

Shape Preserving Incremental Learning for Power Systems Fault Detection

Jose Cordova[✉], Carlos Soto, Mostafa Gilanifar[✉], Yuxun Zhou, Anuj Srivastava,
and Reza Arghandeh[✉], *Senior Member, IEEE*

Abstract—This letter presents a shape preserving incremental learning algorithm that employs a novel shape-based metric called the Fisher–Rao amplitude-phase distance (FRAPD) metric. The combined amplitude and phase distance metric is achieved on a function space from the Fisher–Rao elastic registration. We utilize an exhaustive search method for selecting the optimal parameter that captures the amplitude and phase distance contribution in FRAPD when performing a clustering process. The proposed incremental learning structure based on the shape preserving FRAPD distance metric utilizes continuously updated fault shape templates with the Karcher mean. The seamless updating of abnormal events enhances the clustering performance for power systems fault detection. The algorithm is validated using the actual data from real-time hardware-in-the-loop testbed.

Index Terms—Shape-based data analysis, incremental learning, event detection, fault detection, power distribution networks.

I. INTRODUCTION

THE ADVENT of advanced measurement devices such as Phasor Measurement Units (PMU) provides higher resolution and granularity data from power systems. These advances introduce more opportunities for using various data analysis methods in different applications for detecting and analyzing complex events. Examples of these applications are: topology detection [1], fault detection [2]–[4], and network modeling [5] in power distribution systems. Classical methods such as the traveling wave-based [6] and the impedance-based

methods [7], [8] are among the most established methods for power systems fault detection. However, these methods depend heavily on prior knowledge of the network topology [9].

With recent advancements in machine learning and statistical inference, data-driven fault detection methods are gaining more attention. In the literature, Neural Networks (NN) [10] and Support Vector Machines (SVM) [11] are among the most used methods for fault detection in power systems. However, accurate fault detection requires retraining the model which may require a significant amount of time for convergence [9]. Using incremental learning can accelerate the time needed for convergence through continuous updating of the training dataset.

The shape-based data analysis typically examines different datasets in a function space that provides information on the data structure by preserving their characteristic shapes. In the problem of fault detection, the shape of fault events may be a key to better diagnose the type of fault events. Cordova *et al.* [3] proposed a fault clustering method using a shape-based distance with square-root velocity function (SRVF). The SRVF transforms the Riemannian space under the Fisher–Rao metric to a standard $\mathbb{L}^2$ norm which makes the geodesic calculations less complex. In that publication, similar fault events are clustered together by the amplitude-distance component obtained from the registration under the Fisher–Rao metric (RFRM) framework. However, the RFRM framework provides a second component, the phase distance, which ostensibly contributes to the representation of a signal in a Hilbert space.

In this letter, we propose a novel metric for joining both the amplitude and phase components resulting from the RFRM by weighting the joint distance component in a Hilbert space. The joint distance component preserves the shape of a signal during the clustering process. Additionally, we propose an approach to find the optimal amplitude-phase weighting parameters using an exhaustive search. The proposed Fisher–Rao amplitude-phase distance (FRAPD) metric utilizes both the Fisher–Rao amplitude distance (FRAD) and the Fisher–Rao phase distance (FRPD). In this letter, the FRAPD combines amplitude and angle shift information that are caused by fault occurrence on voltage and current measurements within power distribution networks.

Moreover, the use of RFRM allows the construction of *fault templates* utilizing the Karcher mean. The fault templates result from time-aligned signals and represent the clustered fault events. They preserve the characteristic phase angle shift and amplitude similarities between different fault types. The templates are incrementally updated as more faults occur.

Manuscript received March 10, 2018; revised May 25, 2018; accepted June 15, 2018. Date of publication July 2, 2018; date of current version July 20, 2018. This work was supported in part by the European Union's Horizon 2020 (ERIGrid) through the Trans-National Access Programme under Grant 654113, and in part by the U.S. DOE Grant RADIANCE GMLC. Recommended by Senior Editor F. Blanchini. (*Corresponding author: Jose Cordova.*)

J. Cordova is with the ECE Department, Florida State University, Tallahassee, FL 32310 USA (e-mail: jdc13b@my.fsu.edu).

C. Soto and A. Srivastava are with the Statistics Department, Florida State University, Tallahassee, FL 32310 USA.

M. Gilanifar is with the IME Department, Florida State University, Tallahassee, FL 32310 USA.

Y. Zhou is with the EECS Department, University of California at Berkeley, Berkeley, CA 94702 USA (e-mail: jdc13b@my.fsu.edu).

R. Arghandeh is with the Department of Computing, Mathematics and Physics, Western Norway University of Applied Sciences, 5020 Bergen, Norway.

Digital Object Identifier 10.1109/LCSYS.2018.2852064

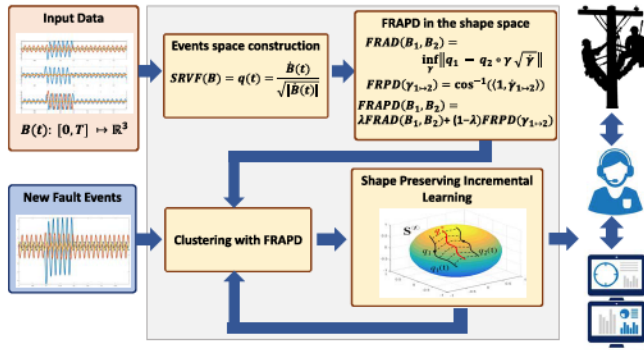


Fig. 1. Overview of the fault detection with elastic registration distances under Fisher-Rao metric.

The contributions and novelties in this letter are as follows:

- We propose a novel shape-based distance function on a Hilbert sphere space for data clustering called Fisher-Rao Amplitude-Phase Distance (FRAPD). The FRAPD combines the weighted amplitude and phase components from the RFRM outcomes under SRVF representation.
- We propose an exhaustive search optimization scheme to select optimal weighting parameters λ in FRAPD for more accurate clustering.
- We propose a shape preserving incremental learning algorithm for power system fault detection by applying our novel FRAPD metric and producing fault templates using the Karcher mean. The incremental updating of fault events enhances the data clustering performance.

This letter is structured as follows: in Section II, we present the computation of the shape-based distance under the Fisher-Rao metric. Additionally, the incremental learning algorithm based on the Karcher mean is explained. In Section III, we describe the real-time use case simulated to validate our framework. Discussions of the fault detection results are presented in Section IV.

II. ELASTIC REGISTRATION UNDER FISHER-RAO METRIC WITH SRVF REPRESENTATION

In this letter, we propose an incremental learning algorithm for fault detection that preserves the shape of fault signals using elastic registration (or alignment) under Fisher-Rao metric (RFRM). The primary objective is to cluster the measurement data streams from different electrical fault conditions using three-phase voltage and current measurements.

The overall architecture of the shape preserving incremental learning algorithm for fault detection is presented in Fig. 1. The components of the proposed methodology are described in five steps:

Step 1. Input Data: The training data sets are recorded fault events that include three-phase voltage and current measurements by PMU devices. Data streams can represent the fault magnitude or angle value (i.e., $V\angle\phi$) (See Section III).

Step 2. Events Space Construction: The fault data represented by the function space $V = V(t) : [0, T] \mapsto \mathbb{R}^3$ (i.e., 3-phase voltage) is then transformed by means of a square-root velocity function (SRVF) into a new event space. In this new event space, the Fisher-Rao metric becomes the L^2 norm. In other words, the SRVF transforms the event space (PMU streams) into a function space where the RFRM can be

implemented in a less complex manner. This will be further discussed in Section II-B.

Step 3. Calculate the Fisher-Rao Amplitude-Phase Distances in the Shape Space: A novel amplitude-phase based pairwise matrix distance is calculated by weighting the contribution of the amplitude and phase shape components of the data streams for the clustering problem. The weighted contribution of each component is tuned by a parameter λ defined in Section II-D.

Step 4. Clustering With Fisher-Rao Amplitude-Phase Distance: The clustering is performed by a hierarchical process applied to the amplitude-phase distance (FRAPD) matrix. As a result, similar events are grouped together for event detection purposes.

Step 5. Shape Preserving Incremental Learning: The fault event clusters are utilized for constructing fault templates that are the representative shapes of fault events. This step is performed by using the Karcher mean. Then, the fault templates created are continuously updated as new faults occur. These constant updates increase the performance of the algorithm in distinguishing different types of faults.

A. Events Space Construction

To perform elastic registration under the Fisher-Rao metric, we first need to transform the function space composed by data streams. The three-phase voltage and current data streams from a time-synchronized measurement device such as a distribution PMU are defined as follows:

$$\begin{aligned} V &= V(t) : [0, T] \mapsto \mathbb{R}^3 \\ I &= I(t) : [0, T] \mapsto \mathbb{R}^3. \end{aligned} \quad (1)$$

where $V = [V_a(t), V_b(t), V_c(t)]'$ and $I = [I_a(t), I_b(t), I_c(t)]'$ with $t \in [0, T]$ and T being the event's length. An event space is defined as a subset of functions $B : [0, T] \mapsto \mathbb{R}^3$ matched with necessary smoothness and integrability conditions. The space $\mathcal{F}$ and the elements of that space, i.e., the signals of length T , are denoted as $B \in \mathcal{F}$. We utilize a functional data analysis approach to interpret the events as shapes. Following the shape-based methodology in [12], we compute the event shape space $\mathcal{S}$ by interpreting the space $\mathcal{F}$ as a Riemannian manifold with the Fisher-Rao Riemannian metric [12]. The event space $\mathcal{F}$ is transformed to a pre-shape space $\mathcal{S}$ with the Square Root Velocity Function operator denoted by $SRVF(\cdot)$. The elements of $\mathcal{S}$ are then denoted by q which is defined as:

$$SRVF(B) = q(t) = \dot{B}(t) / \sqrt{|\dot{B}(t)|}. \quad (2)$$

Consequently, we define a space of the equivalence classes of q under warping scale ($\gamma(t)$) in the time domain of the event signals. The warping function γ is defined as an operator function. Given a pair of functions f_1 and $f_2 \in \mathcal{F}$, we say that $f_1(\gamma(t))$ and f_2 are well aligned. Here, space is a quotient space $\mathcal{S}/\Gamma$, where Γ comprises the set of all domain warping functions $\gamma(t)$. In practice, instead of working with the full space of Γ , we restrict the set to all continuous, weakly increasing boundary preserving warping functions.

B. Fisher-Rao Amplitude-Phase Distances in the Shape Space

The methodology we propose seeks to combine the amplitude and phase metric components into a single weighted

metric that captures the angle shift for different fault events in power systems. This section describes how to calculate the Fisher-Rao amplitude and phase distance (FRAPD) in the event shape space $\mathcal{S}$ using the proposed framework.

The SRVF representation has the advantage of inducing a transformation of the Fisher-Rao Riemannian metric to a standard $\mathbb{L}^2$ norm. We define $\mathbb{L}^2([0, 1], \mathbb{R})$ (or simply $\mathbb{L}^2$) as the set of all square integrable functions, which are SRVFs in our case. Then, for every $q \in \mathbb{L}^2$ there is a function B that has a translation so that $\text{SRVF}(B) = q$. Using the chain rule, it can be shown that by warping a function f using γ , the SRVF of $f \circ \gamma$ becomes:

$$\tilde{q}(t) = (q \circ \gamma)(t) \sqrt{\dot{\gamma}(t)}. \quad (3)$$

where the $\circ$ operator, from here on, denotes a composition function. There is $f(t) \circ \gamma(t) = f(\gamma(t))$ for any given $f(t)$ and $\gamma(t)$. The variable $\tilde{q}$ represents the SRVF of $f \circ \gamma$. This transformation, a group action, will be denoted by $(q, \gamma) = (q \circ \gamma) \sqrt{\dot{\gamma}}$. Using the theory presented in [13], for any two SRVFs $q_1, q_2 \in \mathbb{L}^2$ and any $\gamma \in \Gamma$, we have $\|(q_1, \gamma) - (q_2, \gamma)\| = \|q_1 - q_2\|$. We also have $\|(q_1, \gamma)\| = \|q_1\|$ which is the group action of time warping acting isometrically on elements of the SRVF space. This is another advantage of using the Fisher-Rao metric. In general, $\|(f, \gamma)\| \neq \|f\|$ under the $\mathbb{L}^2$ metric.

Restricting all curves to have the same length allows calculation of the distance between curves in the event space by merely computing the arc-length between two points on the Hilbert sphere. Thus, we restrict the space to $\{q \in \mathbb{L}^2 \mid \int |q|^2 dt = 1\}$ which is an infinite-dimensional Hilbert sphere. This restriction can be easily imposed by scaling each q by dividing by its respective length. Then, the distance between any two events $[B_1, B_2]$, which have been transformed to q_1 and q_2 in the sphere $\mathbb{S}^\infty$ is defined as:

$$\text{FRAD}(B_1, B_2) = \inf_{\gamma} \|q_1 - q_2 \circ \gamma \sqrt{\dot{\gamma}}\|. \quad (4)$$

We will follow the interpretation in [14] and denote it as the *amplitude distance* or *Fisher-Rao amplitude distance* (FRAD). In particular: $\text{FRAD}(B(t)) = \text{FRAD}(B(\gamma(t)))$, where $\gamma : [0, T] \mapsto [0, T]$ is a warping (orientation-preserving diffeomorphism). Therefore, we define an amplitude of a function B as an equivalence class under all time warping:

$$\text{FRAD}(B) = \{B \circ \gamma \mid \gamma : [0, T] \mapsto [0, T], \gamma \in \mathbb{S}^\infty([0, T])\}. \quad (5)$$

The existence of an amplitude difference naturally leads to a *phase difference* in the function space $\mathcal{S}$. The phase difference is the amount of necessary warping γ to align the curve B_1 to B_2 to minimize the amplitude-distance between them after the SRVF transformation. Notice that this notion of amplitude and phase used for time warping should not be confused with the phasor values of voltage and current in power systems. In this context, amplitude and phase represent geometrical parameters of mapped data points on the Hilbert sphere space.

In this letter, we utilized the R package “fdaSRVF” [15] to find an optimal γ function which utilize a dynamic programming approach as follows:

$$\gamma_{1 \rightarrow 2} = \arg \min_{\gamma} d_{\text{SRVF}}(q_1, q_2 \circ \gamma \sqrt{\dot{\gamma}}), \quad (6)$$

where γ represents the optimal warping. The notation $\gamma_{1 \rightarrow 2}$ shows that the warping action is performed from curve 1 to

Algorithm 1 Distance Calculations Under SRVF Framework

Input: Signals B_i with $i = 1, 2, \dots, n$.

Output: Amplitude and phase components distances: $\text{FRAD}(B_i, B_j)$ and $\text{FRPD}(\gamma_{i \rightarrow j})$.

Initialization :

- 1: Define $B_i(t) : [0, T] \mapsto \mathbb{R}^3, i = 1, 2, \dots, n$.
- 2: **for** $i, j = 1, 2, \dots, n$. **do**
- 3: Calculate $\text{SRVF}(B_i)$ and $\text{SRVF}(B_j)$ with (2)
- 4: Compute $\gamma_{i \rightarrow j}$ with (6)
- 5: $\text{FRAD}(B_i, B_j)$ with (4)
- 6: Compute $\text{FRPD}(\gamma_{i \rightarrow j})$ with (7)
- 7: **end for**
- 8: **return** $\text{FRAD}(B_i, B_j)$ and $\text{FRPD}(\gamma_{i \rightarrow j})$

curve 2. The distance in the SRVF function space is represented by d_{SRVF} . The *phase distance* or *Fisher-Rao phase distance* (FRPD) quantifies the amount of warping necessary for alignment. It is measured as the distance from the identity warping function γ_{id} (i.e., the γ function performing no warping) as follows:

$$\text{FRPD}(\gamma_{1 \rightarrow 2}) = \cos^{-1}(\langle \gamma_{id}, \sqrt{\dot{\gamma}_{1 \rightarrow 2}} \rangle). \quad (7)$$

Hence, for every signal B_1 and B_2 , two distances will be calculated: the amplitude distance $\text{FRAD}(B_1, B_2)$ and the phase distance $\text{FRPD}(\gamma_{1 \rightarrow 2})$. The RFRM is summarized in algorithm 1.

C. Clustering With Fisher-Rao Amplitude-Phase Distance

The amplitude (FRAD) and phase distances (FRPD) calculated in the previous section are then weighted to construct a combined metric (FRAPD) that captures both components resulting from the elastic registration under the Fisher-Rao metric. Using the FRAPD, the FRPD and FRAD distances mutually play a role in the classification of electrical faults as follows:

$$\text{FRAPD}(B_1, B_2) = \lambda * \text{FRAD}(B_1, B_2) + (1 - \lambda) * \text{FRPD}(\gamma_{1 \rightarrow 2}). \quad (8)$$

It is clear that for any two proper distances d_1 and d_2 and nonnegative constants a and b , $d = ad_1 + bd_2$ is also a proper metric. Therefore, the FRAPD is a valid metric. We have $\lambda \in [0, 1]$ where $\lambda = 0$ gives us pure phase distance and $\lambda = 1$ gives us pure amplitude distance. To tune the weighting parameter, the optimal value of λ is found by an exhaustive search in the training process of the fault detection algorithm.

D. Shape Preserving Incremental Learning

After clustering the faults using FRAPD, we update the labeled fault dataset $B_i(t) : [0, T] \mapsto \mathbb{R}^3, i = 1, 2, \dots, n$, with a single representative fault template (shape) that will be utilized in the clustering process. The constructed fault templates are continuously updated as new events occur in the electric grid as it is illustrated in Fig. 1. This incremental learning process improves the knowledge of the algorithm in detecting different types of faults.

The representative fault template is computed with the Karcher mean $\mu_{\mathcal{F}} = \int_0^T \mu_q(v) |\mu_q(v)| dv$ as shown in algorithm 2. As opposed to the commonly used arithmetic mean, the Karcher mean is an intrinsic measure of center on the

Algorithm 2 Computing Karcher Mean of Amplitudes Under SRVF Framework

Input: Signals B_i with $i = 1, 2, \dots, n$.

Output: Karcher Mean μ_{KM}

Initialization:

- 1: Define $B_i(t) : [0, T] \mapsto \mathbb{R}^3$, $i = 1, 2, \dots, n$. Calculate $q_i = \text{SRVF}(B_i)$. Let $\mu_0 = \mu_q = q_j$ be the initialization of the Karcher mean for some $q_j \in \{q_i\}$
- 2: **while** $\|\mu_q - \frac{1}{n} \sum q_i\| > \epsilon$ **do**
- 3: Align each q_i to the current Karcher mean estimate with (6)
- 4: Set $\mu_q = \frac{1}{n} \sum q_i$
- 5: **end while**
- 6: Compute Karcher mean by $\mu_{\mathcal{F}} = \int_0^T \mu_q(v) |\mu_q(v)| dv$
- 7: **return** $\mu_{\mathcal{F}}$

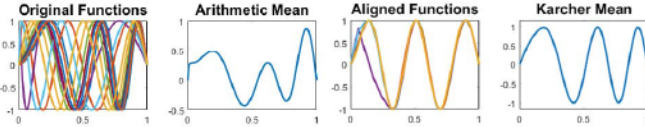


Fig. 2. Comparison of Arithmetic Mean and Karcher Mean.

manifold on which the curves lies. If all curves are bimodal, for instance, the Karcher mean is bimodal as well. In general, the arithmetic mean does not guarantee this property. In other words, the shape of the curves will be preserved by the Karcher mean as shown in Fig. 2. Note that in this graphical example the original functions are sinusoidal curves. The arithmetic mean does not capture the representative shape of the original curves in either height or structure, while the Karcher means does. At first glance, it may look similar to an arithmetic mean, and that is because the SRVF space is flat. A key component of our proposed algorithm is that time warping is done at every iterative step which acts isometrically in the SRVF space.

III. CASE-STUDY

In this section, we describe the datasets and the grid models used for validating our proposed methodology. For the first case study, we have modeled the IEEE 13-nodes test feeder in Real Time Digital Simulator (RTDS) that resembles real-field conditions that come inherently with distribution networks (see Fig. 3). The hardware setup for this feeder can be found in [16].

As it is shown in Fig. 3, we have placed three PMUs on the secondary side of the feeder (node 632) to resemble the scarce presence of monitoring devices in power distribution systems. It is a common practice to put monitoring devices at the start of the feeder measuring currents and voltages. For this case study, the data streams correspond to the magnitudes of both voltage and current measured by the PMUs. Fault events have been placed in three different locations, nodes 645, 691, and 680, as shown in Fig. 3. The events were created with a duration of 0.2 seconds (approx. 12 cycles). Fault impedances for each fault type were changed following a random distribution with a low impedance value ranging between 0.01 and 0.015 ohms.

In our real-time RTDS simulations setup, events for each possible phase combination of the 11 fault types (shown in Fig. 4) were simulated 50 times. This gives a total of 550 events simulated using this realistic model for each location. Furthermore, having three different locations where faults are applied gives a total number of 1650 observations. The phases

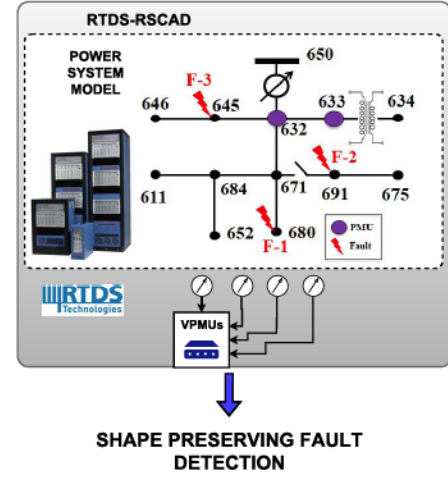


Fig. 3. Experimental setup of IEEE 13-Nodes test feeder model in RTDS/RSCAD.

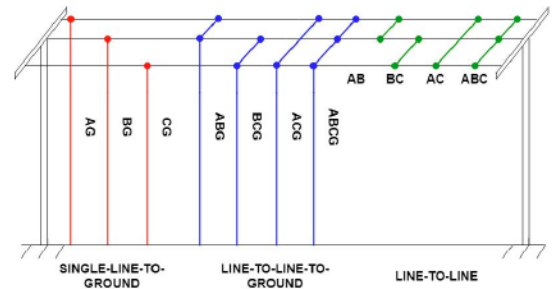


Fig. 4. Fault types simulated in RTDS/RSCAD. Note that fault impedance is not shown for easier visualization.

were labeled as A, B, and C with ground represented by G. Fig. 5 shows an example of a single-line-to-ground (AG) as seen from the feeder node, where the measurement device was located.

To further validate the proposed algorithm, we have performed fault event detection in the IEEE 37-nodes test feeder as a second case study. The fault events simulated with this model were performed in a hardware-in-the-loop setup inside an RT-LAB environment developed by Opal-RT. For more details on the technical setup of this model, we refer the reader to [17]. In a similar setup as with the 13-nodes test feeder, we have placed fault events in three different locations with random fault impedances and one PMU monitoring the feeder. The data streams corresponding to the magnitude and angle of both voltage and current phasors were measured by the PMU. We emulated three different fault types (AG, AB, and ABCG) with 99 events each for a total of 297 fault events. This gives a total of 891 fault events for this feeder.

IV. RESULTS AND DISCUSSION

In this section, we validate the proposed FRAPD metric and we implement the shape preserving incremental learning algorithm for fault detection in power systems using two different datasets achieved from different digital real-time simulators testbeds. The objective of each experimental setup is to accurately cluster together similar fault events by their specific type depicted in Fig. 4. Fault clusters are obtained by a pairwise distance matrix that is based on the novel FRAPD metric introduced in Section II.

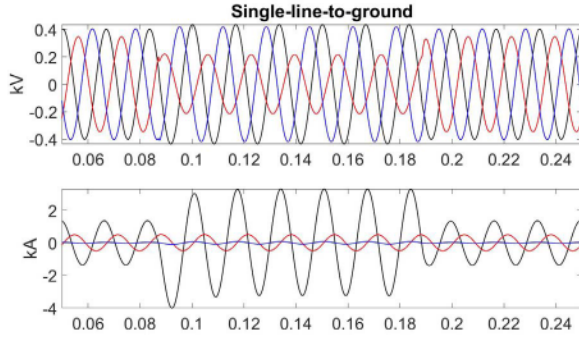


Fig. 5. Single-line-to-ground at node 680 in IEEE 13-nodes test feeder.

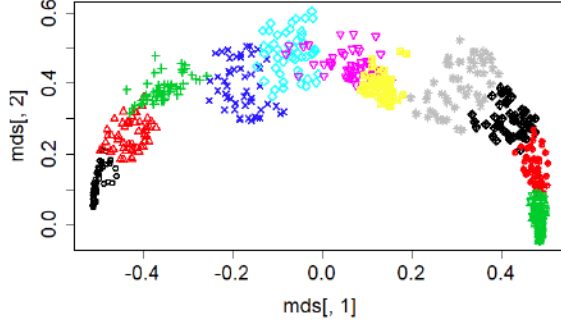


Fig. 6. Clustering 11 fault types at location 680 in the IEEE 13-nodes test feeder.

To better visualize the scatter patterns of clustered data in two dimensions, Fig. 6 shows the multidimensional scaling (MDS) plot resulting from the fault detection at node 680 of the IEEE 13 nodes test feeder. In general, MDS is a tool to reproduce the observed distances in any number of directions for data clusters. The orientations are arbitrary and they are useful for visualizing clusters from an N-dimensional distance matrix. The different clusters have been colored by their true label for better visualization (i.e., 11 fault types). This figure illustrates the weighted clusters under the proposed FRAPD metric, that is, the inclusion of both FRAD and FRPD components of the elastic registration-based algorithm.

To show the accuracy of fault detection using the proposed FRAPD with the incremental learning method, the multivariate accuracy (mACC) for the IEEE 13-nodes model is shown in Fig. 7. The prediction accuracy rate is shown by the diagonal bars while the false positive and negative errors are represented by the off-diagonal bars. Fig. 7a shows the mACC as a result of considering the FRAD only. It can be observed that the FRAD presents accurate results for line-to-line fault classification. Fig. 7b shows the results for fault detection utilizing the FRPD. The results obtained for single-line-to-ground and three-line-to-ground faults are observed to be consistent. The off-diagonal misclassified elements can be explained in Fig. 6. It can be observed that some of the clusters may be too close to the others (have similar shapes). Therefore, there is an overlap of some of the fault events distances over other types of faults. The proposed methodology using novel FRAPD is illustrated in Fig. 7c. This figure shows that the accuracy represented by the diagonal improves significantly concerning the previously mentioned methods.

We summarize the misclassification rate of the FRAPD clustering method for the IEEE 13-nodes test feeder as shown in

TABLE I

CASE I: CONFUSION MATRIX FOR FAULT CLASSIFICATION IN IEEE 13-NODES TEST FEEDER (VALUES ARE IN PERCENTAGE)

Method	FRAPD	B-FRAPD	FRAD	FRPD	SVM	NN
False Positive	6.06	12.54	9.09	7.58	29.82	21.1
False Negative	3.33	9.27	44.36	29.82	29.88	28.33
Total Error	9.39	21.82	53.45	37.39	59.70	50.38

TABLE II

CASE II: CONFUSION MATRIX FOR FAULT CLASSIFICATION IN IEEE 37-NODES TEST FEEDER (VALUES ARE IN PERCENTAGE)

Method	FRAPD	B-FRAPD	FRAD	FRPD	SVM	NN
False Positive	2.69	13.43	0.00	26.93	18.52	33.34
False Negative	1.68	10.25	25.93	0.00	16.16	33.33
Total Error	4.37	23.68	25.93	26.93	34.68	66.67

Table I. Results show that utilizing both the FRAD and FRPD components as a compound FRAPD metric improves the fault detection mACC up to 90.6%. As depicted in Table I, the false positive and false negative errors are considerably low with 6.06% and 3.33% respectively. The use of the combined FRAPD components provides a considerable relative improvement of 75% and 83% over FRAD and FRPD methodologies respectively.

Additionally, to find the optimal value of λ in the FRAPD metric (as shown in Eq. (8)), we perform different optimization approaches. The exhaustive search algorithm outperforms the other methods such as the Bayesian linear regression (B-FRAPD) approach by 57% as enumerated in Table I.

Moreover, we compared the FRAPD classification accuracy with the state-of-the-art machine learning methods. Classification based on a support vector machine (SVR) was configured with the parameter tuning function *tune* from [18]. Also, we have trained a Neural Network (NN) with learning one hidden layer and a threshold of 0.01 configured with [19]. Both machine learning methods work with $B_i(t) : [0, T] \mapsto \mathbb{R}^3, i = 1, 2, \dots, n$ as the signals to perform the classification. As shown in Table I, the proposed FRAPD-based methodology outperforms both SVM and NN, improving the classification accuracy by 85% and 82% respectively.

We implemented our FRAPD-based fault detection algorithm on a second dataset from the IEEE 37-nodes test feeder for more comprehensive validation. This model was discussed in Section III. Table II shows the confusion matrix for the fault classification using IEEE 37-nodes test feeder. It demonstrates that the results are consistent with respect to the 13-nodes test feeder. Results show that the use of FRAPD improves the accuracy as compared to the FRAD and FRPD methods by 83% and 84% respectively. Additionally, we compared our methodology with the B-FRAPD, and machine learning methods SVM and NN. Once again, our proposed methodology outperforms SVM and NN by more than 82% relatively. A crucial observation is that the misclassification total error in this case study is less than the misclassification total error resulting from the 13-nodes test feeder. This is expected, as the 37-nodes test feeder data contains the electrical phase angle along with the magnitudes for voltage and current measurement from the PMUs. The PMU phasor measurements give the shape preserving algorithm more information regarding voltage and current signals.

To show the incremental learning algorithm impact in the fault detection performance, we compared the fault detection

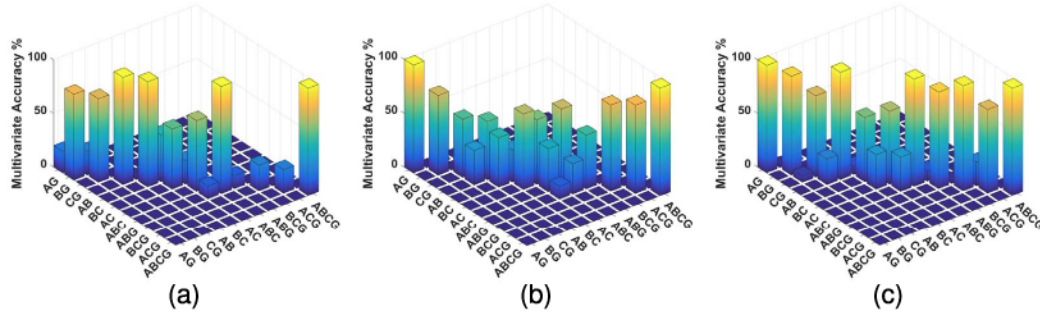


Fig. 7. Confusion Matrix for different methods for the IEEE 13-nodes test feeder. Diagonal terms are correct identifications and off-diagonal ones are misclassifications. mACC for multi-class detection accuracy: (a) Clustering with FRAD; (b) Clustering with FRPD; (c) Clustering with FRAPD distance.

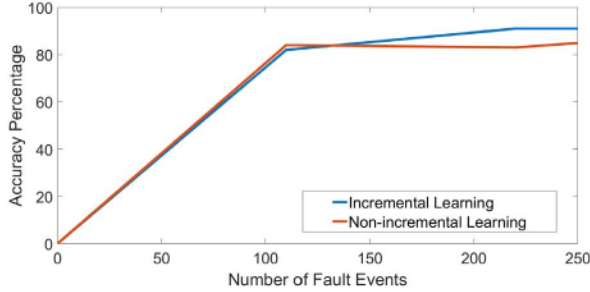


Fig. 8. Incremental algorithm average accuracy percentage (training dataset) for 300 fault events at nodes 645, 680, and 691 in the IEEE 13-Nodes test feeder.

accuracy using the incremental and non-incremental learning approaches as shown in Fig. 8. This figure illustrates that the incremental learning algorithm needs a lesser number of training data to get to 91% overall accuracy as compared to the non-incremental learning. In fact, the incremental learning reaches higher accuracy after 100 samples as opposed to the non-incremental methodology.

V. CONCLUSION

This letter proposes a shape preserving incremental learning fault detection algorithm using Fisher-Rao elastic registration. The new Fisher-Rao amplitude-phase distance metric (FRAPD) was utilized to perform clustering of fault signals with similar characteristics. The results show that the FRAPD shape-based clustering outperformed state-of-the-art machine learning techniques such as SVM and NN by reducing fault detection error up to 85%. Moreover, the combined FRAPD was compared to utilizing the amplitude and phase distance components individually as a shape-based clustering metric. Results show that the FRAPD-based classification improved 75% and 83% respectively.

ACKNOWLEDGMENT

The authors would like to thank Matthias Stifter, Thomas Strasser, and the AIT Austrian Institute of Technology, Vienna for providing laboratory infrastructure. The technical supports from Artemes, PSL, Opal-RT, and Siemens companies are acknowledged.

REFERENCES

- [1] G. Cavarro and R. Arghandeh, "Power distribution network topology detection with time-series signature verification method," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 3500–3509, Jul. 2018.
- [2] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Online learning of contextual hidden Markov models for temporal-spatial data analysis," in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, Las Vegas, NV, USA, 2016, pp. 6335–6341.
- [3] J. Cordova *et al.*, "Shape-based data analysis for event classification in power systems," in *Proc. IEEE Manchester PowerTech*, Manchester, U.K., Jun. 2017, pp. 1–6.
- [4] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Trans. Smart Grid*, to be published, doi: 10.1109/TSG.2017.2681962.
- [5] A. Tbaileh *et al.*, "Graph trace analysis: An object-oriented power flow, verifications and comparisons," *Elect. Power Syst. Res.*, vol. 147, pp. 145–153, Jun. 2017.
- [6] A. Farughian, L. Kumpulainen, and K. Kauhaniemi, "Review of methodologies for earth fault indication and location in compensated and ungrounded MV distribution networks," *Elect. Power Syst. Res.*, vol. 154, pp. 373–380, Jan. 2018.
- [7] M. Hosseini and V. Talavat, "Comparison of impedance based and travelling waves based fault location methods for power distribution systems tested in a real 205-nodes distribution feeder," *Trans. Elect. Electron. Mater.*, vol. 19, no. 2, pp. 123–133, 2018.
- [8] S. Zhang, S. Lin, Z. He, and W.-J. Lee, "Ground fault location in radial distribution networks involving distributed voltage measurement," *IET Gener. Transm. Distrib.*, vol. 12, no. 4, pp. 987–996, Feb. 2018.
- [9] S. S. Gururajapathy, H. Mokhlis, and H. A. Illias, "Fault location and detection techniques in power distribution systems with distributed generation: A review," *Renew. Sustain. Energy Rev.*, vol. 74, pp. 949–958, Jul. 2017.
- [10] M. Valtierra-Rodriguez, R. de Jesus Romero-Troncoso, R. A. Osornio-Rios, and A. Garcia-Perez, "Detection and classification of single and combined power quality disturbances using neural networks," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2473–2482, May 2014.
- [11] Y. Zhou *et al.*, "Abnormal event detection with high resolution micro-PMU data," in *Proc. 19th Power Syst. Comput. Conf.*, Genoa, Italy, 2016, pp. 1–7.
- [12] A. Srivastava and E. Klassen, *Functional and Shape Data Analysis*. New York, NY, USA: Springer-Verlag, 2016.
- [13] A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. S. Marron, "Registration of functional data using Fisher–Rao metric," *arXiv preprint arXiv:1103.3817*, 2011.
- [14] A. Srivastava, W. Wu, S. Kutek, E. Klassen, and J. S. Marron, "Statistical analysis and modeling of elastic functions," *arXiv preprint arXiv:1103.3817*, 2011.
- [15] J. D. Tucker. (2016). *fdasrvf: Elastic Functional Data Analysis, R Package Version 1.6.1*. [Online]. Available: <https://CRAN.R-project.org/package=fdasrvf>
- [16] W. H. Kersting, "Radial distribution test feeders," in *Proc. IEEE Power Eng. Soc. Win. Meeting*, vol. 2. Columbus, OH, USA, 2001, pp. 908–912.
- [17] M. Stifter, J. Cordova, J. Kazmi, and R. Arghandeh, "Real-time simulation and hardware-in-the-loop testbed for distribution synchrophasor applications," *Energies*, vol. 11, no. 4, p. 876, 2018.
- [18] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), R Package Version 1.6-8*, document e1071, TU Wien, Vienna, Austria, 2017.
- [19] S. Fritsch and F. Guenther. (2016). *neuralnet: Training of Neural Networks, R Package Version 1.33*. [Online]. Available: <https://CRAN.R-project.org/package=neuralnet>