



<b>Publication Year</b>	2019
<b>Acceptance in OA</b>	2021-02-16T14:50:58Z
<b>Title</b>	Delving into the Historical Ca ii K Archive from the Kodaikanal Observatory: The Potential of the Most Recent Digitized Series
<b>Authors</b>	CHATZISTERGOS, THEODOSIOS, ERMOLLI, Ilaria, Solanki, Sami K., Krivova, Natalie A., Banerjee, Dipankar, Jha, Bibhuti K., Chatterjee, Subhamoy
<b>Publisher's version (DOI)</b>	10.1007/s11207-019-1532-5
<b>Handle</b>	<a href="http://hdl.handle.net/20.500.12386/30408">http://hdl.handle.net/20.500.12386/30408</a>
<b>Journal</b>	SOLAR PHYSICS
<b>Volume</b>	294

## Delving into the Historical Ca II K Archive from the Kodaikanal Observatory: the Potential of the Most Recent Digitised Series

Theodosios Chatzistergos<sup>1,2</sup> ·  
Ilaria Ermolli<sup>1</sup> · Sami K. Solanki<sup>2,3</sup> ·  
Natalie A. Krivova<sup>2</sup> · Dipankar Banerjee<sup>4</sup> ·  
Bibhuti K. Jha<sup>4</sup> · Subhamoy Chatterjee<sup>4</sup>

© Springer ●●●

**Abstract** Full-disc Ca II K photographic observations of the Sun carry direct information about the evolution of solar-plage regions for more than a century and are therefore a unique dataset for solar-activity studies. For a long time Ca II K observations were barely explored, but recent digitisations of multiple archives have allowed their extensive analysis. However, various studies have reported diverse results partly due to the insufficient quality of the digitised data. Furthermore, inhomogeneities have been identified within the individual archives, which, at least partly, could be due to the digitisation. As a result, some of the archives, *e.g.* that from the Kodaikanal observatory, were re-digitised. The results obtained by different authors who analysed the data from the new digitisation of the Kodaikanal archive differ from each other as well as from those derived from the old digitisation. Since the data were processed and analysed using different techniques, it is not clear, however, whether the differences are due to the digitisation or the processing of the data. To understand the reasons for such discrepancies, we analyse here the data from the two most recent digitisations of this archive. We use the same techniques to consistently process the images from both archives and to derive the plage areas from them. Some issues have been

---

✉ T. Chatzistergos  
chatzistergos@mps.mpg.de, theodosios.chatzistergos@inaf.it

<sup>1</sup> INAF Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monte Porzio Catone, Italy

<sup>2</sup> Max Planck Institute for Solar System Research, Justus-von-Liebig-weg 3, 37077 Göttingen, Germany

<sup>3</sup> School of Space Research, Kyung Hee University, Yongin, Gyeonggi 446-701, Republic of Korea

<sup>4</sup> Indian Institute of Astrophysics, Koramangala, Bangalore 560034, India

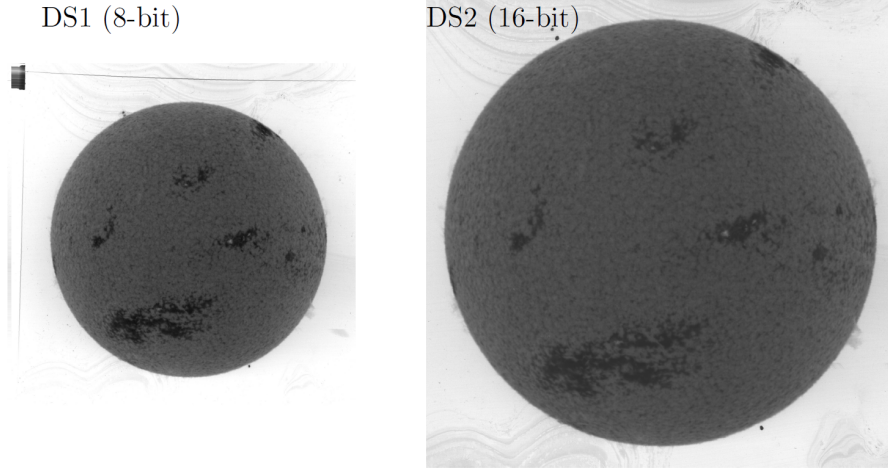
identified in both digitisations, implying that they are intrinsic characteristics of the data. Moreover, errors in timing of the observations plague both digitisations. Overall, the most recent 16-bit digitisation offers an improvement over the earlier 8-bit one. It also includes considerably more data and should be preferred.

**Keywords:** Solar Cycle, Observations; Instrumental Effects

## 1. Introduction

Full-disc photographs of the Sun in the resonance K line of the singly-ionized calcium, Ca II, at 3933.67 Å were first obtained in the early 1890's by Henri Alexandre Deslandres and George Ellery Hale (Hale, 1893) with the spectroheliographs developed at the Paris and Kenwood observatories, respectively. Since then, regular observations with similar instruments have been performed at various sites around the globe, *e.g.* at the Kodaikanal (since 1904), Mt. Wilson (since 1915), Mitaka (since 1917), Coimbra (since 1926), and Arcetri (since 1931) observatories. Ca II K observations with interference filters started later, *e.g.* at the Rome (since 1964), Kandilli (since 1968), and Big Bear (since 1981) observatories. These observations are one of the main sources of information on the long-term changes in the lower solar chromosphere, that is the first thousand kilometres above the temperature minimum. Furthermore, the Ca II K line provides information about solar magnetism. This is due to the large increase in the intensity of the line core when sampling bright magnetic regions (plage; see *e.g.* Skumanich *et al.*, 1984). The potential of the full-disc Ca II K observations to serve as a proxy of the magnetic field (Schrijver *et al.*, 1989; Loukitcheva, Solanki, and White, 2009; Chatzistergos *et al.*, 2019c, and references therein) makes them very valuable for studies of the evolution of solar activity, as well as for analyses of the magnetic activity of stars other than the Sun. Also, irradiance reconstructions and Earth's climate-variability studies can significantly benefit from analyses of Ca II K observations.

Most of the Ca II K observations performed so far are stored in photographic archives, a number of which have recently been digitised (see Chatzistergos, 2017; Chatzistergos *et al.*, 2019a,b). For example, the photographic archives from the Arcetri (1931–1974), Kodaikanal (1904–2007), Kyoto (1926–1969), McMath-Hulbert (1948–1979), Meudon (1893–2002), Mitaka (1917–1974), Mt. Wilson (1915–1985), Rome (1964–1979), and Sacramento Peak (1960–2002) observatories have been made available in digital form. The availability of the Ca II K series as digital data has initiated their extensive exploitation for studies of the long-term variation of the chromospheric magnetic field and for a variety of retrospective analyses. For example, Foukal (1996), Ermolli *et al.* (2009a), Tlatov, Pevtsov, and Singh (2009), Chatterjee, Banerjee, and Ravindra (2016), and Priyal *et al.* (2017) presented plage-area time series derived from the analysis of different archives with distinct image-processing methods. Harvey, Ermolli *et al.* (2009b), and Chatterjee, Banerjee, and Ravindra (2016) produced butterfly diagrams of plage regions from the Mt. Wilson, Arcetri, and Kodaikanal observations, respectively. Moreover, Sheeley, Cooper, and Anderson (2011), Chatterjee,

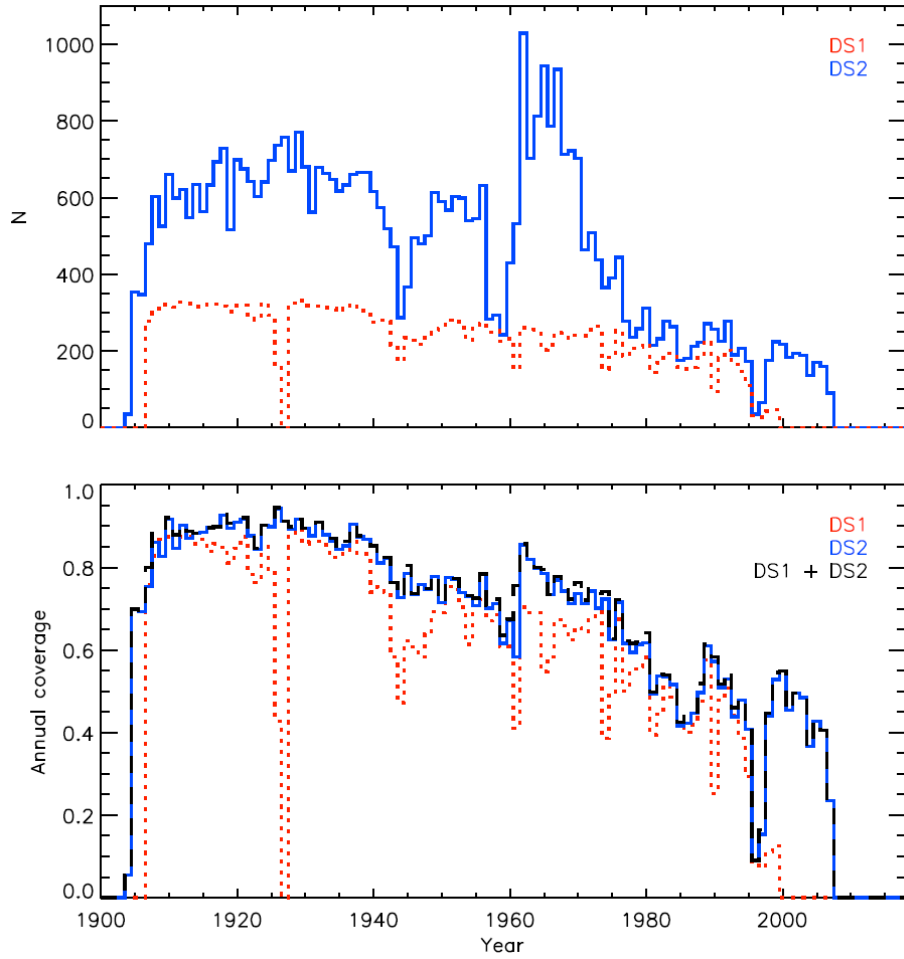


**Figure 1.** Examples of the images from the two digitisations of the photographic Ca II K observations of the Kodaikanal observatory taken on 02 January 1936. Left: 8-bit digitisation from data set 1, or DS1 in short; Right: 16-bit digitisation from DS2. The images are shown to their full range of values and were not compensated for the ephemeris. The DS1 image is shown to scale with the DS2 image (i.e. equal number of pixels per cm), while the DS2 has been cropped to a width of  $2R + R/6$ , where  $R$  is the solar radius in pixels.

Banerjee, and Ravindra (2016), and Pevtsov *et al.* (2016) produced Carrington maps from Ca II K observations of the Mt. Wilson and Kodaikanal archives. Recently, Chatterjee, Mandal, and Banerjee (2017) analysed the supergranulation scale variation in historical Kodaikanal Ca II K data following a previous similar analysis of modern data by Ermolli, Berrilli, and Florio (2003) with Rome Precision Solar Photometric Telescope (Rome/PSPT, hereafter) observations.

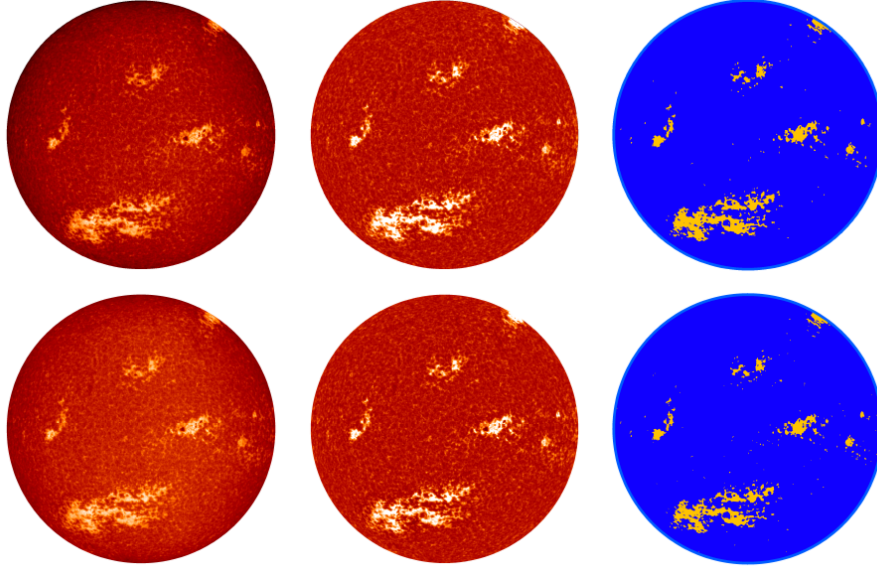
Most studies performed on historical Ca II K data confirm some well-known characteristics of past cycles, *e.g.* they all report that Solar Cycle 19 showed remarkably high plage coverage and a broad latitudinal distribution of active regions. They also agree on the overall increase of solar activity over the first half of the 20th century and the decrease over the last decades. However, they also show significant inconsistencies, such as large differences in absolute values and in the detected short- and long-term trends (Ermolli *et al.*, 2018; Chatzistergos *et al.*, 2019a). Such differences are seen both, between the results derived from different archives, and between the same archives but processed and analysed with different methodologies.

Some of these discrepancies can be ascribed to the varying quality of the analysed data partly arising from the digitisation. Most of the series were digitised independently at different periods and using various set-ups. Furthermore, some series underwent multiple digitisations, *e.g.* the Kodaikanal, Meudon, Mitaka, and Mt. Wilson archives. The two main reasons for the re-digitisation of the various series were: i) the availability of higher quality digitising devices than those employed in the older digitisations and ii) various problems identified in some of the earlier digitised series.



**Figure 2.** Number of images per year (*top panel*) and annual coverage (*bottom panel*) of the Kodaikanal data in the two digitised series. DS1 is shown in dotted red, DS2 in solid blue, while the annual coverage by the two archives together is shown in dashed black.

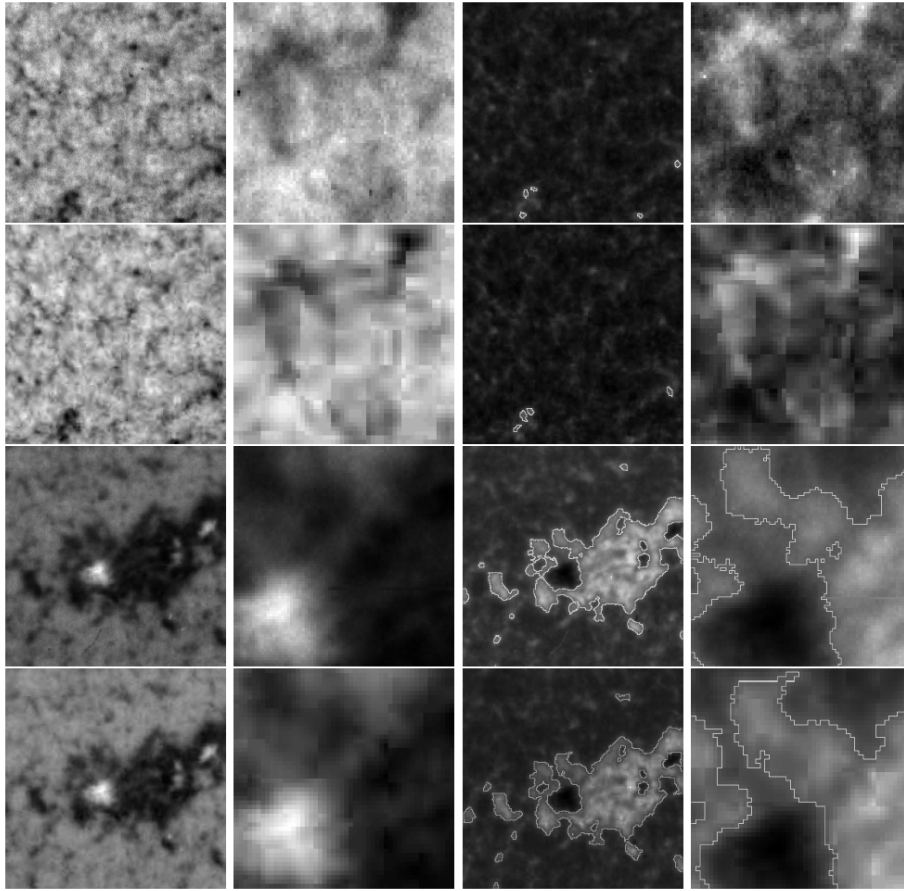
Among the photographic archives of Ca II K observations, the one from the Kodaikanal observatory has probably the largest collection of images, covering more than a century with nearly daily observations between 1904 and 2007. It is thus a particularly important archive. The Kodaikanal photographic archive underwent three digitisations over the last three decades (Kariyappa and Sivaraman, 1994; Makarov *et al.*, 2004; Priyal *et al.*, 2014), although only the last two digitisations included the entire solar disc. In Chatzistergos *et al.* (2019a) we consistently analysed with the same method the data from the former digitisation of the Kodaikanal archive along with the archives from Arcetri, McMath-Hulbert, Meudon, Mitaka, Mt. Wilson, Schauinsland, and Wendelstein to derive plage areas. We used the Kodaikanal and Mt. Wilson series as reference to construct two composites of plage areas over the entire 20th century. The two composites



**Figure 3.** Selected processing steps applied to images from the two digitisations of Kodaikanal observations taken on 02 January 1936 and shown in Figure 1. The DS2 (16-bit) image is shown in the *upper row*, while the DS1 (8-bit) one in the *lower row*. The columns show the original density image, photometrically calibrated and limb darkening corrected image, and segmentation mask, respectively. The raw images are shown over the entire range of values found within the solar disc, while the limb-darkening-compensated images are saturated at contrast values of  $[-1,1]$ . The segmentation masks show the QS regions in blue and the plage regions in orange.

showed different absolute levels and discrepancies over different Solar Cycles. To understand whether some of these differences are coming from the digitisation and can potentially be resolved with the new digitisation of the Kodaikanal data and to best exploit the potential of one of the most important Ca II K series, we here analyse the quality and intrinsic differences in the images derived from the two most recent digitizations of the full-disc Ca II K observations from the Kodaikanal observatory. We process the data from both digitisations consistently to derive plage areas and to compare the results to each other, as well as to those presented in the literature.

This article is structured as follows: in Section 2 we give an overview of the data employed in this study and describe the methods used to process them. We compare the images from the two digitisations in Section 3 where we also discuss characteristics of the series. We present our results of plage areas from the last digitisation of the Kodaikanal data in Section 4 and compare them to other series reported in the literature. Finally we draw our conclusions in Section 5.

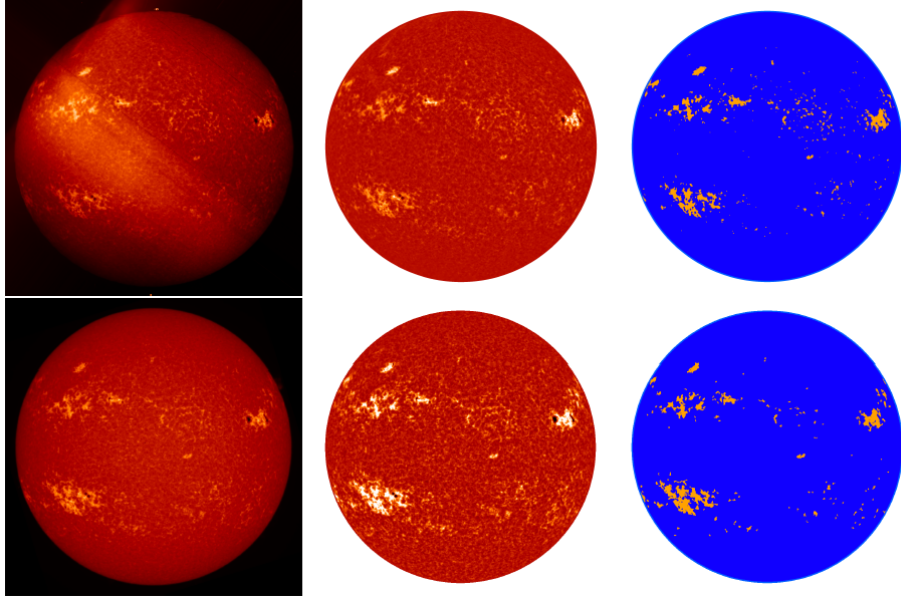


**Figure 4.** Magnified sections from the raw DS1 and DS2 negative images (*first two columns*) and calibrated and limb-darkening compensated images (*last two columns*) of the Kodaikanal observation taken on 02 January 1936 (see Figures 1 and 3) displaying a quiet Sun region (*top 2 rows*) and a plage region with a sunspot (*bottom 2 rows*). The image from DS2 is shown in *rows 1 and 3*, while the one from DS1 is shown in *rows 2 and 4*. The sections in *columns 1+3* and *2+4* have widths of  $260''$  and  $65''$ , respectively, which correspond to 200 and 50 pixels for the DS1 image, respectively. The images are shown over their full range of values. *Over-plotted contours* show the plage regions identified with the method applied in our study.

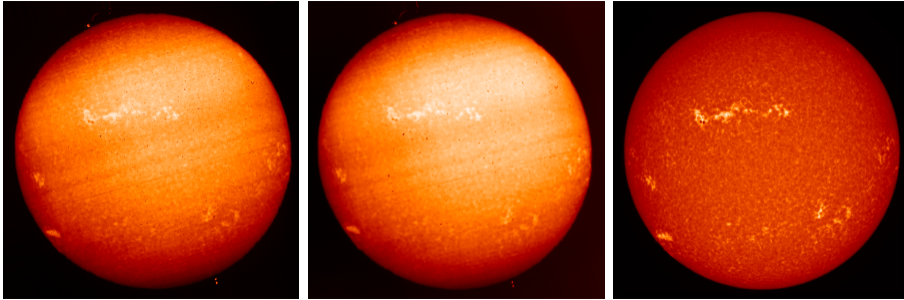
## 2. Data and Processing

### 2.1. Data

We analyse the data from the two most recent digitisations of the photographic full-disc Ca II K Kodaikanal observations. These data were taken with a spectroheliograph having a nominal bandwidth of  $0.5 \text{ \AA}$ . The first dataset (DS1, hereafter) was obtained by Makarov *et al.* (2004) by scanning 22,158 photographic observations taken from 1907 to 1999 with a linear array of 900 pixels. The data were stored in JPG format as  $1800 \times 1800 \text{ pixel}^2$  images, with 8-bit



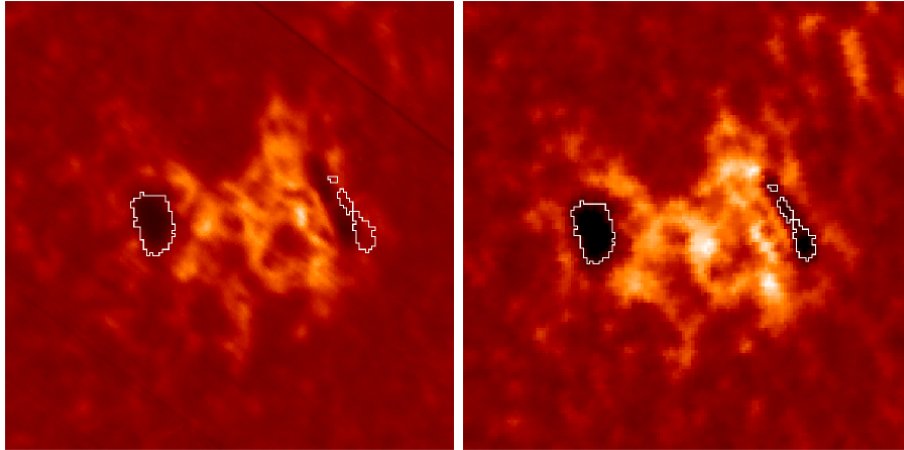
**Figure 5.** Comparison of observations from DS2 (*upper row*) and Rome/PSPT (*lower row*), both taken on 27 May 2000. Shown are the raw images (*left column*), processed and limb-darkening compensated images (*middle column*), and the segmentation masks (*right column*). The raw images are shown over the entire range of values found within the solar disc, while the limb-darkening compensated images are saturated at  $[-0.5, 0.5]$ . In the segmentation masks the QS regions are coloured blue and the plage regions orange.



**Figure 6.** Raw density images from DS1 (*left*), DS2 (*middle*), and Rome/PSPT (*right*) taken on 03 April 1999 illustrating saturated plage regions in the Kodaikanal data.

accuracy, and an average pixel-scale of  $1.3''\text{pixel}^{-1}$ . The second dataset (DS2, hereafter) was derived by Priyal *et al.* (2014) by scanning 48,928 photographic observations taken from 1904 to 2007 with a CCD camera with the resolution of  $4096 \times 4096 \text{ pixel}^2$ . The data were stored in FITS format with 16-bit accuracy and an average pixel scale of  $0.9''\text{pixel}^{-1}$ . Figure 1 shows the two digitisations of a single Kodaikanal photographic plate taken on 02 January, 1936. The raw negative images are displayed to their full range of values, while the image from





**Figure 7.** Enlargement of observations displayed in Figure 5 from DS2 (*left*) and Rome/PSPT (*right*) showing an active region. The images are saturated at  $[-0.5, 0.9]$ . The contours mark the location of sunspots as identified in the Rome/PSPT image. See Section 3.1 for more details.

the DS1 is shown to scale to the DS2 one showing equal number of pixels per cm.

Figure 2 shows the number of images per year in DS1 and DS2. Only a sub-sample of the available photographic plates were scanned for DS1, while almost all of the existing plates have been scanned for DS2. We find only 13,835 images of each set referring to the same solar observations, being scans of the same photographic plate. The fraction of days within each year with at least one observation from either Kodaikanal series is also shown in Figure 2. We find 21,405 days with at least one observation from both archives (out of 21,746 and 26,492 days in DS1 and DS2, respectively), as demonstrated by the mismatch between the coverage by DS2 and the total coverage by DS1 and DS2 together. Besides, we notice that there are seven years (1907, 1909, 1911, 1913, 1955, 1973, and 1993) over which DS1 has a better coverage than DS2. These discrepancies can, at least partly, be explained by errors in reporting both the date and timing of the observation in the digitised files. Such errors in the digitised series are not surprising considering the large number of plates and should most likely affect images from both digitisations. We find that the errors in the dates are limited to only a few images, while the errors in the time of the observations affect a considerable amount of data. In this regard, it is worth noting that images from DS2 include almost the entire plate, thus allowing us to compare the date on the plate to the one passed in the meta-data of the file. This is not the case for the images from DS1, which have been cropped to include the solar disc and only a small area outside of the disc. Furthermore, there is an inconsistency in both datasets in the time format. Time is given as Indian Standard Time up to the 1960's and as Coordinated Universal Time afterwards. For our study, all times from both datasets were corrected to be in Indian Standard Time. Note that this choice does not affect the results presented in the following, since the analysis was done on daily mean values.

The radius of the solar disc varies with time in both series, but it has an average value of 685 and 1095 pixel for DS1 and DS2, respectively. A heliostat was employed at the Kodaikanal Observatory, which allows determining the orientation of the solar disc based on the date and time of the observation. However, the orientation of the plates during the digitisation introduced another orientation that needs to be taken into account to orient the images. This angle is usually small and in general random. Markings have been introduced on the plates before the scanning (Priyal *et al.*, 2014) to identify the north and south poles of the solar disc. However, the information about which side corresponds to the east or west is lost. The information from the regions of the original plates lying outside the solar disc seems to have been maintained in DS2. In DS1, however, there are images where these regions have been saturated (*e.g.* Figure 1). Various artefacts, such as scratches and emulsion holes, appear in both digitisations. Many other artefacts, such as dust or hairs, are found at different locations in images of the two digitisations. One should also note that the DS2 series was generated more than a decade after the DS1 one, so natural degradation of the photographic plates is also partly responsible for this discrepancy. Moreover, since the 8-bit digitisation was performed with a linear array, this caused a few images to have rows that are offset in the  $x$ -direction and hence have distorted discs. This issue is resolved in the 16-bit digitisation, where a CCD camera was used.

Both sets of digitised images require instrumental calibration of the digitisation camera before any further processing. Such a calibration has not been applied to the DS1 (Ermolli *et al.*, 2009a) data, but it has partially been applied to the DS2 data and the standard data for instrumental calibrations were stored. In particular, the dark current was removed with a built-in program of the scanning device, while a lab-sphere illuminating a white surface was used to measure the flat-field of the CCD. The same surface was used to support the photographic plates during their scanning. Inspection of the available data shows that this surface does not always cover the same area as the plate. As a result, information outside the disc is lost when dividing the digitised observation by the corresponding flat-field image. Furthermore, the exposure time of the flat was not constant and was different from the one used for scanning the plates. Since the same CCD recorded the image with different gain in four quadrants, small errors in the photometric calibration of the data could affect the calibrated images showing variations over the quadrants and other residual inhomogeneities. In addition, flat-field images are expected to change over the course of the digitisation, and it is important to have such images created close to the scanning time of each image. However, flat-field images are missing from several folders in the archive or the included flat image does not always account completely for the difference in the quadrants.

We also use modern CCD-based observations taken with the Rome/PSPT as a comparison because there is partial overlap with both digitised Kodaikanal series. PSPT is located at Monte Porzio Catone and is operated by the INAF Osservatorio Astronomico di Roma (Ermolli *et al.*, 1998, 2007). Observations with Rome/PSPT started in May 1996 and continue to the present. The observations used here are taken with an interference filter centred at the Ca II K line with

bandwidth of  $2.5 \text{ \AA}$ . The images have dimensions of  $2048 \times 2048 \text{ pixel}^2$  and are stored in FITS files after the standard instrumental calibration (Ermolli *et al.*, 2007). Therefore, such data can be used to investigate the differences between photographic and CCD observations, as well as between observations taken with a spectroheliograph and an interference filter.

## 2.2. Methods

In our study we analysed raw DS1 images without calibration of the digitising device, raw DS2 data divided by the flat image taken closest in time to the scanning time, and calibrated Rome/PSPT data. Some key characteristics of these archives are listed in Table 1.

We used the DS1 and Rome/PSPT data already processed with the technique described by Chatzistergos *et al.* (2018a,b, 2019a) and in the present study we applied exactly the same processing to the data from DS2. In brief, the images were converted to density images and then photometrically calibrated with a calibration curve (CC, hereafter). The CC was derived by relating the centre-to-limb variation (CLV, hereafter) measured in quiet-Sun regions (QS, hereafter) on the historical observations to a standard reference of QS CLV as measured in modern Rome/PSPT Ca II K observations, and linearly extrapolated to the non-QS regions. Contrast images were constructed by removing the limb darkening, which was determined in an iterative process. This includes application of a running-window median filter and polynomial fitting along rows, columns, and radial locations after the bright features had been removed. The contrast images were then segmented with a multiplicative factor to the standard deviation of the QS intensity values. This factor was determined with a method based on the approach of Nesme-Ribes, Meunier, and Collin (1996). The multiplicative factor for identifying plage was chosen to be 8.5. This choice resulted in plage areas from the Rome/PSPT that are comparable to those from the SRPM segmentation scheme (Fontenla *et al.*, 2009) applied to the same Rome/PSPT data. We emphasize that for this work the precise value of the segmentation parameter is of no importance, since the main goal here was to process all series consistently to understand some of the difference between the results presented in the literature. Figure 3 shows examples of the density images and the processed images (calibrated and limb-darkening compensated images) as well as the segmentation masks for the observation shown in Figure 1. Images with severe artefacts (*e.g.* missing parts of disc) were excluded from analysis as was done by Chatzistergos *et al.* (2018a, 2019a).

## 3. Characteristics of Datasets

### 3.1. Image Comparison

Here we compare the images from DS1 and DS2 to study the differences due to the various digitisation set-ups. As an example, Figure 3 shows the same solar observations in the two datasets. The full-disc images appear rather similar; however, a more detailed inspection reveals significant differences. Figure

4 displays enlarged areas of the observation shown in Figure 1 including a QS and a plage region with a sunspot. The different magnifications correspond to regions with a width of 200 and 50 pixels, respectively, in the DS1 image. Due to the lower spatial resolution of the data of DS1 compared to those of DS2, the pixel size in DS1 is slightly larger than in DS2 when projecting the solar disc in the same physical dimensions. There are evident compression effects in the DS1 images, manifested as smoothed  $\approx 8 \times 8$  pixel<sup>2</sup> regions. The images from DS2 reveal much finer solar structures. Furthermore, the effect of compression seems to affect the QS regions in DS1 more than the plage area.

Figure 4 shows also the enlarged areas of the calibrated and limb-darkening-compensated images. Contours outline the regions that were identified as plage in each image. The regions identified as plage appear larger and coarser in the DS1 data compared to the DS2 ones.

We now compare the images from DS2 to those from the Rome/PSPT. Figure 5 shows an image from DS2 and Rome/PSPT before and after selected processing steps. The DS2 and Rome/PSPT datasets overlap during the period 1996–2007. Over that period, we generally notice saturated regions in the DS2 data, while the QS CLV in DS2 data is stronger than in the Rome/PSPT images. One example of an image with saturated regions in the DS2 data is shown in Figure 6 along with the image from the DS1 and Rome/PSPT of the same day. The images from both DS1 and DS2 display saturated regions, while the DS2 image is blurred compared to the DS1 one, suggesting that it might have been taken out of focus. The contrast in the calibrated and CLV-compensated DS2 image is lower than that in the Rome/PSPT one, which is contrary to the expectation due to the narrower bandwidth of the Kodaikanal observations. This suggests that Kodaikanal observations over that period suffer from severe stray-light effects. The identified plage regions from both images appear very similar, yet with some differences. More small regions can be identified in the DS2 image compared to the Rome/PSPT one, while large plage regions appear more extended in the Rome/PSPT image than in the DS2 one. We note, however, that the regions appear different to some degree due to the time difference between the observations of DS2 and Rome/PSPT, which were taken with a time difference of about six hours on average, with the Kodaikanal one preceding the Rome/PSPT observation.

Figure 7 shows a magnified section of a plage region from the processed images displayed in Figure 5. This sub-array was extracted close to the limb and the plage region is smaller but more squeezed in DS2 than in the Rome/PSPT observation. We also marked with contours the regions identified as sunspots in the Rome/PSPT image, by picking the regions that lie below  $\bar{C} - 3\sigma$ , where  $\bar{C}$  is the mean contrast over the disc and  $\sigma$  is the standard deviation of contrast values within the disc. The same contours are over-plotted in the DS2 image, scaled using the radius ratio of the two images. The sunspot regions cover roughly the same area, despite the difference in nominal bandwidth used by the two observatories (0.5 Å and 2.5 Å for Kodaikanal and Rome/PSPT data, respectively). This suggests that the Kodaikanal observations, at least the ones overlapping with Rome/PSPT, might have been taken with a broader bandwidth than the nominal one (see also Chatzistergos *et al.*, 2019a). However, these could also be,

**Table 1.** Characteristics of the analysed archives. See Section 3.2 for more details.

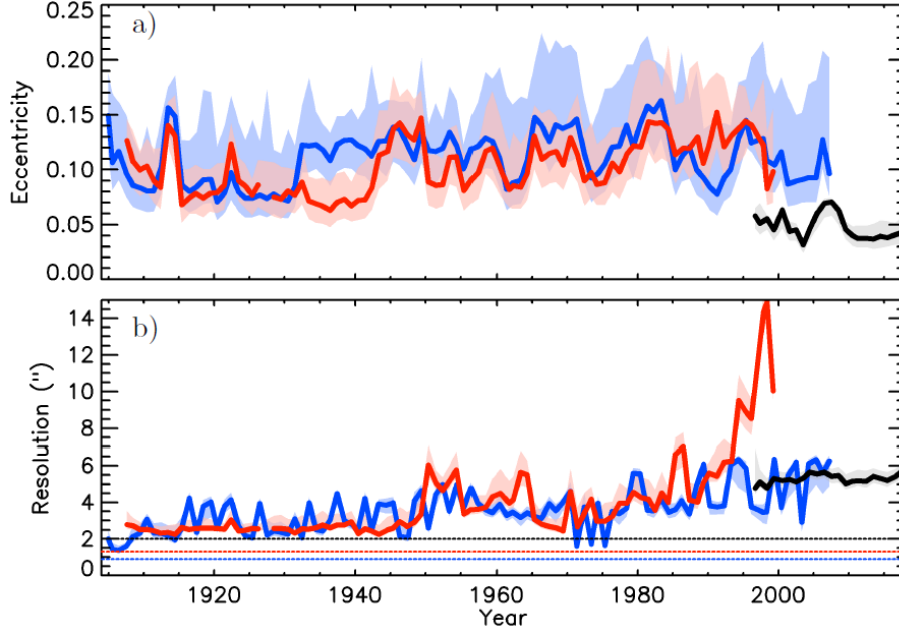
	DS1	DS2	Rome/PSPT
Period covered	1907–1999	1904–2007	1996–2018
Number of images	22158	48928	3292
Number of images used	19291	45519	3292
Number of days	21746	26492	3287
Number of days used	18963	25781	3287
Spectral bandwidth [ $\text{\AA}$ ]	0.5	0.5	2.5
Pixel scale [ $''\text{pixel}^{-1}$ ]	1.3	0.9	2
Disc eccentricity	$0.10 \pm 0.05$	$0.11 \pm 0.06$	$0.05 \pm 0.02$
Spatial resolution [ $''$ ]	$3.5 \pm 1.6$	$3.3 \pm 1.2$	$5.3 \pm 0.7$
Inhomogeneities	$0.07 \pm 0.02$	$0.07 \pm 0.03$	$0.03 \pm 0.02$
$\sigma_D$ raw images	$0.09 \pm 0.04$	$0.23 \pm 0.06$	-
$\sigma_C$ processed images	$0.06 \pm 0.02$	$0.05 \pm 0.02$	$0.06 \pm 0.02$
$\sigma_C^{\text{QS}}$ processed images	$0.020 \pm 0.008$	$0.019 \pm 0.010$	$0.021 \pm 0.003$
Reduced $\chi^2$	$0.002 \pm 0.002$	$0.002 \pm 0.002$	-

at least partly, differences due to the use of a filter and a spectroheliograph in the Rome/PSPT and DS2 data, respectively. The exact effect of the bandwidth on the sunspot areas would require further investigation. Furthermore, the contrast of the sunspot regions in the DS2 is reduced compared to that in the Rome/PSPT images. This might be due to stray-light or underexposure over the very dark regions of the disc.

### 3.2. Characteristics of Time Series

To assess the quality of the two Kodaikanal series and to compare it to that of modern data, we consider different image attributes that allow us to describe both global and local properties of the digital data. In particular, following Ermolli *et al.* (2009a) we study the eccentricity of the solar disc, the spatial resolution, the large scale inhomogeneities, similarity of QS CLV to that from the Rome/PSPT, as well as density/intensity characteristics. Table 1 summarizes the results obtained, which are discussed in more detail in the following.

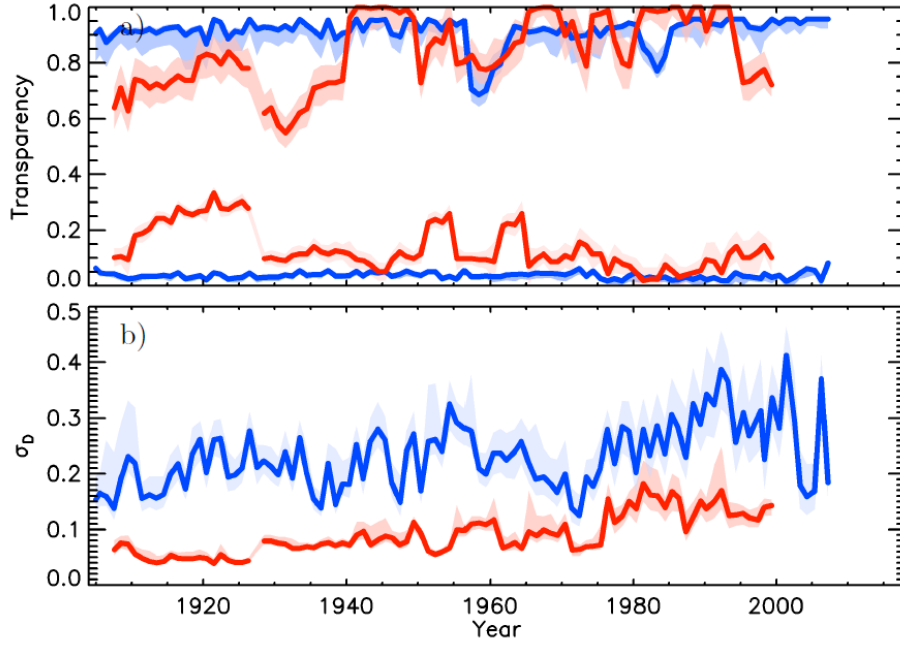
We computed the eccentricity of the recorded solar disc as  $e = \sqrt{1 - (R_{\min}/R_{\max})^2}$ , where  $R_{\min}$  and  $R_{\max}$  are the smallest and largest measured radii. However, the disc in the historical data does not always have an elliptical shape, rather an irregularly distorted shape due to problems with the drive of the spectroheliograph (affecting data from both DS1 and DS2) or with the digitising device (mainly affecting data from DS1). It is important to have an estimate of the eccentricity of the solar disc, since Chatzistergos *et al.* (2019a) demonstrated that the uncertainty in the derived plage areas increases with the disc eccentricity. The eccentricity of the solar disc from the DS1 and DS2 images is shown in Figure 8a). Annual averages are plotted along with the asymmetric  $1\sigma$  interval. We get an average value of  $\bar{e} = 0.10 \pm 0.05$  for the DS1 and  $\bar{e} = 0.11 \pm 0.06$  for the



**Figure 8.** Solar disc eccentricity (*top panel*) and spatial resolution (*bottom panel*) computed for DS1 (red), DS2 (blue), and Rome/PSPT (black) data as a function of time. Shown are annual mean values (solid lines) along with the asymmetric  $1\sigma$  interval (shaded surfaces). The horizontal dotted lines in the *lower panel* indicate the values for the average pixel scale of the data in the archives.

DS2 series. For both DS1 and DS2,  $e$  and  $\sigma_e$  increase with time. The value of  $e$  is relatively low prior to 1940 (DS1) and 1930 (DS2), but it is higher and varying afterwards. Ermolli *et al.* (2009a) had also studied the eccentricity of the disc in the DS1 and reported qualitatively the same results. In particular, they also found an increase in  $e$  and  $\sigma_e$  with time, although the jumps in the results in the early period were less pronounced in their case. They found a mean eccentricity of  $\bar{e} = 0.12 \pm 0.06$ , which is slightly higher than derived here. The eccentricity we derive for DS1 and DS2 is larger than to that for the Rome/PSPT data, which is on average  $\bar{e} = 0.05 \pm 0.02$ . This value is consistent with the value of  $\bar{e} = 0.04 \pm 0.03$  reported by Ermolli *et al.* (2009a). The maximum (RMS) error in the plage areas for the average disc eccentricity found for Kodaikanal data is 0.013 (0.0005), while for Rome/PSPT it is 0.002 (0.0003) as reported by Chatzistergos *et al.* (2019a). The eccentricities for the two Kodaikanal series are similar, although the one for DS2 is higher for most of the time, as can be seen from Figure 8a). Figure 8a) shows all data, but this is seen also when doing the same plot for the data that have the same date and time. In principle this can be due to a small inclination of the plate when it was digitised, or issues with the code to identify the limb, which might have been affected by artefacts.

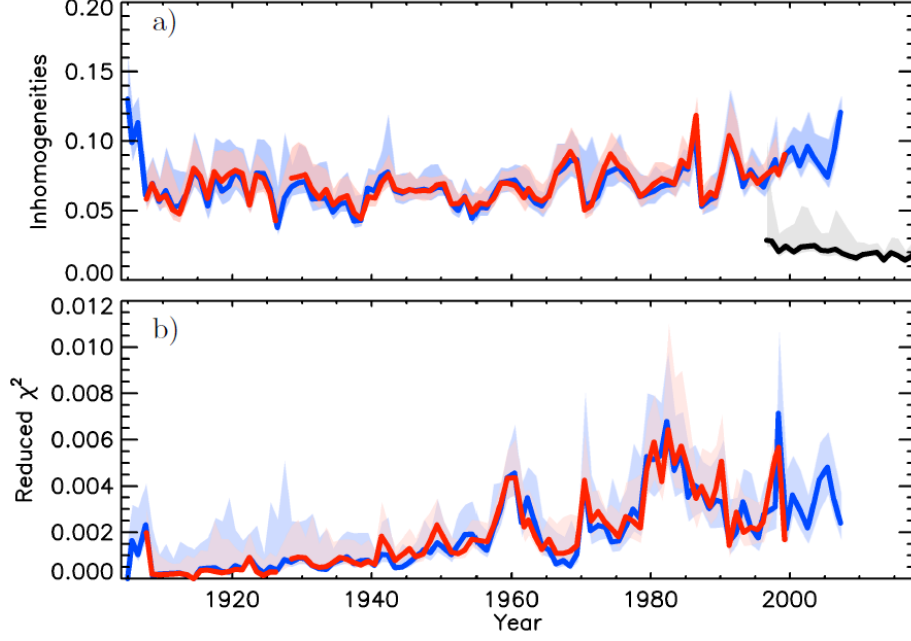
To study the spatial resolution of the images, we evaluate the frequency at which 98 % of the power spectral density is taken into account. The computation was performed on  $64 \times 64$  disc sub-arrays of quiet-Sun regions. We randomly



**Figure 9.** The maximum and the minimum values of transmittance (*top panel*) and the standard deviation of density values,  $\sigma_D$ , (*bottom panel*) within the solar disc for the DS1 (red) and DS2 (blue) data as a function of time. Shown are annual mean values (solid lines) along with the asymmetric  $1\sigma$  interval (shaded surfaces). These quantities are not defined for the Rome/PSPT data, which are given directly in units of intensity.

positioned 100 such segments within the inner  $R/3$  of the disc, and the average value from all the segments was adopted. This method is similar to the approach used by Ermolli *et al.* (2009a). However, we can potentially get a better estimate of the average spatial resolution of the data than Ermolli *et al.* (2009a), considering that the solar observation was not recorded instantaneously, but rather in strips with variable spatial resolution. The derived spatial resolution is shown in Figure 8b). We find the resolution for DS1 and DS2 in general to be roughly the same and slowly getting worse with time. Around 1980, the spatial resolution of DS1 degrades, sharply reaching an average value of  $15''$ . For the data from DS2, the resolution remains between 3 and  $6''$ . The average values over the whole period are  $3.5 \pm 1.6''$  and  $3.3 \pm 1.2''$  for DS1 and DS2, respectively. The spatial resolution of DS2 after 1990 is at similar levels as that of the Rome/PSPT data, but it is considerably worse for DS1 data. This is most likely due to an issue with the digitisation of DS1 data. The spatial resolution of the Rome/PSPT data is  $5.3 \pm 0.7''$  and roughly constant over the whole period. These values are consistent with the  $3.3 \pm 0.1''$  and  $5.0 \pm 0.4''$  for DS1 and Rome/PSPT, respectively reported by Ermolli *et al.* (2009a).

To compare the dynamic range of the data in the two digitisations, Figure 9a) shows the maximum and minimum values of the raw negative images, in units of transmittance, over the solar disc in the DS1 and DS2 data. The values from each

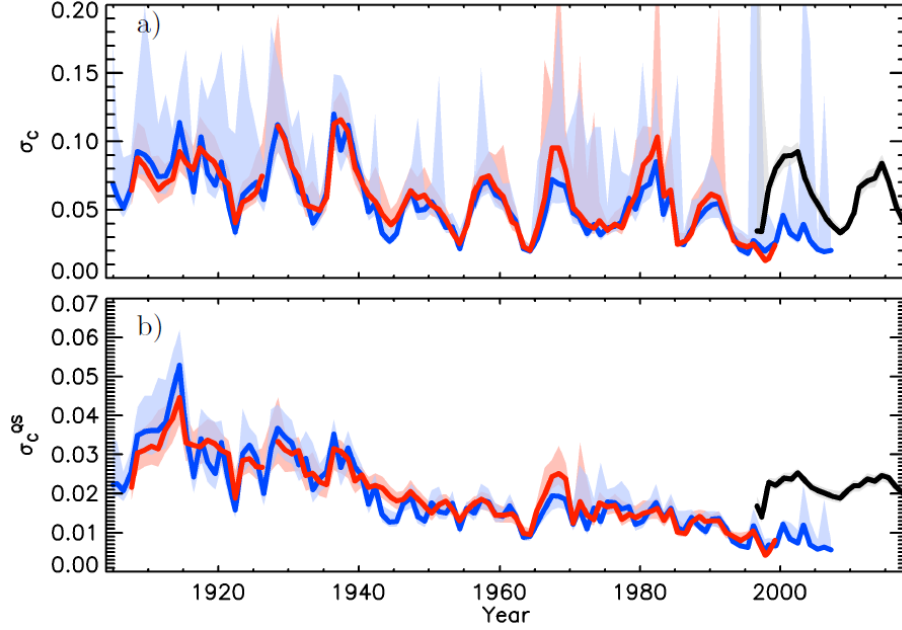


**Figure 10.** Identified large scale inhomogeneities (*top panel*) and reduced  $\chi^2$  of the fit to the curve obtained by relating the measured density QS CLV from the Kodaikanal data to a reference QS CLV from Rome/PSPT data (*bottom panel*) for DS1 (red), DS2 (blue), and Rome/PSPT (black, only in the top panel) data as a function of time. Shown are annual mean values (solid lines) along with the asymmetric  $1\sigma$  interval (shaded surfaces). The  $\chi^2$  is not applicable to the Rome/PSPT data.

dataset were normalised to the maximum value from the respective digitisation. On average, DS2 is much more stable, with the exception of two periods around 1959 and 1984 during which the maximum values decrease. DS1 shows a larger variation with time with abrupt jumps in the transmittance values and gaps. There are periods where the maximum transmittance value is found within the solar disc of DS1 data, hinting at saturation of the low-density regions. This suggests that DS2 has been digitised in a more consistent manner than DS1. Figure 9b) shows the standard deviation of the density values,  $[\sigma_D]$ , over the solar disc from both digitisations. The standard deviation in DS2 is consistently higher than that in DS1. In both datasets a slight increase of the standard deviation with time is observed, which could be due to an increase in CLV because the employed bandwidth became broader than before or the observations were not centred at the core of the line.

Next, we assess whether and how strongly the images are affected by the large-scale inhomogeneities and artefacts. For this, we compute the relative difference of the image background calculated by the image processing to the QS CLV. The image background is a 2D surface map that includes the QS CLV as well as all identified large-scale inhomogeneities and artefacts. This is determined with the iterative process described by Chatzistergos *et al.* (2018a). To make





**Figure 11.** The standard deviation of the contrast values,  $[\sigma_C]$ , over the whole solar disc (*top panel*) and in the quiet Sun only,  $[\sigma_C^{QS}]$ , (*bottom panel*) for the calibrated and limb-darkening compensated DS1 (red), DS2 (blue), and Rome/PSPT (black) data as a function of time. Shown are annual mean values (solid lines) along with the asymmetric  $1\sigma$  interval (shaded surfaces).

the results between the different datasets comparable, we rescaled (using cubic interpolation) only for this test all images to the same dimensions, such that the radius is always 350 pixel. Figure 10a) shows the values of the relative difference of the background to the QS CLV we get for DS1, DS2, and Rome/PSPT data. The level of inhomogeneities increases with time for both the DS1 and DS2 data. The level of inhomogeneities in the Rome/PSPT decreases with time and at  $0.03 \pm 0.02$  is lower than the values from both Kodaikanal series.

From the data produced during the image processing, we also analysed the  $\chi^2$  of a linear fit to the curve obtained by relating the measured density QS CLV to the logarithm of the reference intensity QS CLV from Rome/PSPT data. The reduced  $\chi^2$  from this fit can be considered as an indication for instrumental changes, *e.g.* in the bandwidth or central wavelength (see Chatzistergos *et al.*, 2019a, for a discussion on this). Figure 10b) shows the derived reduced  $\chi^2$  from the fit for images from DS1 and DS2. There are no significant differences between the resulting values and their variation with time in the two series. The  $\chi^2$  of the fit increases with time for both DS1 and DS2, hinting at a gradual degradation of image quality rather than the digitisation.

We have also analysed the contrast values in the calibrated and limb-darkening compensated images. Figure 11a) shows the standard deviation of contrast values over the entire solar disc:  $\sigma_C$ , for DS1, DS2, and Rome/PSPT. Unsurprisingly,  $\sigma_C$  over the disc qualitatively follows the solar-cycle variability. However,

it is less pronounced before 1920, while the amplitude of  $\sigma_C$  over Solar Cycles 18, 19, 21, and 22 is lower than the others suggesting some variations in the instrument parameters, or data quality. The values from the Rome/PSPT data are at similar levels to those derived for both DS1 and DS2 for most Solar Cycles. Figure 11b) shows the same but only for the QS regions:  $\sigma_C^{QS}$ . The standard deviation of the QS decreases with time in both DS1 and DS2 series. This can be due to increased underexposure with time or broadening of the bandwidth, if the observation is centred more towards the wing of the line than its core, or due to observational and instrumental effects such as worsening seeing at the observing site, degradation of the spectroheliograph, and changes in the quality of the plates. The value for the standard deviation of the QS for the Rome/PSPT is always higher than for both Kodaikanal series. This again lends support to the argument that the effective bandwidth of the Kodaikanal over the overlapping period might be broader than the one of the Rome/PSPT or that the Kodaikanal observations were off-centre over that period. However, we note that this is not a conclusive test since any of the instrumental or observational issues mentioned above can also contribute to the change of the standard deviation of the QS with time.

Overall, we find both DS1 and DS2 series to exhibit almost the same characteristics and temporal behaviour, suggesting that the studied characteristics are intrinsic to the original Kodaikanal data and are not artefacts of the digitisation. These include the worsening of the spatial resolution, the increase of the disc eccentricity, the enhancement of the large-scale inhomogeneities, and the change of the CLV with time. Hence, these should be ascribed to the original photographic observations and not to image issues introduced by the digitisation. DS2 shows a more consistent distribution of transparency values, suggesting that the digitisation was performed more consistently than for DS1. The data after 1990 show an improved spatial resolution in DS2 data compared to those in DS1.

## 4. Plage Areas

### 4.1. Results from DS1, DS2, and Rome/PSPT

We have derived plage areas from DS1 and DS2 processed in the same way. The results are shown in Figure 12 along with plage areas from the Rome/PSPT data and the sunspot areas by Balmaceda *et al.* (2009)<sup>1</sup>. We give a tentative RMS error in the derived plage areas from the Kodaikanal and Rome/PSPT data of 0.0025 and 0.0003, respectively, in fraction of the disc area. The value for Kodaikanal data is the sum of the RMS errors due to the disc ellipticity and errors of the processing of the images to perform the photometric calibration as evaluated by Chatzistergos *et al.* (2018a, 2019a). The value for Rome/PSPT data is acquired by considering only the effect of the disc ellipticity. We note, however, that this is not a strictly defined, formal error for the derived plage

<sup>1</sup>The series has been extended to 31 May 2017 and is available at [www2.mps.mpg.de/projects/sun-climate/data.html](http://www2.mps.mpg.de/projects/sun-climate/data.html)

areas, but our best estimate based on our analysis of synthetic data with our method.

The derived plage areas for DS2 data are distinctly higher than those from DS1 for Solar Cycles 15, 18, and 20 and slightly larger in Cycle 17, while in Cycles 21 and 22 the opposite is the case. The two series have RMS difference of 0.006 and Pearson coefficient of 0.95 when daily values are considered (shown in Table 2). However, when considering only the observations that have been taken at the same day and same time the RMS difference becomes 0.005 and the Pearson coefficient 0.98. In Figure 12 we also show the annual plage areas for the common days in DS1 and DS2. The RMS differences between the annual values for all data and only the common days in the two series are 0.0004 and 0.003 for DS1 and DS2, respectively. The smaller effect on the results for DS1 comes as no surprise considering that it is the series with the fewer observations. The agreement between the areas from the two datasets becomes worse for Solar Cycles 20 and 21 when considering only data taken on the same day and time, while it slightly improved for Cycle 22 and remained unchanged for all other cycles. Hence, the differences in the computed plage areas do not stem from differences in sampling of the original photographs. We note, however, that potential errors in the dating of the images could affect our results. Comparing DS1 and DS2 to the sunspot areas we find that Solar Cycles 21 and 22 in the plage areas fall differ from expectation, being too high relative to the previous cycles. Solar Cycle 18 in DS2 is between Cycle 17 and 19 which is in agreement with the ranking of the Solar Cycles in the sunspot areas. This is not the case for Solar Cycle 18 in DS1.

We notice that the results for Rome/PSPT are in relatively good agreement with those from both the DS1 and DS2 series within the period of overlap. The areas from DS2 over the maximum of Solar Cycle 23 are slightly lower by 0.003 than that from the Rome/PSPT, while we also notice an increase in the plage areas of DS2 in 2004, something not seen in the areas from Rome/PSPT or the sunspot areas. We found merely 26 days with data from all three datasets, DS1, DS2, and the Rome/PSPT in the period 19 May 1997 to 29 May 1999. Figure 13 shows the areas in disc fractions derived from all three datasets, by considering only the days available in all three archives. For that period, the areas from both DS1 and DS2 lie mostly below the one from Rome/PSPT except for two and four days for which DS1 and DS2 give greater plage areas, respectively. The RMS difference between the areas derived from DS1 and DS2 to Rome/PSPT is 0.011 and 0.008, respectively, while the maximum absolute difference is 0.033 and 0.019, respectively.

Restricting the comparison between DS2 and Rome/PSPT gives 757 days of overlap over the period 23 April 1997 to 10 September 2007. The difference between the plage areas derived from DS2 to that from Rome/PSPT is shown in the lower panel of Figure 13. We get an average RMS difference of 0.01, while the maximum absolute difference reaches up to 0.08. We notice an annual variation in the differences between the series, with Rome/PSPT giving higher plage areas during Winter periods than DS2, while the opposite occurs during the Summer.

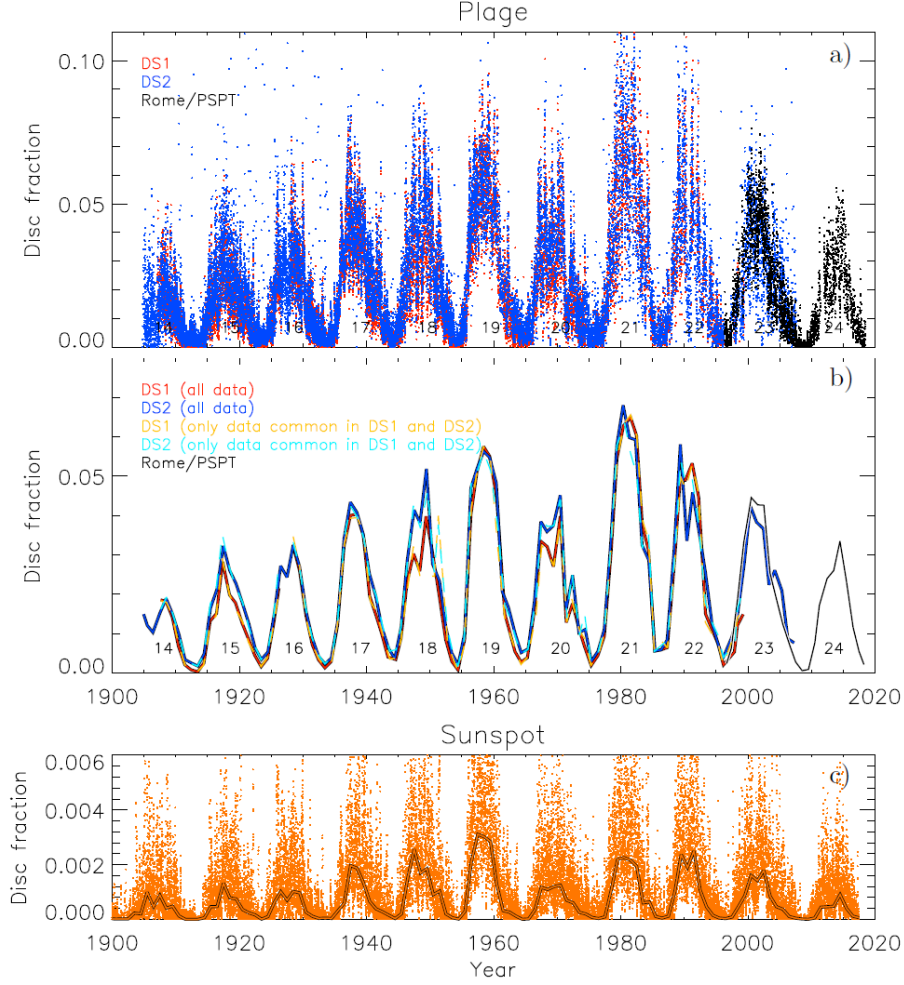
Figure 14 shows scatter plots between the derived plage areas from DS2 and those from DS1 and Rome/PSPT. We find a good agreement between all series, with linear correlation of 0.97 and 0.94 for DS1 and Rome/PSPT, respectively.

## 4.2. Comparison to Other Results

Figure 15 shows scatter plots between the plage areas derived by us from the DS2 data and the various published series obtained from Kodaikanal observations. In particular we consider the series by Kuriyan, Muralidharan, and Sampath (1983) derived from the physical photographs, the series by Ermolli *et al.* (2009a) and Tlatov, Pevtsov, and Singh (2009) from the DS1 data, and the series by Chatterjee, Banerjee, and Ravindra (2016), Priyal *et al.* (2017), and Singh *et al.* (2018) from the DS2 data. Note that the series by Kuriyan, Muralidharan, and Sampath (1983) and Tlatov, Pevtsov, and Singh (2009) are only available as annual values and most likely include different selections of observations. For all the other series, we consider only the days common with the data from DS2 we use here. Besides comparing the daily values, we also compute and compare annual median values. However, we note that potential errors in the dates and times of the different original archives as well as the copies of them used by the respective authors, affect our results. The scatter between the various series and ours is rather significant. Table 2 lists the RMS difference and the Pearson coefficient between the various series when daily values are used. The best agreement is found between our results and those by Chatterjee, Banerjee, and Ravindra (2016) and Singh *et al.* (2018), although when annual values are used then the agreement is better with the series by Ermolli *et al.* (2009a) and Tlatov, Pevtsov, and Singh (2009). However there are hints of a non-linearity between our plage areas and those by Tlatov, Pevtsov, and Singh (2009). In Chatzistergos *et al.* (2019a) we did a similar comparison, but considering the above series and our results derived from analysis of DS1. Comparing the scatter plots in Figure 15 and those of Chatzistergos *et al.* (2019a) we find an improvement in the match between our series and that from Chatterjee, Banerjee, and Ravindra (2016) with linear correlation factor increasing to 0.94 compared to 0.9 for the annual values. Our results for the plage areas from DS2 data show worse agreement with the series by Ermolli *et al.* (2009a) and Tlatov, Pevtsov, and Singh (2009) from analysis of DS1 data than our results from DS1 data. However, we also notice a worse agreement with our results from DS2 data and those by Priyal *et al.* (2017) and Singh *et al.* (2018) from analysis of DS2 data than our results from DS1 data.

We now discuss various factors that are responsible for part of the differences. Such factors are the definition of the centre coordinates and radius, the fraction of the disc used for the normalisation of the plage areas, as well as the processing techniques including the photometric calibration.

For example, different definitions of radius and centre coordinates affect the results of all series. Priyal *et al.* (2014) defined the radius and centre coordinates for the DS2 by manually selecting three points at the limb. This information was not stored, but rather was used to crop and centre the images. These cropped images were then used by Chatterjee, Banerjee, and Ravindra (2016), Priyal *et al.* (2017), and Singh *et al.* (2018) to derive their plage areas. Ermolli *et al.* (2009a) and Tlatov, Pevtsov, and Singh (2009) defined the radius and centre coordinates for DS1 independently. Chatzistergos *et al.* (2019a) used the radius estimates for DS1 made by Ermolli *et al.* (2009a), but we corrected a few errors. Here for



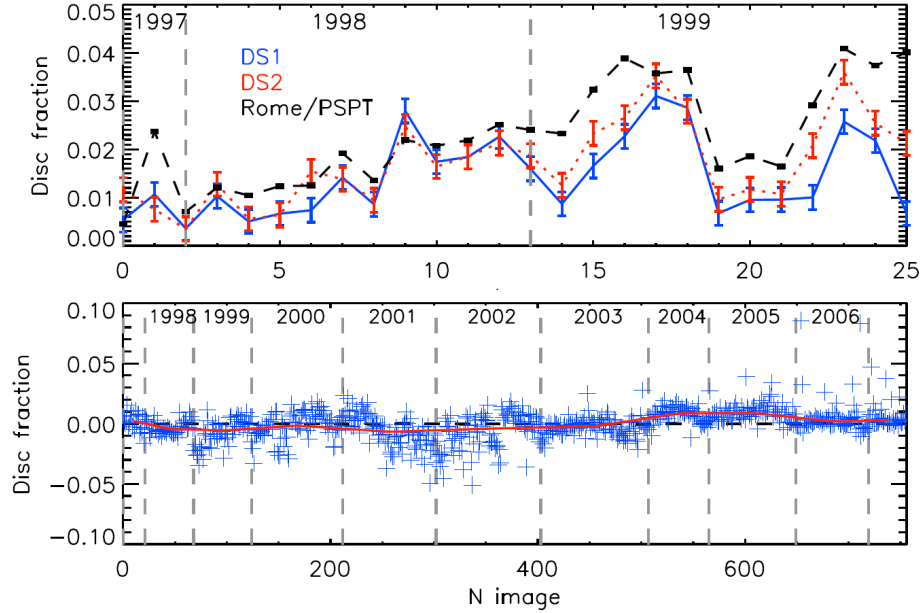
**Figure 12.** Fractional disc coverage by plage (Panels **a** and **b**) as a function of time, derived with the same processing and segmentation parameters from DS1 (red), DS2 (blue), and Rome/PSPT (black) images. Panel **a** shows daily values, while panel **b** displays annual mean values (solid lines). The dashed lines in panel **b** show the annual values for DS1 (orange) and DS2 (light blue) when only the common days in DS1 and DS2 are considered. Also shown (Panel **c**) are daily (dots) and annual (solid line) values of sunspot areas from Balmaceda *et al.* (2009).

DS2 we determined the radius from scratch by using the method described by Chatzistergos *et al.* (2019a).

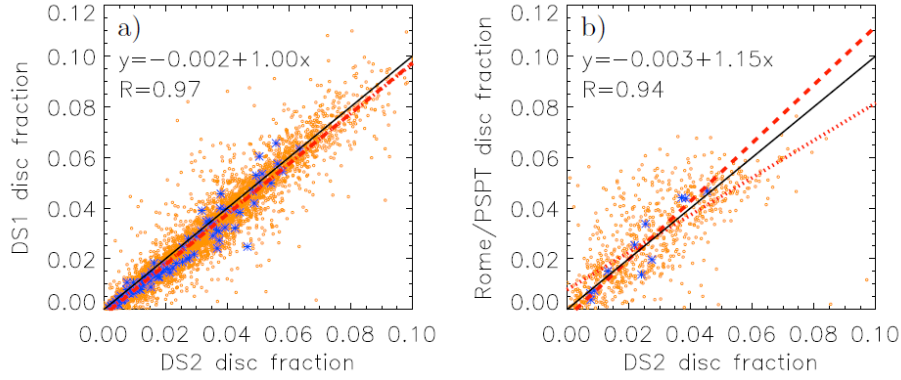
Studies in the literature also differ in the fraction of the disc area that was used to normalise the identified plage areas. For our study we used the disc area reaching out to  $0.98R$  (i.e., 96% of the total area), while Ermolli *et al.* (2009a), Chatterjee, Banerjee, and Ravindra (2016), and Priyal *et al.* (2017) used the disc up to  $0.97R$  (94% of the total area) Priyal *et al.* (2014) used the disc up to  $0.985R$  (97% of the total area). Figure 16 shows an example

**Table 2.** Quantification of the agreement between different plage area series. The values above the diagonal are the RMS differences, while those below the diagonal are the Pearson coefficients, both computed for the common days. The number of common days is given within the brackets. The abbreviations CEA16, EEA09, PEA17, and SEA18 refer to the Chatterjee, Banerjee, and Ravindra (2016), Ermolli *et al.* (2009a), Priyal *et al.* (2017), and Singh *et al.* (2018) series, respectively.

	DS1	DS2	CEA16	EEA09	PEA17	SEA18
DS1	-	0.006 [18387]	0.013 [12657]	0.012 [18454]	0.012 [10992]	0.009 [14370]
DS2	0.955 [18387]	-	0.011 [15479]	0.015 [20385]	0.015 [13112]	0.010 [16357]
CEA16	0.886 [12657]	0.896 [15479]	-	0.020 [13492]	0.019 [12499]	0.014 [12940]
EEA09	0.879 [18454]	0.838 [20385]	0.845 [13492]	-	0.013 [11593]	0.013 [14511]
PEA17	0.827 [10992]	0.784 [13112]	0.777 [12499]	0.780 [11593]	-	0.011 [12063]
SEA18	0.885 [14370]	0.867 [16357]	0.857 [12940]	0.827 [14511]	0.854 [12063]	-



**Figure 13.** *Top:* Plage areas in disc fraction for 26 images taken on the same days found in DS1 (dotted red), DS2 (solid blue), and the Rome/PSPT (dashed black) series plotted against the number of the image. The error bars denote the RMS error in the derived plage areas due to the disc ellipticity and the processing to photometrically calibrate the images (the latter is applicable only to the Kodaikanal data) as found by Chatzistergos *et al.* (2019a). *Bottom:* Difference of plage areas derived from the 757 images common to DS2 and Rome/PSPT series plotted against the number of the image. Blue plus signs denote individual values, while the red solid line is for the annual mean value. The dashed vertical lines separate the years, which are written at the top of each panel.



**Figure 14.** Scatter plots between the plage area values derived from images of DS2 ( $x$ -axis) and those from images ( $y$ -axis) of DS1 (Panel **a**) and Rome/PSPT (Panel **b**). Blue asterisks (orange dots) show the annual (daily) values. The solid black lines have a slope of unity and represent the expectation value. The dashed (dotted) red lines are linear fits to the annual (daily) data. Also shown are the corresponding parameters of the linear fits to the annual values and the linear correlation coefficients of the annual values.

observation from Kodaikanal where the regions considered in the various studies have been marked. Notice that the various studies defined the centre of the solar disc differently. To get an error estimate for using a slightly different normalising area, we repeated the segmentation of DS1 data by considering the area of the disc up to  $0.97R$ . We found relative differences in the derived plage areas when considering the disc up to  $0.97R$  to  $0.98R$  that are  $0.01 \pm 0.04$ .

Processing artefacts contribute more to the systematic differences in the various results presented in the literature. For example, Figure 17 shows the observation from Figure 16 calibrated with our method and with those used by Ermolli *et al.* (2009a); Priyal *et al.* (2014); Chatterjee, Banerjee, and Ravindra (2016). The QS regions in the image processed with our method are more uniform compared to the others. The images processed by Ermolli *et al.* (2009a) and Priyal *et al.* (2014) show remaining large-scale inhomogeneities that can affect the plage area determination with these methods. In the image processed by Chatterjee, Banerjee, and Ravindra (2016) the large-scale inhomogeneities have been accounted for, but the contrast of the plage regions has been suppressed, causing the immediately surrounding areas of large plage regions to become much darker. Thus a smaller part of the large plage areas will be considered as plage, while some network elements might be counted as plage.

## 5. Conclusions

The photographic archive of full-disc Ca II K observations of the Kodaikanal observatory is a valuable source of information on past solar activity. We have compared the two more recent digitisations of this archive to understand if there are any differences in the results and if these differences are responsible for the partly conflicting results presented in the literature. For this, we have processed

images from the two digitisations consistently and applying the same technique. Thus, we have applied the methods developed and tested by Chatzistergos *et al.* (2018a) to the 16-bit data series as well. We applied the same processing on modern CCD-based data from Rome/PSPT.

The plage areas derived from DS1 and DS2 and their variation with time are rather similar to each other. Many of the issues previously reported about the varying quality of the Kodaikanal data are found to apply to the data from the new digitisation too, implying that they are intrinsic characteristics of the physical archive. These are an increase with time in the disc eccentricity coupled with a worsening spatial resolution, growing large-scale inhomogeneities, and change of the QS CLV with time. However, we found the quality of the DS1 data after 1990 to deteriorate more than that of the DS2 data. This can introduce significant errors when cross-calibrating plage area series from different archives or for irradiance reconstructions. Furthermore, both digitisations of Kodaikanal data seem to suffer from errors in the meta-data, especially concerning the observational date and time of the plates. This issue plagues all analyses from these data.

We find a good match between the plage areas derived from the Kodaikanal archive and those from the Rome/PSPT, with the Kodaikanal areas being slightly lower. This is in agreement with a drift in the quality of the Kodaikanal data and considering that Rome/PSPT has a nominal bandwidth which is five times broader than that of the Kodaikanal observatory. The plage areas presented in the literature from the various digitisations of Kodaikanal data show significant differences. We suggest that the diverse employed methods to calibrate the data as well as the different definitions of the recorded solar radius are the main reasons for the discrepancies among the various published results.

Overall, we found the new digitisation of the Kodaikanal archive to offer an improvement in the image quality over the 8-bit series. It also has more than doubled the available images. With over 48,000 images, the new Kodaikanal series is possibly the richest currently available Ca II K archive, so that it has a great potential to improve our understanding of solar activity, especially when the various issues affecting the series have been properly addressed.

**Acknowledgments** The authors thank the Kodaikanal and the Rome Solar Groups. The newly digitized Kodaikanal data as presented here (DS2) are available at [kso.iap.res.in/](http://kso.iap.res.in/). D. Banerjee. and I. Ermolli thank the International Space Science Institute (Bern, Switzerland) for supporting the International Team 420 "Reconstructing solar and heliospheric magnetic field evolution over the past century". T. Chatzistergos acknowledges a postgraduate fellowship of the International Max Planck Research School on Physical Processes in the Solar System and Beyond. This work was supported by grants COST Action ES1005 "TOSCA", FP7 SOLID, 01LG1909C from the German Federal Ministry of Education and Research, and by the BK21 plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea. This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 824135 (SOLARNET). We thank the anonymous referee for the valuable comments which helped to improve this manuscript. This research has made use of NASA's Astrophysics Data System.

#### **Disclosure of Potential Conflicts of Interest**

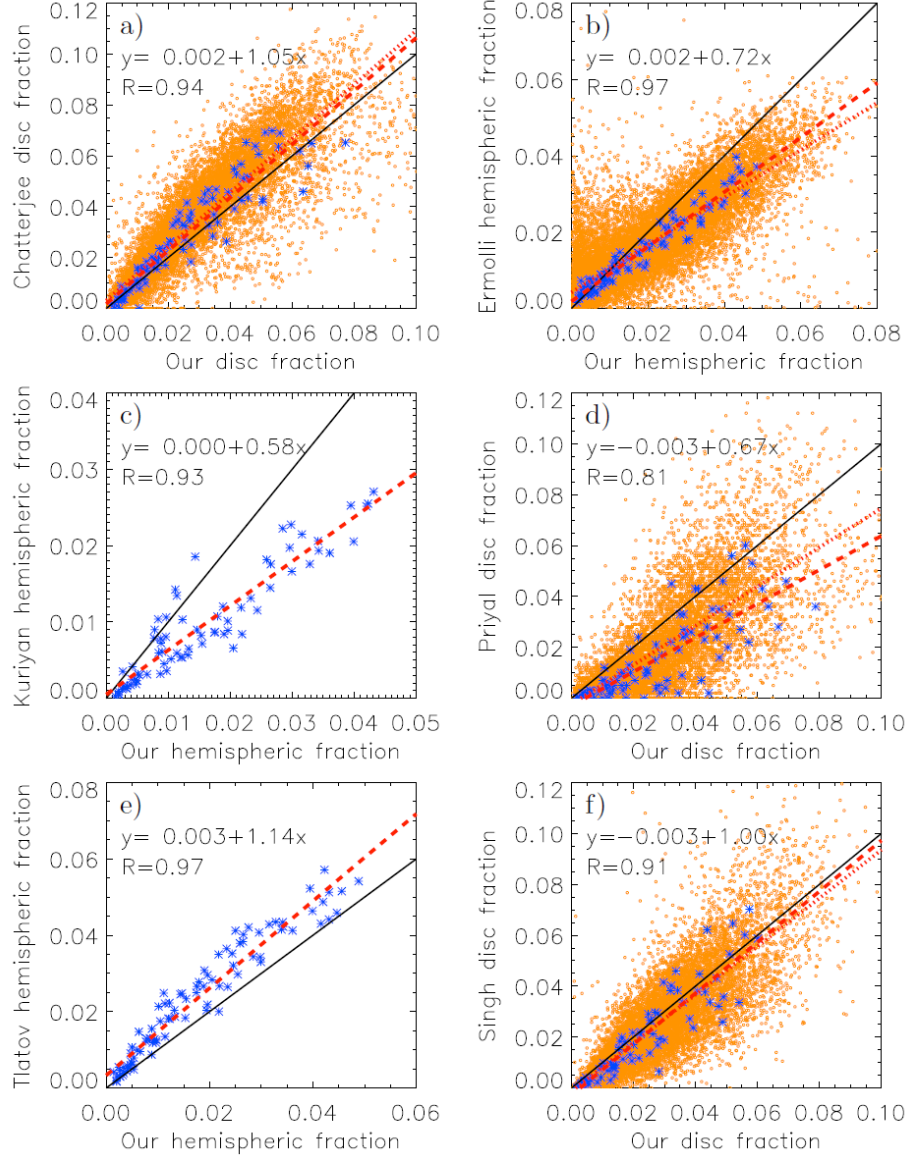
The authors declare that they have no conflicts of interest.



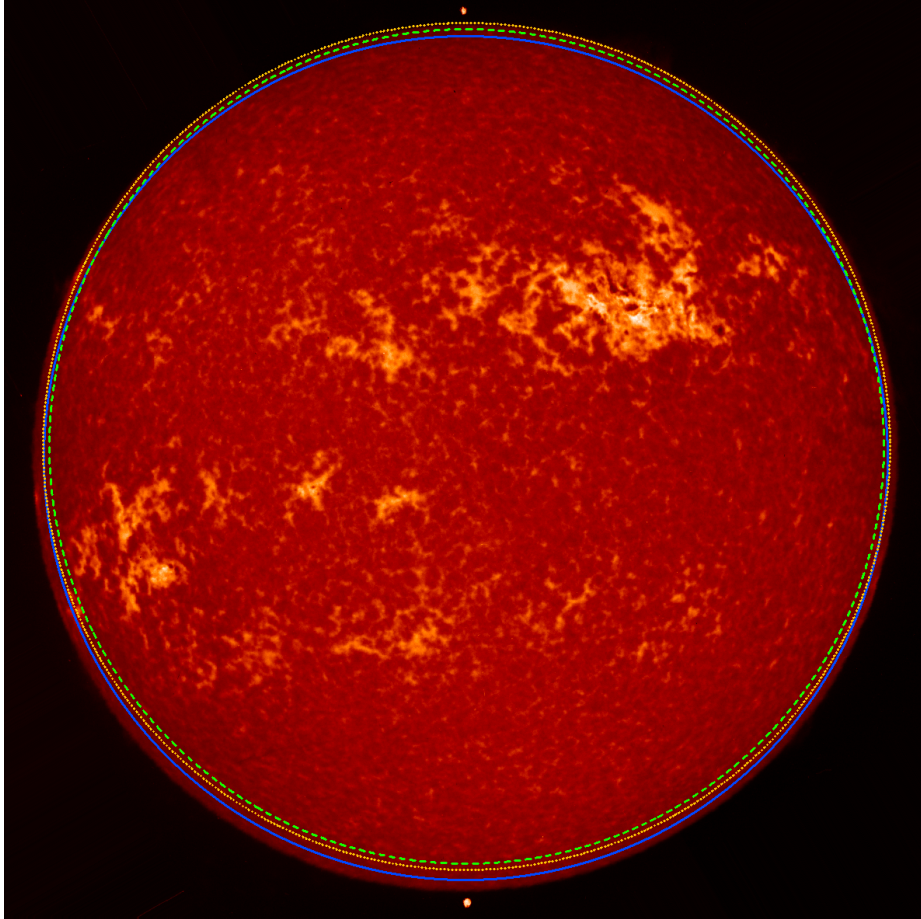
## References

- Balmaceda, L.A., Solanki, S.K., Krivova, N.A., Foster, S.: 2009, A homogeneous database of sunspot areas covering more than 130 years. *Journal of Geophysical Research: Space Physics* **114**(A7). DOI.
- Chatterjee, S., Banerjee, D., Ravindra, B.: 2016, A butterfly diagram and carrington maps for century-long Ca II K spectroheliograms from the Kodaikanal observatory. *The Astrophysical Journal* **827**(1), 87. DOI.
- Chatterjee, S., Mandal, S., Banerjee, D.: 2017, Variation of Supergranule Parameters with Solar Cycles: Results from Century-long Kodaikanal Digitized Ca II K Data. *The Astrophysical Journal* **841**, 70. DOI.
- Chatzistergos, T.: 2017, *Analysis of historical solar observations and long-term changes in solar irradiance*, PhD thesis, Uni-edition, ??? ISBN 978-3-944072-55-5.
- Chatzistergos, T., Ermolli, I., Solanki, S.K., Krivova, N.A.: 2018a, Analysis of full disc Ca II K spectroheliograms - I. Photometric calibration and centre-to-limb variation compensation. *Astronomy & Astrophysics* **609**, A92. DOI.
- Chatzistergos, T., Ermolli, I., Krivova, N.A., Solanki, S.K.: 2018b, Ca II K spectroheliograms for studies of long-term changes in solar irradiance. 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, 125. DOI.
- 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. *Astronomy & Astrophysics* **625**, A69. DOI.
- Chatzistergos, T., Ermolli, I., Falco, M., Giorgi, F., Guglielmino, S.L., Krivova, N.A., Romano, P., Solanki, S.K.: 2019b, Historical solar Ca II K observations at the Rome and Catania observatories. In: *Il Nuovo Cimento* **42C**, 5. DOI.
- Chatzistergos, T., Ermolli, I., Solanki, S.K., Krivova, N.A., Giorgi, F., Fabrizio, Y., Yeo, Kok Leng: 2019c, Recovering the unsigned photospheric magnetic field from Ca II K observations. *A&A* **626**, A114. DOI.
- Ermolli, I., Berrilli, F., Florio, A.: 2003, A measure of the network radiative properties over the solar activity cycle. *Astronomy and Astrophysics* **412**, 857. DOI.
- Ermolli, I., Fofi, M., Bernacchia, C., Berrilli, F., Caccin, B., Egidi, A., Florio, A.: 1998, The Prototype Rise-Pspt Instrument Operating in Rome. *Solar Physics* **177**, 1.
- Ermolli, I., Criscuoli, S., Centrone, M., Giorgi, F., Penza, V.: 2007, Photometric properties of facular features over the activity cycle. *Astronomy and Astrophysics* **465**, 305. DOI.
- Ermolli, I., Solanki, S.K., Tlatov, A.G., Krivova, N.A., Ulrich, R.K., Singh, J.: 2009a, Comparison Among Ca II K Spectroheliogram Time Series with an Application to Solar Activity Studies. *The Astrophysical Journal* **698**, 1000. DOI.
- Ermolli, I., Marchei, E., Centrone, M., Criscuoli, S., Giorgi, F., Perna, C.: 2009b, The digitized archive of the Arcetri spectroheliograms. Preliminary results from the analysis of Ca II K images. *Astronomy and Astrophysics* **499**, 627. DOI.
- 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.
- Fontenla, J.M., Curdt, W., Haberreiter, M., Harder, J., Tian, H.: 2009, Semiempirical Models of the Solar Atmosphere. III. Set of Non-LTE Models for Far-Ultraviolet/Extreme-Ultraviolet Irradiance Computation. *The Astrophysical Journal* **707**, 482. DOI.
- Foukal, P.: 1996, The Behavior of solar magnetic plages measured from Mt. Wilson observations between 1915-1984. *Geophysical Research Letters* **23**, 2169. DOI.
- Hale, G.E.: 1893, Solar photography at the Kenwood Astro-physical Observatory. *Memorie della Societa Degli Spettroscopisti Italiani* **21**, 68.
- Harvey, K.L.: The Cyclic Behavior of Solar Activity. In: *Proceedings of the National Solar Observatory/Sacramento Peak 12th Summer Workshop* **27**.
- Kariyappa, R., Sivaraman, K.R.: 1994, Variability of the solar chromospheric network over the solar cycle. *Solar Physics* **152**, 139. DOI.
- Kuriyan, P.P., Muralidharan, V., Sampath, S.: 1983, Long-term relationships between sunspots, Ca-plages and the ionosphere. *Journal of Atmospheric and Terrestrial Physics* **45**, 285. DOI.
- Loukitcheva, M., Solanki, S.K., White, S.M.: 2009, The relationship between chromospheric emissions and magnetic field strength. *Astronomy and Astrophysics* **497**, 273. DOI.

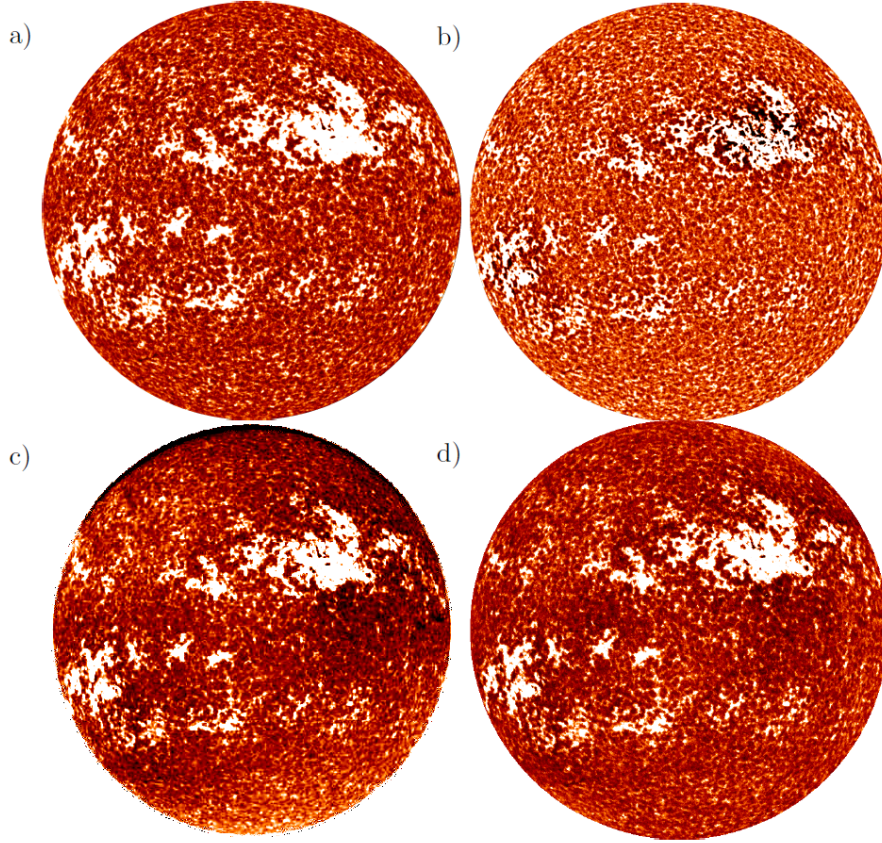
- Makarov, V.I., Tlatov, A.G., Singh, J., Gupta, S.S.: 2004, 22-years magnetic cycle in polar activity of the Sun. In: Stepanov, A.V., Benevolenskaya, E., Kosovichev, A.G. (eds.) *Multi-Wavelength Investigations of Solar Activity, Proceedings of the International Astronomical Union* **223**, Cambridge University Press, Cambridge, UK, 125. ISBN 1743-9221. DOI.
- Nesme-Ribes, E., Meunier, N., Collin, B.: 1996, Fractal analysis of magnetic patterns from Meudon spectroheliograms. *Astronomy and Astrophysics* **308**, 213.
- Pevtsov, A.A., Virtanen, I., Mursula, K., Tlatov, A., Bertello, L.: 2016, Reconstructing solar magnetic fields from historical observations. I. Renormalized Ca K spectroheliograms and pseudo-magnetograms. *Astronomy and Astrophysics* **585**, A40. DOI.
- Priyal, M., Singh, J., Ravindra, B., Priya, T.G., Amareswari, K.: 2014, Long Term Variations in Chromospheric Features from Ca-K Images at Kodaikanal. *Solar Physics* **289**, 137. DOI.
- 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**(7), 85. DOI.
- 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. *The Astrophysical Journal* **337**, 964. DOI.
- Sheeley, N.R. Jr., Cooper, T.J., Anderson, J.R.L.: 2011, Carrington Maps of Ca II K-line Emission for the Years 1915-1985. *The Astrophysical Journal* **730**, 51. DOI.
- Singh, J., Priyal, M., Sindhuja, G., Ravindra, B.: 2018, Variations in Ca-K line profiles and Ca-K line features as a function of latitude and solar cycle during the 20th century. *Proceedings of the International Astronomical Union* **13**(S340), 23. DOI.
- Skumanich, A., Lean, J.L., Livingston, W.C., White, O.R.: 1984, The sun as a star - Three-component analysis of chromospheric variability in the calcium K line. *The Astrophysical Journal* **282**, 776. DOI.
- Tlatov, A.G., Pevtsov, A.A., Singh, J.: 2009, A New Method of Calibration of Photographic Plates from Three Historic Data Sets. *Solar Physics* **255**, 239. DOI.



**Figure 15.** Plage areas presented in the literature ( $y$ -axis) versus the ones derived here from DS2 ( $x$ -axis): (a) Chatterjee, Banerjee, and Ravindra (2016) from DS2; (b) Ermolli *et al.* (2009a) from DS1; (c) Kuriyan, Muralidharan, and Sampath (1983) from the actual photographs; (d) Priyal *et al.* (2017) from DS2; (e) Tlatov, Pevtsov, and Singh (2009) from DS1; (f) Singh *et al.* (2018) from DS2. Blue asterisks (orange dots) show the annual (daily) values. The solid-black lines have a slope of unity. The dashed (dotted) red lines are linear fits to the annual (daily) data. Also shown are the corresponding parameters of the linear fits to the annual values and the linear correlation coefficients of the annual values.



**Figure 16.** Raw Kodaikanal observation taken on 20 January 1938. Circles enclose the areas considered by Chatterjee, Banerjee, and Ravindra (2016, dashed green), Priyal *et al.* (2017, dashed green), Priyal *et al.* (2014, dotted yellow), and in this work (solid blue).



**Figure 17.** Calibrated images of Kodaikanal observation taken on 20 January 1938 (shown in Figure 16): (a) with our method, and with the methods by (b) Chatterjee, Banerjee, and Ravindra (2016), (c) Priyal *et al.* (2014), and (d) Ermolli *et al.* (2009a). The images processed with our method and by Ermolli *et al.* (2009a) are given as contrast values and are saturated at the same level  $[-0.02, 0.02]$ . The image processed by Chatterjee, Banerjee, and Ravindra (2016) is in arbitrary units, while the one by Priyal *et al.* (2014) was provided as a JPG image. Therefore, these images are saturated such that the plage regions visually appear similar to the saturated image with our method.