Stochastic Analysis of Solar Flare Duration

Saifuddin Ahmed Jillani *
Department of Physics, Institute of Space & Planetary Astrophysics, University of Karachi, Pakistan

M. Ayub Khan Yousufzai *
Department of Applied Physics, Institute of Space & Planetary Astrophysics, University of Karachi, Pakistan

ABSTRACT

In this communication, stochastic analysis of Solar Flare Duration (SFD) has been implemented. A best model is selected for forecasting after being different methods of residual analysis. This study establishes a quantitative aspect of solar activity according to their duration for the data points covering the specified period of the occurrence of solar flares between Jan 2000–Mar 2006. The main data source for this study is National Oceanic and Atmospheric Administration (NOAA).

INSPEC Classification : A9600, A9660, A9690

Keywords : Coronal Mass Ejection (CME), Solar Proton Event (SPE), Stochastic analysis, Solar Flare Duration (SFD)

*The material presented by the authors does not necessarily portray the viewpoint of the editors and the management of the Institute of Business and Technology (IBT) or Karachi Institute of Power Engineering.

© JICT is published by the Institute of Business and Technology (IBT).
Ibrahim Hydri Road, Korangi Creek, Karachi-75190, Pakistan,
INTRODUCTION:
The means of solar flare involves the reconnection of magnetic field lines. When two oppositely directed magnetic field lines come close to each other they partially annihilate each other and there is an eruption of magnetic energy take place. This is associated with streams of solar particles mostly protons and high energy radiations in the form of X-rays and ã-rays energy bands. The finite resistance in the plasma results in joule heating of the gas, causing temperature to reach $10^7$ K. Such high temperature can produce X-ray and gamma ray emission. The sun’s tumultuous magnetic fields provide the fuel of flares. Scientists generally agree that the energy released in a flare must be stored in the sun’s magnetic fields. Flares erupt from the part of the sun called active regions, where solar magnetic fields are much stronger than average. These areas are most easily identified by the presence of sunspots—those dark-looking patches host the most intense magnetic field on the sun. The solar flares that are the serious threat to our terrestrial climate are consist of gradual phase. These type of flares have duration greater than 1 hour and they are mostly associated with solar protons. The major terrestrial effects include ozone layer depletion, radio wave black-out, danger for orbiting satellites, polar cap absorption and auroras [1-7].

METHODOLOGY:

The important aspect of scientific study is based over the idea of a model. A model is a small copy or the representation of an existing or planned object. It is also may be the supposed structure of something. Also a model can define the real situation of a system.

MODEL SELECTION:
The investigation is not only for the adequate representation of the observed data but also for the prediction that is helpful for other nobserved samples. There are two major types of forecasting models: time series and regression models. In the first type, prediction of the future is based on past values of variable and/or past errors. The objective of such time-series forecasting methods is to discover the pattern in the historical data series and extrapolate that pattern into the future.

Autoregressive / Integrated / Moving Average (ARIMA) models have been studied extensively by George Box and Gwilym Jenkins (1976). Autoregressive (AR) models were first introduced by Yule (1926) and later generalized by Walker (1931), while moving average (MA) models were first used by Slutzky (1937).

In most forecasting situations, accuracy is treated as the intervening
measure for selecting a forecasting method. In many instances, the word accuracy refers to goodness of fit, which in turn refers to how well the forecasting model is able to reproduce the data that are already known. In time series modeling it is possible to use a subset of the known data to forecast the rest of the known data, enabling one to study the accuracy of forecasts more directly [8-9].

TEST OF RANDOMNESS & STATIONARITY:

The autocorrelations for 1, 2, 3, 4, -----, k time lags can be found and denoted by \( r_k \), as follows.

\[
r_k = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(Y_{i-k} - \bar{Y})}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}
\]  \hspace{1cm} (1)

Theoretically, all autocorrelation coefficients for a series of random numbers must be zero.

As shown by Anderson (1942), Bartlett (1946), Quenouille (1949), and others, the autocorrelation coefficients of random data have a sampling distribution that can be approximated by a normal curve with mean zero and standard error \( 1/\sqrt{n} \) [10-11]. Since the total data points are 75 the standard error is \( 1/\sqrt{75} = 0.115 \). This means that 95 percent of all sample-based autocorrelation coefficients must lie within the range specified by the mean plus or minus 1.96 standard errors. That is the data series can be concluded to be random if the calculated autocorrelation coefficients are within the limits.

\[-1.96(0.115) \leq r_k \leq +1.96(0.115) \]  \hspace{1cm} (2)

\[-0.2254 \leq r_k \leq 0.2254 \]

![Fig 1 Autocorrelation of solar flare duration](image)
Fig 1. shows the autocorrelation coefficients for the data of solar flare duration for time lags of 1, 2, 3, ----, 15. The two dashed lines are the upper and lower 95 percent confidence limits for a random series (-0.2254, +0.2254). All of the autocorrelation coefficients do not lie within these limits, confirming the presence of a pattern in the series. Success in time-series analysis depends in large part on interpreting the results from autocorrelation analysis and being able to distinguish what is pattern and what is randomness in the data [12-13].

IDENTIFYING MOVING AVERAGE ASPECT OF ARIMA MODEL
As the partial autocorrelations drop to random values after first 2 or 3 times lags instead decline to zero exponentially, therefore it is suggested that the true generating process is not the MA one.

![Partial Autocorrelation Function](image)

Fig 2 Partial Autocorrelation of solar flare duration

STOCHASTIC APPROACH:
A Stochastic process is a system expressing a phenomenon or experiment developed in some time with random variables. Such a process is also probabilistic. Modeling these processes can be done using autoregressive (AR), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) techniques.

The first thing to note is that most of time series are non-stationary, and the AR and MA aspects of an ARIMA model refer only to a stationary time series. A time series is said to be stationary if there is no systematic change in mean (no trend) which we obtain after first differencing of the data [14-15].
Fig 3 Original data series of SFD

Fig 4 Data series of SFD after first differencing

Fig 5 Best fit model for forecasting of SFD

Fig 6 Model (5, 1, 0) depicts forecasted values of SFD
The formulation of ARIMA (5,1,0) model associated with its equations are as follows.

**ARIMA (5, 1, 0) Model:**

\[ X_t = \mu + X_{t-1} + \beta_1 (X_{t-1} - X_{t-2}) + \beta_2 (X_{t-2} - X_{t-3}) + \beta_3 (X_{t-3} - X_{t-4}) + \beta_4 (X_{t-4} - X_{t-5}) + \beta_5 (X_{t-5} - X_{t-6}) \]  

\[
\hat{\beta}_1 = \frac{\sum_{t=2}^{N} X_{t-1} X_{t-4}}{\sum_{t=2}^{N} X_{t-1}^2}, \quad \hat{\beta}_2 = \frac{\sum_{t=3}^{N} X_{t-1} X_{t-2}}{\sum_{t=3}^{N} X_{t-1}^2}, \quad \hat{\beta}_3 = \frac{\sum_{t=4}^{N} X_{t-2} X_{t-3}}{\sum_{t=4}^{N} X_{t-2}^2}, \quad \hat{\beta}_4 = \frac{\sum_{t=5}^{N} X_{t-3} X_{t-4}}{\sum_{t=5}^{N} X_{t-3}^2},
\]

\[
\hat{\beta}_5 = \frac{\sum_{t=6}^{N} X_{t-4} X_{t-5}}{\sum_{t=6}^{N} X_{t-4}^2}
\]

\[
\hat{\beta}_1 = -0.4720 \pm 0.11544, \quad \hat{\beta}_2 = -0.1786 \pm 0.12409, \quad \hat{\beta}_3 = -0.0153 \pm 0.13110
\]

\[
\hat{\beta}_4 = -0.2968 \pm 0.13019, \quad \hat{\beta}_5 = -0.3317 \pm 0.12275
\]

Using the best fit model from equation (1) the solar flare duration for 76th data point of the year April 2006, can be computed as,

\[ X_{76} = 62.357 + X_{75} - 0.4720 (X_{75} - X_{74}) - 0.1786 (X_{74} - X_{73}) - 0.0153 (X_{73} - X_{72}) - 0.2968 (X_{72} - X_{71}) - 0.3317 (X_{71} - X_{70}) \pm 0.6235 \]

\[ X_{76} = 62.357 + 85.6899 \]

Thus the forecast values as given in the table 1 are manipulated accordingly.

**Table 1: Forecast of Solar Flare Duration (SFD) from ARIMA (5, 1, 0) for April 2006-Jan 2007.**

<table>
<thead>
<tr>
<th>Period</th>
<th>Forecast</th>
<th>Lower 95.00%</th>
<th>Upper 95.00%</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>85.69</td>
<td>55.77</td>
<td>115.61</td>
<td>17.94</td>
</tr>
<tr>
<td>77</td>
<td>96.18</td>
<td>62.35</td>
<td>130.02</td>
<td>20.29</td>
</tr>
<tr>
<td>78</td>
<td>97.53</td>
<td>59.61</td>
<td>135.45</td>
<td>22.74</td>
</tr>
<tr>
<td>79</td>
<td>89.72</td>
<td>47.51</td>
<td>131.94</td>
<td>25.32</td>
</tr>
<tr>
<td>80</td>
<td>92.72</td>
<td>49.56</td>
<td>135.88</td>
<td>25.88</td>
</tr>
<tr>
<td>81</td>
<td>86.91</td>
<td>43.10</td>
<td>130.71</td>
<td>26.27</td>
</tr>
<tr>
<td>82</td>
<td>85.35</td>
<td>39.31</td>
<td>131.40</td>
<td>27.61</td>
</tr>
<tr>
<td>83</td>
<td>88.95</td>
<td>41.71</td>
<td>136.19</td>
<td>28.33</td>
</tr>
<tr>
<td>84</td>
<td>89.32</td>
<td>40.20</td>
<td>138.44</td>
<td>29.46</td>
</tr>
<tr>
<td>85</td>
<td>89.26</td>
<td>37.51</td>
<td>141.00</td>
<td>31.03</td>
</tr>
</tbody>
</table>
RESIDUAL ANALYSIS:
Coefficient of Determination $R^2$:
The coefficient of determination ($R^2$) can be obtained by the following equation.

$$R^2 = 1 - \frac{SS_E}{SS_y} \quad (4)$$

where $SS_E$ = Residual sum of squares
$SS_y$ = Total sum of squares

The forecast error of the selected model have been calculated by the following formulas.

Mean Absolute Forecast Error (MAFE)

$$MAFE = \frac{\sum |y_t - \hat{y}_t|}{m} \quad (5)$$

where $y_t$ is the actual value of Y observed at time t and $\hat{y}_t$ is the forecast value of Y for time t.

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{\sum \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{m} \times 100 \quad (6)$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum (y_t - \hat{y}_t)^2}{m}} \quad (7)$$

where $m$ is the number of time periods for which forecasts have been made.

Table 2 Residual Analysis of Model ARIMA (5,1,0)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>62.85 %</td>
</tr>
<tr>
<td>MAFE</td>
<td>10.74</td>
</tr>
<tr>
<td>MAPE</td>
<td>23.71</td>
</tr>
<tr>
<td>RMSE</td>
<td>17.32</td>
</tr>
</tbody>
</table>

CONCLUSION:
The autocorrelation of stationary data should drop to zero after the second or third time lag, while for a non-stationary series they are significantly different from zero for several time periods, as we have found in our data series. A first differencing produce satisfactory result for converting the non-stationary series to a stationary one. The moving average aspect of the model has been discarded after considering the
partial autocorrelation. Finally, after doing residual analysis it has been suggested that ARIMA (5, 1, 0) is the adequate model for the forecasting of solar flare duration (SFD). The coefficient of determination $R^2$ for a reliable model should be greater than 50% and for this selected model it is 62.85% that confirms its adequacy.

Acknowledgement:
The presenter of this communication thanks the organizers of the conference for providing the opportunity to impart this piece of information. I also appreciate duty leave from University of Karachi to attend and present this paper.

REFERENCES:
[7] Robert C. Whitten; Sheo S. Prasad: Ozone in the free Atmosphere