Graphical models method - implementation to coupling processes in the atmosphere

Kateřina Podolská, Petra Koucká Knížová, Jaroslav Chum, Michal Kozubek, and Dalia Burešová
Institute of Atmospheric Physics CAS, Prague, Czech Republic, kapo@ufa.cas.cz

IAP

EGU2020-4805

Abstract

Internal atmospheric waves interact with themselves and/or with the undisturbed atmospheric flow, creating very complicated dynamical system with long-range dependencies. We suspect that regional character of the atmosphere at tropospheric heights may be crucial for explanation of the three different dependencies of foF2 on F10.7cm. We employ multivariate statistic methods applied to daily observational data which were obtained using mid-latitude ionosondes for the investigation of these relationships.

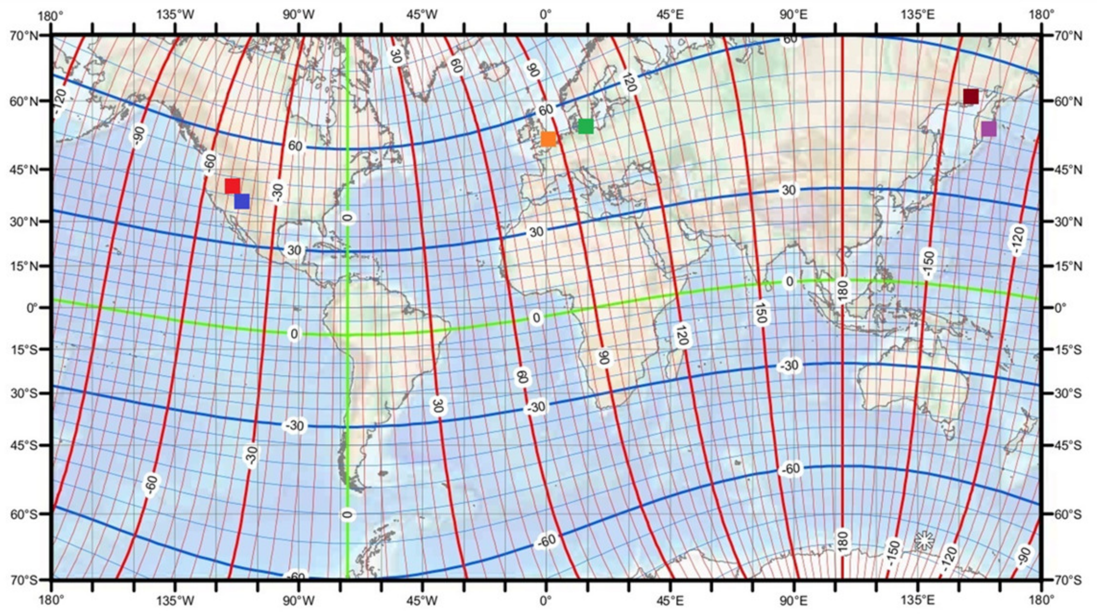
We consider specific and characteristic atmospheric wave generation that correspond to particular climatology of each location “European”, “American” and “Far East”. Specific conditions of each region, involving meteorological phenomena of the location as spectrum of atmospheric wave generation and their propagation. We consider significant difference in low atmosphere climatology as a key explanation of the three classes of ionospheric response to the F10.7cm on long time-scales and suggest that climatology of the troposphere must be taken into account for modelling of the ionospheric response.
Our aim is also to demonstrate that conditional independence graph (CIG) models, representing a robust method of multivariet statistical analysis, are useful for finding a relation between the ionosphere and space weather. This method appears to be more appropriate than correlation analysis between foF2 and geomagnetic and solar indices, especially for longitudinal data for which the characteristics may change over time or time series is interrupted. This method seems more effective to us than correlation analysis or scale analysis.

The final results of our analysis by CIG show that the dependence time shifts were clearly identified, namely +0 day shift in all cases, and +3 / +4 / +5 day shift in the dependence on the solar cycle phase and geographical longitude. Here, we would like to point out that we have analyzed data from ionospheric stations in a rather short span of latitudes, all stations belong to mid-latitudes (41.4° N – 54° N) and one would expect either practically same response/dependence of foF2 to F10.7 cm for longitudes and/or geomagnetic dependence as solar and geomagnetic forcing is considered as the most important. However, we have not identified any significant geomagnetically dependence with respect to selected stations and their geomagnetic location. Knowing that ionosphere is strongly coupled with lower-laying atmosphere, we come to the conclusion that climatology of the troposphere may come into play and be responsible for the difference in time-dependencies and time-lags in ionospheric response to the external solar forcing.

Data and geographical delineation

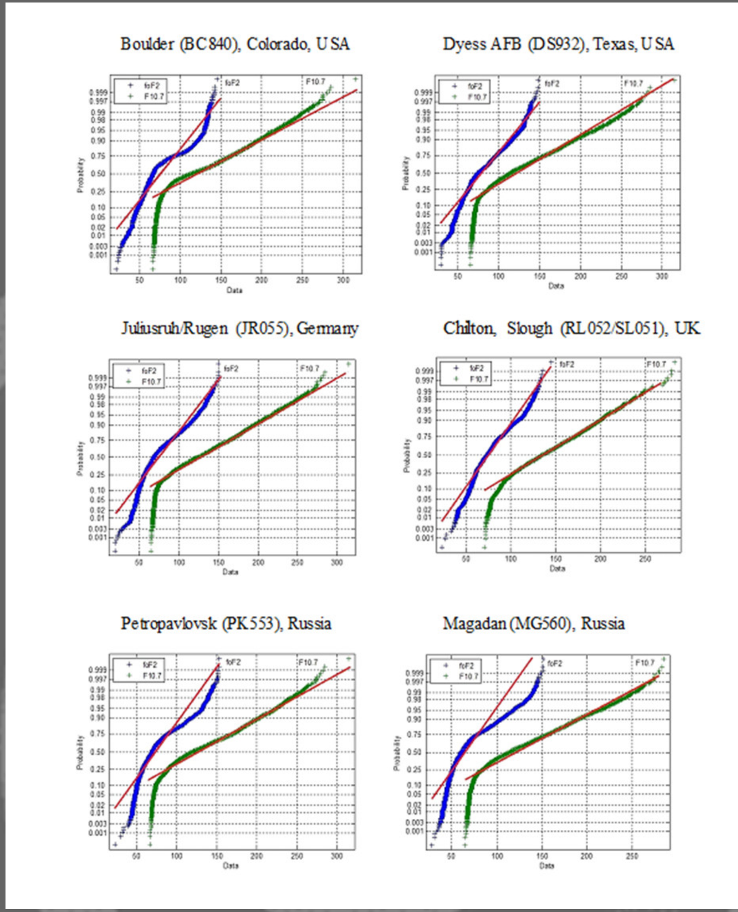
We used 3-hour LT noon average daily foF2 data measured during the period 1994 – 2009 by mid-latitude ionosondes in three longitudinal areas (100°W-110°W / 0°-10°E / 100°E- 160°E). We choose data measured by these 6 mid-latitude ionosondes to compare the influence of geographical longitude with pairwise dependence foF2, F10.7 and shifted F10.7 time series.

US/UK World Magnetic chart, Epoch 2010, Geomagnetic coordinates



Geomagnetic chart by UKDSSC

Station	Geographic		Geomagnetic (Year 2000)	
	Latitude	Longitude	Latitude	Longitude
Boulder (BC840), Colorado, USA	40.0 N	105.3 W	48.2 N	39.8 W
Dyess AFB (DS932), Texas, USA	32.5 N	99.7 W	41.4 N	32.1 W
Juliusruh/Rugen (JR055), Germany	54.6 N	13.4 E	54.0 N	99.6 E
Chilton, Slough (RL052/SL051), UK	51.6 N	1.3 W	53.8 N	83.6 E
Petropavlovsk (PK553), Russia	53.0 N	158.7 E	45.5 N	138.4 W
Magadan (MG560), Russia	60.0 N	151.0 E	51.5 N	146.9 W
according to https://www.ukssdc.ac.uk/gbdc/station-list.html				

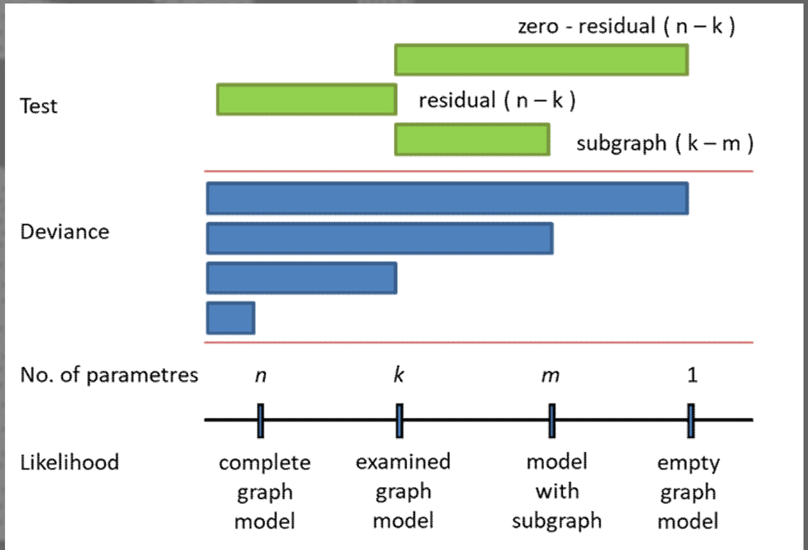


Normal probability plots of foF2 LT noon 5-hour averages vs. and F10.7 daily data. 1994 – 2009

Conditional independence graphs method

Graphical model **with graph *G* is a system of dependency distribution of a random vector *X*, fulfilling conditional independencies given by graph *G*.**

In order to build a graphical model of conditional independencies, it is necessary to determine pair correlations between time series represented by graph vertices first. Pair correlations will be determined from an inverse matrix of covariant matrix of the estimated model residues.



Deviance is defined as $dev^f(G) = 2(l_S - l_G)$,
where l_S is the maximum of logarithmic likelihood function in the saturated model and l_G is the maximum of logarithmic likelihood function in the model with graph *G*. The deviance of a graphical model and graph *G* is then $dev(G) = n\{(S\widehat{D})^T - \ln[\det(S\widehat{D})] - k\}$,
where $\widehat{D} = \widehat{V}^{-1}$ a $\widehat{V}$ is the maximum likelihood estimation of variance matrix *V* in graphical model with graph *G* based on random selection from a multivariate normal distribution.
The estimations $\widehat{D}$ a $\widehat{V}$ are with probability 1 defined unambiguously. The selective variance matrix *S* is the maximum likelihood estimation of variance matrix *V* in a saturated complete graph. The deviance has an asymptotic χ-square distribution and is a test statistic, whether the model with graph *G* matches the alternatively saturated model. The number of degrees of freedom *f* for χ-square distribution of deviance of model with graph *G* equals the number of missing edges in graph *G* and depends on probability distribution of data.

Conditional independence graphs method

Testing the graphical model using deviance

The final graphical model of conditional independencies is then selected according to the following procedure (Whittaker, 1990: s. 241–246):

i. Test of independency of logarithmic data (log (K/K_{t-1}) = log K_t - log K_{t-1}) is run, based on a differentiated sign test, where K_t is the value of realisation of random vector in time *t* and K_{t-1} is the value of realisation of random vector in time (*t-1*).

ii. Test of the normality of logarithmical series.

iii. The system of k-variate normal distributions for *X* is determined and the set of parameters described. For a multivariate normal distribution, the independency is characterized by a variance matrix *V = var(X)* or by its inversion *D*.

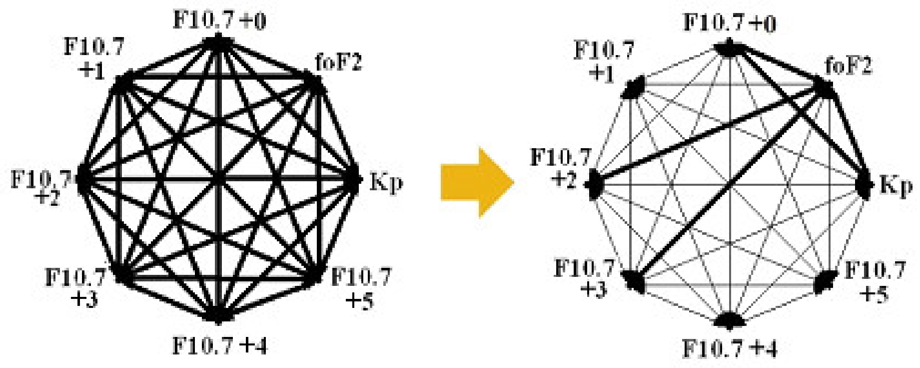
iv. Selection of a graphical model, which will be tested, whether it matches the data.

v. Construction of the likelihood function.

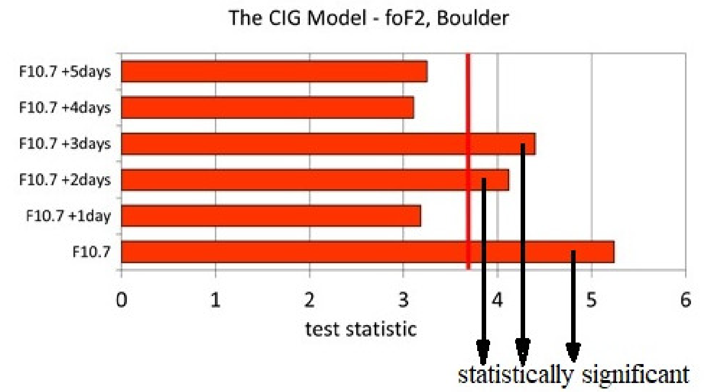
vi. Unknown parameters are determined by the maximization of likelihood function through a set of these parameters. The requirements for the likelihood function are determined by the chosen graphical model.

vii. A test if the selected graphical model and the data match. The test statistic (deviance) equals double difference of the maximized logarithmic likelihood functions. The first one is maximized without restriction, the second one with restriction given by the selected graphical model, because deviance has χ2 - distribution. Therefore, we can assess, if the graphical model fits the data.

Example of CIG computation, Boulder (BC840)



The thin edges denote complete graph of *foF2*, *F10.7* and *F10.7* time shifted timeseries, the bold black ones, edges of minimalised graph. The vertices are labelled by the number of days of the time shift. Edges between vertices which correspond to time shifted series $F10.7_{Shift}$ are trivial, and thus were excluded during calculation.



The conditional independence graphs (CIG) model computation

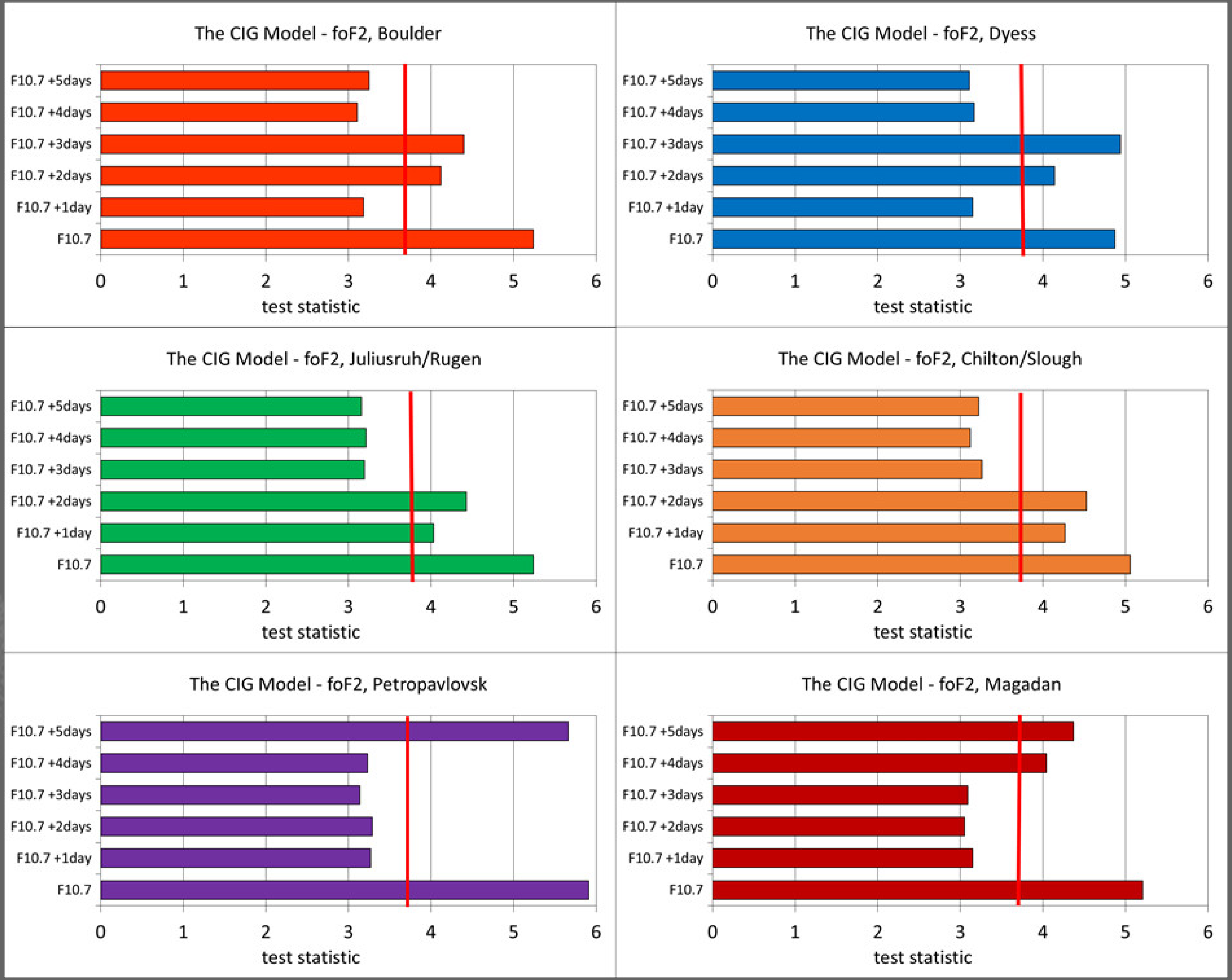
Results of the conditional independence graphs (CIG) model computation

The F10.7 shifted time series graph vertices are labelled by the number of days of the shift.

In our selection of minimized model, all edges with computed deviance less than 3.84 (the 5% point of $\chi^2(1)$) were excluded from the entire graph.

Computed total models deviances and test p-values for all surveyed time series in time period 1994 – 2009.

The stopping threshold for this calculation was 0.05.



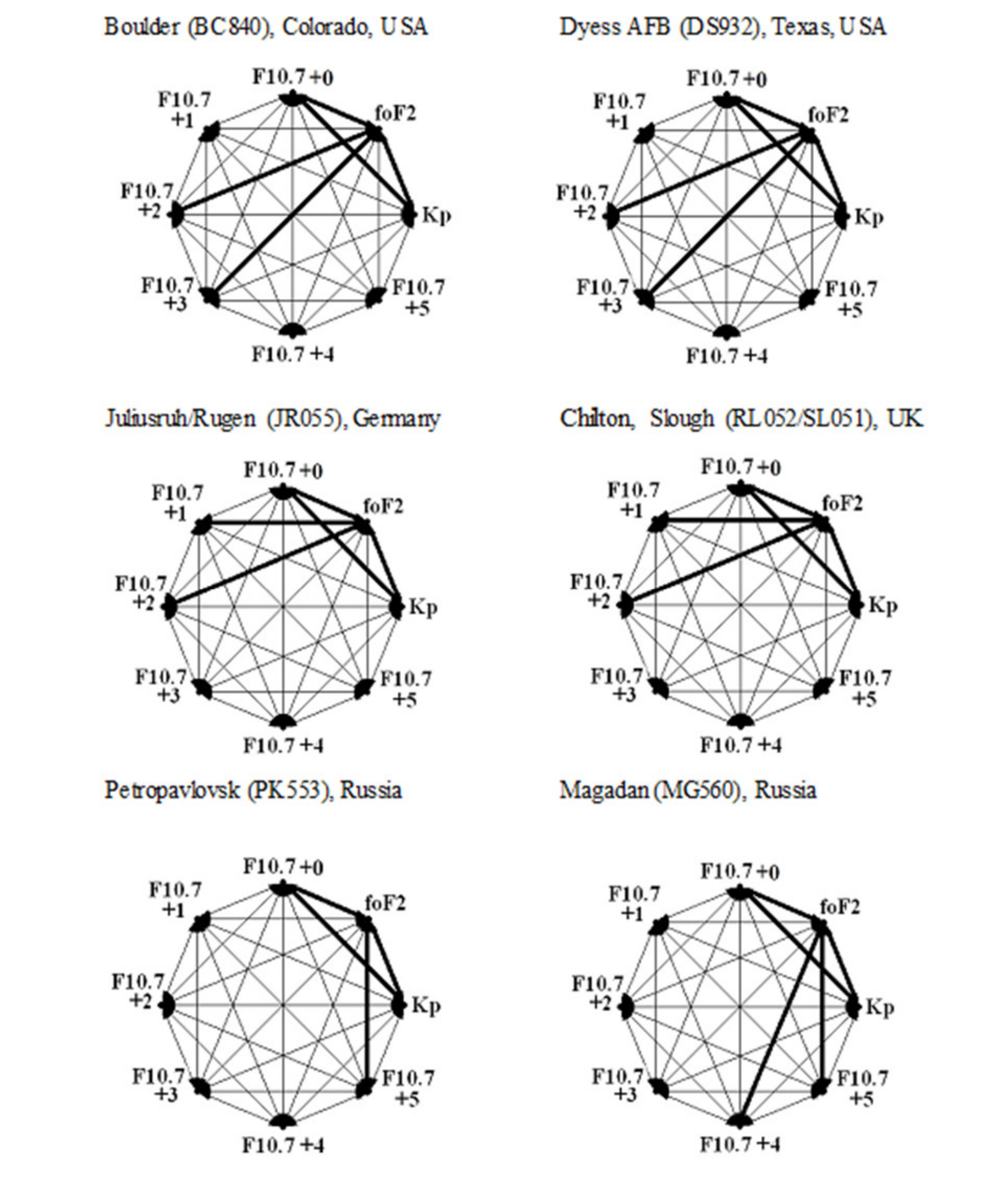
The conditional independence graphs (CIG) model computation

Minimalized conditional independence graphs (CIG), models computation results.

The *F10.7* shifted time series graph vertices are labelled by the number of days of the shift.

In the selection of minimized model, all edges with computed deviance less than 3.84 (the 5% point of $\chi^2_{(1)}$) were excluded from the entire graph.

The stopping threshold for this calculation was 0.05.



The conditional independence graphs (CIG) model computation

Summary of conditional independence graphs (CIG) model computation, sorted by their geomagnetic latitude

Station	Deviance of minimalized graphical model	F10.7	F10.7 +1	F10.7 +2	F10.7 +3	F10.7 +4	F10.7 +5	Kp	Edge number	test p-value
Dyess AFB (DS932)	4.86	✓	✗	✓	✓	✗	✗	✓	4	0.047
Petropavlovsk (PK553)	3.91	✓	✗	✗	✗	✗	✓	✓	3	0.039
Boulder (BC840)	7.56	✓	✗	✓	✓	✗	✗	✓	4	0.037
Magadan (MG560)	6.47	✓	✗	✗	✗	✓	✓	✓	4	0.049
Chilton, Slough (RL052)	9.14	✓	✓	✓	✗	✗	✗	✓	4	0.033
Juliusruh (JR055)	7.38	✓	✓	✓	✗	✗	✗	✓	4	0.039

The initial graphical model for data best fitting is initially selected as an entire graph. The graph vertex represents the component of a random vector (*F10.7, foF2*), the edge between two vertices represents the dependence of these variables (green checks). In our selection of minimized model, all edges with computed deviance less than 3.84 (the 5% point of $\chi^2_{(1)}$) were excluded from the entire graph (red crosses).
This method appears to be more appropriate, especially for longitudinal data, than correlation between *foF2* and geomagnetic and solarindices depend on length of timescale.

Conclusions

- We used the directed conditional independence graphs to emphasize the qualitative assumptions underlying the complex models. The final results of our analysis by conditional independence graphs show that the dependence time shifts between *foF2* and *F10.7* were clearly identified, namely +0 day shift in all cases, and +1, +2-3, and +4-5 day shifts in European, American and east Asia sector, respectively.
- The CIG method has been found to be more eligible for this purpose than correlation analysis between *foF2* and geomagnetic and solar indices because of its relative insensitivity to data gaps or long term changes of data series. CIG method seems more effective than wavelet-based cross-correlation analysis (Roux, 2012) or scale analysis when solving this problem.
- In the minimized graphical model, the interconnections between all series of indices were not identified, so the effect was not complex but varies according to the station region climate type.
- Taking into account that ionosphere is strongly coupled with lower-laying atmosphere, we hypothesized that climatology of the troposphere may be responsible for the difference in time-dependencies and time-lags in ionospheric response to the external solar forcing.

Acknowledgements

The support under the grant 18-01969S and 18-01625S by the Czech Science Foundation is acknowledged.
This work was supported in part by ESA through Contract 4000126709/19/NL/IS “VERA“.

Data

http://www.wdc.rl.ac.uk
http://spidr.ngdc.noaa.gov
http://www.ngdc.noaa.gov
ftp://igs.ensg.ign.fr/pub/igs/products/ionosphere

References

Bannerjee, A. (2011), Probabilistic graphical models for climate data analysis, paper presented at Supercomputing Workshop on Climate Knowledge Discovery (CKD-SC11), IEEE Comput. Soc., Seattle, Wash., 13 Nov.

Edwards, D. (2000). Introductions to Graphical Modeling,2nd edition. NY: SpringerVerlag. (includes MIM software).

Jordan M.I. : Graphical models. Statistical, Science 19; 140–155, 2004.

Koller, D., and N. Friedman (2009), Probabilistic Graphical Models–Principles and Techniques, 1st ed., 1280 pp., MIT Press, Cambridge, Mass.

Lauritzen, S.L. (1996). Graphical Models. NY: Oxford Science Publications.

Roux, S. G.; Abry, P.; Koucká Knížová, P. and Mošna, Z. (2012) Ionosphere fluctuations and global indices: A scale dependent wavelet-based cross-correlation analysis. Journal of Atmospheric and Solar-Terrestrial Physics, ISSN: 1364-6826, Vol: 90, Issue: 1, Page: 186-197.

Tapping, K. F. (2013), The 10.7 cm solar radio flux (F10.7), Space Weather, 11, 394–406, doi:10.1002/swe.20064.

Whittaker J.: Graphical models in applied multivariate statistics, Wiley, New York, 1990.