Turnitin Report_Aditya Kumar_09-05-2024

by Aditya Kumar

Submission date: 09-May-2024 09:20AM (UTC+0530)
Submission ID: 2374804079
File name: 2001EE85_Aditya_Kumar-1.pdf (1.36M)
Word count: 5894
Character count: 31410

DETECTION OF ONSET DISTURBANCE FOR INERTIA ESTIMATION USING KERNEL DENSITY ESTIMATION

A Project Report Submitted

by

ADITYA KUMAR

(2001EE85)

In Partial Fulfilment

of the Requirements for the award of the degree

BACHELOR OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY PATNA

7TH MAY 2024

THESIS CERTIFICATE

This is to certify that the thesis titled **DETECTION OF ONSET DISTURBANCE FOR INERTIA ESTIMATION USING KERNEL DENSITY ESTIMATION**, submitted by **Aditya Kumar**, to the Indian Institute of Technology, Patna, for the award of the degree of **Bachelor of Technology** is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. S.K Parida
Supervisor
Associate Professor
Dept. of Electrical Engineering
IIT-Patna, 800 013

Place: Patna

Date: May 7, 2024

ACKNOWLEDGEMENTS

I want to express my gratitude to Dr. S.K.Parida, my BTP supervisor, for his invaluable guidance throughout the project. His support in conceptualizing the project and providing essential resources to tackle the problem statement was highly appreciated.

ABSTRACT

This project focuses on the detection of disturbances in a 2-area power system for estimating inertia using machine learning techniques. The project investigates an old paper that uses a fixed threshold detection method to isolate the inertial response from the overall frequency response and introduces a new method for the same. The data collection phase involved simulating a transfer function model of a 2-area system and collecting time series data of the center of inertia frequency from 15 signals, each with 1500 samples. The project first explored the use of different algorithms for disturbance detection, like "Density-Based Spatial Clustering of Applications with Noise (DBSCAN)" and "Kernel Density Estimation (KDE)". After testing both algorithms, we decided to go with Kernel Density Estimation.

The model was fed with the dataset one by one and plotted the density and histogram plots of log likelihood values for each signal. We also interpolated the detected outliers over the centre of inertia frequency plot to show the onset of disturbance.

During simulation, the disturbance was given at 10s, and the model's average percentage error in onset detection was coming out to be around **2.5%**.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Power systems play a pivotal role in modern society, providing the backbone for the distribution and utilization of electricity. The stability and efficiency of these systems are of utmost importance to ensure the uninterrupted supply of electricity to consumers. In this context, the estimation of inertia in power systems is a critical aspect that directly impacts the system's stability and response to disturbances. Inertia, in the context of power systems, refers to the ability of rotating machines, such as generators and turbines, to resist changes in speed. This property is crucial for maintaining the stability of the system and ensuring a continuous supply of electricity.

One of the key challenges in power system operation is the detection and mitigation of disturbances. Disturbances can arise from various sources, such as sudden changes in load, faults in the system, or other external factors. Detecting these disturbances early and accurately is essential for preventing cascading failures and ensuring the overall stability of the system. Traditional methods for detecting disturbances often rely on fixed threshold detection, which has several limitations, including the need for careful selection of thresholds and susceptibility to noise.

To address these challenges, this study focuses on the development of a novel approach for the detection of disturbances in power systems using machine learning techniques. Specifically, the study explores the use of "Density-Based Spatial Clustering of Applications with Noise (DBSCAN)" and "Kernel Density Estimation (KDE)" for anomaly detection and onset detection of disturbances in power systems. These techniques offer advantages over traditional methods by providing a data-driven approach that can adapt to the specific characteristics of the system.

The research methodology involves the simulation of a transfer function model of a power system consisting of two distinct areas and the collection of data on disturbance signals, center of inertia frequency, and disturbance magnitude. This data is then used to train and evaluate the KDE model for anomaly detection. The study also investigates the impact of different parameters, such as bandwidth and kernel type, on the performance

of the models. The study's outcomes are anticipated to offer valuable perspectives on the use of machine learning methodologies in detecting disturbances within power systems. This research aims to enhance the field's understanding of power system stability and reliability by investigating novel methods for disturbance detection. Through the application of machine learning techniques, we aim to augment power systems' capability to identify and manage disturbances, leading to improved stability and efficiency in these essential infrastructure systems.

1.1 Inertia in power systems

In power systems, the inertia of machines, like generators and turbines, is crucial for stability. It refers to the energy stored in large rotating generators and some industrial motors, which gives them the tendency to remain rotating. This property helps stabilize the system's frequency, buffering against sudden demand or supply changes. When large loads connect or disconnect suddenly, machine inertia absorbs the impact, allowing the system to stabilize.

Moreover, inertia aids the system's response to disturbances. It stores kinetic energy, helping the system ride through disruptions. This acts as a "shock absorber," preventing widespread failures. Overall, machine inertia is vital for grid stability, ensuring reliable operation. Managing and understanding inertia is key to efficient power system operation.

In conclusion, machine inertia is vital for power system stability. It helps maintain frequency during load changes and aids in the system's response to disturbances. Managing inertia is crucial for a reliable power supply.

1.2 Centre of Inertia Frequency

The center of inertia (COI) frequency is a crucial parameter in power systems, reflecting the balance between generation and demand. It represents the weighted average frequency of synchronous generators in the grid, indicating system stability. Fluctuations in COI frequency suggest imbalances in generation and load, potentially leading

to instability and blackouts. Monitoring and controlling COI frequency are vital for grid stability, helping operators make real-time adjustments to maintain system balance.

COI frequency is influenced by factors like generation-load balance, control system responses, and disturbances. Deviations from the nominal frequency signal cause imbalances, prompting corrective actions. Operators rely on COI frequency to assess grid health and manage generation-load balance for stability. It's also used in control and protection schemes to mitigate disturbances and prevent widespread outages.

In conclusion, the center of inertia frequency is pivotal for grid stability and reliability. It serves as a critical indicator of system health, guiding operators in maintaining balance and mitigating disruptions. Understanding and controlling COI frequency are essential for ensuring the efficient and reliable operation of power systems.

$$\text{COI frequency} = \frac{\sum_{i=1}^{n} f_i \cdot H_i}{\sum_{i=1}^{n} H_i}$$

In the above formula, H_i is the inertia constant of the ith generator, expressed at a common VA base. f_i is filtered individual frequency measurements.

1.3 The Swing equation

The swing equation is a critical component of power system analysis, providing insight into the dynamic behavior of synchronous generators within a multi-machine system. It describes the electro-mechanical interaction between the mechanical and electrical components of a generator, offering a fundamental understanding of power system stability.

The swing equation is represented by the formula:

$$\frac{2 \cdot H_{\text{sys}} \cdot S \cdot \frac{df}{dt}}{f_0} = P_m - P_e$$

where:

H_sys: system inertia constant, reflecting the ability of the system to store and release energy.

S: apparent power of the system.

f: frequency

f_0: standard system's frequency.

P_m: input as mechanical power to the generator.

P_e: output as electrical power from the generator.

This equation illustrates how changes in mechanical power input and electrical power output affect the system frequency, offering valuable insights into the stability of the power system. The traditional approach based on the swing equation, which estimates the inertia constant by analyzing the frequency response following a known disturbance, is a commonly utilized method in power system analyses. [2].

CHAPTER 2

LITERATURE REVIEW

The paper[1] delves into the intricate realm of inertia estimation within microgrid systems, particularly focusing on the challenges posed by the integration of renewable energy sources. In the context of modern power systems, the increasing penetration of renewable energy generation, such as wind and solar, has brought about a paradigm shift in the dynamics of grid stability and frequency control. The traditional synchronous generators, which have long been the backbone of grid stability, are now being supplemented and, in some cases, replaced by these intermittent and non-synchronous renewable sources.

One of the primary issues addressed in the paper is the accurate estimation of inertia in microgrid systems. Inertia, a crucial parameter in power systems, plays a vital role in maintaining grid stability and frequency regulation. However, with the rise of renewable energy sources that do not inherently possess inertia like conventional synchronous generators, accurately quantifying and incorporating inertia into the system becomes a challenging task. The paper proposes a novel method that leverages variable order polynomial fitting for frequency measurement to enhance the accuracy of inertia estimation in microgrids.

The study underscores the limitations of conventional swing equation-based methods for estimating inertia, especially in the presence of virtual inertia support from renewable sources. By segregating the estimation of synchronous and non-synchronous inertia, the proposed method aims to provide a more nuanced and precise understanding of the system's dynamics. This distinction is crucial for the effective control and operation of microgrid systems, where a mix of synchronous and non-synchronous generation sources coexist.

Detecting the onset of a disturbance is critical for accurately fitting the curve of the calculated frequency response. In [1], [3], the threshold value of rate of change of frequency was 0.05 Hz/s and was employed to pinpoint the exact start of the event. To reduce the impact of measurement errors or noise, a 50 ms moving average filter was

applied to the RoCoF value. Any sample that surpassed this threshold was classified as a disturbance.

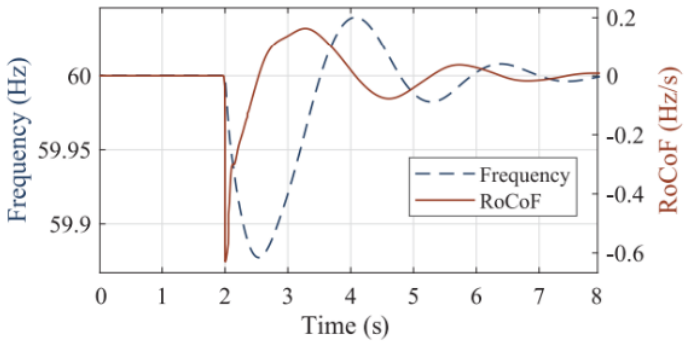


Figure 2.1: Frequency and RoCoF vs time [1]

Furthermore, the paper emphasizes the significance of curve fitting for power to calculate inertia accurately. By analyzing the frequency curve fit and the order of the polynomial used for fitting, the study sheds light on the trade-offs involved in capturing the inertial response within the duration of the curve fit. It highlights the need for a careful balance between the duration of the event and the order of the polynomial to achieve an acceptable fit and an accurate estimation of inertia.

The study described in the paper utilizes simulation studies conducted in the MATLAB/Simulink environment. It incorporates "Type 3" and "Type 4" wind turbines, selected for their widespread use and ability to implement virtual inertia controls, as well as PV arrays operating at their "Maximum Power Point (MPP)" without energy storage reserves. This approach reflects a practical and realistic method for modeling renewable energy generation within microgrid systems. By integrating multiple units of these sources at the 13.8 kV substation, the study allows for the adjustment of renewable generation penetration levels, mirroring real-world scenarios.

Moreover, the paper draws attention to the importance of accurately estimating inertia in microgrid systems to differentiate between estimations of synchronous and non-synchronous inertia. By providing a detailed analysis of the parameters and ratings of synchronous generators used in the study, the research offers insights into the technical aspects of inertia estimation and its implications for grid stability. The comparison of inertia estimates for different disturbance sizes and WTG penetration levels underscores

the impact of accurate inertia estimation on system performance and reliability.

In conclusion, the paper contributes to the existing body of knowledge on inertia estimation in microgrid systems with renewable energy integration. By proposing a novel method that addresses the challenges posed by virtual inertia and the coexistence of synchronous and nonsynchronous generation sources, the study paves the way for more accurate and effective control strategies in microgrid environments. The findings underscore the importance of accurate inertia estimation for ensuring grid stability, frequency regulation, and overall system reliability in the era of increasing renewable energy integration.

CHAPTER 3

DATA COLLECTION FOR THE MACHINE LEARNING APPROACH

The data collection process for this project involved simulating a transfer function model of a two-area power system and generating synthetic data to mimic real-world scenarios. The system was modeled using mathematical equations and simulation software to replicate the behavior of generators, loads, and transmission lines. The simulation was designed to introduce step disturbances of varying magnitudes to the system, simulating different levels of faults or disturbances that could occur in an actual power system.

For each simulation run, time series data was collected for the center of inertia frequency. The center of inertia frequency is a key parameter that reflects the overall balance between generation and demand in the system.

The data collection process resulted in a dataset consisting of 15 signals, each containing 1500 samples. We applied a 0.35 pu, 0.65 pu, and 0.85 pu step disturbance. They were considered low, medium, and high-level disturbances, respectively. For each disturbance, we varied the duration of the disturbance from 100 ms to 500 ms.

The simulation process was repeated multiple times to generate a diverse dataset that captures a range of possible scenarios in a power system. This dataset will be used to train and evaluate machine learning models for the detection of disturbances and the estimation of inertia in power systems. The synthetic nature of the data allows us to control and manipulate the parameters of the disturbances, providing a controlled environment for testing and validation of the models.

Overall, the data collection process was crucial in providing the necessary inputs for the development and evaluation of our models. The synthetic data generated from the simulation process closely resembles real-world scenarios, ensuring that our models are trained and tested on realistic data. This approach allows us to build robust and accurate models for disturbance detection and inertia estimation in power systems.

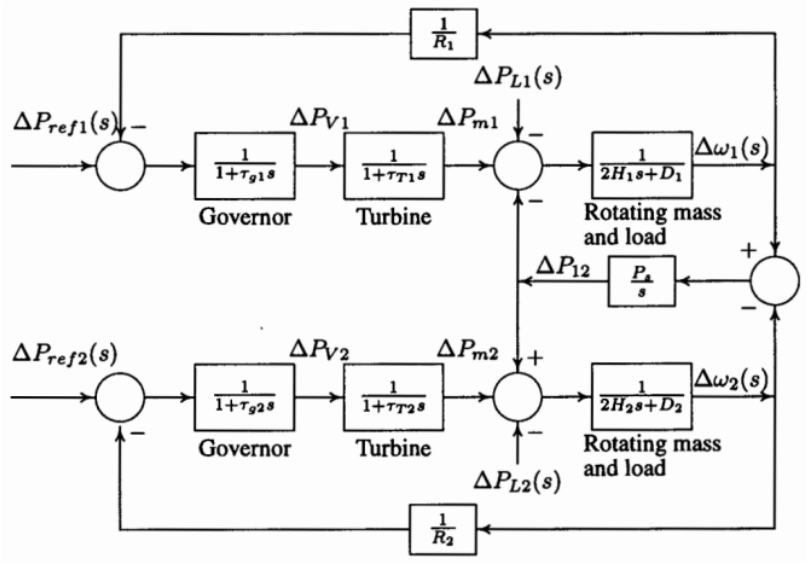


Figure 3.1: Transfer function model of a two-area system

CHAPTER 4

KERNEL DENSITY ESTIMATION (KDE)

After exploring several algorithms like "K-means clustering", "Gaussian Mixture Model (GMM)", and "Density-Based Spatial Clustering of Applications with Noise (DBSCAN)", we decided to go with Kernel Density Estimation.

4.1 Introduction

Kernel Density Estimation (KDE) is a statistical technique used to estimate the probability density function (PDF) of a continuous random variable without assuming a specific parametric form for the distribution. It is particularly useful for visualizing and analyzing the distribution of data when the underlying distribution is not known. KDE is based on the idea of placing a kernel (a smooth, symmetric function) at each data point and summing these kernels to estimate the PDF.

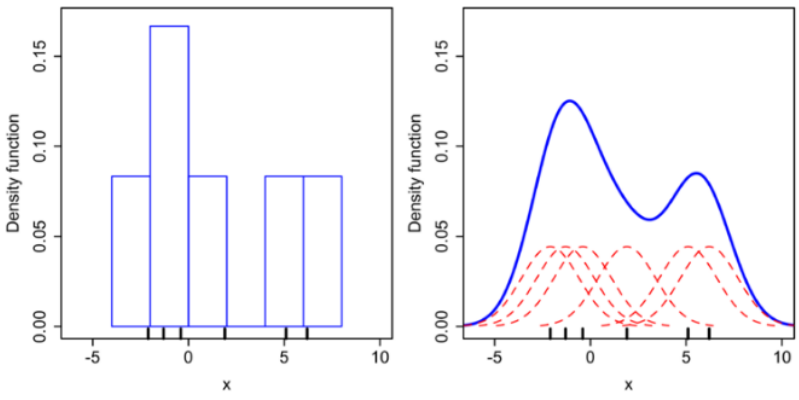


Figure 4.1: Comparison of 1D histogram and KDE

Mathematically, the KDE estimate of the PDF at a point x is given by:

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$$

where:

f_h is the estimated PDF at point x

n is the number of data points,

h is the bandwidth parameter that controls the smoothness of the estimate,

x_i are the data points,

K is the kernel function

4.2 Choice of the Kernel function

Here are some common choices for kernel functions used in kernel density estimation (KDE) along with their mathematical formulations:

1. **Gaussian Kernel:** $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
2. **Uniform Kernel:** $\frac{1}{2}$ for $-1 \leq x \leq 1,$ and 0 otherwise
3. **Exponential Kernel:** $\frac{1}{2} e^{-|x|}$
4. **Epanechnikov Kernel:** $\frac{3}{4}(1 - x^2)$ for $-1 \leq x \leq 1,$ and 0 otherwise

For our onset detection, we used the Gaussian kernel, as it is well-suited for cases where the underlying distribution of the data is expected to be smooth and continuous. The Gaussian kernel assigns weights to data points based on their distance from the point of interest, with closer points receiving higher weights.

4.3 Calculation of the bandwidth

The bandwidth determines the width of the kernel[4] and affects the smoothness of the KDE estimate. Smaller bandwidths lead to more sensitive estimates that capture fine details in the data but may be noisy, while larger bandwidths result in smoother estimates but may over smooth the data. So, we need an algorithm that chooses the optimal bandwidth value and avoids over-smoothing and under-smoothing.

There are several methods for calculating the bandwidth parameter in kernel density estimation (KDE), each with its own characteristics and suitability for different types of data. Some common methods include:

1. **Scott's Rule [5]:** Scott's Rule is a widely used method for bandwidth selection in kernel density estimation (KDE). Scott's Rule is given by $h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5}$, where $\hat{\sigma}$ is the standard deviation of the data and n is the number of data points.

2. **Silverman's Rule [6]:** Silverman's Rule is another popular method for bandwidth selection in KDE. It is based on the inter quartile range of the data and provides a robust estimate of the bandwidth parameter. Silverman's Rule is given by
$$h = 0.9 \cdot \min(\hat{\sigma}, IQR/1.35)n^{-1/5}$$
where $\hat{\sigma}$ is the standard deviation of the data, n is the number of data points and IQR is the inter quartile range of the data.

3. **Cross-Validation:** Cross-validation represents a more advanced approach to bandwidth selection, which entails dividing the data into sets of training and validation data. The goal is to choose the bandwidth that minimizes the mean squared error (MSE) of the KDE when applied to the validation set.

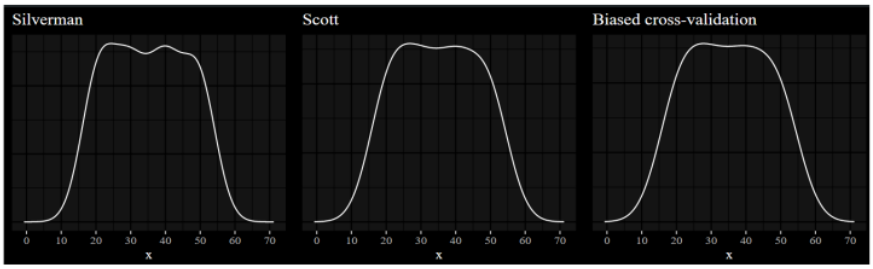


Figure 4.2: Comparison of various bandwidth selectors

For our onset detection, we chose **Scott's rule** for the estimation of bandwidth.

4.4 How it will help in detecting the onset of disturbance?

Kernel Density Estimation (KDE) can help in detecting the onset of disturbance by modeling the underlying probability distribution of the data. In the context of detecting the onset of disturbance in a power system, KDE can be applied to features such as centre of inertia frequency to estimate the probability density function (PDF) of these features under normal operating conditions.

When a disturbance occurs, such as a fault or a sudden change in load, it causes deviations in these features from their normal patterns. These deviations can be detected

as outliers in the estimated PDF. By setting an appropriate threshold based on the PDF, KDE can identify these outliers, which correspond to the onset of disturbances. This allows for the early detection of disturbances and enables timely corrective actions to maintain the stability of the power system.

Overall, KDE provides a flexible and data-driven approach to modeling the underlying distribution of system features, making it well-suited for detecting anomalies like the onset of disturbances in power systems.

CHAPTER 5

METHODOLOGY

5.1 Data Preparation

Data preparation for this project involved simulating the transfer function model of a two-area system. A total of 15 signals were generated, each comprising approximately 1500 samples. These were time series signals representing center of inertia frequency. The signals were sequentially fed into our model for analysis and anomaly detection. This step was crucial to ensure that the model received a diverse set of inputs that accurately reflected the dynamics of the two-area system, enabling it to learn and detect anomalies effectively. As the data were from a simulation model, it didn't require any kind of pre-processing.

5.2 Bandwidth Calculation

In KDE, the bandwidth parameter determines the width of the kernel function used to smooth the data. Scott's rule offers a simple and effective way to calculate this bandwidth automatically. It takes into account the number of data points (n) and the standard deviation ($\hat{\sigma}$) of the data. The formula $h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5}$ implies that the bandwidth decreases as the number of data points or the standard deviation increases, ensuring that the kernel is wider for sparse or more dispersed data and narrower for dense or tightly clustered data. This adaptive approach helps in finding an optimal bandwidth that balances between capturing the underlying distribution's details and avoiding over-smoothing or under-smoothing. In our case, h varies for different dataset.

5.3 KDE model fitting

To fit a KDE model, the first step is to select a kernel function, such as Gaussian, which determines the shape of the probability density estimate. Next, a bandwidth parameter

needs to be chosen, which controls the smoothness of the estimated density. In our case, we have chosen Gaussian as our kernel function.

Once the kernel function and bandwidth are chosen, the KDE model is fitted to the data by placing a kernel function at each data point and summing the contributions from all kernel functions to estimate the density at any given point in the data space. The resulting KDE model provides a smoothed estimate of the underlying probability density function, which can be used for anomaly detection by identifying regions of low density as potential anomalies.

5.4 Log Likelihood Calculation

In Kernel Density Estimation (KDE), the log likelihood values are calculated to assess the likelihood of observing the data given the KDE model. The log likelihood for a single data point x_i is given by the formula:

$$\log(\hat{f}(x_i)) = \log\left(\frac{1}{n}\sum_{j=1}^{n} K_h(x_i - x_j)\right)$$

where $\hat{f}(x_i)$ is the estimated density at point x_i, n is the number of data points, x_j are the data points, and K_h is the kernel function with bandwidth h.

To calculate the log likelihood values for all data points, this formula is applied to each data point in the dataset. These log likelihood values can then be used to assess the overall fit of the KDE model to the data and to compare different models or bandwidths. Lower log likelihood values indicate a better fit of the model to the data, as they represent a higher likelihood of observing the data given the model.

5.5 Visualization of Log Likelihood Density Values

Before determining a threshold for anomaly detection, we visually inspected the log likelihood density values computed by the Kernel Density Estimation (KDE) model. This step was crucial in understanding the distribution of the data and assessing the

goodness of fit of the KDE model.

The density plot displayed the estimated probability density function of the data based on the KDE model. This plot offered a visual representation of the shape of the distribution, illustrating regions of high and low density. By observing the density plot, we could discern patterns in the data and identify potential areas of interest for anomaly detection.

Additionally, the histogram plot depicted the distribution of the log likelihood density values. This plot provided a more detailed view of the density values, allowing us to identify any outliers or anomalies that might be present in the data. By examining the histogram plot, we could gain insights into the overall distribution of the log likelihood values and make informed decisions regarding the selection of a threshold for anomaly detection.

Overall, visualizing the log likelihood density values through density and histogram plots was instrumental in understanding the underlying distribution of the data and guiding the selection of an appropriate threshold for anomaly detection.

5.6 Threshold Selection

To identify anomalies, we selected a threshold based on the log likelihood scores. **25th** percentile was chosen as threshold, indicating the value below which 25% of the data falls. This approach was chosen to capture data points that deviate significantly from the majority of the dataset, as indicated by their lower log likelihood scores. Points with log likelihood scores below this threshold were flagged as anomalies or outliers, suggesting that they were less likely to have been generated by the underlying distribution modeled by the KDE. This method provided a systematic way to detect and isolate potential anomalies in the dataset.

5.7 Onset Detection

After setting the threshold, we identified anomalies by selecting data points with log likelihood scores below this threshold. These anomalies denote instances in the time

series where the data significantly deviates from the expected distribution. Such deviations suggest the presence of disturbances or anomalies in the system at those specific time points. By pinpointing these anomalies, our method helps in detecting and potentially addressing irregularities or issues in the system, contributing to improved system monitoring and maintenance.

5.8 Visualizing outliers over center of inertia frequency vs time plot

For visualizing the results, we utilized the Plotly package to create a plot of f_coi vs. time. Over this plot, we highlighted the outliers detected by our anomaly detection method and marked the onset value. Additionally, we calculated the percentage error, which provides insight into the accuracy of our onset detection. The average the percentage error was approximately 2.5%, indicating the average deviation of our predicted onset values from the actual onset times. This metric helps in assessing the performance and reliability of our anomaly detection approach.

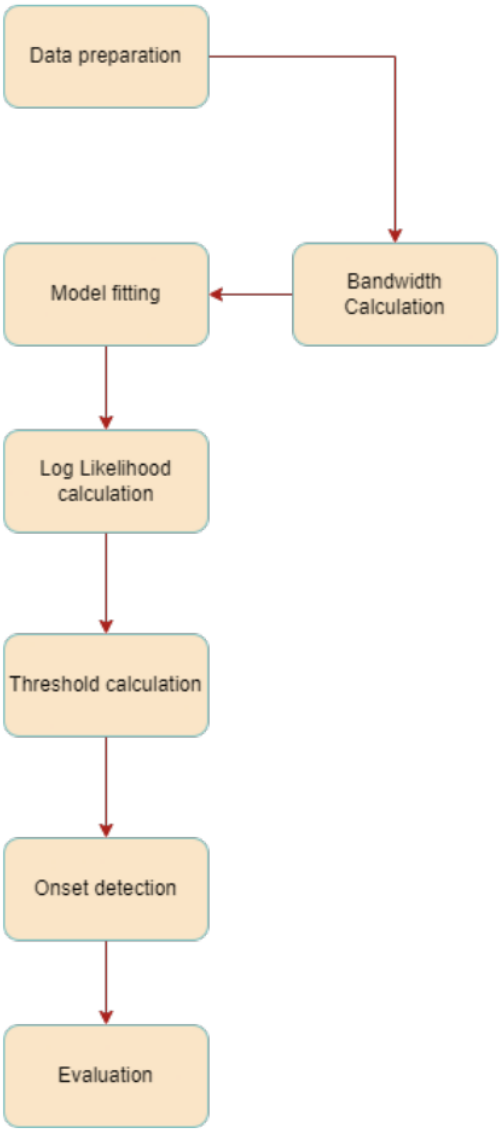


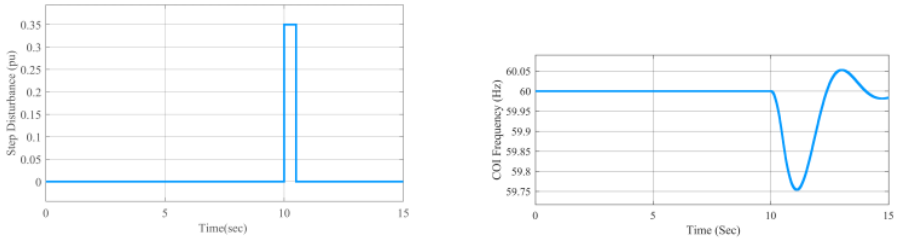
Figure 5.1: Flowchart of the implementation

CHAPTER 6

RESULTS

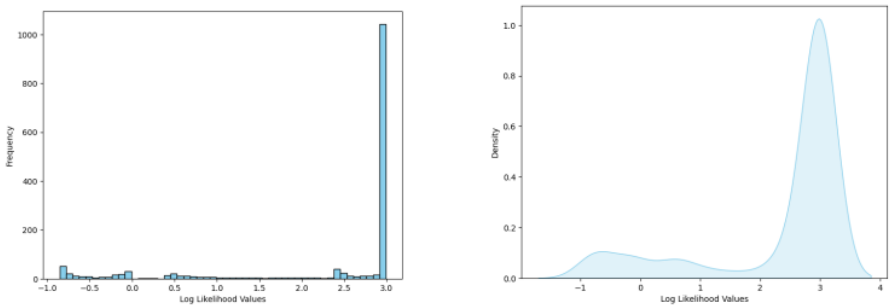
The implementation was divided into 3 cases.

In the first case, a 0.35-pu step disturbance was applied for 500 ms as shown in Figure 6.1(a), and the corresponding center of inertia frequency (COI frequency) vs. time graph was obtained as shown in Figure 6.1(b). The model was fit with the time series data of the center of inertia frequency. The model gave log-likelihood values, which are the logarithm of the estimated density at each data point. This log-likelihood represents the likelihood of the observed data under the estimated density model. The histogram plot as shown in Figure 6.2(a) and density plot as shown in Figure 6.2(b) of log-likelihood values help in finding an appropriate percentile and kernel for the model by providing insights into the distribution of log-likelihood values. For this case, the 25th percentile was chosen, and a plot was obtained as shown in Figure 6.3. This figure shows the same COI frequency vs. time plot, but with outliers marked in red and an arrow that points out the first anomalous point. This first anomalous point is the onset value for disturbance, which in this case was **10.24sec**. Further, the duration of the disturbance was changed to get better insight into the results. This change ranges from 100 ms to 500 ms.



(a) Step Disturbance (b) COI Frequency vs time

Figure 6.1: Step disturbance and COI Frequency signal plot when 0.35pu disturbance was applied for 500ms



(a) Histogram plot (b) Density plot

Figure 6.2: Histogram and Density plot of log likelihood values when 0.35pu disturbance was applied for 500ms

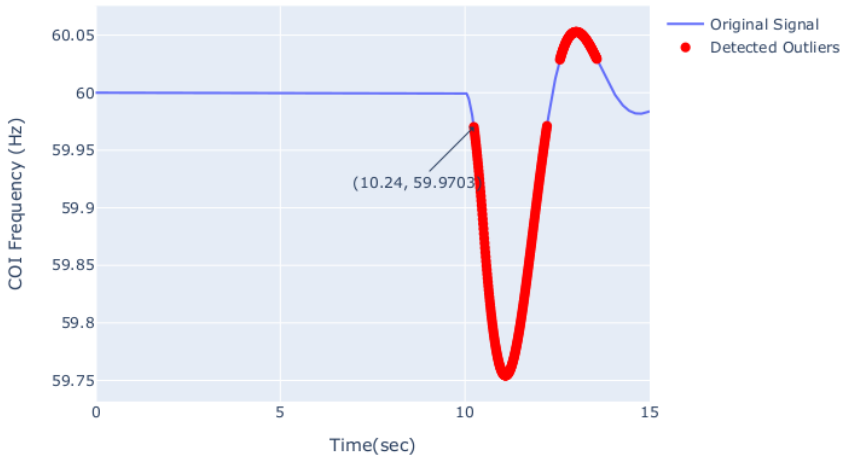
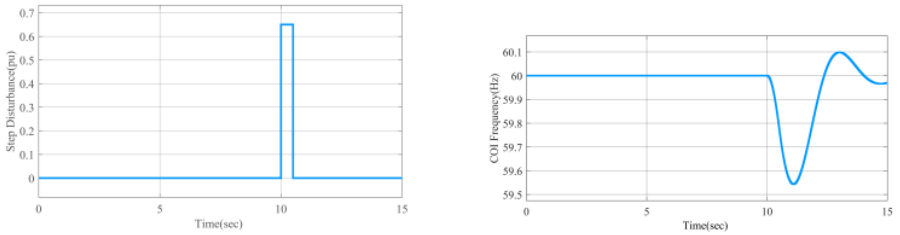


Figure 6.3: COI Frequency vs time plot with outliers marked as red when 0.35pu disturbance was applied for 500ms

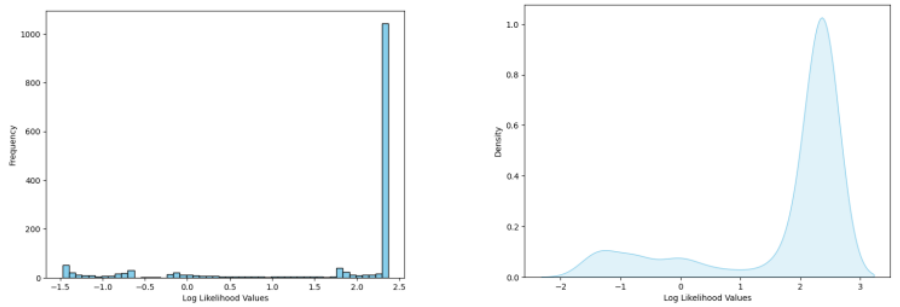
Similarly, in the second case, 0.65-pu disturbance was applied for 500ms shown in figure 6.4(a) and same insights were taken. In this case, the onset of disturbance was detected at **10.2sec**, a bit more earlier than previous case as shown in figure 6.6.

Finally, in the third case, 0.85-pu disturbance was applied for 500ms shown in figure 6.7(a). In this case, the onset of disturbance was detected at **10.19sec**, a bit more earlier than previous case as shown in figure 6.9.



(a) Step Disturbance (b) COI Frequency vs time

Figure 6.4: Step disturbance and COI Frequency signal plot when 0.65pu disturbance was applied for 500ms



(a) Histogram plot (b) Density plot

Figure 6.5: Histogram and Density plot of log likelihood values when 0.65pu disturbance was applied for 500ms

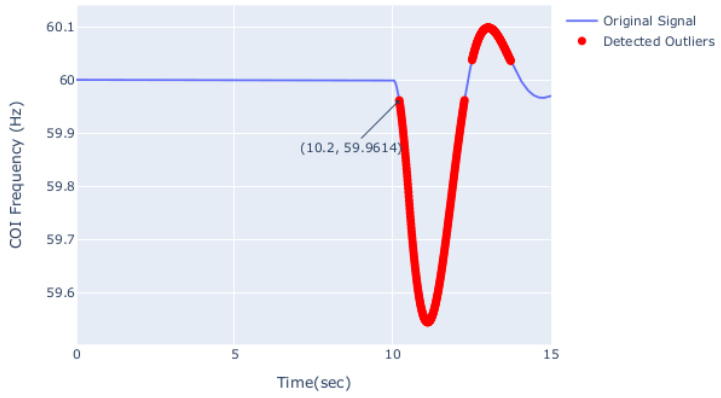
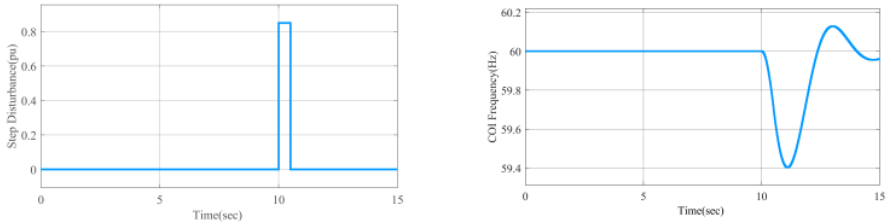
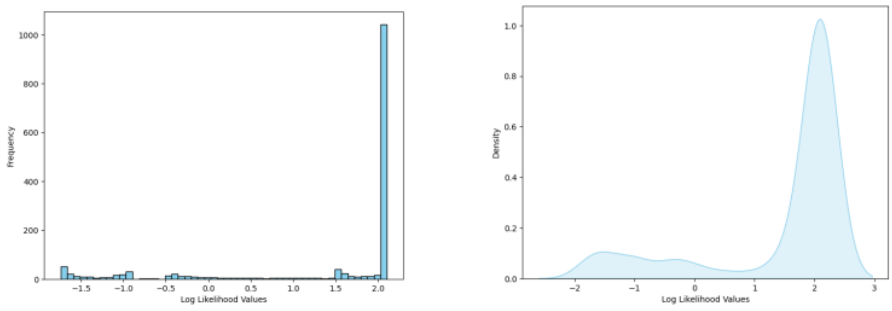


Figure 6.6: COI Frequency vs time plot with outliers marked as red when 0.65pu disturbance was applied for 500ms



(a) Step Disturbance (b) COI Frequency vs time

Figure 6.7: Step disturbance and COI Frequency signal plot when 0.85pu disturbance was applied for 500ms



(a) Histogram plot (b) Density plot

Figure 6.8: Histogram and Density plot of log likelihood values when 0.85pu disturbance was applied for 500ms

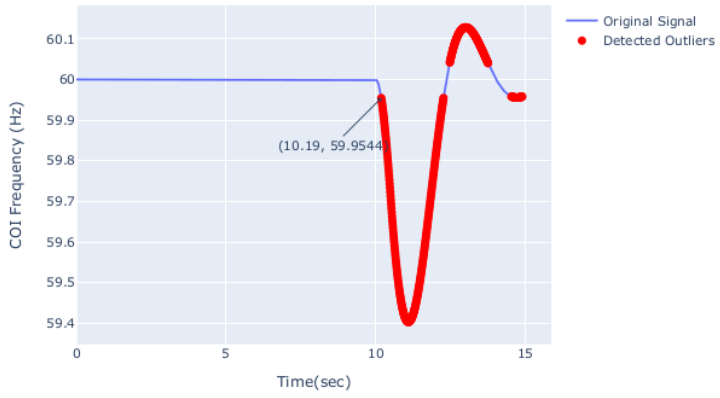


Figure 6.9: COI Frequency vs time plot with outliers marked as red when 0.85pu disturbance was applied for 500ms

Table 6.1 shows time at which onset was detected for each of the cases. From this table, conclusion can be drawn that as the magnitude of disturbance varied from 0.35 to 0.65 and finally to 0.85 the detection was done accurately, making the model robust from normal variations or noise in the data. It also tells us the importance of duration of disturbance, shorter the duration, the later the detection.

Step Disturbance (pu)	Time Duration (ms)				
	100	200	300	400	500
0.35	10.33s	10.3s	10.28s	10.26s	10.24s
0.65	10.3s	10.27s	10.24s	10.22s	10.2s
0.85	10.28s	10.25s	10.23s	10.21s	10.19s

Table 6.1: Onset disturbances predicted when different disturbance were applied for different time duration

Table 6.2 shows the percentage error calculated for each of the cases. The onset given during the simulation was 10 seconds. The percentage error value ranges from 1.9% to 3.3%. The average percentage error was around 2.5%.

Step Disturbance (pu)	Time Duration (ms)				
	100	200	300	400	500
0.35	3.3%	3%	2.8%	2.6%	2.4%
0.65	3%	2.7%	2.4%	2.2%	2%
0.85	2.8%	2.5%	2.3%	2.1%	1.9%

Table 6.2: Percentage error between actual onset disturbance and predicted onset disturbance

CHAPTER 7

CONCLUSION AND FUTURE WORK

In this work, the application of kernel density estimation (KDE) has been explored for the detection of the onset of disturbances in a power system, specifically focusing on estimating inertia. We collected simulated time series data from a 2-area system, comprising 15 signals with 1500 samples, each of which included the center of inertia frequency. KDE was applied to this data to estimate the likelihood of observing each data point under a Gaussian distribution, which helps in identifying anomalies or disturbances in the system. Through visualization and analysis of the log likelihood values, we were able to detect anomalies corresponding to disturbances in the system. Major insights were taken, like that as we increase the disturbance or duration of disturbance, detection becomes easier.

Moving forward, there are several avenues for further research and improvement in this work. Firstly, we plan to refine the threshold selection process for anomaly detection, possibly by incorporating domain knowledge or optimizing the threshold based on specific system characteristics. Additionally, we aim to explore the use of different kernel functions and bandwidth selection methods to improve the accuracy and robustness of the anomaly detection algorithm. Furthermore, we intend to validate the performance of the KDE-based anomaly detection approach on real-world data to assess its practical applicability and effectiveness. Finally, we plan to investigate the integration of other machine learning techniques and data preprocessing methods to enhance the overall performance of the anomaly detection system.

REFERENCES

[1] C. Phurailatpam, Z. H. Rather, B. Bahrani, and S. Doolla, "Measurement-based estimation of inertia in ac microgrids," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 3, pp. 1975–1984, 2020.

[2] T. Inoue, H. Taniguchi, Y. Ikeguchi, and K. Yoshida, "Estimation of power system inertia constant and capacity of spinning-reserve support generators using measured frequency transients," *IEEE Transactions on Power Systems*, vol. 12, no. 1, pp. 136–143, 1997.

[3] P. Wall and V. Terzija, "Simultaneous estimation of the time of disturbance and inertia in power systems," *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 2018–2031, 2014.

[4] C. Heinz and B. Seeger, "Cluster kernels: Resource-aware kernel density estimators over streaming data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 7, pp. 880–893, 2008.

[5] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, aug 1992. [Online]. Available: http://dx.doi.org/10.1002/9780470316849

[6] B. W. Silverman, "Density estimation for statistics and data analysis," 1986.

Turnitin Report_Aditya Kumar_09-05-2024

ORIGINALITY REPORT

7%	3%	5%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Chitaranjan Phurailatpam, Zakir Hussain Rather, Behrooz Bahrani, Suryanarayana Doolla. "Measurement Based Estimation of Inertia in AC Microgrids", IEEE Transactions on Sustainable Energy, 2019 Publication	1%
2	Maja Muftić Dedović, Adnan Mujezinović, Nedis Dautbašić, Ajdin Alihodžić, Adin Memić, Samir Avdaković. "Estimation of Power System Inertia with the Integration of Converter-Interfaced Generation via MEMD during a Large Disturbance", Applied Sciences, 2024 Publication	1%
3	Submitted to University of Edinburgh Student Paper	1%
4	"Learning from Data Streams in Evolving Environments", Springer Science and Business Media LLC, 2019 Publication	<1%
5	www.javatpoint.com	

	Internet Source	<1%
6	Sivasubramanian Ramanathan, Kulsoom Syed, Tejaswini Chavan. "Data-driven visual analytics of Human Mobility data and green cover using Image Processing for Smart Cities", 2023 3rd International Conference on Intelligent Technologies (CONIT), 2023 Publication	<1%
7	digi.lib.ttu.ee Internet Source	<1%
8	Aviti Thadei Mushi, Owdean Suwi, Jackson J. Justo. "Overview of renewable energy power system dynamics", Elsevier BV, 2024 Publication	<1%
9	"Artificial Intelligence Applications and Innovations", Springer Science and Business Media LLC, 2018 Publication	<1%
10	Submitted to NOVA School of Business and Economics Student Paper	<1%
11	arxiv.org Internet Source	<1%
12	Submitted to University College London Student Paper	<1%

13	posoco.in Internet Source	<1%
14	Christoph Heinz. "Cluster Kernels: Resource-Aware Kernel Density Estimators over Streaming Data", IEEE Transactions on Knowledge and Data Engineering, 07/2008 Publication	<1%
15	www.nber.org Internet Source	<1%
16	stax.strath.ac.uk Internet Source	<1%
17	www.studymode.com Internet Source	<1%
18	Hassan, Md. Rafiul, Kotagiri Ramamohanarao, Joarder Kamruzzaman, Mustafizur Rahman, and M. Maruf Hossain. "A HMM-based adaptive fuzzy inference system for stock market forecasting", Neurocomputing, 2013. Publication	<1%

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		