha, Likith Kumar M V, Suma A P

***Abstract*—The smart grid is a promising solution for modernizing electrical power systems in a digital world. However, cybersecurity threats pose significant risks to the grid's integrity and resilience. False Data Injection Attacks (FDIA) have shown serious risks to disrupt the power system infrastructure. SHA-SARIMAX is a novel framework designed to address these challenges by combining distinct elements of Secure Hash Algorithm-256 (SHA256) and SARIMAX. Cryptographic hashing techniques and advanced modelling of SHA-SARIMAX actively detect and recover corrupted values in time series datasets affected by FDIA. This approach detects and recovers corrupted values in time series datasets affected by FDIA, leveraging cryptographic hashing techniques and advanced modelling. The proposed framework achieves a 99.90% accuracy rate from the original corrupted data. Thus, the SHA-SARIMAX demonstrates its applicability and suitability for implementation within smart grids, specifically for addressing the recovery of original data from the corrupted data affected by FDIA.**

***Index Terms*—SHA256, FDIA, SARIMAX, Regeneration, Cyber security**

I. Introduction

A smart grid is a power distribution system which is sophisticated that facilitates the two-way transmission of energy and data using digital communication technology. Its purpose is to modernize traditional electricity grids by incorporating information and communication technologies (ICTs). In contrast to traditional grids, smart grids facilitate the transmission of large data volumes using digital means, eliminating the limitations of high-voltage transmission cables [1]. Smart grid deployment encompasses a range of electrical components, including transmission lines, transformers, and substations. It also facilitates efficient energy storage and exploitation of renewable power production and demand response. [2] [3]. Further, integrating digital communication in smart grids provides a wide range of benefits, such as improved reliability, efficient monitoring, and streamlined electricity transfer [4]. However, the digitalization of smart grids has given rise to considerations about data security, making smart grid security a key challenge. Preserving the integrity and efficiency of smart grid operations demands the critical protection of data exchanged between electrical and digital systems [5].

Among the numerous cybersecurity threats, False Data Injection (FDI) attacks have risen to prominence due to their potential to disrupt critical infrastructure [6]. These attacks involve the injection of false or manipulated data into the smart grid control systems or data streams, leading to incorrect decision-making and system instability. The 2015 Ukraine Blackout 2015 [7] was a cyberattack involving a sophisticated FDIA on the country's electricity grid. Adversaries manipulated sensor data to deceive the power system's control mechanisms, leading to widespread power outages. The attack demonstrated the potential vulnerability of critical infrastructure to cyber threats. This attack aims to cause the system operator to take control actions that can negatively impact the power system's physical or economic functioning [8]. The Texas Big Freeze [9] of 2021 significantly impacted the smart grid, exposing vulnerabilities of physical and cyber elements. The power grid failures disrupted power supply, data communication, and control systems, emphasizing the importance of enhancing smart grid security. [10] Dragonfly cyber-espionage campaign targeted global energy companies during 2016. They performed False data injection attacks to breach energy systems.

The various statistical techniques and models help to detect anomalies and irregularities in the time series data. These techniques focus on identifying data points that deviate significantly from the expected patterns, such as outliers or unexpected abnormalities in values [11]. Moreover, statistical analysis aids in demonstrating abnormal trends or sudden changes in the data, which might signify the presence of false or manipulated information. The main issue with statistical analysis for detecting False Data Injection Attacks (FDIA) is the potential for false truths and false lies. These false truths occur when legitimate data patterns are mistakenly alerted as sceptical, guiding to unwanted alerts and possible disruptions in grid operations [12].

On a different note, machine learning techniques have emerged as a powerful tool in recent years for identifying FDIA in smart grid networks due to their ability to estimate and identify anomalies in complex and dynamic data [13] [14] [15] [16]. Deployment of machine learning and deep learning-based models in smart grids encounters challenges due to the reliance on extensive and high-quality training datasets. Acquiring labelled data for both normal and attack situations in smart grids can be difficult, and the data

Manuscript received July 31, 2023; revised July 9, 2024.

Anand Srivatsa is Associate Professor in the Department of Electronics and Communication Engineering, The National Institute of Engineering, Mysuru, Karnataka, India. (Phone: 7406967186; e-mail:anand.srivatsa@nie.ac.in.; ORCID-https://orcid.org/0000-0002-2202-4076)

T Ananthapadmanabha is Director and Professor of Mysore University, School of Engineering, Manasagangotri Campus, Mysuru, Karnataka, India. (e-mail: drapn2015@gmail.com)

Likith Kumar M V is Associate Professor in Department of Electronics and Electrical Engineering, The National Institute of Engineering, Mysuru, Karnataka, India. (ORCID: 0000-0003-1235-5602, e-mail: likith@nie.ac.in)

Suma A P is Assistant Professor in Information Science Engineering at Mysore University, School of Engineering Manasagangotri Campus, Mysuru, Karnataka, India. (ORCID:https://orcid.org/0009-0008-6732-4591, e-mail: sumapadmanabh@gmail.com)

may be subject to noise, missing values, or inconsistencies, which can influence the model's effectiveness [17] [18]. Additionally, the real-time nature of smart grid operations demands prompt inference from the machine learning models. However, deploying computationally intensive deep learning models on resource-constrained devices or edge nodes may introduce delays, potentially compromising the system's responsiveness [19]. Researchers also utilized blockchain for the identification of potential weakness in the smart grid [20].

In summary, this research focuses on the need of promptly identifying and recovering from false data injection attacks (FDIA) in smart grids, in order to maintain the robustness and dependability of the grid. The study emphasizes the necessity of real-time datasets to detect FDIA by continuously monitoring grid parameters for anomalies indicating attacks. Consequently, we proposed a novel framework named SHA-SARIMAX which integrates cryptographic hashing techniques and advanced modelling to detect and recover corrupted values in time series datasets. The framework's effectiveness is tested using a real-time dataset generated from CESCOM, Mysuru, Karnataka power grid station.

II. Literature Survey

This section summarises the challenges, latest solutions, and ongoing investigations in the domain of cyber security for smart grid infrastructure.

The smart grid represents a transformative advancement in the energy sector, providing numerous advantages such as enhanced energy management, enhanced efficiency, and reduced environmental impact. However, the exponential increase in the interconnectivity and digitization within the smart grid infrastructure has opened up a new range of security challenges [21].

Tange et al. [22] examined the problems of state estimation and false data injection detection in a smart grid setting, where measurements are affected by colored Gaussian noise. The noise is modeled using an autoregressive process in order to assess the condition of power transmission networks. A spotter based on the Generalized Likelihood Ratio Test (GLRT) is designed to identify fake data injection assaults. However the colored Gaussian noise does not fully capture all real-world noise characteristics. Guangdou et al. [22] introduced a spatiotemporal detection method for identifying and assessing false data injection attacks. They utilized the cubature Kalman filter, while Gaussian process regression was employed to examine spatial correlations. These techniques were applied to monitor and record the dynamic properties of the state vector, enhancing the accuracy of false data attack detection. Nevertheless, many detection approaches need a significant amount of historical data for practicing or are strongly impacted by system factors, rendering them unsuitable for large-scale distribution systems [23].

Wang et al. [24] proposed a reliable two-tier detection system for FDIA known as Kalman Filter and Recurrent Neural Network (KFRNN). The first phase involves using the Kalman filter and RNN to forecast the state, proficiently capturing both linear and nonlinear characteristics of the data. In the second step, the results of two base learners are combined using the fully connected layer and back-propagation. However, the strategy only concentrates on the fluctuating temporal variations in the power grid's condition [25].

Huang et al. [26] introduced the idea of matrix separation and a computational approach for identifying FDIA (Fault Detection, Isolation, and Accommodation). The proposed approach investigates the low-rank feature of the non-invasive computation matrix. Additionally, the attack matrix's structural sparsity characteristic is used to differentiate between genuine and altered data [27].

Aladag et al. [28] conducted data corruption attacks on the MNIST character detection dataset, a widely used benchmark in the field. A generative model, specifically an AutoEncoder, is employed to build more secure classification models using the poisoned dataset. Nevertheless, more examination is necessary to verify the model's efficacy in real-life situations and when confronted with increasingly advanced forms of assaults. Truong et al. [29] highlight the potential risk of backdoor poisoning, which leads to the creation of incorrect machine learning models that act as causative attacks, inducing specific errors not only during training but also in the model's functioning. The study identifies the adversarial challenge of distinguishing between the roles of developers and adversaries when corrupting the dataset.

Kewei et al. [30] proposed a secure two-phase authentication protocol for secured measurements from isolated intelligent grid devices. The framework employed an intelligent reader as a connector between the lone device and the advanced grid cloud, taking into account the physical limitations of the devices. The research incorporates security analysis to showcase the framework's efficacy in countering common threats and to verify its compliance with hardware limitations. However, the method suffers from a lack of security and real-time implementation [31]. Yuancheng et al. [32] included a physical model into their study, which successfully obtains accurate measurements by using a Generative Adversarial Network. This approach effectively captures the differences between actual measurements and their ideal counterparts, enabling accurate recovery of state estimation data may have been tampered with by FDI attacks. However, Kalman based techniques are susceptible to adversarial attacks and require extensive training to enhance their robustness. [33].

Yang et al. [34] generated two deliberate false data injection attacks on the forward and backward parths of a control system in a cyber-physical system, with the intention of compromising the smart grid. The controllable parameters provide flexibility in adjusting the attack's impact on the system's performance. Further, inverse estimation and the Kalman filtering method are proposed as a defence system to counter forward and feedback attacks. Sargolzaei et al. [35] developed a controller to detect FDI attacks using a Kalman filter to estimate agent states ensembled with Neural Network architecture to respond to anomalies used by FDI attacks, which increases accuracy in the detection of FDI attacks.

After analyzing the existing research, it is evident that the smart grids involve extensive operations and transmissions of time series data across domains. In real-world scenarios, protecting data during transmission between information nodes within the smart grid environment becomes crucial to ensure data confidentiality and integrity. Additionally, the system should be able to detect and recover from false data

injection attacks, ensuring the identification and regeneration of any corrupted values that may compromise the grid's reliability and security.

A. Paper Contribution

The primary contributions of this research study are as follows:

1) Utilized real time CESCOM dataset for evaluation of our proposed SHA-SARIMAX algorithm.
2) Building encoded text for transmitting and receiving node data to identify the files corrupted by FDIA using SHA-256.
3) Design and Implement a Time series-based SARIMAX algorithm for recovering the corrupted file.
4) Performance evaluation of proposed SHA-SARIMAX using different records of CESCOM dataset.

III. Background

A. Secured Hashing Algorithm

In this study, the transmission phase involved utilizing the Hash-based Message Authentication Code (HMAC) Secure Hashing Algorithm (SHA) framework to create a hash function capable of generating a 256-bit (32-bytes) hash value, known as a message digest, from the provided input data. This message digest served as a crucial means of verifying the integrity of the transmitted data at the receiver's end. Figure 1 illustrates the block diagram of HMAC framework.

To ensure secure data transmission, both the data and its corresponding message digest were transmitted to the receiver, utilizing the receiver's Public-key on the receiving node. After receiving the data, the receiver creates a new message digest for the received data. Subsequently, the generated message digest was compared with the one received during transmission for identifying the corrupted data.

B. Time Series algorithm and SARIMAX

In this research, we considered Time Series (TS) algorithms to recover the corrupted files. TS algorithm is a computational technique designed to analyze and model data points in documented order, specifically for time-dependent data. In our present scenario of the data collected from CESCOM, Mysuru, We will utilize the TS algorithm to forecast the potential value at any point in the middle of the time series data. Further, we utilised the popular SARIMAX model of the TS forecasting model in this research work.

SARIMAX is an extension of the traditional SARIMA model, specifically designed to incorporate exogenous variables into the forecasting process. Although SARIMA is efficient in modelling and forecasting data with seasonal patterns and dependencies, its adequacy might be limited when external factors influence the time series. SARIMAX addresses this limitation by introducing exogenous variables alongside past values and seasonal patterns to capture the complexity of the time series.

The SARIMAX model includes multiple components such as seasonal autoregressive (SAR), seasonal moving average (SMA), non-seasonal autoregressive (AR), non-seasonal moving average (MA), integration (I), and exogenous variables (X). with the consideration of all these components, the model becomes a comprehensive tool for time series forecasting. It enables analysts to account for both internal patterns and external influences, making it more robust and versatile in various real-world scenarios.

IV. Methodology

Figure 3 depicts the suggested SHA-SARIMAX structure for detecting and restoring damaged files from FDIA. The process flow for the working of the proposed SHA-SARIMAX framework is as follows:

1) In this research, we adopted a hashing approach to process the data rows.
2) The computed hash values are appended into a file at the sending node.
3) To ensure data security, a reliable hashing algorithm is employed to compute the values for each feature, considering every 30 samples or rows of data.
4) During the data transmission, two files were sent to the receiving node.
5) One of the files contained the generated values collected at the node.
6) Another file contains securely encrypted hash values, intended for decryption at the receiving node.
7) The hash values are regenerated and appended to a new file at the receiving node.
8) To verify data integrity, the reproduced hash file, containing the regenerated hash values, is compared with the one received from the transmitter.
9) If any discrepancies are found, the model would apply the Time Series algorithm to the attribute(s) that exhibited different hash values.

A. Transmission Phase

The transmission node algorithm involves the following steps: The datafile is read, ensuring proper data. A Hash file is created to store the hashed values of the features. A loop processes 'n' records till the end of the dataset, generating Hash values for each feature and updating the Hash file accordingly. Additionally, the algorithm calculates the minimum, maximum, mean, and median values for the 'n' records. A Hash value is computed for each row in the dataset. After processing all 'n' records, the file is closed. The original data file and the Hash file are then transmitted using a private-public key cryptosystem, ensuring secure data transfer.

B. Receiving Phase

The received data file undergoes another hashing process to generate a new hash file at the receiver's end. The complete steps are discussed in Figure 3. According to Figure 3 the generated hash file is compared with the received hash file to ensure data integrity. Employing the TS algorithm, we examined all the row values associated with the corrupted feature and processed them with the TS algorithm. The dataset incorporates the regenerated value. It replaces values for all features that have changed, and this process continues until thoroughly examining the last row. After decryption, the receiver deciphers the received data. Figure 4 compares sender Hash files at the transmission node and receiver hash

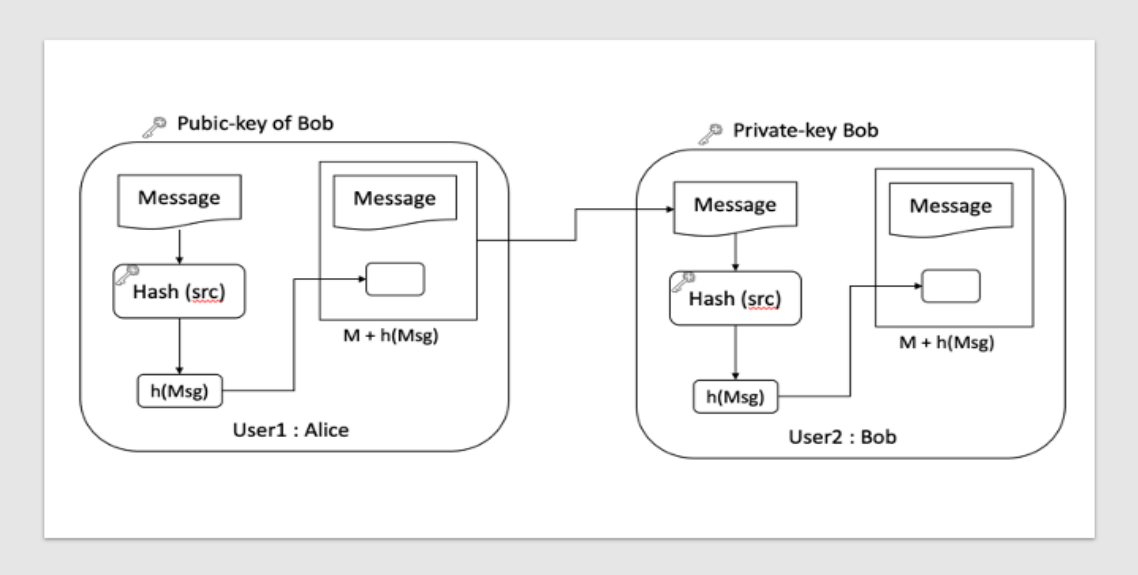


Fig. 1: HMAC Algorithm

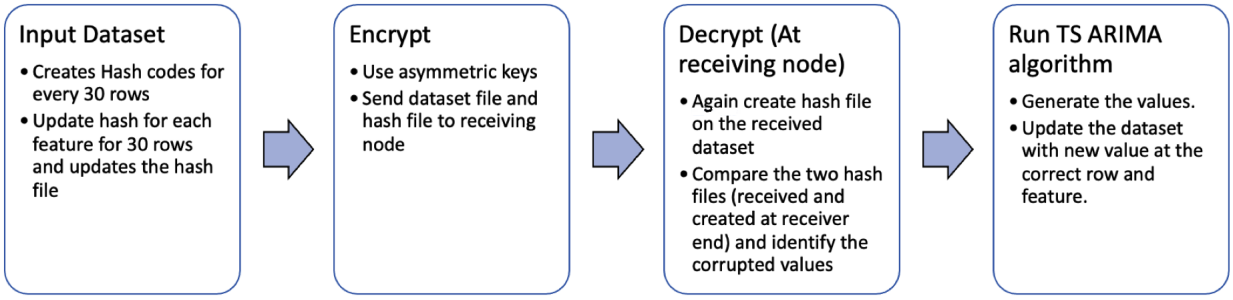


Fig. 2: Proposed SHA-SARIMAX process flow

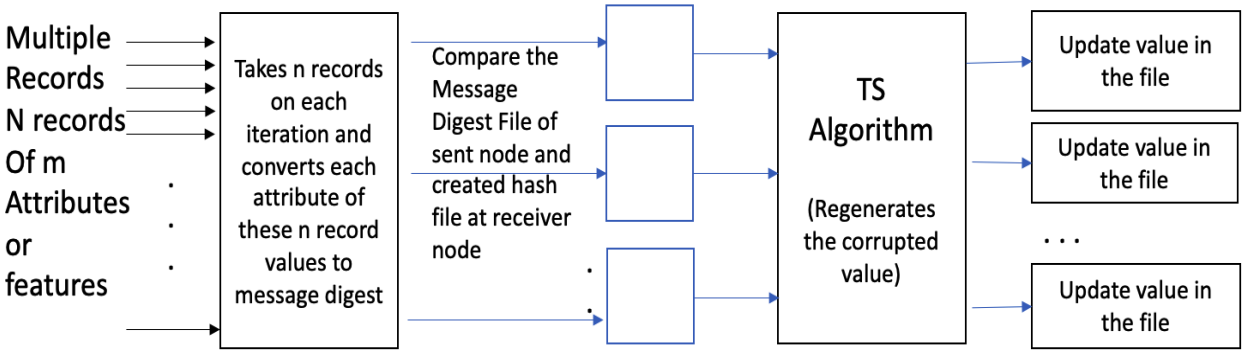


Fig. 3: Receiving Node infrastructure

files at the receiver node. Further, a comparison between the message digest and the SHA is obtained for each row. If a discrepancy is found in the resulting hash value, our statistical time series model regenerates the affected values. The complete explanation is depicted in Algorithm 2.

The motive behind employing the TS Algorithm is that training a Neural model on one dataset may not apply to another dataset acquired from a different transmission node or substation. Each substation may have distinct data, such as Phase Voltage, Active Power, Reactive Power, Current, and Power Consumption, making it unique. Therefore, knowledge gained from one station cannot be directly applied to predict or recover values for data from another station.

C. Simulations Results and Discussions

The efficacy of the proposed SHA-SARIMAX framework was assessed in this research work using the real-time CESCOM, Mysuru dataset. Initially, we investigate the correlation between each of the features. Notably, we find correlations between Phase-Voltage on the 66KV Transformer and 11KV on the same Y line ('66HV1Y-B_PHVOLT vs '11LV1Y-B_PHVOLT'), as well as correlations between certain Power values ('11LV1IB' vs '11LV1IR'). These

Algorithm 1 Transmission of secured data

```
procedure DATA TRANSMISSION(A)
    Read the datafile with columns having proper data.
    Create a Hash file
    for every n records do
        Create Hash values of all the features
        Update the hash file.
        Update min, max, mean and median values of the n records
        Create a hash value for each row
    end for
    Close the file
    Send the original file and Hash file using private-public key cryptosystem.
end procedure
```

Algorithm 2 Receiving and analysing secured data

```
procedure DATA RECEPTION(A)
    Decrypt the received file
    Rerun to get or the hash file fthe received datafile
    for every n records do
        Compare the two hash files generated vs received file for n rows for each feature.
        Get the exact feature value which is corrupted.
        Run the TS algorithm to regenerate the value
        Update the file.
    end for
    Close the file
end procedure
```

findings indicate that changes in Phase voltage (66HV1Y-B_PHVOLT) have an impact on the Current, which, in turn, affects the (11KVY-BPHVOLT). Figure 5 depicts the correlation matrix for the regenerated values. This correlation matrix graph serves as a confirmation that the regenerated values closely match the original data and remain within acceptable ranges.

During this assessment, we focused on the last four corrupted files and compared the observed values with the algorithm's predicted values. For this purpose, the CESCOM dataset is splitted into 50, 100 and 250 records. Table I illustrates the results obtained from statistical analysis of the proposed algorithm. The analysis revealed that for the 250-record subset, the proposed SHA-SARIMAX algorithm exhibited the highest average prediction accuracy, achieving a remarkable 99.90% prediction accuracy in power units (p.units). This performance metric surpassed the results obtained for other subsets of records.

Further, to validate the algorithm, the results are plotted using Standardized residuals. Standardized residuals play a pivotal role in conducting essential model diagnostics, enabling the evaluation of model fit, verification of underlying assumptions, and detecting outliers or irregularities in time series data [36]. Notably, in figure 5, the residuals of the 250-record subset are observed to be closer to the prediction line, signifying the strong goodness of fit achieved by the proposed SHA-SARIMAX algorithm.

Figure 7 illustrates the histogram representing records of 30, 50, 100, 250, and 1440. These observations interpret the distribution and density of values within the active voltage range. The histogram indicates that the dataset under scrutiny is a real-time dataset, evident from its non-normalized and skewed distribution. Consequently, the restoration of corrupted values is skewed for certain values within a specified range. To address this non-normalization in restoring corrupted values, this research introduces the TS algorithm utilizing SHA-SARIMAX.

Figure 8 displays a correlogram graph that demonstrates the alignment of Auto Correlation Function (ACF) values with the closest observed values. The process of normalizing the error value depends on the number of records considered. If the error value is identified within a few record values, the normalization to the correct value occurs more quickly. Conversely, if the error is located towards the end of the records, the normalization process extends over 10 to 20 records. This observation indicates that a smaller number of records requires more time for the corrected value to be normalized.

In conclusion, after conducting TA analysis on the CESCOM, Mysuru dataset data from a substation node, the following observations were derived: utilizing values generated within the specified time interval from the dataset file, the ARIMA model learned from a subset of data and forecasted the best-fit value for one of the features, ensuring it fell within the range of the values for that attribute. The complete simulation diagnosis is presented in Figure 6a. Following this, the procedure was iteratively replicated for other modified attributes of the dataset. For illustration, when we performed the experiment with the feature 'JTYNGR66HV1Y-B_PHVOLT', the forecasted values consistently aligned with the expected values. The Mean Squared Error (MSE) of our forecasts was calculated to be 0.01, while the Root Mean Squared Error (RMSE) stood at 0.09 as shown in Figure 6b.

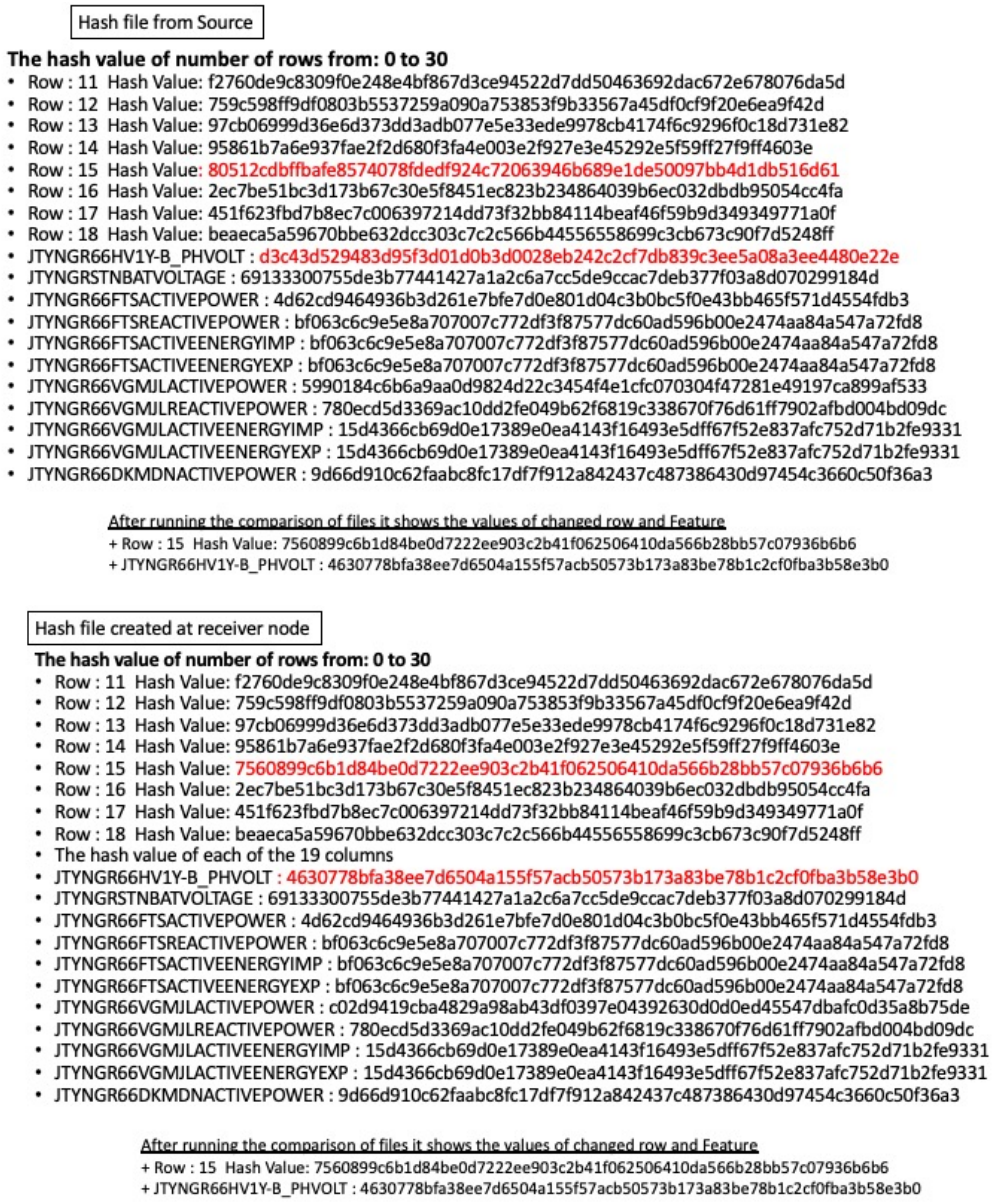


Fig. 4: Comparison of a Hash file generated from transmitter and receiver end

V. Conclusion

Smart grid introduces a promising solution for modernizing electrical power systems to align with the demands of an increasingly digitized world. However, this transformation leads to the emergence of FDIA as a significant cybersecurity threat. To tackle these challenges, we proposed SHA-SARIMAX, a novel framework specifically designed to address FDIA in time series data.

The proposed framework combines distinct elements of SHA and SARIMAX to effectively detect and recover corrupted values in time series datasets affected by FDIA. This work also outline the process of identifying corrupted values through the creation of a hash file and subsequently identifying the corresponding feature and row to regenerate the lost values. We made certain assumptions, including the integrity of the sent hash file, the values remaining within an expected range during the specific time series, and the absence of data fluctuations. Finally, our regeneration of values achieves an accuracy level of nearly 100th decimal places for individual features considered in our proposed model. Future work will explore alternative ML models beyond the TS algorithm to regenerate and compare the accuracy of values, ultimately optimizing the solution for FDIA detection and recovery in TS data.

References

[1] Pankaj Gupta, Ritu Kandari, and Ashwani Kumar. “An introduction to the smart grid-I”. In: *Advances in Smart Grid Power System*. Elsevier, 2021, pp. 1–31.

[2] Ussama Assad et al. “Smart grid, demand response and optimization: A critical review of computational methods”. In: *Energies* 15.6 (2022), p. 2003.

[3] Muhammed Zekeriya Gunduz and Resul Das. “Cyber-security on smart grid: Threats and potential solutions”. In: *Computer networks* 169 (2020), p. 107094.

[4] Light Zaglago Bashir Jimoh and Jose Rodolpho de Oliveira Leo. “Drivers of smart grid technology in Ghana”. In: *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science*. 22-24 October 2019, San Francisco, USA, pp. 204–209.

[5] Shahid Tufail et al. “A survey on cybersecurity challenges, detection, and mitigation techniques for the smart grid”. In: *Energies* 14.18 (2021), p. 5894.

TABLE I: Performance analysis of proposed SHA-SARIMAX algorithm for the prediction of corrupted files

Record Subset	Observed Values (power units)	Predicted Values	Accuracy (%)	Average accuracy (%)
30 records	66.5829	65.53592	98.4274817	98.34246
	66.5829	65.52315	98.4083042	
	66.5829	65.51948	98.4027997	
	66.6206	65.37563	98.1312375	
50 Records	65.9141	65.53592	100.292082	99.90934
	66.0509	65.52315	99.6235054	
	66.0509	65.51948	99.9410394	
	66.017	65.37563	99.7807157	
100 Records	65.6954	65.63	99.9004086	99.36161
	65.6954	65.61568	99.8786095	
	65.6954	65.1712	99.2020297	
	65.6078	64.60094	98.4654023	
250 Record	65.6954	65.67752	99.9727363	**99.7681**
	65.6954	65.63568	99.909053	
	65.6954	65.6712	99.9631177	
	65.6078	65.10094	99.2275074	

[6] Riyadh Rahef Nuiaa et al. "Enhancing the Performance of Detect DRDoS DNS Attacks Based on the Machine Learning and Proactive Feature Selection (PFS) Model." In: *IAENG International Journal of Computer Science* 49.2 (2022), pp. 511–524.

[7] D Alert. "Cyber-attack against ukrainian critical infrastructure". In: *Cybersecurity Infrastruct. Secur. Agency, Washington, DC, USA, Tech. Rep. ICS Alert (IR-ALERT-H-16-056-01)* (2016).

[8] Gaoqi Liang et al. "The 2015 Ukraine blackout: Implications for false data injection attacks". In: *IEEE Transactions on Power Systems* 32.4 (2016), pp. 3317–3318.

[9] Dan Esposito Gimon and Eric. *The texas big freeze: How much were markets to blame for widespread outages?* June 2021. URL: https://www.utilitydive.com/news/the-texas-big-freeze-how-much-were-markets-to-blame-for-widespread-outages/601158/.

[10] TeamSymantec. *Dragonfly: Western energy sector targeted by Sophisticated Attack Group*. 2017. URL: https://symantec-enterprise-blogs.security.com/blogs/threat-intelligence/dragonfly-energy-sector-cyber-attacks.

[11] Usman Inayat et al. "Cybersecurity enhancement of smart grid: Attacks, methods, and prospects". In: *Electronics* 11.23 (2022), p. 3854.

[12] Junbo Zhao et al. "Short-term state forecasting-aided method for detection of smart grid general false data injection attacks". In: *IEEE Transactions on Smart Grid* 8.4 (2015), pp. 1580–1590.

[13] Mohammad Ashrafuzzaman et al. "Detecting stealthy false data injection attacks in the smart grid using ensemble-based machine learning". In: *Computers & Security* 97 (2020), p. 101994.

[14] Mostafa Mohammadpourfard et al. "Ensuring cybersecurity of smart grid against data integrity attacks under concept drift". In: *International Journal of Electrical Power & Energy Systems* 119 (2020), p. 105947.

[15] Mario R Camana Acosta et al. "Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks". In: *IEEE Access* 8 (2020), pp. 19921–19933.

[16] Shuoyao Wang, Suzhi Bi, and Ying-Jun Angela Zhang. "Locational detection of the false data injection attack in a smart grid: A multilabel classification approach". In: *IEEE Internet of Things Journal* 7.9 (2020), pp. 8218–8227.

[17] Mohamed Amine Ferrag and Leandros Maglaras. "DeepCoin: A novel deep learning and blockchain-based energy exchange framework for smart grids". In: *IEEE Transactions on Engineering Management* 67.4 (2019), pp. 1285–1297.

[18] Hamed Haggi, Meng Song, Wei Sun, et al. "A review of smart grid restoration to enhance cyber-physical system resilience". In: *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)* (2019), pp. 4008–4013.

[19] Sajjad Khan et al. "Short-term electricity price forecasting by employing ensemble empirical mode decomposition and extreme learning machine". In: *Forecasting* 3.3 (2021), p. 28.

[20] Ke Yuan et al. "Blockchain Security Research Progress and Hotspots." In: *IAENG International Journal of Computer Science* 49.2 (2022), pp. 433–444.

[21] Jianguo Ding et al. "Cyber threats to smart grids: Review, taxonomy, potential solutions, and future directions". In: *Energies* 15.18 (2022), p. 6799.

[22] Guangdou Zhang et al. "Spatio-temporal correlation-based false data injection attack detection using deep convolutional neural network". In: *IEEE Transactions on Smart Grid* 13.1 (2021), pp. 750–761.

[23] Junjun Xu et al. "A secure forecasting-aided state estimation framework for power distribution systems against false data injection attacks". In: *Applied Energy* 328 (2022), p. 120107.

[24] Yufeng Wang et al. "KFRNN: An effective false data injection attack detection in smart grid based on Kalman filter and recurrent neural network". In: *IEEE Internet of Things Journal* 9.9 (2021), pp. 6893–6904.

[25] Yinghua Han et al. "False data injection attacks detection with modified temporal multi-graph convolutional network in smart grids". In: *Computers & Security* 124 (2023), p. 103016.

[26] Keke Huang et al. "False data injection attacks detection in smart grid: A structural sparse matrix separation method". In: *IEEE Transactions on Network Science and Engineering* 8.3 (2021), pp. 2545–2558.

[27] Yikun Huang and Haolin He. "Advance learning technique for the electricity market attack detection". In: *Computers and Electrical Engineering* 100 (2022), p. 107865.

[28] Merve Aladag, Ferhat Ozgur Catak, and Ensar Gul. "Preventing data poisoning attacks by using generative models". In: *2019 1St International informatics and software engineering conference (UBMYK)*. IEEE. 2019, pp. 1–5.

[29] Loc Truong et al. "Systematic evaluation of backdoor data poisoning attacks on image classifiers". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020, pp. 788–789.

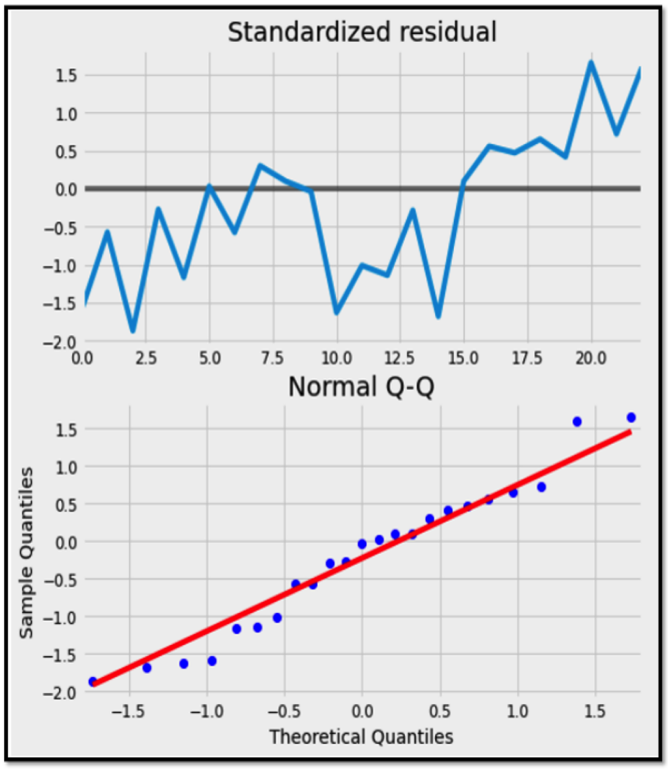
[30] Kewei Sha, Naif Alatrash, and Zhiwei Wang. "A secure and efficient framework to read isolated smart grid devices". In: *IEEE Transactions on Smart Grid* 8.6 (2016), pp. 2519–2531.

[31] Keyan Abdul-Aziz Mutlaq et al. "Symmetric Key Based Scheme for Verification Token Generation in Internet of Things Communication Environment". In: *EAI International Conference on Applied Cryptography in Computer and Communications*. Springer. 2022, pp. 46–64.

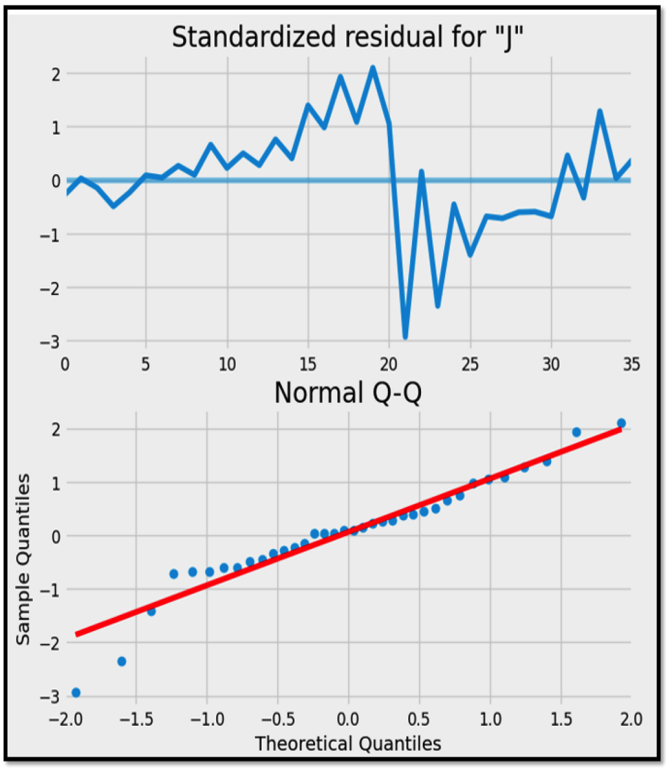
[32] Yuancheng Li, Yuanyuan Wang, and Shiyan Hu. "Online generative adversary network based measurement recovery in false data injection attacks: A cyber-physical approach". In: *IEEE Transactions on Industrial Informatics* 16.3 (2019), pp. 2031–2043.

[33] Lei Cui et al. "Detecting false data attacks using machine learning techniques in smart grid: A survey". In: *Journal of Network and Computer Applications* 170 (2020), p. 102808.

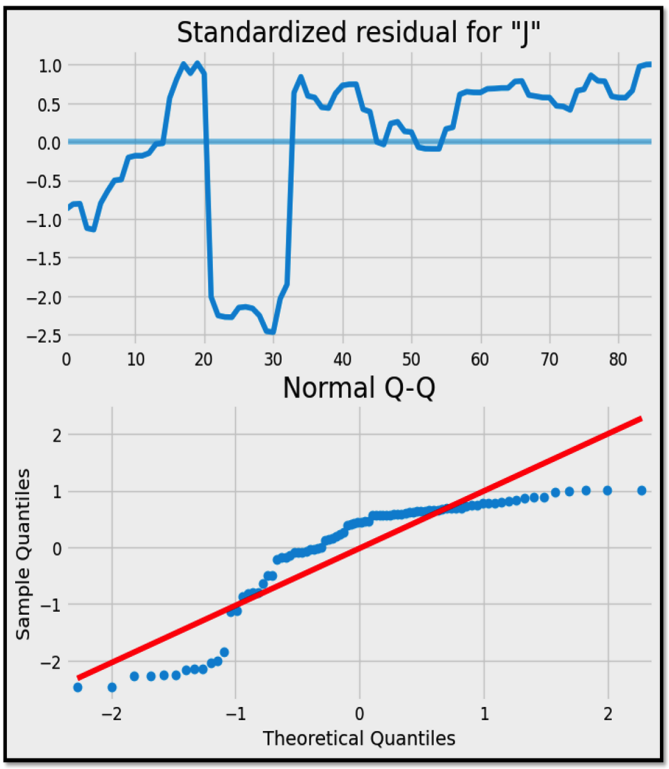
[34] Janghoon Yang. "A controllable false data injection attack for a cyber physical system". In: *IEEE Access* 9 (2021), pp. 6721–6728.



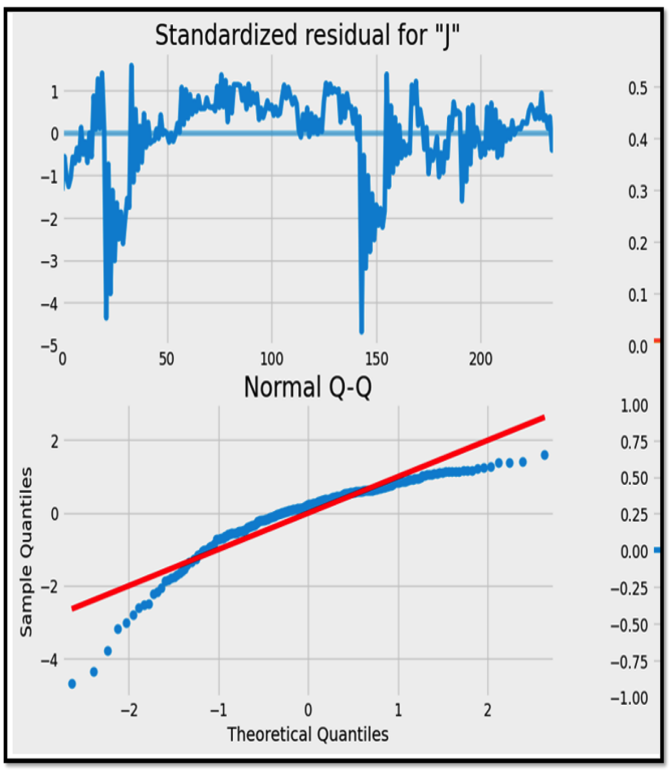
(a) 30 Records



(b) 50 Records



(c) 100 Records

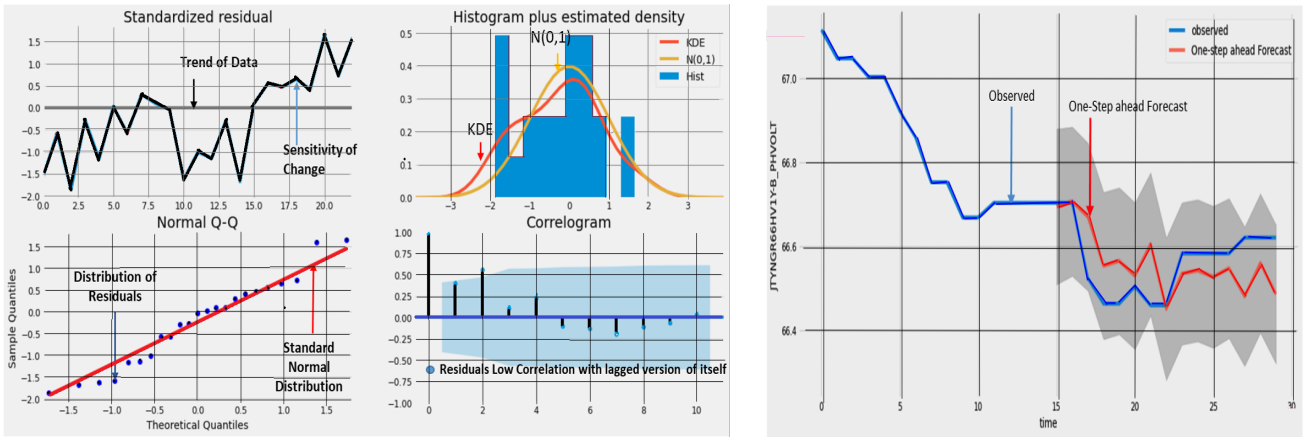


(d) 250 Records

Fig. 5: Standard Residuals of CESCOM dataset

[35] Arman Sargolzaei et al. "Detection and mitigation of false data injection attacks in networked control systems". In: *IEEE Transactions on Industrial Informatics* 16.6 (2019), pp. 4281–4292.

[36] Cindy Feng, Longhai Li, and Alireza Sadeghpour. "A comparison of residual diagnosis tools for diagnosing regression models for count data". In: *BMC Medical Research Methodology* 20.1 (2020), pp. 1–21.



(a) Result diagnostics of proposed SHA-SARIMAX

(b) Predicted value at 15th row

Fig. 6: Simulation Analysis of proposed SHA-SARIMAX

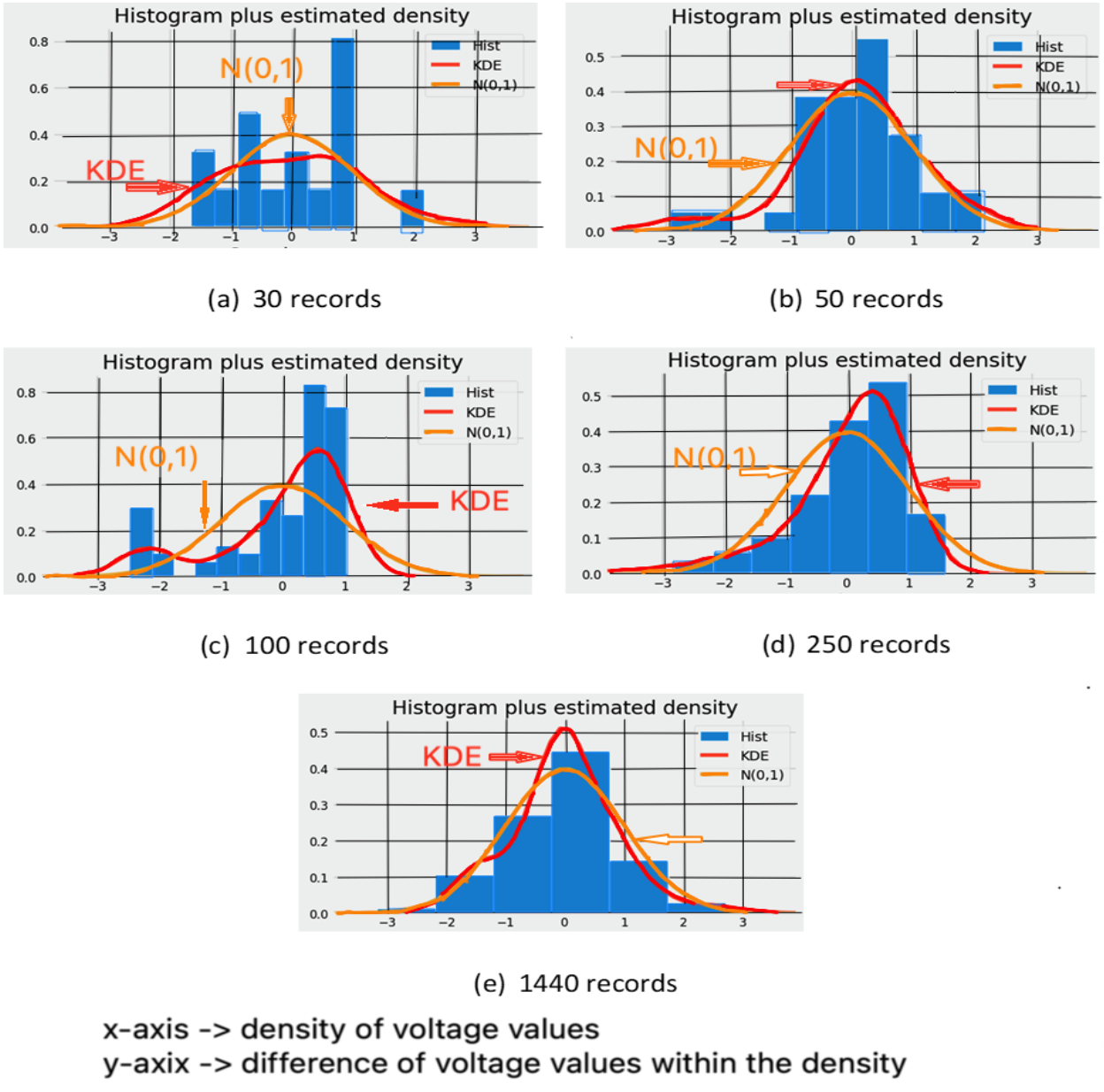


Fig. 7: Histogram density of 30, 50, 100, 250 and 1440 records

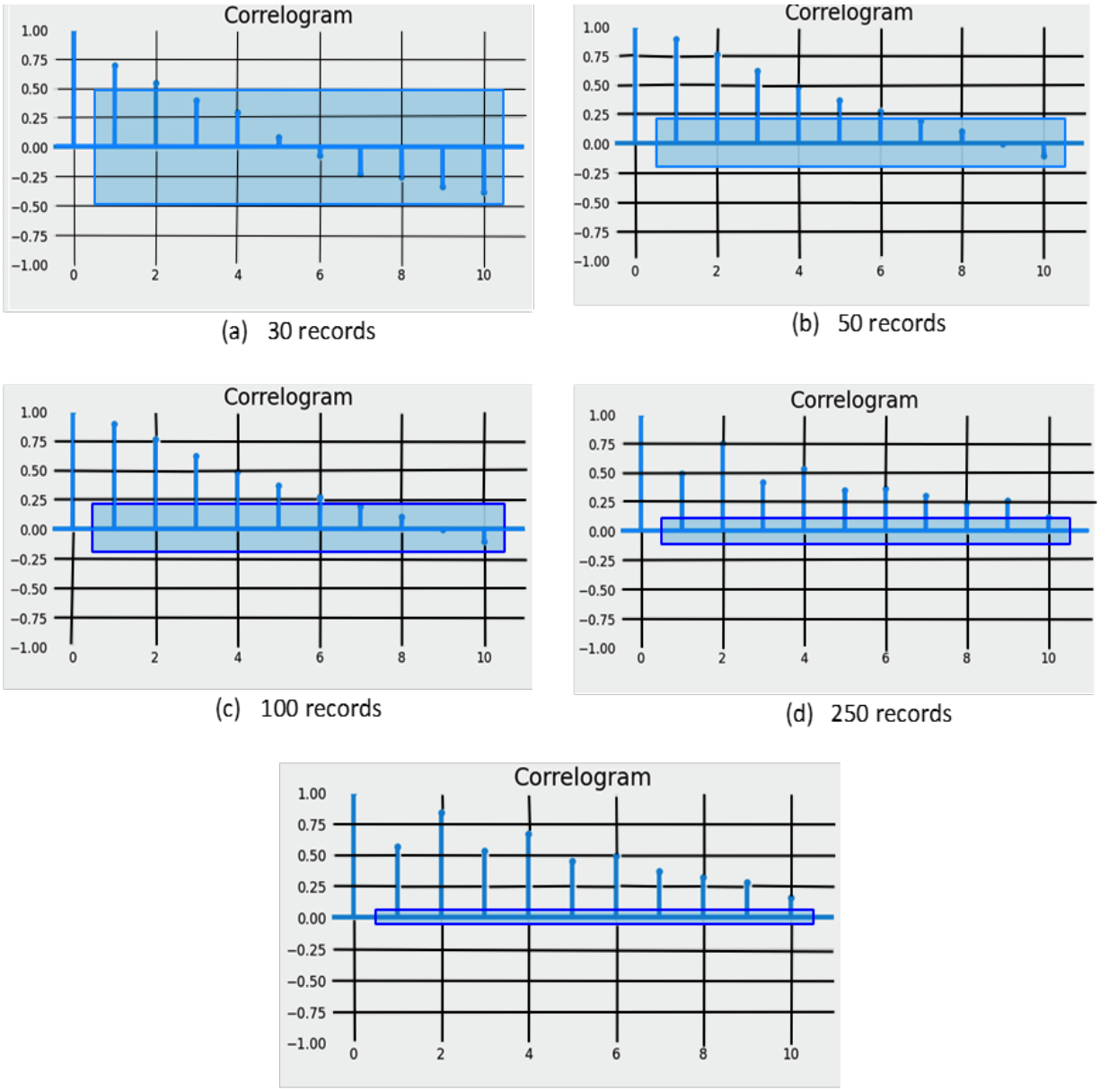


Fig. 8: Correlogram of 30, 50, 100, 250 and 1440 records (X-axis: Number of records, Y-axis Reactive Power distribution of each TS based record)