Astronomy & Astrophysics

Kodaikanal digitized white-light data archive (1921–2011): Analysis of various solar cycle features[*]

Sudip Mandal[1], Manjunath Hegde[1], Tanmoy Samanta[1], Gopal Hazra[1,2], Dipankar Banerjee[1,3], and Ravindra B.[1]

[1] Indian Institute of Astrophysics, Koramangala, 560034 Bangalore, India
  e-mail: `sudip@iiap.res.in`
[2] Department of Physics, Indian Institute of Science, 560012 Bangalore, India
[3] Center of Excellence in Space Sciences India, IISER Kolkata, 741246 Mohanpur, West Bengal, India

ABSTRACT

Context. Long-term sunspot observations are key to understanding and predicting the solar activities and its effects on space weather. Consistent observations, which are crucial for long-term variations studies, are generally not available due to upgradation/modification of observatories over the course of time. We present data for a period of 90 yr acquired from persistent observation at the Kodaikanal observatory in India.
Aims. We aim to build a uniform sunspot area time series along with their positions for a 90-yr period between 1921 and 2011, as obtained from the newly digitized and calibrated white-light images from the Kodaikanal observatory. Our aim is to compare this new time series with known sources and confirm some of the earlier reported results with additional new aspects.
Methods. We use an advanced semi-automated algorithm to detect the sunspots form each calibrated white-light image. Area, longitude and latitude of each of the detected sunspots are derived. Implementation of a semi-automated method is extremely necessary in such studies as it minimizes the human bias in the detection procedure.
Results. Daily, monthly, and yearly sunspot area variations, obtained from the Kodaikanal, compared well with the Greenwich sunspot area data. We find an exponentially decaying distribution for the individual sunspot area for each of the solar cycles. Analyzing the histograms of the latitudinal distribution of the detected sunspots, we find Gaussian distributions, in both the hemispheres, with centers at ~15° latitude. The height of the Gaussian distributions are different for the two hemispheres for a particular cycle. Using our data, we show clear presence of Waldmeier effect, which correlates the rise time with the cycle amplitude. Using the wavelet analysis, we explored different periodicities on different time scales present in the sunspot area times series.

Key words. Sun: activity – sunspots – Sun: magnetic fields – Sun: photosphere – catalogs

1. Introduction

The Sun has a strong and complex magnetic field that is very dynamic and changes over various time scales. The magnetic activity governs the solar radiative energy, radio flux, coronal mass ejections (CMEs), solar wind, heliospheric magnetic field and, consequently, cosmic ray flux at Earth. Sunspots are the enhanced magnetic patches on the solar photosphere and act as an excellent proxy for the eleven-year solar cycle and the magnetic activities. The number of sunspots observed on the solar surface varies periodically, going through strong successive maxima and minima with a periodicity of approximately 11 yr, known as the "sunspot cycle" or "solar cycle". Sunspot observations in modern era were first carried out by Galileo Galilei using a telescope, though there are reports of sunspot observations using the naked eye even before this (see Vaquero 2007, for an exhaustive review on this). Sunspots that are clearly visible in white-light images are recorded by different observatories around the world during different epochs of the last century. The Royal Observatory of Greenwich (RGO) has a white-light data archive with observations from 1874 to 1976. This database has continued to be updated since 1976 from the Solar Optical Observing Network

(SOON) observations. (RGO officially transferred the sunspot observing program to the Debrecen Heliophysical Observatory in Hungary toward the end of 1976.)

Kodaikanal Solar Observatory[1] (Hasan et al. 2010) has been observing the Sun since 1904. This century-long observation has been carried out in three different wavelength bands; white-light (since 1904), Ca-K (since 1907), and in H-alpha (since 1912). Among these three wavelengths, white-light observations have been taken without any changes in the telescope optics since 1918. This makes the data set very efficient for long-term solar studies. Kodaikanal white-light images have been used in the past by many authors to study the time evolution of different sunspot parameters. Sivaraman et al. (1993) used the white-light data for a period of 45 yr (1940–1985) and compared their results with Mount Wilson white-light data. In the subsequent years, various aspects of the sunspot and their variation with the solar cycles have been studied using this data set. Solar rotation rates (Howard et al. 1999; Gupta et al. 1999), axial tilt of the sunspot groups (Sivaraman et al. 1999), tilt angle and size variation of the sunspot groups (Howard et al. 2000), and rotational rate variation with the age of the sunspot groups (Sivaraman et al. 2003) are some of the highlighted works produced from this data set.

---
[1] Official website of the observatory is `https://kso.iiap.res.in//`

Sunspots are closely linked to large-scale events such as solar flares and CMEs, which directly affect the space-weather and earth's atmosphere. Each sunspot cycle is unique in terms of the cycle amplitude and the duration. This makes it interesting to predict the forthcoming solar activities by analyzing a particular cycle. Long-term solar studies are thus a unique opportunity to establish any empirical relations between different parameters of the solar cycles and enhance the accuracy of predictions about following cycles. In this paper, we present the results from the digitized and calibrated Kodaikanal white-light sunspot data, which covers the period 1921 to 2011. Sections 2 and 3 briefly explain the details of the Kodaikanal white-light telescope and the digitization of the photographic plates, whereas Sect. 4 describes the semi-automated algorithm used for sunspot detection. Different aspects of the sunspot area time series, such as daily sunspot variation (Sect. 5.1), north-south asymmetry (Sect. 5.2.1), and latitudinal distributions (Sect. 5.2.2) are presented in the subsequent sections. We discuss the correlations of the different cycle parameters with the cycle amplitude in Sect. 6 and explore different periodicities present in the sunspot area time series in Sect. 7. A summary is presented and conclusions are drawn in Sect. 8.

2. Telescope and photographic plates

Regular observations at the Kodaikanal observatory began in 1904 using a white-light telescope with a 10-cm aperture lens and a f/15 light beam. Between 1912 and 1917, the objective lens was changed several times and in 1918, a 15-cm achromatic lens was installed. This new configuration produced a 20.4 cm size image of the Sun in the image plane. Photographic plates were used to capture the image. The same telescope has been used since 1918 up until 2011 (observations with this telescope continue to be made even today but those images are yet to be calibrated and thus they are not part of the data used in this paper) to take regular white-light observation of the sun. More details of the telescope optics and its initial changes are described in Sivaraman et al. (1993).

Initial white-light images from 1904 were stored in Ilford special Lantern photographic plates, which had a size of 25.4 cm × 25.4 cm. In 1975, these plates became unavailable and thus the old plates were replaced with high-contrast film of the same size. These photographic plates and the films containing the solar images have been stored carefully in individual paper envelopes at the observatory. Also, a logbook, containing the observational details for each observation, has been carefully preserved.

3. Digitization and calibration of the white-light images

Normally, the observations are taken in the morning before 10 IST when the seeing conditions at the site is good (see Bappu 1967 for details of observing conditions at Kodaikanal observatory). In most cases, we have one image per day. For a better use of the scientific data, a new program for high-resolution digitization and calibration of the solar images recorded in the photographic plates and films was initiated at Kodaikanal in the last five years. The first representative results (for three solar cycles) from this digitization were presented in Ravindra et al. (2013). The raw images have been digitized by a 4k × 4k scientific grade CCD camera using a uniform source sphere digitizer unit. Calibration of the digitized data includes flat-fielding, conversion from emulsion density to relative density, identification

of solar radius and the Sun center along with rotation correction (see Ravindra et al. 2013, for details of each process). Calibrated white-light data covers the period from 1920 to 2011. This includes solar cycle 16 to cycle 24 (still ongoing). On average, 290 plates per year have been acquired. Here we must emphasize the fact that though the optics has remained unchanged for the past 90 yr, the image quality may have gone through several changes over this period for various reasons, such as variable seeing, change of storage device from photographic plates to films, and so on. These aspects of the time series for the Kodaikanal white-light digitized images have been studied thoroughly by Ravindra et al. (2013).

We would also like to mention that for the period from 1905 to 1920, there are some issues with the raw images, such as two perpendicular lines on the top of the images (one for N-S and another for E-W line of the Earth) and variation in image size. These cause difficulty in correctly orientating the images. This part of the calibration of 15 yr, starting from 1905 to 1920, could not be handled by the current version of the code and has to be handled separately, a task that we hope to take up in near future.

4. Sunspot detection

Sunspot area varies over the different epochs of a solar cycle. This increment and decrement of the sunspot area with the solar maxima and minima has been analyzed in the past. In this paper, for the first time, we are presenting the sunspot area variation for almost 90 yr using Kodaikanal white-light digitized data. To study the change in the sunspot area with the solar cycle, we detected the sunspots using a semi-automated algorithm.

We used a modified version of the "sunspot tracking and recognition algorithm" (STARA, Watson et al. 2009) to identify the sunspots on our calibrated white-light images. This modified version of the code uses two intensity thresholds, in contrast to a single threshold used in the original code. The working principle of this code is explained in detail in Ravindra et al. (2013). Although this procedure detects the sunspots automatically but occasionally picks up dirt and the east-west line in the image. These artifacts have been removed manually by checking every contour drawn on the detected spots and using a cursor to rejects those which do not belong to a sunspot. Also, we have modified the code to get the position information (longitude and latitude from the "center of gravity" method) of the detected sunspots, which was not implemented in the code used in Ravindra et al. (2013).

Figure 1 shows a representative image of Kodaikanal white-light data where the detected sunspots are shown with the contours overplotted on them. In the inset of Fig. 1, we show a zoomed-in view of one of the sunspots in the image. In this zoomed-in image, we notice the umbral and the penumbral area boundaries along with the contour; which covers the whole-spot area perfectly. We refer to the "whole-spot" area as the "sunspot area" hereafter.

5. Results

5.1. Sunspot area variation

Using the method described above, we detected the sunspots using the Kodaikanal white-light images within a period from 1921 to 2011. For every detected sunspot, we have the area, longitude and latitude information as an output from our modified detection algorithm. Panel a) in Fig. 2 shows the daily sunspot area obtained from Kodaikanal. From the plot, we immediately recover the 11-yr sunspot cycle. Also, we notice that the individual
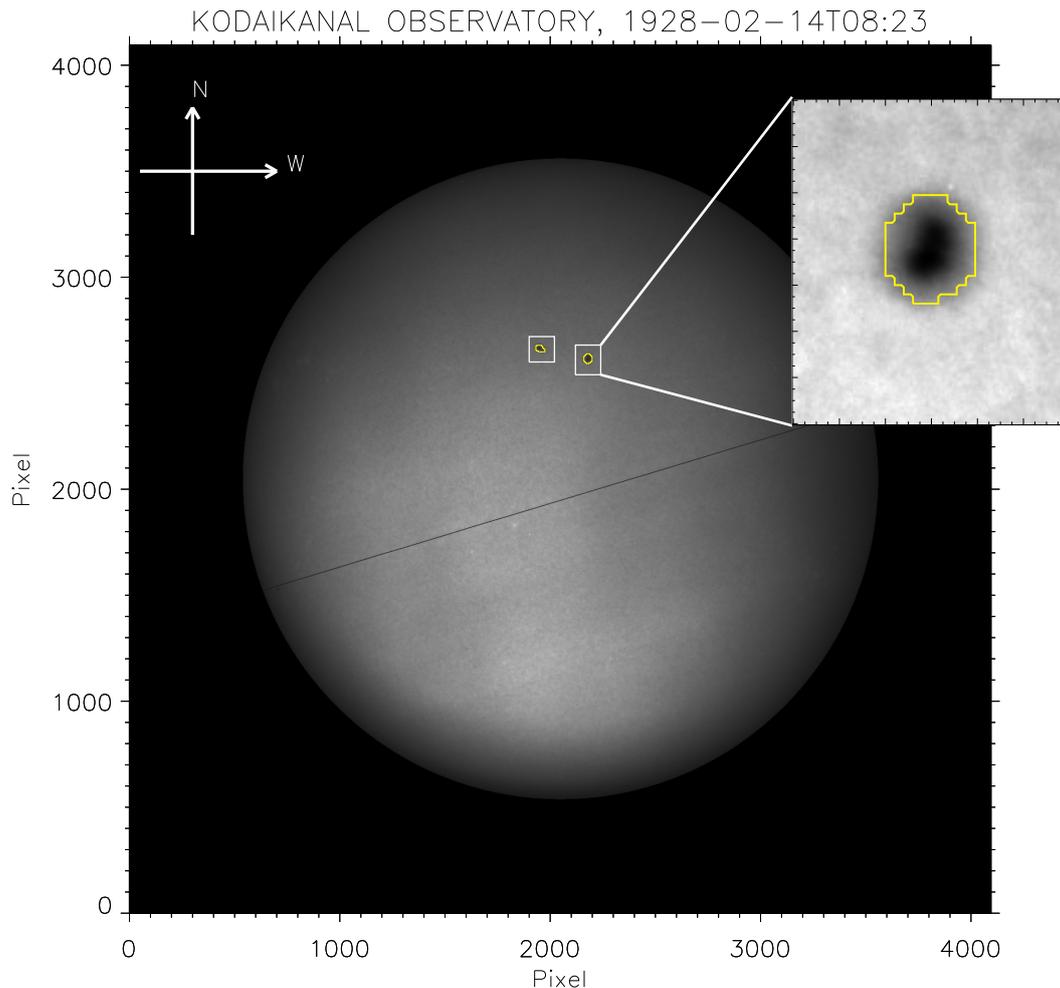
Fig. 1. A representative image from the Kodaikanal white-light digitized data. Two detected sunspots are highlighted by two white rectangular boxes. In the *inset*, we show a zoomed-in view of one of the sunspots. The yellow contours represent the detected area for the sunspots.

cycle amplitudes vary over the entire time series. Now to compare our results, we have used the Greenwich sunspot area data (plotted in panel b) in Fig. 2) which is available online[2]. We see a very good match between our data and the Greenwich time series and hence validate our detection method as well as establish the quality of the data. The time variation of the monthly averaged sunspot area obtained from the Kodaikanal is plotted in red in Fig. 3. The Greenwich data, for the same time period, is overplotted in blue. The fluctuations seen in the monthly averaged data get further smoothed out in the yearly averaged plot shown in red in Fig. 4. From the overplotted Greenwich data (in blue) we notice a good overall match between the two data sets. There are certain differences, however. For example, some of the double-peak behaviors seen in Kodaikanal data are not so prominent in the Greenwich data, and vice-versa. Also, we see that the amplitudes of the two data sets differ to the highest degree at the time of cycle maxima, and that the Greenwich data has higher area values (at the solar maximum) compared to the Kodaikanal data, in most of the cycles. For example, from Fig. 4 we notice that, near the time of maxima of cycle 18, Kodaikanal area values are significantly lower than the Greenwich values, though the pattern remains the same for the two data sets. Due to poor photographic plate conditions, the semi-automated code did not detect all the spots around this cycle maxima, which results in

an underestimation. We should also mention here that the Greenwich data have been cataloged using sunspot observations from different observatories around the globe (Cape of Good Hope, Kodaikanal, and Mauritius) to supplement missing days or bad data. Thus, the RGO data are affected by different scaling in order to incorporate data points from various observatories. This may also lead to slight differences in values in comparison to the Kodaikanal data, which have been compiled from a single observatory for the whole 90-yr study period. To improve the statistics further, we will compare original images from different sources in our forthcoming studies. This involves one-to-one cross-calibration of the sunspot images from the Kodaikanal and the Greenwich observatories (on the simultaneous observational days). Such a study will also help towards generating a consistent sunspot-area time-series that can be used by the solar community.

A further comparison and relevant discussion of the Kodaikanal data with the Debrecen data base have been added to Appendix B.

To quantify the correlation between the Kodaikanal and the Greenwich data, we plot the scatter diagram in panel a) in Fig. 5, for the yearly averaged area values. The correlation coefficient, obtained in this case, is 0.94. We also plot a similar scatter diagram for the sunspot number obtained from Greenwich data along with Kodaikanal sunspot area and we obtain a correlation coefficient of 0.93. Here, we must emphasize that such a high
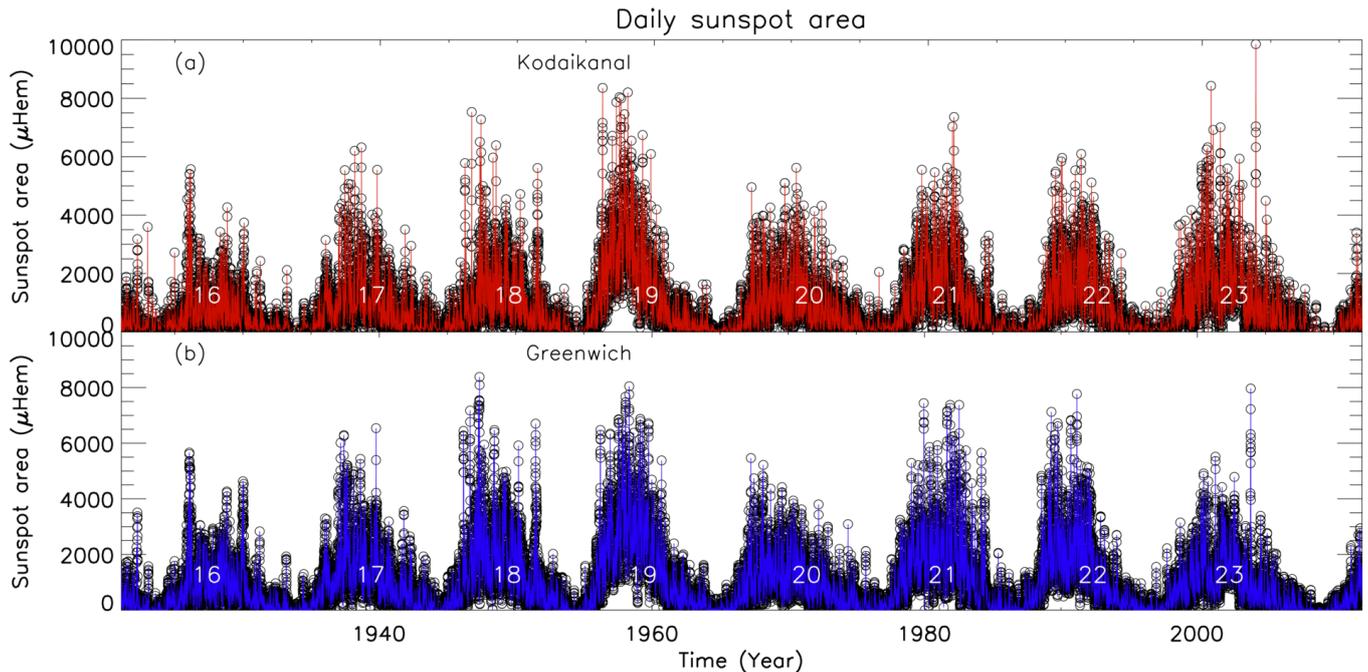
---

[2] http://solarscience.msfc.nasa.gov/greenwch.shtm

Daily sunspot area

Fig. 2. Panel **a)**: daily sunspot area from the Kodaikanal white-light images. In panel **b)**, we plot the same but for the Greenwich data. Individual sunspot cycle numbers are printed on both the panels.

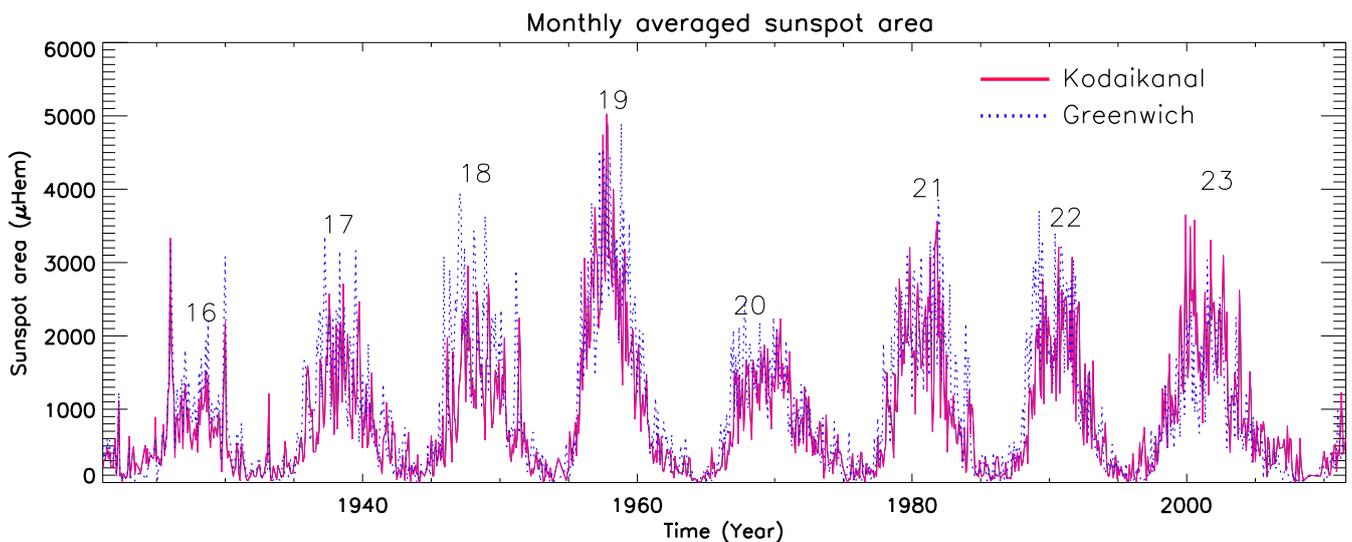## Monthly averaged sunspot area

Fig. 3. Monthly averaged sunspot data obtained from Kodaikanal (plotted in red) along with the data from Greenwich for the same period (plotted in blue).

correlation of the sunspot area with the sunspot number demonstrates the efficacy of using the sunspot area as an alternate proxy for the solar cycle.

To validate our detection algorithm further, we compare our results with Sivaraman (2000). These authors calculated the sunspot area from the Kodaikanal white-light images by using hand-drawn contours. Here, we must mention the fact that these authors did not provide the data in tabular format in their paper. We used the IDL routine *read_jpeg.pro* to extract the information from Fig. 2 in Sivaraman (2000). In panel a) of Fig. 6, we plot monthly averaged sunspot area during the period 1940–1985 for both cases. Careful consideration of the plot reveals that there are three notable deviations in three different months. We thoroughly checked our analysis at those locations and did not find over-estimation or under-estimation of the area values.

As mentioned earlier, there is no written records of the previous work and thus it is impossible for us to comment on the results obtained by Sivaraman (2000). Also, in panel b), we show the scatter plot between the two area values and obtain a correlation of 0.84. This relatively high correlation value indicates the goodness of our semi-automated detection algorithm.

5.2. Sunspot area distribution

Sunspots come in a variety of sizes and also vary with the different phases of a solar cycle. We investigated the size distribution using each sunspot area obtained from the Kodaikanal white-light images for the entire 90-yr period as well as for individual solar cycles. In panel a) of Fig. 7, we plot the histogram of
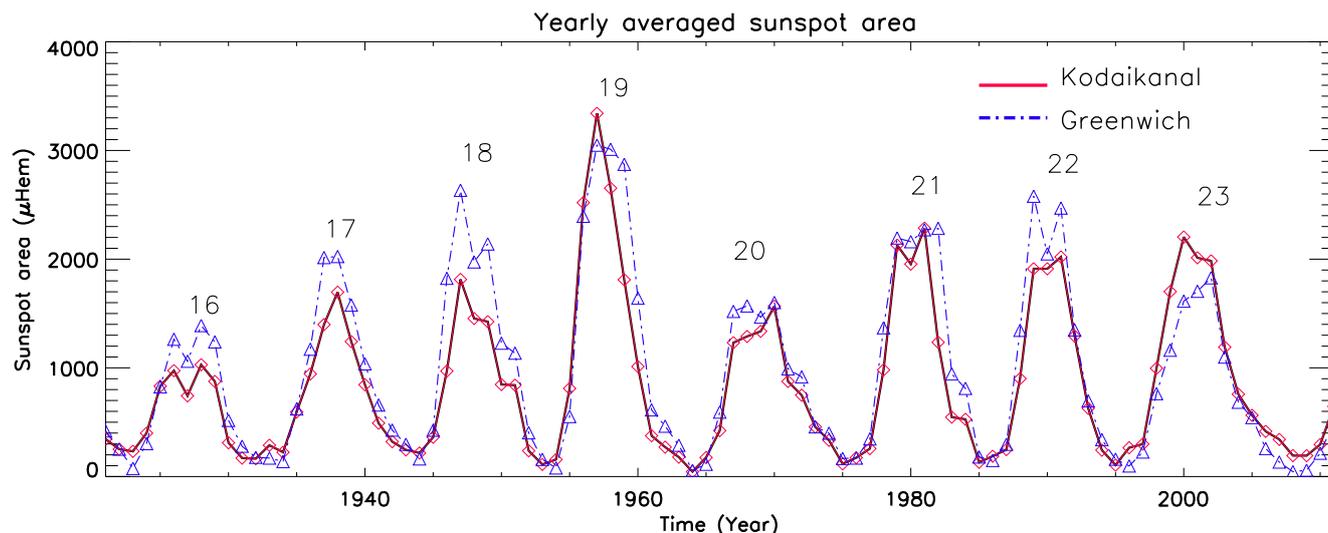
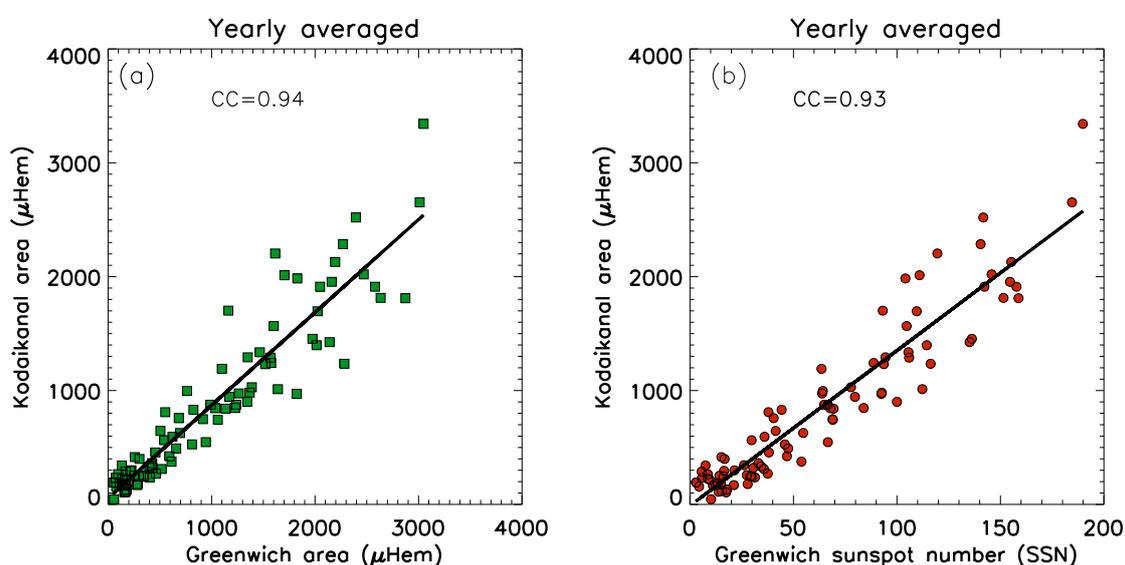**Fig. 4.** Yearly averaged sunspot area from Kodaikanal and Greenwich plotted in red and blue, respectively.

Fig. 5. Panel **a**): scatter plot between the yearly averaged Kodaikanal and Greenwich area measurements. Panel **b**): scatter plot using the yearly averaged smoothed sunspot number (SSN) from Greenwich along with the obtained Kodaikanal sunspot area. Correlation coefficients (CC) are printed on both panels.
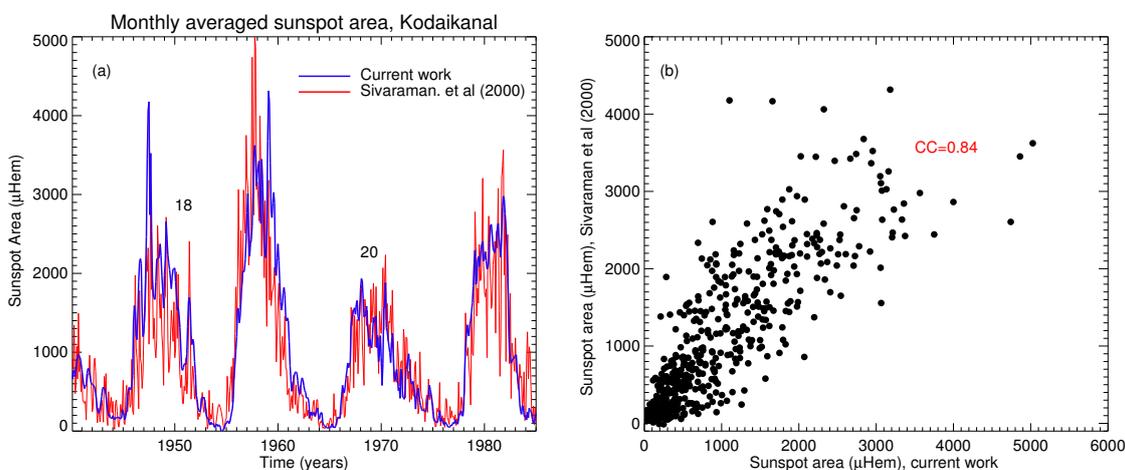

Fig. 6. Panel **a**): comparison between the manual (red line) and the automatic (blue line) detection of sunspot area from Kodaikanal white-light images for a period of 1940–1985. Panel **b**): scatter plot of these two area values.
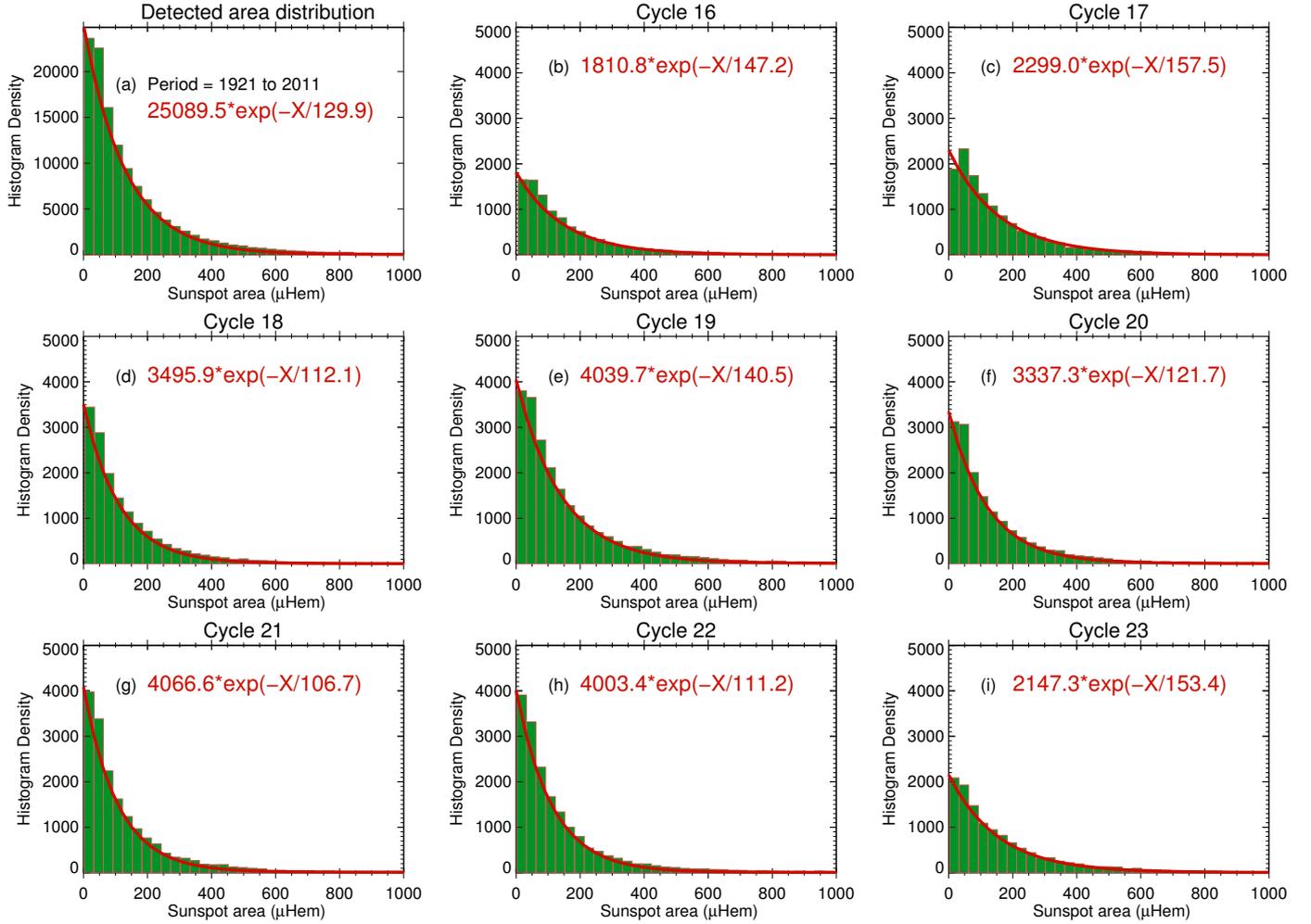
Fig. 7. Histograms (with a bin-size of 30 μHem) of the individual area values (sizes) obtained from Kodaikanal white-light images. The histogram for the full data set is shown in panel **a)**. Panels **b)–i)** show the area distributions for the individual solar cycles (cycle 16 to 23).

the area values of each sunspot for the full 90-yr period. We see a decaying exponential pattern in the histogram profile for the whole data set (panel a)) as well as for individual cycles (panel b)–i)). We fit the histograms with a decaying exponential function of the form, $Y = A_0 \exp^{-\left(\frac{X}{B}\right)}$, as shown by a red solid line in each panel. Analyzing the histograms for individual cycle, we notice that the number of large sunspots has a strong correlation with the overall cycle amplitude, that is, the stronger cycles (e.g., cycle 19) have a larger number of large sunspots compared to the weaker cycles (e.g., cycle 16). The smaller sunspots (area <30 μHem) do not seem to follow any certain rule. Cycles 19 and 21 have different cycle amplitudes but the number of small sunspots is approximately the same, whereas cycles 22 and 23 have similar amplitudes but the number of small sunspots in cycle 23 is almost half that of cycle 22.

In panels a) and b) of Fig. 8, we plot the full width at half maximum (FWHM) and the A_0 values obtained from the fitted curves from each of the histograms. The FWHM here is defined as the value where the amplitude drops to $1/e$ times A_0. We notice some periodicity in the A_0 values but our data series is not sufficiently long to conclude any significant number. Also, we notice a weak anti-correlation between the two parameters, A_0 and FWHM.

In Fig. 7, we show that the individual sunspot area distribution for each cycle is following an exponentially decreasing

function. Next, we investigate the same but for each phase of an individual cycle. We divided each cycle into two-year bins and plotted the area distribution for each cycle in Fig. 9. We notice that the exponential trend remains in all the phases of a cycle. Though, at the time of cycle maxima, the number of smaller as well as bigger sunspots increases but the distribution remains the same. Another important aspect is that not only does the number of big sunspots increase during the maxima of the cycle, but the number of the sunspots with smaller sizes also increases.

We also fitted these area histograms with a lognormal function and found a good match with earlier results. These plots and their descriptions are given in Appendix A.

5.2.1. North-South area distribution

Sunspots do not appear symmetrically in both the solar hemispheres and this phenomena is popularly known as "North-South" (N-S) asymmetry. This asymmetric behaviour in the sunspot area has been studied in the past (Carbonell et al. 1993; Oliver & Ballester 1994; Ballester et al. 2005; Badalyan & Obridko 2011). We re-investigate the N-S asymmetry using our sunspot data extracted from the Kodaikanal white-light images. In panel a) of Fig. 10, we plot the time variation of the monthly averaged sunspot area for the individual
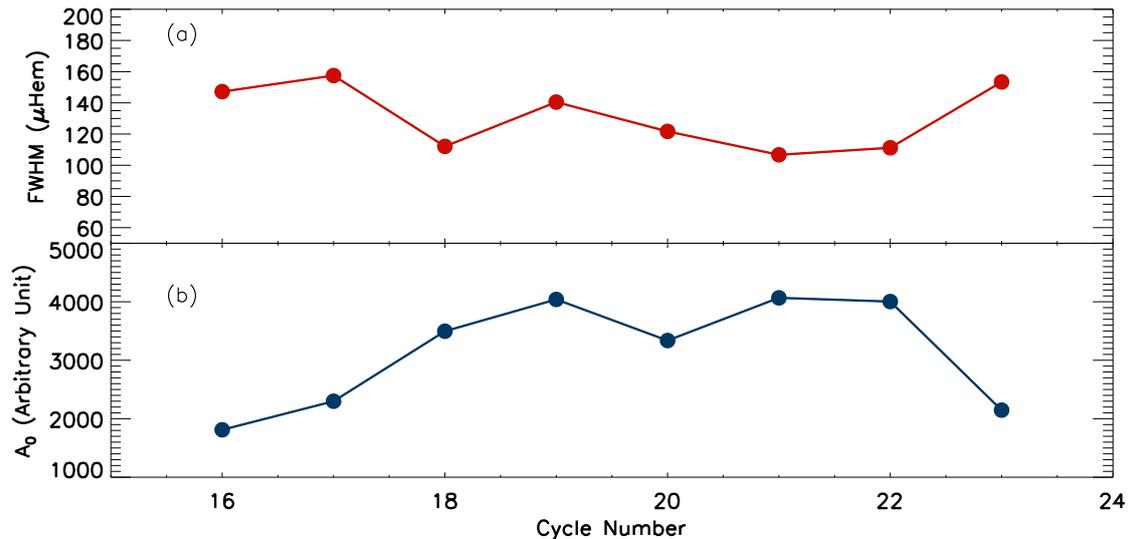
Fig. 8. Variations of the parameters FWHM (panel **a**)) and A_0 (panel **b**)) from the fitted exponential functions.

hemispheres (in green and red for the northern and southern hemisphere, respectively).

We also plot the yearly averaged sunspot area for each hemisphere in panel b). Close inspection of panel b) reveals that the double peak behavior is very often restricted to one hemisphere only. In cycle 19, we see a clear double peak in the northern hemisphere whereas no such behavior is seen in the southern hemisphere for that cycle. Same behavior can be seen for cycle 21. In cycle 22 and 23, it is the southern hemisphere that shows the double peaks.

From the plots, we see that the sunspot area evolution with the solar cycle is not symmetric in its two hemispheres. There are different ways to quantify this asymmetricity. The two most popular methods are the "absolute asymmetry" parameter, defined as $A_a = (N - S)$, and the "normalized asymmetry" parameter defined as $N_a = (N - S)/(N + S)$, where N and S are the sunspot area in the northern and southern hemispheres.

We compare the values of A_a and N_a for our data set. In the top panel of Fig. 11, we plot the yearly averaged sunspot area where, as in the bottom panel of Fig. 11, we plot the A_a and N_a. We notice that the two parameters evolve differently with the progress of the solar cycle. The N_a reaches its maximum value at the time of solar minimum whereas the A_a peaks at the solar maximum of each cycle. Such variation of the parameters can be explained by the fact that, at the time of solar minimum, there are very few spots appearing at the surface of the sun, making the value of the N_a very high, close to +1 or –1, depending on the appearance of the sunspot in the hemispheres. The advantage of using A_a over N_a, especially for time series analyses, has been thoroughly studied in the past (Ballester et al. 2005; Temmer et al. 2006).

5.2.2. Butterfly diagram and latitudinal area distribution

In this section, we explore the latitudinal variation of the emergent sunspots over the solar cycle evolution. In Fig. 12, we plot the time evolution of the latitude values of the detected sunspots from Kodaikanal. We also divided the sunspots into four size bins and overplotted them in different colors on the same plot (different size bins are printed on the plot).

Figure 12 represents the "butterfly diagram", which shows that the sunspots appear at higher latitude in the beginning of the solar cycle and then the activity belt drifts towards the equator as the cycle progresses. Careful inspection reveals that the larger spots (area ≥1000 μHem) appear at the time of the solar maximum. Also, we see that the number of these larger sunspots depends on the strength of the cycle. For example, the strongest cycle, cycle 19, has a larger number of spots larger than 1000 μHem compared to a weaker cycle, such as cycle 16.

Next, we explore the distribution of the number of the detected sunspots with their latitude. Panel a) in Fig. 13 shows the histogram for the period 1921 to 2011. We clearly see a "bi-model" distribution with two bell-shaped curves separated around the zero latitude. Such a distribution is also seen for the individual cycles (panel b)–i)). We use a bin of 2° in latitude to incorporate the fact that a sunspot may slightly change its latitude during its lifetime. We fit each of the peaks with a Gaussian function as shown in red thick line in all the panels in Fig. 13. Maximum height (h) and center (c) of each of the fitted Gaussians is printed in individual panels. Apart from the Gaussian distributions, we also notice from the histograms that the height of the Gaussian peaks are different in the two hemispheres. Panel a) of Fig. 14 shows the variation of the parameter "h" with the cycle number for both hemispheres. From the figure, we see that for most of the cycles, the northern hemisphere dominates over the south. It must be emphasized here that this plot is effectively another representation of the N-S asymmetry. From panel b) of Fig. 14, we see that maximum sunspots appear at a latitude of approximately 14.5° on either side. We also notice a strong deviation of this center value from the average for the cycle 19, the strongest cycle of the 20th century. Recently, Cameron & Schüssler (2016) explained the behavior of these fitted Gaussian parameters using a solar dynamo model and also predicted the value of the turbulent diffusivity acting on the toroidal magnetic field.

6. Reproducing irregular features of solar cycles

Amplitude and length of a solar cycle vary from cycle to cycle. Based on various century-long sunspot observations, it has been found that these irregularities of the solar cycles have some definite properties (e.g., the Waldmeier effect and the correlation between decay rate and cycle amplitude). Waldmeier (1935) reported that during the rising phase of a solar cycle, the rise
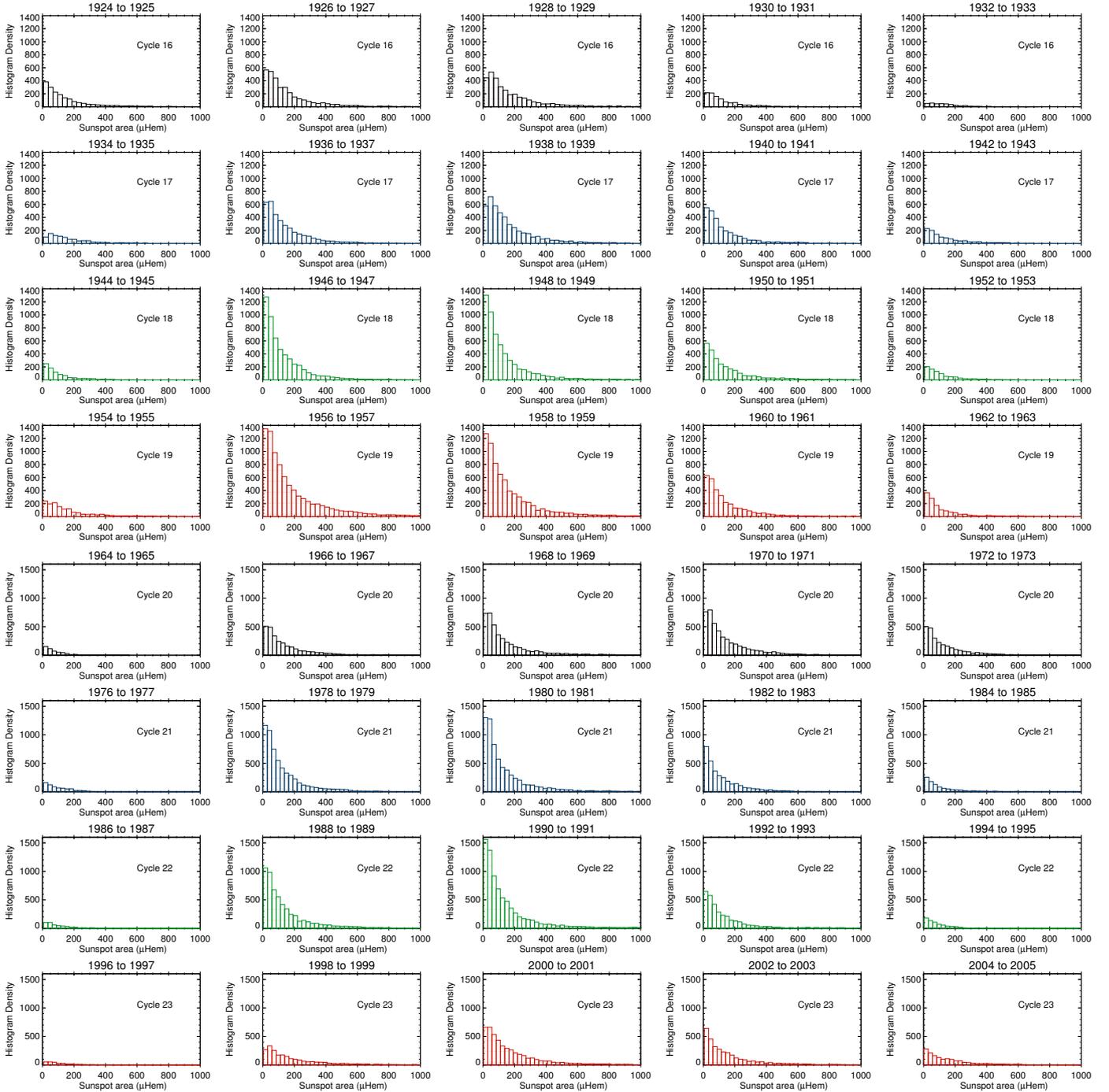
Fig. 9. Sunspot size distribution at different phases of the solar cycle for each individual cycle (cycles 19–23). The bin-size is 30 μHem.

time is anti-correlated with the peak amplitude of that cycle, that is, the stronger cycles take less time to rise and vice-versa. There is also a strong correlation between the rising rate and the peak amplitude of a cycle. These relations are known as the Waldmeier effect. Karak & Choudhuri (2011) reproduced these correlations using Greenwich sunspot data and classified these two properties as Waldmeier Effect 1 (WE1) and Waldmeier Effect 2 (WE2), respectively. During the descending phase of a solar cycle, some important correlations are also reported. Recently, using the Greenwich sunspot data, Hazra et al. (2015) found a correlation between decay rate and the peak amplitude of the cycle. These authors have also shown that, apart from the sunspot area data, such correlation also exists in sunspot number

and 10.7 cm radio flux data. Another important property during the decaying phase is the correlation between the decay rate during the late phase of a cycle and the amplitude of the following cycle (Yoshida & Yamagishi 2010; Hazra et al. 2015). This is also important in terms of predicting the next cycle amplitude. Though this correlation is statistically not so strong, but it still provides an approximate estimation of the next cycle amplitude. Hazra et al. (2015) have interpreted these relations in terms of a flux transport dynamo model.

In this section, we have calculated various correlations during ascending and descending phase of the solar cycles from the Kodaikanal sunspot area data and reproduced all the irregular features of the solar cycles discussed above.
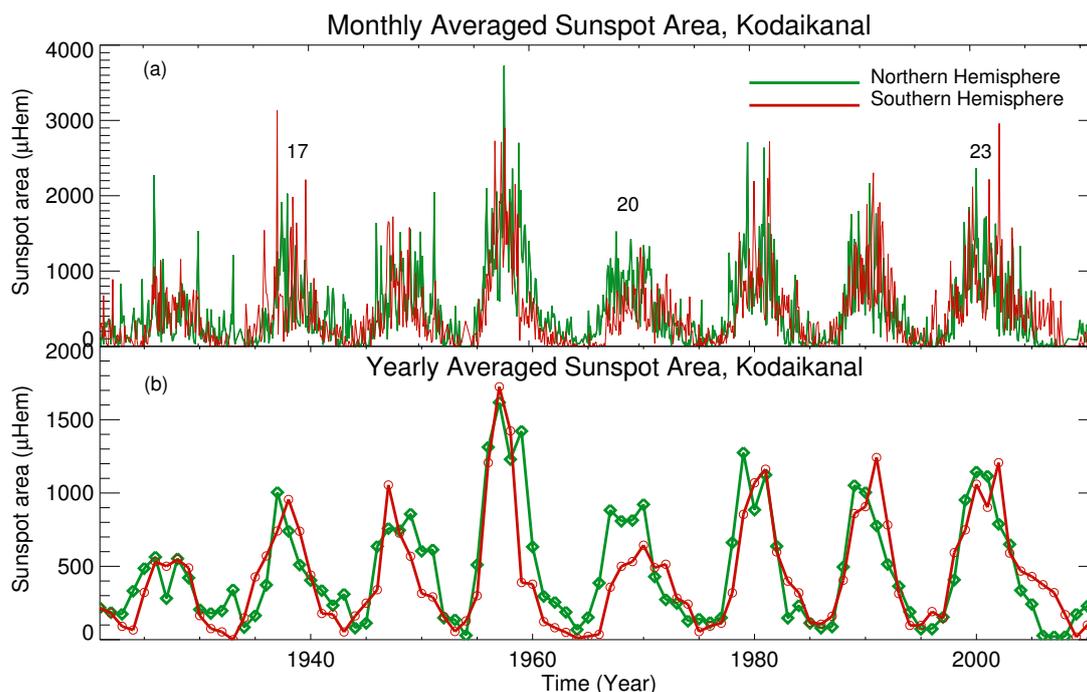
Fig. 10. Panel **a)**: monthly averaged sunspot area variation with time for the northern (shown in green) and southern (shown in red) hemisphere. Panel **b)**: same variation but for yearly averaged data.
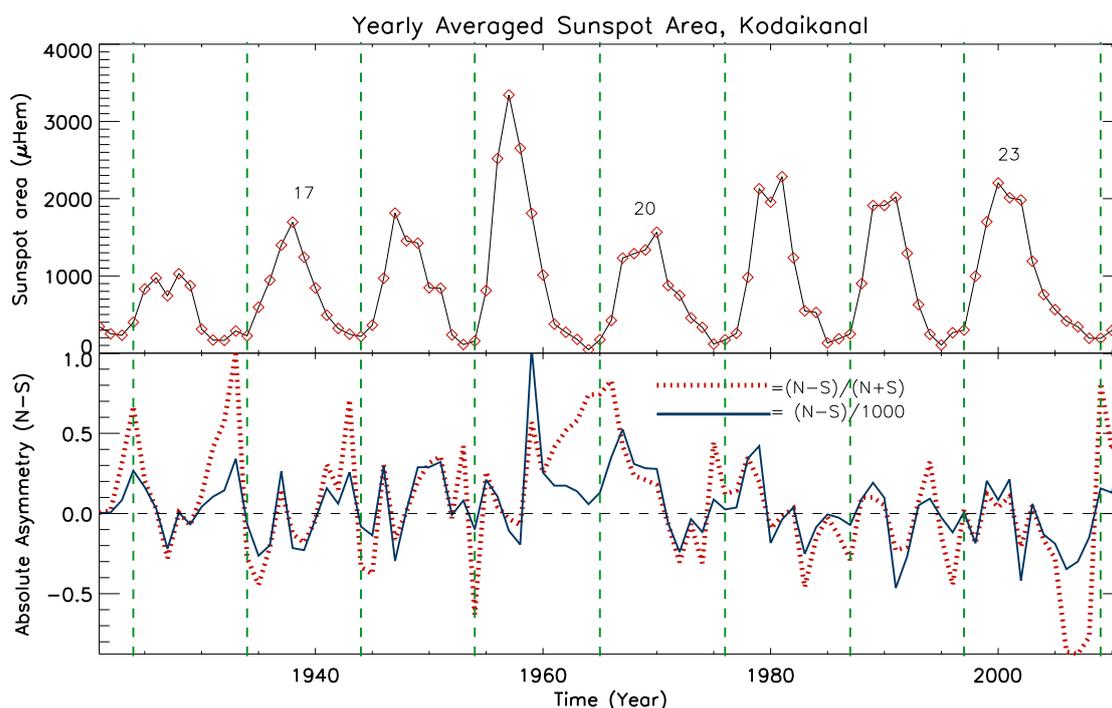

Fig. 11. *Top panel*: yearly averaged sunspot area from Kodaikanal. In the *bottom panel*, we compare the two parameters used to compare the north-south asymmetry from the yearly averaged data. The normalized asymmetry (N-S)/(N+S) is shown by a red dotted line whereas the blue solid line represents the absolute asymmetry parameter.

6.1. Waldmeier effect: correlations in ascending phase of the solar cycle

Although the irregular properties of the solar cycles are well established, doubts are often expressed regarding their existence and some believe they may simply be statistical artifacts (Dikpati et al. 2008). Reconfirmation of the Waldmeier effect from a completely different sunspot area data set is not only

important for the prediction of the peak amplitude when a cycle is in the early stage but it is also important for validation of our existing dynamo models. Time variation of daily sunspot area data or even the monthly averaged sunspot area data are so scattered that it is very difficult to calculate the required quantities, such as rise time and rise rate, to validate their specific properties. Hence, we smoothed our data using a Gaussian filter with *FWHM* = 1 yr and *FWHM* = 2 yr as shown in Figs. 15a
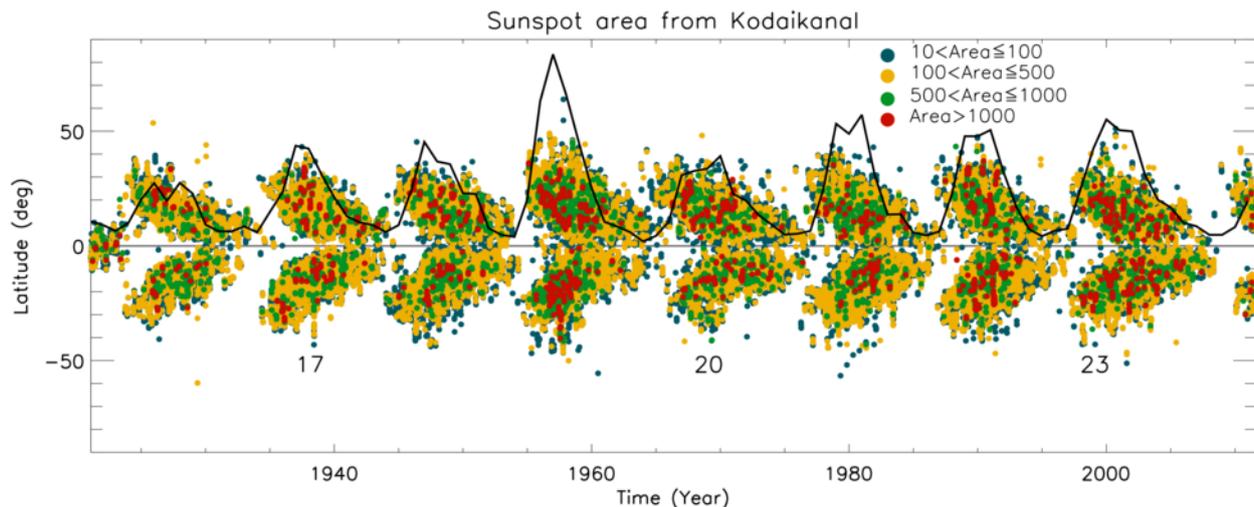
Fig. 12. Butterfly diagram created using the individual sunspot area values (sizes) and their locations obtained from Kodaikanal white-light images. Cycle numbers are also printed on the plot. The solid black line represents the yearly averaged sunspot area (scaled).
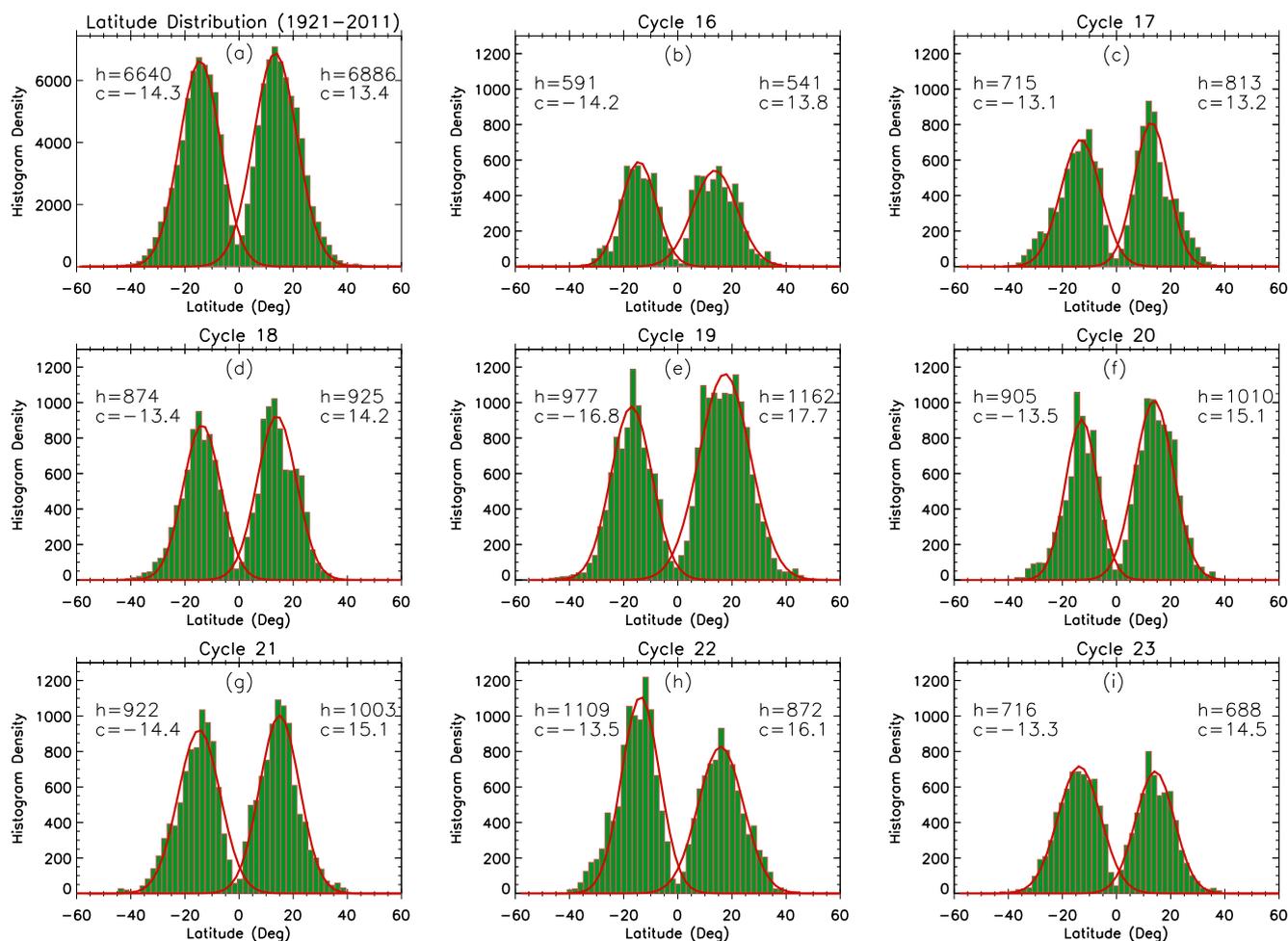

Fig. 13. Histograms of the latitude distribution of the individual sunspot area values (sizes), from the Kodaikanal, are shown for both data sets and for individual cycles. Fitted Gaussian in each histogram is overplotted in red.

and b, respectively. Red and green vertical dotted lines represent the position of the minima calculated for one- and two-year averaged data sets. Closer inspection of Fig. 15 reveals that the determination of the solar minimum for each cycle is always very sensitive to the filtering window and minima calculated using

1-yr averaged data are not the same as those calculated using the 2-yr averaged data. This is also true for determination of solar maxima. It was also found that there is always an overlap between two successive cycles and the position of cycle minima depends on this overlap (Cameron & Schüssler 2007). Thus,
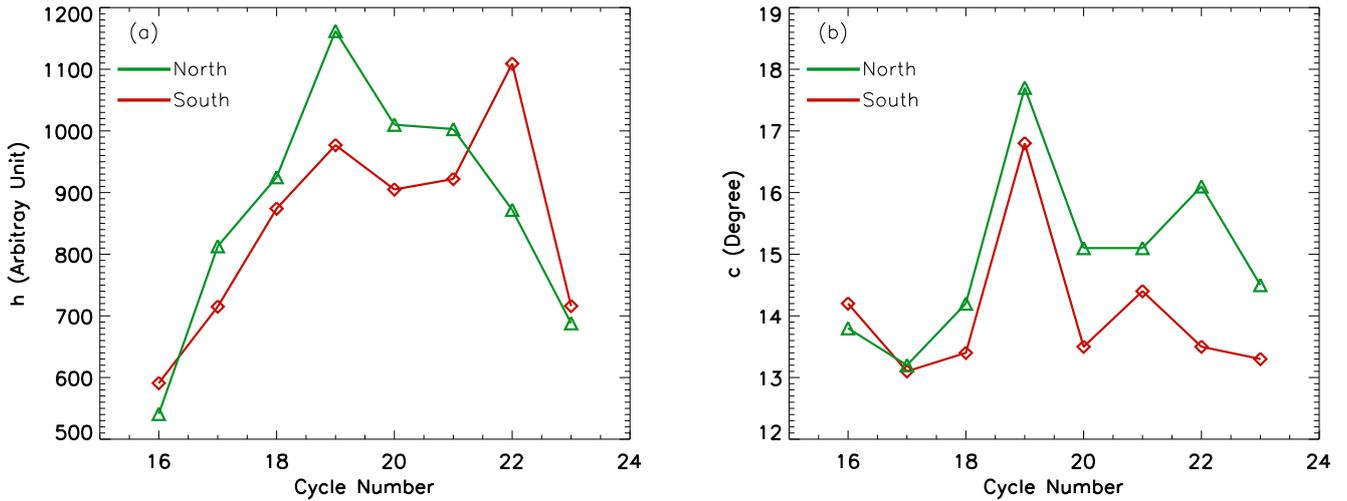
Fig. 14. Panel **a)**: height **h)** of the fitted Gaussian functions (shown in Fig. 13) for northern and southern hemispheres in green and red colors, respectively. Panel **b)**: center values **c)** of these Gaussians.
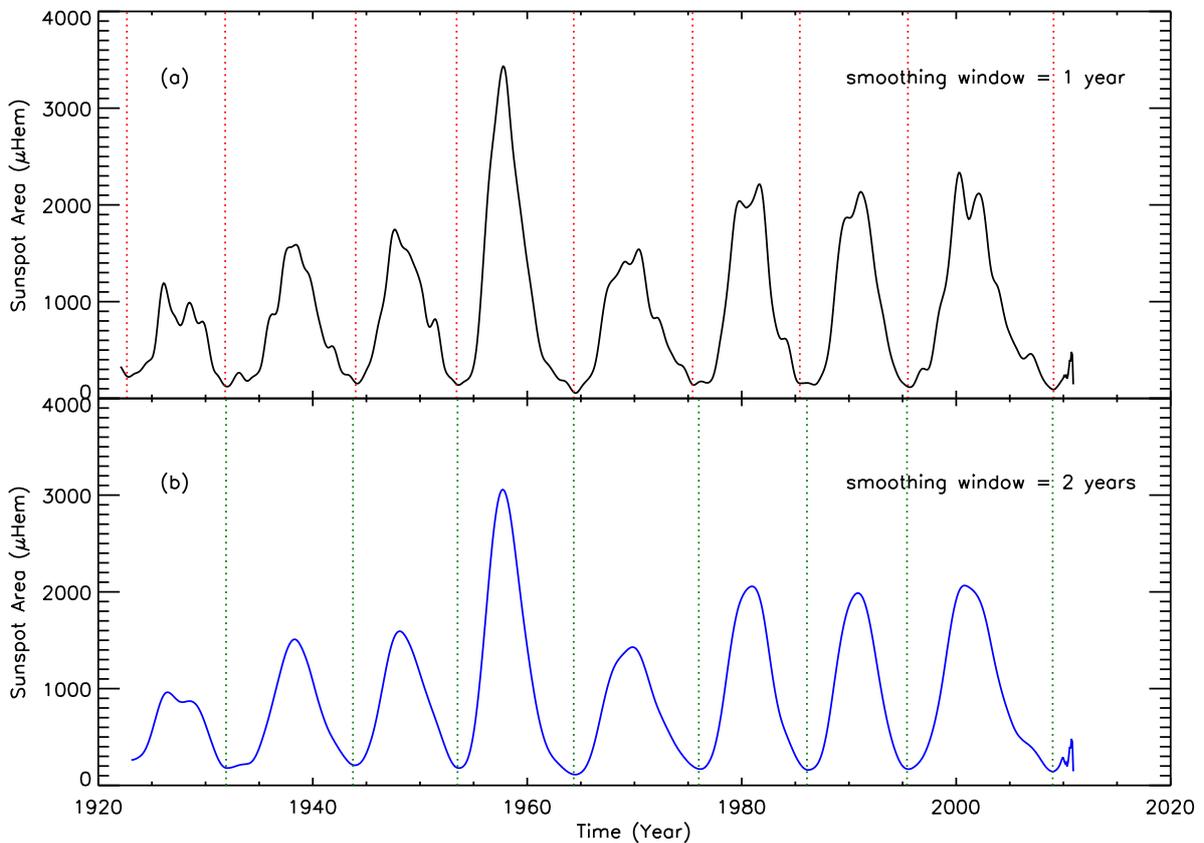

Fig. 15. Panel **a)**: smoothed (with a Gaussian of *FWHM* = 1 yr) solar cycle from Kodaikanal sunspot area data. Panel **b)**: same as panel **a)** but with a FWHM of 2 yr.

defining the solar cycle minima and maxima becomes very difficult. Following Karak & Choudhuri (2011), we calculated the rise time by taking the time period during which the solar cycle reaches 0.8P from 0.2P, where P = peak amplitude of the cycle. Rise rate is defined as the slope between two points separated by one year and the first point is chosen one year before the cycle maxima. All the plots shown in this section have been generated using a data set smoothed by a Gaussian filter with a FWHM of 2 yr. Panel a) of Fig. 16 shows the plot between the rise time and the peak amplitude of the cycle. We see an

anti-correlation ($r = -0.66$) between these two parameters and this effect is known as the WE1. Cycle 16 in panel a) of Fig. 16 has not been considered because the minimum of this cycle is not well captured after two years of averaging. Therefore, the position of 0.2P is not well defined in this case. In panel b) of Fig. 16, the correlation between the rise rate and cycle amplitude is calculated. It is evident from the Fig. that WE2 is also well reproduced with correlation coefficient $r = 0.99$. Here we should mention that the correlation coefficient, which we have calculated here, is only statistically significant when using a large number of data
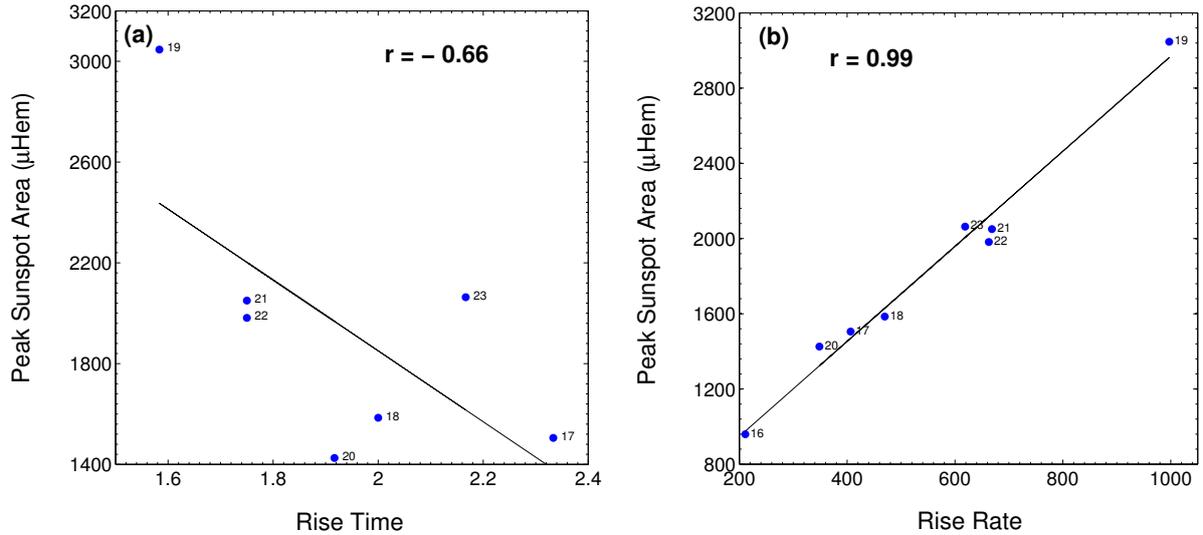
Fig. 16. Panel **a)**: scatter plot of rise time (in years) and the peak amplitude of the same solar cycle (*FWHM* = 2 yr). The solid line represents the best linear fit. Panel **b)**: a scatter plot of rise rate (μHem/year) and the peak amplitude of the same solar cycle (*FWHM* = 2 yr). Solar cycle numbers are also marked beside each of the blue circles.

points. In our case, we have only eight data points corresponding to eight cycles and one should calculate the "adjusted Pearson correlation coefficient" (r_{adj}) defined as

$$r_{adj}^2 = 1 - \left(1 - r^2\right)\frac{n-1}{n-p-1}, \tag{1}$$

where r is the normal correlation coefficient, n is the number of data points, and p is the independent variables. The calculated adjusted coefficients for $r = -0.66$ (WE1) and $r = -0.99$ (WE2) become $r_{adj} = -0.57$ and $r_{adj} = 0.99$, respectively. This shows, therefore, that the adjusted correlation coefficients do not vary significantly for high correlation coefficients and thus we continue with the normal Pearson correlation coefficients.

In Table 2, we summarize the different correlation coefficient values that are calculated using Gaussian with a FWHM of 1 or 2 yr.

6.2. The correlations in the descending phase of the solar cycle

In the previous section, we mentioned that the presence of small fluctuations and the double peaks in the solar cycle make it difficult to identify the actual minimum and maximum of a cycle. To avoid this discrepancy, we have excluded the actual cycle minimum and the maximum in our decay rate calculation. The final decay rate of a cycle was determined by averaging over the individual decay rates calculated at four different locations. The first individual decay rate is determined by taking a slope between two points separated by one year and the first point is taken one year before the cycle minimum. Similarly the last individual decay rate near solar maximum is also calculated by taking two points with one year interval and the first point is chosen one year after the solar maximum. The other two decay rates are calculated at locations in between positions of last and first decay rate points.

In panel a) of Fig. 17, we show the correlation between the final decay rate of a cycle and the cycle amplitude using 2-yr average sunspot area data. The correlation between decay rate calculated during the late phase of a cycle and the amplitude of the following cycle is shown in panel b) of Fig. 17. We obtain

a correlation coefficient $r = 0.75$ for the correlation between decay rate and cycle amplitude for the same cycle, whereas we get $r = 0.41$ for decay rate at late phase with following cycle amplitude. There are only seven points (cycle 23 is omitted) in panel b) of Fig. 17 due to the fact that the correlation is defined with the amplitude of the following cycle with the decay rate at the late phase of the previous cycle. The correlation coefficients obtained from one year averaged data (*FWHM* = 1 yr) are also given in Table 1. From the values given in Table 1, we notice that only the correlation between rise time and peak amplitude of the cycle is sensitive to the smoothing window (though the statistical trend is the same) whereas the other coefficients are of comparable values.

We mentioned that each cycle is different from the previous in terms of cycle amplitude and duration. In the following section, we explore the periodic or quasi-periodic nature of these variations using the wavelet analysis.

7. Periodicities in the sunspot area data

Sunspot cycle has an average periodicity of 11 yr, though there have been various reports of periodicities shorter than 11 yr (Hathaway 2015). Two of the most significant and persistent periods (shorter than 11-yr) seem to be approximately 155 days and 1.3 yr (Carbonell & Ballester 1990; Krivova & Solanki 2002). The origin of the 11-yr sunspot cycle is now believed to be governed by the global solar dynamo mechanism but the origin of these shorter periodicities has, until now, remained poorly understood.

In this section, we study the periodic variations of different time scales using the monthly averaged sunspot area data obtained from the Kodaikanal Observatory. We use the wavelet analysis tool (Torrence & Compo 1998) to study the periodic and quasi-periodic variations in the sunspot area time series. The wavelet analysis is a useful tool for examining the presence of localized oscillations, both in time and frequency domains. Figure 18 shows the wavelet analysis on the monthly averaged sunspot area time series. The top panel shows the time variations of the monthly averaged sunspot area whereas the bottom left-hand panel shows the wavelet power spectrum. The global wavelet power, defined as the time-averaged wavelet

Table 1. Specifications of Kodaikanal white-light images.

Parameter	Description
Observation period	1904 to 2011
Telescope	15 cm objective with f/15 beam
Original storage	Plates/films
Total number of plates	31800 (1904 to 2011)
Average number of plates per year	290
Calibration period	1921 to 2011
Final image size	4k × 4k
Pixel scale	0.62″
Available image format	FITS and JPG
Raw images available at	http://kso.iiap.res.in

Table 2. Correlation coefficients at different phases of a solar cycle.

Smoothed data with *FWHM*	Correlations during rising phase of the cycle between		Correlations during decaying phase of the cycle between	
	Rise time and cycle amplitude (WE1)	Rise rate and cycle amplitude (WE2)	Decay rate and cycle amplitude	Decay rate at late phase and cycle amplitude of following cycle
1 yr	−0.63	0.92	0.57	0.56
2 yr	−0.66	0.99	0.75	0.41
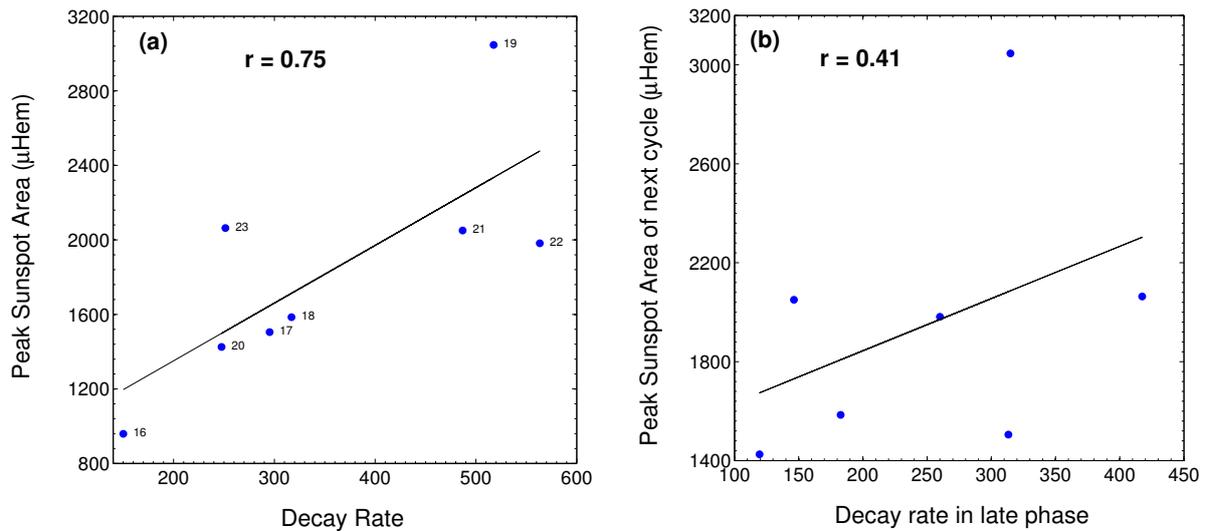
Fig. 17. Panel **a**): scatter plot of decay rate vs. peak amplitude of the same solar cycle (*FWHM* = 2 yr). The best linear fit is shown with the solid line. Panel **b**): scatter plot of decay rate at late phase vs. peak amplitude of the following cycle.

power, is shown in the bottom right-hand panel. The dotted line on top of the global wavelet power shows the significance level of 99%, calculated assuming white noise (Torrence & Compo 1998). The global wavelet power shows a dominant period of 11 yr with the presence of weak power concentrations at 22 and 5.5 yr. The 11-yr period is the strongest period present in the sunspot activity and is probably governed by the global dynamo mechanism (see Charbonneau 2010).

Short-term periodicities between 1 and 4 yr are often referred to as quasi-biennial oscillations (QBOs). QBOs were previously found in different solar proxies, such as in solar wind data and cosmic rays (see Table 1 in Bazilevskaya et al. 2014: for a complete list). QBOs are also observed in the frequency shifts as inferred from the helioseismology data (Broomhall & Nakariakov 2015). Periodicities in such a variety of proxies suggest that

these small period oscillations are genuine phenomena and have origins connected to the solar dynamo.

The 22-yr cycle is probably a manifestation of the Gnevyshev-Ohl (even-odd) rule, which refers to a pattern of alternating higher and lower cycle amplitudes than the average cycle amplitude (Gnevyshev & Ohl 1948; Charbonneau et al. 2007). The 5.5-yr period, apart from being a harmonic of the 11-yr cycle, is closely related to the double peak behavior of the solar cycle (Gnevyshev 1967, 1977; Georgieva 2011).

The sunspot cycle is highly dominated by the 11-yr periodicity. To explore the presence of the short-term periodicities in the time series, one needs to suppress the 11-yr period from the time series. To remove the longer periods, we used the empirical mode decomposition (EMD) technique (Huang et al. 1998). Time series, after removing the longer periods, is shown
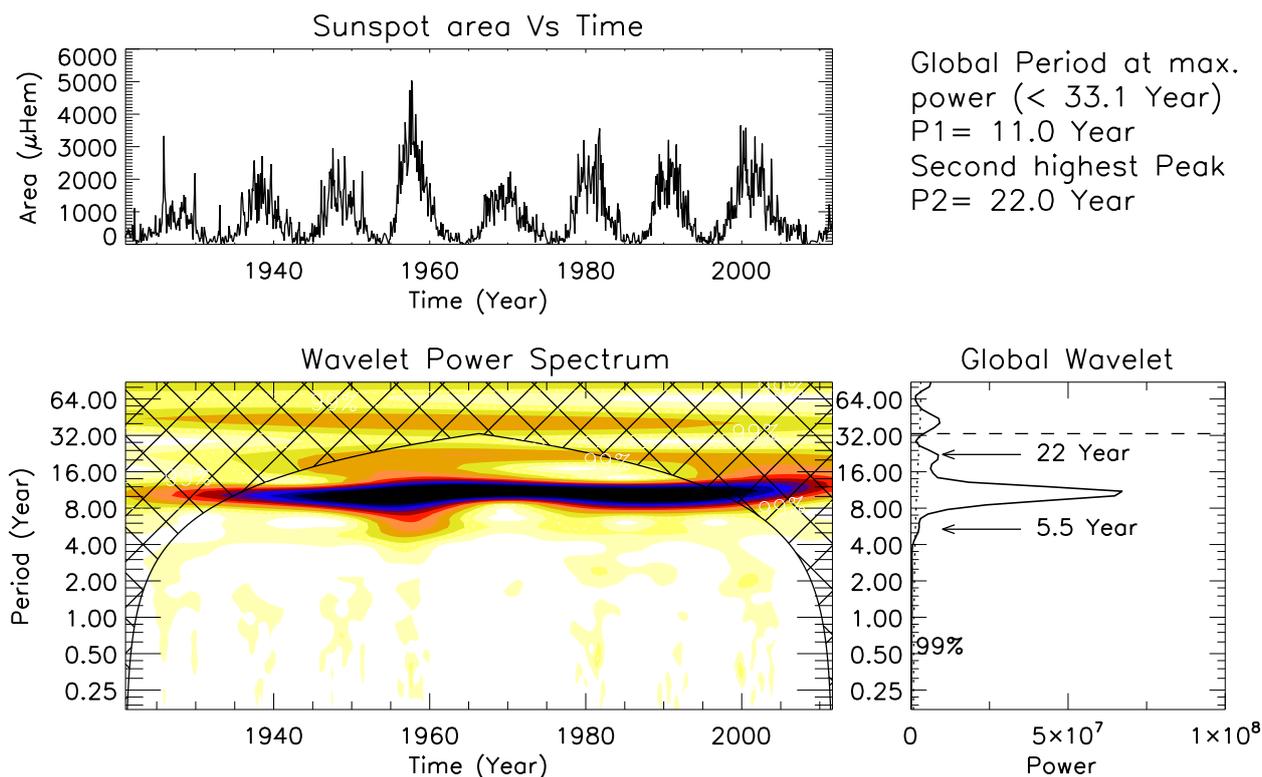
Sunspot area Vs Time

Fig. 18. *Top panel*: variation of monthly averaged sunspot area data from Kodaikanal with time. *Bottom left panel*: wavelet power spectrum. Cross-hatched regions above the wavelet power spectrum show the cone of influence (COI). *Bottom right-hand panel*: global wavelet power. The maximum measurable period is ∼33 yr, which is shown by a horizontal dashed line. The dotted line above the global wavelet power plot shows the significance level of 99%. The significant periods as measured from the global wavelet power are printed on the panel.
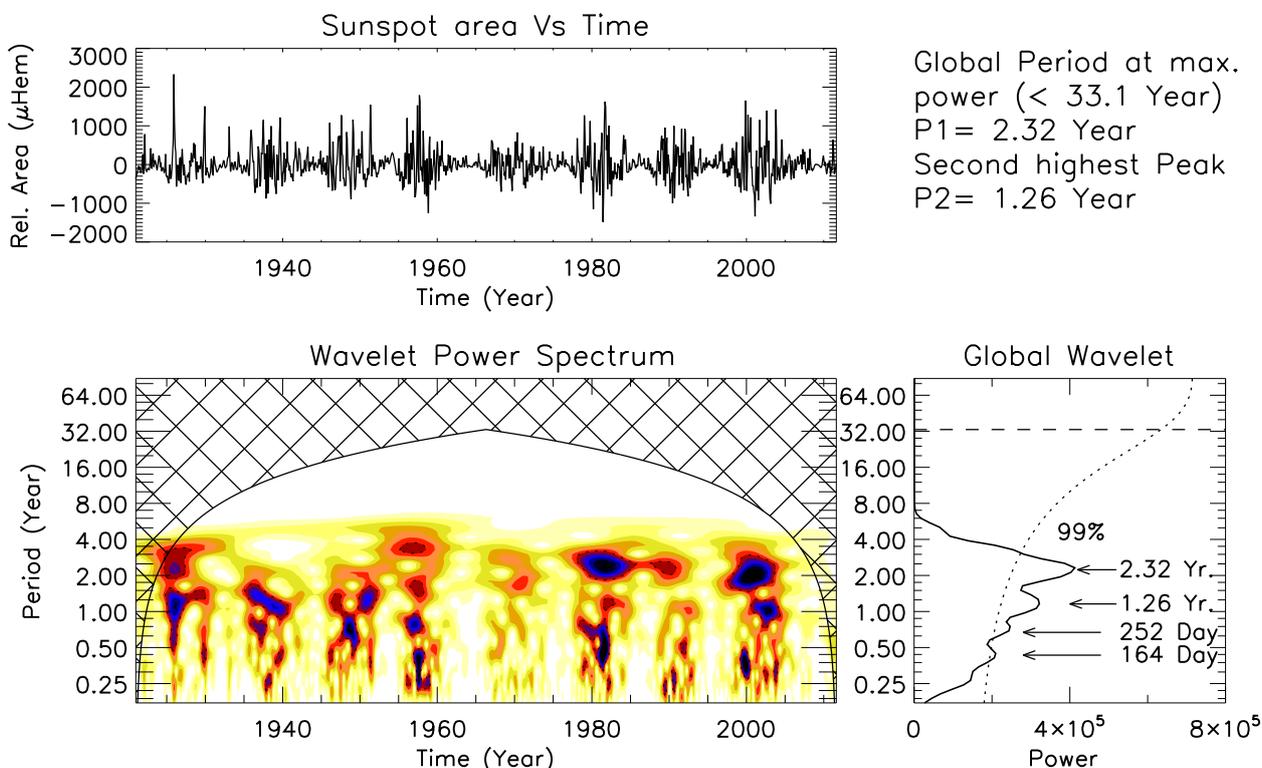
Fig. 19. *Top panel*: variations of sunspot area with time after removing low-frequency components from the time series. The different panels are the same as in Fig. 18.
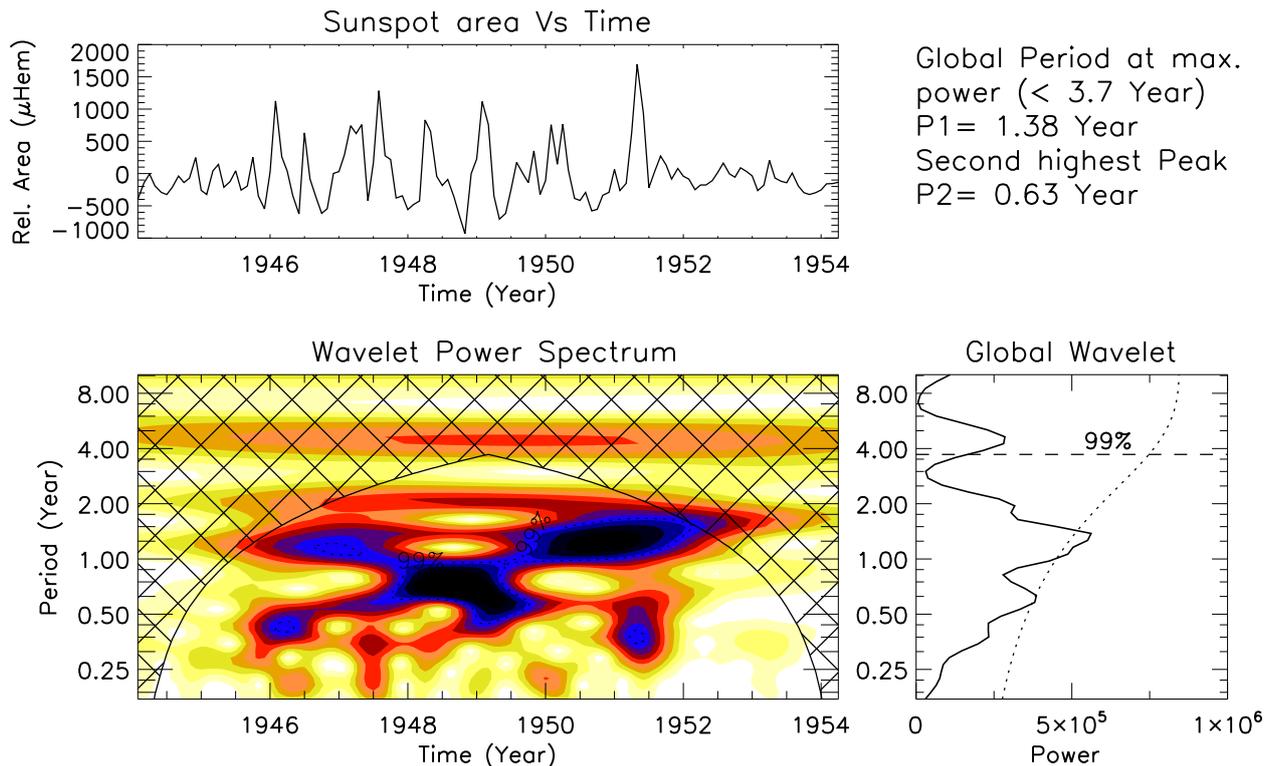
Fig. 20. *Top panel*: variations of sunspot area with time (solar cycle 18) after removing the low-frequency components from the time series. The panels are the same as in Fig. 18.

Table 3. Periodicities in the individual solar cycles.

Cycle number	Strongest period	Second strongest period	Third strongest period
16	3.26 Y	1.26 D	230 D
17	1.26 Y	2.32 Y	164 D
18	1.38 Y	230 D	1.95 Y
19	1.95 Y	274 D	3.28 Y
20	3.28 Y	1.64 Y	164 D
21	2.76 Y	252 D	164 D
22	3.00 Y	1.95 Y	1.06 Y
23	1.95 Y	354 D	124 D

in the top panel of Fig. 19. We filtered out the periods longer than ~4 yr to find the periodic and/or quasi-periodic variations within the solar cycles. The results of the wavelet analysis are shown in Fig. 19. It shows that the shorter-period oscillations are evolving very rapidly compared to the longer-period oscillations. It also shows that the shorter-period oscillations are prominent around the solar maximum (similar to Zhan et al. 2006). The global wavelet plot shows significant powers at 2.32 yr, 1.26 yr, 252 days and 164 days. The presence of 150 to 160-day periodicity in the solar activity is widely reported (Lean & Brueckner 1989; Lean 1990; Carbonell & Ballester 1990; Oliver et al. 1998; Krivova & Solanki 2002; Ballester et al. 2002) along with the 1.3 yr period (Lockwood 2001; Krivova & Solanki 2002; Obridko & Shelting 2007). Zhan et al. (2006) found the existence of 2.21-yr and 246-day periodicity in the sunspot area data obtained from the Royal Greenwich Observatory. We have also analyzed the individual cycles separately and the obtained periodicities are listed in Table 3. One such case, for cycle 18, is shown in Fig. 20.

8. Summary and conclusion

In this paper, we present a century-long white-light digitized data from the Kodaikanal observatory. The advantage of this data set is that the data have been acquired with the same optics since 1918. Such a consistent data set is extremely useful for studying long-term variations in solar activities. We implemented a semi-automated algorithm for sunspot detection throughout this data set. Thus, the effect of human bias on the detection method is minimized and our confidence on the results obtained from it is enhanced.

We present the sunspot area variation over different solar cycles and find a very good correlation with Greenwich data. Sunspot area values obtained using hand-drawn contours on the

Kodaikanal white-light images by Sivaraman (2000) match our semi-automated detections well.

We find that the sunspot sizes follow an exponential distribution for each solar cycle. Such distribution does not change at different phases of the solar cycle though the number of sunspots with different area values seems to increase or decrease according to the cycle maximum or minimum, respectively.

We measured the hemispheric asymmetry associated with the sunspot area. By comparing the two popular methods of ascribing the asymmetricity, we show that the two methods give different results in-terms of peak times of these parameters. Also, the double peak behavior seen during the solar cycles often comes from the double peaks seen in one hemisphere only. Latitudinal distribution of the detected sunspots shows that most of the sunspots appear at approximately $15°$ latitude in each of the hemispheres with a Gaussian-like distribution function (Jiang et al. 2011). The peak heights of the Gaussian distributions for each hemisphere for a particular cycle are different and show long term variation. From the "butterfly diagram", we show that the larger sunspots (area $\geq 1000\ \mu$Hem) appear closer to solar maxima and their numbers are directly proportional to the cycle strength, whereas no such relationship exists for the smaller sunspots (area $\leq 30\ \mu$Hem). Using this area time series, we confirm the "Waldmeier effect", which relates the rise or decay rate of a cycle to the current and following cycle amplitudes. To find the different periodicities present in the individual solar cycles, we performed a wavelet analysis. We find periodicities as long as 22 yr and as small as 124 days.

The entire time series (1921–2011) from Kodaikanal, including the area, latitude, and longitude information corresponding to each of the detected sunspots is provided as supplementary material (available at the CDS). A separate list containing the spotless days is also available at the CDS. We are planning to complete the calibration process of the initial 15 yr (from 1905 to 1920) in future. The time series will then be updated (with proper version number) as soon as we complete the process.

References

Badalyan, O. G., & Obridko, V. N. 2011, New Astron., 16, 357
Ballester, J. L., Oliver, R., & Carbonell, M. 2002, ApJ, 566, 505
Ballester, J. L., Oliver, R., & Carbonell, M. 2005, A&A, 431, L5
Bappu, M. K. V. 1967, Sol. Phys., 1, 151
Baranyi, T., Győri, L., & Ludmány, A. 2016, Sol. Phys., 291, 3081
Baumann, I., & Solanki, S. K. 2005, A&A, 443, 1061
Bazilevskaya, G., Broomhall, A.-M., Elsworth, Y., & Nakariakov, V. M. 2014, Space Sci. Rev., 186, 359
Bogdan, T. J., Gilman, P. A., Lerche, I., & Howard, R. 1988, ApJ, 327, 451
Broomhall, A.-M., & Nakariakov, V. M. 2015, Sol. Phys., 290, 3095
Cameron, R., & Schüssler, M. 2007, ApJ, 659, 801
Cameron, R. H., & Schüssler, M. 2016, A&A, 591, A46
Carbonell, M., & Ballester, J. L. 1990, A&A, 238, 377
Carbonell, M., Oliver, R., & Ballester, J. L. 1993, A&A, 274, 497
Charbonneau, P. 2010, Liv. Rev. Sol. Phys., 7, 3
Charbonneau, P., Beaubien, G., & St.-Jean, C. 2007, ApJ, 658, 657
Dikpati, M., Gilman, P. A., & de Toma, G. 2008, ApJ, 673, L99
Georgieva, K. 2011, ISRN Astronomy and Astrophysics, 2011, 437838
Gnevyshev, M. N. 1967, Sol. Phys., 1, 107
Gnevyshev, M. N. 1977, Sol. Phys., 51, 175
Gnevyshev, M. N., & Ohl, A. I. 1948, Astron. Zh., 25, 18
Gupta, S. S., Sivaraman, K. R., & Howard, R. F. 1999, Sol. Phys., 188, 225
Győri, L., Baranyi, T., & Ludmány, A. 2011, in Physics of Sun and Star Spots, eds. D. Prasad Choudhary, & K. G. Strassmeier, IAU Symp., 273, 403
Hasan, S. S., Mallik, D. C. V., Bagare, S. P., & Rajaguru, S. P. 2010, Astrophys. Space Sci. Proc., 19, 12
Hathaway, D. H. 2015, Liv. Rev. Sol. Phys., 12, 4
Hazra, G., Karak, B. B., Banerjee, D., & Choudhuri, A. R. 2015, Sol. Phys., 290, 1851
Howard, R. F., Gupta, S. S., & Sivaraman, K. R. 1999, Sol. Phys., 186, 25
Howard, R. F., Sivaraman, K. R., & Gupta, S. S. 2000, Sol. Phys., 196, 333
Huang, N. E., Shen, Z., Long, S. R., et al. 1998, Proc. Roy. Soc. London Ser. A, 454, 903
Jiang, J., Cameron, R. H., Schmitt, D., & Schüssler, M. 2011, A&A, 528, A82
Karak, B. B., & Choudhuri, A. R. 2011, MNRAS, 410, 1503
Krivova, N. A., & Solanki, S. K. 2002, A&A, 394, 701
Lean, J. 1990, ApJ, 363, 718
Lean, J. L., & Brueckner, G. E. 1989, ApJ, 337, 568
Lockwood, M. 2001, J. Geophys. Res., 106, 16021
Obridko, V. N., & Shelting, B. D. 2007, Adv. Space Res., 40, 1006
Oliver, R., & Ballester, J. L. 1994, Sol. Phys., 152, 481
Oliver, R., Ballester, J. L., & Baudin, F. 1998, Nature, 394, 552
Ravindra, B., Priya, T. G., Amareswari, K., et al. 2013, A&A, 550, A19
Sivaraman, K. R. 2000, JA&A, 21, 149
Sivaraman, K. R., Gupta, S. S., & Howard, R. F. 1993, Sol. Phys., 146, 27
Sivaraman, K. R., Gupta, S. S., & Howard, R. F. 1999, Sol. Phys., 189, 69
Sivaraman, K. R., Sivaraman, H., Gupta, S. S., & Howard, R. F. 2003, Sol. Phys., 214, 65
Temmer, M., Rybák, J., Bendík, P., et al. 2006, A&A, 447, 735
Torrence, C., & Compo, G. P. 1998, Bull. Am. Meteorological Soc., 79, 61
Vaquero, J. M. 2007, Adv. Space Res., 40, 929
Waldmeier, M. 1935, Mitt. Eidgen. Sternw. Zurich, 14, 105
Watson, F., Fletcher, L., Dalla, S., & Marshall, S. 2009, Sol. Phys., 260, 5
Yoshida, A., & Yamagishi, H. 2010, Ann. Geophysicae, 28, 417
Zhan, L.-S., He, J.-M., Ye, Y.-l., & Zhao, H.-J. 2006, Chin. Astron. Astrophys., 30, 393

Appendix A

Sunspot area distribution has been studied in the past by Bogdan et al. (1988), Baumann & Solanki (2005). These authors used a lognormal distribution function to fit the area data. We also follow the same and fitted the data with a lognormal function defined as

$$y = \frac{1}{\sqrt{2\pi}\sigma x} \exp{-\frac{[\ln(x) - \mu]^2}{2\sigma^2}} \tag{A.1}$$

where μ and σ are the mean and the standard deviation, respectively. The fits on the histograms of the sunspot area distributions for the full period and for the individual cycles are shown

in Fig. A.1. Here, we must emphasize the fact that the lognormal distribution first increases and then decreases exponentially as $\exp(\mu - \sigma^2)$. Since the initial increment of the lognormal distribution is restricted very close to the origin of the histogram, it almost mimics the exponentially decaying distribution, as shown previously in Fig.7.

Also, we plot the same quantity as in panel a) of Fig. A.1 but in a natural logarithmic scale in Fig. A.2. We also fit data with a lognormal function as shown with the solid blue line in the figure. The lognormal function appears as a parabola in a log-log space.
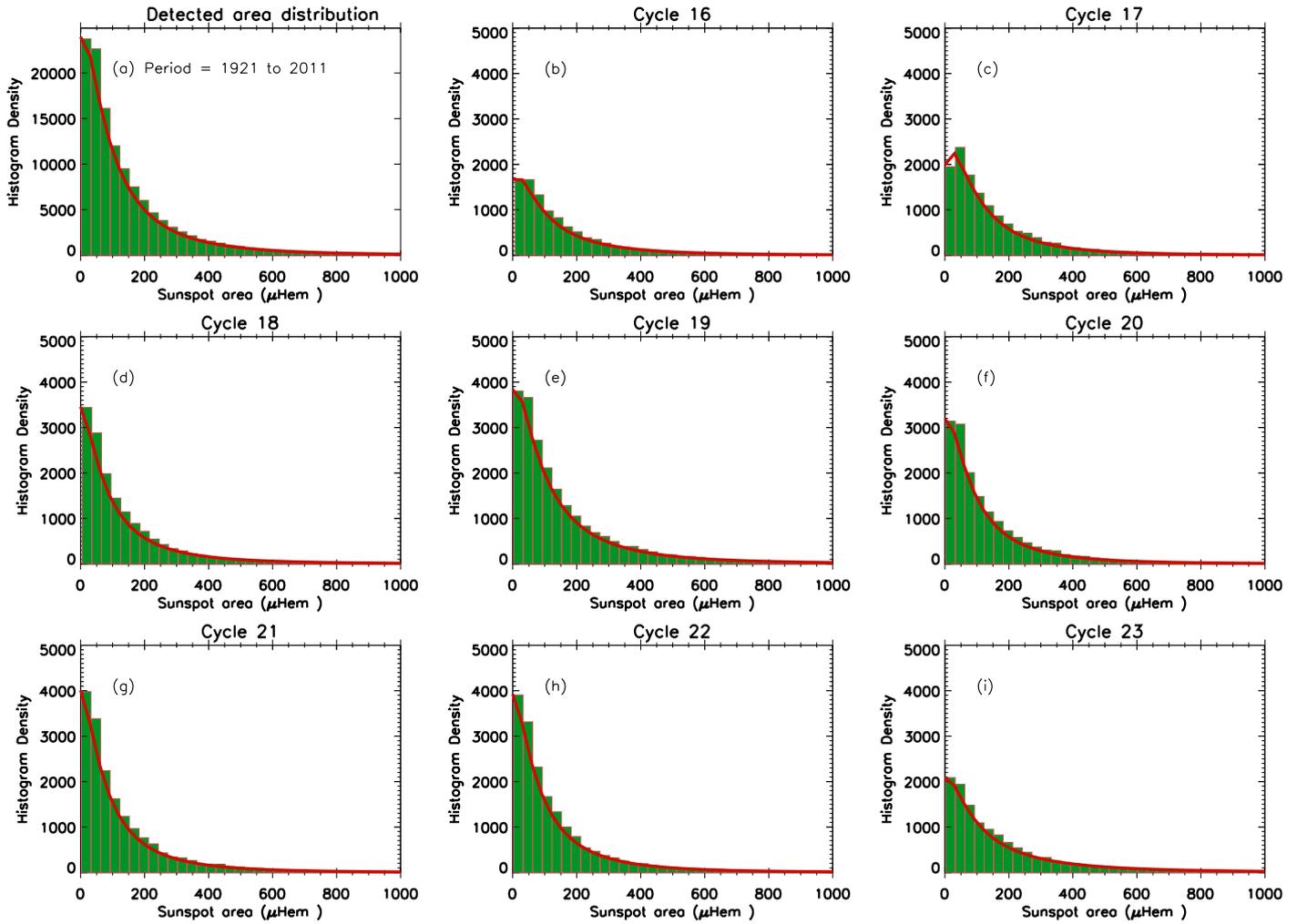

Fig. A.1. Histograms (with a bin size of 30 μHem) of the individual area values (sizes) obtained from Kodaikanal white-light images. The histogram for the full data set is shown in panel **a)**. Panels **b)–i)**: area distributions for the individual solar cycles (cycle 16 to 23).

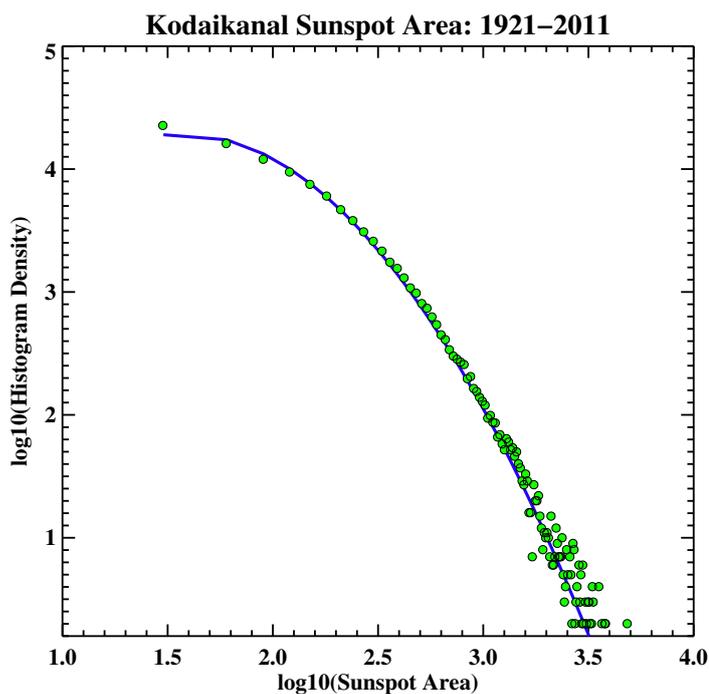**Fig. A.2.** Same as panel **a)** of Fig. A.1 but in a log-log scale. The blue solid line shows the fitted lognormal function.

Appendix B

We also compare our Kodaikanal sunspot area values with the Debrecen data (Győri et al. 2011; Baranyi et al. 2016) which covers the period from 1974 to 2016 (and continuing). Though some of the gaps in the data have been filled in using data from other observatories, the catalog is mainly based on Debrecen and Gyula full-disk white-light observations. We further compute the correlation coefficient between the Kodaikanal and the Debrecen data sets and it is equal to 0.97.
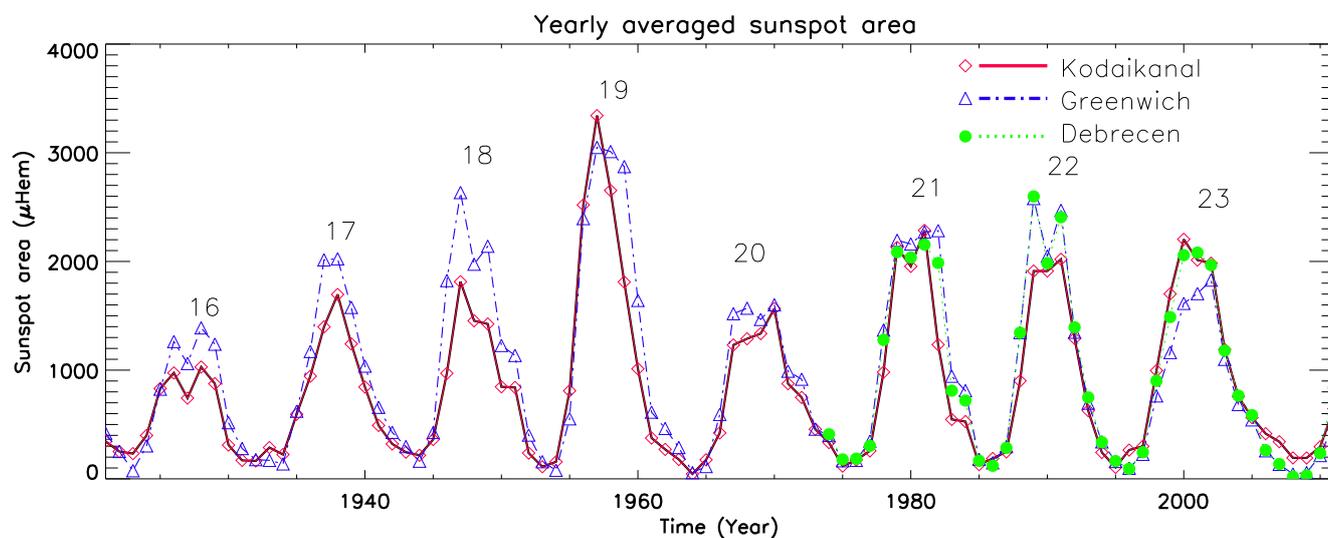

Fig. B.1. Yearly averaged sunspot-area values from Kodaikanal, Greenwich, and Debrecen sunspot catalogs.