

EGFC: Evolving Gaussian Fuzzy Classifier from Never-Ending Semi-Supervised Data Streams – With Application to Power Quality Disturbance Detection and Classification

Daniel Leite
Federal University of Lavras
Department of Automatics
Lavras, Brazil
daniel.leite@ufla.br

Leticia Decker
University of Bologna
INFN Bologna
Bologna, Italy
leticia.deckerde@unibo.it

Marcio Santana, and Paulo Souza
Federal Center for Technological Education
CEFET-MG
Minas Gerais, Brazil
{marciosantana, paulovitor}@cefetmg.br

Abstract—Power-quality disturbances lead to several drawbacks such as limitation of the production capacity, increased line and equipment currents, and consequent ohmic losses; higher operating temperatures, premature faults, reduction of life expectancy of machines, malfunction of equipment, and unplanned outages. Real-time detection and classification of disturbances are deemed essential to industry standards. We propose an Evolving Gaussian Fuzzy Classification (EGFC) framework for semi-supervised disturbance detection and classification combined with a hybrid Hodrick-Prescott and Discrete-Fourier-Transform attribute-extraction method applied over a landmark window of voltage waveforms. Disturbances such as spikes, notching, harmonics, and oscillatory transient are considered. Different from other monitoring systems, which require offline training of models based on a limited amount of data and occurrences, the proposed online data-stream-based EGFC method is able to learn disturbance patterns autonomously from never-ending data streams by adapting the parameters and structure of a fuzzy rule base on the fly. Moreover, the fuzzy model obtained is linguistically interpretable, which improves model acceptability. We show encouraging classification results.

I. INTRODUCTION

Power system disturbance detection and classification are broad and difficult research issues. Detection and classification are generally handled by using machine learning and computational intelligence methods [1]–[3]. The number of correlated attributes potentially carrying important information about disturbances is high. As real-time monitoring of all attributes is infeasible, a method to extract the most prominent ones to assist detection is fundamental. A further issue concerns the frequent occurrence of new situations. The emergence of new patterns in power system data affects the performance of classification systems. Novelty may arise since the involved signals are mutually related and time-varying. Moreover, the superposition of different events in different intensities may generate never-before-seen data, which tends to confuse conventional offline-trained classification systems [4], [5].

Online disturbance-detection methods deal with the occurrence of new patterns in data streams. New behaviors

should be captured by adaptive or evolving models, namely, models supplied with incremental learning algorithms [6]–[9]. Generally, the amount of data in power applications is large. Therefore, storing the data for further offline development of equation-based or intelligent models, and statistical analyses, is quite often time-consuming, useless, or even impossible. Adaptive and evolving modeling from online data streams are distinct concepts. *Adaptive models* (parametrically adaptive models from the control theory) are suitable to cope with smooth, gradual changes of system parameters and statistical properties of data (concept drift). However, when an adaptive model is changed to learn a new behavior, the knowledge about some previous behaviors tends to be partially lost. This phenomenon is known as *catastrophic forgetting* [10]. Abrupt changes of the values of a variable or parameter (concept shift) require both parametrical and structural adaptation of models. Such higher level of flexibility of models, which includes the incremental update of the model structure, outlines a broad research area known as *evolving intelligence* [11]–[14].

Evolving fuzzy rule-based models are appropriate for online detection and classification in nonstationary data stream environments, such as that found in power systems [5], [6], [8]. Incremental fuzzy clustering algorithms have been used for constructing rule-based evolving classifiers. These algorithms are capable of determining the classifier structure and parameters from scratch based on online data. Development and use of rule-based evolving systems have grown in the last decade. Successful applications of these systems in complex real-world problems, including control, prediction, classification, identification, and function approximation, are found [4], [11]–[15]. A further advantage of evolving fuzzy systems is that they may provide linguistically-appealing granular information [13], [14], [16], that is, these systems may explain their results or actions. Online structural adaptation of a fuzzy model to handle nonstationarities is pursued by adding, merging, and removing rules from a knowledge base [4].

This paper addresses a new fuzzy modeling framework

we called Evolving Gaussian Fuzzy Classification (EGFC) framework. EGFC is a semi-supervised variation of evolving granular rule-based approach [13], [17] for the construction of nonlinear and time-varying classifiers – being unsupervised and supervised learning the boundary cases. We aim to detect and classify anomalies, broadly speaking, and power quality disturbances as a particular application example. An EGFC model employs Gaussian membership functions to associate numerical input data with classes. Its incremental learning algorithm provides a dynamic classifier with simple math and linguistic rules describing its decisions. The set of EGFC rules represents the essence of a data stream. From a point of view, the EGFC approach consists in looking to stream data, and deciding between coarser or more detailed granules to achieve a better classification accuracy, and provide decision making support. Spikes, notching, harmonics, and oscillatory-transient types of disturbances in power systems are taken into consideration in the present study.

A hybrid method to extract highly discriminative attributes to be used as inputs of the EGFC model is also addressed. The method combines a Hodrick-Prescott (HP) filter [18] and Discrete Fourier Transform (DFT) [19] applied over a landmark window of voltage data to provide attributes that help disturbance classification. In particular, the method is different from those addressed in related power-system literature as HP analysis provides a smooth nonlinear representation that is sensitive to long and short-term changes. In other words, a series of smooth nonlinear trends are obtained, whereas the information removed from the original data is maintained separately and can be accessed for analysis. In comparison to S-transform-based attribute extraction methods [2], [20], often used in related literature, the proposed hybrid HP-DFT method has proven to be faster and effective – important characteristics for high-frequency data-stream processing.

The rest of this paper is organized as follows. Section II describes the HP and DFT methods for attribute extraction. Section III presents the data-stream-based semi-supervised learning algorithm and the Evolving Gaussian Fuzzy Classifier, EGFC, proposed for a broad class of classification and online anomaly-detection problems from numerical data. Section IV describes the methodology for generating the data and developing the classifier. Results are given in Section V. The conclusion is outlined in Section VI.

II. ATTRIBUTE EXTRACTION

We describe an approach to extract attributes that indicate the occurrence of disturbances in voltage data. Fundamentally, an HP filter and the DFT are applied to raw voltage data within a time window. A more discriminative set of attributes facilitates model interpretation, reduces data overfitting, and may produce better results due to the elimination of attributes and noise that may mislead online systems [13], [16].

A. Hodrick-Prescott Filter

An HP filter decomposes a signal, y_t , into its trend, τ_t , and cyclical and random components, C_t , such that $y_t = \tau_t + C_t$.

It is equivalent to a cubic spline smoother, with the smoothed portion in τ_t [18]. In essence, low-frequency fluctuations are separated from the original data. The separation hypothesis is that the low-frequency variability represents the long-term trend, whereas the high-frequency variability means random phenomena. The HP filter is for the first time used for power system disturbance classification in this paper.

The HP filter extracts the trend, which is stochastic, but with smooth variations over time that are uncorrelated with other variations. The idea is to minimize the functional

$$J = \sum_{t=1}^T C_t^2 + \lambda \sum_{t=2}^{T-1} (\Delta^2 \tau_t)^2, \quad (1)$$

with respect to τ_t , in which $C_t := y_t - \tau_t$, and y_t , $t = 1, 2, \dots, T$, denotes the underlying signal; $\Delta^2 := (1 - L)^2$; L is the lag operator, e.g., $Lx_t = x_{t-1}$; T is the amount of data samples; and λ penalizes the variability of the trend component. Parameter λ is the smoothing parameter; it controls the variation of the growth rate of the trend. The first term of (1) is the sum of deviations of the signal concerning the square trend, a measure of the degree of fit. The second term is the sum of squares of the second differences of the trend component, a measure of the degree of smoothness. See [18] for details. The fourth section of this paper provides practical examples of the HP decomposition specifically for disturbances detection.

B. Discrete Fourier Transform

The Fourier Transform is one of the most used frequency-domain signal processing tools. The idea is that any periodic signal can be described by a sum of sines and cosines [19]. When measured data are available, a frequency spectrum is generated by discrete Fourier transformation from

$$DFT(f_n) = \frac{1}{N} \sum_{k=0}^{N-1} x_k e^{-j2\pi f_n k \Delta t}, \quad (2)$$

in which x_k is a discrete signal; Δt corresponds to time intervals; and $N := T/\Delta t$ is the number of samples – being T the total time interval. Additionally, $f_n = n/T$, $n = 0, 1, \dots, N - 1$, are the frequency components.

The discrete Fourier transform is an invertible linear transformation. The inverse is given by

$$x_k = \frac{1}{\Delta t} \sum_{f_n=0}^{(1-N)/T} DFT(f_n) e^{j2\pi f_n k \Delta t}. \quad (3)$$

See [19] for details on discrete Fourier transforms.

C. Root Mean Square Voltage

The effective (RMS) value is a measure of the magnitude of a variable quantity. RMS values can be calculated for a sequence of discrete values. The effective voltage of an alternating current circuit may provide evidence of some types of disturbances, e.g., sag, swell, and interruptions.

The RMS value of a sinusoidal waveform \mathbf{x} for a set of N samples, $\mathbf{x} = [x_1, x_2, \dots, x_N]$, is

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{k=1}^N x_k^2}. \quad (4)$$

This is an equivalent direct value able to produce the same power as that of the original waveform.

III. EVOLVING GAUSSIAN FUZZY CLASSIFIER FROM NEVER-ENDING SEMI-SUPERVISED DATA STREAMS

A. Preliminaries

We present EGFC, a semi-supervised evolving classifier derived from the online granular-computing framework of Leite et al. [13], [17]. EGFC employs Gaussian membership functions to cover the data space with fuzzy granules (local models), and associate new numerical data to class labels. Granules are scattered in the data space wherever needed to represent local information. EGFC overall response comes from the fuzzy aggregation of local models. A recursive algorithm constructs its rule base, and updates local models to deal with novelties. EGFC addresses issues such as unlimited amounts of data and scalability [13].

Local EGFC models are created if the new data are sufficiently different from the current knowledge. The learning algorithm can expand, retract, delete, and merge granules on occasion. Rules are reviewed according to inter-granular relations. EGFC provides nonlinear, nonstationary, and smooth boundaries among classes. This paper particularly addresses a 5-class disturbance classification problem.

Formally, let an input-output pair (\mathbf{x}, y) be related through $y = f(\mathbf{x})$. We seek an approximation to f to estimate the value of y given \mathbf{x} . In classification, y is a class label, a value in a set $\{C_1, \dots, C_m\} \in \mathbb{N}^m$, and f specifies class boundaries. In the more general, semi-supervised case, C_k may or may not be known when \mathbf{x} arrives. Classification of never-ending data streams involves pairs $(\mathbf{x}, C)^{[h]}$ of time-sequenced data, indexed by h . Nonstationarity requires evolving classifiers to identify time-varying relations $f^{[h]}$.

B. Gaussian Functions and Rule Structure

Learning in EGFC does not require initial rules. Rules are created and dynamically updated depending on the behavior of a system over time. When a data sample is available, a decision procedure may add a rule to the model structure or update the parameters of a chosen rule.

In EGFC models, a rule R^i is

IF $(x_1 \text{ is } A_1^i)$ AND ... AND $(x_n \text{ is } A_n^i)$
THEN $(y \text{ is } C^i)$

in which x_j , $j = 1, \dots, n$, are input attributes, and y is the output (a class). The data stream is denoted $(\mathbf{x}, y)^{[h]}$, $h = 1, \dots$. Moreover, A_j^i , $j = 1, \dots, n$; $i = 1, \dots, c$, are Gaussian membership functions built from the available data; and C^i is the class label of the i -th rule. Rules R^i , $i = 1, \dots, c$, form the rule base. The number of rules, c , is variable, which is a notable characteristic of the approach since guesses on how many data partitions exist are needless [4], [13].

A normal Gaussian function, $A_j^i = G(\mu_j^i, \sigma_j^i)$, has height 1 [16]. It is characterized by the modal value μ_j^i and dispersion σ_j^i . Characteristics that make Gaussians appropriate include: (i) easiness of learning and changing, i.e., modal values and dispersions are updated straightforwardly from a data stream; (ii) infinite support, i.e., since the data are priorly unknown, the support of Gaussians extends to the whole domain; and (iii) smooth surface of fuzzy granules, $\gamma^i = A_1^i \times \dots \times A_j^i \times \dots \times A_n^i$, in the n -dimensional Cartesian space – obtained by the cylindrical extension of one-dimensional Gaussians, and the use of the minimum T-norm aggregation [16].

C. Adding Rules to the Evolving Fuzzy Classifier

Rules may not exist *a priori*. They are created and evolved as data are available. A new granule γ^{c+1} and the rule R^{c+1} that governs the granule are created if none of the existing rules $\{R^1, \dots, R^c\}$ are sufficiently activated by $\mathbf{x}^{[h]}$; i.e., $\mathbf{x}^{[h]}$ brings new information. Let $\rho^{[h]} \in [0, 1]$ be an adaptive threshold that determines if a new rule is needed. If

$$T\left(A_1^i(x_1^{[h]}), \dots, A_n^i(x_n^{[h]})\right) \leq \rho^{[h]}, \quad \forall i, \quad i = 1, \dots, c, \quad (5)$$

in which T is any triangular norm, then the EGFC structure is expanded. The minimum (Gödel) T-norm is used in this paper. If $\rho^{[h]}$ is equal to 0, then the model is structurally stable, and unable to follow concept shifts. In contrast, if $\rho^{[h]}$ is equal to 1, EGFC creates a rule for each new sample, which is not practical. Structural and parametric adaptability are balanced for intermediate values (stability-plasticity tradeoff) [21].

The value of $\rho^{[h]}$ is crucial to regulate how large granules can be. Different choices impact the accuracy and compactness of a model, resulting in different granular perspectives of the same problem. Section III-E gives a Gaussian-dispersion-based procedure to update $\rho^{[h]}$.

A new granule γ^{c+1} is initially represented by membership functions, A_j^{c+1} , $j = 1, \dots, n$, with

$$\mu_j^{c+1} = x_j^{[h]}, \quad (6)$$

and

$$\sigma_j^{c+1} = 1/2\pi. \quad (7)$$

We call (7) the Stigler approach to standard Gaussian functions, or *maximum approach* [22], [23]. The intuition is to start big, and let the dispersions gradually shrink when new samples activate the same granule. This strategy is appealing for a compact model structure.

In general, the class C^{c+1} of the rule R^{c+1} is initially undefined, i.e., the $(c+1)$ -th rule remains unlabeled until a label is provided. If the corresponding output, $y^{[h]}$, associated to $\mathbf{x}^{[h]}$, becomes available, then

$$C^{c+1} = y^{[h]}. \quad (8)$$

Otherwise, the first labeled sample that arrives after the h -th time step, and activates the rule R^{c+1} according to (5), is used to define its class, C^{c+1} .

In case a labeled sample activates a rule that is already labeled, but their labels are different, then a new (partially overlapped) granule and a rule are created to represent new information. Partially overlapped Gaussian granules tagged with different labels tend to have their dispersions reduced over time by the parameter adaptation procedure (Sec. III-D). The modal values of the Gaussian granules may also drift, if convenient for a more suitable decision boundary.

With this initial parameterization, preference is given to granules balanced along their dimensions, rather than granules with unbalanced geometry. EGFC realizes the principle of the balanced granularity [24], but allows the Gaussians to find more appropriate places and dispersions.

D. Incremental Parameter Updating

Updating the EGFC model consists in: (i) reducing or expanding Gaussians $A_j^{i^*}$, $j = 1, \dots, n$, of the most active granule, γ^{i^*} , considering labeled and unlabeled samples; (ii) moving granules toward regions of relatively dense population; and (iii) tagging rules when labeled data are available. Adaptation aims to develop more specific local models [25], and provide pavement (covering) to new data.

A rule R^i is candidate to be updated if it is sufficiently activated by an unlabeled sample, $\mathbf{x}^{[h]}$, according to

$$\min \left(A_1^i(x_1^{[h]}), \dots, A_n^i(x_n^{[h]}) \right) > \rho^{[h]}. \quad (9)$$

Geometrically, $\mathbf{x}^{[h]}$ belongs to a region highly influenced by the granule γ^i . Only the most active rule, R^{i^*} , is chosen for adaptation in case two or more rules reach the $\rho^{[h]}$ level for the unlabeled $\mathbf{x}^{[h]}$. For a labeled sample, i.e., for pairs $(\mathbf{x}, y)^{[h]}$, the class of the most active rule R^{i^*} , if defined, must match $y^{[h]}$. Otherwise, the second most active rule among those that reached the $\rho^{[h]}$ level is chosen for adaptation, and so on. If none of the rules are apt, a new one is created (Sec. III-C).

To include $\mathbf{x}^{[h]}$ in R^{i^*} , EGFC's learning algorithm updates the modal values and dispersions of the corresponding membership functions $A_j^{i^*}$, $j = 1, \dots, n$, from

$$\mu_j^{i^*}(\text{new}) = \frac{(\varpi^{i^*} - 1)\mu_j^{i^*}(\text{old}) + x_j^{[h]}}{\varpi^{i^*}}, \quad (10)$$

and

$$\sigma_j^{i^*}(\text{new}) = \left(\frac{(\varpi^{i^*} - 1)}{\varpi^{i^*}} \left(\sigma_j^{i^*}(\text{old}) \right)^2 + \frac{1}{\varpi^{i^*}} \left(x_j^{[h]} - \mu_j^{i^*}(\text{old}) \right)^2 \right)^{1/2}, \quad (11)$$

in which ϖ^{i^*} is the number of times the i^* -th rule was chosen to be updated. Notice that (10)-(11) are recursive and, therefore, do not require data storage. As σ^{i^*} defines a convex region of influence around μ^{i^*} , very large and very small values may induce, respectively, a unique or too many granules per class. An approach is to keep $\sigma_j^{i^*}$ between a lower, $1/4\pi$, and the Stigler, $1/2\pi$, limits.

E. Dispersion-Based Time-Varying ρ -Level

Let the activation threshold, $\rho^{[h]} \in [0, 1]$, be time-varying. The threshold assumes values in the unit interval according to the overall average dispersion

$$\sigma_{avg}^{[h]} = \frac{1}{cn} \sum_{i=1}^c \sum_{j=1}^n \sigma_j^{i[h]}; \quad (12)$$

c and n are the number of rules and attributes, so that

$$\rho(\text{new}) = \frac{\sigma_{avg}^{[h]}}{\sigma_{avg}^{[h-1]}} \rho(\text{old}). \quad (13)$$

As mentioned, rules' activation levels for an input $\mathbf{x}^{[h]}$ are compared to $\rho^{[h]}$ to decide between parametric or structural changes of an EGFC model. In general, EGFC starts learning from an empty rule base, and without knowledge about the properties of the data. Practice suggests $\rho^{[0]} = 0.1$ as starting value. The threshold tends to converge to a proper value if the classifier structure achieves a level of maturity and stability. Nonstationarities and new classes guide $\rho^{[h]}$ to values that better reflect the needs of the current environment.

F. Merging Similar Granules

Similarity between two granules with the same class label may be high enough to form a unique granule that inherits the essence of both. Analysis of inter-granular relations requires a distance measure between Gaussians. Let

$$d(\gamma^{i_1}, \gamma^{i_2}) = \frac{1}{n} \left(\sum_{j=1}^n |\mu_j^{i_1} - \mu_j^{i_2}| + \sigma_j^{i_1} + \sigma_j^{i_2} - 2\sqrt{\sigma_j^{i_1} \sigma_j^{i_2}} \right) \quad (14)$$

be the distance between γ^{i_1} and γ^{i_2} . This measure considers the information specificity, that is, in turn, inversely related to the Gaussians' dispersion [23]. For example, if the dispersions $\sigma_j^{i_1}$ and $\sigma_j^{i_2}$ differ one from another, rather than being equal, the distance between the underlying Gaussians is larger.

EGFC may merge the pair of granules that presents the smallest value of $d(\cdot)$ for all pairs of granules. Both granules must be either unlabeled or tagged with the same class label. The merging decision is based on a threshold value, Δ , or expert judgment regarding the suitability of combining such granules to have a more compact model. For data within the unit hypercube, we suggest $\Delta = 0.1$ as default, which means that the candidate granules should be quite similar.

A new granule, say γ^i , which results from γ^{i_1} and γ^{i_2} , is built by Gaussians with modal values

$$\mu_j^i = \frac{\frac{\sigma_j^{i_1}}{\sigma_j^{i_2}} \mu_j^{i_1} + \frac{\sigma_j^{i_2}}{\sigma_j^{i_1}} \mu_j^{i_2}}{\frac{\sigma_j^{i_1}}{\sigma_j^{i_2}} + \frac{\sigma_j^{i_2}}{\sigma_j^{i_1}}}, \quad j = 1, \dots, n, \quad (15)$$

and dispersion

$$\sigma_j^i = \sigma_j^{i_1} + \sigma_j^{i_2}, \quad j = 1, \dots, n. \quad (16)$$

These relations take into account the granular uncertainty to find an appropriate location and size to the resulting granule. Merging minimizes redundancy [4], [13].

G. Deleting Rules

A rule is removed from the EGFC model if it is inconsistent with the current environment. In other words, if a rule is not activated for a number of iterations, say h_r , then it is deleted from the rule base. However, if a class is rare, e.g., a type of power quality disturbance is unusual, then it may be the case to set h_r to infinity and keep the inactive rules. Removing rules periodically helps to keep the model updated.

H. Semi-Supervised Learning from Data Streams

The semi-supervised learning procedure to construct and update EGFC models along their lifespan is given below.

EGFC: Online Semi-Supervised Learning

```

1: Initial number of rules,  $c = 0$ ;
2: Initial meta-parameters,  $\rho^{[0]} = \Delta = 0.1$ ,  $h_r = 200$ ;
3: Read input data sample  $\mathbf{x}^{[h]}$ ,  $h = 1$ ;
4: Create granule  $\gamma^{c+1}$  (Eqs. (6)-(7)), unknown class  $C^{c+1}$ ;
5: FOR  $h = 2, \dots$  DO
6:   Read  $\mathbf{x}^{[h]}$ , calculate rules' activation degree (Eq. (5));
7:   Determine the most active rule  $R^{i^*}$ ;
8:   Provide estimated class  $C^{i^*}$ ;
9:   // Model adaptation
10:  IF  $T(A_1^i(x_1^{[h]}), \dots, A_n^i(x_n^{[h]})) \leq \rho^{[h]} \forall i, i = 1, \dots, c$ 
11:    IF actual label  $y^{[h]}$  is available
12:      Create labeled granule  $\gamma^{c+1}$  (Eqs. (6)-(8));
13:    ELSE
14:      Create unlabeled granule  $\gamma^{c+1}$  (Eqs. (6)-(7));
15:    END
16:  ELSE
17:    IF actual label  $y^{[h]}$  is available
18:      Update the most active granule  $\gamma^{i^*}$  whose class  $C^{i^*}$  is equal to  $y^{[h]}$  (Eqs. (10)-(11));
19:      Tag unlabeled active granules;
20:    ELSE
21:      Update the most active  $\gamma^{i^*}$  (Eqs. (10)-(11));
22:    END
23:  END
24:  Update the  $\rho$ -level (Eqs. (12)-(13));
25:  Delete inactive rules based on  $h_r$ ;
26:  Merge granules based on  $\Delta$  (Eqs. (14)-(16));
27: END

```

IV. METHODOLOGY

We describe the methodology to generate power system disturbances. We give examples of disturbances, and a flowchart that connects the DFT-HP attribute extraction and EGFC.

A. Online Monitoring System

Voltage data from a 13.8kV grid are produced according to the IEEE standard [26]. The fundamental and sampling frequencies are 60Hz and 15,360Hz. Thus, 256 samples per

cycle are given. This sampling rate is sufficient to characterize most of the disturbances in power systems, including spikes, notching, harmonics, and oscillatory transient. Gaussian white noise is added to give different signal-to-noise (SNR) ratio,

$$SNR = 20 \log \frac{\beta}{\sqrt{2}\sigma} \text{ [dB]}, \quad (17)$$

in which β is the amplitude of the original voltage signal; σ is the standard deviation of the Gaussian noise; and dB means deciBel – one tenth of Bel.

Voltage data from power systems usually have an SNR from 40 to 70dB. We evaluate the evolving classifier subject to 20, 40, and 60dB SNR – being 20dB the harshest stochastic scenario. Ten thousand voltage waveforms are generated:

- Class 1: 2,000 waveforms without disturbances;
- Class 2: 2,000 waveforms with spikes;
- Class 3: 2,000 waveforms containing notching;
- Class 4: 2,000 waveforms with harmonics;
- Class 5: 2,000 waveforms with oscillatory transient.

Time windows based on constant time intervals between landmarks are considered. Windows of different lengths (1, 4 and 10 cycles of the fundamental) are assessed in different experiments. The greatest peaks and valleys of the fundamental voltage in a window are rescaled in the range $[0, 1]$. A phase angle within $[-\pi, \pi]$ is randomly assigned to the starting point of a waveform. Waveforms are subject to noise (17).

Voltage data within a window are fed to HP-DFT attribute extraction. Then, a vector of input data is formed and provided to the EGFC model. EGFC estimates a class, and then uses the input vector – accompanied or not by a label – to update its parameters and structure. For each window, this procedure is repeated. A general flowchart of the power quality monitoring and classification system is shown in Fig. 1.

Four disturbance indicators compose an input vector $\mathbf{x}^{[h]}$ of the EGFC model. They are

- x_1 : Amplitude of the fundamental (60Hz) voltage component, obtained by DFT over the data in a time window;
- x_2 : Minimum value of the voltage cyclical component after HP decomposition over the data in a time window;
- x_3 : Maximum value of the voltage cyclical component after HP decomposition over the data in a time window;
- x_4 : Effective value of the voltage cyclical component after HP decomposition over the data in a time window.

The EGFC estimated output, \hat{y} , is a class $\hat{C} = \{1, 2, 3, 4, 5\}$. Constructing and updating an EGFC model is a fully online process based on partially-labeled data.

B. Generating Disturbances

Disturbances, viz., spike, notching, harmonics, and oscillatory transient (common power-system phenomena), are added to the fundamental voltage subject to an SNR .

Spikes or surges are fast, short-duration voltage transients caused by lightning strikes, power outages, short circuits, power transitions in large equipment, to mention some [27]. A spike usually lasts from 1 to $30\mu s$, and may reach over 1,000V.

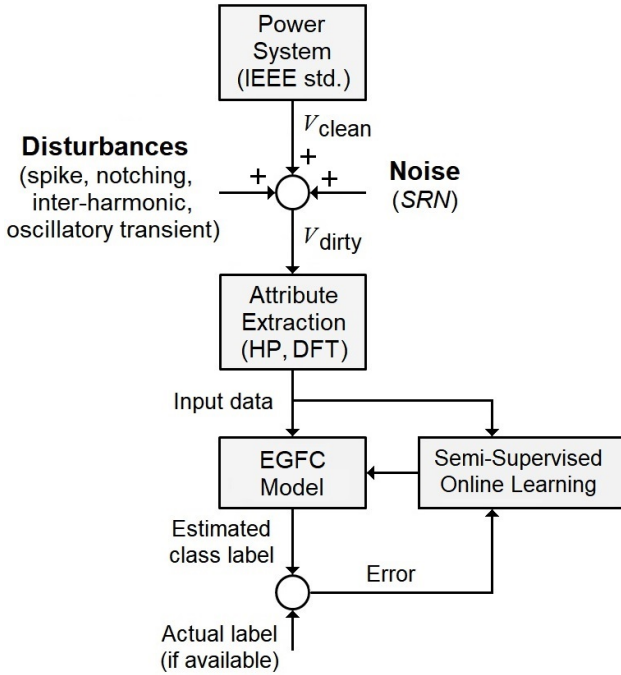


Fig. 1. Evolving Disturbance Detection and Classification System

For example, a motor when switched off can generate a spike of 1,000V. Spikes can degrade wiring insulation and destroy electronic devices. Some common-mode voltage spikes may not be detected by surge protection equipment.

To generate spike we choose a random starting point during the first voltage cycle. The spike peaks after 10 samples, and extinguishes after 20 samples. Its maximum amplitude is a random number in $[1, 1.5]$ pu or $[-1.5, -1]$ pu. The spike repeats in subsequent cycles for window lengths larger than one cycle, see example in Fig. 2.

Notching is a periodic disturbance, a switching lasting less than 0.5 cycles. It is caused by the normal operation of electronic devices and three-phase converters. Notches occur when the current commutates from one phase to another. The severity of a notch is given by the source and isolating inductances of a converter, the magnitude of the current, and the point being monitored. The frequency components associated with notching can be quite high [28]. To generate notching we choose a random starting sample in $[10, 40]$. The disturbance extinguishes after 9 samples; it repeats 8 times per cycle, 23 samples after the previous occurrence. The maximum amplitude is $[-0.5, -0.05]$ pu or $[0.05, 0.5]$ pu, see Fig. 2.

Harmonics are sinusoidal voltages with frequencies that are integer multiples of the fundamental. Distortion arises as current sources that inject harmonic currents into the power system cause nonlinear voltage drops across the system impedance. Harmonic currents result from the normal operation of nonlinear electronic devices and loads on the system. Harmonic distortion is a growing concern for customers and for the overall power system due to an increasing number of

power electronics equipment [26]. To generate harmonics, random values in $[0.008, 0.016]$ pu, $[0.02, 0.04]$ pu, $[0.005, 0.01]$ pu, $[0.023, 0.046]$ pu, $[0.003, 0.006]$ pu, and $[0.02, 0.04]$ pu are chosen, respectively, for the second to the seventh harmonic. The start point of each harmonic is independent one another and may have any phase angle in $[-\pi, \pi]$, see Fig. 2.

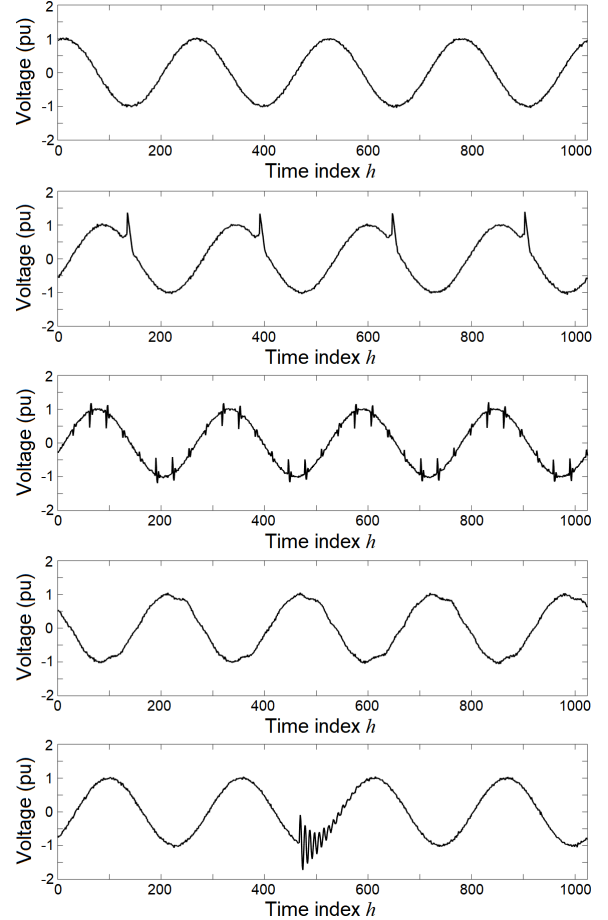


Fig. 2. Examples of 4-cycle voltage waveforms of each class, with $SNR = 30$. From top to bottom: no disturbance (class 1); spikes (class 2); notching (class 3); harmonics (class 4); oscillatory transient (class 5)

Oscillatory transient is a sudden change of the voltage steady-state condition that includes rapid changes of positive and negative polarity values. Transients are almost always due to some type of switching event. Power electronic devices can produce oscillatory transients as a result of commutation and RLC snubber circuits [26]. We choose a random sample in a time window as start point of a transient. The transient is an exponentially damped sinusoid whose start amplitude is in $[0.45, 1]$ pu in relation to the fundamental component. Its frequency is a random value in $[1000, 2500]$ Hz. The damping coefficient is a random value in $[400, 1000]$, see Fig. 2.

C. Classification Accuracy

Classification accuracy is computed recursively from

$$Acc(new) = \frac{h-1}{h} Acc(old) + \frac{1}{h} \tau, \quad (18)$$

in which $Acc \in [0, 1]$; $\tau = 1$ if $\hat{C}^{[h]} = y^{[h]}$ (right estimate). Otherwise, $\tau = 0$ (wrong class estimate).

The average number of granules or rules over time, c_{avg} , is a measure of model concision. Recursively,

$$c_{avg}(new) = \frac{h-1}{h} c_{avg}(old) + \frac{1}{h} c^{[h]}. \quad (19)$$

V. RESULTS AND DISCUSSIONS

We evaluate the EGFC approach. No prior knowledge about the data and power system is assumed. Classification models are developed from scratch, based on data streams.

A. Preliminary Results on Feature Extraction

We use DFT and HP filtering to extract four disturbance indicators from the voltage waveform. The DFT provides x_1 , the amplitude of the fundamental component. The HP filter decomposes the original waveform into trend and cyclical components. We focus on the cyclical component to obtain x_2 , x_3 and x_4 , i.e., the minimum, maximum and effective values of the decomposed waveform. We achieved encouraging results using an HP smoothing coefficient of 256×10^3 as a great portion of noise is isolated from the fundamental. Figure 3 shows typical values of attributes, $\mathbf{x} = [x_1 \ x_2 \ x_3 \ x_4]$, extracted from waveforms of each disturbance class. The examples consider 4-cycle time windows and an SNR of 30.

Notice in Fig. 3 that the value of x_1 changes slightly, up and down, respectively, in the spike and notching scenarios, which may help the evolving classifier to distinguish these classes. The HP cyclical component for the case without disturbance shows an initial transient such that the absolute values of x_2 and x_3 are significantly different one another, which helps the classifier to recognize this class. This phenomenon also happens for the spike scenario, with greater unbalance between the values of x_2 and x_3 . Attribute x_4 tends to zero rapidly for the notching case. The absolute values of x_2 and x_3 in the harmonic scenario are similar. The same happens in the oscillatory transient case, but with higher individual amplitudes. Therefore, \mathbf{x} carries important subtleties.

B. EGFC Results for Labeled Data Streams

We look for an evolving classifier based on a data stream. The default meta-parameters are used (Sec. III-H). Table I shows the results averaged over 5 runs for 9 datasets extracted from voltage waveforms based on window lengths of 1, 4, and 10 cycles; and an SNR of 20, 40, and 60dB. Each dataset consists of 10,000 4-attribute samples related to a target class $C^{[h]} \in \{1, 2, 3, 4, 5\}$. The classes mean ‘no disturbance’, ‘spikes’, ‘notching’, ‘harmonics’, and ‘oscillatory transient’.

Table I shows that the SNR is irrelevant to the classifier performance using the set of attributes \mathbf{x} chosen. The accuracy can be relatively higher in noisier conditions, e.g. 20dB. This is an interesting feature of the proposed monitoring system. In contrast, a very small window length can degrade system performance significantly. The 4-cycle scenario seems more attractive than the 10-cycle one as the system is able to analyze a higher amount of windows at the price of a small

reduction of the classification accuracy. As at least two 4-cycle windows are processed during a 10-cycle period, if the system provides wrong classification for the data extracted from the first window, it can still detect the disturbance class from the data of the other window. Therefore, in practice, the 92.79%-accuracy 4-cycle-based EGFC system can be more efficient than the 94.24%-accuracy 10-cycle-based one. The number of rules in the model structure over the learning steps, and the CPU time in a quad-core i7-8550U with 1.80GHz and 8GB of RAM are similar in all scenarios.

TABLE I
EGFC PERFORMANCE IN MULTICLASS CLASSIFICATION OF POWER SYSTEM DISTURBANCES (99% CONFIDENCE)

SNR	Cycles	$Acc(\%)$	# Rules	Time (s)
60dB	10	93.17 ± 0.72	9.15 ± 1.73	1.80 ± 0.25
	4	87.47 ± 0.60	9.52 ± 1.73	1.72 ± 0.21
	1	64.41 ± 0.27	10.77 ± 1.50	1.97 ± 0.23
40dB	10	92.98 ± 1.30	8.98 ± 0.99	1.69 ± 0.10
	4	88.33 ± 0.74	9.01 ± 1.46	1.67 ± 0.12
	1	63.90 ± 0.39	10.58 ± 0.60	1.96 ± 0.16
20dB	10	94.24 ± 0.21	8.38 ± 0.54	1.59 ± 0.08
	4	92.79 ± 1.07	8.70 ± 0.89	1.65 ± 0.12
	1	67.10 ± 0.64	9.69 ± 1.36	1.87 ± 0.16

Figure 4 shows a typical example of evolution of the ρ -level, accuracy, and number of EGFC rules. The final granules, at $h = 10,000$, are also illustrated. Class-2 data (spike disturbance) spread over a larger area, and require four granules to be represented, whereas the remaining classes require a single granule. Figure 5 shows the confusion matrix obtained. Class 1 (no disturbance) and Class 4 (harmonics), followed by Class 4 and Class 5 (transient), are those responsible for the 7.7% overall error. Additional attributes should be considered in the future to address these particular hesitations.

C. EGFC Result in Semi-Supervised Online Scenario

We changed the proportion of unlabeled data from 0% to 100% considering the harshest 20dB problem. Figure 6 shows average EGFC results for 5 runs for each case. EGFC benefits of all information of the data stream, including that from unlabeled samples. Conventional and evolving classifiers that operate on a supervised basis by simply discarding unlabeled data cannot deal with small fractions of labeled data with reasonable accuracy (as shown by the right-side points of the graph). The left and right extremes of the plot indicate full supervision and non-supervision. In all cases the final result is a partition of data into classes. EGFC is not significantly affected by fractions of unlabeled data. For data extracted from 4-cycle windows, the performance of pure classification and clustering were 92.79% and 86.12%, respectively.

VI. CONCLUSION

We propose a hybrid attribute-extraction method combined with an evolving Gaussian fuzzy model for power quality disturbance detection and classification. A Hodrick-Prescott filter and the Discrete Fourier Transform applied over window of voltage data have provided attractive attributes for disturbance

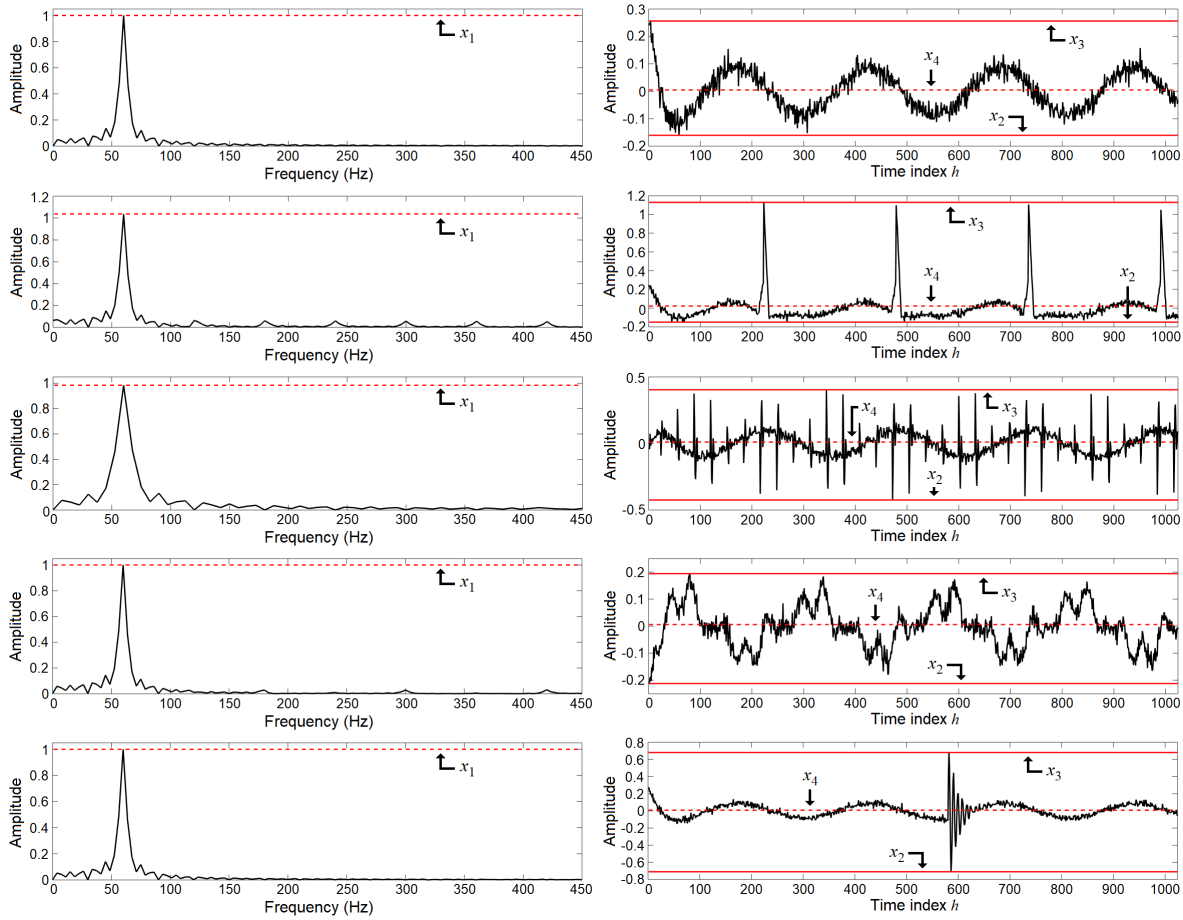


Fig. 3. Typical examples of attributes $\mathbf{x} = [x_1 \ x_2 \ x_3 \ x_4]$ extracted from 5 different voltage waveforms using DFT (left column) and HP filter (right column) considering 4-cycle time windows and an SNR of 30. From top to bottom: row 1 - no disturbance (class 1); row 2 - spikes (class 2); row 3 - notching (class 3); row 4 - harmonics (class 4); row 5 - oscillatory transient (class 5)

discrimination. Common types of disturbances, namely, spikes, notching, harmonics, and oscillatory transient were analyzed. Data generation agrees with the IEEE standard for power quality disturbances. A landmark window containing one, four, and ten voltage cycles as well as signal-to-noise ratio ranging from 20 to 60dB were evaluated. The evolving modeling approach, EGFC, has shown to be efficient for multi-class online classification. Its fuzzy rule-based structure, Gaussian membership functions, and granularity are updated over time driven by the data stream. Online model adaptation has shown to be essential to deal with time-varying systems, such as power systems subject to disturbance patterns.

Harmonics followed by oscillatory transient have shown to be more challenging to be distinguished compared to spikes and notching. The signal-to-noise ratio does not affect EGFC performance significantly. The use of 4 voltage cycles for attribute extraction provided an average EGFC accuracy of 92.8% in the harshest 20dB noise case, and therefore was considered ideal. Changing the proportion of unlabeled data from 0% to 100% made the EGFC performance reduce from 92.8% to 86.1%. Therefore, EGFC is applicable to and robust

to clustering and classification. New types of disturbances can be studied in the future. The EGFC semi-supervised learning framework shall be analyzed considering synthetic examples of concept change and anomaly detection problems.

REFERENCES

- [1] P. D. Achlerkar, S. R. Samantaray, M. S. Manikandan. "Variational Mode Decomposition and Decision Tree Based Detection and Classification of Power Quality Disturbances in Grid-Connected Distributed Generation System." *IEEE T Smart Grid*, 9(4), p. 3122-3132, 2018.
- [2] J. Li, Z. Teng, Q. Tang, J. Song. "Detection and Classification of Power Quality Disturbances Using Double Resolution S-Transform and DAG-SVMs." *IEEE T Instrum Meas*, 65(10), p. 2302-2312, 2016.
- [3] A. S. Sobrinho, R. A. Flauzino, L. H. Liboni, E. C. Costa. "Proposal of a Fuzzy-based PMU for Detection and Classification of Disturbances in Power Distribution Networks." *Int J Elec Power*, 94, p. 27-40, 2018.
- [4] I. Skrjanc, J. A. Iglesias, A. Sanchis, D. Leite, E. Lughofer, F. Gomide. "Evolving Fuzzy and Neuro-Fuzzy Approaches in Clustering, Regression, Identification, and Classification: A Survey." *Information Sciences*, 490, p. 344-368, 2019.
- [5] D. Leite. "Comparison of Genetic and Incremental Learning Methods for Neural Network-Based Electrical Machine Fault Detection." In: Lughofer E., Sayed-Mouchaweh M. (Eds), *Predictive Maintenance in Dynamic Systems*, p. 231-268, Springer: Cham, 2019.
- [6] G. Andonovski, S. Blazic, I. Skrjanc "Evolving Fuzzy Model for Fault Detection and Fault Identification of Dynamic Processes." In: *Predictive Maintenance in Dynamic Systems*, p. 269-285, Springer: Cham, 2019.

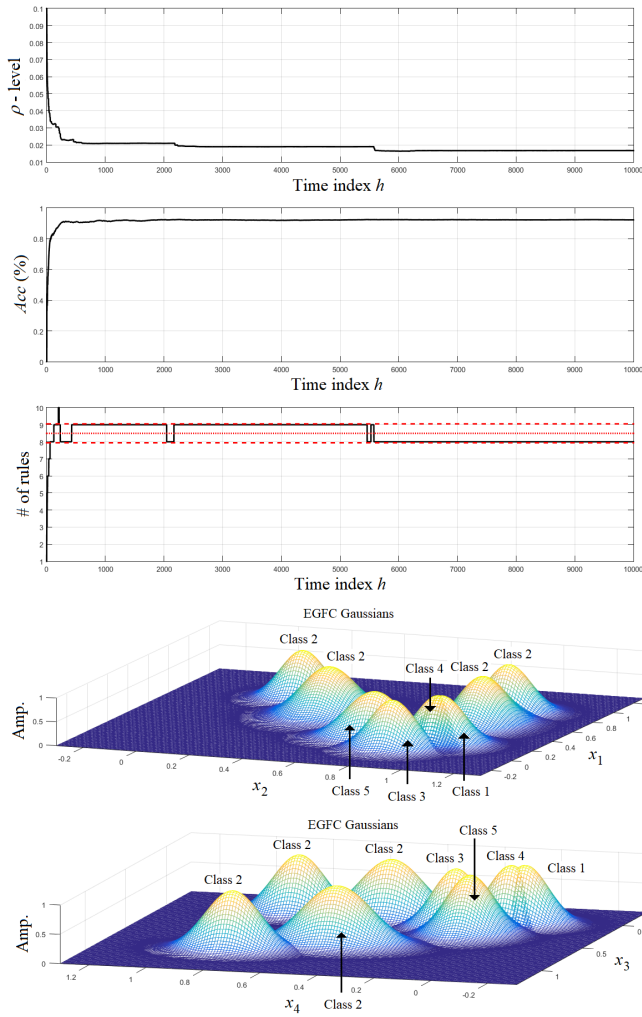


Fig. 4. Evolution of the ρ level, EGFC accuracy, and number of rules. The bottom plots show the final shape of 4-dimensional Gaussian granules

[7] Z. Hu, Y. Bodyanskiy, O. Tyshchenko, O. Boiko. "A Neuro-Fuzzy Kohonen Network for Data Stream Possibilistic Clustering and its Online Self-Learning Procedure." *Appl Soft Comput*, 68, p. 710-718, 2018.

[8] S. Silva, P. Costa, M. Santana, D. Leite. "Evolving Neuro-Fuzzy Network for Real-Time High Impedance Fault Detection and Classification." *Neural Comput & Applic*, 14p. DOI: 10.1007/s00521-018-3789-2, 2018.

[9] P. Souza, et al. "Evolving Fuzzy Neural Networks to Aid in the Construction of Systems Specialists in Cyber Attacks." *J Intell Fuzzy Syst*, 36(6), p. 6743-6763, 2019.

[10] J. Kirkpatrick, et al. "Overcoming Catastrophic Forgetting in Neural Networks." arXiv:1612.00796 [cs.LG], 13p. 2017.

[11] P. Angelov, D. Filev, N. Kasabov. *Evolving Intelligent Systems: Methodology and Applications*. Wiley: New York, 2010.

[12] E. Lughofer, M. Sayed-Mouchaweh (Eds). *Learning in Nonstationary Environments: Methods and Applications*. Springer: New York, 2012.

[13] D. Leite. *Evolving Granular Systems*. PhD Thesis, State University of Campinas (UNICAMP), 2012.

[14] L. Cordovil, P. Coutinho, I. Bessa, M. D'Angelo, R. Palhares. "Uncertain Data Modeling Based on Evolving Ellipsoidal Fuzzy Information Granules." *IEEE T Fuzzy Syst*, DOI:10.1109/TFUZZ.2019.2937052, 2019.

[15] P. C. L. Silva, H. J. Sadaei, R. Ballini, F. G. Guimares. "Probabilistic Forecasting with Fuzzy Time Series." *IEEE Trans Fuzzy Syst*, 14p. DOI: 10.1109/TFUZZ.2019.2922152, 2019.

[16] W. Pedrycz, F. Gomide. *Fuzzy Systems Engineering: Toward Human-Centric Computing*. Wiley: Hoboken - New Jersey, 2007.

	1697 17.0%	0 0.0%	0 0.0%	231 2.3%	3 0.0%	87.9% 12.1%
1						
2	0 0.0%	1995 20.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	1 0.0%	1998 20.0%	1 0.0%	5 0.1%	99.7% 0.3%
4	303 3.0%	0 0.0%	1 0.0%	1768 17.7%	219 2.2%	77.2% 22.8%
5	0 0.0%	4 0.0%	1 0.0%	0 0.0%	1773 17.7%	99.7% 0.3%
	84.9% 15.1%	99.8% 0.2%	99.9% 0.1%	88.4% 11.6%	88.6% 11.4%	92.3% 7.7%
	1	2	3	4	5	
	Target Class					

Fig. 5. Typical EGFC confusion matrix based on a 4-cycle time window, an SNR of 20dB, and labeled data stream

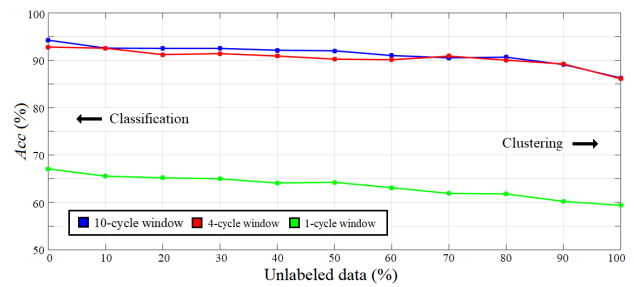


Fig. 6. EGFC performance using different proportions of unlabeled data

[17] D. Leite, R. Ballini, P. Costa, F. Gomide. "Evolving Fuzzy Granular Modeling from Nonstationary Fuzzy Data Streams." *Evolving Systems*, 3(2), p. 65-79, 2012.

[18] R. Hodrick, E. Prescott. "Postwar U.S. Business Cycles: An Empirical Investigation." *J of Money, Credit, and Banking*, 29(1), p. 1-16, 1997.

[19] B. P. Lathi. *Linear Systems and Signals*. Oxford U. Press, 2nd ed. 2004.

[20] M. V. Reddy, R. Sodhi. "A Rule-based S-Transform and AdaBoost based Approach for Power Quality Assessment." *Electr Pow Syst Res*, 134, p. 66-79, 2016.

[21] D. Leite, P. Costa, F. Gomide. "Evolving Granular Neural Networks from Fuzzy Data Streams." *Neural Netw*, 38, p. 1-16, 2013.

[22] S. M. Stigler. *A modest proposal: a new standard for the normal*. *The American Statistician*, Vol. 36-2, JSTOR, 1982.

[23] D. Leite, G. Andonovski, I. Skrjanc, F. Gomide. "Optimal Rule-based Granular Systems from Data Streams." *IEEE T Fuzzy Syst*, 28(3), p. 583-596, 2020.

[24] X. Wang, W. Pedrycz, A. Gacek, X. Liu. "From Numeric Data to Information Granules: A Design Through Clustering and the Principle of Justifiable Granularity." *Knowl-Based Syst*, 101, p. 100-113, 2016.

[25] R. Yager. "Measures of Specificity over Continuous Spaces under Similarity Relations." *Fuzzy Set Syst*, 159, p. 2193-2210, 2008.

[26] IEEE Recommended Practice for Monitoring Electric Power Quality. IEEE Power & Energy Society. IEEE Std 1159-2009, 2009.

[27] D. O. Johnson, K. A. Hassan. "Issues of Power Quality in Electrical Systems." *Int J Energy Power Eng*, 5(4), p. 148-154, 2016.

[28] IEEE Recommended Practice and Requirements for Harmonic Control in Electric Power Systems. IEEE TDC, IEEE Std 519-2014, 2014.