Prophet-Based Research on the Medium and Long-Term Forecast Method of the F10.7 Flux of the Sun

Shuqin Zhu
Beijing Jiaotong University Software College, Beijing, China
Email: zhusq0307@163.com

Abstract. The F10.7 flux of the sun is an important parameter that characterizes the level of solar activity. However, due to the long-term periodicity and short-term randomness of solar activity, it is difficult to obtain accurate prediction results for F10.7 using statistical methods. The Prophet algorithm is based on time series decomposition and machine learning fitting. It can deal with the situation where there are some outliers in the time series, and it can also deal with the problem of partial missing values. F10.7 is a typical time series data, composed of two parts: time and observations, and has a history of nearly one hundred years of observation. It is inevitable that there will be some outliers and missing values in the observation process. Prophet's data processing characteristics make it suitable for the requirements of the solar F10.7 observation data. Through reasonable selection of change points, it can realize the forecast of the future and the forecast of the seasonal trend, and finally realize the model fitting. The experimental results show that using the Prophet model to predict the consistency of F10.7 data and real data can reach more than 90%.

Keywords. Sun activity; F10.7; Prophet.

1. Introduction
Solar activity is a diverse and varied movement phenomenon in the solar atmosphere. The solar activity has a great impact on the earth, which is mainly reflected in: (1) Solar storms disturb the earth’s atmosphere, affect short-wave radio communication, and even cause a short interruption of short-wave radio; (2) High-energy charged particles disturb the earth’s magnetic field, causing a “magnetic storm” phenomenon, causing the magnetic needle to vibrate violently. Correctly indicate the direction; (3) The high-energy charged particle stream rushes into the high-altitude atmosphere of the two poles at a high speed, which will produce auroral phenomena; (4) The disturbance of the earth’s magnetic field by the solar storm causes natural disasters, such as earthquakes, floods and droughts. When the sun erupts, it will suddenly throw out a large number of protons and electrons, which will severely disturb and destroy the ionosphere on the earth, which will cause the interruption of shortwave radio communication, which will cause the ship and aircraft to lose contact with the ground, and pose a threat to the lives of astronauts traveling in space. The main manifestations of solar activity are sunspots, light spots, spectral spots, flares, prominences, coronal holes, etc. The radio waves from the sun mainly come from the plasma confined by the magnetic field in the active area of the corona. The radio wave labeled F10.7 is a radio wave with a wavelength of 10.7 cm, which is close to the peak position of the solar radio wave being observed [1]. The F10.7 index of the sun refers to the radio radiation flux index of the sun with a wavelength of 10.7 cm (2800 MHz), and the unit is sfu (solar luminous flux unit). 1sfu=1x10^(-22) W·m^2·Hz^(-1), measured once a day (determined by the solar radio emission intensity in the 100 MHz bandwidth). The diffusion scale it exhibits comes from the non-radiative heat of the plasma...
confined by the magnetic field in the active area of the corona, and it shows the degree of solar activity well. The F10.7 cm record of the sun can be traced back to 1947, and is the longest record of solar activity available except for the sunspot record. Due to the good correlation between F10.7 flow and solar activity intensity, convenient observation and long observation history, it is an important parameter that characterizes the level of solar activity. Due to the importance of the solar F10.7 index, it has become the focus of solar activity forecasting.

The forecast of the sun is of great significance to the forecast of other elements of the space environment and the protection of space environmental hazards. Among them, the mid-term forecast of solar activity is the focus of forecast research. Solar activity exhibits cyclical characteristics of 11 years on a long time scale and 27 days on a medium and short time scale. It is a periodic time series data. The medium-term forecast model of F10.7 index is mainly based on time series model. Wang [2] et al. realized the medium-term forecast of F10.7 index according to the law of the sun's rotation, which has a good root mean square error. Liu [3] et al. used an autoregressive time series model to achieve a mid-term forecast of the F10.7 index, which is suitable for periods of low solar activity levels, because the changes in solar activity are relatively stable at this time. Zhong [4] et al. used the singular spectrum analysis method to analyse the F10.7 index, which can well tap the inherent periodicity of F10.7. Yang [5] et al. applied the LSTM (Long Short-term Memory, LSTM) method to the F10.7 medium-term forecast. It has a good performance in both low and high years of solar activity. The average error is within 10% in the high year and the average error is within 10% in the low year. The annual average error is within 2%, indicating that the machine learning method has a better performance in forecasting the F10.7 index. According to the correlation between the area of the active area of the sun and the F10.7 index, Ye [6] et al. proposed and verified the F10.7 forecast formula according to the area size classification, and verified the results with the F10.7 forecast of high and low years. The correlation coefficients between the predicted results and the measured values are 0.9318 and 0.9295, respectively. Elena [7] et al. develop a method called RESONANCE (Radio Emissions from the Sun: ONline ANalytical Computer-aided Estimator) for the prediction of the 13-month smoothed monthly mean F10.7 and F30 indices 1-24 months ahead. This method is based on Kalman filter and has good universality. The Prophet Algorithm is an open source data analysis tool provided by Facebook [8]. The algorithm can not only deal with the situation of some outliers in the time series, but also the situation of some missing values, and it can also predict the future trend of the time series almost automatically. The Prophet algorithm is based on time series decomposition and machine learning fitting. Among them, the open source tool pyStan is used when fitting the model, so the results that need to be predicted can be obtained in a faster time. Based on this observational fact, the medium-term forecast of the F10.7 index for the next 27 days is carried out based on the Prophet method. Using a continuous long period of F0.7 data as training data, build a Prophet prediction model to predict the F10.7 index of solar activity in the next 27 days.

2. Method

2.1. Solar Cycle

The solar activity week is also called the sunspot activity week, which is determined according to the changing law of the number of sunspots. From the long-term record of the relative number of sunspots, the average value of the relative number of sunspots clearly shows a periodicity of about 11 years, with the shortest being 9.0 years and the longest being 13.6 years [9], as shown in figure 1.
The reason for the formation of the solar cycle is the power of heat generated inside the sun, that is, how much heat is produced per unit time is stable. This heat will increase the temperature and expand the gaseous matter, but under the action of universal gravitation, the gaseous matter on the sun has a high density and can be close to liquid, so the intermolecular force is very significant, unlike the general gas molecules on the surface of the earth. The force between them is very weak. Generally, gas expands “uniformly” as the temperature rises, and the expansion curve is smooth. However, from liquid to gas, the curve of volume change with temperature has a sharp increase. It may also be that when the internal temperature of the sun rises to a certain value, a certain substance or several substances will have a rapid expansion phenomenon, which will initiate the rapid upward flow of internal substances one by one. At this time, the sun is active. period. At this time, electromagnetic energy (electromagnetic waves and magnetic fields) and thermal energy will be released, and local explosions, vortices, etc. will be formed after colliding with the colder material above, which are phenomena such as flares and sunspots. These are significantly increased during the active solar period than in the inactive period. After the release of heat and electromagnetic energy, the internal temperature of the sun drops, and at this time it returns to a relatively quiet period. This is similar to intermittent hot springs on the earth. Under certain conditions of geothermal release power, the temperature gradually rises. When the temperature reaches a certain temperature, the water rapidly vaporizes and expands, causing the hot springs to surge periodically. After the surge, the heat is released, and the temperature of the lower part is lowered. At this time, it enters a quiet period again.

In the first few years of the solar active week, sunspots continue to be produced, and the activity becomes more and more intensified. The year when the number of sunspots reaches a maximum is called the maximum year of solar activity (peak year); in the following years, the sunspot activity gradually weakens, there are fewer sunspots, and the years when there are very few sunspots are called minimum years of solar activity (valley years). There is a solar cycle between two adjacent minimal years.

2.2. Principle of Prophet
In the field of time series analysis, there is a common analysis method called Decomposition of Time Series, which divides the time series $y_t$ into several parts, which are the trend item $T_t$, seasonal item $S_t$, and the remaining item $R_t$. In other words, for all $t \geq 0$, there are.

$$y_t = T_t + S_t + R_t$$

or

$$y_t = T_t \times S_t \times R_t$$

which is

$$\ln y_t = \ln T_t + \ln S_t + \ln R_t$$
In the Prophet algorithm, the holiday effect $h_t$ is additionally considered. The Prophet algorithm obtains the predicted value of the time series by fitting these items, and then finally adding them up.

2.2.1. Trend Item. In the Prophet algorithm, the trend term has two important functions, one is based on the logistic function (logistic function), and the other is based on the piecewise linear function (piecewise linear function). In a real time series, the trend of the curve will definitely not remain the same. At certain specific times or with a certain potential periodic curve, there will be changes. At this time, some scholars will study the detection of change points. It is also called change point detection. In Prophet, it is necessary to set the position of the change point, and the trend and trend of each segment will also change according to the situation of the change point. There are two methods in the program, one is to specify the position of the change point by manual designation; the other is to automatically select it through an algorithm. In the default function, Prophet will select 25 ($n$ changepoints) change points, and then set the change point range to the first 80% (changepoint_range), that is, change points will be set in the first 80% of the time series.

2.2.2. Seasonal Trend. Time series usually show seasonal changes with seasonal changes such as days, weeks, months, and years, which is also called cyclical changes. In mathematical analysis, the periodic function in the interval can be expressed by the function of sine and cosine.

2.2.3. Holiday Effect. In the real environment, in addition to weekends, there are also many holidays, and different countries have different holidays. Since each holiday has a different degree of influence on the time series, different holidays can be regarded as independent models, and different before and after window values can be set for different holidays, indicating that the holiday will affect the time series for a period of time before and after.

2.3. Prophet Usage

Prophet algorithm workflow:
(1) Enter the time stamp and corresponding value of a known time series;
(2) Enter the length of the time series to be predicted;
(3) Output the future time series trend.
(4) The output result can provide necessary statistical indicators, including fitting curve, upper bound and lower bound, etc.

In general, the offline storage format of time series is timestamp and value. If more, you can provide time series ID, label and other content. Therefore, the time series stored offline are usually in the following form. Among them, date refers to a specific time stamp, category refers to a specific time series id, value refers to the value of this category time series under date, and label refers to a manually labeled label (“0” means Abnormal, “1” means normal, “unknown” means there is no mark or human judgment is unclear).

The time series required by fbprophet is also in this format, as long as you use a csv file to store two columns, the name of the first column is “ds”, and the name of the second column is “y”. The first column indicates the timestamp of the time series, and the second column indicates the value of the time series. Through the calculation of prophet, yhat, yhat_lower, and yhat_upper can be calculated, which respectively represent the predicted value of the time series, the lower bound of the predicted value, and the upper bound of the predicted value.

3. Experimental Results

The experiment is based on the Prophet algorithm, and uses the F10.7 data from January 1, 2016 to March 31, 2018 as the training set, and the data from April 1, 2018 to April 27, 2018 as the test set. Finally, the predicted and actual results obtained are shown in figure 2. It can be seen that the predicted value is in good agreement with the actual value.
According to the experimental results, the Prophet algorithm is applied to the solar F10.7 index forecast, which has good time efficiency and forecast accuracy, and can be used as a reference for the mid-term solar activity.

4. Summary
This paper uses the Prophet algorithm to forecast the solar F10.7 index, but the Prophet algorithm only considers the annual periodicity, and does not consider the periodicity of the solar active week. Further improvements can be made in subsequent studies.

Acknowledgments
Thanks to NOAA’s National Geophysical Data Center for providing historical data on F10.7 of the Sun.

References
[1] Tapping K F 2013 The 10.7 cm solar radio flux (f10.7) Space Weather 11 (7) 394-406.
[2] Wang H B Xiong J N and Zhao C Y 2015 The mid-term forecast method of solar radiation index Chinese Astronomy and Astrophysics 39 (2) 198-211.
[3] Liu S Q, Zhong Q Z, Wen Jing and Dou X K 2010 Xiankang modeling research of 10.7 cm solar radio flux 27-day forecast (I) Chinese Journal of Space Science 30 (1) 353.
[4] Zhong Q Z, Liu S Q, He J X and Gong J C 2005 Application of singular spectrum analysis to solar 10.7 cm radio flux 27-day forecast Chinese Journal of Space Science 25 (3) 41.
[5] Yang X, Zhu Y G, Yang S G, Wang X J and Zhong Q Z 2020 Application of LSTM neural network in F10.7 solar radio flux mid-term forecast Chinese Journal of Space Science 40 (2) 353.
[6] Ye Q, Song Q and Xue B S 2019 F10.7 index forecasting method based on area statistics of solar active regions Chinese Journal of Space Science 39 (5) 352.
[7] Elena P, Tatiana P, Astrid V Stijn L, Benjamin V and Tim F 2021 Medium-term predictions of F10.7 and F30 cm solar radio flux with the adaptive kalman filter The Astrophysical Journal 254 (9).
[8] Taylor S J and Letham B 2017 Forecasting at scale The American Statistician 72 (1) 37-45.
[9] Roy S, Prasad A, Panja S C, Ghosh K and Patra S N 2019. A search for periodicities in F10.7 solar radio flux data Aстрономический вестник 53 (3) 240-240.