NUMERICAL EXPERIMENTS ABOUT FORECAST OF THE STRONG GEOMAGNETIC DISTURBANCES

ASIMOPOLOS Laurențiu, ASIMOPOLOS Natalia-Silvia,
BALEA Bogdan, ASIMOPOLOS Adrian-Aristide

Abstract. The presented numerical experiments belong to three different methodological categories, with the same purpose, but with different approaches. The common goal is the prediction of geomagnetic disturbances and the methodologies used comparatively are Fourier spectral deconvolution, autoregressive models on time series and recurrent Long Short Term Memory (LSTM) neural networks that are capable of long-term dependence. The data used in these experiments are the geomagnetic data recorded at the Surlari Geomagnetic Observatory as well as other planetary observatories located at different latitudes and longitudes. In the experiments we also used planetary geomagnetic indices, downloaded from specialized sites.

Keywords: neural networks, Fourier spectral deconvolution, autoregressive models, geomagnetic disturbances.

Rezumat. Experimente numerice privind prognoza perturbațiilor geomagnetice puternice. Experimentele numerice prezentate fac parte din trei categorii metodologice diferite, avand același scop, dar cu abordări diferite. Scopul comun îl reprezintă prognoza perturbațiilor geomagnetice, iar metodologiile folosite comparativ sunt deconvoluția spectrală Fourier, modelele autoregresive pentru serii de timp și rețelele neuronale recurente Long Short Term Memory (LSTM) care sunt capabile să înregistreze dependențe pe termen lung. Datele folosite în aceste experimente sunt datele geomagnetice înregistrate la Observatorul Geomagnetic Surlari cât și la alte observatoare planetare situate la latitudini și longitudini diferite. In cadrul experimentelor am folosit și indici geomagnetici planetari, descărcați de pe site-urile de specialitate.

Cuvinte cheie: rețele neuronale, deconvoluție spectrală Fourier, modele autoregresive, perturbații geomagnetice.

INTRODUCTION

A time series is a sequence of geomagnetic observations taken sequentially in time. Many sets of data appear as time series in such fields as economics, business, engineering, the natural sciences (especially geophysics and meteorology), and social sciences. An intrinsic feature of a time series is that, typically, adjacent observations are dependent. The nature of this dependence among observations of a time series is of considerable practical interest. Time series analysis is concerned with techniques for the analysis of this dependence. This requires the development of stochastic and dynamic models for time series data and the use of such models in important areas of application (BISGAARD & KULAHCI, 2011; BOX et al., 2016; JENKINDS et al., 2016). In the field of geomagnetism, papers have been published on numerical developments of time series for data acquired in the Surlari Geomagnetic Observatory (ASIMOPOLOS et al., 2011; 2012; ASIMOPOLOS & ASIMOPOLOS, 2018). The data used in this paper come from the most prestigious geomagnetic databases (***. http://www.intermagnet.org; ***. http://www.noaa.gov; ***. https://www.spaceweatherlive.com)

For the analysis methodology in MATLAB we can list KIM, 2017 and ***. https://www.mathworks.com.

TIME SERIES OF GEOMAGNETIC FIELD

Time series analysis requires the development of stochastic and dynamic models for time series data and the use of such models in some application. This analysis means:

- The forecasting of future values of a time series from current and past values.
- The determination of the transfer function between series of inputs data and effect on the series of output data of a system.
- The use of indicator input variables in transfer function models to represent and assess the effects of unusual intervention events on the behaviour of a time series.
- The examination of interrelationships among several related time series variables of interest and determination of appropriate multivariate dynamic models to represent these joint relationships among the variables over time.
- The design of simple control schemes by means of which potential deviations of the system output from a desired target may, as far as possible, be compensated by adjustment of the input series values.

Geomagnetic observations and geomagnetic indices are available at discrete, equispaced intervals of time and constitute input data.

To calculate the best forecasts, it is necessary to specify their accuracy, so that, for example, the risks associated with decisions based upon the forecasts may be calculated. The accuracy of the forecasts may be expressed by calculating probability limits on either side of each forecast.

Methods for estimating transfer function models based on deterministic perturbations of the input, such as sinusoidal changes of magnetic field, have not always been successful. This is due to the fact that, for perturbations of a geomagnetic variations, that are relevant and tolerable, the response of the system may be masked by uncontrollable disturbances such as geomagnetic storms.

For many problems in physical and environmental sciences, such as geomagnetic field, time series data may be available on several related variables of interest. A more informative and effective analysis is often possible by considering individual series as components of a multivariate or vector time series and analysing the series jointly. For k-related time series variables of interest in a dynamic system, we may denote the series as z_{1t}, z_{2t},…, z_{kt}, and let $\boldsymbol{Z}_t = (z_{1t},\ldots, z_{kt})'$ denote the $k \times 1$ time series vector at time t.

Methods of multivariate time series analysis are used to study the dynamic relationships among the several time series that comprise the vector $\boldsymbol{Z}t$. This involves the development of statistical models and methods of analysis that adequately describe the interrelationships among the series. Two main purposes for analysing and modelling the vector of time series jointly are to gain an understanding of the dynamic relationships over time among the series and to improve accuracy of forecasts for individual series by utilizing the additional information available from the related series in the forecasts for each series. No phenomenon is totally deterministic, because unknown factors with fast variations can occur such as Sun activity, variable solar wind velocity, the Earth's Magnetosphere conditions and so on, that can throw a missile slightly off course. In many issues, we have to consider a time-dependent phenomenon, such as cyclic geomagnetic variations, in which there are many unknown factors and for which it is not possible to write a deterministic model that allows exact calculation of the future behaviour of the phenomenon. Nevertheless, it may be possible to derive a model that can be used to calculate the probability of a future value lying between two specified limits. Such a model is called a probability model or a stochastic model. The models for time series that are needed, for example, to achieve optimal forecasting and control, are in fact stochastic models. It can make a difference between the probability model or stochastic process and the time series. Thus, a time series z_1, z_2,…, z_N of N successive observations is regarded as a sample realization from an infinite population of such time series that could have been generated by the stochastic process.

An important class of stochastic models for describing time series, comprises stationary models. Stationary models assume that the process remains in statistical equilibrium with probabilistic properties that do not change over time, in particular varying about a fixed constant mean level and with constant variance. However, forecasting has importance in Earth sciences (such as geomagnetism), where many time series are often better represented as nonstationary and, in particular, as having no natural constant mean level over time.

MACHINE LEARNING AND LONG SHORT-TERM MEMORY

Machine Learning is the technique used to find (or learn) a model from the data. It is suitable for problems that involve intelligence, such as image recognition and speech recognition, where physical laws or mathematical equations fail to produce a model. Once the Machine Learning process finds the model from the training data, we apply the model to actual field data. This process is illustrated in Fig. 1. The vertical flow of the figure indicates the learning process, and the trained model is described as the horizontal flow, which is called inference.

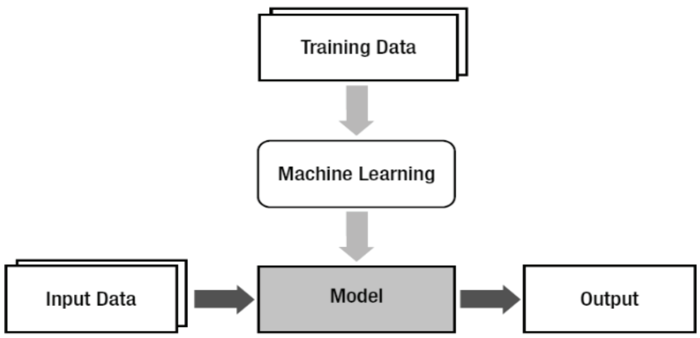

Figure 1. Applying a model based on field data.

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network that is capable of long-term dependencies.

Conceptually, a recurring LSTM unit tries to “remember” all the past knowledge the network has seen so far and to “forget” the irrelevant data. This is done by introducing different layers with activation functions called “gates” for different purposes. Each recurrent LSTM unit also maintains a vector called an “internal cell state” that conceptually describes the information that was chosen to be retained by the previous recurrent LSTM unit. An LSTM network comprises four different gates for different purposes, as described below (Fig. 2):

- Forget gate: determines the extent to which previous data can be forgotten.
- Input gate: determines the information to be written in the internal cell state.
- Input modulation gate: it is considered a part of the input gate and is used to modulate the information that the input gate will write on the internal cell state internally adding nonlinearity to the information and normalizing the information. This is done to reduce learning time, ensuring faster convergence. Although the actions of this gate are less important than the others and are often treated as a concept that offers finesse, it is good practice to include this in the structure of the LSTM unit.

• Output gate: determines which output (next hidden layer) is generated from the internal state of the LSTM unit.

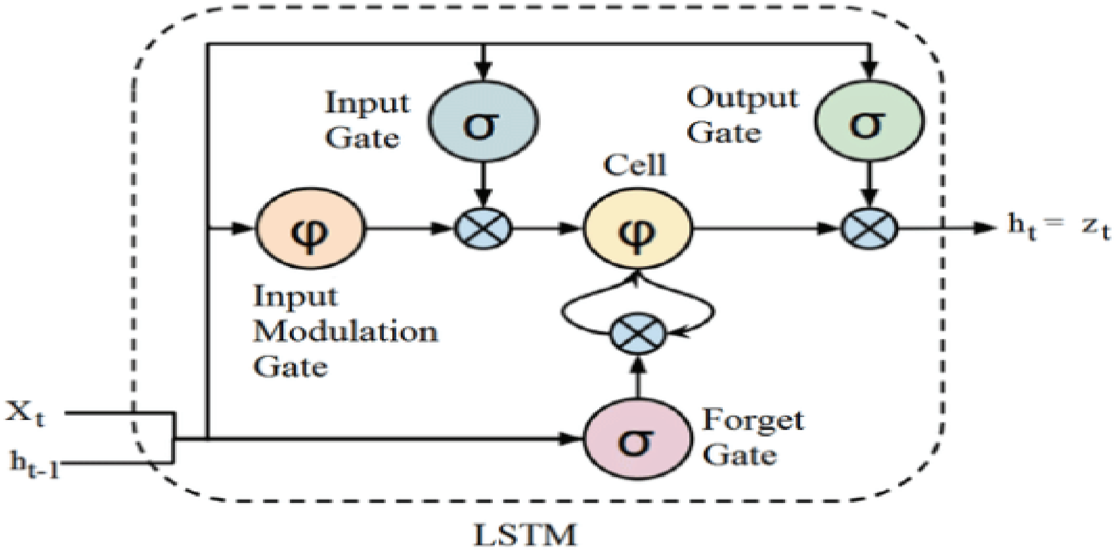

Figure 2. The architecture of an LSTM cell.

STAGES OF IMPLEMENTING THE SOLUTION OF LSTM

The implementation of the solution for predicting certain geomagnetic parameters is implemented in the MATLAB language, using the Deep Learning Toolbox. It provides a framework for the design and implementation of Deep Learning networks, with the development of convolutional neural networks (ConvNets, CNN) and recurrent LSTM networks for the classification or regression of image, time series and text data.

Also, in addition to using the MATLAB environment, the solution can be accessed, modified or improved in the Jupyter Notebook computing environment. The stages of implementing the solution of LSTM are:

1. The data can be read as follows:

```
format short
omni_data = readtable('fisier.txt');
omni_data.Properties.VariableNames = {'Var1' 'Var2' 'Var3' 'Var4' 'Var5' 'Var6' 'Var7' 'Var8'};
data = table2cell(omni_data(:,8));
data = [data{:}];
```

2. Training data partitioning - 70% and testing - 30%:

```
numTimeStepsTrain = floor(0.7*numel(data));
dataTrain = data(1:numTimeStepsTrain+1);
dataTest = data(numTimeStepsTrain+1:end);
```

3. Data normalization

```
mu = mean(dataTrain);
sig = std(dataTrain);
dataTrainStandardized = (dataTrain - mu) / sig;
```

4. Preparation of training data

```
XTrain = dataTrainStandardized(1:end-1);
YTrain = dataTrainStandardized(2:end);
```

5. Defining the Long Short Term Memory network

```
numFeatures = 1;
numResponses = 1;
numHiddenUnits = 150;
layers = [ ...
   sequenceInputLayer(numFeatures)
   lstmLayer(numHiddenUnits)
   fullyConnectedLayer(numResponses)
   regressionLayer];
options = trainingOptions('adam', ...
   'MaxEpochs',150, ...
   'GradientThreshold',1, ...
   'InitialLearnRate',0.005, ...
   'LearnRateSchedule','piecewise', ...
```

```
    'LearnRateDropPeriod',125, ...
    'LearnRateDropFactor',0.2, ...
    'Verbose',0, ...
    'Plots','training-progress');
```

6. Training of the LSTM network

```
net = trainNetwork(XTrain,YTrain,layers,options);
```

7. Normalization of test data

```
dataTestStandardized = (dataTest - mu) / sig;
XTest = dataTestStandardized(1:end-1);
```

8. Effective prediction of the variable selected for analysis, based on the observed data

```
net = resetState(net);
net = predictAndUpdateState(net,XTrain);
YPred = [];
numTimeStepsTest = numel(XTest);
for i = 1:numTimeStepsTest
    [net,YPred(:,i)] = predictAndUpdateState(net,XTest(:,i),'ExecutionEnvironment','gpu');
End
```

9. Calculation of the square error

```
%unstandardize data
YPred = sig*YPred + mu;
YTest = dataTest(2:end);
rmse = sqrt(mean((YPred-YTest).^2))
```

10. Display of observed data (YTest) and forecasted data (YPred)

```
figure
plot(dataTrain(1:end-1))
hold on
idx = numTimeStepsTrain:(numTimeStepsTrain+numTimeStepsTest);
plot(idx,[data(numTimeStepsTrain) YPred],'.-')
hold off
ylabel("Values")
title("Forecast")
legend(["Observed" "Forecast"])
```

Geomagnetic signals are the convolution product of the atomic stationary signal mono-frequential of different amplitudes associated to phenomena with a very broad band of periodicities and nondeterministic signals associated with geomagnetic disturbances and non-periodic phenomena. Among analysis processes used for discrete series of geomagnetic data with different lengths and sampling rates, we can state that the moving average works as a low pass filter in frequency or high pass in time. By eliminating high frequency components (depending on mobile window size used) preferential periodicities greater than a given value can be studied. Signal linearization provides information on the linear trend of the entire analysed series. Thus, for very long data series, we extracted the variation slope for each geomagnetic component, separately. The numeric derivative of signal versus time proved to be a very reliable indicator for geomagnetic disturbed periods. Thus, the derivative value may be increased by several orders of magnitude during periods of agitation in comparison to calm periods. Time-frequency analyses allow to identify the frequency characteristics of the signal at a time. For this, a mobile window has to be chosen, moving along the signal from time t_0 to any position t_i, on temporal axis. The frequency content of each window is analysed, finally obtaining the frequency spectrum well localized in time. The advantage of the wavelet transformations versus the Fourier transform is the possibility to analyse discrete data sets that have some gaps or irregular variations, as the geomagnetic data.

For the experimental part, we used an LSTM model to predict the DST values 1h, 2h, and 3h ahead, and a Gaussian Process classifier to predict the level of the geomagnetic storms. The training dataset has four months of hourly DST values, starting from March 1st, 2001 until June 30th, 2001. This period includes the geomagnetic storm event that occurred between March 31st – 1st April, as shown in Fig. 3, which also shows spectral and wavelet analysis. We also analysed the period December 7-26, 2006 through both the Fourier spectrum, wavelet (Fig. 4), GP (Fig. 5) and LSTM (Fig. 6). In our numerical experiments on the forecast of strong geomagnetic disturbances we also used spectral analysis and wavelet analysis (Figs. 3 and 4). Fourier and wavelets technics allow local analysis of magnetic field components through variable frequency windows and different functions. Windows with longer time intervals allow us to extract low-frequency information, average times of windows lead to extraction of medium frequency information, and very narrow windows highlight the high frequencies or details of the analysed signals. The model of the disturbed geomagnetic field (through DST index) is composed of periodic oscillations plus non-periodic oscillations given by the impact of solar wind on the terrestrial magnetosphere. The purpose of wavelet analysis is to build orthonormal bases composed of wavelets that can reconstruct the geomagnetic parameters.

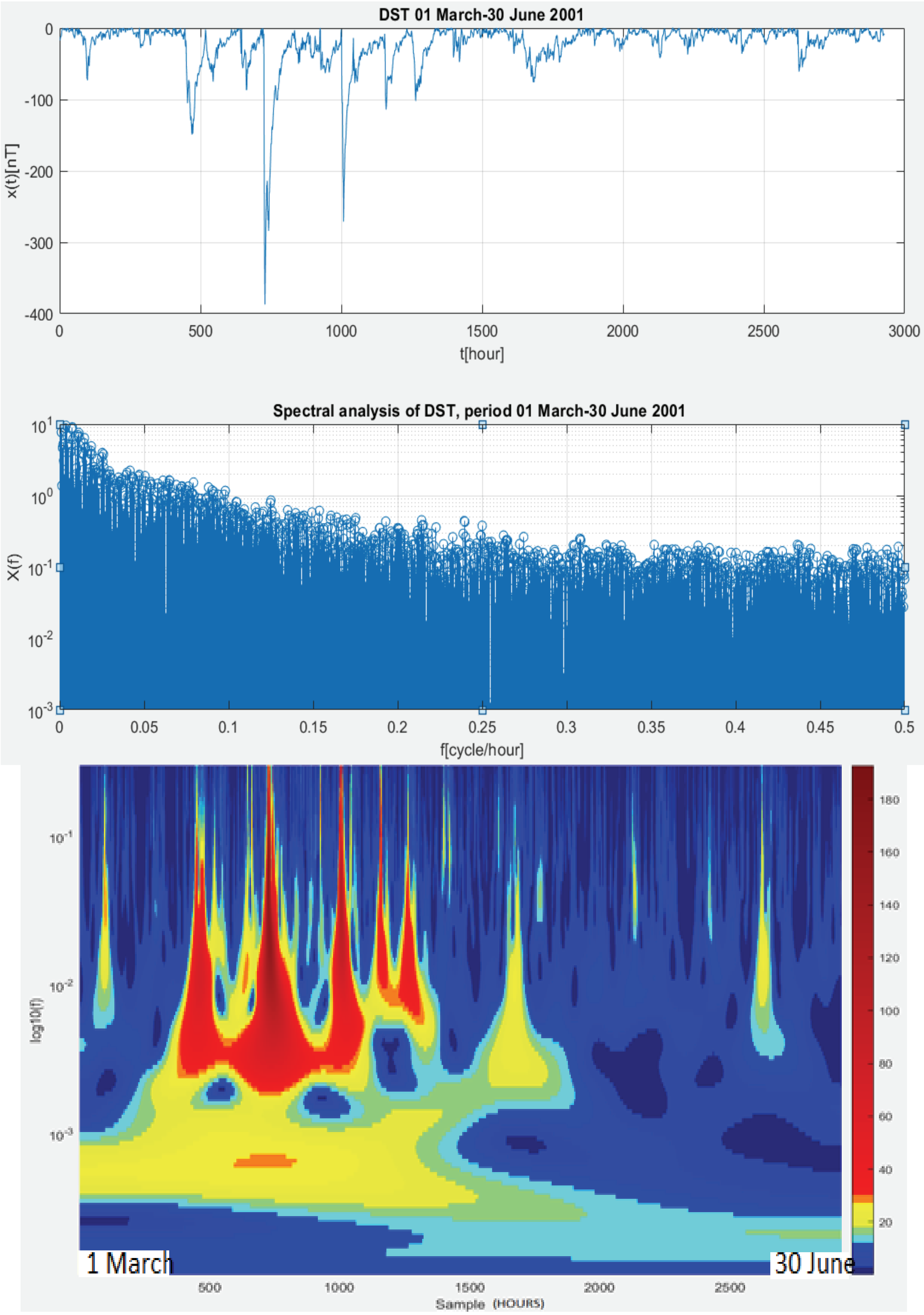

Figure 3. Training data, 1st March 2001 – 30th June 2001 (up), spectral analysis (middle), wavelet analysis (down).

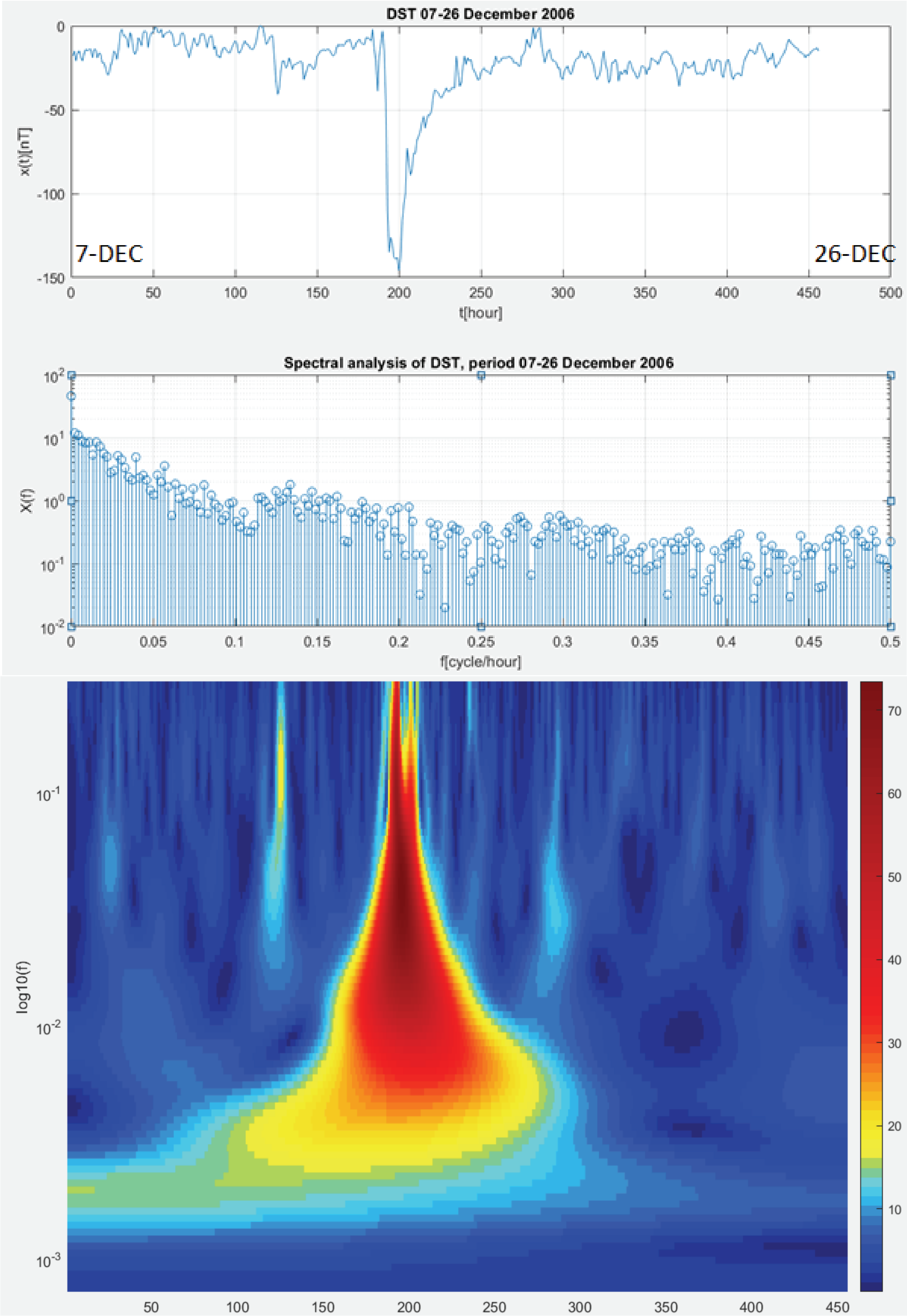

Figure 4. Training data, 7-26 December 2006 (up), spectral analysis (middle), wavelet analysis (down).

The Gaussian Process multiclass classifier is used for geomagnetic storm prediction depending on the level of activity. For instance, for Dst < -250 nT there is a super geomagnetic storm, for -50 nT < Dst < -250 nT there is an intense storm, and for Dst > -50 nT a moderate storm. For a 3h ahead prediction, we obtained a 0.98 ROC AUC, as presented in Fig. 5.

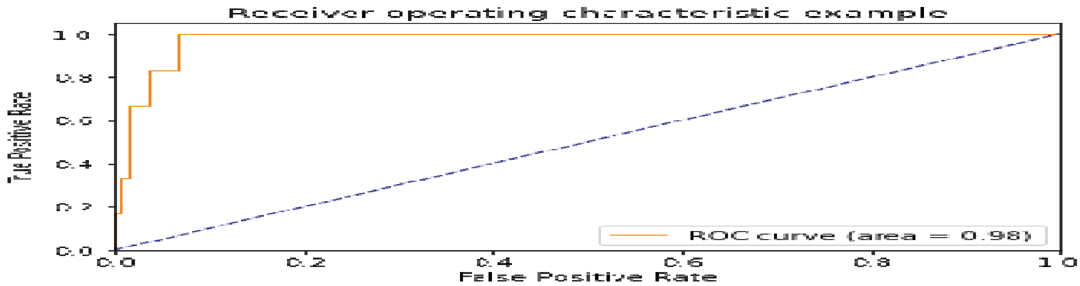

Figure 5. ROC curve of the Gaussian Process classifier.

The geomagnetic activity predicted by the LSTM model (3h ahead) is presented in Fig. 6.

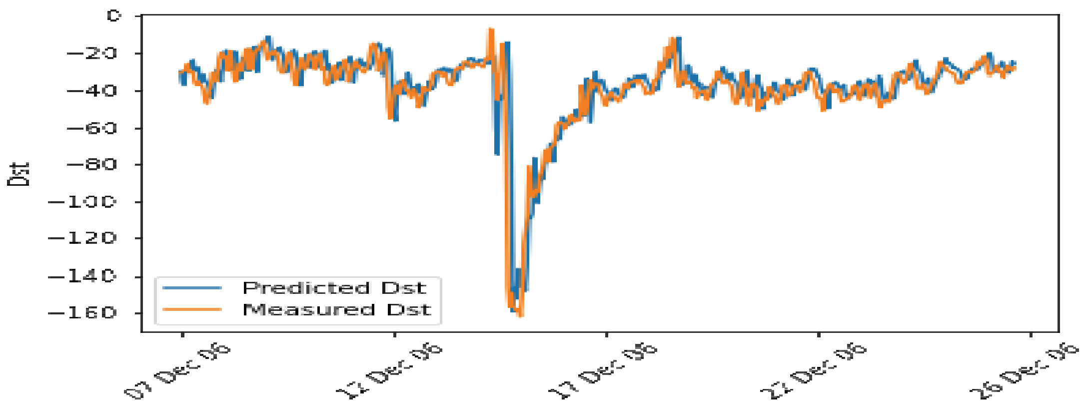

Figure 6. The 3h ahead estimation of the Dst for geomagnetic storm, between 7th December 2006 – 26th December 2006.

CONCLUSIONS

Forecasting time series data is an important topic in many domains, inclusive in geomagnetism. Gaussian Processes (GP) are a method designed to solve regression and probabilistic classification problems. The prediction interpolating the observations is probabilistic (Gaussian) so that one can compute empirical confidence intervals and decide, based on them, if one should refit (online fitting, adaptive fitting) the prediction in some region of interest.

Also, different kernels can be specified and provided, but it is also possible to specify custom kernels. GP can lose efficiency in high dimensional spaces – namely when the number of features exceeds a few tens. Traditionally, there are several techniques to effectively Auto Regressive Integrated Moving Average (ARIMA) with its many variations, that has demonstrated its outperformance in precision and accuracy of predicting the next lags of time series.

Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) that enables support for time series and sequence data in a network. LSTM performs additive interactions, which can help improve gradient flow over long sequences during training. LSTM are best suited for learning long-term dependencies.

We used hourly mean of the DST index to represent the axially symmetric disturbance magnetic field at the dipole equator on the Earth's surface (data from http://www.noaa.gov and http://www.noaa.gov), for major disturbances in Dst which are negative, namely decreases in the geomagnetic field. These are produced mainly by the ring equatorial current system in the magnetosphere. Positive DST indices are caused by the compression of the Earth's magnetosphere through the increase of solar wind pressure.

ACKNOWLEDGEMENT

We address our thanks to the Romanian Ministry of Education and Research for financing of the projects:

The realization of 3D geological / geophysical models for the characterization of some areas of economic and scientific interest in Romania, with Contract no. 49N / 2019.

Institutional capacities and services for research, monitoring and forecasting of risks in extra-atmospheric space, acronym SAFESPACE, project No. 16PCCDI/2018, within PNCDIII.

REFERENCES

ASIMOPOLOS L., NICULICI E., SĂNDULESCU A. M., ASIMOPOLOS N. S. 2011. Comparisons of geomagnetic data measured in Romania with the data of International Geomagnetic Reference Field 11 (IGRF 11). *Surveying Geology & Mining Ecology Management*. Albena. **2**: 25-32.

ASIMOPOLOS L., SĂNDULESCU A. M., ASIMOPOLOS N. S., NICULICI E. 2012. *Analysis of data from Surlari National Geomagnetic Observatory*. Editura Ars Docendi. București. 96 pp.

ASIMOPOLOS N. S. & ASIMOPOLOS L. 2018. Study on the high-intensity geomagnetic storm from march 2015, based on terrestrial and satellite data, Micro and Nano Tehnologies & Space Tehnologies & Planetary Science. *Surveying Geology & Mining Ecology Management*. Albena. **18**: 593-600.

BOX G. E. P., JENKINS G. M., REINSEL G. C., LJUNG G. M. 2016. *Time series analysis - Forecasting and Control, Fifth Edition*. John Wiley & Sons. Hoboken. 709 pp.

BISGAARD S. & KULAHCI M. 2011. *Time series analysis and forecasting by example*. John Wiley & Sons. Hoboken, 382 pp.

KIM P. 2017. *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*. Apress. New York. 162 pp.

***. http://www.intermagnet.org (accessed January 21, 2020).

***. http://www.noaa.gov (accessed January 28, 2020).

***. https://www.spaceweatherlive.com (accessed February 03, 2020).

***. https://www.mathworks.com (accessed February 14, 2020).

Asimopolos Laurențiu
Geological Institute of Romania
1st Caransebeş Street, 012271 - Bucharest, Romania.
E-mail: laurentiu.asimopolos@igr.ro, asimopolos@gmail.com

Asimopolos Natalia-Silvia
Geological Institute of Romania
1st Caransebeş Street, 012271 - Bucharest, Romania.
E-mail: natalia.asimopolos@igr.ro, asi_nata@yahoo.com

Balea Bogdan-Valeriu
Geological Institute of Romania
1st Caransebeş Street, 012271 - Bucharest, Romania.
E-mail: natalia.asimopolos@igr.ro, asi_nata@yahoo.com

Asimopolos Adrian Aristide
University POLITEHNICA of Bucharest
Faculty of Transports, 313 Splaiul Independentei, 060042 - Bucharest, Romania.
E-mail: adrian.asimopolos@gmail.com

Received: March 28, 2020
Accepted: August 12, 2020