USING VLF TIME SERIES FROM THE INFREP NETWORK FOR THE STUDY OF PRE-SEISMIC RADIO ANOMALIES

[1]Manilo Monaco, [2]Giovanni Nico, [3]Pier Francesco Biagi, [4]Anita Ermini, [5]Aleksandra Nina, [6]Mario G.C.A. Cimino, [6]Gigliola Vaglini

[1] Department of Information Engineering, University of Florence, Florence, Italy
[2] Institute of Applied Mathematics, National Research Council, Bari, Italy
[3] Department of Physics, University of Bari, Bari, Italy
[4] Department of Industrial Engineering, University of Tor Vergata, Rome, Italy
[5] Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia
[6] Department of Information Engineering, University of Pisa, Pisa, Italy

ABSTRACT

This work presents an application of the Perceptually Important Points (PIP) technique for the analysis of VLF time series. The aim of the analysis is to detect anomalies with respect to the normal variations of the data trends. Such anomalies could reveal possible radio precursors of the earthquake. Since 2009, several radio receivers have been installed throughout Europe in order to realize the INFREP European radio network for studying the VLF (10-50 kHz) and LF (150-300 kHz) radio precursors of earthquakes. The time series used for experiments was collected during the Dodecanese islands earthquakes (MW=5.6 and MW=5.7) occurred on January 30, 2020.

***Index Terms*—** Ionosphere, VLF signal, Earthquake, Time series, Perceptually important points

1. INTRODUCTION

Different factors, such as meteorological conditions, solar bursts and geomagnetic activity, affect VLF/LF radio waves propagation. At the same time, variations of some parameters in the ground, in the atmosphere and in the ionosphere, occur during the preparatory phase of earthquakes. Such variations can produce disturbances in the propagation of the previous signals along their radio paths, which are the radio precursors. Since 2009, several VLF/LF radio receivers have been installed throughout Europe in order to realize a European radio network for studying the VLF/LF radio precursors of earthquakes, called the INFREP network [1-4].

At the end of January 2020 an intense seismic crisis occurred on Dodecanese islands; the main earthquakes (Mw = 5.6 and Mw = 5.7) happened on January 30. This seismic activity occurred in the "sensitive" zone of the INFREP receiver located in Cyprus.

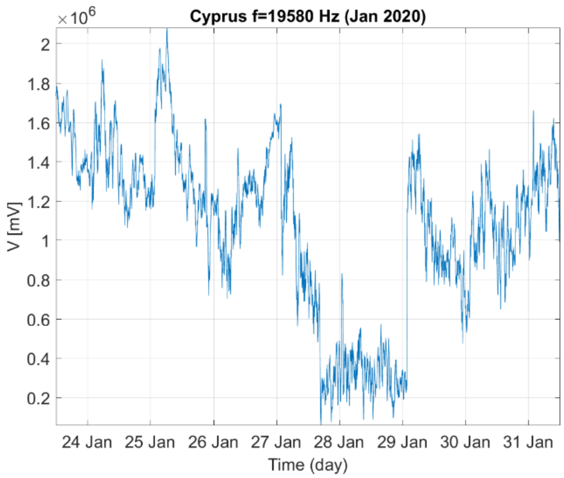
The visual analysis of the raw radio data revealed clear pre-seismic anomalies on the three VLF radio signals (19.58, 20.27, 23.40 kHz) collected by the Cyprus receiver and crossing the zone of the previous seismic activity. The temporal trend of these VLF signals, from the beginning of January 24 till the end of January 31, 2020 is shown in Fig. 1. The analysis of daily day/night trend of these signals points out a clear anomaly during the night of January 29, 2020, one day before the occurrence of the main shocks of the seismic crisis.

In this work an approach based on computational intelligence is experimented for anomaly detection. In the first phase of the approach, a time series segmentation is carried out, to remove the noise and divide the series into discrete segments revealing the major dynamics. As a segmentation method, the Perceptually Important Points (PIP) technique is applied. The PIP is able to preserve the shape of the data [5][6]. In the second phase, the time series of PIPs is statistically compressed using the median for each time window. Finally, the set of median points is separated into low and high dynamics using the K-means clustering algorithm. As a consequence, anomalies are automatically located outside the generated clusters. The effectiveness of the approach has been experimented.

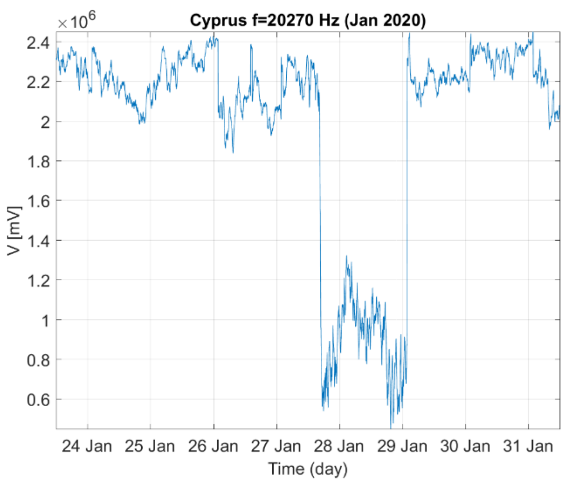
2. INFREP NETWORK

The International Network for Frontier Research on Earth-quake Precursors (INFREP) network currently consists of nine receivers located as follows: one in Austria, Cyprus, Italy, Portugal and Serbia, two in Romania and Greece.

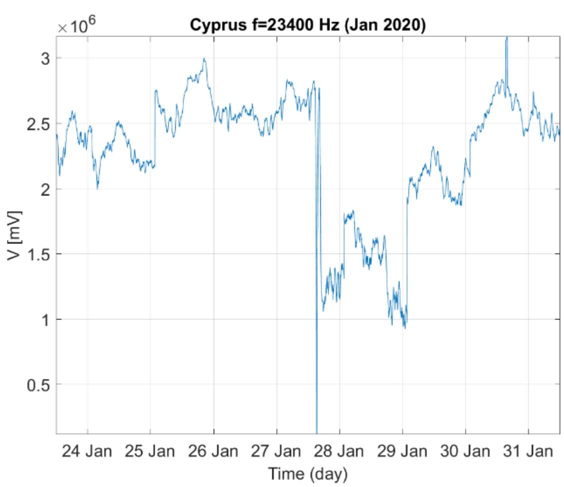
The radio receivers measure the intensity (electric field strength) of radio signals radiated by existing VLF-LF broadcasting stations in the bands VLF (20 - 80 kHz) and LF (150 - 300 kHz), with one-minute sampling rate. The location of the transmitters and receivers are available on the INFREP web site (www.infrep-network.eu).



(a)



(b)



(c)

Fig. 1 - VLF signals collected by the Cyprus receiver from 24 to 30 January 2020: (a) 19.58 kHz, (b) 20.27 kHz and (c) 23.40 kHz. An anomalous decrease appears on 29 January.

The data collected are transmitted every day to the server. The different temporal trends of signals collected at the receivers are also available on the INFREP web site.

3. METHODOLOGY

The problem on how to represent a time series in multi-resolution is fundamental, because time series data is hard to manage in its original structure. Moreover, before processing and analyzing time series, computational intelligence applications is needed to reduce the high dimensionality of the data while retaining the information associated with important points. For example, financial analysis often depends on the shape of data and the significance of points to identify technical patterns. Most time series dimensionality reduction approaches, such as principal component analysis (PCA), discrete Fourier transform (DFT), discrete wavelet transform (DWT), piecewise aggregate approximation (PAA), often lose important data points and fail to retain the general shape of the time series after compression [7]. A method to dynamically segment time series into subsequences of unequal length, preserving the shape of the data, is based on choosing points that are important in the human vision identification process. These data points are known as Perceptually Important Points (PIPs). PIPs were initially introduced to exploit salient points from a price series and then have also been used to identify specific technical patterns [5]. In the context of data mining, PIPs have been used mainly for purposes of dimension reduction (time series representation) and for clustering reasons (see [8] and references therein).

Fig. 2 shows the workflow for detecting anomalies in VLF time series. Specifically, the input is made by the VLF time series $v_i(t)$, i=1,2,3, collected at three different frequencies. A windowing is then carried out, selecting the values that are enclosed in the current sliding interval, characterized by size τ and shift $\partial\tau$. Let $\bar{v}_{i,n}(t)$ be the resulting windowed signal extracted with the n-th window. A segmentation is applied, in which the input time-series is divided into a sequence of σ-1 linear segments, generating the time series $\tilde{v}_{i,n}(t)$. As an example, Fig. 3 shows 10 points located with a PIPs identification algorithm on a time series initially made up of 241 samples. Starting with the time series $\bar{v}_{i,n}(t)$, the algorithm allows to identify σ data points: $\tilde{v}_{i,n,1},\ldots,\tilde{v}_{i,n,\sigma}$, via a reordering process of points in the input data series by their importance. The first two PIPs found are the first and last points of $\bar{v}_{i,n}(t)$. In order to identify the next PIPs, the most classical algorithms calculate the distance between all remaining points and two adjacent PIPs, and label as the next PIP the one with the maximum distance. Three metrics are generally used as a distance, namely Euclidean distance (ED), perpendicular distance (PD) and vertical distance (VD) [8].

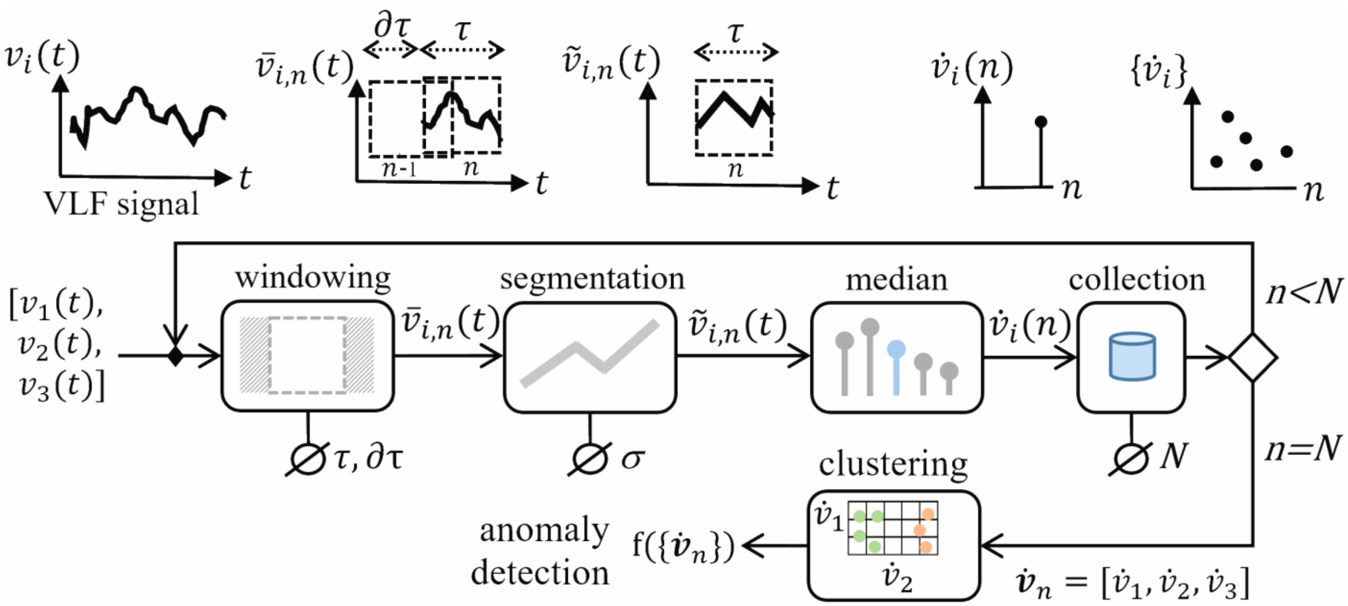


Fig. 2 –Workflow for detecting anomalies in VLF time series

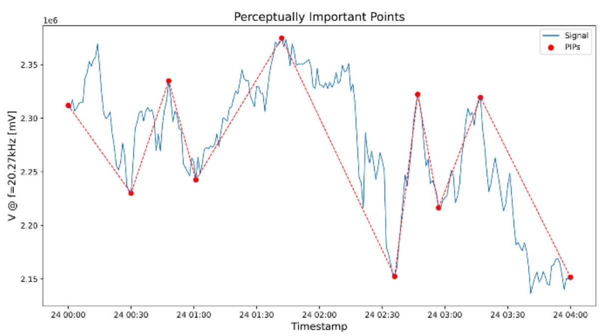


Fig. 3 - PIPs identified on a time series of 241 samples

In this paper, a particular approach has been used to locate the next PIPs. First, the slope of the line joining two PIPs is considered, to identify the main trend of the subsequence. Then, the next one or two (possibly) PIPs are determined, considering the maximum percentage retracement of the main trend. A retracement is defined as a temporary reversal of the trend of a time series. The PIPs identification process continues until σ points in $\bar{v}_{i,n}(t)$ are attached to a list. If σ+1 points are identified, then the first PIP is discarded as the least significant, since it is the oldest. Subsequently, the median of the σ PIPs $\tilde{v}_{i,n,1},\ldots, \tilde{v}_{i,n,\sigma}$ is computed. In the case in point, a vector of three medians is delivered for each n-th window, i.e., $\dot{\boldsymbol{v}}_n \equiv [\dot{v}_1(n), \dot{v}_2(n), \dot{v}_3(n)]$.
Median vectors are continuously collected and stored. When N vectors are available, the anomaly detection process can start. In general, an anomaly detection approach is comprehensive if it is able to both detect and identify the emerging anomalies and characterize the essence of these anomalies by visualizing the structures discovered within data in a way that is understandable to the end-user [9]. In this paper, the *k*-means clustering algorithm is used as a basis for anomaly detection and characterization. Indeed, clustering is an effective way to visualize and understand the variation of dynamics within data. Specifically, *K*-means is a fast, iterative algorithm that has been used in many clustering applications. Formally, it is a point-based clustering method that starts with the cluster centers (centroids) arranged at arbitrary positions and proceeds by moving at each step the centroids in order to minimize the clustering error. The method has the disadvantage of being sensitive to the initial positions of the centroids. Therefore, in order to obtain near optimal solutions using the *k*-means algorithm, several runs must be scheduled changing the initial positions of the cluster centers [10]. Actually, in our approach the clustering is iteratively applied on overlapping sliding windows of N points. As a consequence, after the first clustering, the subsequent ones can be initialized with the previous centroids.
In particular, the clustering is configured to detect 2 cluster: normal and anomalous vector. As a result, two types of analyses can be carried out. In the *offline analysis*, the colored clusters are represented in 3D space as animated frames (one frame per clustering). With this kind of analysis, the labeling generated by clusters exhibit the differences over time between normal and anomalous behaviors [11]. In the *online analysis*, an anomaly score is calculated, in order to detect anomalous events, i.e, a drift in the cluster composition. In our approach, the silhouette index [12] has been used as an aggregated value of anomaly. It measures how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette index ranges from −1 to

+1: high values indicate that the point matches to its own cluster and not to the others. If most points have a high value, the clustering configuration is well defined. In contrast, overlapped clusters exhibit low or negative values. In our approach, as anomaly index the average silhouette index of all points is used to assess an unfolding anomaly.

An important aspect related to the horizontal scalability of this approach, is that different sources of information can be derived from VLF signals, originating a multimodal decision-making process. In order to keep the 3D visualization effective, the context-based clustering can be considered. The context is a sub-aspect of the domain, related to different factors or sub-regions of the monitored sources. A suitable architecture for this purpose is based on receptive fields [13].

4. RESULTS

In order to show the potential of the approach of Fig.2, in this section a pilot case is discussed. Specifically, the window size τ is set to 90 samples, the shift $\partial\tau$ is set to 1, and the segmentation σ is set to 30 PIPs. Fig. 4 shows the result of the cluster analysis. Here, the coordinates of each point are given by the signal amplitude at the three frequencies 19.58 kHz, 20.27 kHz and 23.40 kHz (see Fig. 1). In figure different colors are used to denote points referring to the eight days from January 24 to January 31. Two well separated clusters are clearly visible. The anomalous cluster is the group of points at bottom right, characterized by the green and red points grouped in the days 28 and 29 January.

The purpose of this paper is to introduce a novel analysis perspective of pre-seismic radio anomalies based on VLF time series. A small case study has been briefly discussed to show the potential of the approach. Further research is necessary to achieve significant results. To this aim, future work will focus on further experimentation and investigation, as well as on further integration with other methods.

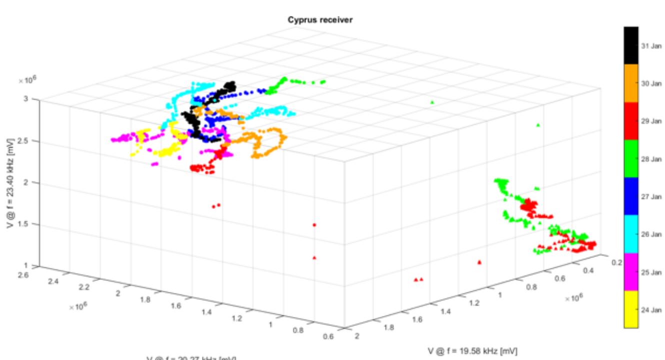


Fig. 4 - Cluster analysis applied to the VLF time series using *k*-means algorithm and two clusters

REFERENCES

[1] P.F. Biagi, T. Maggipinto, F. Righetti, D. Loiacono, L. Schiavulli, T. Ligonzo, A. Ermini, I.A. Moldovan, A.S. Moldovan, A. Buyuksarac, M. Buyuksarac, H.G. Silva, M. Bezzeghoud and M.E. Contadakis, "The European VLF/LF radio network to search for earthquake precursors: setting up and natural/man-made disturbances," *Natural Hazards and Earth System Sciences*, Vol. 11(2), pp. 333-341, February 2011.

[2] P. Dolea, O. Cristea, P. V. Dascal, I. A. Moldovan and P.F. Biagi, "Aspects regarding the use of the INFREP network for identifying possible seismic precursors," *Physics and Chemistry of the Earth*, Vol. 85-86(1), pp. 34-43, 2015.

[3] A. Nina, S. Pulinets, P.F. Biagi, G, Nico, S,T. Mitrović, M. Radovanović, L.Č.Popović, "Variation in natural short-period ionospheric noise, and acoustic and gravity waves revealed by the amplitude analysis of a VLF radio signal on the occasion of the Kraljevo earthquake (Mw = 5.4)", *Science of The Total Environment*, Vol. 710, Paper N. 136406, March 2020.

[4] M.E. Contadakis, D.N. Arabelos, G.N. Vergos, C. Skeberis, T.D. Xenos, P.F. Biagi, E.M. Scordilis, "Ionospheric turbulence from TEC variations and VLF/LF transmitter signal observations before and during the destructive seismic activity of August and October 2016 in Central Italy", *Annals of Geophysics*, Vol. 63(5), PA546, 2020.

[5] T. Fu, F. Chung, R. Luk, C. Ng, "Stock Time Series Pattern Matching: Template-based vs. Rule-based Approaches", *Engineering Applications of Artificial Intelligence*, Elsevier, Vol. 20(3), pp. 347-364, April 2007.

[6] M.G.C.A. Cimino, F. Dalla Bona, P. Foglia, M. Monaco, C.A. Prete, G. Vaglini, "Stock price forecasting over adaptive timescale using supervised learning and receptive fields." International Conference on Mining Intelligence and Knowledge Exploration. Springer, Cham, 2018.

[7] T. Fu, F. Chung, R. Luk, C. Ng, "Representing Financial Time Series Based on Data Point Importance", *Engineering Applications of Artificial Intelligence*, Elsevier, Vol. 21(2), pp. 277-300, March 2008.

[8] P. E. Tsinaslanidis, D. Kugiumtzis, "A Prediction Scheme Using Perceptually Important Points and Dynamic Time Warping", *Expert Systems with Applications*, Elsevier, Vol. 41(15), pp. 6848-6860, 1 November 2014.

[9] H. Izakian, W. Pedrycz, "Anomaly Detection and Characterization in Spatial Time Series Data: a Cluster-Centric Approach", *IEEE Transactions in Fuzzy Systems*, IEEE, Vol. 22(6), pp. 1612-1624, 24 January 2014.

[10] A. Likas, N. Vlassis, J. J. Verbeek, "The Global K-Means Clustering Algorithm", *Pattern Recognition*, Elsevier, Vol. 36(2), pp. 451-461, February 2003.

[11] E. Keogh, J. Lin, A. Fu, "HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence", *In Proceedings of the Fifth IEEE International Conference on Data Mining*, IEEE, Houston, TX, USA, pp. 226-233, 27-30 November 2005.

[12] A. Dudek, "Silhouette Index as Clustering Evaluation Tool". In *Conference of the Section on Classification and Data Analysis of the Polish Statistical Association* pp. 19-33. 2019, Springer, Cham.

[13] M.G. Cimino, W. Pedrycz, B. Lazzerini, and F. Marcelloni, "Using multilayer perceptrons as receptive fields in the design of neural networks". *Neurocomputing*, 72(10-12), pp.2536-2548, 2009.