Fast Dst computation by applying deep learning to Swarm satellite magnetic data

Gianfranco Cianchini [a], Alessandro Piscini [a,*], Angelo De Santis [a], Saioa A. Campuzano [b]

[a] *Istituto Nazionale di Geofisica e Vulcanologia (INGV), Via di Vigna Murata 605, 00143 Roma, Italy*
[b] *Instituto de Geociencias IGEO (CSIC-UCM), Ciudad Universitaria, Madrid 28040, Spain*

Abstract

Dst (Disturbance Storm Time) is an hourly index of magnetic activity computed from the measured intensity of the globally symmetrical equatorial electrojet (Ring Current) obtained by a series of near-equatorial geomagnetic observatories. We selected and trained an Artificial Neural Network (ANN) to give the estimation of the Dst index through the magnetic data measured by the Swarm three-satellite mission. From November 2014 to December 2019, we selected a balanced number of quiet and disturbed days, to get the most uniform set of Dst index values as possible. We then collected a big data collection of Swarm magnetic signals, confined to three very narrow belts of low-to-mid latitude: this choice allows it to better resemble the geographic distribution of the geomagnetic observatories contributing to the calculation of Dst. We also extended the analysis to mid latitude locations to increase the number of satellite samples. Once we determined by means of simulations the best network topology, we trained the network and tested its capabilities. The outcomes show that the ANN is able to give a reliable fast estimation of the Dst index directly from Swarm satellite magnetic data.

1. Introduction and motivation

The processes underlying the geomagnetic activity are rather complex. Geomagnetic storms have a great impact on the terrestrial space environment, which can affect power lines, spacecraft operations, and telecommunications (Ayala Solares et al., 2016). In turn, they can affect electronic devices and Global Positioning Systems (GPS) on Earth causing a lot of damage to the technological infrastructures. A geomagnetic index describes the intensity of the geomagnetic disturbance for a certain period of time, all over the globe or regionally. There are sets of indices used to describe such phenomena in Geomagnetism: some of them describe the general level of geomagnetic activity at global scale, as Kp or Ap (planetary indices); and some others describe a particular type of geomagnetic activity in specific areas, such as the Dst or AE (Auroral Electrojet) indices, useful for characterizing equatorial or auroral region magnetic activity, respectively. In this work, we focused on the estimation of the Dst index because it characterizes the geomagnetic storm evolution (Wanliss, 2004).

The hourly Dst index represents a measure of the intensity of the symmetrical part of the ring current near the equator (Wanliss, 2004). This current is mainly due to the drift of both the positive and negative charged particles bouncing back and forth polarward along the magnetic field lines resulting in a net current lying in the equatorial plane and circulating clockwise around the Earth (when viewed from the north) at a distance of 3 to 8 Earth radii (R_E).

The Dst is calculated from the horizontal magnetic field component at four low-latitude magnetic observatories: Honolulu, San Juan, Hermanus[1] and Kakioka, properly chosen because they are well distributed in longitude and quite far from both the auroral and equatorial electrojets (Wanliss, 2004).

Since it is conceived to measure the perturbations caused near the equator by the current systems flowing above the ionosphere, both Solar Quiet (Sq) contribution of the day-side currents in the ionosphere and the secular variation due to internal sources are considered for each observatory. In particular, for taking into account the secular variation, the baseline for each observatory is calculated through a quadratic fit to the yearly average values of the five quietest days of each month for each year, according to the procedure described in Sugiura and Kamei (1991) (see also Temerin and Li, 2015): hence, its definitive value requires the end of the current year and is released after many months. Meanwhile, the provisional Dst is usually delivered by the World Data Center for Geomagnetism (Kyoto - Japan, http://wdc.kugi.kyoto-u.ac.jp/).

The complexity behind its calculation and the delay in its release are the reasons why we explored the possibility of giving a rapid evaluation of the Dst index exploiting the property of ANN at application stage, once it completed its training phase. In other words, starting from the virtually quasi real-time magnetic data, measured by the European Space Agency (ESA) Swarm satellite mission, we try to find out whether it is possible to retrieve the Dst definitive values as accurately as possible by means of a properly designed artificial neural network (ANN). Although recently a Swarm-based Dst index has been derived by Balasis et al. (2019), as far as we know, this is the first time that a nonlinear approximation method as the ANN is applied to Swarm magnetic data for the purpose of estimating the Dst index.

ANNs have demonstrated their ability to model nonlinear physical systems (Rumelhart et al., 1995) involving complex behaviours, taking into account any nonlinear relationship between the explanatory and dependent variables, especially when they use multiple layers or Deep Learning architecture (Lek and Guegan, 1999; Witten et al., 2017).

ANNs seem to perform more accurately than other techniques such as statistical ones, especially if functional space is particularly complex and the input data might have different statistical distributions (Benediktsson et al., 1990, 1993; Schalkoff, 1992).

Cybenko (1989), Hornik et al. (1990) and Hornik (1991) demonstrated the capability of feed-forward architecture of approximating any continuous function: this capability goes by the name of "universal approximation theorems" (Kratsios, 2021).

Their ability comes from a process that allows them to learn from data and make predictions: at the start, the model represented by the ANN is fit on a *training* dataset (a set of input–output pairs of tuples as examples) used to compute the "weights" of connections between neurons, through a "supervised" learning and an optimization method. The fitted model is then used to predict the response to a second dataset (called the *validation* dataset): this dataset is held back from the training dataset and is used to give an unbiased evaluation of the model skills, thus helping in tuning the parameters of the model. Usually, a final dataset (called *test* dataset) is used to provide the score of the optimum model obtained in the training stage. At the end of these processes, i.e. once the so-called *learning phase* has been completed and the model is optimized, it can be applied very quickly to new data, in order to drastically reduce the computational load required for data management (e.g. James et al., 2013 for more details on this topic).

The approximation properties of the ANNs are useful in many fields of application, for example the forecasting of some physical phenomena in which it is necessary to solve direct or inverse problems. In particular, the technique has been proved successful in Remote Sensing (Atkinson and Tatnall, 1997; Mas and Flores, 2008) by addressing various geophysical issues, such as: retrieve surface wind speed (Krasnopolsky et al., 1995), height-resolved ozone recoveries (Müller et al., 2003; Iapaolo et al., 2007), temperature estimates (Butler et al., 1996), estimate of humidity profiles (Blackwell, 2005; Cabrera-Mercader and Staelin, 1995); and more recently, for the quantitative estimate of volcanic emissions (Picchiani et al., 2011, Piscini et al., 2014a, Piscini et al., 2014b), the monitoring of volcanic activity for the lava flow detection (Piscini and Lombardo, 2014) and the earthquake damage assessment (Piscini et al., 2017).

Recent studies have shown the use of several different kinds of ANNs for diverse geomagnetic indices forecast, by making use of the solar wind parameters and indices from interplanetary magnetic field measured by the Advanced Composition Explorer (ACE) spacecraft at the libration point L1: e.g. the Bala model, based on ANN, was introduced to forecast three indices: Kp, Dst, and AE at 1 h, 3 h and 6 h in advance (Bala and Reiff, 2012). Other studies have introduced sophisticated techniques/algorithms: for example long short-term memory (LSTM) algorithm for Dst and Kp estimation (e.g. Myagkova et al., 2021,2017; Efitorov et al., 2018; Tan et al., 2017); multiscale radial basis function (MSRBF) (Wei et al., 2007); Support Vector Machine (SVM) and Distance Correlation (DC) (Lu et al., 2016). Correctly, their approach tried to model the dependence of the geomagnetic effects on their sources, i.e. the solar activity, from the large variation of the solar wind parameters and the Interplanetary Magnetic Field (IMF) orientation. A very recent noticeable improvement of this approach is represented by the direct inclusion of the observations of radiative phenomena on

---

[1] Now called SANSA Space Science, previously the Hermanus Magnetic Observatory (HMO).

the solar disk by Tasistro-Hart et al. (2021), which also provides a brief review of articles devoted to forecasting the Dst index.

As explained before, our aim and consequently our approach to the problem of the geomagnetic indices estimation is different. In fact, in recent times many Low Earth Orbit (LEO) satellite missions, like Ørsted (1999), CHAllenging Minisatellite Payload (CHAMP, 2000–2010) and the newly Swarm (2013) from ESA, have been developed for a better study and understanding of the geomagnetic field closest to the Earth by collecting a global mapping of it. Our aim is to explore the feasibility of deriving an estimation of the magnetic index Dst, as accurately as possible, from a neural network trained to recognize and model the relationship between the magnetic components measured by LEO ESA Swarm satellites (at least one of them) and that global index. In addition, we want to find out what degree of reliability we get by using a neural network approach for magnetic index retrieval, because with this type of networks, as soon as the magnetic data of the satellite missions (present and future) are released, it would be possible to provide a quick estimate of the Dst index.

The article is organized as follows: Section 2 presents the data (from selection to transformation) and methods (neural architecture) adopted; Section 3 describes the results of the ANN and the test to unseen data; Section 4 discusses the main results and Section 5 mentions the most important conclusions and some possible future developments.

2. Data and methods

The ESA Swarm satellite mission (Friis-Christensen et al., 2006) has been successfully orbiting since November 2013: at the time of writing this manuscript, the three identical satellites of the mission are still monitoring the geomagnetic field and its variations: each satellite has both scalar and vector magnetometers, together with additional complementary instruments to monitor the ionospheric plasma density, temperature and composition. The three satellites are often indicated as Alpha, Bravo, and Charlie, or indifferently Swarm A, Swarm B and Swarm C. The orbital configuration is selected to achieve the various mission objectives: Alpha and Charlie fly close each other at a separation of about 150 km at an altitude of about 445 km and Bravo satellite is on a higher orbit of about 517 km (at May 2021) and almost perpendicular to the other two.

Generally, ESA provides Swarm data within 3–4 days from acquisition time. In this paper, we made use of the magnetic field data with a resolution of 1 Hz from April 2014 to April 2020.

Hornik et al. (1990) showed that the ANNs in general, and the multilayer perceptrons (MLPs) in particular, are capable to approximate any kind of highly nonlinear functions, once they are trained to generalise those functions when new unseen data are presented (Cybenko,1989).

A MLP is composed of a net of interconnected perceptrons (Rosenblatt, 1962), the so-called "neurons" (the nodes), which constitute a nonlinear mapping between the input $[x_i]$ and the output vectors $[y_j]$. Each node is connected to the others through weights and output signals, the latter being the response of the simple nonlinear function, the transfer or activation function, to the sum of the inputs to the node. When the network combines the inputs (also named "features") to obtain a given output, the ANN "learns" the unknown function $y = f(x)$ that mimics the desired response (Principe et al., 2000).

A back-propagation neural network (BPNN) is a MLP consisting of an input layer with nodes representing input variables to the problem, an output layer with nodes representing the dependent variables (i.e. what is being modelled), and one or more hidden layers containing nodes to help capture non-linearities in the data, in which the neurons can be fully or partially interconnected (Hecht-Nielsen, 1990).

This work uses a BPNN with several hidden layers, which falls into the so-called "Deep Neural Network" (Witten et al., 2017, chap.10). Such networks are widely used for their flexibility and adaptability in modelling a broad spectrum of problems in many application areas (Rumelhart et al., 1986). Moreover, Gardner and Dorling (1998) and Hsieh and Tang (1998) showed how this kind of ANNs can be highly effective in the solution of atmospheric inverse problems, and successful in classification and regression problems as well (Tasistro-Hart et al., 2021). For example, for remote sensing applications, usually, the input layer collects data values (the spectral signature of the respective pixel of the image) such as radiances or brightness temperatures from different spectral bands, with the number of nodes equal to the number of bands; meanwhile the output layer provides the corresponding retrieved geophysical parameter value. In this perspective, as typical for an ANN (Nilsson, 1996), we can consider the Dst index as the result (output) of a nonlinear mapping through the network (layers) of the magnetic measurements acquired by the magnetometer on-board the Swarm satellite along with some other features (input), which are as much as possible representative of the underlying physics: we will describe all of them in detail below.

One of the most important preliminary phases of the machine learning approach is the "feature extraction": this stage aims at eliminating the redundancy present in input data, and facilitating generalization during the training phase. Therefore, we focused on the minimum number of data that reflect the dynamics of the external driving forces and environmental conditions and capture the essential patterns, at the same time.

First, from the World Data Center for Geomagnetism (http://wdc.kugi.kyoto-u.ac.jp/qddays/index.html) in the time interval between November 2014 and December 2019, we selected all disturbed days (five per month), and around the 20% of quiet - and quietest - days in the same

period: in this way we tried to reduce the predominance of Dst values corresponding to quiet days with respect to the less numerous disturbed ones. Then, in order to disrupt any possible time dependence to sequentiality and to further mix the distribution of Dst values, we shuffled this list of dates.

Then, according to the shuffled list of dates, we selected magnetic data (X, Y, Z components) of the Swarm C satellite only: the reason why we did not consider Swarm A is that it does not add significant information, being it practically parallel to its twin satellite C. For what concerns Swarm B, in this work we did not consider making use of its data in order to possibly avoid unnecessary bias or complexity, due to different measurement conditions (higher altitude and different local time) with respect to the other satellites.

To resemble the geographic distribution of the four low-latitude magnetic observatories (Honolulu, San Juan, Hermanus and Kakioka) participating to the usual Dst computation, we limited to data within three narrow belts of latitude, each with 0.1° width at 20°, 30° and 40° both north and south (Fig. 1).

Since we are interested in the effects of the external contributions to geomagnetic signals only, we removed from the three components X, Y, Z the correspondent internal (core and lithosphere) sources computed through the CHAOS-7.3 model (Finlay et al., 2020; this version was the most up-to-date available when preparing the datasets): in this way we removed the contribution of the longest wavelengths, including the secular variation of the geomagnetic field. For illustrative purposes, we plot the residual magnetic components of both the ANN training and vali-

dation datasets, distributed as shown in Fig. 2: it is noteworthy the symmetry around zero of both the residual Y and Z components. On the contrary, the residual X appears asymmetrically distributed: this can be easily explained by the particular sensitivity of this horizontal component to variations or increases in the ring current.

Swarm satellite trajectories are designed to drift moving across the different local times during its mission while moving along a quasi-polar orbit. Therefore, we needed to cope with the different conditions affecting the measurements spanning six years: dayside, nightside, seasonal solar position etc. To this purpose, we tried to reduce those differences, or at least to take them into account, by transforming the geographic coordinates into magnetic latitude, longitude and local time in the altitude adjusted corrected geomagnetic (AACGM) coordinate system (Shepherd, 2014; Laundal and Richmond, 2017) which, together with the satellite height above the surface and the residuals magnetic components (X, Y, Z), constitute the "features" to give as input to the network. Please, note that according to that reference system, AACGM coordinates are undefined in some small regions close to the magnetic equator, the largest of which lies between Africa and South America (see Laundal and Richmond, 2017 for more details). Therefore, we systematically removed all data falling within those regions, which however constituted a very negligible fraction of the total.

In conclusion, taking into account the above considerations and to include both the dynamics of the external driving forces and the environmental measurement conditions, we have selected the following seven variables as input "features" to the network: the magnetic latitude, longitude
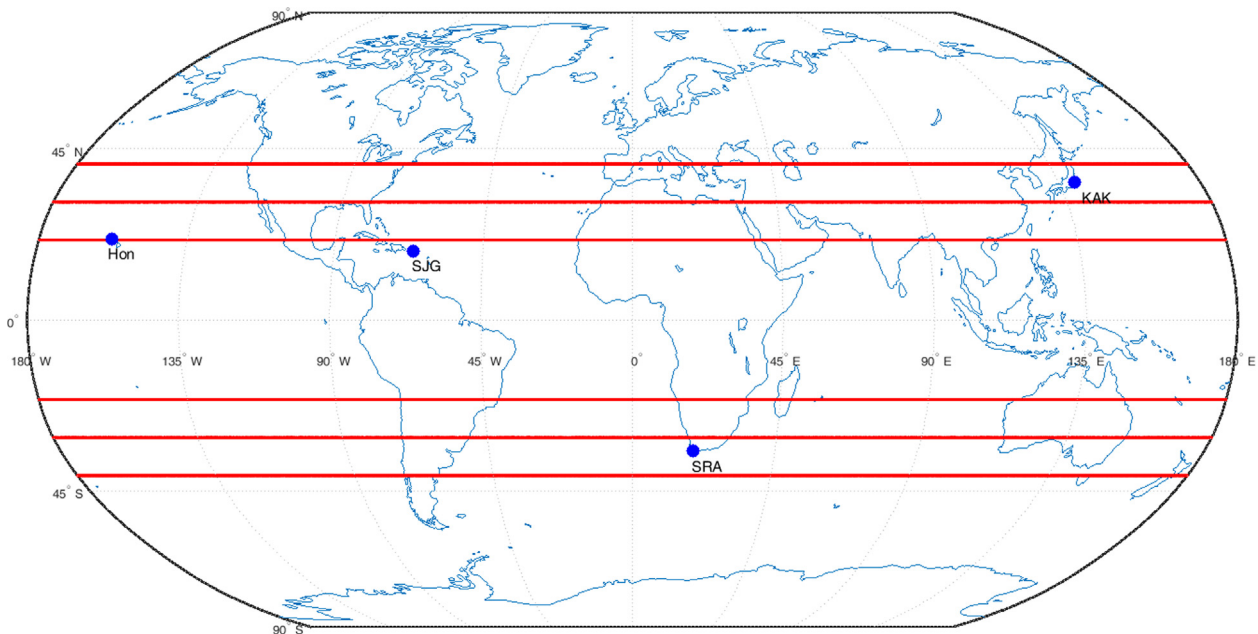

Fig. 1. The geographical distribution of the whole input data: the blue dots represent the location of the four geomagnetic observatories involved in the usual computation of the Dst index. The red lines are the thin bands of latitude where the data inputs were selected to overlap the latitude of the ground observatories. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
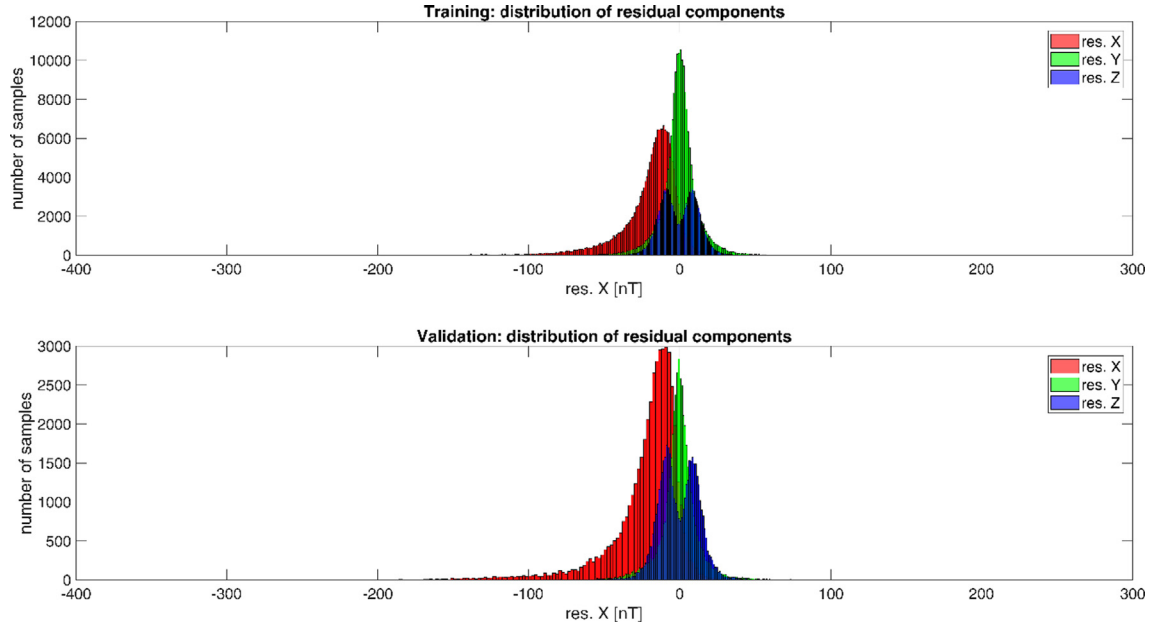
Fig. 2. The distribution of the three residual magnetic components for both training and validation datasets after removing the internal (core and crust) contribution from CHAOS-7.3 model. Note that the residual Y and Z components are symmetrically distributed around zero. On the contrary, the residual X shows an evident asymmetry towards the negative values, reflecting the effect of the magnetic disturbance caused certainly (but not only) by the enhancement of the ring current.

and local time (MLat, MLon and MLT, respectively), and the satellite height together with the three geomagnetic component residuals (res. X, Y and Z).

As output for the ANN, we considered the Dst values, each associated with the corresponding input tuple. As a whole, the dataset consists of more than 236,000 tuples: about 189,000 and 47,000 for the training and validation stage, respectively.

The distributions of the Dst index for the two datasets are given in Fig. 3. We notice that the differences in the tails of their distribution evidence the predominance of small values (quiet days) compared to the larger ones: of course, this is due to the limited number of disturbed days. This limited number is also the reason why we intervened manually when the most disturbed day values were accidentally included in the validation dataset, so the network could learn to recognize cases of more severe magnetic storms.

We used the MLP which trains using backpropagation, with no activation function in the output layer (i.e. using the identity function), and the *tanh* as the activation function for the rest of the nodes. We also chose the Levenberg-Marquardt *backpropagation* algorithm as the solver, which uses the Jacobian for calculations, because it is one of the fastest backpropagation algorithms, particularly useful when dealing with tens of thousands of training samples.

One of the main issues in Neural Network is the selection of the "optimal" network architecture: with this term, we mean a collective name for network topology, training algorithm and parameters, among others (Twomey and Smith, 1995).

In order to select the optimal network topology for our purpose, we performed a sort of "trial and error" preliminary test to assess it, to avoid both overfitting and poor generalization capability.

Through a MATLAB© (Matlab, 2020) script conceived for this scope, we first divided data into two sets, which we call the "Training" and the "Validation" dataset, respectively: the first one was used during the training stage, when the software further created three subsets (called training, testing and validation in the proportion of 70%, 20% and 10%, respectively, as described in the Introduction); the second one was used to validate the results, i.e. to test the capability of the network to generalize, since those inputs were never presented to the network before. We emphasize that not only do the datasets include both magnetically quiet and disturbed periods, but that they are also homogeneous with each other, all belonging to the same six-year time span; in addition, we point out that the Validation dataset is used to determine the best topology, as described shortly after.

Then, in order to find the best network, we recursively explored the performances of a series of topologies (with one, two and three hidden layers) each with the same input and output datasets: we proceeded from one to three hidden layers by progressively increasing the number of neurons in each layer: from 7 to 60 for the one-layer, and from 3 to 25 (for each layer) of the two hidden layer network. Instead, for the three-layer network, we chose a symmetric scheme ($-2i - i$ neurons, with i ranging from 1 to 9) for two reasons: first, because a systematic variation in the
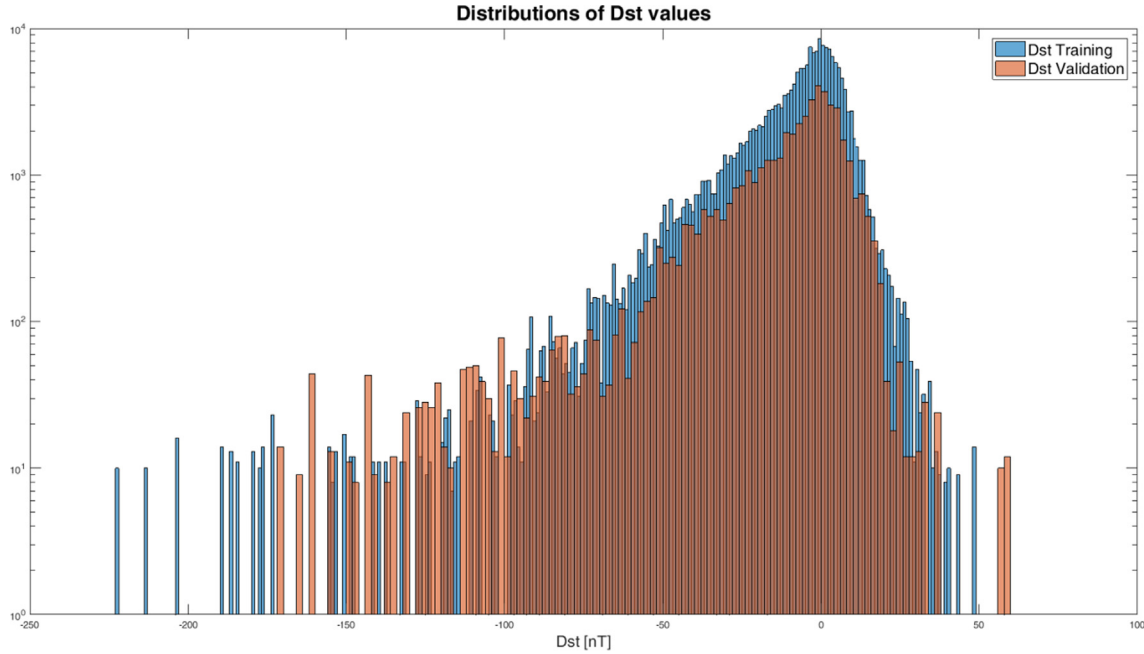
Fig. 3. Linear-log distribution of the Dst index collected for both the training and validation stages. These two datasets are almost identical, but the lowest Dst values are associated with the training (see main text for more details).

neuron numbers in each of the three hidden layers would have been very time-consuming; second, because symmetry is important in a number of situations, for instance, in order to reproduce some particular nonlinear phenomena (Bishop, 1995).

In literature (e.g. Twomey and Smith, 1995), there are different parameters used to measure the performance of the networks: among them, the Mean Absolute Error (MAE)

$$MAE = \frac{\sum_{k=1}^{N} |y_k - p_k|}{N}, \tag{1}$$

with y_k being the actual value, p_k the predicted one and N the number of the samples; the Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE), this latter being simply the square root of MSE:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{k=1}^{N} (y_k - p_k)^2}{N}} \tag{2}$$

We chose to measure the performance of a trained NN by means of the RMSE computed according to Eq. (2).

For each combination, we collected and compared the RMSE both for the Training and Validation datasets: the resulting performances, as a function of the different topology and number of neurons, are shown in Fig. 4.

The three panels of Fig. 4 show the RMSE obtained during Training (black line) and Validation (red line) for one, two and three hidden layers. Focusing on the performance on validation (red curve) for both the one and three hidden layer networks, we see that they show a common behaviour: a decrement which reaches a minimum, followed by a rising. The minimum corresponds to 27 neurons

and 7–14-7 for the networks with one and three hidden layers, respectively. On the contrary, for the network with two hidden layers we can notice a cycling behaviour between minima and maxima along a general stable plateau: that is why we decided to continue our investigation of the first and last topologies, i.e. the more interesting ones. Therefore, we compared the performances of the one hidden and three hidden layers with the aforementioned number of neurons (27 and 7–14-7, respectively), corresponding to the minimum of the RMSE for the validation dataset: incidentally, it is worth noting that the number of neurons in those networks results in a very similar number of model parameters (around 250).

As resumed in Table 1 and shown in Fig. 5 (which collects the scatter plots between the actual and the predicted Dst values for both topologies for the Training dataset) the general behaviour of the networks is very similar during the learning phase. Although the standard deviation of their outputs and the R coefficient are all comparable with each other, slight differences seem to emerge when visually comparing them: at the lower Dst values, the dispersion of data is apparently less for the three-hidden than for the one-hidden layer network. For practically equivalent results in terms of performance, we decided to choose the network that better "predicted" the lower Dst values, corresponding to the most disturbed magnetic conditions: therefore, we selected the one with three hidden layers made of 7, 14 and 7 nodes, respectively.

3. Results

The generalization capability and retrieval accuracy of our ANN was evaluated for some distinct independent
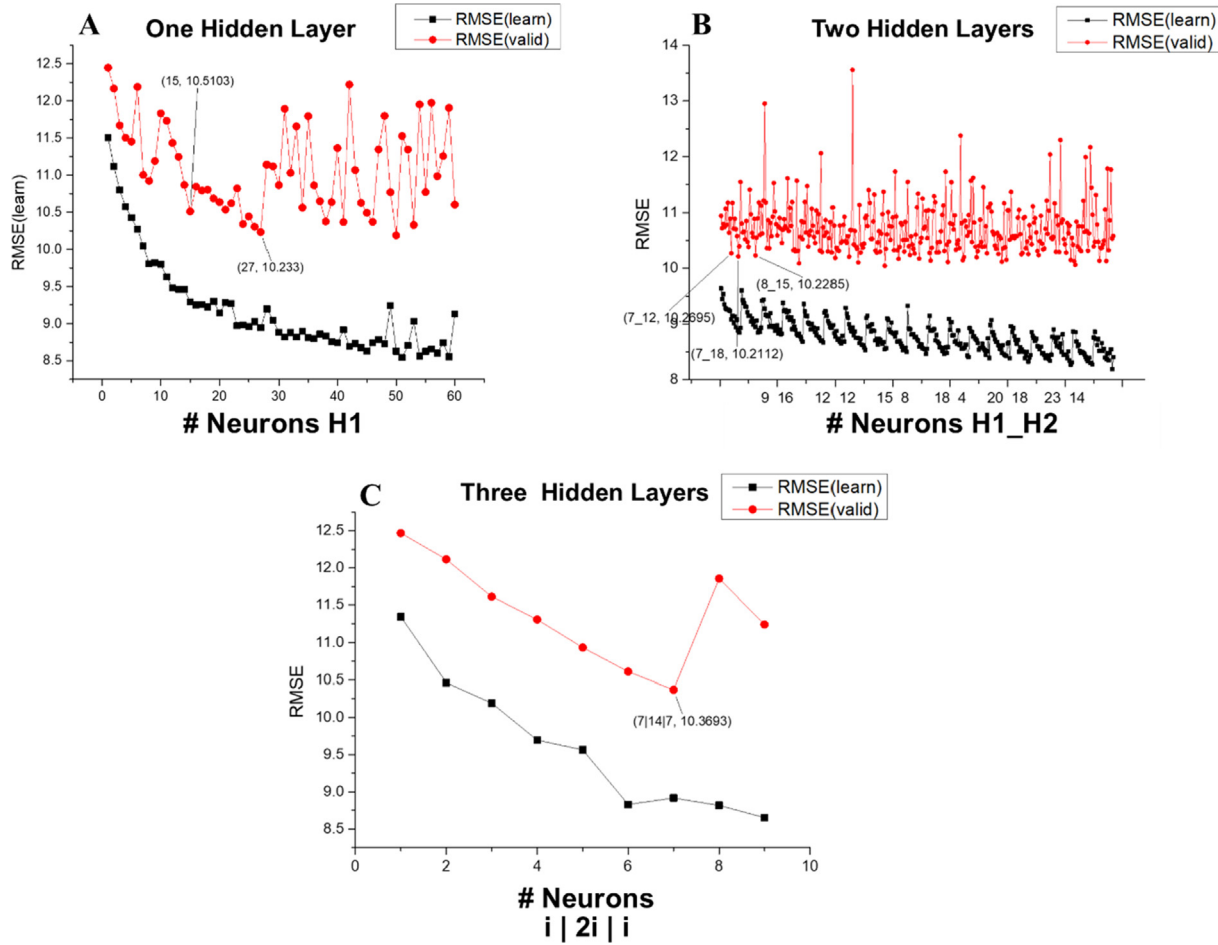
Fig. 4. The topology selection through training-validation trials and evaluation of the RMSE for the different cases: training (black) and validation on the unseen dataset (red). We focused on the validation results because our aim was the "best case network", in terms of its ability to generalize. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Comparison of the results for the two network topologies in terms of number of samples, RMSE, standard deviation (target and model) and correlation coefficient.

		Samples	RMSE	Sigma (target Dst)	Sigma (output Dst)	R
One Hidden	**Training**	~189 k	8.95	18.77	16.52	0.879
Three Hidden	**Training**	~189 k	8.92	18.77	16.54	0.882

datasets, including both quiet and disturbed samples in order to evaluate ANN performance in different electromagnetic conditions. To do so, we compared the prediction error for both the Training and the Validation stages shown in Fig. 6. As a rule of thumb, it is commonly assumed that the accuracy of a regressive neural network is assessed when the RMSE is less than the standard deviation (STD) of the target dataset $\{y_k\}$: a comparison of their values in Table 2 shows that this is the case for our network, being the RMSE much less than the standard deviation.

It is noticeable that the application of the ANN model to the validation set shows the RMSE value, 10.15, well

below the corresponding values of the target standard deviation (STD), 22.71.

Another way to verify whether the network generalizes well is by means of the error distribution: the error is meant as the difference between the actual and the computed Dst values $(y_k - p_k, k = 1...N)$. Looking at the error histograms for both the Training and Validation stages (Fig. 7) what emerges is that in most cases the absolute value of the error is less than around 10 nT for both cases (56% and 59%, respectively); in addition, around 94% of the outcomes stay confined within the interval $[-17, 17]$ nT: this may indicate a good generalization capability of the network.

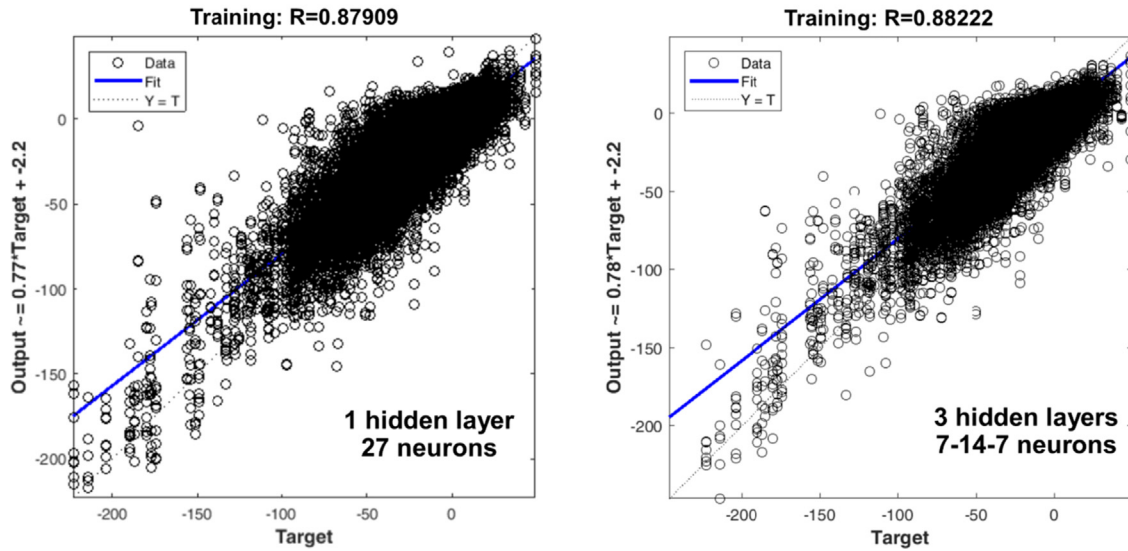## Topology Performance Comparison

Fig. 5. Scatter plot between the target and the modelled Dst for two different topologies: (right) the results of the network with 1 hidden layer; (left) the result of the network with 3 hidden layers.
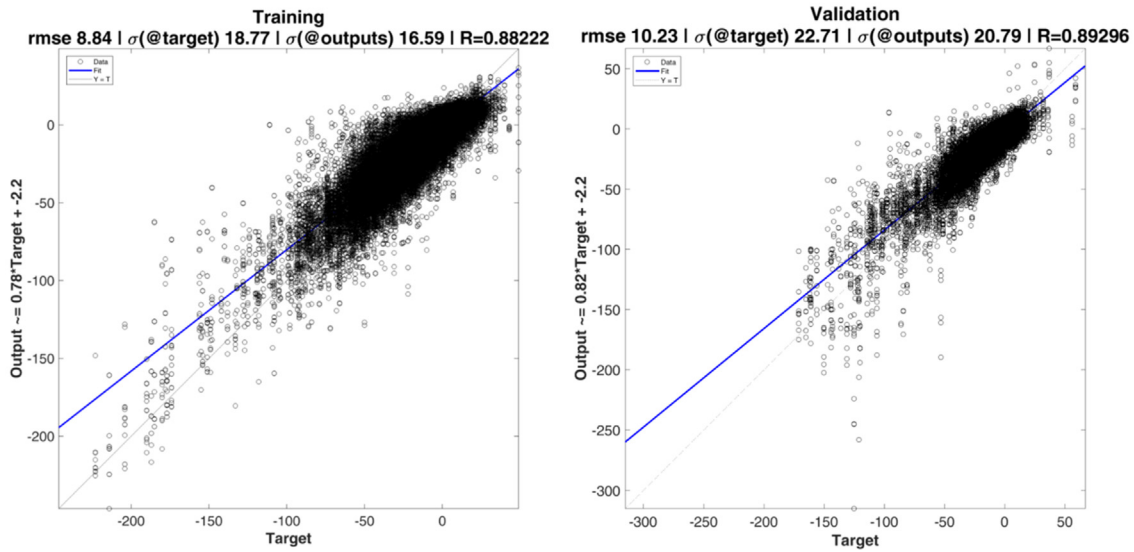

Fig. 6. Scatter plot between Dst target and model for training set (left) and validation set (right).

Table 2
Summary of training and validation sets in terms of number of samples, RMSE, standard deviation (target and model) and correlation coefficient.

	Samples	RMSE	Sigma (target Dst)	Sigma (output Dst)	R
Training	~189 k	8.84	18.77	16.59	0.882
Validation	~47 k	10.15	22.71	20.95	0.8959

In order to test further the network performance in an operational fashion, we applied our ANN model to other unseen data. We selected a total of six days in 2020, three quiet and three disturbed, and prepared the corresponding input from the Swarm C data. We then applied the network to each corresponding dataset: the selected dates and the associated results are listed in the next Table 3, and the Fig. 8 and Fig. 9 show the regression (right) and the distribution of the prediction error (left) for each test case.

First, looking at Table 3 we can notice that just in two cases the RMSE is lower than the standard deviation of the target: the 20 April and 31 March 2020, which are interestingly classified as geomagnetically disturbed. Results about the first case (20 April) are depicted in Fig. 8A,
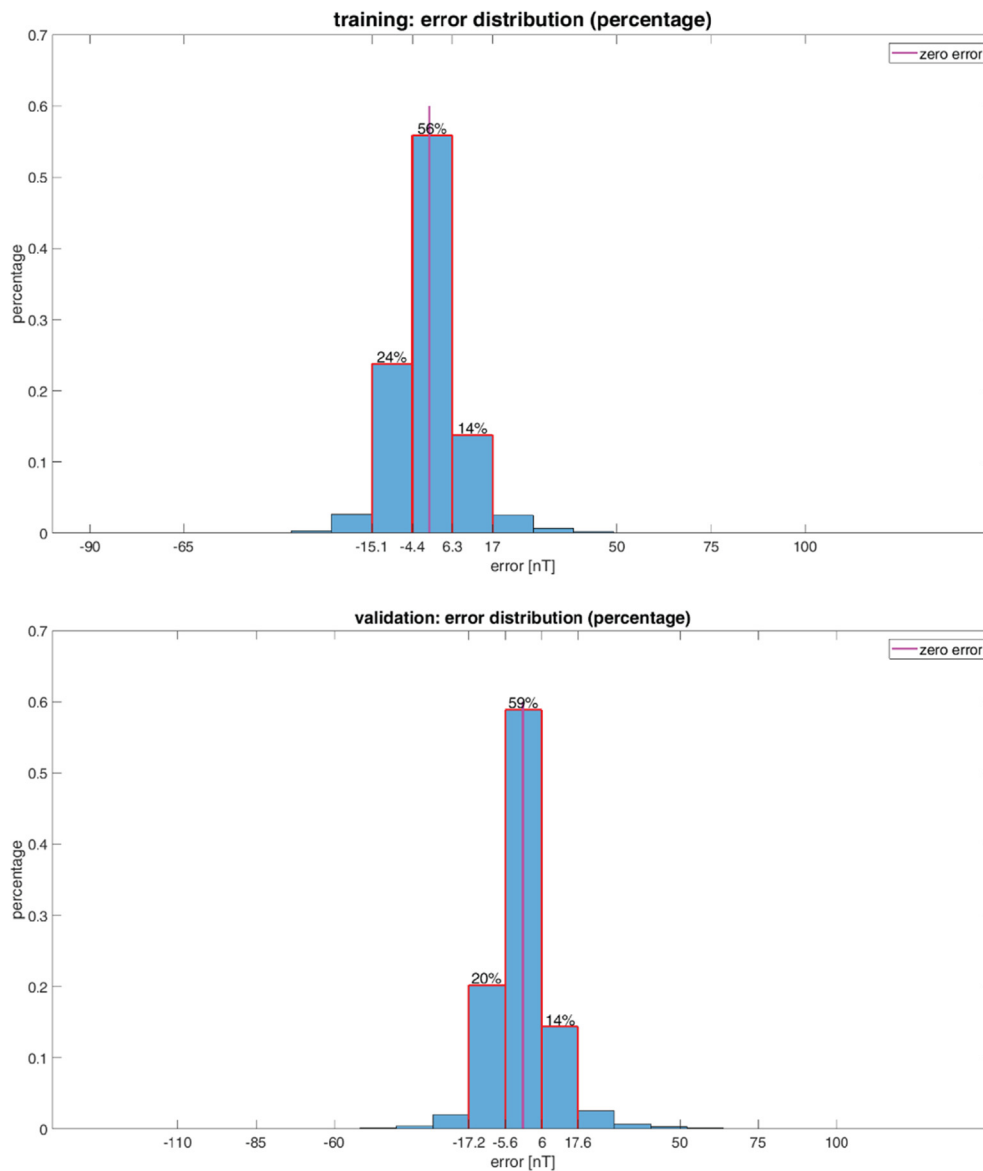
Fig. 7. Distribution and percentage of the prediction error of the network on the training (seen) and validation (unseen) datasets. It is noticeable that almost 56% (59%) of the cases are distributed around zero (±10 nT); 94% (93%) stay within the interval [-15 17] ([-17 17]) nT for the training (validation) dataset.

Table 3
Summary of the results of the ANN on unseen data for three quiet (Q) and three disturbed (D) days in 2020.

Type *Disturbed* *Quiet*	Date	Samples	R	RMSE	Sigma (target Dst)	Sigma (output Dst)
Q	2020 02 03	281	0.45	6.21	4.40	4.81
Q	2020 03 14	294	0.06	4.35	2.17	3.49
Q	2020 04 30	288	0.72	7.75	3.95	3.82
D	2020 04 20	297	0.86	12.23	23.92	20.68
D	2020 03 31	287	0.69	7.47	9.35	8.91
D	2020 02 07	290	0.22	7.06	4.66	6.40

which shows a good accordance between target and model, and the error distribution is centered on quite small values.

By looking at the figures of the results for the three disturbed days, we note that the greater the absolute value of

the Dst is, the more the network is able to follow its trend. This is particularly true for the case of the 20 April and 31 March 2020: noticeably, they are the two cases in which the RMSE is lower than the standard deviation of the target.
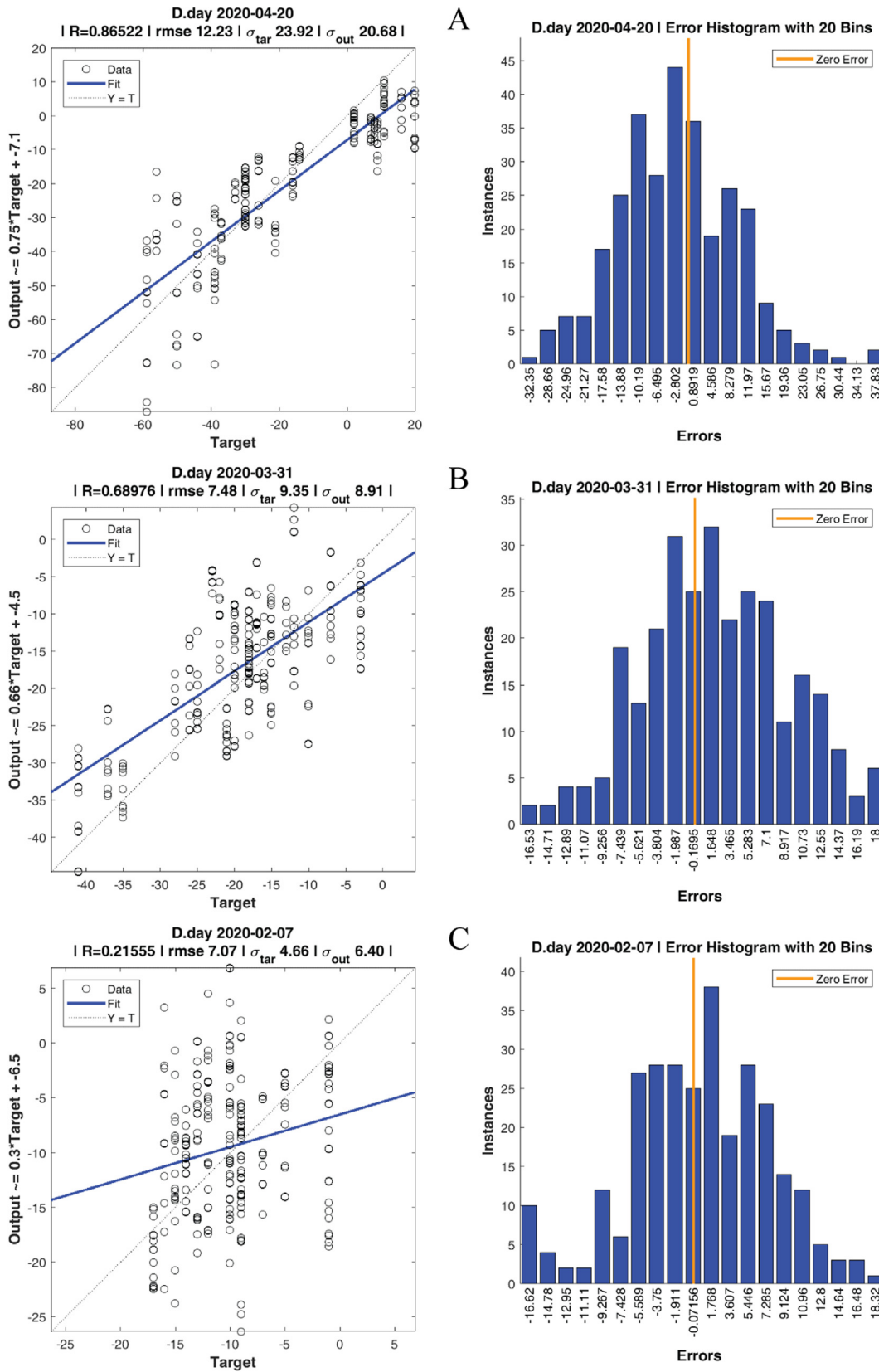
Fig. 8. Test cases for disturbed days (04–20, 03–31 and 02–07). (Left) Scatter plot and regression parameters; (right) histogram of the prediction error in nT.
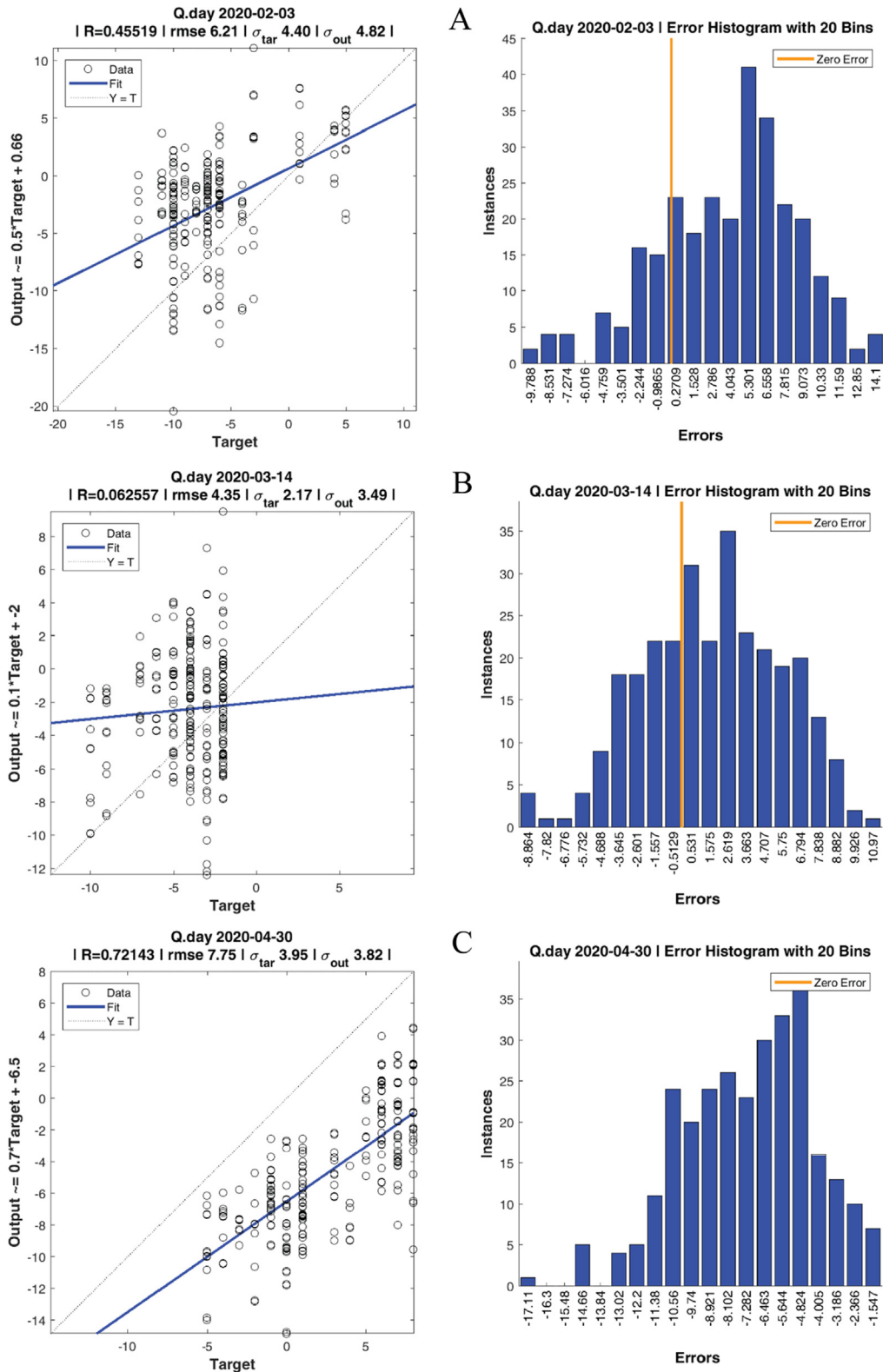
Fig. 9. Test cases for quiet days (02–03, 03–14 and 04–30). (Left) Scatter plot and regression parameters; (right) histogram of the prediction error in nT.

Not only that: even when the alignment along the bisector tends to get lost (i.e. when the "disturbances" become less intense) we find that the error does not differ from the interval ± 20 nT that we identified with the average prediction error (cfr. Fig. 7).

Regarding the behavior of the network during the inspected quiet days (when the Dst absolute values are ≲ 20 nT), we note that, although the regression seems to lose its meaning, the tendency to overestimate seems more prominent and the RMSE is comparable (but always

higher) to the target standard deviation, in general the response of the network falls within the range of uncertainty mentioned above for Training and Validation (i.e. the errors distribution are centered on small values); in one case (30 April, Fig. 9C) the distribution is not centered around zero, rather we find an unbalanced distribution towards overestimated values. These results bring to consider that the obtained ANN does not show a good generalization capability for quiet geomagnetic cases. Nonetheless, the analysis puts in evidence how the neural network model seems more reliable in monitoring a geomagnetic disturbed situation.

As we reported in the Introduction section, Balasis et al. (2019) derived a Swarm-based Dst index. The authors showed that during three periods in 2015 (precisely, in March, June and December) when geomagnetic storms occurred, the Swarm-derived Dst index was able to follow the standard ground-based Dst with a correlation greater than 0.9 (Balasis et al., 2019). In order to test the performance of our results with Swarm-based indices computed by other methodologies, we considered the Dst derived by Balasis et al. (2019).

From our dataset spanning from 2014 to 2019, we selected the same three months of the 2015 analysed by Balasis et al. (2019), i.e. March, June and December. Then, we computed the correlation coefficient of the results of the neural network (Dst_{NN}) with the standard ground-based Dst. Please, note that due to the way we selected our dataset, discontinuity in our time series was unavoidable.

Table 4. compares our results with those found by Balasis et al. (2019).

It is noteworthy that the results of the present study show a very good compatibility with the other Swarm-based index by Balasis et al. (2019), despite the discontinuity in the data series.

4. Discussion

We expected that the ANN model we obtained will retrieve, in particular, good estimation of the "disturbed" Dst values: when we look at the three scatter plots (and regression curves) in Fig. 8 (the three unseen disturbed days in 2020) we could appreciate the ANN capability to give good response to large external contributions, concluding that our neural network model seems more reliable in monitoring a geomagnetic disturbed situation.

But we are also interested in better understanding the causes of both the wide dispersion of the outputs at Dst

values less than −100 nT (more evident in Fig. 6, where the output of the Validation dataset is represented), and the spread of the results for small Dst values ("quiet hours"). Indeed, although we can see there is a quite good response of the network, as the output values distribute along the bisector with a relatively small spreading, nonetheless this dispersion increases around −50 nT and rapidly gets very noticeable from Dst values less than −100 nT: in particular, we note an asymmetry with the predominance of overestimated outputs.

To look for an explanation, this outperformance behavior was analyzed more in detail for the more numerous Validation dataset. First, we inspect the relationships between the results and the ANN input variables, by considering separated datasets, as depicted in Fig. 10. There, we considered three distinct groups: the first one comprises those outputs whose differences (Δ) with true values are between + 20 and −20 nT (light blue), independently of the target Dst. Note that this interval is comparable to the width of the distribution around zero of the prediction error, so it is reasonable to consider this group as "good results" when compared with the other two groups. These latter are selected starting from the values of the target Dst less than −100 nT, with an absolute difference |Δ| from the true value greater than 20 nT: they are further distinguished according to whether they are overestimated (dark red) or underestimated (red).

To understand the origin of even large deviations from the true values, we focus on the input data of the three datasets and their specific characteristics and mutual differences. In Fig. 11, we have represented, for comparison, the normalized distributions for the Central, High (the overestimated) and Low (the underestimated) input data points.

In general, the distributions of the residual Y and Z magnetic components and of the magnetic spatial coordinates MLat and MLon (Fig. 11B, C, D and E, respectively) are compatible with each other. Curiously, a significant difference for the three data groups emerges when inspecting the distributions of both the residual component X (Fig. 11A) and the magnetic local time MLT (Fig. 11F).

Indeed, by looking at Fig. 11A, we notice that the centroids of the distributions of the residual X corresponding to the overestimated and underestimated data are largely shifted towards the increasingly negative values if compared to the central group: almost certainly this indicates that the residual X component strongly drives the results, but we cannot exclude the action

Table 4
Correlation coefficients between the Dst and the DstNN, the network output, in the periods of March, June and December 2015. These months practically correspond to those analysed by Balasis et al. (2019), periods affected by strong geomagnetic storms. It is noteworthy both the good correlation for all time intervals and the good agreement with the previous result by Balasis et al. (2019).

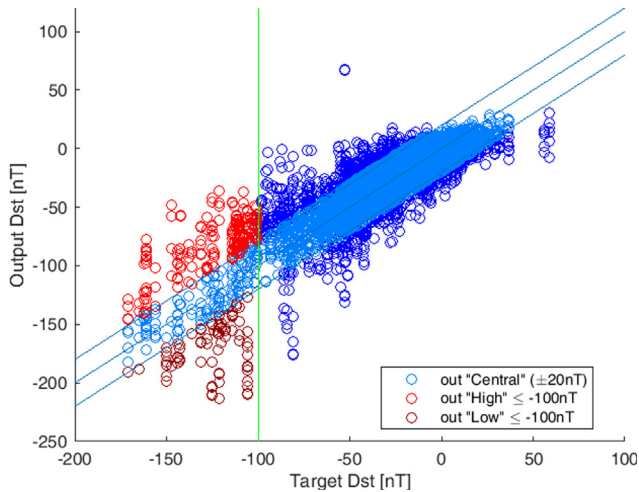|	Dst_{NN} vs Dst			Swarm-based Dst (Balasis et al. (2019))
Timespan	March 2015	June 2015	December 2015	2015
Correlation	0.945	0.930	0.918	≳ 0.90

Fig. 10. Scatter plot target vs model considering three distinct subsets: the first includes those values which are close to the actual corresponding values, i.e. the difference Δ lays within the ± 20nT interval (light blue); the other subsets comprise the values higher than −100 nT (very disturbed conditions) indicated with the green vertical line. In turn, they are divided depending on whether their differences with the actual value (Δ) are greater (red) or lower (dark red) than ± 20 nT, respectively, for further analyses and considerations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of other causes, such as (among others) the uncertainties accompanying the use of the model CHAOS-7.3 used to obtain the residuals. In fact, we also notice that many data points of the "central" residual X components are associated with very low values of Dst and yet the corresponding distribution maintains not far from zero, i.e. does not go towards low values. For what concerns the relationship of the three inspected datasets with MLT in Fig. 11F, an evident asymmetry emerges for the two High and Low groups of dispersed data: the concentration of their distributions around 8–10 and 22–24 MLT. We may hypothesize that this behaviour is related to the effects of the onset and termination of the large ionospheric currents and photochemical processes driven by the direct exposure to sunlight.

In light of the result in Fig. 11A, we first moved further to inspect the correlation between the residual geomagnetic X component, Dst target and output, respectively for the three dataset described above, a relation depicted in Fig. 12, Fig. 13 and Fig. 14.

Fig. 12 shows the scatter plot between both the actual and modelled Dst, and the residual X component for the first dataset (the "good" one): we notice that they show a good correlation, although the model is characterized by a slightly minor spreading, maybe due to low noise in the input signal or to a smoothing capability of the ANN.

When inspecting the same correlation with the residual X component of the two other datasets in Fig. 13 and Fig. 14 (the over- and underestimated groups), we notice a greater dispersion for the actual than for the modelled Dst: the model reveals a high correlation

with X component of geomagnetic field and it seems as if the network carries out a sort of signal filtering, as internal noise removing. Since the residual X component depends only on the external component of the geomagnetic field because of the internal components removal, it seems that the neural model behaves more coherently with variation of magnetic field induced from the external sources. We believe that this aspect deserves more attention and much analysis in the future for a better comprehension of the input data and of the results of the network.

After analyzing the relationships between the residual X components and the Dst (input and output) and having also noted that the points of the three datasets are distributed uniformly in space (Fig. 15), we decided to turn our attention to the possible mutual dependence of those two quantities on the magnetic local time MLT, i.e. the other input parameter whose distributions for the three groups (Fig. 11F) showed an evident differentiation.

Fig. 16 and Fig. 17 show the dependence on MLT by the residual X component and target Dst, respectively. They both show the presence of bunches corresponding to particular MLT: 7–8; 10–11; 19–20 and 22–24. Looking at Fig. 17, we observe that the clusters are characterized by both central and under-overestimated network estimates. Therefore, the performance of the ANNs does not depend on MLT.

Since the introduction of Dst, more data are available and now we know that complex and rapid interactions and large particle fluxes take place during the diverse phases of a magnetic storm. That is why some authors question the adequacy of Dst hourly index to represent the ring current effect and its rapid variations (e.g. Campbell, 2004), although some others found that compared to the 1-minute SYM-H index, Dst may be used as the low-resolution version of that (Wanliss and Showalter, 2006).

Therefore, when evaluating the output of the network, we need to bear in mind that many difficulties may arise. First, we must consider that complex behavior marks both the solar activity and the indices introduced to describe their geomagnetic effects, and the hasty variations in their characteristics make it difficult to model and predict them (Mirmomeni et al., 2010). The procedure of determining both the baseline and the average of the four values of Dst computed by each observatory, together with the removal of Solar quiet (Sq) currents and of the asymmetric part of the ring current, may introduce a smoothing of the rapid change of the external conditions: added to the fact that the given value is hourly, this can determine a certain degree of inadequacy of this index to follow the rapid variations which, on the contrary, are accurately measured by the satellite instruments: we argue that the residual magnetic components feeding the ANNs are certainly affected by the enhancement of the ring current during the magnetic storms, as evidenced by the scatter plot between Dst and the residual X component in Fig. 13 and Fig. 14; but we also believe that the hourly Dst index is not able to describe the effects of other different ionospheric phenomena which,
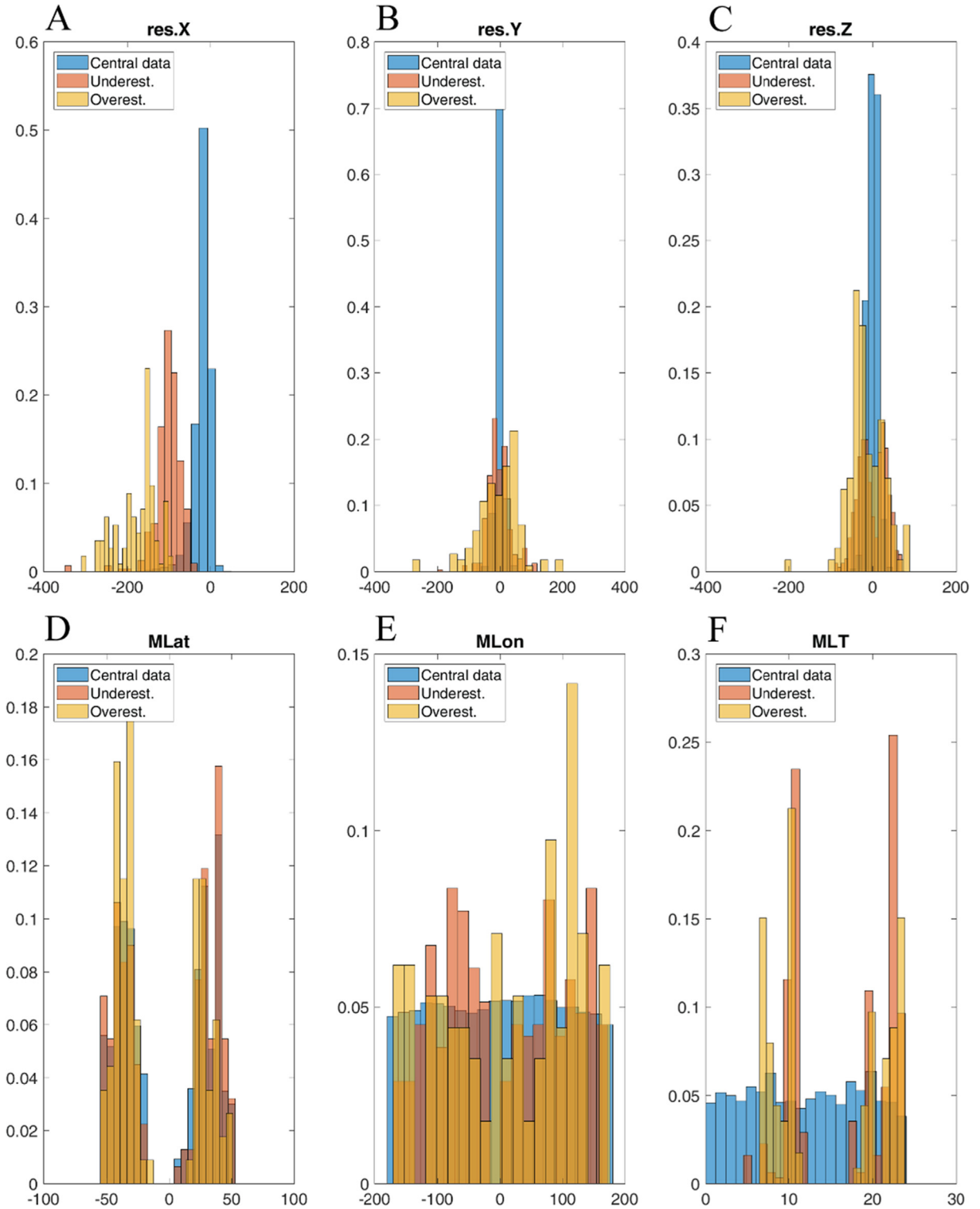
Fig. 11. Histograms of the inputs: the three residual magnetic components (X, Y and Z) and of the three magnetic time–space coordinates: magnetic latitude (MLat), longitude (MLon) and local time (MLT) corresponding to the three groups of data points (named Central, High and Low) described in Fig. 10. We notice that the distributions of the residual Y and Z components are mutually compatible, meaning that we should not expect they give origin to the large dispersion. That is not the case of the residual X component: they differ consistently and this difference may be at the base of the large dispersion for large negative values of Dst. For what concerns the space–time coordinates, we notice that their distributions are mutually compatible, meaning that we should not expect they are the origin of the large dispersion, except for MLT (see main text for more details).
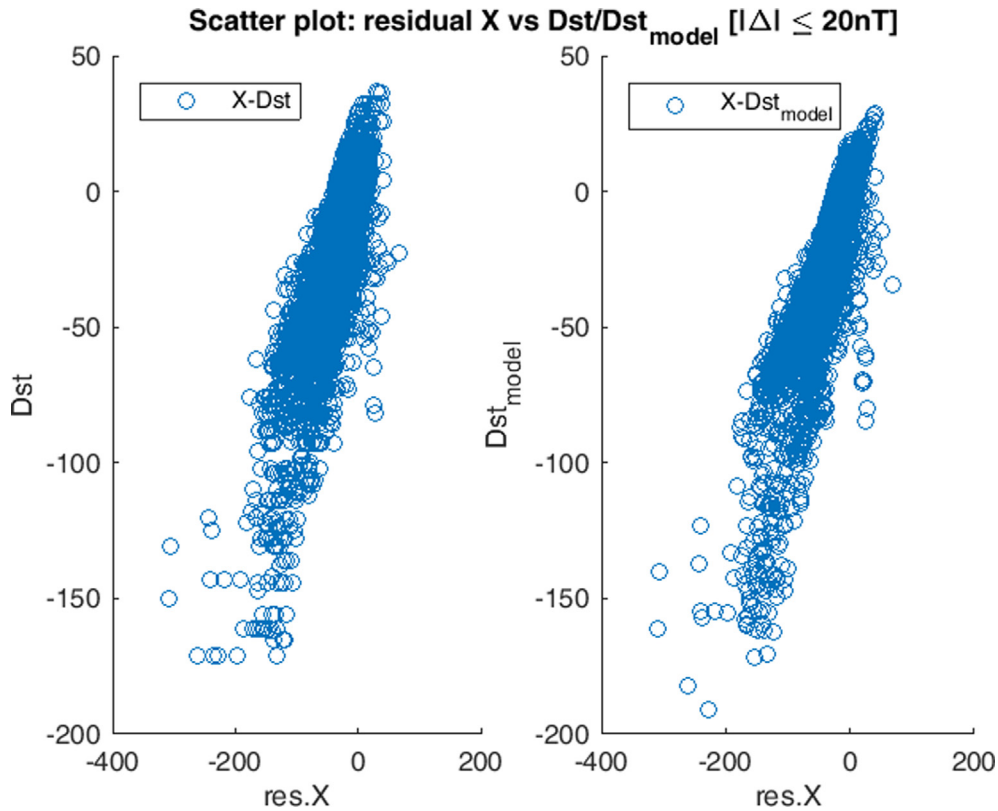
Fig. 12. Scatter plot between residual X component and the Dst (target and modelled) for the "Central" (light blue) values on Fig. 10. It is remarkable that they show almost the same behavior. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

acting on small spatial and time scale, may superpose to the larger spatial and time scale of the ring current, giving origin to both the dispersion in the retrieved Dst and the wrong estimation.

Another element to be considered is the model used for the removal of the internal contribution of the field, supposed to be well represented in space and almost constant in the time interval of data selection. Nonetheless, each model carries a certain degree of uncertainty. Therefore, it is possible, for example, that the discrepancies between the modelled values and the real ones may have the effect of feeding the dispersion of the network results.

5. Conclusions

We explored the possibility to instruct an ANN to provide a very quick estimation of the Dst geomagnetic index, on the basis of the near real time magnetic measurements collected continuously by the ESA Swarm constellation of satellites.

The proposed method represents a reliable new approach for Dst index retrieval and due to its fast application in the operational stage it also demonstrates its usefulness in near real time ionospheric irregularities monitoring.

However, we realized that for the ANN to be effective it is necessary to pre-treat the data so to remove the internal field contributions (core and crustal) by means of the

CHAOS-7.3 model, resulting in residual components certainly affected by the external sources; as well as to take into account the different local hours explored by the satellite, which of course can lead to different effects on the measurements.

Since the study has highlighted the imperfect correlation of Dst with the residual X component of the geomagnetic field in some "extreme" situations, for the future there are other further steps we can take thanks to the experience we got with this work: it may be appropriate to consider Dst samples consistent with the residual X component of the geomagnetic field, to obtain a more performing retrieval/forecast; but we think also it is worth considering a high-resolution index as the SYM-H, because it should have the ability to better follow the rapid variations we expect to emerge in magnetic data.

It would be worth considering the possibility of using the geomagnetic index F10.7, based on solar flux, as a more suitable output of the neural model.

Furthermore, we envisage the possibility to increase the input dataset by including Swarm B satellite magnetic data: differently from Swarm A (which is practically parallel and very close in space and time to Swarm C), it flies at a higher altitude with different orbital parameters. Thus, simultaneously with Charlie, Bravo explores regions of the ionosphere that differ in local time and magnetic conditions, enriching the diversity of the dataset.
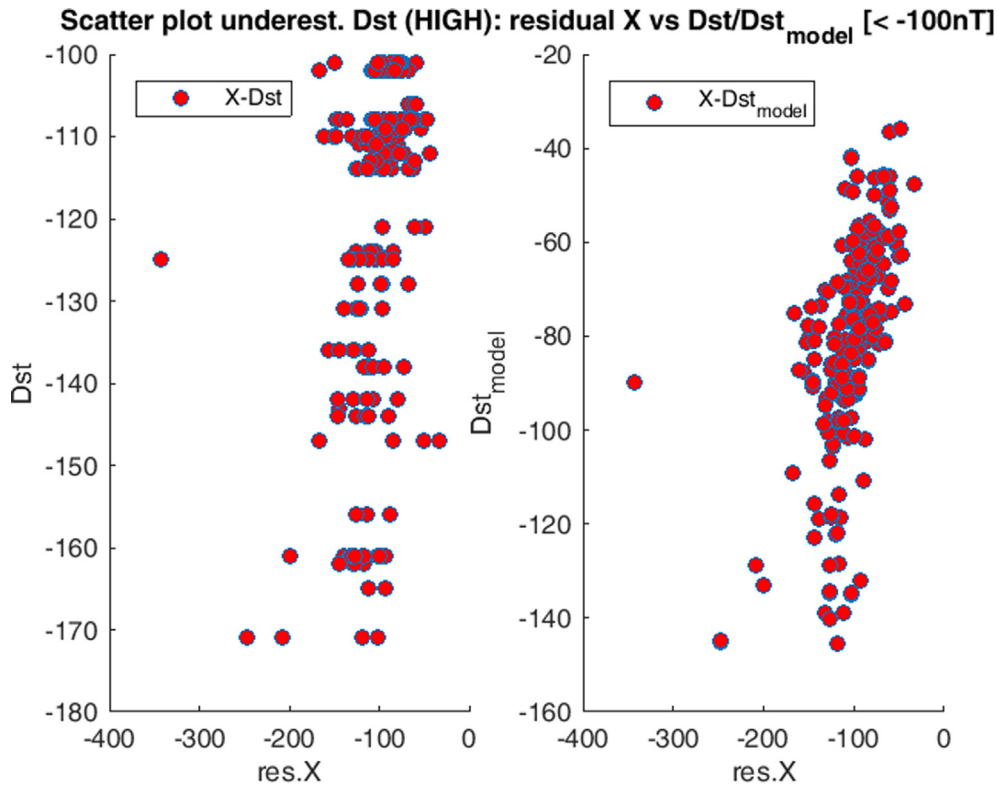
Fig. 13. Scatter plot between the residual X component and Dst (target and modelled) for the values indicated as "High" (red) on Fig. 10. Their comparison puts in evidence a higher correlation between the residual X component and the modelled Dst than the actual one. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
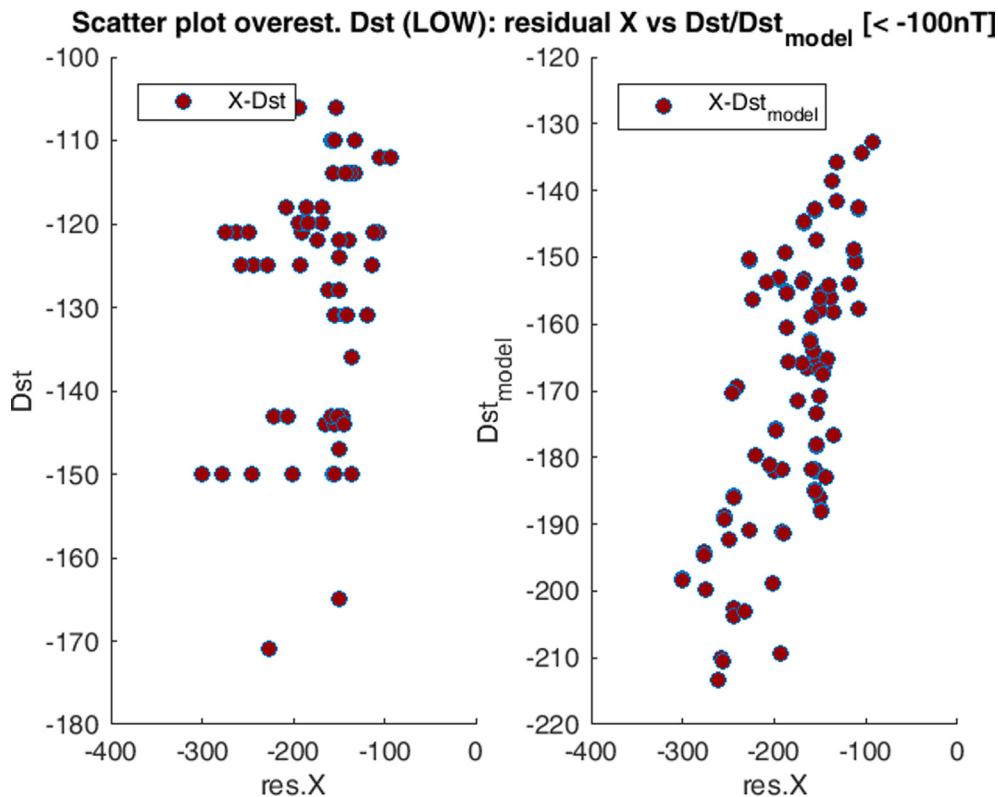

Fig. 14. Scatter plot between the residual X component and Dst (target and modelled) for the values indicated as "Low" (dark red) values on Fig. 10. Their comparison puts in evidence a higher correlation between the residual X component and the modelled Dst than the actual one. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
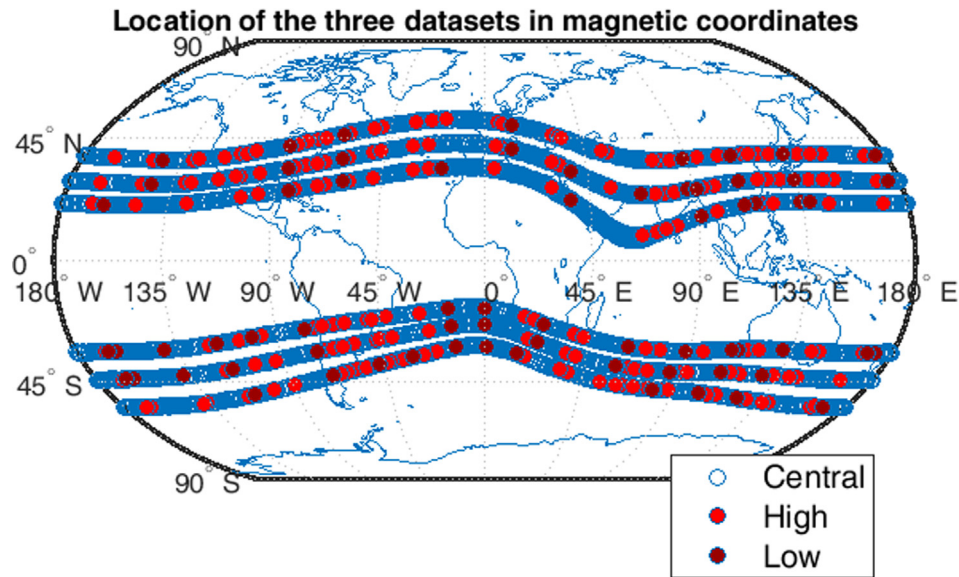
Fig. 15. In magnetic coordinates, the location of data belonging to the datasets named Central (considered as "good"), High and Low, these latter resulting in under- and overestimated Dst values. It is evident that data are distributed uniformly.
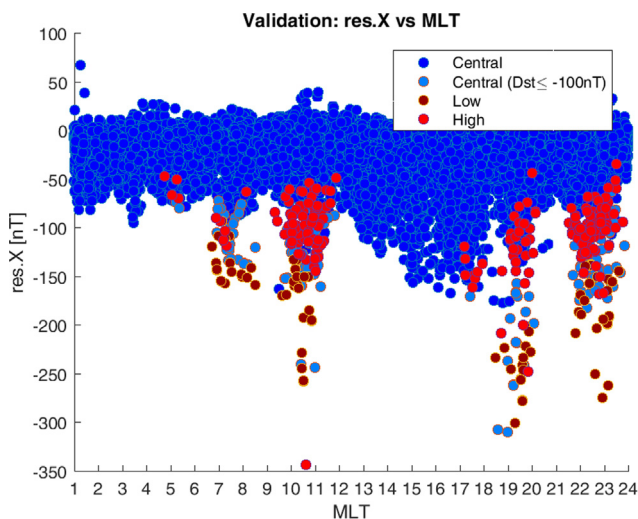

Fig. 16. Distribution of input residual X corresponding to the three groups datasets in function on MLT, selected for Dst values below −100 nT.
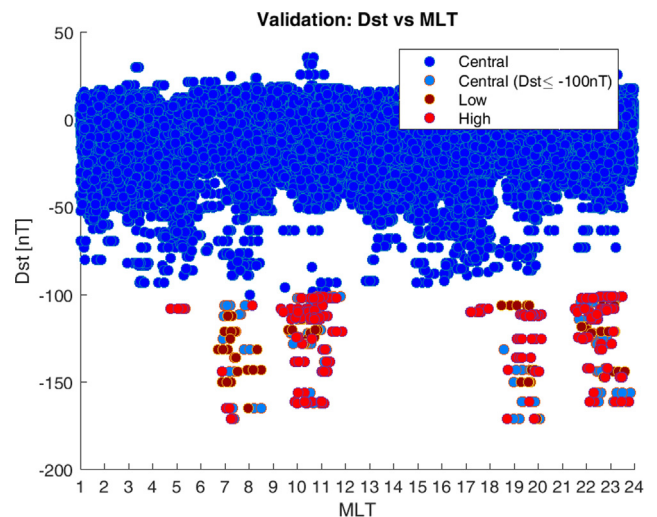

Fig. 17. Distribution of target Dst corresponding to the three groups datasets in function on MLT, selected for Dst values below −100 nT.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Atkinson, P.M., Tatnall, A.R.L., 1997. Introduction Neural networks in remote sensing. Int. J. Remote Sens. 18 (4), 699–709. https://doi.org/10.1080/014311697218700.

Ayala Solares, J.R., Wei, H.-L., Boynton, R.J., Walker, S.N., Billings, S. A., 2016. Modeling and prediction of global magnetic disturbance in near-Earth space: A case study for Kp index using NARX models. Space Weather 14, 899–916. https://doi.org/10.1002/2016SW001463.

Bala, R., Reiff, P., 2012. Improvements in short-term forecasting of geomagnetic activity. Space Weather 10, S06001. https://doi.org/10.1029/2012SW000779.

Balasis, G., Papadimitriou, C., Boutsi, A.Z., 2019. Ionospheric response to solar and interplanetary disturbances: a Swarm perspective. Phil. Trans. R. Soc. A 377, 20180098.

Benediktsson, J.A., Swain, P.H., Ersoy, O.K., 1993. Conjugate gradient neural networks in classification of multisource and very high dimensional remote sensing data. Int. J. Remote Sens. 14, 2883–2903.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford.

Blackwell, W.J., Chen, F.W., 2009. Neural Networks in Atmospheric Remote Sensing. Artech House, Norwood, MA, USA.

Butler, C.T., Meredith, R.V.Z., Stogryn, A.P., 1996. Retrieving atmospheric temperature parameters from DMSP SSM/T-1 data with a neural network. J. Geophys. Res. 101 (D3), 7075–7083.

Cabrera-Mercader, C.R., Staelin, D.H., 1995. Passive microwave relative humidity retrievals using feed forward neural networks. IEEE Trans. Geosci. Remote Sens. 33, 1324–1328.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Contr. Sign. Syst. 2, 303–314. https://doi.org/10.1007/BF02551274.

Efitorov, A.O., Myagkova, I.N., Shirokii, V.R., et al., 2018. The Prediction of the Dst-Index Based on Machine Learning Methods. Cosm. Res. 56, 434–441.

Finlay, C.C., Kloss, C., Olsen, N., et al., 2020. The CHAOS-7 geomagnetic field model and observed changes in the South Atlantic Anomaly. Earth Planets Space 72, 156. https://doi.org/10.1186/s40623-020-01252-9.

Friis-Christensen, E., Lühr, H., Hulot, G., 2006. Swarm: A constellation to study the Earth's magnetic field. Earth Planets Space 58, 351–358. https://doi.org/10.1186/BF03351933.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. Atmos. Environ. 32, 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R. Springer, New York.

Hecht-Nielsen. Neurocomputing. Addison-Wesley, Reading, MA, 1990.

Hsieh, W.W., Tang, B., 1998. Applying neural network models to prediction and data analysis in meteorology and oceanography. Bull. Am. Meteorol. Soc. 79 (1), 1855–1870.

Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks 3 (5), 551–560. https://doi.org/10.1016/0893-6080(90)90005-6.

Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Netw. 4 (2), 251–257.

Iapaolo, M., Godin-Beekman, S., Del Frate, F., Casadio, S., Petitdidier, M., McDermid, I.S., Leblanc, T., Swart, D., Meijer, Y., Hansen, G., Stebel, K., 2007. GOME ozone profiles retrieved by neural network techniques: a global validation with lidar measurements. J. Quant. Spectrosc. Radiat. Transfer 107 (1), 105–119. https://doi.org/10.1016/j.jqsrt.2007.02.015.

Krasnopolsky, V.M., Breaker, L.C., Gemmill, W.H., 1995. A neural network as a nonlinear transfer function model for retrieving surface wind speeds from the Special Sensor Microwave Imager. J. Geophys. Res. 100 (C6), 11033–11045. https://doi.org/10.1029/95JC00857.

Kratsios, A., 2021. The Universal Approximation Property. Ann Math Artif Intell. https://doi.org/10.1007/s10472-020-09723-1.

Laundal, K.M., Richmond, A.D., 2017. Magnetic Coordinate Systems. Space Sci. Rev. 206, 27–59. https://doi.org/10.1007/s11214-016-0275-y.

Lek, S., Guegan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecol. Model. 44 (1), 1259–2126.

Lu, J.Y., Peng, Y.X., Wang, M., Gu, S.J., Zhao, M.X., 2016. Support Vector Machine combined with Distance Correlation learning for Dst forecasting during intense geomagnetic storms. Planet. Space Sci. 120, 48–55. https://doi.org/10.1016/j.pss.2015.11.004.

Mas, J.F., Flores, J.J., 2008. The application of artificial neural networks to the analysis of remotely sensed data. Int. J. Remote Sens. 29, 617–663.

MATLAB, 2020. version 9.9.0.1538559 (R2020b) Update 3. The Math-Works Inc., Natick, Massachusetts.

Mirmomeni, M., Lucas, C., Moshiri, M., Araabi, B.N., 2010. Introducing adaptive neurofuzzy modeling with online learning method for prediction of time-varying solar and geomagnetic activity indices. Expert Syst. Appl. 37 (12), 8267–8277. https://doi.org/10.1016/j.eswa.2010.05.059.

Myagkova, I., Shiroky, V. and Dolenko, S.. Prediction of geomagnetic indexes with the help of artificial neural networks. E3S Web Conf. Volume 20, 2017 VIII International Conference "Solar-Terrestrial Relations and Physics of Earthquake Precursors". Doi: https://doi.org/10.1051/e3sconf/20172002011.

Myagkova, I.N., Shirokii, V.R., Kalegaev, V.V., et al., 2021. The Use of Coupling Functions in the Forecasting of the Dst-Index Amplitude with Adaptive Methods. Geomagn. Aeron. 61, 138–147. https://doi.org/10.1134/S0016793220060092.

Müller, M.D. et al., 2003. Ozone profile retrieval from Global Ozone Monitoring Experiment (GOME) data using a neural network approach (Neural Network Ozone Retrieval System (NNORSY)). J. Geophys. Res. 108 (D16), 4497. https://doi.org/10.1029/2002JD002784.

Nilsson, Nils J., 1996. Introduction to Machine Learning - An Early draft of a proposed textbook, Robotics Laboratory Department of Computer Science. Stanford University, Stanford, CA 94305, USA.

Picchiani, M., Chini, M., Corradini, S., Merucci, L., Sellitto, P., Del Frate, F., Stramondo, S., Piscini, A. (2011). Volcanic ash retrieval from IR multispectral measurements by means of Neural Networks: an analysis of the Eyjafjallajokull eruption, Proc. IEEE IGARSS 2011, Vancouver (Canada), July 24-29.

Piscini, A. and Lombardo, V. (2014a). Volcanic hot spot detection from optical multispectral remote sensing data using artificial neural networks, Geophys. J. Int., 196 (3), 1525-1535, DOI:10.1093/gji/ggt506, 2014.

Piscini, A., Picchiani, M., Chini, M., Corradini, S., Merucci, L., Del Frate, F., Stramondo, S. (2014b). A Neural Network approach for the simultaneous retrieval of volcanic ash parameters and SO2 using MODIS data, Atmos. Meas. Tech. Discuss., 7, 3349–3395, doi:10.5194/amtd-7-3349-2014.

Piscini, A., Carboni, E., Del Frate, F., Grainger, R.G., 2014. Simultaneous retrieval of volcanic sulphur dioxide and plume height from hyperspectral data using artificial neural networks. Geophys. J. Int. 198 (2), 697–709. https://doi.org/10.1093/gji/ggu152.

Piscini, A., Romaniello, V., Bignami, C., Stramondo, S., 2017. A New Damage Assessment Method by Means of Neural Network and Multi-Sensor Satellite Data. Applied Sciences 7 (8), 781. https://doi.org/10.3390/app7080781.

Principe, J.C., Euliano, N.R., Lefebvre, W.C., 2000. Neural and adaptive systems: fundamentals through simulations. Wiley, New York.

Rosenblatt, R., 1962. Principles of Neurodynamics. Spartan Books.

Rumelhart, D.E., Hinton, G.E. and Williams,R.J. (1986). Learning internal representation by error propagation. In McClleland J.L. Rumelhart, D.E., editor, Parallel Distributed Processing: Exploration in the Microstructure of Cognition, volume i, chapter 8, pages 318–362. MIT Press.

Rumelhart, D.E., Durbin, R., Golden, R. and Chauvin Y. (1995). Backpropagation: the basic theory. In Yves C. Rumelhart, D.E., editor, Backpropagation: Theory, Architecture, and Applications, pages 1–34. Lawrence Erlbaum, NJ, 1995.

Schalkoff, R., 1992. Pattern Recognition: Statistical. Wiley, Structural and Neural Approaches.

Shepherd, S.G., 2014. Altitude-adjusted corrected geomagnetic coordinates: Definition and functional approximations. J. Geophys. Res. Space Physics 119, 7501–7521. https://doi.org/10.1002/2014JA020264.

Sugiura, M. and Kamei, T., 1991. Equatorial Dst index 1957–1986, IAGA Bull., 40. by A. Berthelier and M. Menville (Int. Serv. Geomagn. Indices Publ. Off., Saint Maur, 1991).

Tan, Y., Hu, Q., Wang, Z., Zhong, Q., 2017. Geomagnetic index Kp forecasting with LSTM. Space Weather 16, 406–416. https://doi.org/10.1002/2017SW001764.

Tasistro-Hart, A., Grayver, A., Kuvshinov, A., 2021. Probabilistic geomagnetic storm forecasting via deep learning. J. Geophysical Res.: Space Phys. 126. https://doi.org/10.1029/2020JA028228, e2020JA028228.

Temerin, M., Li, X., 2015. The *Dst* index underestimates the solar cycle variation of geomagnetic activity. J. Geophys. Res. Space Phys. 120 (7), 5603–5607. https://doi.org/10.1002/2015JA021467.

Twomey, J.M., Smith, A.E., 1995. Performance Measures, Consistency, and Power for Artificial Neural Network Models. Mathl. Comput. Modelling 21 (1/2).

Wanliss, J.A., 2004. Nonlinear variability of SYM-H over two solar cycles. Earth Planet Sp 56, e13–e16. https://doi.org/10.1186/BF03352507.

Wanliss, J.A., Showalter, K.M., 2006. Highresolution global storm index: Dst versus SYM-H. J. Geophys. Res. 111, A02202. https://doi.org/10.1029/2005JA011034.

Wei, H.L., Zhu, D.Q., Billings, S.A., Balikhin, M.A., 2007. Forecasting the geomagnetic activity of the Dst index using multiscale radial basis function networks. Adv. Space Res. 40 (12), 1863–1870. https://doi.org/10.1016/j.asr.2007.02.080.

Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2017. Deep learning. Data Mining 417–466. https://doi.org/10.1016/b978-0-12-804291-5.00.