This is a repository copy of *Predicting the loci of solar eruptions*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/123162/

Version: Accepted Version

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Predicting the Loci of Solar Eruptions

N. Gyenge[1,2,3] **& R. Erdélyi**[1,3]

[1]Solar Physics and Space Plasmas Research Centre (SP^{2}RC),
School of Mathematics and Statistics, University of Sheffield
email: n.g.gyenge@sheffield.ac.uk

[2]Debrecen Heliophysical Observatory (DHO), Konkoly Observatory,
Research Centre for Astronomy and Earth Sciences
Hungarian Academy of Sciences, Debrecen, P.O.Box 30, H-4010, Hungary
[3]Dept. of Astronomy, Eötvös L. University, Pázmány P. sétány 1/A
Budapest, H-1117, Hungary

Abstract.

The longitudinal distribution of solar active regions shows non-homogeneous spatial behaviour, which is often referred to as Active Longitude (AL). Evidence for a significant statistical relationships between the AL and the longitudinal distribution of flare and coronal mass ejections (CME) occurrences is found in Gyenge et al. 2017 (*ApJ*, 838, 18). The present work forecasts the spatial position of AL, hence the most flare/CME capable active regions are also predictable. Our forecast method applies Autoregressive Integrated Moving Average model for the next 2 years time period. We estimated the dates when the solar flare/CME-capable longitudinal belts face towards Earth.

Keywords. Sun: flares, Sun: coronal mass ejections (CMEs), Sun: sunspots, Sun: rotation

1. Motivation

The existing flare and CME forecasting tools are usually based on the local dynamics of active regions, such as the magnetic topology (e.g., Korsós et al. 2014, Korsós & Erdélyi 2016). However, certain global phenomena are also able to provide opportunities for predicting the locus of solar eruptive events, such as the longitudinal inhomogeneous properties of active regions (e.g., Bumba et al. 1965; Bogart 1982; Berdyugina et al. 2006; Balthasar 2007; Zhang et al. 2011). In our previous study, Gyenge et al. 2017 (hereinafter GY17), we found that the most complex active regions appear near the Active Longitude (AL). We concluded that the source of the most probably flare/CME-capable active regions is at the AL. In this paper, we combine our recently developed AL tracking method by Autoregressive Integrated Moving Average (ARIMA) model for forecasting the longitudinal position of the AL numerous Carrington Rotations (CRs) in advance.

2. Forecasting the Active Longitude

We employ the Debrecen Photoheliographic Data (DPD) sunspot catalogue for determining the position of the AL. The DPD covers 540 Carrington Rotations, which equals to about five solar cycles. Panel *A* of Figure 1 shows the migration of the AL. The horizontal axis represents time. The vertical axis shows the Carrington Phase ($CP = L/360$), which is the longitudinal position L of the most significant sunspot group clusters in a certain CR. The clustering method (DBSCAN) groups together points that are relatively closely packed together in a high-density region and it marks outlier points that stand
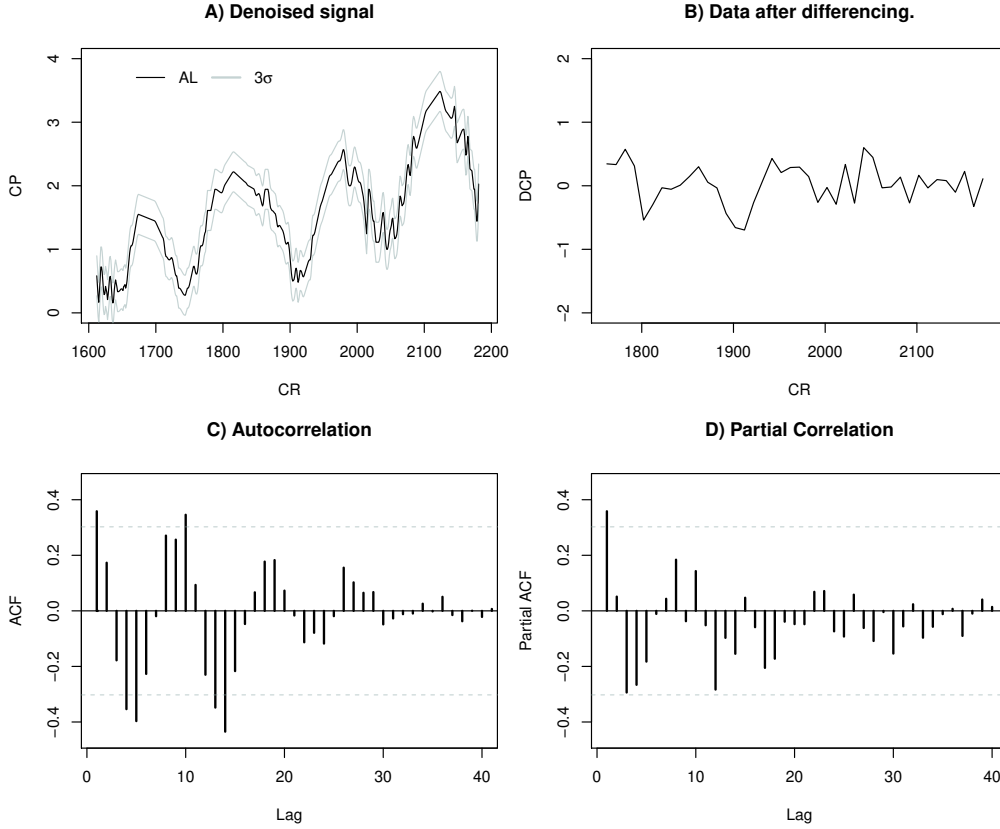
Figure 1: Migration of the AL (Panel A), the normalised denoised signal (Panel B), the autocorrelation (Panel C) and partial correlation (Panel D) of the differentiated data for the northern hemisphere are shown. One lag represents around a year.

alone in low-density regions (Ester et al. 1996). The longitudinal location of the clusters represent the position of the AL. The AL identification method developed based on an empirical approach, described in GY17. The data is distinguished by the hemispheres.

Seasonal ARIMA model (e.g. Ho & Xie. 1998) is used for projecting the future values of time series, derived from the parameters ARIMA(p,d,q) $\times$ (P,D,Q). The parameters p and P are the number of non-seasonal and seasonal autoregressive (AR and SAR) terms. The parameters q and Q are the non-seasonal and seasonal moving average (MA and SMA) terms. Finally, the parameters d and D are the non-seasonal and seasonal differences (Box et al. 1968).

The application of the ARIMA model requires stationarity data; if y_t is considered as a stationary time sample, then the distribution of any subsample $(y_t, ..., y_s)$ is independent on t for all s (Chatfield 1975). The signal is denoised by applying a moving average with 3 CRs window (Panel A of Figure 1). The black line the marks actual position of the AL clusters and the grey belt within the migration demonstrates the mean squared error. The data is non-stationary if a pronounced trend is present. Dickey-Fuller (ADF) tests are used to study the trend-stationarity. The null hypothesis of the test is that the unit root is present in a time series. The alternative hypothesis could mean trend-stationarity.
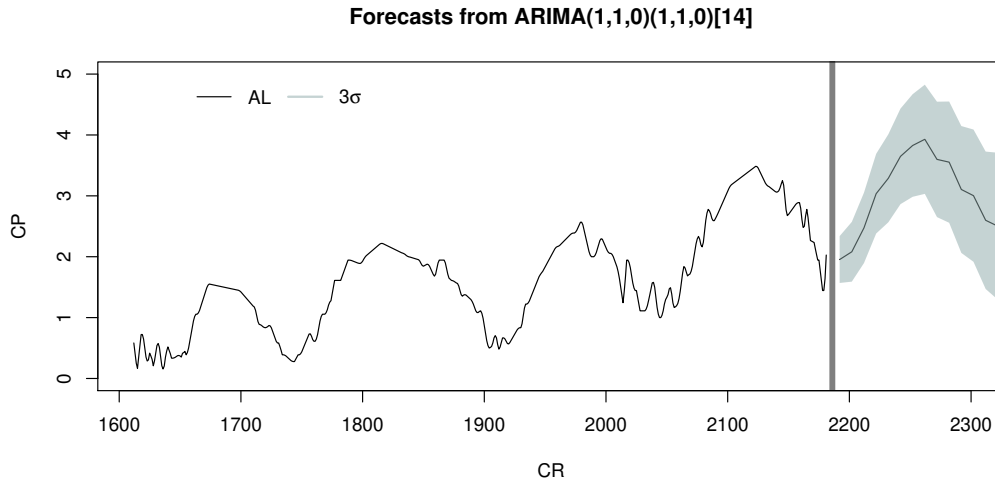
Forecasts from ARIMA(1,1,0)(1,1,0)[14]

Figure 2: Forecasting the AL for the norther hemisphere. The vertical axis is the Carrington Phase (CP) and the horizontal axis stands for the Carrington Rotation (CR). The 3σ standard deviation is displayed by the grey band.

The results of the test also suggest that the data is non-stationary (northern hemisphere data: p-value $= 0.6351$ and southern hemisphere: p-value $= 0.3054$).

The trend-stationarity data can be archived by differencing. Now, the first difference of the time sample means the change between consecutive data points, and it is written as $y'_t = y_t - y_{t-1}$. Seasonality makes it so that the mean of the observations is not constant, but instead it evolves according to a cyclical pattern. The seasonally differenced series has a similar definition but the difference here means the difference between an observation and the corresponding observation from the previous "AL-cycle": $y'_t = y_t - y_{t-n}$, where, n is the "AL-cycle". In GY17, we found that the pattern of the AL migration does not correspond well with the 11-year solar cycle. The lengths of the cycloid AL pattern iterates between 8 years and 14 years. However, there is an 11-year cyclic behaviour on average. Panel B of Figure 1 shows the seasonal differenced and a first differenced data to obtain stationary time sample. The ADF tests confirms the results obtained with differencing and seasonal differencing (northern hemisphere: p-value $= 0.09696$ and southern hemisphere: p-value $= 0.04897$). We found that seasonal differencing $D = 1$ and differencing $d = 1$ are useful terms for the ARIMA model estimation.

The next step of the parametric estimation of the ARIMA model is the examination of the autocorrelation (ACF) and partial autocorrelation (Partial ACF) of the differenced data. Applying ACF and Partial ACF provides some initial guess value about the appropriate ARIMA model parameters (Box et al. 1994). Due the down-sampled data, lag 10 corresponds to 140 CR, which equals to the 11-year solar cycle. Panel C of Figure 1 (ACF) shows a harmonic wave decay pattern and the Partial ACF (Panel D of Figure 1) cuts off quickly. The ACF also reveals the existence of the fluctuation at positive lag around lag 10 (also 20 and 30). If the Partial ACF displays a sharp cut-off while the ACF decays more slowly, this behaviour clearly suggests an AR signature $p = 1$ (Box et al. 1994). The seasonal spikes in the ACF seem to be clear, they fade away in the ACF and cut off after lag 1 in the Partial ACF. This is the sign of a seasonal SAR model $P = 1$ (Box et al. 1994). So, it seems ARIMA(1,1,0) $\times$ (1,1,0) is a good model for estimating the future values of the time series to the northern hemisphere (Figure 2). Similar method-
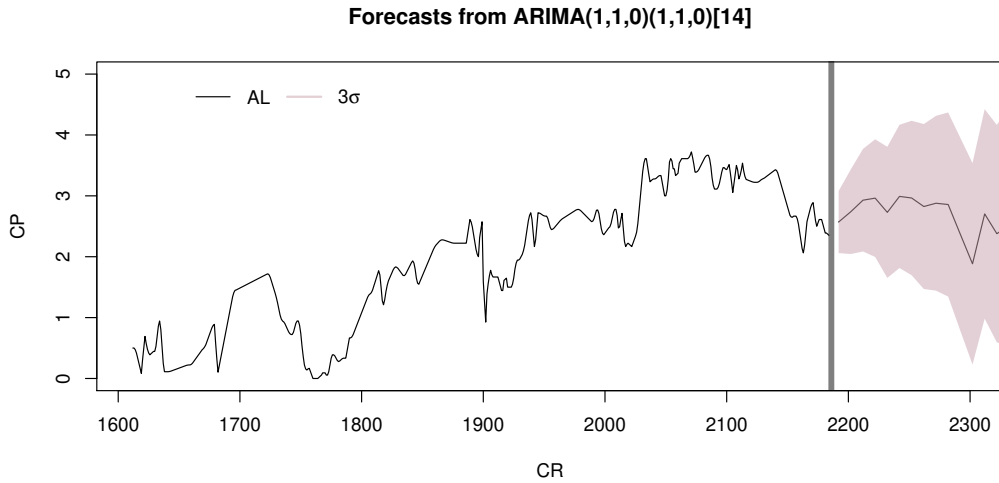
Forecasts from ARIMA(1,1,0)(1,1,0)[14]

Figure 3: Forecasting the AL for the southern hemisphere. The vertical axis is the Carrington Phase (CP) and the horizontal axis stands for the Carrington Rotation (CR). The 3σ standard deviation is displayed by the red band.

ology is applied for the southern hemisphere data (Panel B of Figure 3), however, the statistical error is an order of magnitude larger than in the other hemisphere. Due the large error the souther hemisphere data does not seem to provide reliable forecast.

3. Conclusion

By applying our method outlined in this short report, we are able to forecast the potential flare and/or CME sources several CRs in advance. We predict that the enhanced CME and flare active longitudinal belt will face towards Earth in weeks 41, 42, 46, 50 in 2017 and weeks 2, 5, 9, 13, 17, 20, 24, 27, 35, 39 in 2018. These estimates indicate 60% of flare and CME activity and the highest probability of fast CME occurrence (see GY17).

Acknowledgement

References

Balthasar, H., 2007, *A&A*, 471, 281
Berdyugina, S.V., Moss, D., Sokoloff, D., Usoskin, I.G., 2006, *A&A*, 445, 703
Bumba, V. et al. 1965, *ApJ*, 141, 14921517.
Bogart, R.S., 1982, *Sol. Phys*, 76, 155
Box G E P, Jenkins G M *Applied Statistics*, 17(2), 91-109
Box, G E P, Jenkins, G M, Reinsel G C *Time Series Analysis*, Forecasting and Control
Chatfield C. *Chapman and Hall*, London (see also, 6th ed., 2003)
Ester, M., H. P. Kriegel, J. Sander, and X. Xu 1996, pp. 226231
Gyenge, N.; Singh, T.; Kiss, T. S.; Srivastava, A. K.; Erdélyi, R. 2017, *ApJ*, 838, 1, 10 pp
Ho, S. L. & Xie, M. *Computers & industrial engineering*, 35, 213-216
Korsós, M.B. & Erdélyi, R. 2016, *ApJ*, 823, 2, 11 pp
Korsós, M.B., Baranyi, T., Ludmány, A. 2016, *ApJ*, 789, 2, 7 pp

Zhang, L.Y., Mursula, K., Usoskin, I.G., Wang, H.N. 2011, *JASTP*, 73, 2-3, p. 258-263.