

A machine learning approach for automated ULF wave recognition

Georgios Balasis^{1,*}, Sigiava Aminalragia-Giamini^{1,2}, Constantinos Papadimitriou^{1,2},
Ioannis A. Daglis^{1,3}, Anastasios Anastasiadis¹, and Roger Haagsmans⁴

¹ Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Metaxa and Vas. Pavlou St., Penteli, 15236 Athens, Greece

² Space Applications & Research Consultancy, SPARC, 8 Kleissovias str., 106 77 Athens, Greece

³ Department of Physics, National and Kapodistrian University of Athens, Panepistimiopolis, Zografos, 15784 Athens, Greece

⁴ European Space Research & Technology Centre, ESTEC, Keplerlaan 1, Postbus 299, 2200 AG Noordwijk, The Netherlands

Received 31 August 2018 / Accepted 17 March 2019

Abstract—Machine learning techniques have been successfully introduced in the fields of Space Physics and Space Weather, yielding highly promising results in modeling and predicting many disparate aspects of the geospace environment. Magnetospheric ultra-low frequency (ULF) waves can have a strong impact on the dynamics of charged particles in the radiation belts, which can affect satellite operation. Here, we employ a method based on Fuzzy Artificial Neural Networks in order to detect ULF waves in the time series of the magnetic field measurements on board the low-Earth orbit CHAMP satellite. The outputs of the method are validated against a previously established, wavelet-based, spectral analysis tool, that was designed to perform the same task, and show encouragingly high scores in the detection and correct classification of these signals.

Keywords: machine learning / ULF waves / LEO satellites

1 Introduction

Artificial Neural Networks are information processing systems based on the principle of simple processing units being interconnected in a network and working in parallel as a whole. The original base for this kind of systems was a modeling of the way biological neural networks function, such as those found in nature in the central nervous systems of animals. They stemmed from the field of Artificial Intelligence and are an important category of systems both for real-life applications but also for theoretical studies and the furthering of our understanding of the functions of the animal and human brain. Many types of neural networks have been employed in various and seemingly disparate fields and applications including, but not limited to, medical applications for the assessment of disease possibility, image processing, the real time monitoring of systems especially in hazardous environments and the prediction of the behavior of complex systems like financial, biological or weather systems. They are very useful when a decision making process needs to consider many variables,

interdependent phenomena which affect an overall outcome, and processes that are highly complex, hidden or unknown (Anderson, 1995).

In space physics or space weather in particular, artificial neural networks (ANN) have been used for the forecasting purposes of geomagnetic activity indices (e.g., Dst, Kp, AE etc.) as well as ground effects of the geomagnetic activity (e.g., geomagnetically induced currents – GICs). For instance, the Dst index has been widely used as a proxy of the ring current strength and consequently of the intensity of geomagnetic storms, the most complex phenomenon of magnetospheric dynamics (Daglis, 2006). Several researchers have developed Dst index forecasts using measurements of interplanetary magnetic field (IMF) and/or solar wind plasma parameters. Therefore, a number of empirical models exists, based on differential equations (e.g., Burton et al., 1975; O'Brien & McPherron, 2000; Temerin & Li, 2002), and on ANN (e.g., Lundstedt et al., 2002; Pallochia et al., 2006; Wei et al., 2007). Temerin & Li's (2002, 2006) scheme makes successful Dst predictions with a time step of 10 min whereas the Dst index has time step of one hour by using real-time data from ACE as the input to their model (http://lasp.colorado.edu/space_weather/dsttemerin/dsttemerin.html). The forecasting model by Lundstedt et al. (2002) consists of a

*Corresponding author: gbalasis@noa.gr

recurrent neural network that has been optimized to be as small as possible without degrading the accuracy. It is driven solely by hourly averages of the solar wind magnetic field component B_z , particle density, and velocity, which means that the model does not rely on observed Dst. In an evaluation based on more than 40,000 hours of solar wind and Dst data, it was shown that this model had smaller errors than other models in operational use. Similar attempts have also been made for the prediction of the K_p , based on both past values of the index itself and additional solar wind parameters, that not only exhibit great prediction scores, but also help to elucidate the relation between the magnetospheric response and the characteristics of the incoming solar wind (Wing et al., 2005; Wintoft et al., 2017). The system identification approach, including neural network methodologies, has also been applied to modeling the radiation belts (for a recent review see Boynton et al., 2016). In a recently published book (Camporeale et al., 2018) on the use of machine learning techniques for space weather, there has been a series of chapters devoted to the application of ANN for: determining magnetospheric conditions (Bortnik et al., 2018), reconstruction of plasma electron density from satellite measurements (Zhelavskaya et al., 2018), and classification of magnetospheric particle distributions (Souza et al., 2018).

Magnetospheric ultra-low frequency (ULF) waves can be excited by the solar wind with a scale size of the whole magnetosphere and these waves can have a strong impact on charged particle dynamics in the radiation belts (Mann, 2016). For instance, radial diffusion caused by ULF waves may accelerate electrons with MeV energies in the radiation belts. These electrons may penetrate spacecraft shielding, generate a build-up of static charge within electrical components resulting in subsystem damage, and ultimately can even cause the total loss of Earth-orbiting satellites (Mann, 2016). Therefore, studies of the excitation, propagation and global characteristics of ULF waves, as well as their impact on MeV electron dynamics remain an active area of research (Mann, 2016). There has been also a number of studies addressing the monitoring of ULF waves from low-Earth orbit (LEO) satellites (for recent reviews see Balasis et al., 2015; Papadimitriou et al., 2018) as well as from the ground (Bogoutdinov et al., 2018). Here, we employ a method based on ANN in order to detect ULF wave events in the topside ionosphere with CHAMP satellite. The outputs of the method are validated against a previously established, wavelet-based, spectral analysis tool (Balasis et al., 2013), that was designed to perform the same task, and show encouragingly high scores in the detection and correct classification of these signals.

The present paper is organized as follows: Section 2 introduces the basic theory of the specific type of ANN used in this study, while Section 3 describes the dataset and the classification methodology. In Section 4, we present and discuss our results. Finally, Section 5 summarizes our findings.

2 Adaptive neuro-fuzzy inference system

The building block of a neural network is the neuron, the analogue of the biological neural cells. They are simple units which consist of a linear input part which accepts multiple weighted scalar inputs and sums them much like the synapses

and dendrites function in reality. This sum is then forwarded to the non-linear part which is the characteristic activation function of the neuron. The signal produced from the input sum having activated the function is then forwarded to multiple outputs (Bishop, 1995).

Having the building block of a network the next step is the way the units are organized and interconnected. This, in artificial neural networks is done in general by ordering neurons in layers where each layer feeds-forward its information to the next one. Here the similarities with nature begin to thin as there is no way yet to viably reproduce what occurs in reality. Figure 1 shows the general structure of an artificial neural network compared to a photograph of actual, biological neurons, to showcase how they have been the inspiration behind the design of such artificial structures (<http://www.alanturing.net/>).

The typical structure of neural networks has three distinct types of layers (Bishop, 1995):

1. The input layer part which receives data and forwards it to the next hidden layer.
2. The hidden layer(s) which applies non-linear activation functions on the weighted sums of the previous layer.
3. The output layer that sums the weighted output of the previous hidden layer and gives the final output.

The neurons of these layers are interconnected in such a way that each neuron of a layer transfers its information to every neuron of the next one with a different “synaptic” weight. This architecture and the fact that neural networks alter their structure (the interconnection weights) during the training process is what allows them to be powerful tools that exhibit complex global behavior even though they consist of quite simple building blocks.

At the initialization phase of most networks, each synapse connecting two neurons can have a weight between zero and one, zero being a “dead” synapse meaning that these two neurons effectively do not communicate and one being the strongest connection. During training the weights are altered and reconfigured to create the final structure of the net, so even if seemingly all neurons communicate with those of the next layer the true picture is one of many different connections carrying information not in an identical way but adjusted so that the network can create a mapping of the sample area with which it was trained and will subsequently function in Reed & Marks (1999).

Adaptive Neuro-Fuzzy Inference System (ANFIS) (Shing & Jang, 1993) is a special type of system which incorporates elements from models based on fuzzy rules and models based on neural networks. Systems of this type are considered universal approximators. The ability to approximate non-linear functions has its core in the training of a set of fuzzy conditional statements in the form of “If-Then” rules through which the data input is processed. The structure of a typical ANFIS network with two inputs and one output is shown in Figure 2 (<http://www.alanturing.net/>).

The application of these “If-Then” rules is how these systems differ compared to more traditional neural networks. As an example, a very simple version of such a system would consist of a network with an additional input layer, that tests each input parameter and decides whether it satisfies one rule or another, e.g., if the given input x_i is numerically closer to a certain value c_1 or to another value c_2 , by checking their

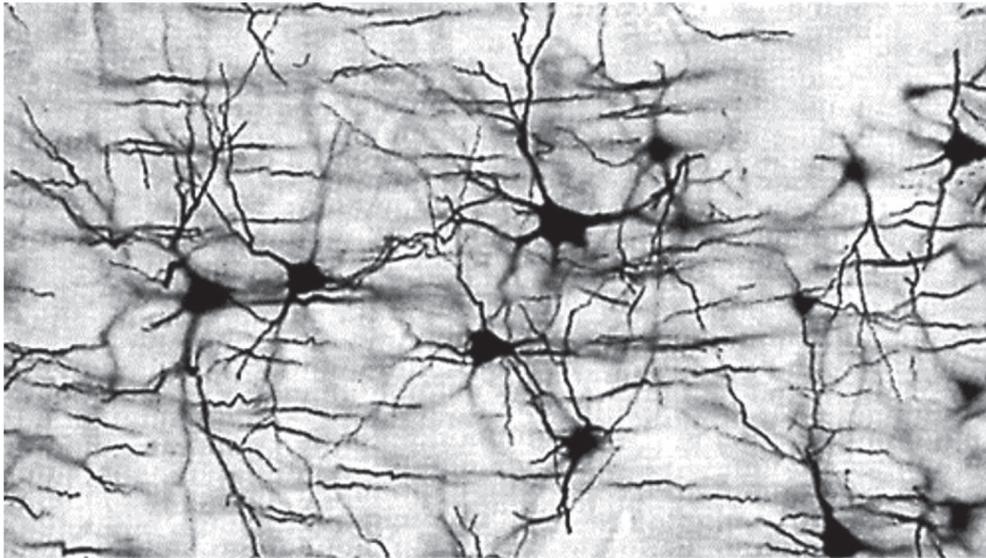
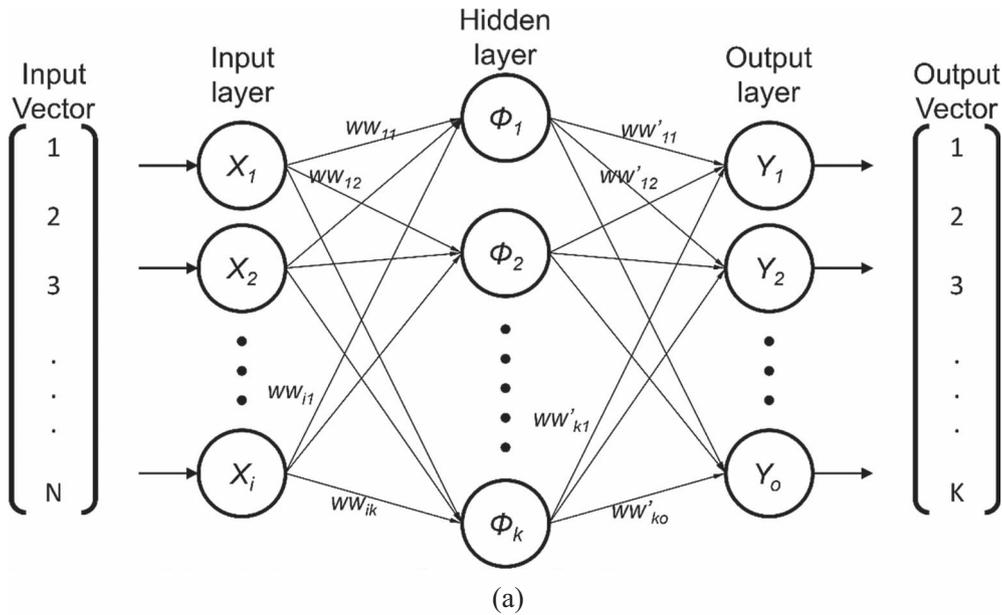


Fig. 1. (a) Structure of an artificial neural network. (b) Biological neurons with their interconnections (from <http://www.alanturing.net/>).

absolute difference. Depending on this result, the neuron would then apply a specific weighting factor on that input, e.g., 0.1 if $|x_i - c_1| < |x_i - c_2|$ and 0.9 otherwise, and then propagate the result to the next layer. That would be the case of an “exclusive if” statement, but ANFIS take this line of reasoning further by additionally utilizing the concept of “fuzziness”. In this sense, the “If–Then” rules are not satisfied in a Boolean manner (true or false) but rather in a quantitative manner. Following the previous example, each neuron in the “fuzzy–If” layer can now test not only if an input parameter lies close to a certain value, but measure this distance and depending on it, generate two weighting factors for each input. All these weighting factors are then summed and normalized in order to determine the final weight according to which each of the rules will be applied to the inputs.

In an actual implementation, an ANFI system is comprised by five basic layers and typically uses only two rules. Layer 1 performs the “fuzzy–If” condition that was described above, usually using a Gaussian as a membership function, with a different center and spread (mean and standard deviation of the Gaussian function) for each rule, and the result is the output of the Gaussian function based on the distance of each input value from the corresponding center, yielding two output weights for each input. Layer 2 multiplies the generated weights that correspond to the same rule, producing the weights w_1 and w_2 which represent the firing strength of each rule, while Layer 3 normalizes these weights (by dividing with their sum). Having now established the strength of each rule, Layer 4 applies functions f_1 and f_2 , weighted by their corresponding w_s , to the initial, unadulterated versions of the inputs x_i and the results

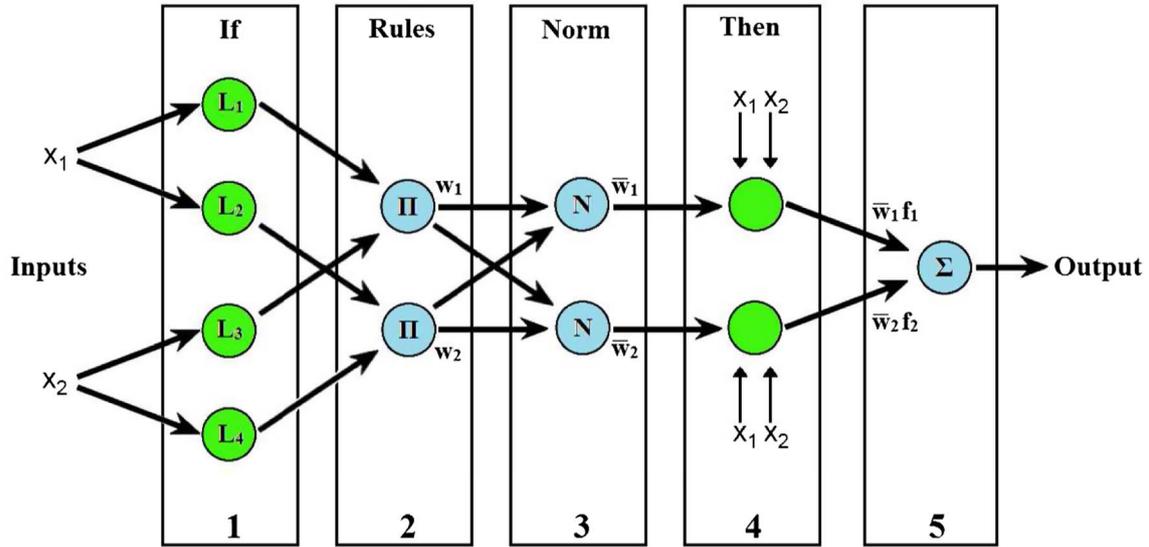


Fig. 2. The basic structure of an ANFIS system (from <http://www.alanturing.net/>).

are summed at the final Layer 5 to provide the final output as a single number.

When considering what is called a first-order Sugeno type system (Sugeno, 1985) the functions f_1 and f_2 are:

$$f_1 = p_1x_1 + q_1x_2 + r_1,$$

$$f_2 = p_2x_1 + q_2x_2 + r_2,$$

and the application of these functions according to the If–Then rules can be summarized as:

- Rule 1: If x_1 is L_1 and x_2 is L_3 Then f_1 .
- Rule 2: If x_2 is L_2 and x_2 is L_4 Then f_2 .

The set of premises $\{L_i\}$, i.e., the parameters of the Gaussians that determine the extent of the membership of each $\{x_i\}$ to the corresponding rule and the consequent parameters $\{p_i, q_i, r_i\}$ with which the system builds the If–Then rules is the “knowledge base” of the system which is formed during training. Training is performed with a set of patterns from which certain “process parameters”, selected by the user, have been extracted. The exact nature and type of these process parameters depends strongly on the data and the application for which the system is being purposed for. In some cases, they can be simple and straightforward, e.g., the value of a certain observable at a specific moment in time, while in others they can be some derived statistical property based on the data of the pattern or even more complicated outputs of mathematical operations that are applied on the pattern. These input vectors of process parameters are fed into the system along with the desirable output to train the ANFIS network, usually with a hybrid, least-squares and back-propagation gradient descent, learning algorithm. In the case of classification, the networks are trained to give two (or more) specific numeric outputs whenever an input belonging to a specific class is given. For example, in the case of two categories, an arbitrary threshold can be set that if the output is negative the pattern is classified into class A and if

positive into class B. One important advantage of ANFIS networks compared to other neural networks is that because of their architecture they generally require a smaller number of patterns to be trained.

The ANFIS neural networks were implemented using the framework of the MATLAB programming environment, in which the main parameter that determines the ANFIS network’s behavior is the number of training epochs, which typically ranged from 20 to 100 for the training sessions that we performed.

3 Data and methodology

The dataset used is a database of magnetic field measurements, obtained by the CHAMP (CHALLENGING Minisatellite Payload) satellite. CHAMP was a highly successful German LEO mission, which launched in July 2000 into an almost circular, near polar ($i = 87^\circ$) orbit with an initial altitude of 454 km (Reigber et al., 2005). CHAMP flew for 10 years, until September 2010, but for the first few years of the mission (up to 2002) the satellite was performing orbit corrections maneuvers, while at the last years (i.e., after 2007) the satellite was flying in quite lower altitudes than at the beginning of the mission, due to the effect of ionospheric drag. For these reasons, we decided to limit the time period of the data that were retrieved to five years, namely from the beginning of 2003 up until the end of 2007. We have used the magnetic field data as measured by the Fluxgate Magnetometer (FGM) instrument on board CHAMP. The CHAMP FGM data (1 s sampling rate) of the total magnetic field time-series are segmented into mid-latitudinal tracks, i.e., the parts of the orbit for which the satellite lies in the region from -55° to $+55^\circ$ in magnetic latitude. This is done in order to exclude the influence of polar field aligned currents that might affect the measurements (Ritter et al., 2013). Moreover, the magnetic field time series is filtered using a high-pass Butterworth filter with a cutoff frequency of 16 mHz (Williams & Taylors, 1988), to focus only on the contributions of external

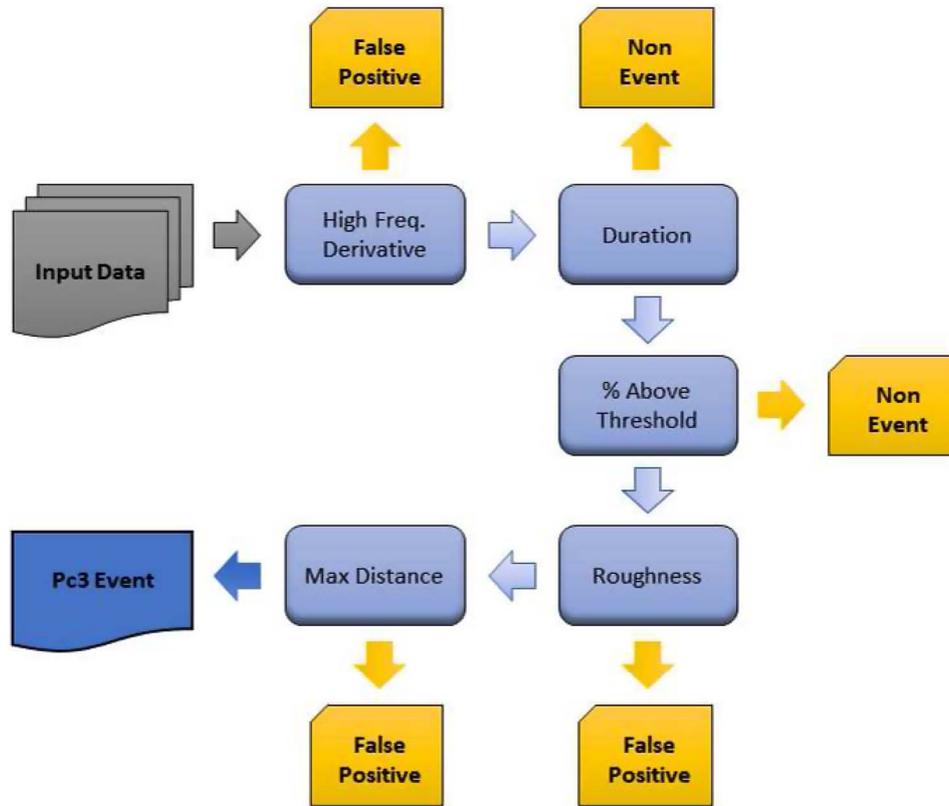


Fig. 3. Schematic representation of the tests to which each candidate event is subjected and the corresponding category to which it is assigned if it fails them.

ULF waves and ignore the slower varying changes in the field due to the satellite flying through areas of different magnetic field strength.

The aim of this study was to detect ULF wave events in the Pc3 frequency range (from 20 to 100 mHz approximately), and in order for any machine learning method to be applied, there should be a training and a test dataset. To construct these datasets we initially applied a wavelet-based analysis method for the detection of Pc3 ULF wave events (Balasis et al., 2013, 2015) on the CHAMP-FGM tracks, employing the Morlet mother function for the application of the wavelet transform. The method detects ULF wave events by examining the wavelet spectra of the total magnetic field time-series in each track and identifying time periods of consecutive significant activity. Naturally, not all signals that exhibited prominent wave power in the Pc3 range were true ULF pulsations, as the measurements were sometimes contaminated by short lived anomalies, such as spikes or abrupt discontinuities due to instrument errors, which created broad band responses on the wavelet spectra and thus exhibited large wave power within the frequency range of our interest. In addition to that, the satellites' low altitude forced it to pass through the topside ionosphere, a layer which is often-times plagued by turbulent plasma, e.g., from rising flux tubes in near-equatorial, night-side areas, a phenomenon that is known as Equatorial Spread F (ESF) irregularities and are most commonly referred to as "plasma bubbles" (Stolle et al., 2006; Park et al., 2013). Such instabilities also left their mark on the magnetic field time series, a mark that happens to also have an imprint in the Pc3 frequency range.

In order to rid our studies from all those non-ULF signals, we devised a method to classify tracks in three main categories, namely actual Pc3 Wave Events, Non-Events i.e., background noise without significant wave activity and False Positives (FP), e.g., anomalous signals due to spikes, discontinuities, etc. Figure 3 shows a schematic of the wavelet method and Figure 4 shows tracks with examples of an Event, Non-Event and False Positive. In principle, the wavelet method works by testing each candidate signal against a series of criteria and either assigning it to one of the predetermined categories, if it fails to meet one of them, or classifying it as a real Pc3 Wave Event, if it successfully passes all of them. Such criteria were: the duration of the signal that exhibits total wavelet power in the Pc3 range above a certain threshold, the percentage of points that lie above a given value, the "roughness" of the magnetic field series (measured by calculating the average of the second time derivative at points of local extrema in the series), the maximum difference of the series from its mean, or the maximum value of the first time derivative of its wavelet power at the highest frequency (100 mHz). All these values had to lie within certain thresholds for the test to be deemed as successfully passed, thresholds which were estimated heuristically by testing the results of this method on tracks that were selected and classified by human operators. In the original method there is also a fourth category named Plasma Instabilities (PIs), which consists of events that coincide with ESF-related phenomena, but detection of these signatures requires additional information in the form of electron density time series from the planar Langmuir probe (PLP) instrument. Since the present study focuses only

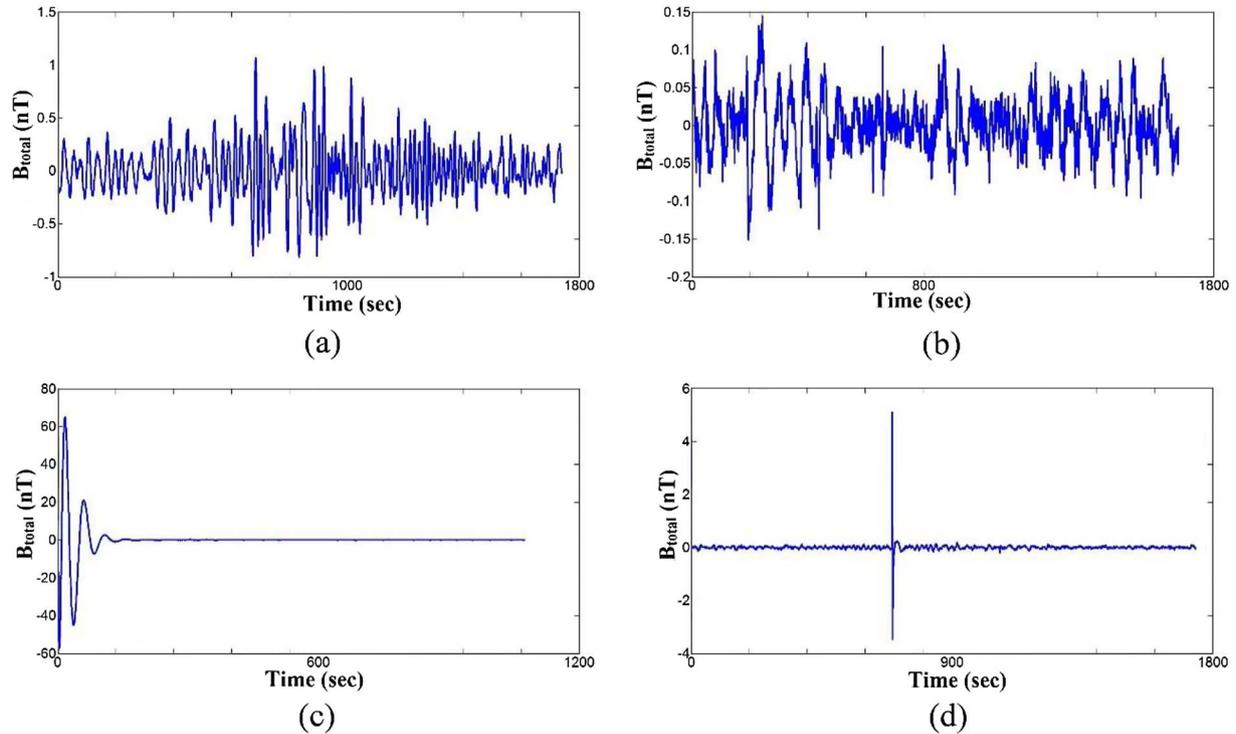


Fig. 4. (a) Event, (b) Non-Event, (c) Type I False Positive, (d) Type II (spike) False-Positive.

on the analysis of the magnetic field series, this category could not be included, although the data that were plagued by PIs were still part of the dataset.

By applying this method to CHAMP data, we produced a database of classified mid-latitude tracks of the time-series of the total magnetic field. The tracks were further separated to Dayside and Nightside subsets, due to the fact that PIs are predominantly present in the Nightside tracks. This can create errant results in the classifying scheme for Nightside tracks and is something that has to be taken into account when evaluating classifying results. As we show further below, this is the most likely explanation for the differences between Dayside and Nightside classification results.

In order to utilize this dataset for the training and testing of the ANFIS classifiers we extract from each track time series a number of statistical measures, which represent the signal in an N -dimensional space. These N -dimensional feature vectors, produced from each track, are used as the input to the ANN system. A total of 13 different parameters are calculated and used in two different subsets. These can be seen in Table 1. Most of these parameters are simple statistical quantities of the time series itself or of properties that are themselves derived from the times series. The only exception is the “fractal dimension” which corresponds to the dimensionality of the series according to the box counting method (Falconer, 1990).

We have structured the ANN system in the arrangement of a cascade binary classifier as seen in Figure 5. Two different neural networks are used to classify the tracks in the three categories. The first tier network separates the FP tracks from the Events and Non-Events and in essence it functions as a “cleaner” of the dataset removing all anomalous signals before the categorization. If the track is not classified as a FP it is

forwarded to the next tier and the second network separates Events from Non-Events.

The reason we have used this structure is two-fold. First, it allows a specialized approach at each tier relative to its task. It is obviously more challenging for a network to distinguish between three or even more categories of signals than only two. A simultaneous classification of all three categories would require a parametric representation in an M -dimensional feature space with which the ANN would be able to distinguish all the categories in one step. Conversely with the scheme applied here the task becomes much simpler as each tier has to answer a simple “yes-or-no” question allowing for little room in ambiguity in the outputs of the ANN. Furthermore our scheme allows for the different parametric description of the signal at each tier which is better suited for the distinction that is to be made resulting in task-specific networks able to better capture the discriminative qualities yielding more accurate results. Indeed, in the first tier parameters 1–9 are used for the distinction of FPs while in the second tier all thirteen are used for the distinction of Events. Second, this structure is modular and it can be easily expanded with further classifications such as the distinction of different categories of False Positives, e.g., types of artificial errors, or even the detection of the previously discussed Plasma Instabilities as a subcategory of Events. This modularity is a significant advantage in any such system as it does not require fundamental changes, like the retraining of networks to successfully incorporate a new category, as would be the case in a system with a singular classifier. The only significant disadvantage of the structure is the possible “bleed-through” of a misclassified signal to a next tier. If for example an FP signal is not classified correctly in the first tier then it will be forwarded to the next one to be classified as an Event or a Non-Event.

Table 1. The parameters used with the ANN in the classifying system.

Name	Description
1. Energy	Sum of squared values of signal.
2. Frequency roughness	Mean of the absolute value of the second derivative of the signals' % Fourier Spectrum.
3. Variance	Variance of the track.
4. Max-minus-median	Maximum value minus median value.
5. Fraction over half-max	The fraction of the signal's values that are greater from 40% of its maximum value.
6. RMS	Root mean square value of the track.
7. Fractal dimension	The fractal dimension of the signal with values 1–2.
8. Stdev of variances	The standard deviation of the variance values of the signal using a sliding window of 100 measurements.
9. Maximum	Maximum value of track.
10. Zero-cross area variance	The variance of the integral from one zero-crossing of the signal to another.
11. Zero-cross norm area variance	The variance of the integral from one zero-crossing of the signal to another normalized by it duration.
12. Zero-cross area difference mean	The mean difference of consecutive integrals from one zero-crossing of the signal to another.
13. Zero-cross area difference variance	The variance of the differences of consecutive integrals from one zero-crossing of the signal to another.

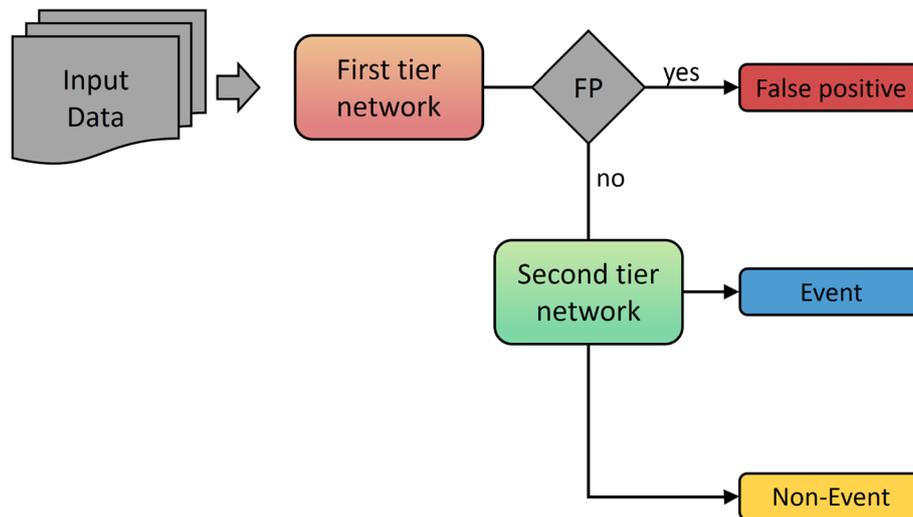


Fig. 5. The classifier system using ANN which is structured in a two-tier arrangement.

4 Results and discussion

To test the ANFIS we constructed an extensive database of tracks, spanning the five year time period from 2003 up to 2007, which were classified by means of the wavelet method. Table 2 summarizes this classification process by showing the numbers of the detected signals that were attributed to each category. For the training of the networks we visually examined thousands of those identified signals, from both geomagnetically active and quite times and carefully handpicked 200 signals from each category (for both Dayside and Nightside tracks) that we considered as the most representative. In this manner we created an initial pool of signals, which we believe reflects, as accurately as possible, the variety with which Pc3 wave events and False Positives manifest in the data. Obviously, no such collection can ever be perfect, but we spent considerable effort to produce an ensemble that is in agreement with our accumulated experience. From this pool we then composed the training datasets.

Table 2. Tracks of each category from Dayside and Nightside.

Tracks	Dayside	Nightside
False Positives	672	812
Events	4026	634
Non-Events	23,406	21,265

The first tier network, which is used to root out the False Positive signals, was trained with a set consisting of 200 FPs and a joint class of 150 Events and 150 Non-Events, all manually selected from the predefined handpicked signals. The second tier network, which performs the actual Pc3 Event detection, was trained with all the 200 handpicked Events and 200 Non-Events. Both ANFIS implement the first-order Sugeno type system that was described above, with two rules and linear activation functions, and the only significant difference is in the input parameters that they use (parameters 1–9 for the first tier and 1–13 for the second). Each tier network was trained with its

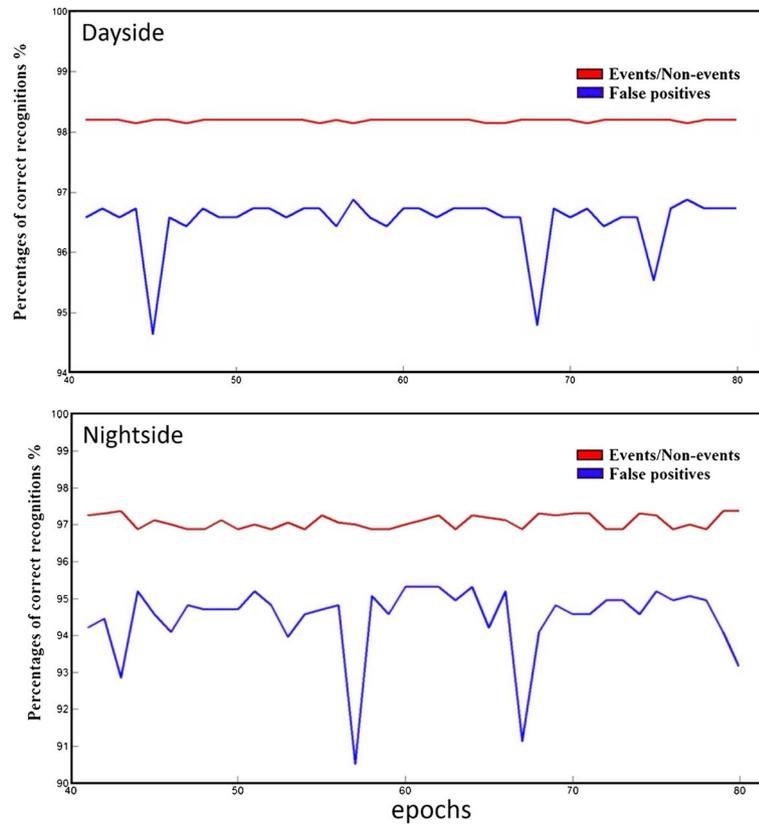


Fig. 6. Results of first stage classification for Dayside tracks.

corresponding training dataset for a wide range of epochs and then tested against the full five year dataset of wavelet classified tracks. The best results were found for a number of training epochs in the range of 40–80, in agreement with the typical behavior of ANFIS which usually require tens or few hundreds of epochs for a successful training. Figure 6 shows the accuracy scores (percentage of correct classifications) achieved from the network for the first tier in Dayside and Nightside.

As seen in Figure 6 the ANFIS network achieves correct classifications consistently higher than 95% for both Dayside and Nightside data while small variations appear for the different training epochs. The False Positive category shows slightly lower percentages of successful classification than the joint Events/Non-Events category, in both Dayside and Nightside, which implies that a small number of tracks classified as False Positives in the wavelet method are recognized as Events/Non-Events. On the other hand extremely few of the Events/Non-Events are recognized as False positives and the percentages in this category are above 98% and 97% for Dayside and Nightside, respectively. Finally, we note the small but consistent differences in both categories between Dayside and Nightside. As mentioned above Nightside tracks contain Plasma Instabilities which are not considered here. It is possible their inclusion in the training and/or testing datasets is the reason for the slightly lower performance in the Nightside data.

For the second tier network the same approach was followed and an investigation on the performance relative to the training epochs was made in the range of 40–80. The classification performance of the second tier is dependent on the

performance of the first as we forward each of the outputs of the first stage to that of the second. This results in a 40×40 grid where each cell is the overall performance in each of the Events and Non-Events category stemming from the combination of classifications in both tiers. Figures 7–10 show these combinations for Events and Non-Events in both Dayside and Nightside tracks just to highlight that for a wide range of training epochs, the performance of the system lies consistently above the 90% level.

As it can be seen the overall performance for Events and Non-Events remains consistently very close or above 95% for both categories in both Dayside and Nightside. It is noteworthy that the overall performance for Events and Non-Events is quite close to the percentages of the first tier showing decreases of 1–3%. The scores are calculated relative to the initial absolute number of Events and Non-Events, not only those that were forwarded from the first tier, and thus this means the second tier network achieves performances well above 95% and close to 100%. The best performing networks with their highest classification performances are shown in Table 3. Translating these results into Heidke Skill Scores (Heidke, 1926) yields values of 0.85 for the Dayside and 0.50 for the Nightside, both showing significant detection capabilities (since they are both > 0 with 1 being the perfect score). This also highlights the reduced identification prowess of the method in the Nightside tracks, since there are far more Non-Events in the Nightside than actual Pc3 Events and so even an accuracy score of 94.8% for the detection of Non-Events will yield a significant number of falsely classified tracks.

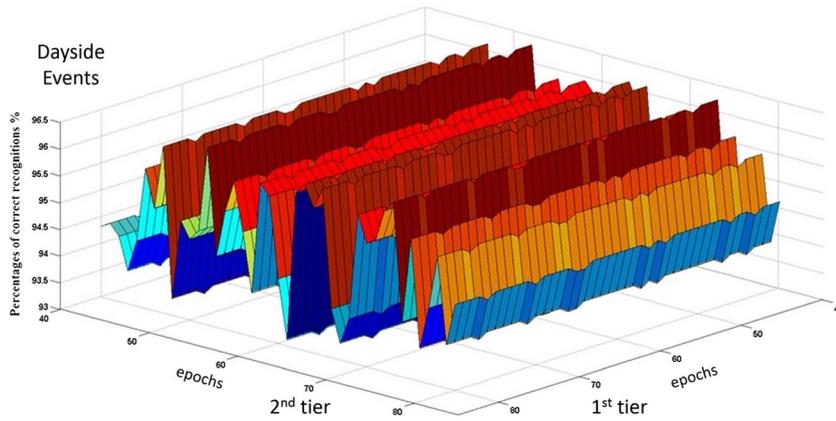


Fig. 7. Overall classification performances from all combinations of networks in 1st and 2nd tier for the category of Events in the Dayside.

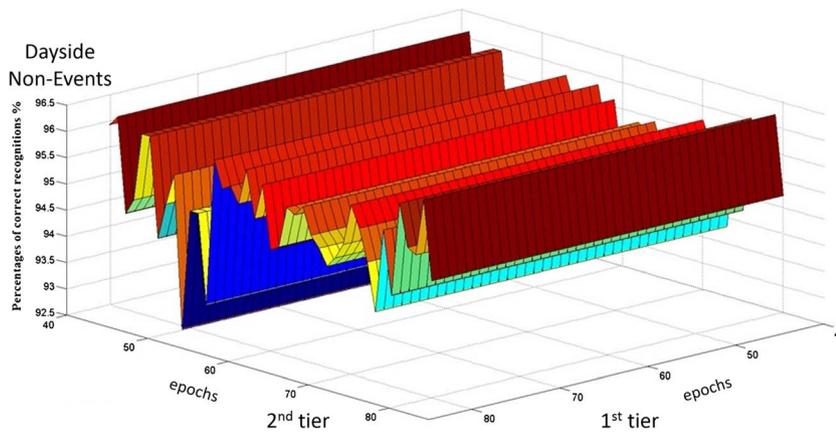


Fig. 8. Overall classification performances from all combinations of networks in 1st and 2nd tier for the category of Non-Events in the Dayside.

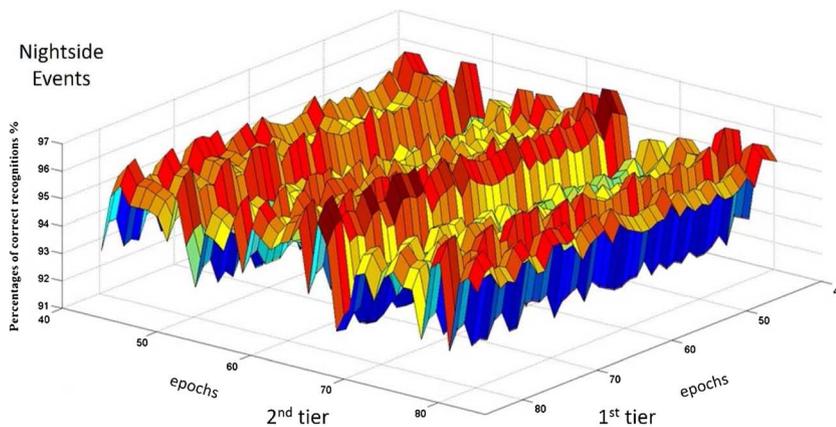


Fig. 9. Overall classification performances from all combinations of networks in 1st and 2nd tier for the category of Events in the Nightside.

It is interesting to note the difference in the overall behavior of the system, when tested against Dayside or Nightside data. As has already been stated, the performance on Dayside data shows slightly better scores for both categories, compared to Nightside data, even though the differences between them

remain small. In Dayside classification the performance of the second tier network seems to not depend on the performance of the first, resulting in the mostly straight ridges in the 3D plot which depend only on the specific performance of the second tier network. On the other hand in the Nightside plot there is

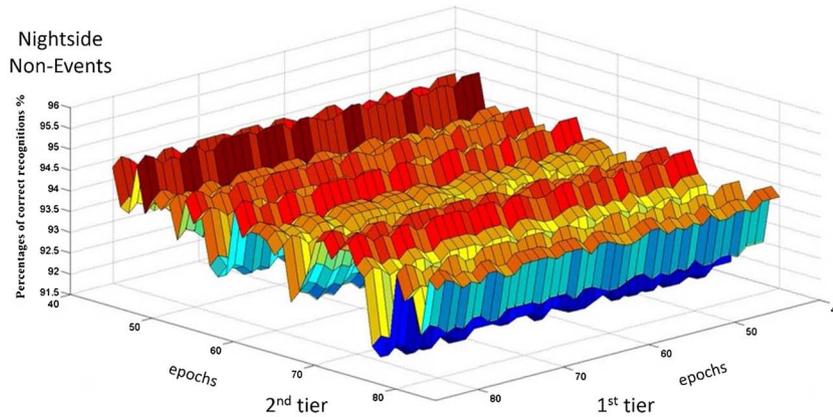


Fig. 10. Overall classification performances from all combinations of networks in 1st and 2nd tier for the category of Non-Events in the Nightside.

Table 3. Highest performing networks and classification scores.

Tracks	Dayside (%)	Nightside (%)
Events	95.7	95.4
Non-Events	96	94.8

clearly a dependence on the classification of the first tier resulting in the more “noisy” image. This is in agreement with the previously discussed existence of Plasma Instabilities which can be forwarded or not from the first to the second tier creating a slight decrease in the classification performance. Similar results were obtained when using this classification scheme against a smaller, human selected, dataset consisting of 193 Events, 240 Non-Events, and 150 False Positive tracks.

5 Conclusions

The two-tiered ANFIS classifier presented here shows encouraging results in the identification and classification of Pc3 ULF waves and thus proves its applicability for real-world applications. The final accuracy rates lie above or close to 95% for a large range of network parameters, which proves both the validity of the method as well as the robustness of the system. These scores were obtained by comparing the outputs of this system against the wavelet-based spectral detection method, a method which obviously is not completely error-free. It is true that sometimes e.g., a False Positive might co-occur with an Event, but it is important to keep in mind that the wavelet based method (that is used as reference) first checks for all other types of signals (false or background) and only if no such features are found, it then concludes that the candidate track is an actual wave event. Therefore, there is a strict order implied in the classification process that promotes all other types first and assigns the label “Event” at last. Since the ANFIS were trained using wavelet-classified tracks, they attempt to replicate this behavior despite following a completely different procedure. In this manner, their accuracy will always be bounded by the accuracy of the wavelet method. Nevertheless, the fact that the ANFIS

were able to replicate the results of the wavelet method almost perfectly, seems to indicate that such methods are capable to replace, or at the very least work in parallel to, more traditional approaches. An example of such a possibility would be a case where a track is checked by the wavelet method and is found to only barely pass the criteria for the identification of a Pc3 event. In such cases, using the ANFIS method could provide a fast and efficient second opinion, which would make the classification process more certain.

The CHAMP satellite can be considered as the predecessor of Swarm, a multi-satellite mission of European Space Agency (ESA), that was launched in November 2013. The ULF wave identification methodology based on neural network approaches presented here can be also applied to data from Swarm in order to observe ULF wave signal distributions in the topside ionosphere. The latter can be very useful when calculating the Swarm ULF wave index (Papadimitriou et al., 2018), which is a potential candidate for a new Swarm Level 2 product related to Space Weather.

The modularity of the system makes further extensions not only possible but straightforward, by adding more nodes in the system’s architecture, nodes which can be either other ANN or even completely different classifiers. One such extension, which we plan for the near future, is the addition of another layer of networks which will attempt to perform a third binary classification of the outputs of the previous stages, in order to properly identify wave signatures that were caused by Plasma Instabilities, although the introduction of more features derived from other data products (such as the Swarm Level 2 product of Ionospheric Bubble Index; Park et al., 2013) will be necessary for that.

Acknowledgements. This work has received funding from the European Space Agency under contract ESTEC 4000103770/11/NL/JA/ef. We also acknowledge support of this work by the project “PROTEAS II” (MIS 5002515), which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund). We gratefully acknowledge Helmholtz Centre Potsdam GFZ, German Research Centre for Geosciences, for supplying data

used in this study. The editor thanks Vyacheslav Pilipenko and an anonymous referee for their assistance in evaluating this paper.

References

- Anderson JA. 1995. *An introduction to neural networks*. MIT Press, Cambridge, MA.
- Balasis G, Daglis IA, Georgiou M, Papadimitriou C, Haagmans R. 2013. Magnetospheric ULF wave studies in the frame of Swarm mission: a time-frequency analysis tool for automated detection of pulsations in magnetic and electric field observations. *Earth Planets Space* **65**(11): 1385–1398.
- Balasis G, Papadimitriou C, Daglis IA, Pilipenko V. 2015. ULF wave power features in the topside ionosphere revealed by Swarm observations. *Geophys Res Lett* **42**: 6922–6930. DOI:10.1002/2015GL065424.
- Bishop C. 1995. *Neural networks for pattern recognition*. Oxford University Press, Oxford.
- Bogoutdinov ShR, Yagova NV, Pilipenko VA, Agayan SM. 2018. A technique for detection of ULF Pc3 waves and their statistical analysis. *Russ J Earth Sci* **18**: ES6006. DOI:10.2205/2018ES000646.
- Bortnik J, Chu X, Ma Q, Li W, Zhang X, et al. 2018. Artificial neural networks for determining magnetospheric conditions. In: *Machine learning techniques for space weather*. Camporeale E, Wing S, Johnson J (Eds.), Elsevier, Amsterdam, The Netherlands, pp. 279–300.
- Boynton RJ, Balikhin MA, Sibeck DG, Walker SN, Billings SA, Ganushkina N. 2016. Electron flux models for different energies at geostationary orbit. *Space Weather* **14**: 846–860. DOI: 10.1002/2016SW001506.
- Burton RK, McPherron RL, Russell CT. 1975. An empirical relationship between interplanetary conditions and Dst. *J Geophys Res* **80**: 4204–4214.
- Camporeale E, Wing S, Johnson J. 2018. *Machine Learning Techniques for Space Weather*. Elsevier, Amsterdam, The Netherlands **454** p. DOI:10.1016/C2016-0-01976-9.
- Daglis IA. 2006. Ring current dynamics. *Space Sci Rev* **124**: 183–202. DOI:10.1007/s11214-006-9104-z.
- Falconer K. 1990. *Fractal geometry: mathematical foundations and applications*. John Wiley, Chichester, pp. 38–47. ISBN 0-471-92287-0.
- Heidke P. 1926. Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geografiska Annaler* **8**: 301–349. DOI: 10.2307/519729.
- Lundstedt H, Gleisner H, Wintoft P. 2002. Operational forecasts of the geomagnetic Dst index. *Geophys Res Lett* **29**(24): 2181. DOI: 10.1029/2002GL016151.
- Mann IR. 2016. Waves, particles, and storms in geospace: an introduction. In: *Waves, particles, and storms in geospace*. Balasis G, Daglis IA, Mann IR (Eds.), Oxford University Press, Oxford, pp.1–14.
- O'Brien TP, McPherron RL. 2000. Forecasting the ring current index Dst in real time. *J Atmos Terr Phys* **62**: 1295–1299.
- Palloccchia G, Amata E, Consolini G, Marcucci MF, Bertello I. 2006. Geomagnetic Dst index forecast based on IMF data only. *Ann Geophys* **24**: 989–999. DOI: 10.5194/angeo-24-989-2006.
- Papadimitriou C, Balasis G, Daglis IA, Giannakis O. 2018. An initial ULF wave index derived from two years of Swarm observations. *Ann Geophys* **36**: 287–299. DOI: 10.5194/angeo-36-287-2018.
- Park J, Noja M, Stolle C, Lühr H. 2013. The Ionospheric Bubble Index deduced from magnetic field and plasma observations onboard Swarm. *Earth Planets Space* **65**: 1333–1344.
- Reed R, Marks RJ. 1999. *Neural smithing: supervised learning in feedforward artificial neural networks*. MIT Press, Cambridge, MA.
- Reigber C, Lühr H, Schwintzer P, Wickert J (Eds.). 2005. *Earth observation with CHAMP: results from three years in orbit*. Springer, Berlin. 628 p.
- Ritter P, Lühr H, Rauberg J. 2013. Determining field-aligned currents with the Swarm constellation mission. *Earth Planets Space* **65**: 1285–1294.
- Shing J, Jang R. 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* **23** (3):665–685.
- Sugeno M. 1985. *Industrial applications of fuzzy control*. Elsevier Science Pub Co., Japan.
- Souza VM, Medeiros C, Koga D, Alves LR, Vieira LEA, Dal Lago A, Da Silva LA, Jauer PR, Baker DN. 2018. Classification of magnetospheric particle distributions via neural networks. In: *Machine learning techniques for space weather*. Camporeale E, Wing S, Johnson J (Eds.), Elsevier, Amsterdam, The Netherlands, pp. 329–353.
- Stolle C, Lühr H, Rother M, Balasis G. 2006. Magnetic signatures of equatorial spread F as observed by the CHAMP satellite. *J Geophys Res* **111**: A02304. DOI: 10.1029/2005JA011184.
- Temerin M, Li X. 2002. A new model for the prediction of Dst on the basis of the solar wind. *J Geophys Res* **107**(A12): 1472. DOI: 10.1029/2001JA007532.
- Temerin M, Li X. 2006. Dst model for 1995–2002. *J Geophys Res* **111**: A04221. DOI:10.1029/2005JA011257.
- Wei HL, Zhu DQ, Billings SA, Balikhin M. 2007. Forecasting the geomagnetic activity of the Dst index using multiscale radial basis function networks. *Adv Space Res* **40**: 1863–1870.
- Williams AB, Taylors FJ. 1988. *Electronic filter design handbook*. McGraw-Hill, New York, USA. ISBN 0-07-070434-1.
- Wing S, Johnson JR, Jen J, Meng C-I, Sibeck DG, et al. 2005. Kp forecast models. *J Geophys Res Space Phys* **110**: A04203. DOI: 10.1029/2004JA010500.
- Wintoft P, Wik M, Matzka J, Shprits Y. 2017. Forecasting Kp from solar wind data: input parameter study using 3-hour averages and 3-hour range values. *J Space Weather Space Clim* **7**: A29.
- Zhelavskaya IS, Shprits YY, Spasojevic M. 2018. Reconstruction of plasma electron density from satellite measurements via artificial neural networks. In: *Machine learning techniques for space weather*. Camporeale E, Wing S, Johnson J (Eds.) Elsevier, Amsterdam, The Netherlands, pp. 301–327.

Cite this article as: Balasis G, Aminalragia-Giamini S, Papadimitriou C, Daglis IA, Anastasiadis A, et al. 2019. A machine learning approach for automated ULF wave recognition. *J. Space Weather Space Clim.* **9**, A13.