

A Fuzzy Curve-based Anomaly Detection and its application to Electromagnetic data

Vyron Christodoulou, Yaxin Bi, George Wilkie

Abstract—The problem of data analytics in real world electromagnetic (EM) applications poses a lot of algorithmic constraints. The process of big datasets, the requirement of prior knowledge, unknown location of anomalies and variable length patterns are all issues that need to be addressed. In this application we address those issues by proposing a Fuzzy Symbolic Representational method with anomaly detection (AD). This method is evaluated against twelve benchmark datasets of different kinds of anomalies and provides promising results based on the use of a new performance metric that takes into account the distance between predicted and actual anomalies. Real-world EM data from the Earth's magnetic field are provided by the SWARM satellite constellation using regions in China, Greece and Peru. The seismic events that occurred in those regions are compared against the SWARM data. Moreover, three other methods: GrammarViz, HOT-SAX and CUSUM-EWMA are also applied to further investigate the possible linkages of EM anomalies with seismic events. The findings further our understanding of real-world data analytics in EM data and seismicity. Some proposals regarding the limitations of available data for the real-world datasets are also presented.

Index Terms—Anomaly Detection, Symbolic Representation, Fuzzy Logic, Electromagnetic Data, SWARM Satellites



1 INTRODUCTION

In recent years the advent of data mining has seen an increase in the interest in discovering complex data structures in order to improve the performance of data analytics techniques. Improvements in time series classification, clustering, mining interesting patterns, data representations, AD and others have all been at the center of data research. Even though a lot of interesting results and improvements are evident in the field, a major issue in processing comes from the continuous growth of data. The fact that big time series data are essentially high-dimensional, causes a high computational cost in both time and storage.

An equally important problem in time series AD is that existing methods do not address the issue of identifying where the location of an anomaly is. Most methods require the user to provide as a minimum a window length parameter or to have prior knowledge of the data. In that sense, our aim is twofold: (a) Address the high dimensionality through a symbolic approximation and (b) the localization of AD. Moreover and equally importantly, we introduce a metric to measure the performance of such algorithms that combines the following: (a) the identified window length, (b) the predicted anomalous location and (c) the true anomalous location. Without such a metric any algorithmic comparison is difficult to be complete.

From our point of view, an AD method can benefit from the symbolic representation of a time series at the pre-processing stage. A symbolic representation with AD can

be used in different scenarios that require the detection of anomalous patterns of various forms. The main benefit is the time series reduction which affords a quick and efficient AD that can also be used to precisely identify the sequence's pattern and anomalous location. High precision combined with little or no prior knowledge is what is needed for in today's AD systems. What is more, there is a need for systems to perform robustly in an ever-changing data stream along with on-line AD capabilities in fields such as social media, networks, fraud detection etc. [10]. There exist many methods that require prior knowledge, however, the main focus in this work is to propose and compare similar methods with little to no parameter tuning for AD. As a consequence, there is an attempt to bridge this gap by bringing together these research areas.

Our aim is to utilize a fuzzy system for the symbolic representation of the signals and join it with an AD method in order to evaluate its capability as a model for the representation and AD in geophysical signals. These facts put several constraints on our problem. To address these limitations the algorithm has to fulfil several criteria:

- Handle Big Datasets with the symbolic representation and subsequent appropriate reduction of the time series to symbols.
- Detect anomalies of variable length and unknown patterns and be flexible in doing so.
- Be robust to noise.
- Anomaly localization.
- Perform all of the above with little prior knowledge.

Therefore, at its first stage, our method will be focused on the quality of the representation and its description of the time series. Specifically, it will be useful to understand how the time series reduction to symbols from the original form, compensates for the loss of information in terms

• Vyron Christodoulou is with the School of Computing and Mathematics Ulster University Jordanstown, Newtownabbey Co. Antrim BT37 0QB, U.K.

E-mail: christodoulou-v@email.ulster.ac.uk

• Yaxin Bi and George Wilkie are with the Ulster University, School of Computing and Mathematics, Jordanstown, Newtownabbey Co. Antrim BT37 0QB, U.K.

E-mail: y.bi@ulster.ac.uk, fg.wilkie@ulster.ac.uk

of AD. It will be shown that the different representation methods in the proposed algorithm, model the specific cases of anomalies differently after the conversion phase. They can better highlight and represent in more detail the nature of the anomalies. Each representation benefits from the fact that it gives weight to a different factor and can model unique kinds of anomalies within the same sequence. This gives the user a better understanding of what kinds of anomalies one has to deal with and from our point of view it will shed more light on the unknown nature of geophysical EM anomalies and their relation to seismic events.

The rest of the paper is organized as follows. In Section 2, the related work is discussed and how this work fits in this scope. Section 3, discusses the data gathering process and a brief description of the real world data is given. Section 4, lays the foundation of the background knowledge. Section 5, provides a step by step description of the process of the overall AD. Section 6, presents and discusses the results and finally Section 7 gives a summary and further possible directions regarding the results.

2 RELATED WORK

There is a large body of literature concerning AD algorithms [19]. Most AD algorithms are classified in five categories: Probabilistic, Reconstruction-based, Domain-based, Distance-based and Information Theoretic. As there is not a universally accepted definition of an anomaly, algorithms focus on different aspects ranging from point, context or pattern anomalies. In our case, the difficulty lies in the unknown nature of the anomaly. We do not know if it is a point anomaly or of a specific pattern. As a result, most AD algorithms are based on evidence and are data-based approaches.

Probabilistic algorithms for AD have been well researched. Their main benefit is that they require no parameter tuning and they are based on a sound theory. The CUSUM, detects variations in the mean of a process and assumes that the generative process is Gaussian. In [26] the focus was on the on-line update of CUSUM's K parameter with the implementation of a runs rules approach. A variety of CUSUM and EWMA combinations have been proposed in different forms in, [28], [27] and [29]. All the above methods improved the change detection in terms of sensitivity of the algorithm but it is known that it is difficult to detect frequency changes. A more complex probabilistic approach for AD was also proposed in [2] where the authors combine martingales with a Geometric Moving Average to detect anomalies. This method, although it does not make use of a user selected window length still needs parameter tuning.

In [18], a combination of a reconstruction and a statistical approach is applied. An Artificial Bee Colony Algorithm for training a predictor that estimates the future value of electron content is used and a statistics based threshold flags the anomalies with promising results.

Distance based approaches such as HOT-SAX [8] are appropriate to our case because they address both the high dimensionality and the anomaly localization problem. For SAX a symbolic representation is combined with an AD method. The authors propose a bruteforce discord discovery (BFDD) and a Heuristic Discord Discovery (HDD) for the AD. A Piecewise Aggregate approximation (PAA) is used to

discretize the original time series but it causes information loss and can miss anomalous patterns. Numerous different representation methods have been spawned following the success of SAX, in [20], [11] using similar principles.

A Piecewise Aggregate Pattern Representation (PARP) method was proposed in [12] in order to address the PAA-induced information loss. Instead of the euclidean distance a new similarity measure is defined to assess the similarity between subsequences. However, this method also involves parameter tuning.

Once it became evident that the parameter tuning is also an obstacle, a less time consuming method with minimal parameter tuning was featured in [24]. The authors proposed a grammar rule based system for AD. Their method, utilizes the SAX approach for symbolic representation and discretization and it is able to find subsequences of variable length. Similarly, in [25] the authors propose SLADE-TS. An algorithm that even though it requires some parameter tuning for the discretization step, it is parameter-less during the AD and can also perform on-line AD. The AD scheme in the previous works follows the same pattern as in our work: (a) Time Series Discretization, (b) Symbolic approximation, (c) Subsequence extraction and last but not least, the AD method.

The same challenge was again addressed in [17]. What is essentially at stake is the discovery of subsequences of varying lengths and patterns, which alludes to a parameter free method. To overcome this problem, a new location-based discord discovery method was envisioned. The authors propose a two pass adaptive window discord discovery (AWDD) approach. In contrast to BFFD, proposed in SAX, which is a one pass approach, this method demonstrates faster processing time but it does not use a symbolic representation step. The detection of patterns of varying length is also the focus of [30], where a dynamic time warping (DTW) approach is used to calculate similarity patterns of varying length. The patterns have been identified by another quadratic regressions algorithm. The authors present a promising performance but their method still needs parameter configuration in the first step and because the second step is highly dependable different parameters lead to very time consuming demands.

In terms of fuzzy logic, the Fuzzy symbolic representation has not been very well investigated. Recently, some work has been carried out in [14]. The authors proposed a fuzzy representation of linear piecewise segments in time series using linear regression in combination with a growing window. Another fuzzy representation concept, this time characterizing temperature records is used in [15]. In this case the Fuzzy System is both a fuzzy descriptor and an anomaly detector. This is achieved by assigning different confidence levels to the output membership function for the AD. In our case such an approach is impossible since there is no prior knowledge of the nature of the anomaly and specifically whether it is of high or low amplitude, if it differs in shape etc. In both approaches however, none of them proceeds to symbolize the fuzzy system's output into symbols by characterising the time series curve.

A clustering-based AD was implemented in [22], where the extracted subsequences are given a linguistic description

and are clustered into normal and abnormal with the use of a fuzzy-c-means algorithm. As seen, a fuzzy system can be used both as a symbolic representation scheme and in AD. However, without prior knowledge of what constitutes an anomaly in the time series sequence, it is difficult to assign confidence levels to the output anomaly MF of the Fuzzy System. Therefore in our case, the Fuzzy System does not perform AD but it encodes the time series to symbols. Following that, the AD brute-force scheme is incorporated with the Fuzzy System. In our case it plays an active role in linking the appropriate Fuzzy representation by configuring its parameters to perform the final AD. Given enough prior knowledge or used in conjunction with another AD scheme the system can achieve very promising results.

3 DATA SELECTION

3.1 Benchmark Datasets

The experimental evaluation is split into two segments: (i) Evaluation under benchmark data and (ii) evaluation under real-world EM data gathered by the SWARM Satellite constellation.

Twelve benchmark datasets were downloaded from the physionet website ¹ that hosts publicly accessible databases from a variety of physiological datasets. The data used are from the chfdb database that contains heart failure data and the mit database that contains arrhythmia data. A hierarchical structure for the selection of the data had to be used in order to evaluate the algorithm under diverse categories. These categories are :

- High-Amplitude
- Low-Amplitude
- Complex Anomalies

Complex anomalies refer to anomalies in frequency, shape, amplitude or all of the above within a single sequence. Finally, the type and exact locations of the anomalies in the datasets were marked by cardiologists.

3.2 Real World Dataset

The electromagnetic (EM) data used in our experiment were gathered by the SWARM satellite constellation. Such new data offer an opportunity for the detection of EM anomalies during or before seismic activity. The three identical satellites follow a polar orbit at a height of 450km for SWARM A and C and at 530km for SWARM B and have a revisit time of 4 days on average. A fundamental question that remains unanswered is whether there is any correlation between seismic events and anomalies in the Earth's geomagnetic field.

The real-world dataset starts from 31st March 2014 until 25th February 2015. In order to eliminate any possibilities of artificial anomaly introduction or anomalies caused by other sources, it was decided that three geographical regions will be investigated. The choice of a region has to conform to the following rules:

TABLE 1: Investigated seismic regions and their coordinates

Location, Magnitude	Epicenter (Lat, Lon)	Control Region (Lat, Lon)	Date
China, 6.2	27.189°N 103.409 °E	27.18 °N 113.4 °E	03-08-2014
Peru, 6.8	14.598 °S 73.571 °W	14.598 °S 83.571 °W	24-08-2014
Greece, 5.8	36.685 °N 23.706 °E	36.68 °N 33.706 °E	29-08-2014
U.K	54.59°N 5.93 °E	N/A	N/A

- Its amplitude has to be above or close to a momentum Magnitude (M_w) of 6 as defined in ². This choice is in-line with the literature [13].
- The seismic event must have occurred approximately in the middle with respect to the time of the available data to allow for an adequate time series dataset before and after to be extracted.
- Belong to a different region.

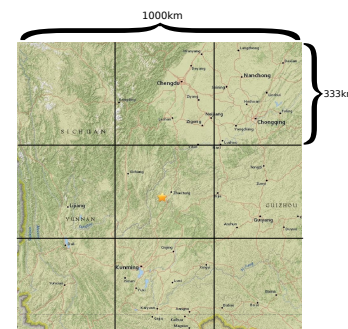


Fig. 1: China, Wupeng region of investigation

Fig. 1 depicts one of the selected regions and its dimensions. The number of readings gathered depend on the number of passes the SWARM satellites have made. The first region is defined as the true anomalous region where the epicenter of the seismic event was. If any anomalies are present they are expected to appear in this region. The second region is adjacent to the region of the seismic event and is regarded as the control region. This region is selected to eliminate the possibility of other seismic events causing EM anomalies to "leak" into the adjacent (control) region we investigate. A third region has also been defined, a seismically dormant region is regarded as the ground truth. The regions separated in main and control for the experiment and their details are shown in Table 1.

Now we can observe some simple rules: If there is an anomaly present in both the control and the anomalous region then a second seismic event is probably the root of it. If there is an anomaly present only in the anomalous region then the seismic event that is currently under investigation is likely the cause of it. If there is an anomaly only in the control region then another event, which we do not examine, likely took place. If anomalies are present in all three regions then the disturbance is due to other unknown sources such as EM storms, human-related emissions etc. and is discounted.

The SWARM observations are in the Common Data Format (CDF) and consist of 22 fields that provide information such as coordinates, altitude, date/time, error correction

1. <https://physionet.org/cgi-bin/atm/ATM>

2. <https://earthquake.usgs.gov/learn/glossary/?term=seismic~moment>

vectors etc. The observations used are from the Vector Field Magnetometer (VFM) of each SWARM satellite. The VFM instrument, measures the earth's geomagnetic field. The geomagnetic field can be best described by three orthogonal components. Let B , be the geomagnetic field. X , is the northerly intensity, Y the easterly and Z the vertical. The Eq. 1 below gives a better understanding on how to convert the three separate values to the EM vector.

$$|\vec{B}| = \sqrt{b_X^2 + b_Y^2 + b_Z^2} \quad (1)$$

The process followed for the data preprocessing is the following:

- Extract the values from the coordinates that belong to the region of study.
- Convert the X, Y, Z intensity vectors to a single vector.
- Subdivide the region of investigation into a grid and create nine different vectors from each grid point.

4 DEFINITIONS OF TERMINOLOGY

Before delving deeper, we have to give the definition of some of the terminology used.

Definition 1. Time Series: A time series $T = t_1, \dots, t_m$ is an ordered set of m real-values.

Definition 2. Subsequence: Given a time series T of length m , a subsequence C of T is a sampling of length $n \leq m$ of adjacent position from T , that is, $C = t_p, \dots, t_{p+n-1}$ for $1 \leq p \leq m - n + 1$.

Definition 3. Non-self match: Given a time series T , a subsequence C of length n beginning at position p and a matching subsequence M beginning at position q , we define that M is a non-self match to C if the condition $p - q \geq n$ holds true .

Definition 4. Anomaly: Given a time series T , the subsequence D of length n beginning at position p is said to be the highest anomaly of T if D has the largest distance to its nearest non-self match.

Definition 5, Euclidean Distance: Euclidean distance (EDist) is a function that takes two subsequences C and M of length m and returns a non-negative result Z , that is said to be the distance between pairs of symbolic values from C and M . The EDist also fulfils the symmetric property, that is $EDist(C,M)=EDist(M,C)$ and it is defined as follows:

$$Z(M, C) = \sqrt{(M_1 - C_1, \dots, M_m - C_m)^2} \quad (2)$$

Definition 6. Triangular Fuzzy Membership Function: A membership function (MF) for a fuzzy set S on the universe of discourse X is defined as $\mu_A : X \in [0, 1]$, where each element of X is mapped to a value between 0 and 1. This value, called the membership value or degree of membership, quantifies the grade of membership of the element in X to the fuzzy set S . A Triangular membership function therefore is defined by a lower limit a , an upper limit b , and a value m , where $a < m < b$.

The Fuzzy System is used as a symbolic approximation system that provides three different outputs, each with its own distinct symbolic approximation. The first is based in the shape conditions of the time series, the second is amplitude-based and uses an equiprobable segmentation of

the y-axis and the last one is amplitude-based but on the equal segmentation of the y-axis. The work is an extension of [23] and while it is known that an equally segmented space introduces a different amount of bias to the symbolic approximation, it addresses the unknown generative process of the signals by making no assumptions as to whether a Gaussian distribution is the generative process.

At the start of the process, each time series subsequence is normalized between $[0,1]$ before attempting to symbolize and before calling the distance function. It is well understood that in all settings, it is meaningless to compare time series of different variance [8] [17].

5 PROPOSED METHOD

5.1 Peak Finding

The core mechanism of the algorithm is its peak finding function. A lot of the functionality of the method depends on the detected peaks that act as one of the inputs to the fuzzy system (the second being the gradient). The standard approach to peak finding is to use a smoothing function. A de-noising function helps smooth the signal in order to avoid getting trapped in local minima that result from noise. The choice of the smoothing function depends on the domain and the nature of the data. In this case, a Savitzky-Golay (S-G) filter was the selected method. It is accepted that an attractive property of the S-G filter is the ability to provide both: (i) a thorough de-noising method, and (ii) preserve the peak-shape of time series signals [4]. For that reason it is now used in medical research studies [1] and this is also the reason it was used as the first step in the proposed algorithm.

Before applying the filter, a min-max normalization is applied to the data seen in Eq. 3.

$$m' = \frac{m - \min(T)}{\max(T) - \min(T)} \quad (3)$$

This normalization maps a value m of a time series T to a value m' in the range of $[0 1]$. The peak finding process creates two vectors, vectors $A \subset T$ for the Amplitude, and $G \subset T$ for the gradient. In that case, vector A consists of i peak elements and vector G consists of g gradient elements. An element $g \in G$ can be computed by the following equation and is used as the second MF input,

$$g = \frac{\partial m'_i}{\partial i} \quad (4)$$

Additionally, the peak distance is a critical parameter for the system since it is the only configurable parameter. The system changes the length of one identified peak to the next adjacent one, with the aim to find the optimal length that will best resemble the shape morphology of the original. With each iteration a different signal is fed to the fuzzy system which in turn provides a new representation. The bruteforce method used provides a way to find a solution to identify the correct peak distance that provides the representation which more accurately resembles the original data.

5.2 Fuzzy System

A Fuzzy Inference System (FIS) processes information in a similar way to human language. It uses imprecise linguistic terms to provide precise (crisp) answers. As such, it is especially useful when a system combines vague terms or uncertainties. Their advantage is the ability to incorporate the human reasoning into a system that responds by turning uncertainty into appropriate actions. Fuzzy reasoning comprises of three steps which will be discussed in detail in the following sections. These steps are : (i) Fuzzification, (ii) Rule Inference and (iii) Defuzzification.

In the proposed system, uncertainty refers to two parameters: (i) the amplitude, A , and (ii) the gradient, G , of the process. Each of these two parameters can be described by fuzzy linguistic terms. The fuzzy terms that are applied to the first input can be seen in Fig. 2.

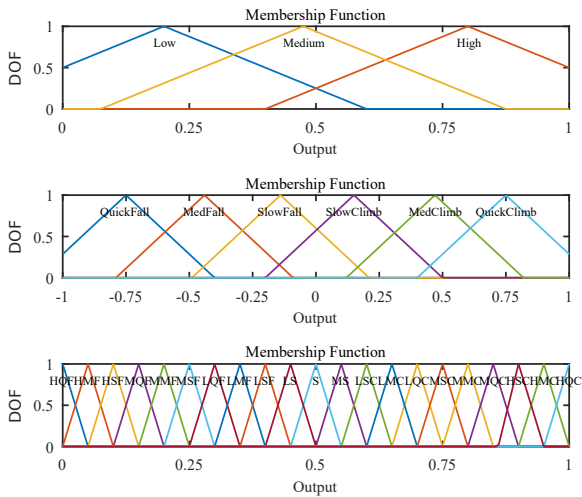


Fig. 2: Top to Bottom: The two input Membership functions used for the Amplitude and the Gradient and the output

Fig.2(top) shows the MF, A , used for the min-max normalized peak amplitude in the interval of $[0,1]$. The gradient MF, G' , shown in the mid figure is set in the interval of $[-1 +1]$ and the output MF, O , that signifies their relationship. The relationship of the MFs leads to a complex rule-base system with twenty one rules in total. A detailed description is given in the following sections.

5.2.1 Fuzzification

The first issue to consider when constructing a FIS is the cardinality of the MFs in the universe of discourse of the input values. A high number of MFs increases the complexity of the system, introduces more rules and increases the time complexity for the computation of the results. Each value of the universe of discourse must belong to at least one of the fuzzy sets so that every input will fire a rule at least once. The optimal choice of the cardinality is based on the application at hand and, in the case of our work it is a matter of precisely modelling the input variables to the system.

A problem that often comes with a FIS is the choice of MF. Choices range from triangular, gaussian, rectangular etc. functions. The right selection of an MF is problem specific and plays a significant role in the overall performance of

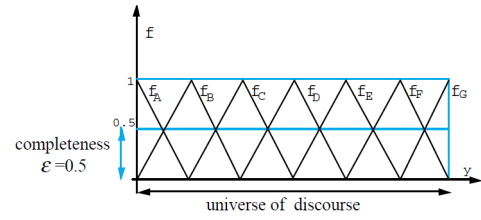


Fig. 3: Completeness of Membership Functions [7]

the system and its fuzzy representation. A rule of thumb for the selection of the MF is to match the behavior of a variable. In this case the peak or the shape are modelled with a triangular MF. Our choice is further justified by the well established fact that in practice triangular functions are able to model most processes very well [16].

An important characteristic of the MFs is their overlap, known as *completeness*. For clarity the definition of completeness is provided.

Definition 7: Completeness The completeness E of a set S of F fuzzy subsets, $S_i(m')$, $i = 1, 2, \dots, F$ that describes the characteristics of the normalized input variable m' is defined as the minimal level of E for which the E -level covers the input variable set T [6].

$$\forall m' \in U \exists 1 \leq i \leq F S_i(m') > 0 \quad (5)$$

Furthermore, a measure of completeness can also be defined that measures the fuzzy partition in Eq. 6.

$$E(m') = \sum_{i=1}^F S_i(m') \quad (6)$$

A visual interpretation of the level of completeness or overlap can also be seen in Fig. 3 where a 0.5 overlap is used between MFs.

Based on some heuristic considerations, authors in [9] argue that a completeness level ranging from 0.25 to 0.5 provide the best results. A completeness level of 0.5 was selected in this work because, for every input value i , there is bound to be a dominant rule with a selected membership grade of 0.5 or above, something that can be verified from Fig.2. A higher level of completeness can lead to redundancy and destabilize the system. It also becomes clear that when a lower level of completeness is selected the universe of discourse is characterized by more regions with a low maximal degree of truth of the rules they activate and the system becomes unstable. Moreover, an incomplete FIS, with no-overlapping MFs in both input and output, reduces the FIS to the function of a boolean logic system [7].

5.2.2 Rule Inference

The uncertainty that is contained in a fuzzy inference system (FIS) is evaluated through simple If-Then rules. The precedent part of these rules conforms to a nonfuzzy condition that is possible to express a conclusion in the antecedent part once this condition holds true. Consequently, a fuzzy rule can be obtained by combining two crisp rules. More specifically, let us consider the rule,

$$\text{If } i \in A' \text{ and } g \in G' \text{ then } o \in O \quad (7)$$

where A', G' are ordinary subsets in the domains A , the amplitude, and G , the gradient respectively. O represents their combination in the domain of W and is the set that describes their relation. The descriptive analysis of the curve in linguistic terms using these rules leads to a complex rule-based system. For example, one rule can be that Low Magnitude (Peak) AND Quick Fall (Gradient) gives an output of LQF (Low Quick Fall), which can be verified by the actual MFs seen in Fig.2.

Assuming there is prior knowledge of the value of g in relation to G' , this kind of information can be formally defined as

$$\pi_g(G') = \begin{cases} 1 & \text{if } g \in G' \\ \kappa < 1 & \text{if } g \notin G' \end{cases} \quad (8)$$

This information expresses the degree of certainty or probability distribution that $g \in G'$. A higher value of κ means that there is a higher probability that $g \notin G'$ and less that $g \in G'$. Conforming to the same model used previously, the rule-based inferring model can also be described as:

$$\pi_{o|i,g}(W, A, G) \leq n(\mu_{A'}(A)) \vee n(\mu_O(W)) \vee n(\mu_g(G')) \quad (9)$$

This translates as: if $i \in A$ it is certain that the value of i is in O and if $g \in G$ then the value of g is in O .

$$\sup_G(\min(\pi(w)), \pi_{o|i,g}(W, A, G)) = \sup_{i|g}(A, G \leq \max(n(\mu_A(u)), \mu_B(u), \kappa)) \quad (10)$$

Now, Eq. 10 means that the probability distribution obtained, expresses that the possible values for o , when the value of i is in A , are in O with a probability of almost equal to 1 and outside O , with a probability at most κ . The fuzzy rule obtained can then be finally translated to a crisp value with an uncertain conclusion of the form: "if $i \in A$ then o is $(1 - \kappa)$ -certain". Table. 2 shows the rules that were used to describe the function of the fuzzy system used for this work [3]. All twenty one rules give a description of input to output and decide which rule is fired each time the two input values of the amplitude and the gradient are input into the system.

TABLE 2: Decision Table for the FIS

AND		Magnitude		
		Low	Average	High
Gradient	Quick Fall	LQF	AQF	HQF
	Average Fall	LAF	AVF	HAF
	Slow Fall	LSF	ASF	HSF
	Stable	LS	Stable	HS
	Quick Climb	LQF	AQC	HQC
	Average Climb	LAV	AAC	HAC
	Slow Climb	LSC	ASC	HSC

5.2.3 Defuzzification

Defuzzification involves the process of transposing the fuzzy outputs to crisp outputs. There are a variety of methods to achieve this, however this discussion is focused in the method utilized in this work. The method, known as Center of Gravity (COG), is a method of averaging the output

values by calculating centroids of sets. The output MFs to which the fuzzy outputs are transposed are restricted to a set containing a single object (singletons).The fuzzy outputs are transposed to their membership functions similarly as in fuzzification. With COG the singleton values of outputs are calculated using a weighted average method. The crisp output is the result and is passed out of the fuzzy inferencing system for the symbolization process.

5.3 Symbolic Representation

After the three steps finish, the representation has produced a single value, i.e 0.45. In fuzzy linguistic terms this is translated to Low Quick Fall. This provides a basic description of the encoded subsequence based on the peak and the gradient that defines it. In this system three different encodings are used. The purpose of using different encodings is the additional information they provide, which comes intuitively. That is, when an anomaly is detected in the shape-based encoding, it implies a shape related anomaly, whereas an anomaly detected in the amplitude encoding, implies an amplitude related anomaly.

With the limitation set by the real world problem, in [21] two different approaches were proposed. The first encoding uses simple equidistant breakpoints to encode the signal based on each value $i' \in A'$, without any assumption. The number of break points were decided based on the empirical evaluation and can be seen in Fig 4. The encoding translates into: [0,0.2]=E, [0.2,0.4]=D, [0.4,0.6]=C, [0.6,0.8]=B and [0.8,1]=A.

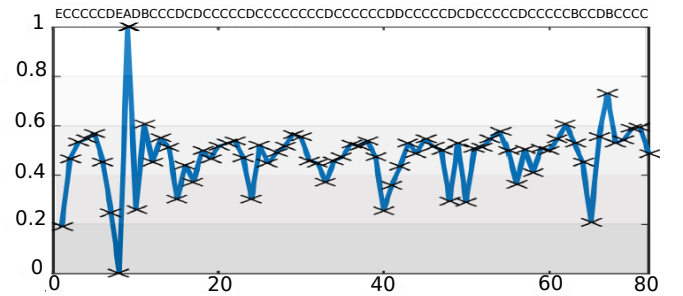


Fig. 4: The equally segmented space for the Amplitude Representation

As a continuation of the experimentation with the real world signals, an additional encoding is proposed. This time the assumption followed for the encoding is that the signal follows a Gaussian distribution. While this may not always be true for EM signals [18], it is required to better understand their nature and eliminate any unidentified possibilities. Therefore, a segmentation of the x-axis has to be obtained based on the Gaussian distribution. We therefore define the breakpoints accordingly.

Definition 8. Breakpoint: A list of numbers that indicate the bounds under the normal density curve from $[\infty, n]$. If m , is a single observation of a sequence T , then it is normally distributed in the non-standardized normal distribution with mean μ and a standard deviation of σ .

Therefore it is easily understood that the values used to calculate the break points are calculated from a Gaussian

Distribution table, scaled by the bounds set by μ and σ . The general formula of the cumulative distribution of the normal distribution that the breakpoints are computed from can be seen in Eq.11

$$F(m) = \int_{-\infty}^m \frac{e^{-(m-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \quad (11)$$

The shape encoding is based on an intuitive description of shapes. By taking into account three consecutive values $o \in O$ that the FIS has produced one can define the most elementary shape in a 2D space. As already mentioned, both amplitude and gradient information are incorporated in the output o . Therefore, eight different configurations can be defined. As a consequence, the whole time series is broken down and analyzed by those shapes. The defined shapes that were used as part of the encoding can be seen in Fig. 5.

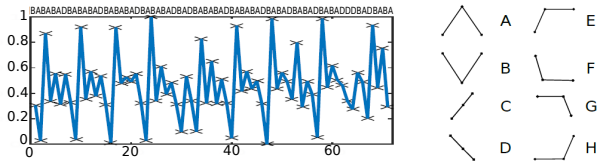


Fig. 5: The Shape representation scheme

5.4 Evaluation Metric

In order to evaluate the efficiency of each algorithm an overview of the used metric is presented in this section. A more detailed description was given in [5]. It is known that precision, recall and F1-Score ignore the True Negative (TN) values and can have a positive bias. Nevertheless, our interest lies specifically in the positive class, meaning the class or set of correctly or incorrectly predicted anomalous cases or the True Positive (TP) and False Positives (FP). Therefore, this set is used as the main factor in the definition of our accuracy measure. In all, the defined accuracy gives the average error rate for the correctly predicted values based on a factor of the window length.

In the case of AD for seismic AD, our interest lies in the localization of the anomaly. In other words this means to be able to define an accuracy metric between the expected actual anomaly (date/time of the seismic event) and that identified by the algorithm. To our knowledge, there has not been a definition of a metric that combines these characteristics. In pursuit of covering this gap, the proposed metric uses the widely known F1-Score, by additionally using the length of the anomaly in the form of a window to compute the difference between the predicted and the true (expected) anomalous locations. The proposed metric is defined by the following equation:

$$R = \sum_1^l \frac{F_1 - (Z_l \times F_1)}{N} \quad (12)$$

where,

$$Z_l = \frac{|t_l - p_l| * W}{2} \quad (13)$$

where, for each anomalous location l a degree of accuracy R is calculated by subtracting the true location, t_l , from the predicted, p_l as a factor of the window length, W . The index, l is defined as the mid-point of the anomaly of length W . The value Z is consequently used to calculate the proportion of F1-score divided by the value N that represents the TP.

5.5 Anomaly Detection

With the symbolic representation of the time series, the time series is converted to one symbolic sequence. A common method to detect anomalies is to define subsequence S of a window length n and calculate the EDist among all the possible subsequences of the same size. A brute-force algorithm specifically tuned for AD is used in our case. In this work, we do not assume any prior knowledge, therefore the use of a brute-force approach for AD without a predefined anomaly length is used for the purpose of our proposed algorithm. The brute-force algorithm is shown in the following figure:

Algorithm 1 Proposed Brute-force Algorithm

```
figure
procedure BFDD(BruteForce(T,n,scale))
  for  $p=1$  to  $T-n+1$  do
    NearestNeighborDist =  $\infty$ 
    for  $q=1$  to  $T-n+1$  do
      if  $p - q \geq n$  then
        if NearestNeighborDist = 0 then
          break
        else  $pdist(T_{(p...p+n-1)}, T_{(q...q+n-1)}) < N$ 
           $N = pdist(T_{(p...p+n-1)}, T_{(q...q+n-1)})$ 
        end if
      end if
    end for // end inner loop
  end for // end outer loop
  if NearestNeighborDist > dist then
     $dist(p) = nearest$ 
     $loc(p) = p$ 
  end if
FinalLocs = actualLocation =  $(T, loc, scale)$ 
```

The algorithm has two loops one inner and one outer. In the outer loop a subsequence is selected and is compared against every subsequence with equal length. The last distance is then stored and the algorithm proceeds to the inner loop. Only the subsequences which have higher distance than the previous ones evaluated in the loop are considered as anomalies. We can make two observations here, the first one is that in fact we do not need to evaluate each and every distance in the inner loop. Once the distance drops to zero, we can safely assume that this is not an anomalous subsequence and proceed to the outer loop. A second addition is to be able to detect more than one anomaly.

Admittedly, a brute-force method is a naive method to traverse through the search space. In this work however, the aim is twofold: (i) Multiple anomalous subsequences can be detected but more importantly variable length subsequences can be detected. Depending on the length n of the subsequence that is input to the brute-force algorithm, one can identify the distance between peaks that is used by the peak finding algorithm. This information indicates the length of

the anomalous subsequence without prior knowledge as in the case of SAX. BFDD, a simple method that guarantees best results, helps understand how a representation method with AD helps identify different kinds of anomalies and how the reduced time series links back to the original raw time series through the anomaly localization.

For HOT-SAX, the only parameter that affects the actual results is the length of the anomalous subsequence. This parameter needs prior knowledge of the dataset. The same is stated in [8] therefore for a true no parameter AD method, the optimization needs to be focused on the length of the subsequence. A SAX word encompasses all the information that the subsequence length gives, since they are linked together. A single symbol is the length of the subsequence divided by the length of the word. In the solution proposed in this work, by using the brute-force algorithm we show an AD method to identifying both the location of the anomaly and the peak distance (window length) for the representation method with minimal prior knowledge.

6 PERFORMANCE EVALUATION

6.1 Benchmark Algorithms

To evaluate the effectiveness of the proposed algorithm the strategy followed in this section is to assess their performance with similar methods that require a symbolic approximation and are considered state-of-the-art. One of the main issues to be addressed in the real datasets is the non-periodicity of the data. Due to the nature of data collection, there is a varying length of gaps between observations. In fact, most algorithms consider periodic data, something that is easy to address with a predefined window for the symbolic approximation. In that respect we decided to evaluate algorithms that use both approaches. Fuzzy and GrammarViz can address both periodic and non-periodic data while HOT-SAX addresses only periodic data.

HOT-SAX utilizes a PAA to compress the signal from the original data points. It requires three parameters: (a) the length of the subsequence to consider, (b) the word length, w , for the encoding and (c) the alphabet a , that segments the distribution space into a Gaussian equiprobable segments. In contrast, GrammarViz does not require a word length parameter.

From a statistical point of view, the CUSUM-EWMA (CE) is not constrained by the length of the anomaly. However, it also requires parameter tuning. The two parameters used are the CUSUM statistic, K , and the number of past values we are going to consider based on the EWMA statistic, λ . The overall objective is to illustrate whether current more complex methods with less parameter tuning are appropriate for AD and how they perform in a constrained real-world environment. All algorithms are compared by using a novel metric, R , introduced in [5] and summarized in Section 5.4, which takes into account the subsequence length, the predicted by the algorithm anomalous location and the true anomalous location.

6.2 Benchmark Datasets

An increasing order of complication based on each kind of anomaly is applied in the benchmark experiments. For the

Fuzzy method, the upper limit for the peak finding distance was set from 10 to 150. Moreover, in HOT-SAX, GrammarViz and the Gaussian-based fuzzy representation the alphabet size is hardcoded to 3. It is known that a higher number of breakpoints makes the AD more complex by unnecessarily increasing the Euclidean distance without carrying any information. In [8] the authors point out that during their experiments hardcoding the alphabet size to 3, produced the best results but in cases with a high variability over time a higher number might be preferable. Both parameters, the subsequence length and the alphabet size, do not affect the results of SAX but affect the efficiency of the algorithm. In GrammarViz there were different results among different starting conditions. Although the length does not affect the AD, it affects the location and it is useful to evaluate the algorithm under different lengths and starting conditions based on our metric.

More specifically, for HOT-SAX the configurations examined were based around the two tunable parameters: the length and the word. The best length/word configuration is selected and compared against our algorithm. A variable anomaly length from 100 to 250 was used as the length N and four different word configurations, based on how many data points are encoded into a single symbol from 3 to 7 were tested. The word length, w is used as the input to the brute-force algorithm.

For GrammarViz the same conditions as in HOT-SAX were selected. As mentioned earlier, the algorithm can find variable length anomalies and the starting condition serves only as a starting point. Lastly, for the CUSUM-EWMA, the EWMA statistic was set to a length from 10 to 100, this ensured that the evaluation of the algorithm used a low to a high memory statistic and the CUSUM statistic was tuned to start from 1 to 100. A varying threshold was set from 0.1 to 1 and the best result was selected.

The use of the benchmark datasets will help us understand the process and find the optimal configuration for all algorithms. In the fuzzy algorithm the configuration leads to an estimate of the length of the anomaly via the peak distance parameter, which will be adjusted for the real-world data accordingly.

6.2.1 High Intensity Anomalies

The high intensity case with similar anomalies as seen in Fig. 6 was proven trivial for all algorithms. In the general case, all anomalies are visible. The core components of each algorithm, the PAA for SAX, SAX for GrammarViz and peak finding for the Fuzzy, are able to easily decipher the interesting cases and accurately model them on to the representation stage.

The CUSUM-EWMA however is inconsistent as it detects many false positives and gives less confident results. However, the dataset shown in Fig. 6 proves challenging even for the more intelligent methods. In that case the CUSUM's AD remains acceptable because it has limited variation in its FP/TP ratio. Even though GrammarViz is based on SAX for its symbolic representation and discretization it offers a better detection scheme. GrammarViz's variable window length can better model the more inconspicuous change in the 2160th data point, pictured in Fig. 6. Both Fuzzy and

SAX had issues with this specific anomaly. The detection for SAX is difficult due to the insignificant change in the mean and a fixed window length does not help to point out the anomaly. As a result, the outcome is not visible in the representation. For the Fuzzy method, some True Positive (TP) values appear due to the inflexibility of the peak finding algorithm when it comes to similar amplitude changes, as there is no threshold or a tolerance level to reduce its peak selectivity in terms of amplitude. However, the anomaly is detected because the periodicity of the signal also changes something that makes it easier for the algorithm to detect it.

Table 3 shows the best accuracy and its configuration. It also provides a picture of how the accuracy metric works. As seen in the magnified graphs around the anomalies in Fig. 6 the algorithm that detects the anomaly in a further point than where the truth is has a worse accuracy calculated by the R metric, Eq. 12.

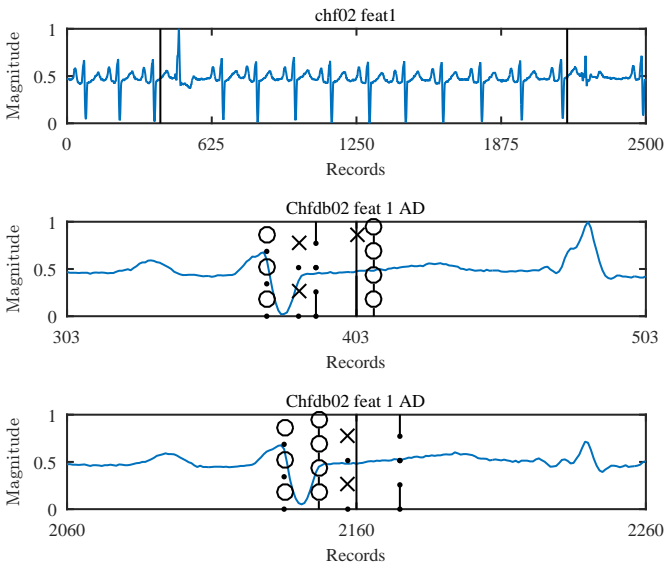


Fig. 6: AD achieved for the chf15 feat1 dataset: SAX(-·-), Normalized Amplitude(×), Amplitude(·-·), Shape(-×-), CUSUM-EWMA(·○), GrammarViz(-○), True anomalous(-)

TABLE 3: The R Metric,Eq. 12 for the High Intensity Case

	Amp. (w,n)	Norm.Amp. (w,n)	Shape (w,n)	SAX (w,n)	CE (k,N)	GrammarViz (w,n)
Chf01 feat1	(21,2)	(48,3)	(24,2)	(133,5)	(2,4)	(179,7)
Accuracy (R)	100	100	99	99	77	84.53
Chf02 feat1	(21,4)	(13,6)	(15,2)	(155,6)	(17,179)	(118,5)
Accuracy (R)	85.5	88.5	66.67	66.67	41.14	85.38
mitdb101 feat2	(13,8)	(30,5)	(25,5)	(200,4)	n/a	(131,4)
Accuracy (R)	97	99	66	80	n/a	98.76
mitdb108 feat2	(38,2)	(128,6)	(126,4,6)	(250,5)	(5,191)	(194,3)
Accuracy (R)	77.2	56.33	67.6	47	51.6	51.56

6.2.2 Low Intensity Anomalies

Fig. 7 provides an example of AD in the low intensity datasets. Consistent with the previous results, SAX achieves lower accuracy even when it detects the anomalies. The issue identified is that with each different configuration, the approximation causes significant loss of information because the mean used by the PAA component tends to smooth the curve. The level of smoothing causes the representation to be

less sensitive to abrupt changes, thus making it more prone to misrepresenting important changes in the time series. Moreover, when the sequence has a high variability as in the case shown in Fig. 7, SAX and GrammarViz misidentifies the interesting pattern. Interestingly, the CE is able to succeed in this example because the process has a dramatic change in its mean which is easy for the statistic based algorithm to detect.

In the same case, even though the fuzzy representation also loses information the algorithm compensates by applying a representation based on the original time series via the peak finding. The use of both the amplitude and additionally the gradient, provides a more robust AD as seen in Table 4.

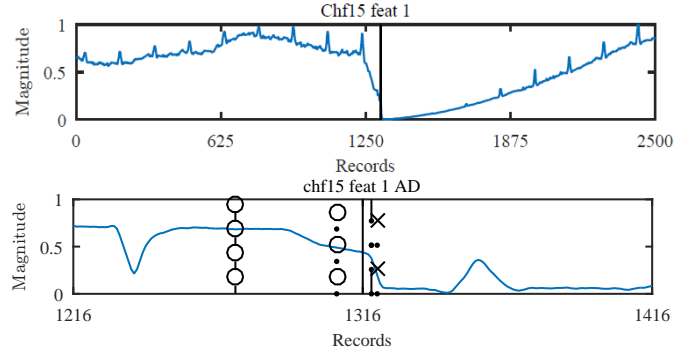


Fig. 7: AD achieved for the chf15 feat1 dataset: SAX(-·-), Normalized Amplitude(×), Amplitude(·-·), Shape(-×-), CUSUM-EWMA(·○), GrammarViz(-○), True anomalous(-)

TABLE 4: The R Metric,Eq. 12 for the Low Intensity Case

	Amp. (w,n)	Norm.Amp. (w,n)	Shape (w,n)	SAX (w,n)	CE (k,N)	GrammarViz (w,n)
Chf01 feat2	(17,2)	(44,2)	(27,5)	(137,4)	(8,186)	(107,6)
Accuracy (R)	100	92	90	100	46	92.18
mitdb101 feat1	(15,10)	(58,6)	(11,4)	(158,4)	n/a	(150,4)
Accuracy (R)	66.67	66.67	99	89	n/a	100
chf15 feat1	(15,10)	(59,2)	(15,4)	(145,5)	(2,119)	(199,4)
Accuracy (R)	97	95	88	78.5	100	66.13
chf15 feat2	(13,8)	(30,5)	(17,9)	(112,4)	(4,164)	(172,4)
Accuracy (R)	89.5	90	92.5	58.9	93	49.42

6.2.3 Complex Anomalies

The complex anomalies proved to be challenging for all algorithms, something that can be seen in Fig. 8. In a similar scenario to the lower case shown in Fig.7, algorithms get trapped in the local trend change around the 1500th data point, seen in Fig. 9 and misidentify the anomalies. However, the variable window in GrammarViz proves to be an important characteristic that improves the detection rate.

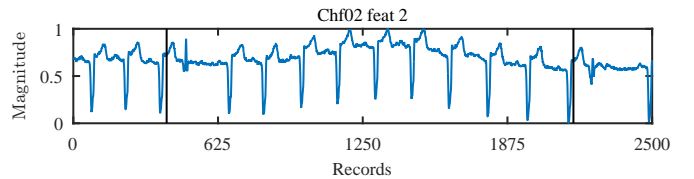


Fig. 8: The chf02 feat 2 dataset with its anomalous subsequences marked

The scenario with three anomalies of variable length and shape is shown in Fig. 9. When presented with different kinds of anomalies within the same dataset, the fuzzy representation retains a more faithful approximation in the encoding stage. Both the amplitude and the gradient, provide the necessary information to better represent differences in frequency and amplitude. The accuracy can be misleading since the shape and the amplitude representations provide different anomalous points. One can identify that in point 4083 the anomaly is amplitude related, whereas in point 10014 the anomaly is shape related. This kind of information is useful to understand the nature of the change in the time series and it can be exploited when processing the real-world datasets. It is important to notice that GrammarViz also works well in contrast to SAX and understand that the variable window provides a good improvement for non periodic signals in AD.

On the other hand, SAX, fails to impress when challenged with an unknown length and a different type of an anomaly other than amplitude related such as frequency related. Dataset mitdb100 provides such a case, and GrammarViz and SAX have some difficulties. It seems more often than not that the SAX can be a hampering factor because its representation mode only takes into account the mean and not any other factors such as the gradient of a subsequence. The CE method, also using the mean for the detection, is unable to perform AD-something that is to be expected. The results can be seen in Table 5.

TABLE 5: The R Metric, Eq. 12 for the Complex Case

	Amp. (w,n)	Norm.Amp. (w,n)	Shape (w,n)	SAX (w,n)	CE (k,N)	GrammarViz (w,n)
chf02 feat2	(51,3)	(54,3)	(36,6)	(150,6)	(4,91)	(179,3)
Accuracy (R)	66.67	66.67	66.67	66.67	58	81.96
mitdb100 feat1	(23,5)	(36,4)	(141,2)	(113,5)	n/a	(195,3)
Accuracy (R)	66.67	66.67	58	66.67	n/a	48.51
mitdb100 feat2	(24,5)	(21,6)	(40,3)	(189,6)	n/a	(173,6)
Accuracy (R)	65.3	65.3	65.3	56.67	n/a	52.25
mitdb108 feat1	(69,3)	(4,9)	(82,4)	(120,4)	(1,3)	(191,3)
Accuracy (R)	73.2	50	47	46.5	61.6	80.14

6.3 Real World Datasets

The real world data were gathered by defining a region similar to the one seen in Fig. 1. Each satellite collects a varying number of readings per date based on how long it stays above the defined region. This is a result of their different orbital trajectories. This behavior causes a data collection problem because it provides inconsistent data. Despite taking great care not to introduce artificial anomalies and making sure to use as many data points as available, it is difficult to eliminate information loss. In Fig. 10 artificial anomalies are clearly visible. In that case taking the median of each date, that has readings, introduces artificial anomalies. The same occurs when we take the mean for each date because of the variable number of readings problem. Four different cases were considered: (a) the mean, (b) the median, (c) the first data point of each day and (d) division by a user set divisor. After some consideration it was decided to utilize method (d) this introduces the smallest number of artificial anomalies. The scale of the experiments requires to run each algorithm in all 189 different time series leading to over one thousand results. Due to space limitations, we present the results from the central grid of each investigated region similar to Fig. 1. That is the most representative region and includes the seismic event where appropriate. Nevertheless, as the experiments were run for all the grids, similar conclusions can be drawn.

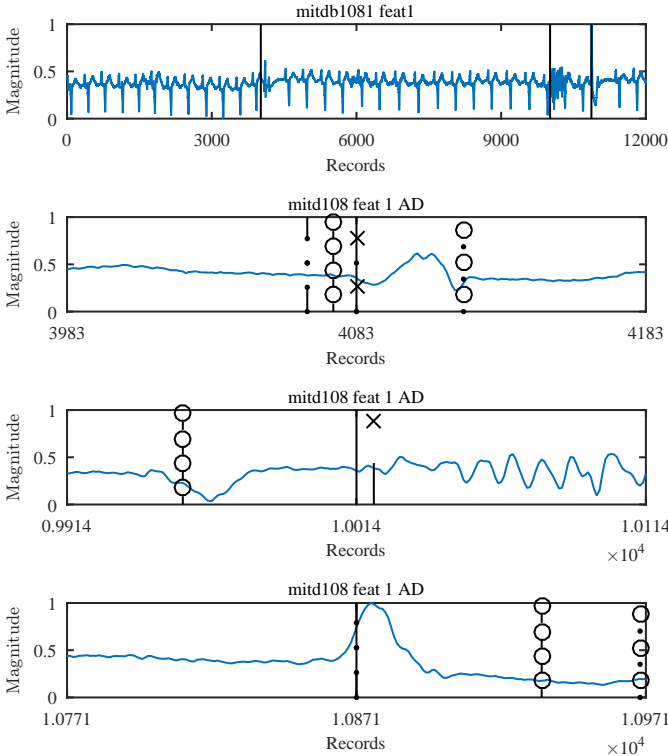


Fig. 9: AD achieved for the chf15 feat1 dataset: SAX(\cdots), Normalized Amplitude(\times), Amplitude(\cdots), Shape(\times), CUSUM-EWMA(\circ), GrammarViz(\circ), True anomalous($-$)

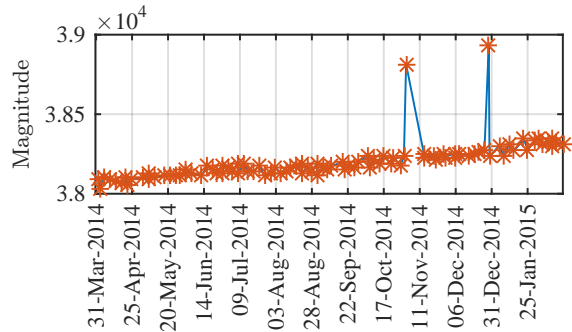


Fig. 10: Artificial anomalies caused by the median

The experiments carried out in this section derive knowledge based on the empirical results from the benchmark datasets. In that case, in each experiment with the benchmark datasets a common theme arises. Each algorithm has two parameters: (a) the subsequence window length chosen and (b) the total time series sequence length. This ratio can be used as a guiding principle for the parameters to be applied in the real world dataset. The ratio ranges from 12 to 20. This means that the anomalous subsequence is 12 to 20 times smaller than the entire sequence's length. In these

datasets we will apply a ratio of 12 times since this is closer to what was used in the smaller datasets with a length of 2,500, which is closer to the number of real world data points we could gather. This means that the window length is set to 6 in both SAX and GrammarViz with the word length set to 4 and the alphabet to 3. Similarly for the Fuzzy algorithm the peak distance is also set to 6.

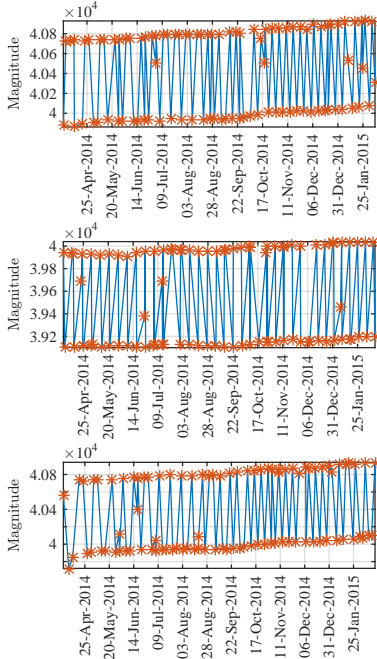


Fig. 11: Swarm A, B, C center grid for U.K

The CE also uses a length that is the memory element of the EWMA statistic and a similar ratio factor is used. This translates for the CUSUM statistic, k , to 1 and for the EWMA $lambda$ to 6.

As seen from the experiments, the results that prevail from the experiments carried in time series do not inspire confidence. All algorithms were tested in the benchmark datasets and their AD capabilities are well documented. However, in Table 6 the algorithms detect anomalies even when they should not, indicating that there is either: (i) a problem with the data collection, (ii) interference from other sources, (iii) False Positives from the algorithms (their drawbacks are already known from the benchmark data) or (iv) other aftershocks that were not taken into account originally. Because of the aforementioned issues, nothing can be suggested about the results with certainty.

In Table 6, the detected anomalous dates by all algorithms are shown. For China, the data range was from February to October. The seismic event occurred on 03/08/2014. Nevertheless, there is no specific pattern among the Swarm data or even among the main and the control regions that can be deciphered. While we know for a fact that all algorithms are able to perform AD convincingly as seen in the benchmark experiments, here we are presented with conflicting results. Similarly, for Greece’s 29/08/2014 seismic event in Fig. 14 and Table 6 a range of February to January can be seen. Lastly, for Peru the data from January to February can also be seen in Fig. 13. The detected anomalies

in Table 6 show no specific pattern once again. As already mentioned, the algorithms were specifically selected to address the unknown nature of the EM signals with respect to their duration, their amplitude and frequency changes. In the case of the real-world data the results fall short but nevertheless there are some interesting points to consider.

The main issue can be attributed to the real-world data sources. Even with the employment of artificial anomaly-avoidance techniques, the data are in reality very sparse. A revisit period of 4 days for a single satellite is not enough for the precise monitoring of EM variations above a specific region. To create an accurate model of normality there needs to be as much data as possible in order to provide high density and high resolution coverage. This will provide an almost continuous time series model that can be used to deliver consistent and confident results for our purpose. Solutions provided by data interpolation, or aggregation techniques, are employed to alleviate the sparse data problem but cause information loss themselves, thereby either unwittingly introducing new anomalies when they should not or causing others to fade. In the end, they are undoubtedly changing the landscape of the patterns of the original time series.

Furthermore, the use of three identical satellites to provide measurements for a single region was originally thought to provide more data if combined. Therefore it would enable us to overcome the problem of monitoring the daily EM variation above a specific region. On the contrary, because the satellites fly on parallels this was not the case. Their orbital paths instead of increasing data availability, produce data that are duplicates in terms of time and can only be used for validation purposes.

Moreover, SWARM B’s data alone cannot be used to counter this problem. Measurements and instrument calibration are different and they cannot be used to provide an immediate solution regarding data sparsity. Most of the issues are caused by the satellites’ data availability which is difficult to overcome. The amount of data available for each date also plays a significant role but it does not affect the processing as much as data availability. This issue can comfortably be addressed by the aggregation methods available as described earlier, and does not change the time series landscape.

7 CONCLUSION AND FUTURE WORK

The work presented, provides guidelines that can be used to: (i) avoid introducing artificial anomalies, (ii) counter any issues with data sparsity and (iii) show how variable length AD methods work in different cases by using a new metric.

The experiments show that the proposed method can model more accurately each kind of anomaly through the separate representation model in its three forms. In the more complex cases both SAX and GrammarViz disappoint because the PAA uses the mean, thus it only picks changes in the amplitude as shown by the experiments. When the anomaly becomes more complex than an amplitude change, even with prior knowledge of the length both SAX and GrammarViz have the same limitations. As for the CE algorithm, it comes as no surprise that it provides good AD

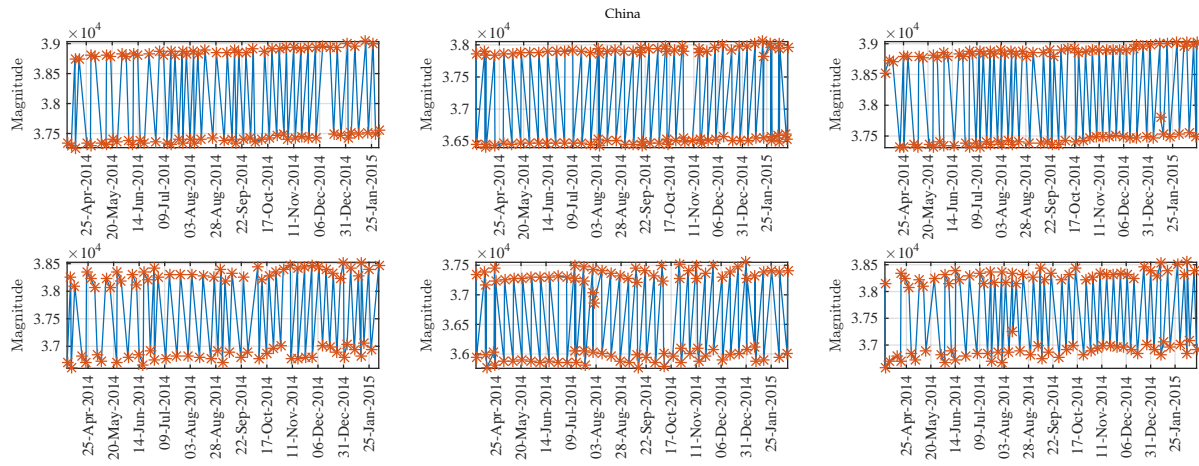


Fig. 12: Time Series for the China region

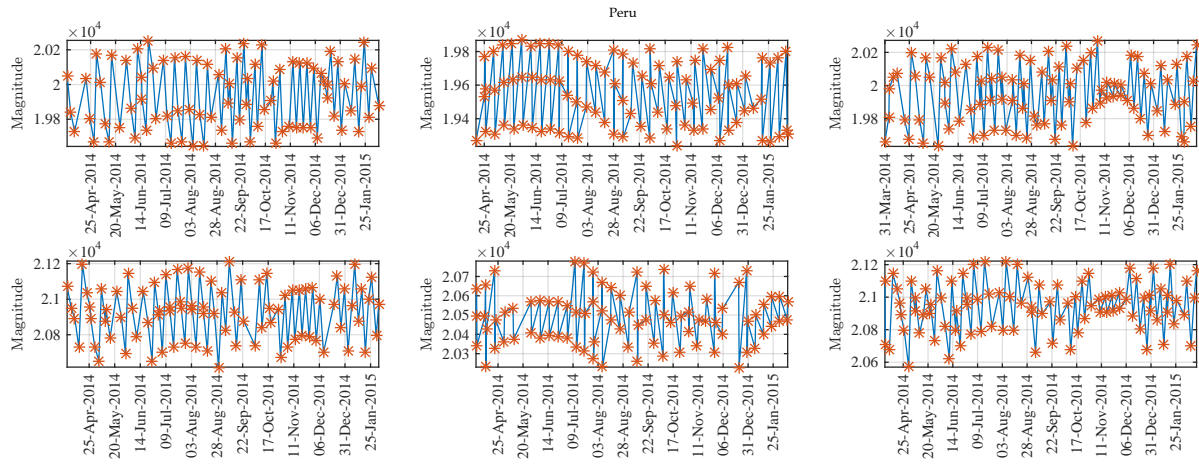


Fig. 13: Time Series for the Peru region

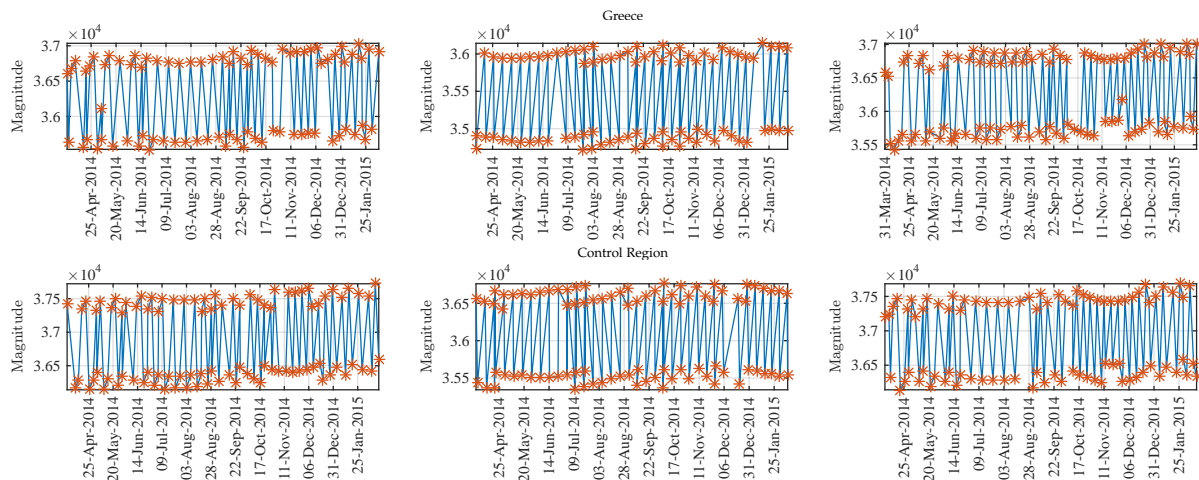


Fig. 14: Time Series for the Greece region

TABLE 6: Dates with EM anomalies as identified by the algorithms

Swarm A						
Method	China		Greece		Peru	
	Main	Control	Main	Control	Main	Control
CE	03-04-2014, 15-11-2014	12-10-2014, 05-08-2015	10-10-2014	12-09-2014, 19-10-2014	02-08-2014, 14-06-2015, 15-08-2014	14-02-2015
Amp	14-02-2015	21-09-2014, 15-08-2014, 21-08-2014	10-08-2014, 20-08-2015	25-09-2014	10-10-2014, 14-04-2015	20-09-2014, 14-02-2015
Amp Norm.	12-06-2014, 24-07-2014	20-05-2014	24-06-2014, 14-05-2014	14-12-2014	08-08-2014, 30-07-2014	29-10-2014
Shape	02-08-2014, 19-08-2014, 11-11-2014	09-09-2014	14-06-2014	24-09-2014, 07-08-2014	29-09-2014	13-09-2014, 03-08-2014, 20-06-14
GrammarViz	27-07-2014	23-05-2014, 17-12-2014, 21-07-2014	20-09-2014, 14-02-2015	20-09-2014, 14-02-2015	11-10-2014	15-07-2014, 14-02-2015
HOT-SAX	22-09-2014, 14-10-2014	09-09-2014	15-08-2014	13-06-2014	12-07-2014, 10-06-2014	05-09-2014
Swarm B						
Method	China		Greece		Peru	
	Main	Control	Main	Control	Main	Control
CE	20-09-2014, 14-02-2015	20-09-2014, 14-02-2015	20-09-2014, 14-02-2015	24-07-2015, 11-08-2014, 14-01-2015	12-01-2015	14-08-2015
Amp	01-09-2014	13-11-2014	20-09-2014, 14-02-2015	20-02-2015, 29-08-2014, 14-10-2014	01-07-2014, 15-07-2014, 14-02-2015	07-07-2015
Amp Norm.	13-08-2014, 10-07-2015, 05-09-2014	11-09-2014	04-06-2014, 19-07-2014	20-09-2014, 14-02-2015	20-09-2014, 14-02-2015	10-08-2014, 19-08-2014
Shape	04-08-2014, 11-08-2014	03-10-2015	01-08-2014, 15-08-2014	19-08-2014	20-08-2014	20-08-2014, 09-09-2014
GrammarViz	08-05-2014, 23-10-2014, 18-09-2014	26-06-2014, 18-09-2014	20-09-2014, 14-02-2015	01-08-2014, 19-08-2014	10-08-2014, 09-09-2014, 21-09-2014	20-08-2014, 14-09-2014
HOT-SAX	20-09-2014, 14-02-2015	15-09-2014, 22-09-2014	10-09-2014	24-09-2014	14-09-2014, 14-02-2015	28-09-2014, 04-11-2014
Swarm C						
Method	China		Greece		Peru	
	Main	Control	Main	Control	Main	Control
CE	20-09-2014, 14-02-2015	20-09-2014, 14-08-2014	20-09-2014, 14-02-2014	06-06-2014	02-10-2014, 14-02-2015	04-08-2014, 21-08-2014
Amp	11-10-2014	20-09-2014, 20-07-2014	10-08-2014	28-08-2014, 14-02-2015	05-09-2014, 14-09-2015	14-10-2015
Amp Norm.	20-09-2014, 05-03-2015	20-09-2014, 14-02-2015	22-09-2014, 06-10-2014	08-09-2014, 08-08-2014	14-07-2014	20-09-2014, 14-02-2015
Shape	20-06-2014, 14-07-2015	25-08-2014, 09-09-2015	20-08-2014, 14-02-2015	29-09-2014	11-01-2015	02-10-2014, 28-09-2015
GrammarViz	10-05-2014, 12-10-2015	30-05-2014, 26-10-2014	19-08-2014, 20-09-2014, 01-09-2014	04-09-2014, 14-08-2014	09-11-2014, 02-19-2015	02-11-2014
HOT-SAX	14-09-2015	15-09-2014	03-06-2014	05-09-2014	15-10-2014, 28-10-2014	02-09-2014, 14-02-2015

in terms of changes in the mean but it underperforms in different anomaly cases.

For the Fuzzy method, an important function to notice is that different anomalies are detected when using the different representations and one can decipher multiple anomaly types. This is a useful feature to understand the type of anomaly depending on the representation used. This can be further exploited in the future to provide a combined metric based on the kind of anomalies that are detected. However, there are also some shortcomings, specifically in the complex cases. Because of the complexity of the fuzzy system and its rule-base, the system becomes oversensitive to changes, something that is to be expected based on the research. One solution would be to exclude some of the descriptive fuzzy membership functions. A better solution that can be investigated in the future is to include adaptive membership functions based on the input variables.

In all, it is unclear whether there is a relationship between EM anomalies and seismic events. Despite the current research suggesting otherwise, in our case it is difficult to assess whether the shortfalls lie in the choice of algorithms, the data quality or both. Nevertheless, it is important to keep the issues separate.

Regarding algorithms, all kinds of anomalies presented in our work should be of equal focus. More specifically, the attention has to be on non-periodic data. The unknown nature of data and its collection methods have to be taken into account. In future work, more unsupervised algorithms that employ (true) free parameter tuning will be the focus of our research.

Furthermore, regarding data collection, a few points are worth mentioning. While satellites observe the whole earth, it can be said that their revisit time and orbit prove to be an insurmountable constraint. This fact alone is an issue in terms of data quality and data sparsity. As such, it might be worthwhile to rethink their use in terms of geomagnetic observations above a specific region.

Potential problems can be solved as in the case of SWARM by setting different orbital paths instead of fixed parallel flights. With the same configuration, data from more satellites can be combined. Satellite data quality and its shortfalls, is something that to our knowledge has not been addressed in our field and this study has provided some valuable insight into a real-world application.

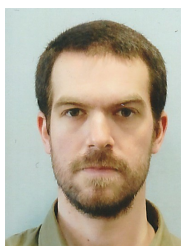
ACKNOWLEDGMENTS

This work is supported by the Dragon 3 program, a collaboration between the European Space Agency and the Ministry of Science and Technology of China. The authors would also like to acknowledge some additional support received from the project funded by the National Natural Science Foundation of China (Grant No. 41374077) and the European Space Agency for making SWARM data available for our research.

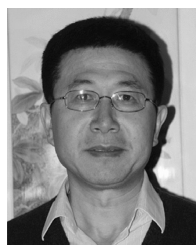
REFERENCES

- [1] Cunha, J. P. S., Choupina, H. M. P., Rocha, A. P., Fernandes, J. M., Achilles, F., Loesch, A. M., Noachtar, S. (2016). NeuroKinect: a novel low-cost 3Dvideo-EEG system for epileptic seizure motion quantification. *PLoS one*, 11(1), e0145669.
- [2] Kong, Xiangzeng, Yaxin Bi, and David H. Glass. "Detecting Seismic Anomalies in Outgoing Long-Wave Radiation Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8, no. 2 (2015): 649-660.
- [3] Dubois, Didier, and Henri Prade. "What are fuzzy rules and how to use them." *Fuzzy sets and systems* 84, no. 2 (1996): 169-185.
- [4] Schafer, R. W. (2011). What is a Savitzky-Golay filter?[lecture notes]. *IEEE Signal processing magazine*, 28(4), 111-117.
- [5] Christodoulou V., Bi Y., Wilkie G., Zhao G. (2016) Proceedings of Dragon 3 Final Results and Dragon 4 Kick-Off, held 4-8 July in Wuhan, China. Edited by L. Ouwehand. ESA-SP Volume 739, ISBN: 978-92-9221-304-6, p.84
- [6] Jin, Y. (2012). *Advanced fuzzy systems design and applications* (Vol. 112). Physica.
- [7] Bouchon-Meunier, B., Dotoli, M., and Maione, B. (1996). On the choice of membership functions in a mamdani-type fuzzy controller.
- [8] Keogh, E., Lin, J., and Fu, A. (2005, November). Hot sax: Efficiently finding the most unusual time series subsequence. In *Data mining, fifth IEEE international conference on* (pp. 8-pp). IEEE.
- [9] MIZUMOTO, M., "Fuzzy controls under various fuzzy reasoning methods", in *Information Sciences* 45, 1988.
- [10] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." *ACM computing surveys (CSUR)* 41, no. 3 (2009): 15.
- [11] Malinowski, Simon, Thomas Guyet, Ren Quiniou, and Romain Tavenard. "1d-sax: A novel symbolic representation for time series." In *International Symposium on Intelligent Data Analysis*, pp. 273-284. Springer Berlin Heidelberg, 2013.
- [12] Ren, H., Liu, M., Li, Z., and Pedrycz, W. (2017). A Piecewise Aggregate Pattern Representation Approach for Anomaly Detection in Time Series. *Knowledge-Based Systems*.
- [13] Petraki, Ermioni, Dimitrios Nikolopoulos, Constantin Nomicos, John Stonham, Demetrios Cantzos, Panayiotis Yannakopoulos, and Sofia Kottou. "Electromagnetic Pre-earthquake Precursors: Mechanisms, Data and Models-A Review." *Journal of Earth Science and Climatic Change* 6, no. 1 (2015): 1.

- [14] Moreno-Garcia, Antonio, Juan Moreno-Garcia, Luis Jimenez-Linares, and Luis Rodriguez-Benitez. "Time series represented by means of fuzzy piecewise lineal segments." *Journal of Computational and Applied Mathematics* (2016).
- [15] Wijayasekara, Dumidu, Ondrej Linda, and Milos Manic. Computational intelligence based anomaly detection for Building Energy Management Systems. No. INL/CON-12-25793. Idaho National Laboratory (INL), 2012.
- [16] Barua, Aditi, Lalitha Snigdha Mudunuri, and Olga Kosheleva. "Why trapezoidal and triangular membership functions work so well: Towards a theoretical explanation." (2013).
- [17] Chuah, Mooi Choo, and Fen Fu. "ECG anomaly detection via time series analysis." In *International Symposium on Parallel and Distributed Processing and Applications*, pp. 123-135. Springer Berlin Heidelberg, 2007.
- [18] Akhoondzadeh, M. "Application of Artificial Bee Colony algorithm in TEC seismo-ionospheric anomalies detection." *Advances in Space Research* 56, no. 6 (2015): 1200-1211.
- [19] Pimentel, M. A., Clifton, D. A., Clifton, L., and Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99, 215-249.
- [20] Li, Guiling, Liping Zhang, and Linquan Yang. "TSX: a novel symbolic representation for financial time series." In *Pacific Rim International Conference on Artificial Intelligence*, pp. 262-273. Springer Berlin Heidelberg, 2012.
- [21] Christodoulou, Vyron, Yaxin Bi, and Guoze Zhao. "A Fuzzy Inspired Approach to Seismic Anomaly Detection." In *International Conference on Knowledge Science, Engineering and Management*, pp. 575-587. Springer International Publishing, 2015.
- [22] Izakian, Hesam, and Ajith Abraham. "Fuzzy C-means and fuzzy swarm for fuzzy clustering problem." *Expert Systems with Applications* 38, no. 3 (2011): 1835-1838.
- [23] Christodoulou, Vyron, Yaxin Bi, and Guoze Zhao. "A Fuzzy Inspired Approach to Seismic Anomaly Detection." In *International Conference on Knowledge Science, Engineering and Management*, pp. 575-587. Springer International Publishing, 2015.
- [24] Senin, Pavel, Jessica Lin, Xing Wang, Tim Oates, Sunil Gandhi, Arnold P. Boedihardjo, Crystal Chen, Susan Frankenstein, and Manfred Lerner. "GrammarViz 2.0: a tool for grammar-based pattern discovery in time series." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 468-472. Springer Berlin Heidelberg, 2014.
- [25] Wang, X., Lin, J., Patel, N., and Braun, M. (2016, October). A Self-Learning and Online Algorithm for Time Series Anomaly Detection, with Application in CPU Manufacturing. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 1823-1832). ACM.
- [26] Riaz, Muhammad, Nasir Abbas, and Ronald JMM Does. "Improving the performance of CUSUM charts." *Quality and Reliability Engineering International* 27, no. 4 (2011): 415-424.
- [27] Zaman, Babar, Nasir Abbas, Muhammed Riaz, and Muhammad Hisyam Lee. "Mixed CUSUM-EWMA chart for monitoring process dispersion." *The International Journal of Advanced Manufacturing Technology* 86, no. 9-12 (2016): 3025-3039.
- [28] Abbas, Nasir, Muhammad Riaz, and Ronald JMM Does. "Mixed exponentially weighted moving averagecumulative sum charts for process monitoring." *Quality and Reliability Engineering International* 29, no. 3 (2013): 345-356.
- [29] Christodoulou, Vyron, and Yaxin Bi. "A combination of CUSUM-EWMA for Anomaly Detection in time series data." In *Data Science and Advanced Analytics (DSAA)*, 2015. 36678 2015. *IEEE International Conference on*, pp. 1-8. IEEE, 2015.
- [30] Leng, Mingwei, Xiaoyun Chen, and Longjie Li. "Variable length methods for detecting anomaly patterns in time series." In *Computational Intelligence and Design*, 2008. *ISCID'08. International Symposium on*, vol. 2, pp. 52-56. IEEE, 2008.



Vyron Christodoulou is a Ph.D. candidate with particular interest in data mining, anomaly detection and fuzzy logic. He received his MSc degree in Artificial Intelligence from Heriot-Watt University, in 2013. Since 2015 he has been working towards his Ph.D. degree in Artificial Intelligence at the School of Computing and Mathematics, Ulster University, Jordanstown, U.K.



Yaxin Bi is a Reader in School of Computing and Mathematics, a core member of the Computer Science Research Institute at Ulster University. His main research interests include multiple supervised and unsupervised machine learning-based classification systems and ensemble methods in conjunction with the Dempster-Shafer (DS) theory of evidence and Rough Set theory. Yaxin has developed a range of big data analytics and decision making with uncertainty methods to

address real world problems, including satellite data exploitation with an emphasis on anomaly/change detection for hazardous analysis and risk resilience; sentiment analysis for opinion mining and cyberbullying detection; and sensor fusion for activity/event recognition in smart environments. Yaxin has been involved in a number of European Frameworks and Space Agency funded projects along with industry cooperation in the past decades and has more than 130 peer-reviewed publications in international journals and conferences. He has served for a number of international conferences as general chair, program co-chair and program committee member. Furthermore Yaxin is an associate editor for *International Journal of Artificial Intelligence Review* as well as a steering committee member for the series of *International Conference on Knowledge Science, Engineering and management (KSEM)* and *Science and Information (SAI) Conference*.



George Wilkie is a fellow of the British Computer Society. He joined the University of Ulster in 1992 after spending 7 years working in various software engineering roles in industry. He is the author of more than 85 research papers in the area of object oriented modelling, complexity metrics and software process improvement, including one book. He was co-chair of the 12th International Workshop on Software Technology and Engineering Practice (STEP) held in Chicago in

September 2004. He held positions as a Visiting Research Scientist at Carnegie Mellon University, Center for Information Systems Engineering, Pittsburgh, USA from October 1999 to October 2001 and Visiting Professor at the Ecole de Technologie Supérieure, University of Quebec in Montreal, Canada between June 2003 and June 2009. He was a charter member of the International Process Research Consortium (www.sei.cmu.edu/iprc), led by the Software Engineering Institute at Carnegie Mellon University and in January 2005 was elected to the position of Technical Architect within this prestigious international research group. After a sustained contribution to the Northern Ireland software industry spanning 20 years, he received the 2006 British Computer Society's Northern Ireland "IT Professional of the Year" award at a ceremony in Belfast Castle. Prior to his time in industry, George read Physics at Queens University Belfast, receiving a 1st Class Honours degree in 1982 and a Ph.D in applied Atomic Physics in 1985.