RESEARCH ARTICLE

A Hybrid Regression-Neural Network (HR-NN) Method for Forecasting the Solar Activity

D. I. Okoh[1,2] (iD), **G. K. Seemala[2]** (iD), **A. B. Rabiu[1]** (iD), **J. Uwamahoro[3]**, **J. B. Habarulema[4,5]** (iD), **and M. Aggarwal[2]**

[1]Center for Atmospheric Research, National Space Research and Development Agency, Abuja, Nigeria, [2]Indian Institute of Geomagnetism, Navi Mumbai, India, [3]College of Education, Department of Mathematics and Science, University of Rwanda, Kigali, Rwanda, [4]South African National Space Agency, Space Science, Hermanus, South Africa, [5]Department of Physics and Electronics, Rhodes University, Grahamstown, South Africa

Abstract The Sun is the major driver of space weather events, and as a result, most applications requiring modeling/forecasting of space weather phenomena depend largely on the activities of Sun. Accurate modeling of solar activity parameters like the sunspot number (SSN) is therefore considered significant for the quantitative modeling of space weather phenomena. Sunspot number forecasts are applied in ionospheric models like the International Reference Ionosphere model and in several other projects requiring prediction of space weather phenomena. A method called Hybrid Regression-Neural Network that combines regression analysis and neural network learning is used for forecasting the SSN. Considering the geomagnetic Ap index during the end of the previous cycle (known as the precursor Ap index) as a reliable measurement, we predict the end of solar cycle 24 to be in March 2020 (±7 months), with monthly SSN 5.4 (±5.5). Using an estimated value of precursor Ap index as 5.6 nT for solar cycle 25, we predict the maximum SSN to be 122.1 (±18.2) in January 2025 (±6 months) and the minimum to be 6.0 (±5.5) in April 2031 (±5 months). We found from the model that on changing the assumed value of precursor Ap index (5.6 nT) by ±1 nT, the predicted peak of solar cycle 25 changes by about 11 sunspots for every 1-nT change in the assumed precursor Ap index.

Plain Language Summary A combination of regression and neural network methodology is used to forecast the solar activity using the current and past observed solar parameters. To test the method, the current solar cycle 24 activity is forecasted using previous solar cycles and then the same method is used to predict the upcoming solar cycle 25. The results are presented in this publication.

1. Introduction

Sunspots are darker, and colder regions on the Sun's surface that are marked by the intense magnetic activity and act as a base to the explosive solar flares and coronal mass ejections from the Sun's corona. The sunspot number (SSN) has long been used as an indicator of solar activity level of the entire visible disc of the Sun (e.g., Duhau, 2003; Hoyt & Schatten, 1998; Kirkwood, 1869) and exhibits a day-to-day variability with periodicity of about 11 years of cycle. It is a measure of the number of sunspots and groups of sunspots present on the surface of the Sun. Its historical record dates as far back as the 17th century. SSN was first introduced in 1848 by Rudolf Wolf, and the measurements of solar activity are available continuously thereon (Kiepenheuer, 1953; McKinnon & Waldmeier, 1987; Waldmeier, 1961). Initially, the SSN measurements were maintained by the Swiss Federal Observatory in Zürich, Switzerland. Now it is provided and maintained by the Solar Influences Data Center in Brussels, Belgium, where monthly updates are available online.

Sunspots have a significant role in the prediction of solar activity and space weather events and are needed to plan long-term space missions. The minimum solar activity (lowest SSN) has also been found associated with atmospheric cooling. In particular, the Dalton minimum (1795–1830), the Maunder minimum (circa 1635–1705), and the Spörer minimum (circa 1450–1550) are considered as associated with significantly cooler climates (e.g., Eddy, 1981). Jiang and Xu (1986) dispute this generalization about the Spörer minimum. During high SSN transient solar phenomena such as coronal mass ejections and solar flares are most frequent (Ramesh, 2010; Uwamahoro & McKinnell, 2013). These phenomena often lead to geomagnetic storm phenomena that represent adverse aspects of space weather.

Sunspot number forecasts have, for example, been applied in projects like the International Reference Ionosphere (IRI), NeQuick, and IRI-Plas (IRI extended to the plasmasphere). These models are usually updated in time intervals of about 2–5 years, and so SSN forecast lead times of about 5 years are sufficient for this application. For this reason, a lot of research resources are put into modeling of the SSN and hence the solar activity. There exist various techniques that have been used in the past for the prediction of the occurrence and amplitude of solar cycles (SCs). Early predictions of the amplitude and timing for SC 24 have been listed by Pesnell (2008), which included 54 different predictions obtained by applying various techniques based on climatology (past behavior), dynamo models, spectral analysis, neural networks, and geomagnetic and solar precursor methods. The different techniques have given the different predictions.

The precursor method is one of the recently developed models for the prediction of amplitude of SC (Conway, 1998; Hathaway, 2010; Svalgaard et al., 2005). These precursor techniques based on the geomagnetic activity near or before the time of solar minimum to forecast SSN in the subsequent SC are gaining the most attention (Bhatt et al., 2009a, 2009b; Feynman, 1982; Gonzalez & Schatten, 1988; Ohl & Ohl, 1979; Thompson, 1993; Wilson et al., 1998). Also, neural network is one of common techniques used to search for relationships among indicators (Herbrichet al.,1999). Gholipour et al. (2005) proposed a method based on spectral analysis and neurofuzzy modeling and made an early prediction of the maximum SSN for cycle 24 to be 145 in 2011–2012. Svalgaard et al. (2005) predicted Cycle 24 to be the weakest in 100 years ($R_{24} = 70 \pm 2$) based on the polar field strength, whereas Dikpati et al. (2006) predicted a strong Cycle 24 ($R_{24} = 155–180$) based on applying a dynamo model. Bhatt et al. (2009a) using the precursor technique found the maximum amplitude of SC 24 to be 111.

Neural network forecasts are derived from nonlinear, statistical algorithms that determine and model complex relationships between inputs and outputs to find patterns in the data that can be extrapolated. Some researches (e.g., Habarulema, 2010; Okoh et al., 2016) have demonstrated that neural networks are efficient for modeling ionospheric variations that depend mostly on the Sun's activities. For parameters like SSN that have shown variations in the SC properties (e.g., the value of SSN at the cycle peaks and the cycle durations), neural networks can be guided in techniques that combine other methods to increase their accuracies. We present here a novel technique (known as the Hybrid Regression-Neural Network, HR-NN) in which the method of regression is used to derive key parameters of a SC, and then these key parameters are subsequently fed in as inputs for a trained neural network procedure from which instantaneous SSN values can be forecasted. The idea in using regression to derive key parameters of the SCs is to provide the neural network with guided information regarding the cycle amplitudes and durations. Ultimately, the method of neural network is used to learn how the SSNs vary through the cycles by relying on the wealth of information obtainable from the large volume of available SSN data. Section 2 describes the data used in the study, section 3 gives the method and results, and discussion and conclusions from the study are provided in sections 4 and 5, respectively.

2. Data Used

There exists several indices that are indicators for the geomagnetic activity: *Kp*, *Dst*, aa, ap, *Ap*, and *AE*. The geomagnetic activity indices, viz., *Ap* and aa, have been shown to be good precursors for forecasting the SSN (e.g., Bhatt et al., 2009a, 2009b; Maris & Oncica, 2006; Uwamahoro et al., 2009). In the present analysis, we exploit the *Ap* index as geomagnetic precursor to estimate the maximum amplitude of upcoming SCs. The *Ap* index provides a daily average level of the geomagnetic activity. The 3-hourly values of *Kp* is converted to ap values, and the average from 8 daily ap values gives the daily *Ap* index of a certain day, with units of nT. Related to the *Kp* index, they are average values of the irregular disturbance levels in the horizontal geomagnetic field components, observed at the selected magnetic observatories worldwide. Definitive values of *Ap* are maintained and provided by GeoForschungsZentrum Potsdam, Germany, on behalf of the International Service of Geomagnetic Indices of the International Association of Geomagnetism and Aeronomy. The *Ap* index is thus a geomagnetic activity index where days with high levels of geomagnetic activity (magnetically disturbed days) have higher values of daily *Ap* indices. The *Ap* index data were obtained from the GSFC/SPDF OMNIWeb interface at https://omniweb.gsfc.nasa.gov. The data from OMNIWeb are available from year 1963, and all available *Ap* index data till year 2017 were used, which makes the *Ap* data set from SC 20 to SC 24.
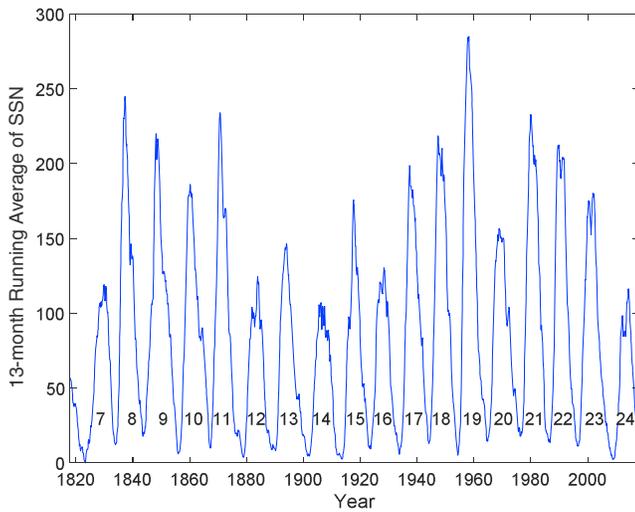
Figure 1. Illustration of solar cycles 7 to 24 using 13-month running average of the sunspot number. The numbers inscribed indicate the respective solar cycle numbers for each cycle.

Sunspot number data used in this work are the revised SSN version 2 (Clette et al., 2015) obtained from the WDC-SILSO, Royal Observatory of Belgium, Brussels (http://www.sidc.be/silso/datafiles). The data are a daily record of SSNs from January 1818 to November 2017. Due to nonavailability of reliable data before 1818 and the need to ensure accuracy of results, SSN for days earlier than 1818 were not used in this work.

3. Methods and Results

3.1. Regression Method

The daily values and monthly averages of the daily SSN are noisy and must be smoothed in some manner in order to obtain significant information like minima, maxima, etc. To ensure smoothness of the data profile, a monthly average of the SSN was constructed, and the 13-month running average was further computed using the formula in equation (1) as defined by Conway (1998). The 13-month running mean is a standard smoothing that is centered on the month in subject and using half-weights for the months at the start and end of the series.

$$R13_i = \frac{1}{12}\left(\Sigma_{j=-5}^{5} R_{i-j} + \frac{1}{2}\left(R_{i-6} + R_{i+6}\right)\right) \tag{1}$$

where i is the index of the month for the 13-month running mean computation; R_i is the monthly average SSN for month i, and j is the index from the preceding to the subsequent months of which the monthly average SSNs are used to compute the running mean. In the same manner, we calculated the monthly mean values of Ap indices (smoothed using 13-month running average as in equation (1)). The Ap index for the month at the end of a cycle is considered as the precursor Ap index for the next cycle.

Figure 1 illustrates the variability of SSN over the period from 1818 to 2017 using the 13-month running average of SSNs derived in this work. The period covers SC numbers 7 to the present 24. Figure 1 shows that there is conspicuous variation in the peak values of the SSN for the cycles varying from SC to cycle. The cycle properties (minimum and maximum values of SSN, the times taken to rise from onset to peak

Table 1
Observed Characteristics of Solar Cycles 7–24

SC	Start year	Peak year	End year	SSN at start	SSN at peak	SSN at end	Rise duration (years)	Fall duration (years)	Total duration (years)
7	1,823.416	1,829.833	1,833.833	1.6	119.2	12.2	6.417	4.000	10.417
8	1,833.916	1,837.167	1,843.500	12.4	244.9	17.6	3.251	6.333	9.584
9	1,843.583	1,848.083	1,855.917	18.1	219.9	6.0	4.500	7.834	12.334
10	1,856.000	1,860.083	1,867.167	6.3	186.2	9.9	4.083	7.084	11.167
11	1,867.250	1,870.583	1,878.917	10.0	234.0	3.7	3.333	8.334	11.667
12	1,879.000	1,883.917	1,890.167	4.1	124.4	8.3	4.917	6.250	11.167
13	1,890.250	1,894.000	1,902.000	9.7	146.5	4.5	3.750	8.000	11.750
14	1,902.083	1,906.083	1,913.500	4.6	107.1	2.5	4.000	7.417	11.417
15	1,913.583	1,917.583	1,923.583	2.5	175.7	9.3	4.000	6.000	10.000
16	1,923.666	1,928.250	1,933.667	9.5	130.2	5.8	4.584	5.417	10.001
17	1,933.750	1,937.250	1,944.083	6.0	198.6	12.9	3.500	6.833	10.333
18	1,944.166	1,947.333	1,954.250	13.0	218.7	5.1	3.167	6.917	10.084
19	1,954.333	1,958.167	1,964.750	5.5	285.0	14.3	3.834	6.583	10.417
20	1,964.833	1,968.833	1,976.167	15.0	156.6	17.8	4.000	7.334	11.334
21	1,976.250	1,979.917	1,986.667	18.4	232.9	13.5	3.667	6.750	10.417
22	1,986.750	1,989.833	1,996.333	14.7	212.5	11.2	3.083	6.500	9.583
23	1,996.416	2,001.833	2,008.917	11.6	180.3	2.2	5.417	7.084	12.501
24	2,009.000	2,014.250	NA	2.5	116.4	NA	5.250	NA	NA

Note. NA is used for characteristics of the present solar cycle (SC 24) that have not been measured at the time of developing this research.
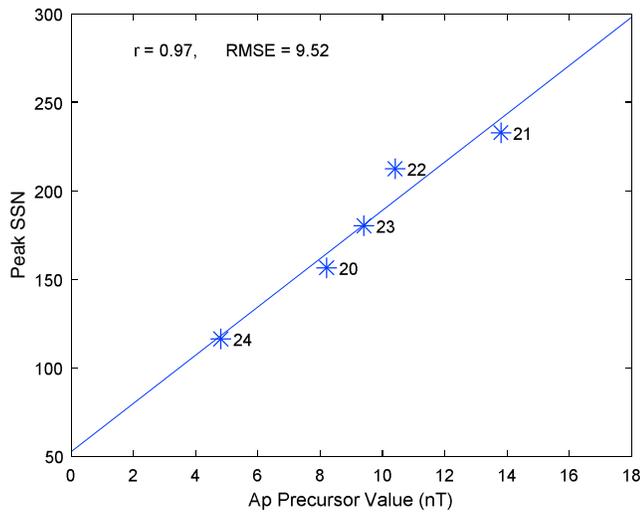
Figure 2. Relationship between the peak of sunspot number and cycle precursor value of *Ap* index for SC 20–24 (RMSE: root-mean-square error).

of cycles, the times taken to fall from peak to end of cycles, and the total durations for each cycle) are found to be different. To investigate the connections between various parameters of the SCs, we used the regression analysis. Table 1 summarizes properties of the SCs as obtained using the computed 13-month running averages. Each SC is distinct. The minimum values of SSN varied between 1.6 and 18.4, while the maximum values varied between 107.1 and 285.0 for different cycles. The SCs have total durations that vary between ~9.5 and ~12.5 years, with a mean value of ~11 years.

3.1.1. Peak Value of SSN in a Cycle Versus Precursor *Ap* Index

There is a long standing observation of the connection between the solar activity and the precursor geomagnetic activity (Bhatt et al., 2009a; Cliver et al., 1999; Gonzalez & Schatten, 1988; Pesnell, 2017). From Ruzmaikin (1985), the relationship between SSN and the corresponding geomagnetic aa/*Ap* index can also be viewed as indirect evidence for the solar dynamo. The state of geomagnetic activity during the end of a solar activity cycle has been found to be a reliable indication of the amplitude of the following solar activity maximum or peak SSN (Cliver et al., 1998; Hathaway & Wilson, 2006). Figure 2 illustrates a plot of the peak values of SSN versus the precursor *Ap* index for SCs 20 to 24 (as *Ap* values are available on website for SC 20–24). The precursor *Ap* index is the value of the *Ap* index at the end of the preceding SC.

Figure 2 shows a direct relationship between the peak of SSN and the precursor *Ap* index, suggesting that higher/lower peak of SSN for next SC is associated with higher/lower *Ap* precursor values; the value of correlation coefficient (*r*) is ~0.97, and the best fitted linear regression is illustrated in equation (2).

$$R_{peak} = 55.56 + 13.65\,Ap \qquad (\pm 9.52) \qquad (2)$$

where R_{peak} is the peak of SSN and *Ap* is the precursor *Ap* index for the cycle. The relationship in equation (2) shows that, at the onset of a SC, the peak value of SSN (R_{peak}) for the cycle can be reliably predicted using the observed precursor *Ap* values. Pesnell (2017) had also illustrated a direct relationship between the peak of solar radio flux at 10.7 cm (F10.7 peak) and the precursor *Ap* index; F10.7peak = 121.1 + 8.5 *Ap*.

3.1.2. Mean Rise Rate Versus Peak Value of SSN

To understand and explore the characteristics of the SCs, we obtained the mean rise rate for each cycle (defined as the mean rate of increase of the SSNs from the onset to the peak of the cycles). The relationship between the mean rise rate and the peak value of SSN was investigated, and the results are shown in Figure 3.

Figure 3 reveals that there is a direct relationship between the mean rise rate and the peak value of SSN for each SC 7 to 24; the rate of increase of SSN is higher if the SSN at the peak of the same cycle is higher, and vice versa. The correlation coefficient between the rise rate and the peak value of SSN is ~0.94, an indication that the connection between the two is strong. The best fitted linear regression is illustrated in equation (3a).

$$R_{\Delta+} = 0.33\,R_{peak} - 16.23 \qquad (\pm 6.20\ \mathrm{yr^{-1}}) \qquad (3a)$$

$R_{\Delta+}$ is the mean rise rate of SSN for the cycle, computed as defined in equation (3b).
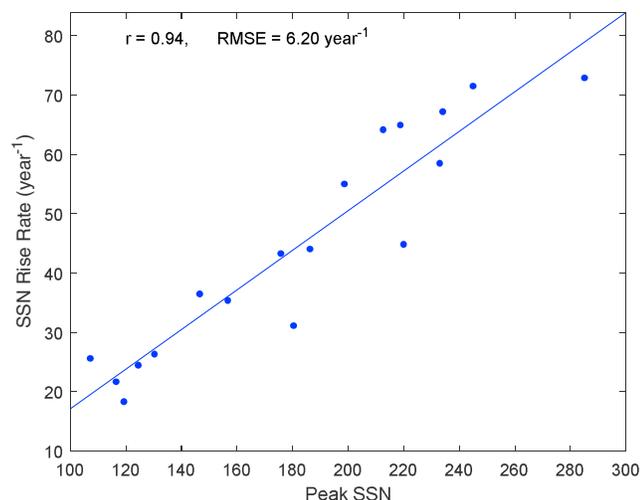
Figure 3. Relationship between the mean rise rate and the peak value of sunspot number for each SC 7–24.
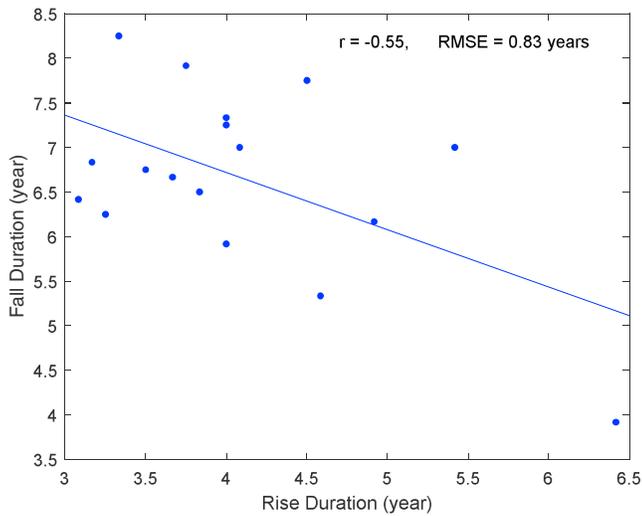
Figure 4. Relationship between the rise duration and the fall duration for each SC 7–24.

$$R_{\Delta+} = \frac{R_{\text{peak}} - R_{\text{onset}}}{T_{\text{rise}}} \tag{3b}$$

where R_{onset} is the SSN at onset of the cycle and T_{rise} is the time taken for the cycle to rise from the onset to the peak of the cycle. Equations (3a) and (3b) reveal the possibility to compute (at the onset of an SC) how long it will take to reach the peak of the cycle from the onset (i.e., the rise duration, T_{rise}) provided that R_{peak} is "known," which can be obtained in advance using the Precursor Ap index using equation (2).

3.1.3. Fall Duration Versus Rise Duration

The relationship between the rise duration (T_{rise}) and the fall duration (the time taken from the peak to the end of the cycle, T_{fall}) for each SC 7 to 24 was also investigated, and the results are as shown in Figure 4.

Figure 4 shows that there exists a moderate inverse correlation (correlation coefficient ~−0.55) between the rise duration and the fall duration, implying broadly that cycles that take longer times to rise will take shorter times to fall, and vice versa. The rise duration and fall duration for SCs vary between 3 to 6.5 years and 5 to 8.5 years, respectively. The best fitted linear regression is shown in equation (4).

$$T_{\text{fall}} = 9.29 - 0.64 \, T_{\text{rise}} \qquad (\pm 0.83 \text{ year}) \tag{4}$$

where T_{fall} is the fall duration. Following from equation (4), it is therefore possible to compute (at the onset of the cycle) the fall duration provided T_{rise} is known, which can be obtained using equation (3b), and hence, the total duration of the cycle ($T_{\text{rise}} + T_{\text{fall}}$) can be calculated.

3.1.4. Mean Fall Rate Versus Mean Rise Rate

Finally, the relationship between the mean rise rate and the mean fall rate (defined as the mean rate of decrease of the Sunspot numbers from the peak to the end of the cycle) was investigated. The results are illustrated in Figure 5. The SSN rise and fall rates vary between 20 to 75 SSN/year and 15 to 40 SSN/year, respectively, during SC 7–24. The figure shows that there exists a positive correlation (correlation coefficient ~0.79) between the mean rise rate and the mean fall rate, implying broadly that cycles with higher rise rates also have higher fall rates, and vice versa. Equation (5a) shows the best fitted linear regression between the mean rise rate and the mean fall rate.

$$R_{\Delta-} = 0.30 \, R_{\Delta+} + 12.84 \qquad (\pm 4.12 \text{ year}^{-1}) \tag{5a}$$

where $R_{\Delta-}$ is the mean fall rate, computed as defined in equation (5b).

$$R_{\Delta-} = \frac{R_{\text{peak}} - R_{\text{end}}}{T_{\text{fall}}} \tag{5b}$$

where R_{end} is the SSN at the end of the cycle. Following from equations (5a) and (5b), it is therefore possible (at the onset of the cycle) to estimate the SSN at the end of the cycle using R_{peak} (equation (2)), T_{rise} (equation (3b)), and T_{fall} (equation (4)).

3.1.5. Summary of Regression Application and the Need for a Neural Network Procedure

From the foregoing sections, it is evident that using the regression method (equations (2) to (5a) and (5b)), the characteristics of an SC (e.g., the peak value of SSN for the cycle, the time it takes to reach the peak, the total duration of the cycle, and the value of SSN at the end of the cycle) can be derived. There is, however, emphasis on the fact that reliable predictions for a cycle can be made only after the onset of the cycle. This is because the precursor Ap index is required, and this can only be obtained after the end of the previous cycle to predict/forecast the forthcoming
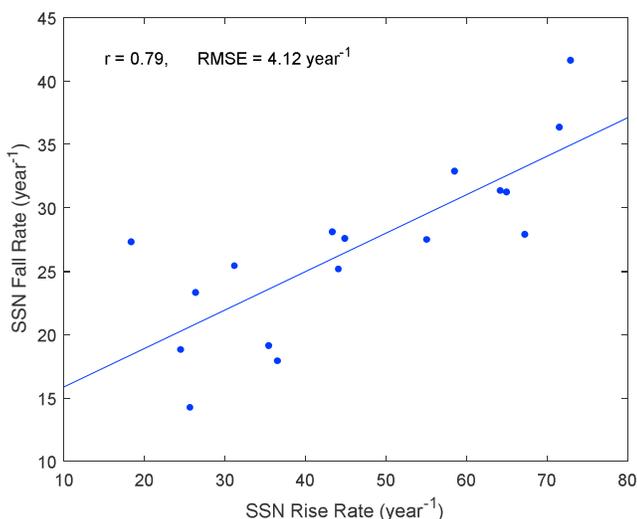
Figure 5. Relationship between the mean rise rate and the mean fall rate for each SC 7–24.

cycle. We emphasize here that the parameters derived using the regression equations were not used as data for the neural network training but were rather used as inputs to the already-trained networks to estimate values of the parameters where observations are not yet made.

While these properties explicitly characterize a SC, they do not provide required information on the progression of SSN from the start of the cycle to its end; there is additional requirement also to be able to predict the value of SSN at any given instance of the cycle. A crude procedure will be to fit this progression using one or more interpolation schemes. There will, however, be enormous considerations to make. For example, the following questions will need to be addressed: What should be the shape of the SC curves? If sinusoids are used, how, and at what locations should the sinusoids be shaped differently since the cycles are not identically shaped? The results may either be oversimplified approximations of the SSNs or exaggeratedly modeled outputs of the SSNs. In either case, the prediction errors will be large.

Hence, a neural network procedure is recommended to accomplish a reliable progression; this is because neural networks have the capability to learn patterns of the SSN variations in time series over the SCs. As a large amount of SSN data, spanning about 200 years (~18 SCs), is available, the neural network can learn the SSN progression patterns over the SCs. Besides building a reliable pattern for the SSN progression in a SC, the neural networks (since they have machine learning capability) will be able to automatically adjust the regression-derived SC properties using information they learn from the training data. So that, for instance, the peak value of the SSN obtained for a cycle using the neural network predictions will not be exactly as that obtained using the regression method; the neural network will adjust the value, within the limits of the regression errors, to conform to information it obtains from the training data.

Several researches (e.g., Ajabshirizadeh et al., 2011; Maris & Oncica, 2006; Uwamahoro et al., 2009) have previously used neural networks in solar physics modeling. A review of these researches(and as also summarized in Pesnell, 2016) shows that time series based neural network training alone is not adequate to obtain reliable results. This is especially true because the available record of SSNs does not contain consistent clues to forecasting the amplitudes and durations of the SCs. For this reason, we supported the neural network model in this work with a regression-guided estimation of the cycle properties that makes use of the precursor value of the *Ap*.

3.2. Neural Network Training Method

Decision about the input parameters to be used for a neural network training largely affects what the network is able to learn and therefore contributes significantly to the accuracy of predictions that will be obtained using the network. Since the SSN has been established to vary in a cyclic period of about 11 years, it is intuitive (and as in nearly all neural network based predictions of the SSN) to include a time series indicator as input for the neural network. To enable the network's learning of time series variations in this work, the year fraction (= year + (month − 1)/12) was used as the first input neuron for the network trainings.

Introducing additional input neurons that characterize each cycle will facilitate the network's learning of the cycle extents (in time and SSN values). To this end, four additional input neurons were introduced to facilitate the networks learning of the cycle extents as described below:

1. The peak of SSN for each cycle: This was introduced to facilitate the networks' learning of the SSN profile heights for each cycle.
2. The rise duration for each cycle: This was introduced to enable the networks' learning of how long it takes to reach the peak of SSN from the cycle onset.
3. The fall duration for each cycle: This was introduced to enable the networks' learning of how long it takes to reach the cycle end from the peak of SSN.
4. A normalized fractional value (which we shall refer to hereafter in this paper as the SC time index, τ_{sc}) defined as given in equation (6).

$$\tau_{sc} = \frac{\text{year fraction} - \text{start year}}{\text{total duration of cycle}} \tag{6}$$

By this definition, τ_{sc} is constrained to have values in the range from 0 (at the beginning of the cycle) to 1 (at the end of the cycle). At any instance in the cycle, τ_{sc} describes and provides an indication of the fraction of
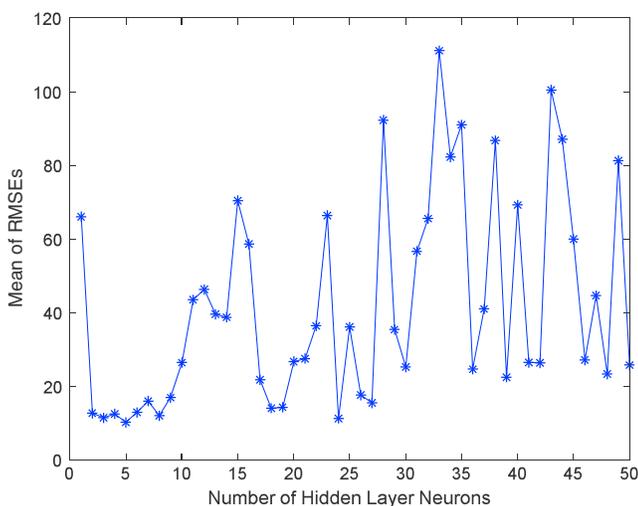
Table 2
Regression and HR-NN Predictions of SC 24 and SC 25 Properties

Method	Start year	Peak year	End year	SSN at start	SSN at peak	SSN at end	Rise duration (years)	Fall duration (years)	Total duration (years)
Regression (SC 24)	2,009.000	2,014.000	2,020.090	2.5	121.1	−0.5	5.000	6.090	11.090
HR-NN (SC 24)	2,009.000	2,013.583	2,020.167	7.4	113.6	5.4	4.583	6.584	11.167
Regression (SC 25)	2,020.250	2,025.043	2,031.266	1.0	132.0	1.1	4.793	6.222	11.015
HR-NN (SC 25)	2,020.250	2,025.000	2,031.250	7.2	122.1	6.0	4.750	6.250	11.000

the cycle that has been completed. This input neuron is introduced to facilitate the neural network's learning of the SSN profile in a progression from the start of a cycle to the end of it. The neural network training procedure is elucidated below:

1. A total of five input neurons were used, namely: the year fraction, the peak of SSN, the rise duration, the fall duration, and the SC time index.
2. Before the trainings, we excluded data for the period of SC 24 from the training data set. This is because of our intention to use the SC 24 data set as test data (to evaluate the performances of the networks after their trainings).
3. A total of 50 neural networks were trained; the difference between them is in the number of hidden layer neurons we applied (we varied the number of hidden layer neurons from 1 to 50). Using a larger number of hidden layer neurons usually leads to better predictions (because the prediction errors will reduce) for data within the range of the training data set. If however, the same network is used to predict data outside the range of the training data set, the errors decrease, and then increase after a certain number of hidden layer neurons. This scenario has been fully discussed later in this section. We define the best network as the one that gives the least prediction error on forecasted data.
4. (a) Each of the trained networks were then used to predict the SSN values for SC 24, and by using corresponding SSN observations for the cycle, we computed the RMSEs using the formula in equation (7).

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\text{prediction}_i - \text{observation}_i\right)^2}{n}} \tag{7}$$

(b) We similarly used each of the trained networks to predict the SSNs at the onset, peak, and end of SC 25. And by referencing corresponding values calculated using the regression equations and the HR-NN, the various characteristics are obtained for SC-24 and 25, respectively, as shown in Table 2.

(c) For each of the 50 trained networks, we computed the mean of the RMSEs obtained in stages 4a and b. Figure 6 shows the mean of the RMSEs vary with the changing number of hidden layer neurons. The mean of the RMSEs for SSN provides performance of a network in forecasting the SSNs for SCs 24 and 25. We found the least RMSE (~10.2) from the network with 5 hidden layer neurons, and hence, we adopted this network.

Figure 6 shows a trend in which the mean of the RMSEs decrease as we increase the number of hidden layer neurons from 1 to 5. And thereafter, the mean of the RMSEs increase with increasing number of hidden layers. This reveals that the network with five hidden layer neurons gives the least prediction errors when used for forecasting. Oscillations in the graph are as a result of random changes in the values of the RMSEs; the RMSEs could get high if the optimization procedure for a neural network leads it to a

Figure 6. Plot showing the variation of the forecasting errors with changing number of hidden layer neurons.
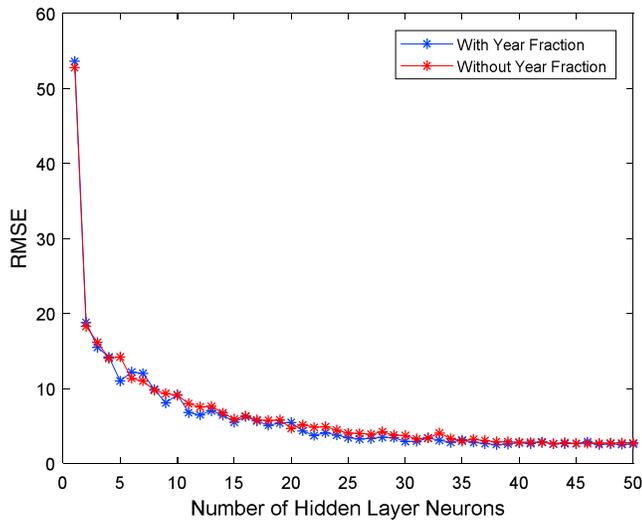
Figure 7. Plot showing the variation of the prediction errors with changing number of hidden layer neurons using a randomly selected test data set.

local minimum (rather than the global minimum). The result is that RMSEs for such networks will be higher. The RMSEs will be lower when the optimization procedure leads the neural network to the global minimum.

To investigate the effect of including the year fraction as an input node, we trained two sets of neural networks that are similar in every way except that one has the year fraction as an input node while the other does not. We tested the networks using a test data set that comprised 15% of randomly selected data which were excluded from the training set. For each set, we trained 50 neural networks (that differ in the number of hidden layer neurons). Figure 7 shows results of how the RMSEs vary with the number of hidden layer neurons. Figure 7 shows that the network with the year fraction included gave an improved performance in terms of reducing the prediction error.

We also use Figure 7 (in comparison to Figure 6) to highlight the effect of using a forecast test data set rather than a randomly selected data set. In Figure 7 (where a randomly selected data set is used), the RMSEs generally keep decreasing as the number of hidden layer neurons increase. This trend suggests that using an excessive number of hidden layer neurons will lead to an improved neural network; this is not correct because using an excessive number of hidden layer neurons will cause the neural network to predict interpolated data so well, whereas the prediction accuracy grows worse for extrapolated data. The randomly selected test data set evidently contains data that are within the time-series range of the training data set. When the neural networks fit the training data set so well, they also fit the interpolated data well; this is especially because of the smoothed SSN data used for the training. However, in using the forecast test data set (Figure 6), the prediction errors are observed to decrease generally to an extent (around when the number of hidden layer neurons is 5–9), and then to generally increase thereafter. Clearly, the figures do also demonstrate the expectation that neural networks perform better on interpolated predictions than on extrapolated predictions; this is why the RMSEs in Figure 6 are relatively higher than those in Figure 7. The scenario we explain is rather one in which, including more number of hidden layer neurons (above some limits), causes the networks to more accurately fit interpolated data than it does for extrapolated data as discussed in Okoh et al. (2015).

Training of the neural networks was done using the Bayesian regularization back-propagation algorithm (Bayes, 1763; Laplace, 1812). The algorithm is implemented within the Levenberg-Marquardt optimization (Levenberg, 1944) as it updates the weights and bias values according to the Levenberg-Marquardt optimization (https://in.mathworks.com/help/nnet/ref/trainbr.html). The Bayesian regularization back-propagation algorithm is admired for its robustness and has been shown to yield better results when compared to standard back-propagation training algorithms (Aggarwal et al., 2005; Burden & Winkler, 2008). We used the MATLAB programming for the development and implementation of the HR-NN model, and the complete program has been made available at Mathworks Central (https://in.mathworks.com/matlab-central/fileexchange/65686-sunspot-number-prediction-forecasting-via-a-hybrid-regression-neural-network-based-model).

All the observed SSN data from SC 7 to 23 are used for the neural network training, and the data of SC 24 are used for testing. The Bayesian regularization algorithm (which is used in this work) does not require a validation data set, as it has its own form of validation built into the algorithm and so no validation data set is used in this work. The point of checking validation is to see if the error on the validation set gets better or worse as training progresses, and the idea is to stop training if the error gets worse. However, the Bayesian error is not just based on how well the model is performing on the data set; it is also based on how large the weights are. The larger the weights, the higher the error. So throughout training, the Bayesian regularization algorithm does not let the network explore larger weights, even though larger weights may lead to the global minimum errors (Foresee & Martin, 1997; MacKay, 1992).

In the MATLAB implementation of this algorithm, the validation stops are disabled by default (Mathworks, 2018a). Normalization of the training data was done using the mapminmax processing function, which is
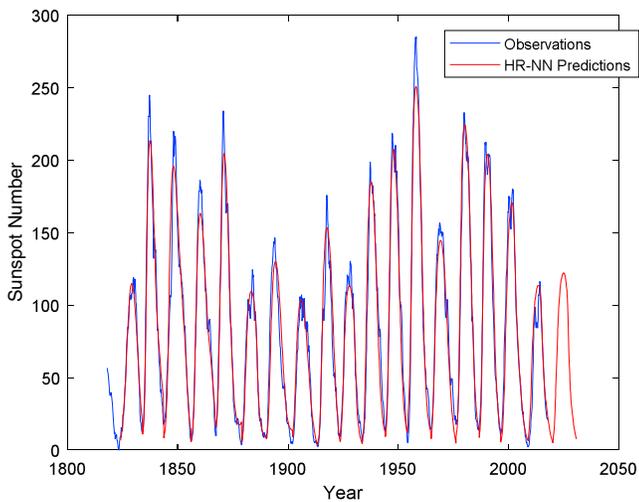
Figure 8. Sunspot number observations and predictions for solar cycles 7–24. Prediction of solar cycle 25 is illustrated based on an estimation of the precursor *Ap* index value as 5.6 nT.

default for the MATLAB Bayesian regularization training algorithm used in this work. The mapminmax function normalizes the training data so that inputs fall in the range [−1, 1] by mapping the minimum and the maximum values to −1 and 1, respectively(Mathworks, 2018b). For the year fraction input node, we intentionally set the minimum and maximum values to 1,818 and 2,031, respectively, to cover the time frame within which the neural network can be used to for SSN predictions. Monthly values of SSN were used for the neural network training, but the final neural network model also allows prediction/forecasting of daily SSN values.

4. Discussion

4.1. Predicting SC 24

Since the greater part of SC 24 has already been witnessed, it may not seem significant to predict SSNs for the cycle. We, however, emphasize that since we did not include data for SC 24 in the neural network training data set, the neural network in this study was used to predict the SSNs for SC 24 in a forecasting mode. And additionally, the neural network is actually used to forecast the SSNs for the remaining part of SC 24.

To predict SC 24, equations (2) to (5a) and (5b) were first used to derive the cycle properties as listed in Table 2. The observed *Ap* index at the end of SC 23 (= 4.8 nT) was used as the precursor *Ap* index, and the observed SSN (= 2.5) was used as the onset value of SSN for the cycle SC 24. Table 2 also summarizes the final SC 24 properties obtained from the HR-NN model. We found that the peak of SSN predicted by using the regression method was 121.1, while the HR-NN model gave a value of 113.6. Whereas the observed peak value of SSN for the cycle is 116.4 (see Table 1). Therefore, the regression prediction is about 5 units greater than the observation, while the HR-NN prediction is only about 3 units lower. Although the difference is not significant, the scenario shows that the neural network procedure can be used to reduce over-shooting of the cycle peaks as estimated using the regression equations.

Another striking feature to note in Table 2 is that the SSN at the end of SC 24 is predicted to be negative; by definition, the SSN should always be positive, and at least zero. This is one example of shortfall in applying regression techniques alone. This problem was taken care of in this work by applying hyperbolic tangent sigmoid transfer functions during the neural network trainings. Transfer functions are used in neural network trainings to compute next layer values from the previous layer values. The hyperbolic tangent sigmoid transfer function or the log-sigmoid transfer function is usually used to avoid negative values in the output of a neural network, especially in cases when the targets are bounded for a physical or mathematical reason (Heath, 2014). This transfer function provided better predictions of the SSN profiles (in terms of shape and prediction accuracy) than the log-sigmoid transfer function.

Figure 8 illustrates the comparison of predictions from the HR-NN model and the observations for SCs 7–24. The values for SC 25 were forecasted (as explained in the next section) based on an estimated precursor *Ap* index for the cycle as 5.6 nT. The RMSEs for predicting the peak and end values of SSNs for SCs 7–24 are 18.2 and 5.5, respectively, and for predicting the rise and fall durations are respectively 6 and 5 months. The maximum error observed in predicting the total duration was 7 months.

4.2. Forecasting SC 25: An Illustration

Again, we emphasize here that the predictions for SC 25 were strictly done for illustrative purpose; the precursor *Ap* index value was estimated to be 5.6 nT by investigating the mean rate of change of the *Ap* index during the last 3 years of SCs 20 to 23 as follows. As the final minimum value of *Ap* will not be known until the very solar minimum, we investigated the phases of SSN and magnetic activity index (*Ap*) during years 1964 to 2017 as shown in Figure 9. The minima in geomagnetic activity tend to occur just after those for the SSN and the geomagnetic activity tends to remain high during the declining phase of each cycle (Hathaway, 2015). Feynman (1982) decomposed the geomagnetic variability into two components—one which is proportional to and in phase with the sunspot cycle (relative SSN component) and the other out of phase with the sunspot
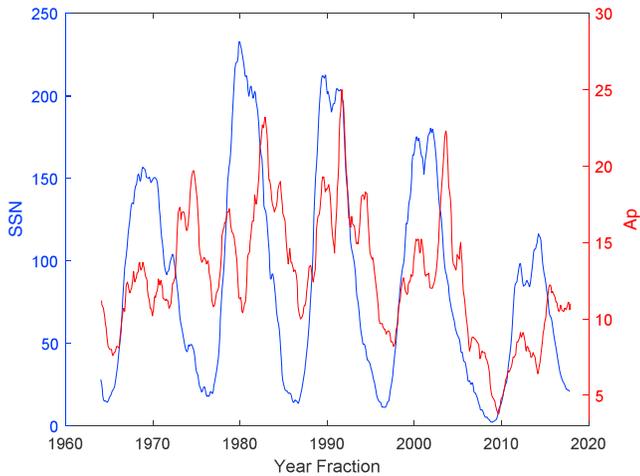
Figure 9. Plot of smoothed sunspot number an *Ap* index values from years 1964 to 2017.

cycle (interplanetary component). The author found a baseline level (with 5.38 as the intercept) of geomagnetic activity that increases with the increase in the SSN. We found from Figure 3 that *Ap* index generally decreases for 3 years during the end of the SCs with a rate of ~1.7 nT/yr. Hence, considering our predicted end of cycle SC 24 to be in March 2020 and *Ap* index during March 2017 as 10.7 nT, we obtained the estimated *Ap* index at the end of SC 24 as 5.6 nT (10.7 − [3 × 1.7 for 3 years]). The onset value of SSN for the cycle was also arbitrarily assumed to 1, which is the minimum it can be. More reliable predictions of the cycle properties can be obtained when we know the actual precursor *Ap* index value and the onset value of SSN. These values can only be known at the end of SC 24. The parameters derived in this section are therefore strictly for illustrative purpose.

Using the above mentioned assumptions, parameters for SC 25 are derived using equations (2) to (5a) and (5b) as shown in Table 2. The parameters (year fraction, peak of SSN, rise duration, fall duration, and SC time index) are further fed into the neural network model to obtain the SSN profile for SC 25 as illustrated in Figure 8. The HR-NN predictions in Figure 8 show that the peak of SSN for SC 25 will be about 122 and will occur in January 2025. Details of parameters for SC 25 as predicted using the HR-NN model are summarized in Table 2.

As the value of the precursor *Ap* index may change from the estimated value of 5.6 nT, we use the HR-NN model to simulate four other scenarios in which the precursor *Ap* index is varied in steps of 0.5 nT around the 5.6-nT value; that is for cases where the precursor *Ap* indices are 4.6, 5.1, 6.1, and 6.6 nT. The results are illustrated in Figure 10. The figure shows that variations in the precursor *Ap* index value will significantly alter the SSN profile for the cycle. The SSN values at the cycle peaks are respectively 111, 117, 128, and 133 for the four scenarios, and occur in the months of September 2024, January 2025, February 2025, and January 2025, respectively. The values show that the predicted peak of SC 25 changes by about 11 sunspots for every 1-nT change in the assumed precursor *Ap* index. Recently, Helal and Galal (2013) forecasted the SC 25 using the precursor technique of spotless events and obtained maximum amplitude and time of rise to be 118.2 and 4.0 years, respectively. While their rise time is only about 9 months different from our value of 4.75 years, simulations from our HR-NN model indicate that the peak value of SSN will agree with their value of ~118 only if the precursor *Ap* index value gets as low as ~5.2 nT.
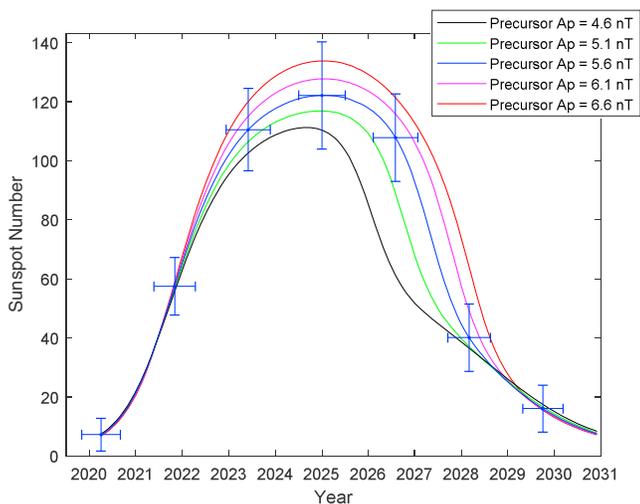
Figure 10. Sunspot number predictions for solar cycle C 25 using the Hybrid Regression-Neural Network model with precursor *Ap* index values varied from 4.6 to 6.6 nT in steps of 0.5 nT.

Recently, using the Advective Flux Transport code, Hathaway and Upton (2016) simulated the evolution of the Sun's polar magnetic fields from early 2016 to the end of 2019 (which is near the expected time of Cycle 24/25 minimum). They found that the average strength of the polar fields near the end of Cycle 24 will be similar to that measured near the end of Cycle 23, and so gave the indication that Cycle 25 will be similar in strength to the current cycle. Similar results were obtained by Cameron et al. (2016) using surface flux transport simulations for the descending phase of Cycle 24 and estimated the value of the dipole moment around year 2020 as 2.5 ± 1.1 G. They observed that this value is comparable to that observed at the end of Cycle 23 (which is about 2 G), and they therefore suggest that Cycle 25 will be of moderate amplitude, not much higher than that of the current cycle. The conclusions from these physics based models are similar to the predictions from our neural network model, and so our method offers a lightweight empirical method that is in agreement with the physics based models. By using some classical statistical relations among feature parameters of SC profiles, Li et al. (2015) predicted that the start of SC 25 will be in November 2019 and that the maximum amplitude of SC 25 will be 109.1 in October 2023. Table 3 (updated from Li et al., 2015) presents a summary of the predictions of the peak of SC 25 from other

Table 3
Predictions of the Peak of SC 25 From Previous Studies

Reference	SSN at peak of cycle	Year of cycle peak
Chistyakov (1983)	121	2,028.5
Kontor et al. (1983)	117	2,024
Du (2006)	102.6 ± 22.4	
Du and Du (2006)	111.6 ± 17.4	
Quassim et al. (2007)	116	2,020
Hiremath (2008)	110 ± 11	2,023
Pishkalo (2008)	112.3 ± 33.4	2,023.4 ± 0.7
Rigozo et al. (2011)	132.1	2,023.3
Abdel-Fattah et al. (2013)	90.7 ± 8	2,020
Hamid and Galal (2013)	118	2,022–2,023
Li et al. (2015)	109.1	2,023.75

previous studies. These predictions consistently indicate that the peak of SC 25 will be moderate when compared to the peaks of the previous cycles. And these results agree with that of the present study.

A major limitation for using the HR-NN model to make SSN predictions for a given SC is that it relies on the precursor Ap index value, which can only be measured at the end of the previous cycle. This means that the model can be reliably used for predictions in a given SC only after the end of the cycle. One way to overcome this limitation is to depend on reliable forecasts of the Ap index. At the moment, the Space Environment Prediction Center, Center for Space Science and Applied Research (http://eng.sepc.ac.cn/index.php) uses an autoregressive method to make 27-day forecasts of the Ap index at http://eng.sepc.ac.cn/ApForecast.php. The short-term nature of the forecasts, however, still places a limitation to the duration over which the forecasted Ap indices are available. The highly erratic nature of geomagnetic activity has made long-term forecasting of the Ap index a challenging task, given the present-day level of available data and information from space science studies.

5. Conclusions

The SSN is an important physical parameter that characterizes the solar activity, and efficient space weather predictions require accurate forecasts of SSN. In this research work, a HR-NN model is presented for forecasting SSN values for upcoming cycle 25 using precursor Ap. The method of regression is used to derive relationships between various parameters of a SC, and then used to forecast parameters for the upcoming SCs. The obtained parameters are further fed as inputs for a neural network procedure to obtain instantaneous predictions of the SSN values.

The neural network procedure was recommended to harness the huge amount of SSN data available for learning the SSN progression patterns over the SCs. And given their machine learning capabilities, the neural networks are also able to automatically adjust the regression-derived SC properties using information they learn from the training data. Five input neurons (year fraction, peak of SSN, rise duration, fall duration, and SC time index) were used for the neural network training.

The HR-NN model was also used to forecast SSN values for the remaining part of SC 24 at a RMSE value of 3.5, and to give indications of the expectations for SC 25. Forecasts by the model show that total duration of SC 24 will be 11.167 years (±7 months), and the end of the cycle will be in March 2020 (± 7 months).

Using an estimated precursor Ap index of 5.6 nT for SC 25, we found the peak SSN ~122.1 (±18.2) of the cycle to occur in January 2025 (±6 months) with a total duration of 11 years (±7 months). Further simulations of the SSNs by varying the precursor Ap index between 4.6 and 6.6 nT showed that peak SSNs for SC 25 will change by about 11 units for every 1-nT change in the assumed precursor Ap index.

The major limitation for the model is that it depends on the Ap index value at the end of one cycle to make SSN predictions for the next cycle, and so reliable SSN predictions can be obtained at the end/onset of the previous/upcoming cycles when the precursor Ap index values have been measured.

References

Abdel-Fattah, A., Hamed, A. I., & Hassan, M. B. (2013). A neuro-fuzzy modeling for prediction of solar cycles 24 and 25. *Astrophysics and Space Science*, *344*(1), 5–11.

Aggarwal, K. K., Singh, Y., Pravin, C., & Puri, M. (2005). Bayesian regularization in a neural network model to estimate lines of code using function points. *Journal of Computer Science*, *1*(4), 505–509.

Ajabshirizadeh, A., Masoumzadeh, J. N., & Abbassi, S. (2011). Neural network prediction of solar cycle 24. *Research in Astronomy and Astrophysics*, *11*(4), 491–496.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Bhatt, N. J., Jain, R., & Aggarwal, M. (2009a). Predicting maximum sunspot number in solar cycle 24. *J. Astronomy & Astrophysics*, *30*(1), 71–77.

Bhatt, N. J., Jain, R., & Aggarwal, M. (2009b). Prediction of the maximum amplitude and timing of sunspot cycle 24. *Solar Physics*, *260*(1), 225–232.

Burden, F., & Winkler, D. (2008). Bayesian regularization of neural networks. In D. J. Livingstone (Ed.), *Artificial neural networks. Methods in molecular biology* (Vol. 458, pp. 23–42). New York City: Humana Press. https://doi.org/10.1007/978-1-60327-101-1_3

Cameron, R. H., Jiang, J., & Schüssler, M. (2016). Solar cycle 25: Another moderate cycle? *The Astrophysical Journal*, *823*(2), L22. https://doi.org/10.3847/2041-8205/823/2/L22

Chistyakov, V. F. (1983). A forecast of solar activity till the year 2030. *Solnechnye Dannye*, *1*, 97–100.

Clette, F., Cliver, E. W., Lefèvre, L., Svalgaard, L., & Vaquero, J. M. (2015). Revision of the sunspot number (s). *Space Weather*, *13*, 529–530. https://doi.org/10.1002/2015SW001264

Cliver, E. W., Boriakoff, V., & Feynman, J. (1998). Solar variability and climate change: Geomagnetic aa index and global surface temperature. *Geophysical Research Letters*, *25*, 1035–1038. https://doi.org/10.1029/98GL00499

Cliver, E. W., Ling, A. G., Wise, J. E., & Lanzerotti, L. J. (1999). A prediction of geomagnetic activity for solar cycle 23. *Journal of Geophysical Research*, *104*, 6871–6876. https://doi.org/10.1029/1998JA900081

Conway, A. J. (1998). Time series, neural networks and the future of the Sun. *New Astronomy Reviews*, *42*(5), 343–394.

Dikpati, M., de Toma, G., & Gilman, P. A. (2006). Predicting the strength of solar cycle 24 using a flux-transport dynamo-based tool. *Geophysical Research Letters*, *33*, L05102. https://doi.org/10.1029/2005GL025221

Du, Z., & Du, S. (2006). The relationship between the amplitude and descending time of solar activity cycle. *Solar Physics*, *238*(2), 431–437.

Du, Z. L. (2006). Relationship between solar maximum amplitude and max-max cycle length. *Astronomy Journal*, *132*(4), 1485–1489.

Duhau, S. (2003). An early prediction of maximum sunspot number in solar cycle 24. *Solar Physics*, *213*(1), 203–212.

Eddy, J. A. (1981). In R. I. Rotberg & T. K. Rabb (Eds.), *Climate and the role of the Sun, in climate and history: Studies in interdisciplinary history* (p. 145). Princeton, NJ: Princeton University Press.

Feynman, J. (1982). Geomagnetic and solar wind cycles, 1900–1975. *Journal of Geophysical Research*, *87*, 6153–6162. https://doi.org/10.1029/JA087iA08p06153

Foresee, F. D., & Martin, T. H. (1997). Gauss-Newton approximation to Bayesian learning. Proceedings of the International Joint Conference on Neural Networks.

Gholipour, A., Lucas, C., Araabi, B. N., & Shafiee, M. (2005). Solar activity forecast: Spectral analysis and neurofuzzy prediction. *Journal of Atmospheric and Solar-Terrestrial Physics*, *67*(6), 595–603. https://doi.org/10.1016/j.jastp.2004.12.001

Gonzalez, G., & Schatten, K. H. (1988). Using geomagnetic indices to forecast the next sunspot maximum. *Solar Physics*, *114*(1), 189–192.

Habarulema, J. B. (2010). A contribution to TEC modelling over Southern Africa using GPS data (PhD thesis). Rhodes University.

Hamid, R. H., & Galal, A. A. (2013). An early prediction of the maximum amplitude of the solar cycle 25. *Journal of Advanced Research*, *4*, 275.

Hathaway, D. H. (2010). The solar cycle. *Living Reviews in Solar Physics*, *7*(1), 65.

Hathaway, D. H. (2015). The solar cycle. *Living Reviews in Solar Physics*, *12*(1), 87.

Hathaway, D. H., & Upton, L. A. (2016). Predicting the amplitude and hemispheric asymmetry of solar cycle 25 with surface flux transport. *Journal of Geophysical Research: Space Physics*, *121*, 10,744–10,753. https://doi.org/10.1002/2016JA023190

Hathaway, D. H., & Wilson, R. M. (2006). Geomagnetic activity indicates large amplitude for sunspot cycle 24. *Geophysical Research Letters*, *33*, L18101. https://doi.org/10.1029/2006GL027053

Heath, H. (2014). How can I avoid negative values in the output of a feedforward network? Retrieved from https://in.mathworks.com/matlabcentral/answers/158570-how-can-i-avoid-negative-values-in-the-output-of-a-feedforward-net

Helal, R. H., & Galal, A. A. (2013). An early prediction of the maximum amplitude of the solar cycle 25. *Journal of Advanced Research*, *4*(3), 275–278. https://doi.org/10.1016/j.jare.2012.10.002

Herbrich, R., Keilbach, M., Graepel, T., Bollmann-Sdorra, P., & Obermayer, K. (1999). Neural networks in economics: Background, applications, and new developments. *Advances in Computational Economics*, *11*, 169–196.

Hiremath, K. M. (2008). Prediction of solar cycle 24 and beyond. *Astrophysics and Space Science*, *314*(1-3), 45–49.

Hoyt, D. V., & Schatten, K. H. (1998). Group sunspot numbers: A new solar activity reconstruction. *Solar Physics*, *181*(2), 491–512.

Jiang, Y., & Xu, Z. (1986). On the Spörer Minimum. *Astrophysics and Space Science*, *118*, 159–162. https://doi.org/10.1007/BF00651121

Kiepenheuer, K. O. (1953). Solar activity. In G. P. Kuiper (Ed.), *The Sun* (pp. 322–345). Chicago: The University of Chicago Press.

Kirkwood, D. (1869). On the periodicity of the solar spots. *Proceedings of the American Philosophical Society*, *11*(81), 94–102.

Kontor, N. N., Lyubimov, G. P., Pereslegina, N. V., & Khotilovskaya, T. G. (1983). A prediction of the sunspot maxima for solar cycles NN 22-44. *Solnechnye Dannye*, *11*, 74–79.

Laplace, P. S. (1812). *Theorieanalytique des probabilities*. Paris: Ve. Courcier.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, *2*(2), 164–168.

Li, K. J., Feng, W., & Li, F. Y. (2015). Predicting the maximum amplitude of solar cycle 25 and its timing. *Journal of Atmospheric and Solar-Terrestrial Physics*, *135*, 72–76.

MacKay, D. (1992). Bayesian interpolation. *Neural Computation*, *4*(3), 415–447.

Maris, G., & Oncica, A. (2006). Solar cycle 24 forecasts. *Sun and Geosphere*, *1*(1), 8–11.

Mathworks (2018a). Bayesian regularization back-propagation. Retrieved from https://in.mathworks.com/help/nnet/ref/trainbr.html

Mathworks (2018b). Configure neural network inputs and outputs. Retrieved from https://in.mathworks.com/help/nnet/ug/configure-neural-network-inputs-and-outputs.html

McKinnon, J. A., & Waldmeier, M. (1987). Sunspot numbers, 1610–1985: Based on "The sunspot activity in the years 1610–1960". World Data Center A for Solar-Terrestrial Physics. Boulder, CO. Report UAG, 0579–-7144.

Ohl, A. I., & Ohl, G. I. (1979). A new method of very long-term prediction of solar activity. In R. Donnelly (Ed.), *Solar-terrestrial predictions proceedings 2* (pp. 258–263). Boulder, CO: NOAA/Space Environment Laboratory.

Okoh, D., Owolabi, O., Ekechukwu, C., Folarin, O., Arhiwo, G., Agbo, J., et al. (2016). A regional GNSS-VTEC model over Nigeria using neural networks: A novel approach. *Geodesy and Geodynamics*, *7*(1), 19–31.

Okoh, D., Yusuf, N., Adedoja, O., Musa, I., & Rabiu, B. (2015). Preliminary results of temperature modelling in Nigeria using neural networks. *Weather*, *70*(12), 336–343.

Pesnell, W. D. (2008). Predictions of solar cycle 24. *Solar Physics*, *252*(1), 209–220.

Pesnell, W. D. (2016). Predictions of solar cycle 24: How are we doing? *Space Weather*, *14*, 10–21. https://doi.org/10.1002/2015SW001304

Pesnell, W. D. (2017). Solar cycle predictions. Retrieved from https://ccmc.gsfc.nasa.gov/RoR_WWW/SWREDI/2017/pesnell_SC_Pred_GSFC_SWx_Jun_2017.pdf

Pishkalo, M. I. (2008). Preliminary prediction of solar cycles 24 and 25 based on the correlation between cycle parameters. *Kinematika i Fizika Nebesnykh*, *24*, 370.

Quassim, M., Fattia, A., & Elminir, H. (2007). Forecasting the peak amplitude of the 24th and 25th sunspot cycles and accompanying geomagnetic activity. *Solar Physics*, *243*(2), 253–258. https://doi.org/10.1007/s11207-007-0447-8

Ramesh, K. B. (2010). Coronal mass ejections and sunspots—Solar cycle perspective. *The Astrophysical Journal Letters*, *712*(1), L77–L80.

Rigozo, N. R., Souza, E. M. P., Evangelista, H., Nordemann, D. J. R., & Echer, E. (2011). Prediction of sunspot number amplitude and solar cycle length for cycles 24 and 25. *Journal of Atmospheric and Solar - Terrestrial Physics*, *73*(11–12), 1294–1299. https://doi.org/10.1016/j.jastp.2010.09.005

Ruzmaikin, A. A. (1985). The solar dynamo. *Solar Physics*, *100*(1-2), 125–140. https://doi.org/10.1007/BF00158425

Svalgaard, L., Cliver, E. W., & Kamide, Y. (2005). Sunspot cycle 24: Smallest cycle in 100 years? *Geophysical Research Letters*, *32*, L01104. https://doi.org/10.1029/2004GL021664

Thompson, R. J. (1993). A technique for predicting the amplitude of the solar cycle. *Solar Physics*, *148*(2), 383–388.

Uwamahoro, J., & McKinnell, L. A. (2013). Solar and interplanetary precursors of geomagnetic storms in solar cycle 23. *Advances in Space Research*, *51*(3), 395–410.

Uwamahoro, J., McKinnell, L. A., & Cilliers, P. J. (2009). Forecasting solar cycle 24 using neural networks. *Journal of Atmospheric and Solar-Terrestrial Physics*, *71*(5), 569–574. https://doi.org/10.1016/j.jastp.2008.12.003

Waldmeier, M. (1961). *The sunspot-activity in the years 1610–1960*. Zurich: Schulthess.

Wilson, R. M., Hathaway, D. H., & Reichmann, E. J. (1998). An estimate for the size of cycle 23 based on near minimum conditions. *Journal of Geophysical Research*, *103*, 6595–6603. https://doi.org/10.1029/97JA02777