








## Butterfly Diagram and other properties of plage areas from Kodaikanal Ca II K Photographs Covering 1904 – 2007

Bibhuti Kumar Jha<sup>1,2,3</sup>  ·  
Theodosios Chatzistergos<sup>4</sup>  ·  
Dipankar Banerjee<sup>2,3,5</sup>  · Ilaria Ermolli<sup>6</sup>  ·  
Natalie A. Krivova<sup>4</sup>  · Sami K. Solanki<sup>4</sup>  ·  
Aditya Priyadarshi<sup>2,3</sup> 

© The author(s) ●●●●

### Abstract

Ca II K observations of the Sun have a great potential for probing the Sun's magnetism and activity, as well as for reconstructing solar irradiance. The Kodaikanal Solar Observatory (KoSO) in India, houses one of the most prominent Ca II K archives, spanning from 1904 to 2007, obtained under the same experimental conditions over a century, a feat very few other sites have achieved. However, the KoSO Ca II K archive suffers from several inconsistencies (e.g., missing/incorrect timestamps of observations and orientation of some images) which have limited the use of the archive. This study is a step towards bringing the KoSO archive to its full potential. We did this by developing an automatic method to orient the images more accurately than in previous studies. Furthermore, we included more data than in earlier studies (considering images that

---

✉ B. K. Jha  
[maitraibibhu@gmail.com](mailto:maitraibibhu@gmail.com)

✉ T. Chatzistergos  
[chatzistergos@mps.mpg.de](mailto:chatzistergos@mps.mpg.de)

- <sup>1</sup> Southwest Research Institute, Boulder, CO 80302, USA
- <sup>2</sup> Indian Institute of Astrophysics, Koramangala, Bangalore 560034, India
- <sup>3</sup> Aryabhata Research Institute of Observational Sciences, Nainital 263000, India
- <sup>4</sup> Max Planck Institute for Solar System Research, Justus-von-Liebig-Weg 3, D-37077 Göttingen, Germany
- <sup>5</sup> Center of Excellence in Space Sciences India, IISER Kolkata, Mohanpur 741246, West Bengal, India
- <sup>6</sup> INAF Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monte Porzio Catone, Italy

could not previously be analyzed by other techniques as well as 2845 newly digitized images), while also accounting for mistakes in the observational date/time. These images were accurately processed to identify plage regions along with their locations, enabling us to construct the butterfly diagram of plage areas from the entire KoSO Ca II K archive covering 1904–2007. Our butterfly diagram shows significantly fewer data gaps compared to earlier versions due to the larger set of data used in this study. Moreover, our butterfly diagram is consistent with Spörer’s law for sunspots, validating our automatic image orientation method. Additionally, we found that the mean latitude of plage areas calculated over the entire period is  $20.5\% \pm 2.0$  higher than that of sunspots, irrespective of the phase or the strength of the solar cycle. We also studied the North-South asymmetry showing that the northern hemisphere dominated plage areas during solar cycles 19 and 20, while the southern hemisphere dominated during solar cycles 21–23.

**Keywords:** Chromosphere, Quiet; Chromosphere, Active; Solar Cycle, Observations;

## 1. Introduction

The discovery of two dominant absorption lines of singly ionized Calcium (Ca II) in the solar spectrum by Joseph von Fraunhofer in 1814 opened a new window to multi-wavelength solar observation. These two lines, known as H and K lines, probe the solar atmosphere from the photosphere to the chromosphere and serve as excellent diagnostics of the chromospheric structure and dynamics (e.g. Linsky and Avrett, 1970; Murabito et al., 2023) as well as of the solar and stellar chromospheric activity (e.g. Eberhard and Schwarzschild, 1913; Wilson, 1968; Sowmya et al., 2021). Moreover, solar full disk Ca II K observations provide unique information on chromospheric plage and network regions, which are observed overlying photospheric magnetic features, such as faculae and network magnetic elements, and hence act as keys for studying the solar surface magnetic field’s evolution (e.g. Skumanich, Smythe, and Frazier, 1975; Schrijver et al., 1989; Harvey and White, 1999; Loukitcheva, Solanki, and White, 2009; Kahil, Riethmüller, and Solanki, 2017; Chatzistergos et al., 2019), in particular when magnetograms are not available. Therefore, Ca II K observations are an important asset for studying the evolution of plage (e.g. Chatterjee, Banerjee, and Ravindra, 2016; Ermolli et al., 2018; Chatzistergos et al., 2019; Chatzistergos et al., 2020a; Chatzistergos et al., 2022; Chatzistergos, 2023; Singh et al., 2022), network (Ermolli et al., 1998; Ermolli, Giorgi, and Chatzistergos, 2022; Berrilli et al., 1999; Chatterjee, Mandal, and Banerjee, 2017), and reconstructing solar magnetism (e.g. Chatzistergos et al., 2019; Chatzistergos, Krivova, and Ermolli, 2022; Mordvinov et al., 2020, and references therein). Furthermore, since the evolution of the solar surface magnetic field affects the Sun’s radiative output, Ca II K observations are also very useful for reconstruction of past total solar irradiance variations (e.g. Chatzistergos et al., 2020c, 2021; Chatzistergos, Krivova, and Ermolli, 2024).

Although the existence of absorption lines in the solar spectrum was known from the beginning of the 19th century (e.g. Secchi, 1871; Tacchini, 1875; Ermolli and Ferrucci, 2021; Ermolli et al., 2023), systematic and photographic observations of the Sun over specific wavelength intervals started only after the invention of spectroheliograph (Hale, 1890, 1891, 1893; Deslandres, 1908). Since then, synoptic observations of the Sun in Ca II K have been performed from many sites around the globe (see Chatzistergos et al., 2020b; Chatzistergos, Krivova, and Ermolli, 2022, and references therein), such as those at Kodaikanal (since 1904; Priyal et al., 2014a), Mt. Wilson (since 1915; Lefebvre et al., 2005), Mitaka (since 1917; Hanaoka, 2013), Arcetri (since 1931; Ermolli et al., 2009) and Meudon (since 1893; Malherbe et al., 2022; Malherbe, 2023). The fact that these observations were until recently available only in the form of physical photographs, has been a hindrance to analysing them. This changed dramatically over the last decades, during which most of the existing Ca II K archives were digitized. By now there is a wealth of digitally available solar Ca II K data (for a review see Chatzistergos, Krivova, and Ermolli, 2022). Among the available archives, Kodaikanal Solar Observatory (KoSO; Hasan et al., 2010) has one of the longest records of solar observations in Ca II K covering more than 100 years (1904–2007). The fact that essentially the same instrumental setup was used for the entire period of observation, makes this dataset an important resource for investigating solar processes over a century-long time scale.

Indeed, the KoSO Ca II K data have been used for various studies, with determining plage areas being the most common one (e.g. Chatterjee, Banerjee, and Ravindra, 2016; Chatzistergos et al., 2019b; Singh et al., 2022). The majority of the past analyses determined disk-integrated quantities and thus did not require any specific orientation of the solar disk. However, there are several applications for which it is essential to know quite accurately the orientation of the solar disk in the images. This includes studies of e.g., the chromospheric solar differential rotation (Singh and Prabhu, 1985; Mishra et al., 2023; Kharayat et al., 2024), the polar network index (Priyal et al., 2014b; Mishra et al., 2024), as well as deriving time-latitude maps, commonly referred to as “butterfly” diagrams useful for studying the evolution of the solar activity and the properties of magnetic regions.

To our knowledge only a few such diagrams have been produced from Ca II K data (Harvey, 1992; Ermolli et al., 2009; Chatterjee, Banerjee, and Ravindra, 2016; Priyal et al., 2017; Bertello, Pevtsov, and Ulrich, 2020), two of which were with KoSO observations. Hindrances for producing such diagrams have been the difficulty in properly orienting the images, as well as difficulties in accurately processing them. The latter has been overcome over the last few years with the development of methods for the high quality processing of photographic full disk data (Chatzistergos, Krivova, and Ermolli, 2022; Chatzistergos et al., 2023). The situation is worse for the image orientation. Previous studies with KoSO data utilized markings on the plates (examples are shown in Fig. 1) to orient the images whether manually clicking on them with the computer mouse (Priyal et al., 2014a) or having an automatic process to identify their locations (Chatterjee, Banerjee, and Ravindra, 2016). However, relying on such markings introduces significant uncertainty and bias in the determined orientation, due to

potential mistakes in the placement of the markings on the plates, and also in the identification of their location in the digital images. Furthermore, not all of the observations include such markings. In particular, the earliest data in the archive, that is before 1907, were not complemented with pole markings at all and thus have not been incorporated in any butterfly diagram yet. Achieving an accurate orientation of the KoSO data is particularly important for any investigation extending back to those times.

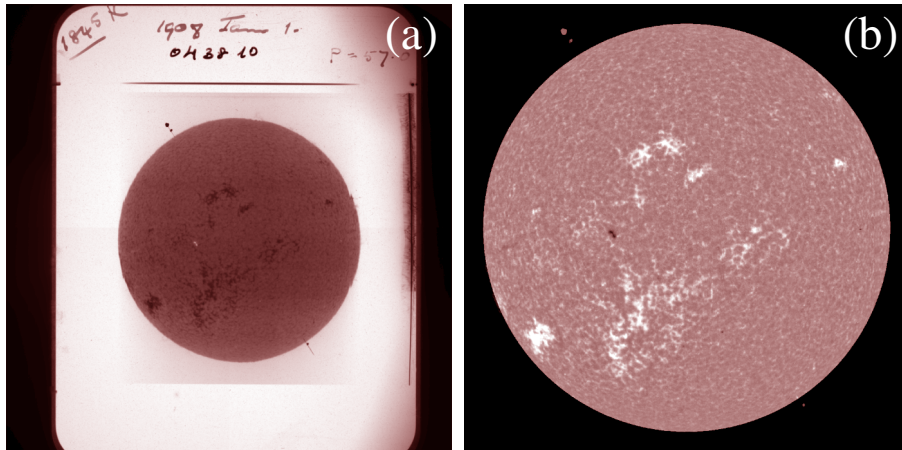
To further facilitate the exploitation of the KoSO data to their full potential, in this study we present an automatic and versatile method we developed to orient the observations without the need to rely on the pole markings. We applied the method to the data accurately processed with the techniques developed by (Chatzistergos et al., 2018, 2019a), and produced a butterfly diagram of plage areas from the entire data collection of KoSO, covering 1904–2007. In Section 2, we will briefly describe the data and the processing we applied on them. In Section 3 we present the updated plage latitude vs. time diagram (i.e. plage butterfly diagram). We conclude with a summary and discussion in Section 4.

## 2. Data and their Processing

The main features of the KoSO Ca II K observations and methods employed to digitize, process and calibrate them are described in several papers, so that we restrict ourselves presenting a summary only.

### 2.1. KoSO Photographic Archive

KoSO is among the first observatories to start regular multi-wavelength surveys of the Sun at the beginning of the 20th century (Hasan et al., 2010), in white light and in the Ca II K line since 1904, and in the H $\alpha$  line since 1912. A Hale spectroheliograph, which produces a 30 mm image of the Sun with the help of a 30 cm Cooke photo-visual triplet, is used to take the Ca II K observations by scanning the disk of the Sun. A Foucault siderostat, containing a single mirror of 46 cm diameter, reflects the beam of light on a 30 cm lens, which is ultimately fed to a spectroheliograph to obtain the observations (Bappu, 1967). This spectroheliograph consists of two prisms which disperse the light, and then a Lyot filter is used to ensure that only a narrow spectral region centered at 393.367 nm wavelength with a pass-band of 0.05 nm passes through it (Bappu, 1967; Priyal et al., 2017). These observations were recorded on photographic glass plates until February 1989, while after that they were taken on photographic films (in the following we will refer to both glass plates and film as plates). There were also a few instances before February 1989 when observations were stored on films. The plates were developed in a pitch-dark room and then they were placed inside an envelope and stored in a humidity-free environment. The date and time of observations ( $T_{\text{OBS}}$ ) were written on the margin of the plates as well as on the envelope before archiving/preservation.



**Figure 1.** (a) Representative image of the photographic plate with the raw negative observation taken on 1st January 1908 at 10:08 IST (04:38 UTC) from the KoSO Ca II K digital archive, (b) the calibrated image of the corresponding observation derived from the careful calibration and processing of the raw negative data. The calibrated image is shown in the range of  $[-0.4, 0.4]$  in contrast values, while the raw image is shown to its full range of values. Note the markings to denote the poles of the Sun, two circles for the North pole and one small circle for the South pole.

## 2.2. Digitization

To allow the larger community to utilize this extensive dataset, it was essential to transform it into digital format. Two such large-scale digitizations have been performed for the KoSO archive (see Chatzistergos et al., 2019b). The first such attempt was restricted to only a sample of the data covering 1907–1999 (Makarov et al., 2004). The digitization was performed with a linear array of 900 pixel and the images were stored in 8-bit JPG file format (Makarov et al., 2004). However, some inhomogeneities were reported in the digitised KoSO data (see Chatzistergos et al., 2019b). Furthermore, an important part of the data collection, that is the data over 1904–1907 and 2000–2007, were completely missing. This led to an effort to re-digitise the KoSO data. The second digitization was performed with a better scanning device than the one used for the first digitisation and it aimed at including the entire data collection covering 1904–2007 (Priyal et al., 2014a; Chatterjee, Banerjee, and Ravindra, 2016). The digitization setup consisted of a 1 m uniform source sphere with a 35 cm opening. The source sphere contained 3–4 current control light sources to provide highly uniform lighting. At the opening of the sphere, a plate holder with a horizontal slider was used to hold the plates. A cryogenic-cooled (at  $-100^{\circ}\text{C}$ )  $4\text{k} \times 4\text{k}$  CCD camera with a 16-bit readout, attached to a vertical slider, was used for digitizing the Ca II K observations taken on photographic plates. The same setup with a film holder instead of a plate holder is used for digitizing the observations taken on photographic films. After digitization, these images were stored in 16-bit FITS (Flexible Image Transport System; Wells, Greisen, and Harten, 1981) format, with a pixel scale of  $\approx 0.9$  arcsec/pixel (an example image

is shown in [Figure 1a](#)). A set of five flat-field images are taken for each month of data after digitization for that particular month, whereas the dark is subtracted during the digitization itself (Priyal et al., 2014a). Finally, the filename of FITS images is given based on the date and time of observation ( $T_{\text{OBS}}$ ), noted from the plate or envelope carrying these plates, which is required for using the data.

The above digitization setup and procedure are now permanently available at KoSO for digitization of photographic observations. Indeed, recently, we found some inconsistencies in the digitized data for six years (1957–1962), and we also recovered a few more plates in the archive missed by both previous digitization drives, mostly between 1957 and 1961. Hence, we have re-digitized all the data for this period with the same setup as used for the latest digitization, in order to remove any data in-homogeneity and incorporate the newly-found plates. In this work, we use the data from the second digitization (Priyal et al., 2014a), along with the re-digitized data over (1957–1962). In total there are 51773 images, which is 29615 more images than analyzed by Makarov et al. (2004); Tlatov, Pevtsov, and Singh (2009); Ermolli et al. (2009) based on the first digitization of the KoSO Ca II K archive, and 2845 more images than analyzed by Priyal et al. (2014a); Chatterjee, Banerjee, and Ravindra (2016) from the second digitization of the KoSO Ca II K archive.

### 2.3. Calibration

We used the KoSO images processed by Chatzistergos et al. (2019b, 2020a) along with newly digitised images over 1957–1962, which were processed with exactly the same techniques as the rest of the data (see Chatzistergos et al., 2018, for details). The processing comprised three main steps: (i) pre-processing; (ii) photometric calibration; and (iii) compensation for limb-darkening and image artifacts. The pre-processing includes the flat-fielding of the CCD-camera used for the digitisation (Chatzistergos et al., 2019b), conversion of transparency (negative) to density (positive) values, and the identification of the solar disk. The latter was achieved by applying a Sobel filter and using a threshold to isolate the rim of the solar disk (Chatzistergos et al., 2020a). An ellipse was fitted to the isolated rim to acquire the values of the center of the ellipse and the semi-major and -minor axes. These were then used to resample the images and return circular solar disks (Chatzistergos et al., 2020a,b). Such corrections were required due to image distortions introduced during the observation (see Chatzistergos et al., 2020a,b).

The photometric calibration is required because the emulsions used to store the photographs have a non-linear response to the incident light. The approach applied to recover the information needed for the photometric calibration is to relate the density centre-to-limb variation (CLV) of the darker parts of the quiet Sun (QS) regions measured in the photographic data to intensity QS CLV from modern CCD-based observations (Chatzistergos et al., 2018, 2019a). In this way we construct a calibration curve for each image separately. Finally, compensation for the limb darkening (Chatzistergos et al., 2018, 2019a) allows to create contrast images ([Figure 1b](#)) of the form  $C_{i,j} = (I_{i,j} - I_{i,j}^{\text{QS}})/I_{i,j}^{\text{QS}}$ , where  $C_{i,j}$ ,  $I_{i,j}$ , and  $I_{i,j}^{\text{QS}}$  are the contrast relative to the quiet Sun, intensity, and

intensity of the quiet Sun for pixels  $(i, j)$ . The map of  $I_{i,j}^{\text{QS}}$  was computed with an iterative approach, involving a 2D running-window median filter as well as polynomial fitting along rows, columns, and radial segments of the solar disk after removing the active regions (Chatzistergos et al., 2018). This step allows for the removal of large scale image artifacts due to the methods applied for the acquisition and storage of the observation. The superior accuracy of the above processing method compared to others from the literature has been extensively discussed by Chatzistergos et al. (2018, 2019a, 2020a,c); Chatzistergos, Krivova, and Ermolli (2022) and Chatzistergos et al. (2023).

#### 2.4. Pole Angle Calculation

The accurate orientation of KoSO Ca II K observations results from application of a few processing steps. First, we describe how we calculate the pole angle (orientation), i.e. the direction of solar North with respect to the vertical axis of a digitized image for a given date and time of observation ( $T_{\text{OBS}}$ ). Assuming that the siderostat at KoSO reflects the sunlight perfectly in the geographical South direction along the meridian, the rotation,  $\alpha$ , of the field of view (FOV) caused by the siderostat is given by (Cornu, 1900):

$$\alpha = 2 \tan^{-1} \left[ K \tan \left( \frac{\text{HA}}{2} \right) \right], \quad (1)$$

where

$$K = \frac{\sin \left( \frac{L-\delta}{2} \right)}{\sin \left( \frac{L+\delta}{2} \right)}.$$

In Equation 1, HA is the hour angle of the Sun (Smart, 1977),  $L$  is the latitude of the observatory, and  $\delta$  is the pole angle ( $90^\circ$  - declination) of the Sun at  $T_{\text{OBS}}$ . To calculate hour angle of the Sun we have used a IDL routine in SolarSoftWare (SSW; Freeland and Handy, 1998) *sunpos.pro* and *eq2hor.pro*. We use the following co-ordinate of the observatory Latitude( $L$ )= $10.23^\circ$  ( $10^\circ 13' 50''$ )N and Longitude =  $77.47^\circ$  ( $77^\circ 28' 7''$ )E (<https://www.iiap.res.in/kodai.location>). The orientation of digitized images is in such a way that, to get the solar North at the top of the images we need to rotate them by the complementary angle of  $\alpha$ . Therefore, the pole angle ( $\Theta$ ) to bring the solar North to the top of the image is given by

$$\Theta = 90^\circ - \alpha. \quad (2)$$

It is obvious from Equations 1 and 2 that, getting  $\Theta$  is straightforward, provided we have the correct  $T_{\text{OBS}}$ .

Prior to the use of Equation 1 and 2 for the calculation of  $\Theta$ , we compared  $\Theta$  calculated with the method of Priyal et al. (2014a). In most cases, the difference amounts to a few degrees ( $\pm 2^\circ$ ), but for  $\approx 0.5\%$  of the data the differences are substantial ( $> 10^\circ$ ). Manual inspection of these cases proved that our approach returned the correct orientation, while the pole markings were erroneously placed on the plates. A more detailed discussion of the comparison can be found in



[Appendix A](#). Note that the above comparison is based on the assumption that all the  $T_{\text{OBS}}$  is correct, which is not the case as we discuss in the following sections.

### 2.5. Inconsistencies in time of observations

The approach described above allows to get an accurate  $\Theta$  if  $T_{\text{OBS}}$  is known precisely. However, some inconsistencies in the date and time of observations of the digitized images have previously been reported (e.g. Chatzistergos et al., 2019b; Jha, 2022), which can affect the image orientation as well as results from any analysis with inexactly oriented data. For instance, this has the potential implication of previously derived plage areas including values that have been assigned to the wrong time periods. It is thus worth trying to tackle those inconsistencies (see Jha, 2022).

We note that, there are systematic inconsistencies in  $T_{\text{OBS}}$  of the files as well as random mistakes. Systematic inconsistencies arose due to conversion from Coordinated Universal Time (UTC) to Indian Standard Time (IST), because of either an incorrect conversion or a change of the convention used when writing the data on the plate/folder. Unfortunately, there are also random mistakes in the date and time of observation, which could have arisen at any stage, either when the date/time was written on the plate/folder or when the plate was digitized and the time was used to name the file. To illustrate these inconsistencies, in [Figure 2a](#) we show the time of observation, extracted from the filenames, as a function of the observational date. In [Figure 2a](#) we notice the following:

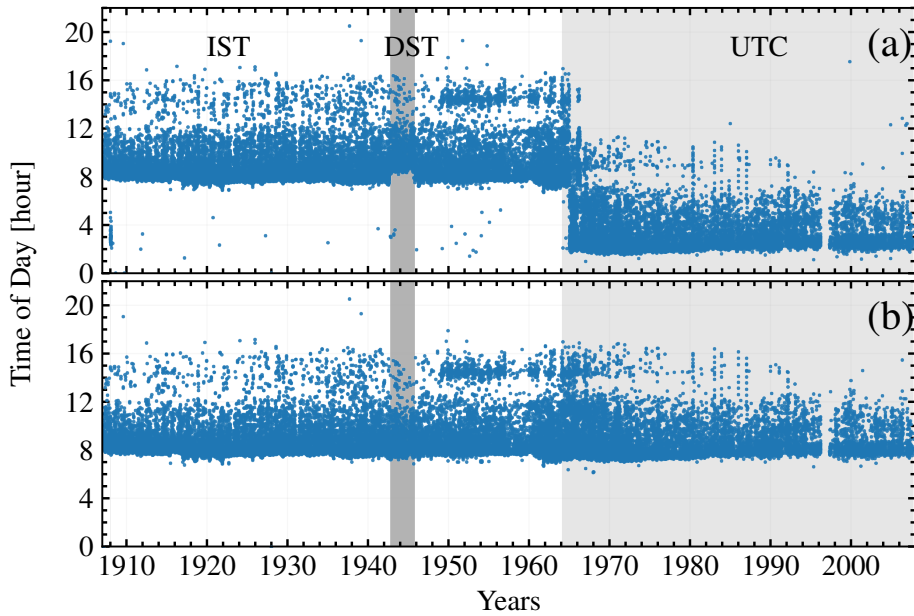
1. Before February 1964 time is recorded predominantly in IST (+05:30 UTC), while UTC was used afterwards. However, we found that in the years 1964 (February on-wards) and 1966 (mostly in the first half), 97% and 43% of data, respectively, have been converted to IST from UTC, during the digitization itself.
2. There is a systematic shift of one hour in the time of observation, from September 1, 1942 to October 15, 1945 due to the Daylight Saving Time (DST; +06:30 UTC)<sup>1</sup> followed in India.
3. From January 1 to January 24, 1908, most of the  $T_{\text{OBS}}$  is in UTC (timestamp in [Figure 1](#) is in UTC), with few exceptions which are in IST.
4. Observation times beyond the usual daylight period (06:30 to 17:00), which are primarily due to mistakes in naming of files during digitization or confusion between IST and UTC.

We have accounted for the above inconsistencies between IST and UTC. However, as we previously mentioned, there are also mistakes in the dates/times of the observations which might have gone unnoticed. Furthermore, we cannot

---

<sup>1</sup><https://web.archive.org/web/20230517151641/https://www.timeanddate.com/time/change/india/kolkata>





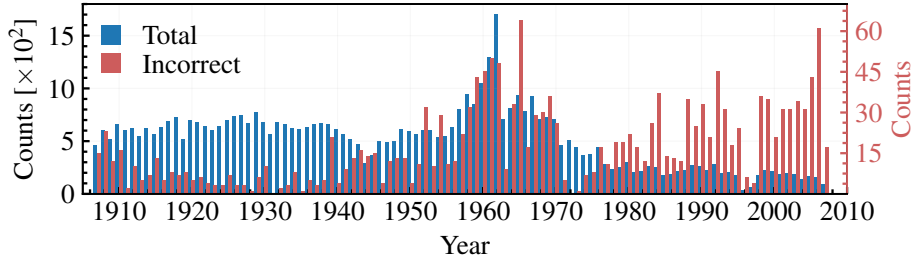
**Figure 2.** (a) Time of observation at KoSO over the period of observations. Different background shadings are used to mark the different time conventions adopted, IST (+05:30 UTC; white), UTC (light grey) and DST (+06:30 UTC; dark grey), (b) same as panel (a) after applying the corrections of time of observations described in Section 2.5.

rule out that some of the isolated outliers in Figure 2a might have been due to mistakes in writing the date/time and not due to conversion between IST and UTC of the series. We note that the data in Figure 2 only cover the period 1907–2007, i.e. they lack the first 3 years of KoSO Ca II K observations, which are short of information on the  $T_{\text{OBS}}$ .

We correct the inconsistencies in the  $T_{\text{OBS}}$  in two strides, each consisting of several steps. These are briefly discussed in the following, while more details are given in Appendix B. In an attempt to bring all  $T_{\text{OBS}}$  values to the same time zone, and since most of the timestamps are in IST, we decided to go with IST. We stress that this refers to the time stored in the header of the FITS files, the time is converted to UTC for all images when computing the butterfly diagram. As the first stride of correction we considered accounting for the major inconsistencies in  $T_{\text{OBS}}$  as they were pointed out earlier. In the second stride, we have attempted to identify and correct residual mistakes in the  $T_{\text{OBS}}$  following Jha (2022, their Section 2.3). In Figure 2b we show the corrected  $T_{\text{OBS}}$  as a function of date of observation. Additionally, we also note that there is gradual change in  $T_{\text{OBS}}$  post 1960 which is an attribute of the early observation in the morning at KoSO, but the reason behind the start of early observation is unknown to us.

After applying all these corrections, we found that  $\approx 3.4\%$  of observations still remain with potentially incorrect  $T_{\text{OBS}}$ (timestamps). The most probable reason for the incorrect  $T_{\text{OBS}}$  of these observations is due to the mistake in noting  $T_{\text{OBS}}$

during the observations. In [Figure 3](#) we show the yearly counts of images with potentially incorrect  $T_{\text{OBS}}$  along with total number of observations. Note the different scales used for the two counts in [Figure 3](#). [Figure 3](#) shows an increase of the number of images with incorrect  $T_{\text{OBS}}$  over the periods 1957–1970 and 1975–2007, which cover Solar Cycles 19–20 and the most recent decades, thus potentially affecting the behavior of the envelope and long-term evolution of the solar activity derived from analysis of KoSO Ca II K observations.



**Figure 3.** Number of Ca II K images per year in the KoSO archive (blue) along with the number of images that we estimate to have incorrect  $T_{\text{OBS}}$  even after our reduction to same time convention (red and ordinate axis on the right-hand side; see [Section 2.5](#)).

## 2.6. Recovering the image orientation over 1904–1906

We found that approx 3.4% of the data suffer from potentially incorrect  $T_{\text{OBS}}$  after conversion of the  $T_{\text{OBS}}$  archived with the observation to the same time convention. Besides, we also note that observations taken before 1907 do not carry a time of observation at all (only the date is available). Furthermore, we observe that the digitized images derived from the KoSO archive may suffer from further orientation issues due to, e.g., how exactly the plate was placed in the digitization device.

To overcome all the above problems, we developed a cross-correlation technique to determine the orientation of the KoSO Ca II K observations. The method tries all possible orientations between two subsequent images and selects the one that minimizes the sum of squared residuals of their contrast values as the correct one. This is done in two iterations. First, all angles are checked with a step of  $1^\circ$ , while the second iteration reduces the step to  $0.1^\circ$  and restricts the search of the angle within  $\pm 1^\circ$  of the optimum value returned during the first iteration. We use the calibrated and limb-compensated images to remove as much as possible the influence of limb darkening and image artifacts, which would have affected the results from the cross-correlation of the images. The images were then resized to have the same dimensions (we note that this was done only for the estimation of the rotation, otherwise the original file size was maintained). We also applied differential rotation to rotate every image in question to  $T_{\text{OBS}}$  of the corresponding image acting as the reference (the one with known orientation) to match their observing times. This is done using *drot\_map.pro*<sup>2</sup> of SSW. To work

<sup>2</sup>See [https://hesperia.gsfc.nasa.gov/ssw/gen/idl/maps/drot\\_map.pro](https://hesperia.gsfc.nasa.gov/ssw/gen/idl/maps/drot_map.pro) for details about the routine.

correctly, this in turn requires that the reference image of each comparison is properly oriented. For orienting the images we used the observation available in the archive that is closest in time and has been already oriented. We restricted the search for such a fiducial image to within a 3-day interval around the time the image to be rotated was recorded. When there were no data available within the desired limits, we used observations from the Meudon dataset (as analysed by Chatzistergos et al., 2020a) to act as reference.

We used the above cross-correlation approach on all data before 1907 to determine the pole (rotation) angle. However, we also used the cross-correlation approach on the data after 1907 in order to validate our pole angle calculation based on Equations 1 and 2 and to identify any images with wrong timing or possible misalignment of the plate during the digitization. To distinguish between potentially correct and wrong orientations we imposed a threshold of  $5^\circ$  between the rotation angle determined with Equations 1 and 2 and the one from the cross-correlation approach. If the difference between the two angles was greater than the threshold, or if the cross-correlation returned a different flipping (e.g., East-West Flipped in the cases where images are scanned from the wrong side), then we checked the observation date in the raw images to identify if there was a mistake in the date/time of the digital file. However, there were still a few instances, where we were unable to correct the observational date and time of the images (e.g. if there was no date information written within the scanned area of the plates) or there was still disagreement between the rotation angle derived with the cross-correlation and the one with Equations 1 and 2 (suggesting that there might have been a mistake in writing the date/time on the plate). In these cases we simply stored the rotation angle determined from the cross-correlation in the FITS files header and used this one instead to align the images appropriately.

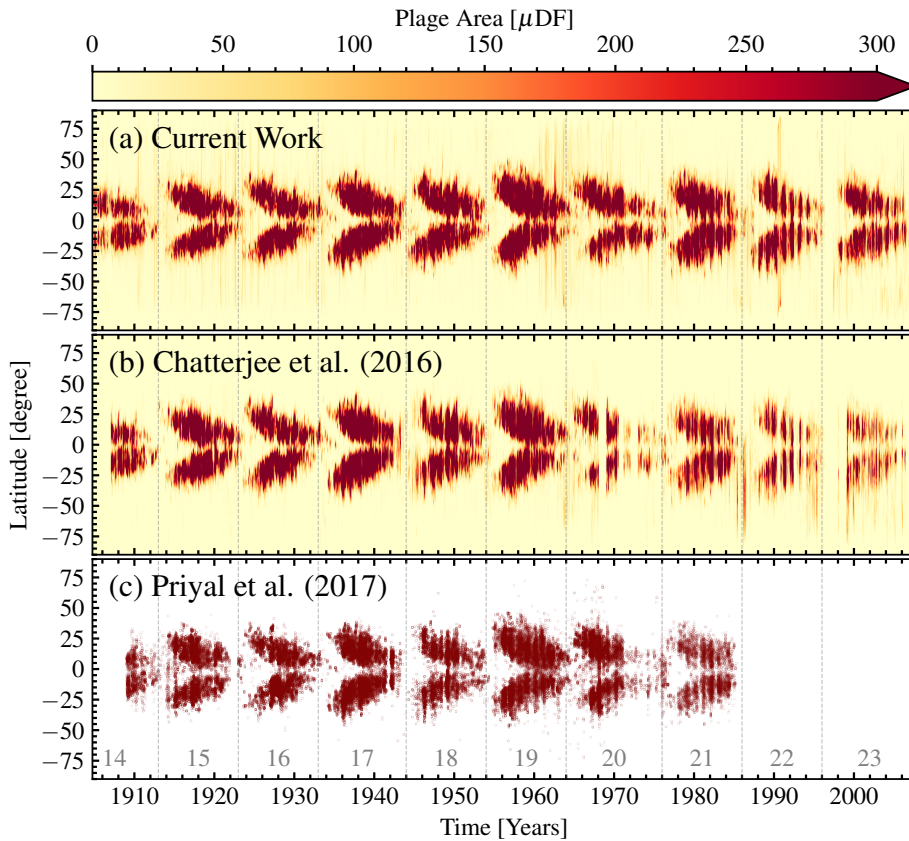
## 2.7. Segmentation

The reliable orientation of the KoSO Ca II K observations accurately processed with the methods developed by Chatzistergos et al. (2018) allows for investigations of plage regions going beyond those that could be achieved in previous studies by e.g. Chatzistergos et al. (2020a). In this light, we also segmented the oriented data following Chatzistergos et al. (2019a). In brief, plage regions were identified with a contrast threshold, which was taken as a multiplicative factor to the standard deviation of the quietest parts of the disk. The level of the quiet Sun regions was identified with a variant of the method by Nesme-Ribes, Meunier, and Collin (1996). This thresholding was used to produce binary maps of the solar disk where the plage regions have the value of one and all other locations the value of 0. We used these masks to compute/derive the butterfly diagram of the daily plage areas. We considered plage areas in millionths of solar disk fraction ( $\mu$ DF) within latitudinal strips of  $1^\circ$ , similarly to (Harvey, 1992). For days when multiple observations exist, we took the average plage area within each latitude bin for all images of that day.

### 3. Results

#### 3.1. Plage Butterfly Diagram

Figure 4a shows the butterfly diagram of plage areas derived from our processing of the KoSO Ca II K observations. It is evident from this figure that similar to sunspots, plage also appear at higher latitudes at the beginning of the solar cycle (SC), and as the SC progresses, they move towards the equator. This behavior is in good agreement with earlier findings of Harvey (1992); Chatterjee, Banerjee, and Ravindra (2016) and Priyal et al. (2017).



**Figure 4.** The plage area millionths of solar disk fraction ( $\mu\text{DF}$ ) for each latitude bin of  $1^\circ$  for each observation, averaged over all the data of the day in case of multiple observations, as a function of time, (a) for current work, (b) generated using the plage regions identified by Chatterjee, Banerjee, and Ravindra (2016) and (c) from Priyal et al. (2017). However, we note that the butterfly diagram with data by Priyal et al. (2017) shows the latitude of the centroid of individual plage regions and not latitudinally averaged areas.

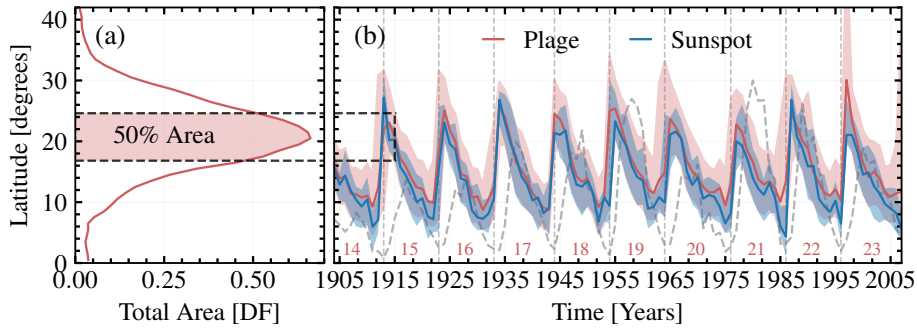
To compare our findings with earlier results from the literature, we have also plotted in Figure 4b and 4c the butterfly diagrams of KoSO plage areas produced by Chatterjee, Banerjee, and Ravindra (2016) and Priyal et al. (2017),

respectively. Both of these studies created their butterfly diagram by defining the centroid latitude of each connected plage region. To make the comparison to our butterfly diagram (which plots areas in latitudinal bins irrespectively of connectivity of plage regions), we use the identified plage region masks provided by Chatterjee, Banerjee, and Ravindra (2016) to create a butterfly diagram in the same way as in this work, i.e., disk fractional area in each latitudinal bin of  $1^\circ$ . However, Priyal et al. (2017) only provides the mean latitude and area for individual plage regions along with the time of observation. Therefore, in Figure 4c, we plot the butterfly diagram as it is presented in their paper. Figure 4 clearly shows the butterfly diagram derived from our processing includes more data (Figure 4a), by covering the entire period of operations of the KoSO and having fewer data gaps, than both the other compared diagrams (Figure 4b and 4c). Both Chatterjee, Banerjee, and Ravindra (2016) and Priyal et al. (2017) ignored many images (mostly after 1960, while Priyal et al. 2017 also ignored all data after 1985) due to their degrading quality, while no data exclusion was performed in our case. Comparison between Figure 4a and 4b also suggests that our processing improves the accuracy of the data representative of the minima at approx 1985 and 1995, i.e. minima SC 21–22, SC 22–23. This could be important when analyzing long-term trends of solar processes.

This further suggests that images taken around 1961 and 1990 still suffer from some inaccuracies and the stripes we see in Figure 4a are due to the presence of artifacts in the digitized images, which are falsely identified as plage regions (Chatzistergos et al., 2019a). These stripes are not present in Figure 4b simply because those data were not considered in the analysis of Chatterjee, Banerjee, and Ravindra (2016) at all. Moreover, our results also appear to be more consistent than the others for the periods 1942–1945 and 1986 in terms of the distribution of plage over latitude. This is due to the correction of the timing of observations applied by us (see Sect. 2.5).

### 3.2. Latitudinal Extent of the Plage

We used the derived butterfly diagram to investigate the latitudinal extent of plage regions with respect to that of sunspots. For sunspot areas we use the recently updated and extended KoSO sunspot area series by Jha et al. (2022), so that we compare chromospheric plage and photospheric sunspot data obtained under nearly the same observing conditions. We calculate the yearly latitude band in which 50% of plage/sunspot area lies. To do this, we combined the plage/sunspot areas in both hemispheres. Then we summed up all the plage/sunspot areas within each  $1^\circ$  latitude bin over a year (see Figure 5a for the plage observed during the year 1915 as an example). The central latitude band in which 50% of the plage area for the selected year is covered is represented by a shaded region in Figure 5a. In Figure 5b, we plot this latitude band for plage regions as a function of time along with the similarly determined latitude band for sunspots. We calculate the area of plage regions in each latitude bin ( $1^\circ$ ), while for sunspots the mean centroid location of a sunspot is used as the latitude and the areas of all sunspots whose centroids lie within a particular latitude bin



**Figure 5.** (a) Normalized latitudinal distribution of plage areas averaged over the year 1915 after combining both hemispheres. The black dashed horizontal lines and red shading mark the bands between 25% and 75% percentile of the area distribution. (b) The yearly variation of median latitude of KoSO plage areas (solid red line) along with the latitude band that includes 50% of each year's Ca II K plage areas (red shaded surface). Also shown is the latitudinal extent of KoSO sunspot areas by Jha et al. (2022, blue curve for annual median values and blue shaded surface for the band covering 50% of the areas). The Grey dashed line represents the scaled annual mean plage area and is over plotted to show the progression of the SC. The **black dashed** rectangle marks the latitudinal band including 50% of plage areas for 1915, the latitudinal distribution of which is shown in panel (a). The numbers at the bottom of the panel denote the conventional SC numbers.

are assigned to that bin. An example of a study making use of the latitudinal distribution of Ca II K plage is given by Hofer et al. (2024).

In Figure 5, we observe that the median value of the latitude of plage regions and the upper end (75 percentile) of the latitude band covered by them are almost always higher than those for sunspots, whereas the lower limit (25 percentile) remains close to that of sunspots. Additionally, we found that the mean latitude of plage areas calculated over the entire period is  $20.5\% \pm 2.0$  higher than that of sunspots. This behaviour is consistently present throughout the observation period, independently of the phase of the SC. Over a comparatively short period of time during late 1990s, plage areas extend to particularly high latitudes. This is unfortunately due to residual artefacts due to the lower quality of the data over this period (Chatzistergos et al., 2019a).

### 3.3. Hemispheric Asymmetry

The Sun shows asymmetry in the magnetic activity manifested in its northern and southern hemispheres. This asymmetry of the Sun (N-S asymmetry, hereafter) has been studied using various indices of solar activity, e.g., sunspot number (Veronig et al., 2021), sunspot area (Vizoso and Ballester, 1990), Ca II K brightness (Bertello, Pevtsov, and Ulrich, 2020), and magnetic flux (Hathaway and Upton, 2016). Here, we utilize the butterfly diagram derived from our processing of the KoSO Ca II K observations to study the N-S asymmetry in chromospheric plage areas. In Figure 6a, we plot the yearly averaged area calculated separately for the northern ( $A_{\text{North}}$ ) and southern hemispheres ( $A_{\text{South}}$ ). In Figure 6b, we also show, as a function of time, the N-S asymmetry



calculated as:

$$\text{N-S Asymmetry} = \frac{A_{\text{North}} - A_{\text{South}}}{A_{\text{North}} + A_{\text{South}}}. \quad (3)$$

To further investigate the relationship between plage and sunspots, we computed the correlation coefficients between the N-S asymmetry observed for plage and sunspots.

We notice that the southern hemisphere dominates during SCs 21–23, while the northern hemisphere dominates from the declining phase of SC 18 until the maximum of SC 20. Earlier SCs appear more balanced between the two hemispheres although for SCs 14 and 18, each hemisphere dominates roughly half of the cycle.

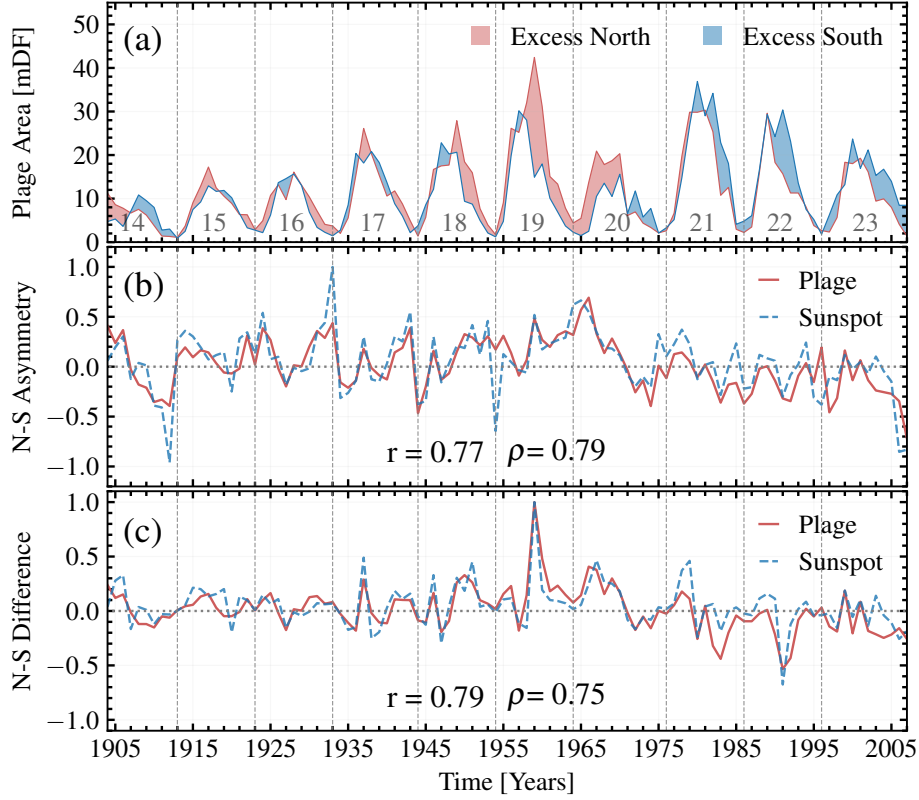
Interestingly, the rise of SC 19 seems to be nearly balanced in the two hemispheres until the peak is reached in the southern hemisphere. Afterwards the plage areas in the northern hemisphere continue rising for another two years, and then the northern hemisphere dominates until the declining phase of SC 20. We also notice that during the descending phase of SC 20 and the following minimum, when we previously (Section 3.1) reported residual artifacts at high latitudes, there is an excess of plage areas in the southern hemisphere. During SCs 16, 17, and 21–23 one hemisphere has double peaks, while the other hemisphere shows a single peak, which is in agreement with the earlier work by Chowdhury et al. (2022) for SCs 16 and 17, whereas they have not used the data after 1980 to discuss SCs 21–23.

To compare our results for the N-S asymmetry of plage areas to that in sunspot areas, we again used the KoSO white-light sunspot area series by Jha et al. (2022). We calculated the asymmetry in sunspot area in the same way as we did for the plage areas, and both are shown in Figure 6b. The two series follow each other quite closely. The main differences we notice is the slightly more balanced sunspot areas over the two hemispheres during SCs 21–23 compared to plage areas. Multiple sharp excursions in the sunspot area asymmetry are seen near activity minima. These are due to the fact that during minima sunspot areas drop to very low values close to 0, while plage areas are always higher than sunspot ones. Nonetheless, the small plage coverages around activity minima likely also lead to (less dramatically) excessive asymmetries in the plage areas at such times.

To avoid such excursions during times of low coverage, in Figure 6c, we display the difference of the yearly averaged plage and sunspot area coverage of the north and south hemispheres (N-S Difference =  $A_{\text{North}} - A_{\text{South}}$ ). To keep the values within the same vertical scale as in Figure 6b, we plot the N-S Difference after normalizing it by the maximum of the respective differences (27.5 mDF and 1031  $\mu\text{Hem}$ , respectively for plage and sunspot). It is evident from Figure 6c that the large values of N-S asymmetry during the SC minima are an artifact of poor statistics. It is interesting to note that the pronounced and comparatively long-lasting dominance of the northern hemisphere during the second half of SC 19 and the first half of SC 20 is seen consistently in both Ca II K plage and sunspot areas. The Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation coefficients for N-S Asymmetry with its sunspot counterpart are  $r = 0.77$  and  $\rho = 0.79$ .



The respective coefficients for N-S Difference and its sunspot counterpart are  $r = 0.79$  and  $\rho = 0.75$ ), respectively.



**Figure 6.** (a) Hemispheric yearly averaged plage area as a function of time. The red (blue) shaded regions represent the period where the northern (southern) hemisphere dominates over the other hemisphere. (b) The N-S asymmetry calculated using Equation 3 for plage and sunspot areas and the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation coefficients are listed at the bottom of the panel whereas (c) shows the normalized difference (with their respective maximum values) of the yearly averaged plage and sunspot area in two hemispheres.

#### 4. Summary and conclusions

KoSO has provided a wealth of solar observations since the beginning of the 20th century in various wavelength intervals, most prominently in the Ca II K line. That being said, these data have not been utilized to their full potential due to inconsistencies in the orientation of the images. This limitation particularly affects studies that require precise locations of features on the Sun (e.g. Mishra et al., 2024), e.g. latitudinal distribution of plage (e.g. Hofer et al., 2024), N-S asymmetry of plage areas, etc. Therefore, in this study we developed and applied a new automated process to orient the KoSO Ca II K observations. Our approach is more robust than previous methods applied to KoSO data, which

relied on manually placed pole markings on the drawings and subsequently manual clicking on the markings to determine the orientation of the observations. Furthermore, we identified the difference between our computed rotation angle to previous estimates to be  $\pm 2^\circ$  assuming there is no inconsistencies in Time of Observation ( $T_{\text{OBS}}$ ). However,  $\approx 0.5\%$  ( $\approx 200$ )

of the images were found to have pole markings at wrong positions, resulting in an error in the rotation angle that is greater than  $10^\circ$ . Additionally, we also corrected the time of observation,  $T_{\text{OBS}}$ , for  $\approx 10\%$  of the data which suffer from incorrect  $T_{\text{OBS}}$ , again leading to an error in the rotation angle. The newly developed method for image orientation, and knowledge of issues with  $T_{\text{OBS}}$ , worked out here for Ca II K observation will be extended for application to KoSO H $\alpha$  and Ca II K prominence observation as all of them are taken with the same instrumental setup. Furthermore, it will also be useful for observations from other archives taken with a similar instrumental setup.

In our analysis we further benefited from more accurate processing to compensate for artifacts affecting the images and to perform the photometric calibration than in previous studies requiring an accurate image orientation. Based on this, we produced the first complete butterfly diagram from KoSO Ca II K data, which covers the entire period of observations between 1904–2007.

We have also compared our butterfly diagram to those previously published in the literature. We showed that due to our more accurate processing we are able to produce a butterfly diagram including more data than previously possible. To remove some inconsistencies we corrected issues with the dates and times of the observations that existed in previously published butterfly diagrams of plage areas from KoSO Ca II K data. We then used the obtained butterfly diagram to compare the latitudinal extend of plage areas to that of sunspots. We found that the mean latitude of plage areas is higher than that of sunspots irrespective of the phase or the strength of the SC. We expect that this behavior has to do with Joy’s law (Hale et al., 1919) and the fact that plage are more strongly concentrated in the following polarities of active regions, which are on average further from the equator than the leading polarity (which typically harbors the larger spots). The longer lifetime of plage also means that they are more affected by diffusion and are also more likely to be dragged by the meridional flow towards the poles than sunspots are.

Finally we also investigated the N-S Asymmetry in plage areas. We found that the northern hemisphere dominated plage areas during SCs 19 and 20, while the southern hemisphere dominated between SCs 21 and 23. This consistent dominance of the southern hemisphere could be an artifact of degraded data quality, but it can also be linked to the long-term variation of the solar magnetic equator as suggested by Pulkkinen et al. (1999). However, further investigation is needed to substantiate such claims. We also compared the N-S Asymmetry of plage areas to that displayed by sunspot areas based on white-light observations obtained at KoSO under the same observing conditions. We found that the asymmetry in plage areas is well correlated ( $r = 0.77$ ,  $\rho = 0.79$ ) with the asymmetry observed in the sunspot areas.

With this study we extended the usability of KoSO Ca II K data, forming one of the most prominent such historical archives, for wider scientific applications.

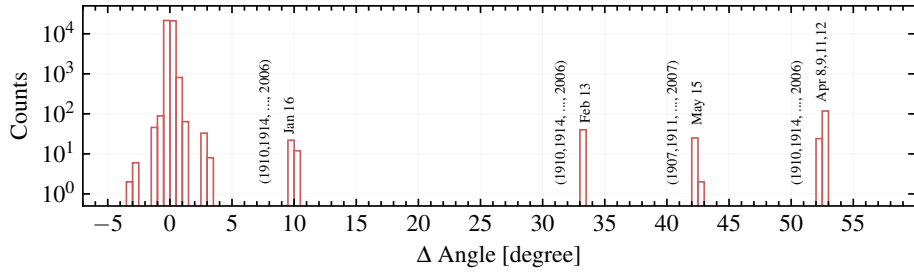
Therefore, corrected data and obtained results will contribute to achieving improved reconstructions of past solar surface magnetism (e.g. Chatzistergos et al., 2019; Chatzistergos, Krivova, and Ermolli, 2022; Mordvinov et al., 2020; Hofer et al., 2024, and references therein) and irradiance variations (e.g. Chatzistergos et al., 2021; Chatzistergos, Krivova, and Yeo, 2023) from Ca II K observations. Information derived from the century-long Ca II K plage butterfly diagram obtained here has the potential to provide valuable inputs to dynamo models and understanding of SC evolution. Besides, the KoSO Ca II K data accurately processed and oriented derived from this study will allow for further research of the properties of the solar rotation and large-scale flows over roughly a century. First results from such studies have been recently presented by Mishra et al. (2023) and more will follow. More importantly, we produced a dataset of KoSO Ca II K accurately oriented observations along with the data for corrected  $T_{\text{OBS}}$  and image orientation will be made available to the public for use in further investigations.

## Appendix

### A. Comparison of pole angle calculation methods

Here we present the comparison of our new automatic method of estimating  $\Theta$  from Equation 1 and 2 to the one used by Priyal et al. (2014a). These authors have calculated  $\Theta$  by manually clicking on the pole markings in the images (see double and single dots right outside the solar limb in Figure 1, which represent the position of the North and South poles, respectively). These pole markings are drawn during the observation or just before digitization using a tabulated value (lookup table) provided at KoSO. To compare  $\Theta$  calculated from these two methods, we firstly used an automatic code to extract  $\Theta$  from the KoSO table ( $\Theta'$  here-after) for a given  $T_{\text{OBS}}$  and then used that  $T_{\text{OBS}}$  to get  $\Theta$  based on Equation 2. We note however, that the rotation angle derived by Priyal et al. (2014a) includes the extra uncertainty of drawing the polar markings on the photographic plates as well as manually clicking on them on the digital images with the computer mouse. In Figure 7, we show the histogram of the differences ( $\Delta\Theta = \Theta - \Theta'$ ) in the computed angles, with bin size of  $0.5^\circ$ . Figure 7 implies that the distribution has the most prominent peak within  $\pm 2^\circ$ , which tells us that, in most cases ( $\approx 99.5\%$ ), our new method is in good agreement with the tabulated (used by Priyal et al., 2014a, and others) one, provided we use the same  $T_{\text{OBS}}$  in both these cases, irrespective of their correctness. This analysis helped us to understand the difference between two methods for a given  $T_{\text{OBS}}$ .

We note that there are cases when these methods differ significantly ( $\Delta\Theta > 10^\circ$ ) from each other. The dates and times of observation ( $T_{\text{OBS}}$ ) corresponding to these peaks are restricted to a few specific dates (marked in Figure 7), which are repeating every four years. We visually inspected the corresponding images to check which of the methods returned the wrong rotation angle. We used the rotation angles from our analysis ( $\Theta$ ) and that calculated from the KoSO table ( $\Theta'$ ), and compared the images to nearby observations. An example is shown in



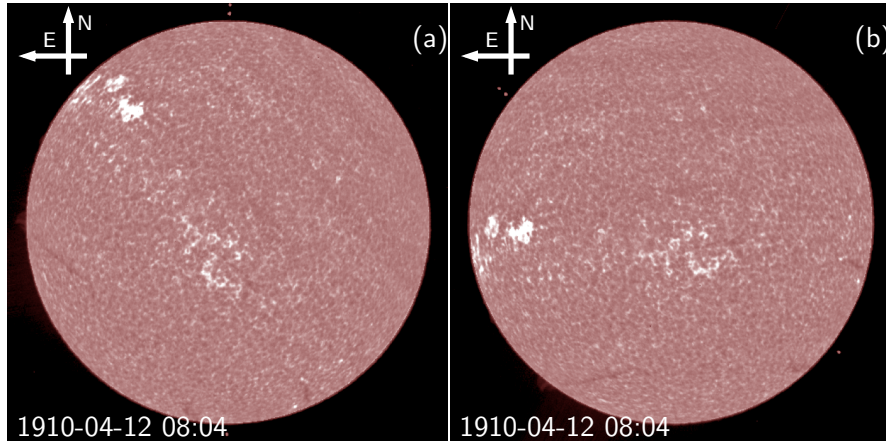
**Figure 7.** Distribution of differences in rotation angle ( $\Delta\Theta$ ) calculated from the [lookup table provided at KoSO](#) and our estimate based on [Equation 2](#). The peaks at  $\Delta\Theta > 10^\circ$  correspond to particular dates repeating every four years. These dates for the first two and last occurrences in the series are written on the right side of the peaks.

**Figure 8.** We noticed that in these images although after the rotation with the angle by  $\Theta'$  ([Figure 8a](#)) the North pole markings (two small dots) appear at the top, the distribution of plage appears tilted. In particular in [Figure 8a](#) plage regions appear at extremely high latitudes, especially considering the phase of the SC of that observation (being in the descending phase of SC 14). The rotation with our method ([Figure 8b](#)) resulted in the North pole markings being at roughly off by  $45^\circ$ , but plage regions appear at more realistic latitudes. To validate which plage region distribution is correct for these observations, we compared them with the observation nearest in time to the one under consideration. This allowed us to ascertain that the pole markings were placed at the wrong positions for these cases. The rather systematic time difference of these observations (repeating every four years) suggests that this was most likely an error in the algorithm when computing the rotation angle before making the table and drawing the pole markings. This shows the robustness of our method to such mistakes. It is also worth noting that the method developed in our study can be applied to accurately derive the orientation of the KoSO observations obtained at other spectral regions than Ca II K e.g. H $\alpha$  and Ca II K prominences.

## B. Manual identification of incorrect $T_{\text{OBS}}$ by comparing with nearest observations.

As mentioned in the main text, the images unfortunately suffer from mistakes in the recorded observational dates and times, which can hamper the performance of our approach to orient them. We discussed ways to mitigate some systematic issues due to such mistakes, but more isolated cases remained. In the following we describe the steps we undertook to try to identify such cases with wrong observational date/time and account for this error.

1. First, we selected an observation (say  $\text{Im}_0$ ) from the archive and using the  $T_{\text{OBS}}$  in the file name we rotated the image according to the pole angle calculated using [Equation 2](#).

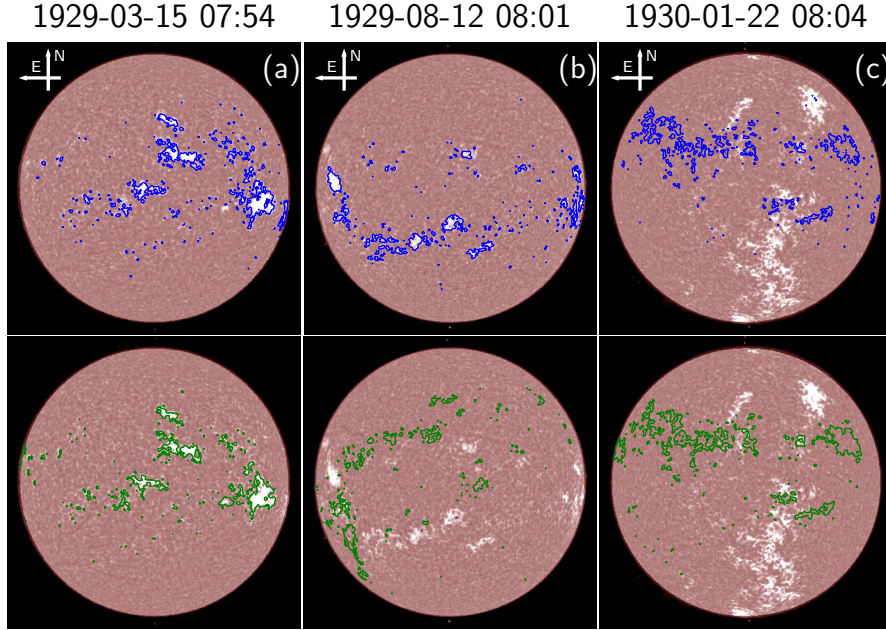


**Figure 8.** KoSO Ca II K observation taken on 12 April 1910. The images were rotated with the angle (a) tabulated by Priyal et al. (2014a) and (b) estimated by us using Equations 1 and 2. The arrows at the top left part of the image denote the expected orientation of the observations. Note that the double dot is the North pole marking drawn by the observer on the photographic plate. These markings were used by Priyal et al. (2014a) to orient the images. The images are shown after the compensation of the limb darkening and removal of large-scale artifacts, and they are displayed in the range of contrast values of  $[-0.4, 0.4]$ .

2. After that, we selected the immediately preceding,  $Im_{-1}$  and immediately following,  $Im_{+1}$ , observations and rotate them accordingly with the same method. While selecting  $Im_{-1}$  and  $Im_{+1}$ , we make sure that the difference in  $T_{OBS}$  with  $Im_0$  is between 0.5 to 3 days, to achieve as best comparison as possible.
3. In the next step, we differentially rotated<sup>3</sup>  $Im_{-1}$  and  $Im_{+1}$  forward and back, respectively, to the time ( $T_{OBS}$ ) of  $Im_0$  and over-plotted the contours of bright regions, using a fixed intensity threshold (see Figure 9a–9c).
4. In the last step, we manually checked the overlap of the plage contours from the various images to that of  $Im_0$ . If either of the contours shows a good overlap then we flag the  $T_{OBS}$  (of  $Im_0$ ) as correct (within the uncertainty of this exercise) otherwise we flag it as potentially incorrect.

Three examples of this procedure are shown in Fig. 9. In particular, the first two cases suggest that the observational date and time of  $Im_0$  is correct since at least one of the contours from  $Im_{-1}$  and  $Im_{+1}$  exhibits a good overlap with the plage regions of  $Im_0$ . However, in Figure 9c, we see that none of the contours shows a good overlap. In general in this case besides the possibility that  $Im_0$  has the wrong  $T_{OBS}$  it is also possible that the  $Im_0$  has correct  $T_{OBS}$  but both the  $Im_{-1}$  and  $Im_{+1}$  have incorrect  $T_{OBS}$ . Therefore to avoid such cases, we marked

<sup>3</sup>Using IDL SolarSoftWare (SSW) routine *drot\_map.pro*.



**Figure 9.** Examples of the procedure to identify mistakes in the observational date/time of the images. The observations were made on 15 March 1929 at 07:54 IST, 12 August 1929 at 08:01 IST, and 22 January 1930 at 08:04 IST. We show the calibrated images in the range of  $[-0.4, 0.4]$  in contrast values. In each case we overplot contours of plage regions from an earlier observation ( $Im_{-1}$ , shown in blue in the top row) and a later observation ( $Im_{+1}$ , shown in green in the bottom row). With these images we show three different cases of overlap: (a) both contours overlapping, thus we marked  $T_{OBS}$  to be correct, (b) only contour of  $Im_{-1}$  is overlapping, again suggesting  $T_{OBS}$  is correct (but  $T_{OBS}$  of  $Im_{+1}$  is most likely wrong), and (c) none of the contours are overlapping hence  $T_{OBS}$  was flagged as incorrect.

all three images as potentially having the wrong  $T_{OBS}$  and we repeated the above mentioned process for all the observations marked in this step as having potentially incorrect  $T_{OBS}$ . When repeating this, we made sure that, this time, we only choose  $Im_{-1}$  and  $Im_{+1}$ , which have been previously flagged as having correct  $T_{OBS}$ .

**Acknowledgements** Kodaikanal Solar Observatory (KoSO) is a facility of the Indian Institute of Astrophysics, Bangalore, India. These data are now available for public use at <http://kso.iiap.res.in> through a service developed at IUCAA under the Data Driven Initiatives project funded by the National Knowledge Network. The authors also thank the observers at the KoSO for their meticulous work at maintaining this archive. BKJ expresses his gratitude to Max Planck Institute of Solar System and Research for the warm hospitality and financial support during his visit. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 101097844 — project WINSUN). T.C. thanks ISSI for supporting the International Team 474 “What Determines The Dynamo Effectivity Of Solar Active Regions?”. This research has made use of NASA’s Astrophysics Data System (ADS; <https://ui.adsabs.harvard.edu/>) Bibliographic Services.

**Author Contribution** BKJ devised the method to orient the observations and corrected the files for mistakes in observational dates/times, while also performing the bulk of the analysis



presented in this study. TC processed the observations to produce the butterfly diagram as well as performed the cross-correlation technique to fix residual issues with the orientation. The manuscript was mostly written by BKJ and TC, while all co-authors provided ideas, feedback about the work and proof read the manuscript.

**Data Availability** The Ca II K data analyzed in this work is available through KoSO website at <http://kso.iiap.res.in>. The calibrated version of Ca II K data can be provided on request. The updated KoSO sunspot area series used is available through <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/JIST0V> (Jha, 2024). The corrected  $T_{\text{OBS}}$  and correct orientation of the Ca II K observation, fractional plage area (used for producing Figure 4), the yearly latitudinal distribution (for Figure 5) along with other analyzed data are available at <https://github.com/bibhuraushan/KoSOCaK> and will be made available through Zenodo.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

- Bappu, M.K.V.: 1967, Solar Physics at Kodaikanal. *Solar Phys.* **1**, 151. DOI. ADS. 4
- Berrilli, F., Ermolli, I., Florio, A., Pietropaolo, E.: 1999, Average properties and temporal variations of the geometry of solar network cells. *Astron. Astrophys.* **344**, 965. ADS. 2
- Bertello, L., Pevtsov, A.A., Ulrich, R.K.: 2020, 70 Years of Chromospheric Solar Activity and Dynamics. *Astrophys. J.* **897**, 181. DOI. ADS. 3, 14
- Chatterjee, S., Banerjee, D., Ravindra, B.: 2016, A Butterfly Diagram and Carrington Maps for Century-long CA II K Spectroheliograms from The Kodaikanal Observatory. *Astrophys. J.* **827**, 87. DOI. ADS. 2, 3, 5, 6, 12, 13
- Chatterjee, S., Mandal, S., Banerjee, D.: 2017, Variation of Supergranule Parameters with Solar Cycles: Results from Century-long Kodaikanal Digitized Ca II K Data. *Astrophys. J.* **841**, 70. DOI. ADS. 2
- Chatzistergos, T.: 2023, Is there a link between the length of the solar cycle and Earth's temperature? *Rendiconti Lincei. Scienze Fisiche e Naturali.* DOI. URL. 2
- Chatzistergos, T., Krivova, N.A., Ermolli, I.: 2022, Full-disc Ca II K observations - A window to past solar magnetism. *Frontiers in Astronomy and Space Sciences* **9**. DOI. URL. 2, 3, 7, 18
- Chatzistergos, T., Krivova, N.A., Ermolli, I.: 2024, Understanding the secular variability of solar irradiance: the potential of Ca II K observations. *Journal of Space Weather and Space Climate* **14**, 9. DOI. URL. 2
- Chatzistergos, T., Krivova, N.A., Yeo, K.L.: 2023, Long-term changes in solar activity and irradiance. *Journal of Atmospheric and Solar-Terrestrial Physics* **252**, 106150. DOI. URL. 18
- Chatzistergos, T., Ermolli, I., Solanki, S.K., Krivova, N.A.: 2018, Analysis of full disc Ca II K spectroheliograms. I. Photometric calibration and centre-to-limb variation compensation. *Astron. Astrophys.* **609**, A92. DOI. ADS. 4, 6, 7, 11
- Chatzistergos, T., Ermolli, I., Krivova, N.A., Solanki, S.K.: 2019a, Analysis of full disc Ca II K spectroheliograms. II. Towards an accurate assessment of long-term variations in plage areas. *Astron. Astrophys.* **625**, A69. DOI. ADS. 4, 6, 7, 11, 13, 14
- Chatzistergos, T., Ermolli, I., Solanki, S.K., Krivova, N.A., Banerjee, D., Jha, B.K., Chatterjee, S.: 2019b, Delving into the Historical Ca ii K Archive from the Kodaikanal Observatory: The Potential of the Most Recent Digitized Series. *Solar Phys.* **294**, 145. DOI. ADS. 3, 5, 6, 8
- Chatzistergos, T., Ermolli, I., Falco, M., Giorgi, F., Guglielmino, S.L., Krivova, N.A., Romano, P., Solanki, S.K.: 2019, Historical solar Ca II K observations at the Rome and Catania observatories. In: *Il Nuovo Cimento* **42C**, 5. DOI. URL. 2
- Chatzistergos, T., Ermolli, I., Solanki, S.K., Krivova, N.A., Giorgi, F., Yeo, K.L.: 2019, Recovering the unsigned photospheric magnetic field from Ca II K observations. *Astron. Astrophys.* **626**, A114. DOI. ADS. 2, 18



- Chatzistergos, T., Ermolli, I., Krivova, N.A., Solanki, S.K., Banerjee, D., Barata, T., Belik, M., Gafeira, R., Garcia, A., Hanaoka, Y., Hegde, M., Klimesš, J., Korokhin, V.V., Lourenço, A., Malherbe, J.-M., Marchenko, G.P., Peixinho, N., Sakurai, T., Tlatov, A.G.: 2020a, Analysis of full-disc Ca II K spectroheliograms - III. Plage area composite series covering 1892–2019. *Astronomy & Astrophysics* **639**, A88. Publisher: EDP Sciences. DOI. URL. [2](#), [6](#), [7](#), [11](#)
- Chatzistergos, T., Ermolli, I., Krivova, N.A., Solanki, S.K.: 2020b, Historical solar Ca II K observations at the Kyoto and Sacramento Peak observatories. *Journal of Physics: Conference Series* **1548**, 012007. Publisher: IOP Publishing. DOI. URL. [3](#), [6](#)
- Chatzistergos, T., Ermolli, I., Giorgi, F., Krivova, N.A., Puiu, C.C.: 2020c, Modelling solar irradiance from ground-based photometric observations. *Journal of Space Weather and Space Climate* **10**, 45. DOI. ADS. [2](#), [7](#)
- Chatzistergos, T., Krivova, N.A., Ermolli, I., Yeo, K.L., Mandal, S., Solanki, S.K., Kopp, G., Malherbe, J.-M.: 2021, Reconstructing solar irradiance from historical Ca II K observations. I. Method and its validation. *Astron. Astrophys.* **656**, A104. DOI. ADS. [2](#), [18](#)
- Chatzistergos, T., Ermolli, I., Krivova, N.A., Barata, T., Carvalho, S., Malherbe, J.-M.: 2022, Scrutinising the relationship between plage areas and sunspot areas and numbers. *Astronomy & Astrophysics* **667**, A167. DOI. URL. [2](#)
- Chatzistergos, T., Ermolli, I., Banerjee, D., Barata, T., Chouinavas, I., Falco, M.: 2023, Analysis of full disc Ha observations over 1909-2022. *Astron. Astrophys.* [3](#), [7](#)
- Chowdhury, P., Belur, R., Bertello, L., Pevtsov, A.A.: 2022, Analysis of Solar Hemispheric Chromosphere Properties using the Kodaikanal Observatory Ca-K Index. *Astrophys. J.* **925**, 81. DOI. ADS. [15](#)
- Cornu, M.A.: 1900, On the Law of Diurnal Rotation of the Optical Field of the Siderostat and Heliostat. *Astrophys. J.* **11**, 148. DOI. ADS. [7](#)
- Deslandres, H.: 1908, Reports of Committees and Papers communicated to the Solar Union. No. 11. Enregistrement de la surface et de l'atmosphère solaire à l'Observatoire de Meudon. *Transactions of the International Union for Cooperation in Solar Research* **2**, 232. ADS. [3](#)
- Eberhard, G., Schwarzschild, K.: 1913, On the reversal of the calcium lines H and K in stellar spectra. *The Astrophysical Journal* **38**, 292. DOI. URL. [2](#)
- Ermolli, I., Ferrucci, M.: 2021, The Legacy of Angelo Secchi at the Forefront of Solar Physics Research. In: Chinnici, I., Consolmagno, G. (eds.) *Angelo Secchi and Nineteenth Century Science: The Multidisciplinary Contributions of a Pioneer and Innovator Historical and Cultural Astronomy*, Springer International Publishing, Cham, 123. ISBN 978-3-030-58384-2. URL. [3](#)
- Ermolli, I., Giorgi, F., Chatzistergos, T.: 2022, Rome Precision Solar Photometric Telescope: precision solar full-disk photometry during solar cycles 23–25. *Frontiers in Astronomy and Space Sciences* **9**. DOI. URL. [2](#)
- Ermolli, I., Fofi, M., Bernacchia, C., Berrilli, F., Caccin, B., Egidi, A., Florio, A.: 1998, The Prototype RISE-PSPT Instrument Operating in Rome. *Solar Phys.* **177**, 1. DOI. ADS. [2](#)
- Ermolli, I., Marchei, E., Centrone, M., Criscuoli, S., Giorgi, F., Perna, C.: 2009, The digitized archive of the Arcetri spectroheliograms. Preliminary results from the analysis of Ca II K images. *Astron. Astrophys.* **499**, 627. DOI. ADS. [3](#), [6](#)
- Ermolli, I., Chatzistergos, T., Krivova, N.A., Solanki, S.K.: 2018, The potential of Ca II K observations for solar activity and variability studies. In: Banerjee, D., Jiang, J., Kusano, K., Solanki, S. (eds.) *Long-term Datasets for the Understanding of Solar and Stellar Magnetic Cycles, IAU Symposium* **340**, Cambridge University Press, Cambridge, UK, 115. DOI. [2](#)
- Ermolli, I., Chatzistergos, T., Giorgi, F., Carrasco, V.M.S., Aparicio, A.J.P.: 2023, Solar Observations by Angelo Secchi. I. Digitization of Original Documents and Analysis of Group number over 1853-1878. [3](#)
- Freeland, S.L., Handy, B.N.: 1998, Data Analysis with the SolarSoft System. *Solar Physics* **182**, 497. Number: 2. DOI. URL. [7](#)
- Hale, G.E.: 1890, Note on Solar Prominence Photography. *Astronomische Nachrichten* **126**, 81. DOI. ADS. [3](#)
- Hale, G.E.: 1891, The Kenwood Physical Observatory. *Pub. Astron. Soc. Pac.* **3**, 30. DOI. ADS. [3](#)
- Hale, G.E.: 1893, Solar photography at the Kenwood Astro-physical Observatory. *Memorie della Societa Degli Spettroscopisti Italiani* **21**, 68. ADS. [3](#)
- Hale, G.E., Ellerman, F., Nicholson, S.B., Joy, A.H.: 1919, The Magnetic Polarity of Sun-Spots. *Astrophys. J.* **49**, 153. DOI. ADS. [17](#)

- Hanaoka, Y.: 2013, Long-term synoptic observations of the Sun at the National Astronomical Observatory of Japan. In: *Journal of Physics Conference Series, Journal of Physics Conference Series* **440**, 012041. DOI. ADS. 3
- Harvey, K.L.: 1992, The Cyclic Behavior of Solar Activity. In: *Proceedings of the National Solar Observatory/Sacramento Peak 12th Summer Workshop, Astronomical Society of the Pacific Conference Series* **27**, San Francisco, 335. URL. 3, 11, 12
- Harvey, K.L., White, O.R.: 1999, Magnetic and Radiative Variability of Solar Surface Structures. I. Image Decomposition and Magnetic-Intensity Mapping. *Astrophys. J.* **515**, 812. DOI. ADS. 2
- Hasan, S.S., Mallik, D.C.V., Bagare, S.P., Rajaguru, S.P.: 2010, Solar Physics at the Kodaikanal Observatory: A Historical Perspective. In: *Magnetic Coupling between the Interior and Atmosphere of the Sun, Astrophysics and Space Science Proceedings* **19**, 12. DOI. ADS. 3, 4
- 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. DOI. ADS. 14
- Hofer, B., Krivova, N.A., Cameron, R., Solanki, S.K., Jiang, J.: 2024, The influence of small bipolar magnetic regions on basic solar quantities. *Astron. Astrophys.* **683**, A48. DOI. ADS. 14, 16, 18
- Jha, B.K.: 2022, Long-term Study of the Sun and Its Implications to Solar Dynamo Models. PhD thesis, Department of Physics, Pondicherry University, Puducherry, India. DOI. URL. 8, 9
- Jha, B.K.: 2024, KoSO Sunspot Area Series V2. DOI. URL. 22
- Jha, B.K., Hegde, M., Priyadarshi, A., Mandal, S., Ravindra, B., Banerjee, D.: 2022, Extending the sunspot area series from Kodaikanal Solar Observatory. *Frontiers in Astronomy and Space Sciences* **9**, 1019751. DOI. ADS. 13, 14, 15
- Kahil, F., Riethmüller, T.L., Solanki, S.K.: 2017, Brightness of Solar Magnetic Elements As a Function of Magnetic Flux at High Spatial Resolution. *Astrophys. J. Suppl.* **229**, 12. DOI. ADS. 2
- Kharayat, H., Singh, J., Priyal, M., Ravindra, B.: 2024, Equator to Pole Solar Chromospheric Differential Rotation Using Ca-K Features Derived from Kodaikanal Data. *The Astrophysical Journal* **968**, 53. DOI. URL. 3
- Lefebvre, S., Ulrich, R.K., Webster, L.S., Varadi, F., Javaraiah, J., Bertello, L., Werden, L., Boyden, J.E., Gilman, P.: 2005, The solar photograph archive of the Mount Wilson Observatory. A resource for a century of digital data. *Memorie della Societa Astronomica Italiana* **76**, 862. URL. 3
- Linsky, J.L., Avrett, E.H.: 1970, The Solar H and K Lines. *Pub. Astron. Soc. Pac.* **82**, 169. DOI. ADS. 2
- Loukitcheva, M., Solanki, S.K., White, S.M.: 2009, The relationship between chromospheric emissions and magnetic field strength. *Astron. Astrophys.* **497**, 273. DOI. ADS. 2
- Makarov, V.I., Tlatov, A.G., Singh, J., Gupta, S.S.: 2004, 22-years magnetic cycle in polar activity of the Sun. In: Stepanov, A.V., Benevolenskaya, E., Kosovichev, A.G. (eds.) *Multi-Wavelength Investigations of Solar Activity, Proceedings of the International Astronomical Union* **223**, Cambridge University Press, Cambridge, UK, 125. ISBN 1743-9221. DOI. URL. 5, 6
- Malherbe, J.-M.: 2023, 130 years of spectroheliograms at Paris-Meudon observatories (1892-2022). *Journal for the History of Astronomy.* DOI. URL. 3
- Malherbe, J.-M., Bualé, I., Crussaire, D., Cornu, F., Corbard, T.: 2022, The potential of Meudon spectroheliograph for investigating long term solar activity and variability. *Advances in Space Research.* DOI. URL. 3
- Mishra, D.K., Routh, S., Jha, B.K., Chatterjee, S., Chatzistergos, T., Banerjee, D., basu, J.: 2023, Differential Rotation of the Solar Chromosphere: A Century-long Perspective from Kodaikanal Solar Observatory Ca II K data. 3, 18
- Mishra, D.K., Jha, B.K., Chatzistergos, T., Ermolli, I., Banerjee, D., Upton, L.A.: 2024, Ca II K Polar Network Index of the Sun: A Proxy for Historical Polar Magnetic Field. *ApJ.* 3, 16
- Mordvinov, A.V., Karak, B.B., Banerjee, D., Chatterjee, S., Golubeva, E.M., Khlystova, A.I.: 2020, Long-term Evolution of the Sun's Magnetic Field during Cycles 15-19 Based on Their Proxies from Kodaikanal Solar Observatory. *Astrophys. J. Lett.* **902**, L15. DOI. ADS. 2, 18

- Murabito, M., Ermolli, I., Chatzistergos, T., Jafarzadeh, S., Giorgi, F., Voort, L.R.v.d.: 2023, Investigating the Effect of Solar Ambient and Data Characteristics on Ca II K Observations and Line Profile Measurements. *The Astrophysical Journal* **947**, 18. DOI. URL. 2
- Nesme-Ribes, E., Meunier, N., Collin, B.: 1996, Fractal analysis of magnetic patterns from Meudon spectroheliograms. *Astron. Astrophys.* **308**, 213. ADS. 11
- Priyal, M., Singh, J., Ravindra, B., Priya, T.G., Amareswari, K.: 2014a, Long Term Variations in Chromospheric Features from Ca-K Images at Kodaikanal. *Solar Phys.* **289**, 137. DOI. ADS. 3, 5, 6, 7, 18, 20
- Priyal, M., Banerjee, D., Karak, B.B., Muñoz-Jaramillo, A., Ravindra, B., Choudhuri, A.R., Singh, J.: 2014b, Polar Network Index as a Magnetic Proxy for the Solar Cycle Studies. *Astrophys. J. Lett.* **793**, L4. DOI. ADS. 3
- Priyal, M., Singh, J., Belur, R., Rathina, S.K.: 2017, Long-term Variations in the Intensity of Plages and Networks as Observed in Kodaikanal Ca-K Digitized Data. *Solar Physics* **292**, 85. Number: 7. DOI. URL. 3, 4, 12, 13
- Pulkkinen, P.J., Brooke, J., Pelt, J., Tuominen, I.: 1999, Long-term variation of sunspot latitudes. *Astron. Astrophys.* **341**, L43. ADS. 17
- Schrijver, C.J., Cote, J., Zwaan, C., Saar, S.H.: 1989, Relations between the Photospheric Magnetic Field and the Emission from the Outer Atmospheres of Cool Stars. I. The Solar CA II K Line Core Emission. *Astrophys. J.* **337**, 964. DOI. ADS. 2
- Secchi, A.: 1871, *Sulla distribuzione delle protuberanze intorno AL disco solare*. URL. 3
- Singh, J., Prabhu, T.P.: 1985, Variations in the solar rotation rate derived from Ca<sup>+</sup> K plage areas. *Solar Phys.* **97**, 203. DOI. ADS. 3
- Singh, J., Priyal, M., Ravindra, B., Bertello, L., Pevtsov, A.A.: 2022, On the Application of the Equal-contrast Technique to Ca-K Data from Kodaikanal and Other Observatories. *Astrophys. J.* **927**, 154. DOI. ADS. 2, 3
- Skumanich, A., Smythe, C., Frazier, E.N.: 1975, On the statistical description of inhomogeneities in the quiet solar atmosphere. I. Linear regression analysis and absolute calibration of multichannel observations of the Ca<sup>+</sup> emission network. *Astrophys. J.* **200**, 747. DOI. ADS. 2
- Smart, W.M.: 1977, *Textbook on Spherical Astronomy*, 6th edn. Cambridge University Press. 7
- Sowmya, K., Shapiro, A.I., Witzke, V., Nèmec, N.-E., Chatzistergos, T., Yeo, K.L., Krivova, N.A., Solanki, S.K.: 2021, Modeling Stellar Ca II H and K Emission Variations. I. Effect of Inclination on the S-index. *The Astrophysical Journal* **914**, 21. DOI. URL. 2
- Tacchini, P.: 1875, Macchie e facole al bordo. Osservazioni dirette e spettroscopiche fatte all'Osservatorio di Palermo nel 1873. *Memorie della Societa Degli Spettroscopisti Italiani* **4**, A81. URL. 3
- Tlatov, A.G., Pevtsov, A.A., Singh, J.: 2009, A New Method of Calibration of Photographic Plates from Three Historic Data Sets. *Solar Phys.* **255**, 239. DOI. ADS. 6
- Veronig, A.M., Jain, S., Podladchikova, T., Pötzi, W., Clette, F.: 2021, Hemispheric sunspot numbers 1874-2020. *Astron. Astrophys.* **652**, A56. DOI. ADS. 14
- Vizoso, G., Ballester, J.L.: 1990, The north-south asymmetry of sunspots. *Astron. Astrophys.* **229**, 540. ADS. 14
- Wells, D.C., Greisen, E.W., Harten, R.H.: 1981, FITS - a Flexible Image Transport System. *Astronomy and Astrophysics Supplement Series* **44**, 363. URL. 5
- Wilson, O.C.: 1968, Flux Measurements at the Centers of Stellar H- and K-Lines. *The Astrophysical Journal* **153**, 221. DOI. URL. 2