

Regression analysis of ionospheric disturbance factors

Yuryi Polozov*, Oksana Mandrikova

Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS, 684034, Kamchatsky krai, Paratunka, Russian Federation

Abstract. Investigation of interactions of the near-planet space parameters, Earth magnetic field and ionospheric parameters are of interest in the tasks of solar-terrestrial physics and applied researches related to space weather. Ionosphere is one of the important factors of space weather. Functioning of modern ground- and satellite-based engineering facilities depends much on its state. The paper makes a statistical estimate and analyzes complex effect from interplanetary magnetic field parameters, solar radiation and geomagnetic data on ionospheric disturbance development. Ionospheric disturbances were estimated on the basis of the method developed by the authors. The method applies wavelet transform and adaptive thresholds. The most significant factors of ionospheric disturbance occurrences were detected during the investigation and their significance was evaluated.

1 Introduction

Investigation of interplanetary environment parameter impact on the Earth's magnetosphere and ionosphere state is of interest in the tasks of solar-terrestrial physics and applied researches associated with space weather. Changes on the Sun, in interplanetary field parameters, in the Earth's magnetosphere and ionosphere make significant effect on engineering system operation and human activity. Ionosphere is one of the natural factors from which the functioning of communication means and ground- and space-based technical facilities depend directly. Solar activity, interplanetary magnetic field changes, geomagnetic disturbances etc. make significant impact on ionosphere parameter (for example, foF2) changes. As such factors, authors (for, example, [1-3]) use interplanetary magnetic field parameters (total $|B|$ and Bz-component), solar flux radio radiation (f10.7), auroral indices of geomagnetic activity (AE), planetary index of geomagnetic activity (Kp) and Dst-index applied as an estimate of the level of geomagnetic disturbances caused by the Sun.

In the paper we apply regression analysis to investigate statistical impact of interplanetary environment parameters and processes in the magnetosphere on ionosphere state and ionospheric disturbance occurrences. To estimate ionosphere state, we use ionospheric critical frequency foF2 data and the characteristics of its disturbance degree, ionospheric disturbance intensity I , which was obtained on the basis of wavelet transform [4-7].

* Corresponding author: up_agent@mail.ru

2 Applied data and their preliminary processing

In the paper we applied foF2 hourly and 15-minute data which were recorded at «Paratunka» site (Kamchatskiy kray, Russia, site coordinates: 53.0 N, 158.7 E). Hourly data of foF2 have been obtained by ionograms manual processing at IKIR FEB RAS since 1969. Records of 15-minute data have been carried out since August 2015 in automatic mode by the means of an ionospheric station. Due to the presence of gaps in the ionospheric data, which are the result of technical and natural factors, we used foF2 time series which did not have gaps for more than a day. Gaps of the length of 24 hours and less were filled in by the median method.

Hourly values of geomagnetic activity AE- and DST-indices (<http://wdc.kugi.kyoto-u.ac.jp>), diurnal values of solar radio radiation f10.7 (<https://omniweb.gsfc.nasa.gov>), 3-hour data of geomagnetic field planetary Kp-index (<http://isgi.unistra.fr>) and hourly data of interplanetary magnetic field |B| and Bz (<https://cdaweb.gsfc.nasa.gov>) were used in regression analysis. Gaps in IMF data were filled in by the nearest-neighbor method [8]. If there was a gap with the duration of 8 hours and more within a day or there was a continuous gap from 4 hours and more, an event was removed from the analysis. The period under analysis is from August 2015 till September 2018. When constructing regression models, we applied the data for the period satisfying the following conditions:

1. IMF Bz-component values reached -8 nT;
2. One-day total value of Kp-index was > 18.

To estimate ionospheric disturbance intensity I , wavelet transform of foF2 data was performed

$$d_{\eta,n} = |\eta|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \Psi\left(\frac{t-n}{\eta}\right) dt, f \in L^2(R), \eta, n \in R, \eta \neq 0 \quad (1)$$

Then we used adaptive threshold functions

$$d_{\eta,n} = \begin{cases} d_{\eta,n}^+, & \text{if } (d_{\eta,n} - d_{\eta,n}^{med}) \geq T_{\eta}^{ad} \\ d_{\eta,n}^-, & \text{if } (d_{\eta,n} - d_{\eta,n}^{med}) \leq -T_{\eta}^{ad} \end{cases}, \quad (2)$$

where $T_{\eta}^{ad} = V * S_{\eta}^t$, the value $S_{\eta}^t = \sqrt{\frac{I}{\Phi - I} \sum_{n=1}^{\Phi} (d_{\eta,n} - \overline{d_{\eta,n}})^2}$, $\overline{d_{\eta,n}}$ and $d_{\eta,n}^{med}$ are the average and the median, respectively. They are calculated within a moving time window of length Φ taking into account the diurnal variation.

The ionospheric disturbance intensity I at time $t = n$ was estimated

$$I_n^{\pm} = \sum_{\eta} |d_{\eta,n}|. \quad (3)$$

In what follows we show an example of estimation of ionospheric disturbance intensity during a moderate magnetic storm on October 13, 2016. Fig. 1 illustrates the graphs of IMF parameters, magnetic indices and ionospheric parameters for the period October 12-16, 2016. According to space weather data (<http://ipg.geospace.ru>), IMF southern component fluctuations were intensified from the end of the day on October 12 till the end of the day on October 14 in the result of arrival of accelerated fluxes from two coronal ejections from a coronal hole. During the magnetic storm under analysis, IMF |B| reached the value of 24.2 nT, IMF Bz took the value of -20 nT (Fig. 1 c,d). During the main phase of the storm, auroral activity significantly increased, AE-index reached the value of 1200 nT (Fig. 1 g). Fig. 1 b illustrates the results of calculation of ionospheric disturbance intensity. Red color shows

intensity positive values I_n^+ (characterize electron concentration increase), blue color shows intensity negative values I_n^- (characterize electron concentration decrease). The results (Fig. 1 b) show anomalous increase in electron concentration before the event. It occurred during the growth of geomagnetic activity indices (Fig. 1 e) and reached the maximum 7 hours before the storm commencement. The ionospheric anomalous increase lasted for about 21 hours. During the magnetic storm recovery phase, electron concentration anomalously decreased. It is likely to be associated with the heating of the thermosphere lower part in the auroral region during the magnetic storm main phase and air transfer towards the equator [1]. The negative anomaly lasted for about 26 hours. Calculation results for 27-day median (blue in Fig. 1 a is foF2 values, green is foF2 median values) agree with these results and show deviations of foF2 values from the corresponding median values before and during the magnetic storm. We should note that estimation results for disturbance intensity I_n^\pm are more illustrative, compared to the median method, and allow us to detect accurately the ionospheric anomalous periods.

As the example shows, disturbances in the ionosphere may last up to tens of hours and more. In their turn, the factors triggering ionospheric disturbances have time shift. For example, it is known [3] that IMF Bz turn to the south occurs several hours or more before ionospheric disturbance commencement. During the analysis we should also take into account that the duration of anomalous changes in interplanetary environment parameters and that of ionospheric disturbances caused by these factors, is different.

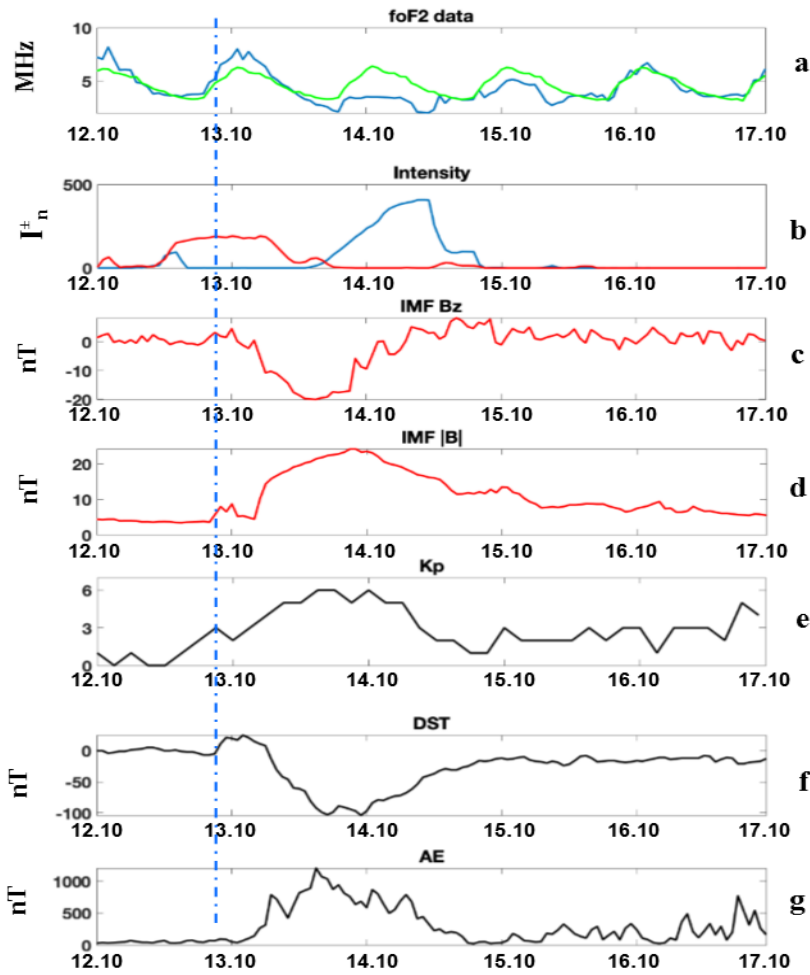


Fig. 1. Geophysical parameter data for the period October 12-16, 2016. Blue color indicates the magnetic storm commencement.

3 Investigation results

3.1 Estimation of correlation coefficients

Correlation coefficient is estimated as

$$r = \frac{\sum_i (x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (z_i - \bar{z})^2}}$$

where $\bar{x} = \frac{1}{n} \sum_i x_i$, $\bar{z} = \frac{1}{n} \sum_i z_i$.

To estimate the significance of the obtained correlation coefficients, we used Student's *t*-criterion [9]. Results of the estimates are presented in Table 1. Analysis of the data from Table 1 shows the presence of linear statistically significant relation between the estimated values and confirms the fact of significant impact of IMF Bz parameters on ionospheric state and ionospheric disturbance occurrences during magnetic storms. In spite of the small values of the estimated correlation coefficients ($\approx 0,5$ for I_n^- values and $\approx 0,3$ for foF2 data), their statistical significance, according to Student's criterion, is very great. We should note that correlation coefficients obtained for I_n^- , exceed significantly the correlation coefficients for foF2 data. That confirms I_n^- consistency and its efficiency for estimation of ionospheric disturbance intensity during negative phases of ionospheric storms.

Table 1. Correlation coefficients of IMF Bz parameters and foF2 data.

No	Correlation coefficient r	t-statistics	Period length (days)	Number of periods	foF2 data	IMF Bz parameter
1	0.44	12.19	3	606	I_n^- maximum	maximum range (negative values)
2	0.46	-12.91	3	606	I_n^- maximum	achieved minimum
3	0.46	14.46	2	781	I_n^- maximum	maximum range (negative values)
4	0.47	-15.15	2	781	I_n^- maximum	achieved minimum
5	0.28	-7.18	3	602	fof2 minimum	maximum range (negative values)
6	0.29	7.49	3	602	foF2 minimum	achieved minimum
7	0.26	-7.66	2	756	foF2 minimum	maximum range (negative values)
8	0.27	7.97	2	756	foF2 minimum	achieved minimum

3.2 Development and analysis of regression models

In the paper we estimated the models of the type

$$Y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b,$$

where Y is the value of I_n^- , x_i are the factors affecting foF2, a_i, b are estimated parameters of the model.

To estimate the statistical significance of the obtained models we applied the following quantities:

- determination coefficient

$$R^2 = 1 - \frac{\sum_i (z_i - \hat{z}_i)^2}{\sum_i (z_i - \bar{z}_i)^2}$$

where $\bar{z} = \frac{1}{n} \sum_i z_i$, z_i, \hat{z}_i are real and calculated values of the explanatory variable;

- F-statistics

$$F = \frac{R^2 / n}{(1 - R^2) / (m - n - 1)}$$

where m is the number of explanatory variables, n is the number of observations;

- Student's distribution to estimate regression coefficient significance

$$t = \frac{B}{Sb}$$

where B is the regression estimated coefficient, Sb is the standard error of regression coefficient.

The following models were obtained:

1. *Model of dual regression (data for the periods of 46 events were used):*

$$Y = -14.86x_{Bz} - 39.24$$

where x_{Bz} is the IMF Bz-component minimum value for a day.

The obtained value of model F-statistics $F = 24.82$, for the given level of significance $\alpha = 0,01$ (probability 99%), indicate the presence of statistically significant linear relation between IMF Bz and I_n^- value. The statistically significant linear relation between the values under the investigation is also confirmed by the obtained value of t-statistics, $t_{empir} = 4,98 > t_{crit} = 2,01$ (for the significance level $\alpha = 0.05$). Determination coefficient of the model $R^2 = 0.36$ indicates the necessity to include additional explanatory variables (factors).

2. *Model of multiple regression (data for the periods of 213 events were used):*

$$Y = -4.14x_{Bz} + 0.9x_{Kp} - 0.02x_{AE} + 0.52x_{f10.7} + 2.25x_{|B|} - 1.16x_{DST} - 54.72$$

where x_{Bz} is the IMF Bz-component minimum, x_{Kp} is the Kp-index maximum diurnal value, x_{AE} is the AE maximum for a day, $x_{f10.7}$ is the f10.7- index value for a day, $x_{|B|}$ is the IMF strength module maximum for a day, x_{DST} is the DST minimum for a day.

The obtained value of model F-statistics value $F = 21.864$, for the given level of significance $\alpha = 0,01$ (probability 99%), indicate the presence of linear relation between the

explanatory variables (factors) and I_n^- value. The obtained values of t-statistics $t_{Bz} = -2.03$, $t_{Kp} = 0.14$, $t_{AE} = -0.71$, $t_{f10.7} = 2.13$, $t_{|B|} = 1.61$, $t_{DST} = -4.25$

indicate the statistically significant influence of IMF Bz parameters, f10.7 values and DST data on I_n^- . Maximum diurnal values of geomagnetic activity Kp-indices, maximum diurnal values of auroral AE-indices and diurnal IMF |B| values, according to the given significance level $\alpha = 0.05$, do not have statistically significant effect. Model determination coefficient $R^2 = 0.389$ exceeded the determination coefficient of the model of dual regression (described above), but indicated that the variables (factors) included into the model explain a small part (about 40 %) from the total number of factors.

3. *Model of multiple regression (data for the periods of 213 events were used):*

$$Y = -6.14x_{Bz} + 0.53x_{f10.7} - 1.05x_{DST} - 49.06,$$

where x_{Bz} is the IMF Bz-component minimum diurnal value, $x_{f10.7}$ is the f10.7-index diurnal value, x_{DST} is the DST diurnal value.

The obtained value of model F-statistics value $F = 45.568$, for the given level of significance $\alpha = 0,01$ (probability 99%), indicate the presence of strong linear relation between the explanatory variables (factors) and I_n^- value. The obtained values of t-statistics indicate the statistically significant influence of all the explanatory variables used in the model, $t_{Bz} = -4.24$, $t_{f10.7} = 2.14$, $t_{DST} = -4.42$, on I_n^- . When we removed three factors (maximum diurnal values of geomagnetic activity Kp-indices, maximum diurnal values of auroral AE-indices and IMF |B| diurnal values, model 2) from the model, the determination coefficient reduced insignificantly, $R^2 = 0.379$.

4. *Model of multiple regression (data for the periods of 213 events were used):*

$$Y = -0.0022x_{Bz} + 0.0321x_{Kp} - 0.0002x_{AE} - 0.0025x_{f10.7} + 0.0013x_{|B|} + 0.016x_{DST} - 0.563,$$

where Y is the maximum diurnal value of $Z = f(i) - f^{med}(i)$ (difference between foF2 and corresponding median value for 27 days), x_{Bz} is the IMF Bz-component minimum diurnal value, x_{Kp} is the Kp-index maximum diurnal value, x_{AE} is the AE maximum diurnal value, $x_{f10.7}$ is the f10.7-index diurnal value, $x_{|B|}$ is the IMF strength module maximum diurnal value, x_{DST} is the DST minimum diurnal value.

Determination coefficient $R^2 = 0.28$, F-statistics $F = 13.394$ were obtained for this model. Values of t-statistics are: $t_{Bz} = -0.1$, $t_{Kp} = 0.46$, $t_{AE} = -0.98$, $t_{f10.7} = -0.95$, $t_{|B|} = 0.09$, $t_{DST} = 5.55$.

According to the estimate results, the variables (factors) included into the model explain a small part (about 28%) from the total number of factors affecting the value of deviation of foF2 values from corresponding median values. The obtained value of F-statistics is significantly lower than that of models 1-3, however, it exceeded the critical level ($F_{crit} = 3.1$, for the significance level of 0.01). From all the explanatory variables included into the model, only DST parameter confirmed its statistical significance.

The obtained results confirmed the statistical significance and consistency of I_n^- . Model analysis also shows that such factors as interplanetary magnetic field Bz-component, solar

radio radiation flux f10.7 values and DST-index, used as an estimate of geomagnetic disturbance level, have statistically significant impact on ionospheric disturbance degree during ionospheric storm negative phases.

4 Conclusions

Based on the estimates of correlation coefficients, we confirmed the linear statistically significant relation between IMF Bz parameters and the introduced measure of ionospheric disturbance intensity I_n^- . The obtained correlation coefficients confirm the efficiency of I_n^- in estimation of ionospheric disturbance intensity during ionospheric storm negative phases.

Based on the developed regression models, we have determined the factors having statistically significant impact on the appearance of anomalous changes in ionospheric parameters during magnetic storms. Statistically significant factors are: interplanetary magnetic field Bz-component, solar radio radiation flux f10.7 values and DST-index. It is insufficient to apply only one factor, for example IMF Bz, in the model. Results of regression analysis showed the consistency of I_n^- value to estimate ionospheric disturbance intensity.

The work was carried out according to the Subject AAAA-A17-117080110043-4 «Dynamics of physical processes in the active zones of near space and geospheres». The authors appreciate the help of the Institutes supporting ionospheric stations (<https://csu.ikir.ru/home>), providing the data of interplanetary magnetic field and solar wind, OMNI database (<https://omniweb.gsfc.nasa.gov/ow.html>), World Data Center for Geomagnetism, Kyoto (<http://wdc.kugi.kyoto-u.ac.jp>) and International Service of Geomagnetic Indices (<http://isgi.unistra.fr>).

References

1. A.D. Danilov, *Advances in Space Research*, **52**, 343–366 (2013)
2. E. Saiz, Y. Cerrato, C. Cid, V. Dobrica, P. Hejda, P. Nenovski, P. Stauning, J. Bochnicek, D. Danov, C. Demetrescu, W. D. Gonzalez, G. Maris, D. Teodosiev, F. Valach, *J. Space Weather Space Clim.* **3**, A26 (2013)
3. I. Tsagouri, A. Belehaki, *J. Space Weather Space Clim.*, **5**, A9 (2015)
4. O.V. Mandrikova, N.V. Fetisova, Y.A. Polozov, I.S. Solovev, M.S. Kupriyanov, *Earth, Planets and Space*, **67** (2015)
5. O. Mandrikova, Yu. Polozov, N. Fetisova, T. Zalyaev, *Journal of Atmospheric and Solar-Terrestrial Physics*, **181**, 116-126 (2018)
6. Yu. Polozov, N. Fetisova *Solar-Terrestrial Relations and Physics of Earthquakes Precursors. E3S Web of Conferences*, **127**, 01003 (2019)
7. O. Mandrikova, Yu. Polozov, B. Mandrikova, *Solar-Terrestrial Relations and Physics of Earthquakes Precursors. E3S Web of Conferences*, **127**, 02002 (2019)
8. R. J. A. Little, D.B. Rubin, *Statistical analysis with missing data* (Wiley-Interscience, N.-Y., 2002)
9. Student, *Biometrika*, **6 (1)** (1908)