A&A 667, A167 (2022)
https://doi.org/10.1051/0004-6361/202244913

Astronomy & Astrophysics

Scrutinising the relationship between plage areas and sunspot areas and numbers

Theodosios Chatzistergos[1,2], Ilaria Ermolli[2], Natalie A. Krivova[1], Teresa Barata[3], Sara Carvalho[4], and Jean-Marie Malherbe[5,6]

[1] Max Planck Institute for Solar System Research, Justus-von-Liebig-weg 3, 37077 Göttingen, Germany
e-mail: chatzistergos@mps.mpg.de
[2] INAF – Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monte Porzio Catone, Italy
[3] Univ. Coimbra, Instituto de Astrofísica e Ciências do Espaço, Department of Earth Sciences, 3030-790 Coimbra, Portugal
[4] Univ. Coimbra, Instituto de Astrofísica e Ciências do Espaço, 3040-004 Coimbra, Portugal
[5] LESIA, Observatoire de Paris, 92195 Meudon, France
[6] PSL Research University, 75006 Paris, France

Received 7 September 2022 / Accepted 14 September 2022

ABSTRACT

Context. Studies and reconstructions of past solar activity require data on all magnetic regions on the surface of the Sun (i.e. on dark sunspots as well as bright faculae–plage and network). Such data are also important for understanding the magnetic activity and variability of the Sun and Sun-like stars. The longest available direct faculae–plage datasets are white-light facular and Ca II K observations going back to 1874 and 1892, respectively. Prior to that time the only direct data available are for sunspots.
Aims. We reassess the relationship between plage areas and sunspot records (areas and numbers) since 1892, to allow reconstructions of facular–plage areas which can be employed for studies going further back in time to the period when solely sunspot observations are available.
Methods. We use the plage areas derived from 38 consistently processed Ca II K archives as well as the plage area composite based on these archives. The considered archives include both the well-known observatories (e.g. Coimbra, Kodaikanal, Meudon, Mt Wilson), and the less explored observatories (e.g. Kharkiv, Mees, and Upice). These data allow us to study the relationship between plage area and sunspot records (areas and number) over a period of 12 solar cycles and for different bandpasses.
Results. We find the relationship between plage and sunspot areas to be well represented by a power-law function; the relationship between the plage areas and the sunspot number is also best fit with a power-law function. We further find that the relationships depend on the bandwidth and the solar cycle strength. The reconstructions with a power-law relationship are in good agreement with the original plage area series, whereas employment of a cycle-strength-dependent relationship improves the reconstructions only marginally. We also estimate the error in the plage areas reconstructed from the sunspot areas or numbers. Performing the same analysis on other previously published plage area series, usually derived from a single archive with diverse processing techniques, returns different results when using different time series. This highlights the importance of applying a consistent processing to the various archives and demonstrates the uncertainties introduced by using previously published series for studies of past solar activity, including irradiance reconstructions.
Conclusions. Our results have implications for past solar activity and irradiance reconstructions and for stellar activity studies, which sometimes assume a linear dependence between plage and sunspot areas.

Key words. Sun: activity – Sun: chromosphere – Sun: faculae, plages – Sun: photosphere – sunspots

1. Introduction

Sunspots and faculae are the most prominent manifestations of solar surface magnetism (Solanki et al. 2006). Sunspots are relatively dark and cool areas on the surface of the Sun, whereas faculae are comparatively small and patchy bright regions usually seen in the vicinity of sunspots and in remnants of active magnetic regions. Faculae were originally discovered in full-disc white light images, where they are seen mostly near the limb. The co-spatial chromospheric features observed in the Ca II K spectral range, called plage, are bright and easily observable over the whole solar disc (Solanki 1993). The connection of facular regions to strong magnetic fields was first pointed out by Babcock & Babcock (1955). Being a manifestation of the same physical process, the concentration of magnetic field at the solar surface, sunspots and faculae are closely associated (Solanki 2003). However, their evolution differs. While individual facular elements generally do not live as long as sunspots, within a given active region facular regions can last much longer than sunspots, partly because sunspot decay products form new facular elements (e.g. Wang et al. 2012).

Sunspots have been observed on the solar disc since antiquity (Vaquero & Vázquez 2009; Arlt & Vaquero 2020). However, regular monitoring of sunspots began only with the advent of the telescope in the early 17th century. Later, Wolf (1850) started compiling sunspot measurements by various observers, and created the first sunspot number series, the Wolf number, later renamed the international sunspot number series (ISNv1), which continues to be produced to this day (Clette et al. 2007). This series goes back to 1818 with daily cadence, while annual values are available back to 1700. More recently, Hoyt & Schatten (1998, HoSc98 hereafter) introduced a different measure of

Table 1. Previous studies of the relationship between sunspot and facular measurements.

Study	Plage data	Sunspot data	Period	Relation		
				Annual	Monthly	Daily
Ca II K plage areas and sunspot number series						
Dorotović et al. (2010)	Co	ISNv1	1996–2006			Quadratic [a]
Foukal (1996)	MW	ISNv1	1915–1984	Linear		
Fligge & Solanki (1998)	MW [b]	ISNv1	1944–1954			Quadratic
Kuriyan et al. (1982)	Ko	ISNv1	1905–1977	Linear		
Bertello et al. (2016)	Ko [c]	ISNv2	1907–1999	Linear	Linear	
Priyal et al. (2017)	Ko	ISNv2	1907–2007		Linear	
Singh et al. (2021)	Ko	ISNv2	1905–2007		Linear	Linear
Singh et al. (2022)	MW	ISNv2	1915–1985		Linear	Linear
Singh et al. (2022)	ML	ISNv2	1998–2015		Linear	Linear
Yeo et al. (2020)	SP [d]	ISNv1, ISNv2	1976–2017			Power law [e]
Yeo et al. (2020)	SP [d]	CEA17, HoSc98	1976–2010			Power law [e]
Ca II K plage areas and sunspot area series						
Foukal & Vernazza (1979)	SGD	SGD	1969–1974			(∼0.5–685) [f]
Schatten et al. (1985)	SGD	SGD	1969–1982		(∼18–40)	
Lawrence (1987b)	SGD	SGD	1974–1985	(∼5–30)		
Foukal (1993)	SGD	RGO	1954–1987	Linear		
Foukal (1996)	MW	RGO	1915–1984	Linear		
Foukal (1998)	MW	RGO	1915–1984	Linear		Quadratic
Tlatov et al. (2009)	MW	RGO	1944–1963		Quadratic	
Tlatov et al. (2009)	Ko	RGO	1944–1963		Quadratic	
Mandal et al. (2017)	Ko [g]	RGO	1907–1965	Linear		
Singh et al. (2021)	Ko	RGO	1905–2007		Linear	Linear
Singh et al. (2022)	MW	RGO	1915–1985		Linear	Linear
Singh et al. (2022)	ML	RGO	1998–2015		Linear	Linear
Chowdhury et al. (2022)	Ko [g]	RGO	1907–1980	Quadratic		
Chapman et al. (1997)	SF1, SF2	SF1, SF2	1988–1995	Quadratic [h]	Quadratic	
Chapman et al. (2001)	SF1, SF2	SF1, SF2	1988–1999		(∼5–40)	
Chapman et al. (2011)	SF1, SF2	SF1, SF2	1988–2009	Linear	Linear	
Steinegger et al. (1996)	SP	VTT [i]	1980			(∼3–30)
Yeo et al. (2020)	SP [d]	BEA09 [j]	1976–2017			power law [e]
Chatzistergos et al. (2020b)	SP	BEA09 [j]	1960–2002	(∼35–55)		
Chatzistergos et al. (2020b)	Ky [k]	BEA09 [j]	1928–1969	(∼13–20)		
Magnetogram facular areas and sunspot area/number series						
Solanki & Unruh (2013)	KP [l]	KP areas	1974–2002			Quadratic
Shapiro et al. (2014)	KP [l]	KP areas	1974–2002			Quadratic
Borgniet et al. (2015)	MDI [m]	SOON [n] areas	1996–2007		(∼10–160)	
Criscuoli (2016)	MDI [m]	ISNv2	1996–2011		Linear	Quadratic

Notes. Columns are: Bibliography entry; Type of plage and sunspot data used; Period covered by the analysed data; Form of the derived relationship (considering sunspot data as a function of plage areas) for annual, monthly, and daily data (values in parentheses are reported values for the ratio of plage to sunspot area). The column corresponding to monthly values is for periods intermediate to annual and daily. Only Tlatov et al. (2009), Bertello et al. (2016), and Singh et al. (2021) used monthly values; Chapman et al. (1997, 2001) and Chapman et al. (2011) used 100-day bins, Shapiro et al. (2014) 58-day bins, and Criscuoli (2016) and Priyal et al. (2017) 6-month averages. See Tables 2 and 3, and the footnotes in this table for information on the abbreviations. [a]Not explicitly stated by the authors, but can be inferred from Fig. 6A in Dorotović et al. (2010). [b]The MW plage area series used in this study was produced by Foukal (1996). [c]The Ko plage area series used in this study was produced by Tlatov et al. (2009). [d]This is the Bertello et al. (2016) disc-integrated 1Å Ca II K composite index with SP and Synoptic Optical Long-term Investigations of the Sun Integrated Sunlight Spectrometer data. The SP data are not the full-disc spectroheliograms used in our study. [e]This is a convolution of a power-law relation with a finite impulse response filter. [f]This was roughly estimated from Fig. 4 in Foukal & Vernazza (1979), who argued that there is large scatter in the data rendering sunspot areas poor indicators of plage areas. [g]The Ko plage area series used in this study was produced by Chatterjee et al. (2016). [h]The authors reported a quadratic relation for annual values; however, they also state that roughly 75% of the variance in the scatter plots is accounted for by the linear term. [i]Refers to data acquired with the Vacuum Tower Telescope at the Teide observatory. [j]Refers to the Balmaceda et al. (2009) sunspot area composite series. [k]Refers to the plage areas from the Kyoto observatory. [l]Refers to facular areas from Kitt Peak magnetograms. [m]Refers to facular areas from the Solar and Heliospheric Observatory Michelson Doppler Imager magnetograms. [n]Refers to the Solar Optical Observing Network.

solar activity based on sunspot counts, namely the number of sunspot groups or group sunspot number (GSN). This enabled the early sunspot observations performed between 1609 and 1700 to be exploited as well. In the last few years, the database of sunspot data has been scrutinised and updated to include more data (e.g. Vaquero et al. 2016; Carrasco et al. 2018, 2020, 2021;

Table 2. Abbreviations used for Ca II K plage datasets.

Dataset	Acronym	Period	Reference
Individual series			
Arcetri	Ar	1931–1974	Ermolli et al. (2009a),
Big Bear	BB	1982–2006	Naqvi et al. (2010)
Coimbra	Co	1925–2019	Lourenço et al. (2019)
Kharkiv	Kh	1952–2019	Belkina et al. (1996)
Kodaikanal	Ko	1904–2007	Priyal et al. (2017)
Mauna Loa [a]	ML	1998–2015	Rast et al. (2008)
McMath-Hulbert	MM	1948–1979	Mohler & Dodson (1968)
Mees	MS	1982–1998	
Meudon	MD	1893–2019	Malherbe & Dalmasse (2019)
Mt Wilson	MW	1915–1985	Lefebvre et al. (2005)
Rome [b]	RP	1996–2019	Ermolli et al. (2007)
Sacramento Peak	SP	1960–2002	Tlatov et al. (2009)
San Fernando CFDT1 [c]	SF1	1988–2015	Chapman et al. (1997)
San Fernando CFDT2 [d]	SF2	1992–2013	Chapman et al. (1997)
Upice	UP	1998–2019	Klimeš et al. (1999)
Composite series			
Solar Geophysical Data series	SGD	1942–1987	
Plage area composite	CEA20	1892–2019	Chatzistergos et al. (2020c)

Notes. The Ca II K plage area series are separated into those resulting from individual archives and composite series. [a]Taken with the Mauna Loa Precision Solar Photometric Telescope, PSPT. [b]Taken with the Rome Precision Solar Photometric Telescope, PSPT. [c]Taken with the Cartesian Full-Disc Telescope, CFDT1. [d]Taken with the Cartesian Full-Disc Telescope, CFDT2.

Hayakawa et al. 2020a,b, 2021; Vokhmyanin et al. 2020). The method of how to include individual sunspot series has also been a matter of debate (Lockwood et al. 2016; Usoskin et al. 2016a). This led to the release of a new version of the international sunspot number, ISNv2 (Clette & Lefèvre 2016), and a number of alternative GSN series (e.g. Lockwood et al. 2014; Cliver & Ling 2016; Svalgaard & Schatten 2016; Usoskin et al. 2016b, 2021; Chatzistergos et al. 2017; Willamo et al. 2017). In addition to sunspot and sunspot group numbers, the areas of sunspots have also been recorded since the early telescopic observations (Arlt et al. 2013; Arlt & Vaquero 2020), albeit considerably less systematically than sunspot numbers, which makes cross-calibration of the individual records extremely challenging. Therefore, the earliest sunspot area measurements typically employed in solar activity and irradiance studies are those from the Royal Greenwich Observatory (RGO) dating back to 1874.

Observations of bright faculae have also been reported since the advent of telescopes (Vaquero & Vázquez 2009). However, due to their low contrast when observed in the continuum, observations of facular regions had long been rather episodic, with regular observations over a long time span potentially carrying information on such regions going back to the late 19th century. More detailed information on facular regions is provided by Ca II K observations, which have been performed at many places around the globe since 1892 (see Chatzistergos 2017; Chatzistergos et al. 2020c, and references therein), and thus provide a very good temporal coverage of the entire 20th century. These observations sample the lower chromosphere where faculae are seen as bright plage regions. Thus, an accurate measurement of plage properties in these observations is of primary interest for reconstructions of past solar magnetism (Chatzistergos et al. 2019d).

Various studies require information about both sunspots and faculae, such as irradiance reconstructions (e.g. Foukal & Lean 1990; Solanki & Fligge 1998; Fligge & Solanki 2000; Krivova et al. 2010; Dasi-Espuig et al. 2016; Wu et al. 2018; Lean 2018), reconstructions of the long-term evolution of the solar magnetic field (e.g. Solanki et al. 2000, 2002; Cameron et al. 2010; Jiang et al. 2011, 2013; Nagovitsyn et al. 2016; Bhowmik & Nandy 2018; Krivova et al. 2021), or analyses of large-scale flow patterns (Rincon & Rieutord 2018, and references therein). However, even if the Ca II K data are used directly, they cover only about 130 years, compared with 400 years of sunspot number data. The Ca II K data series are too short for many purposes. For example, to get a good idea of solar influence on climate, it is important to have solar irradiance reconstructions going back to the Maunder minimum, that is until at least 1700 CE. To do this, it is necessary to reconstruct the plage area starting from sunspot data. A first step in that

Table 3. Acronyms used in this work to refer to various sunspot datasets, listed separately for the sunspot area, ISN, and GSN series.

Data series	Acronym	Period
Sunspot areas		
Royal Greenwich Observatory	RGO	1874–1976
Mandal et al. (2020a) [a]	MEA20	1874–2019
ISN		
Clette et al. (2007) [b]	ISNv1	1700–2014
Clette & Lefèvre (2016) [b]	ISNv2	1700–2019
GSN		
Svalgaard & Schatten (2016) [b]	SvSc16	1610–2015
Chatzistergos et al. (2017) [a]	CEA17	1739–2010

Notes. [a]The series is available at `http://www2.mps.mpg.de/projects/sun-climate/data.html` [b]Available at `https://www.sidc.be/silso/datafiles`

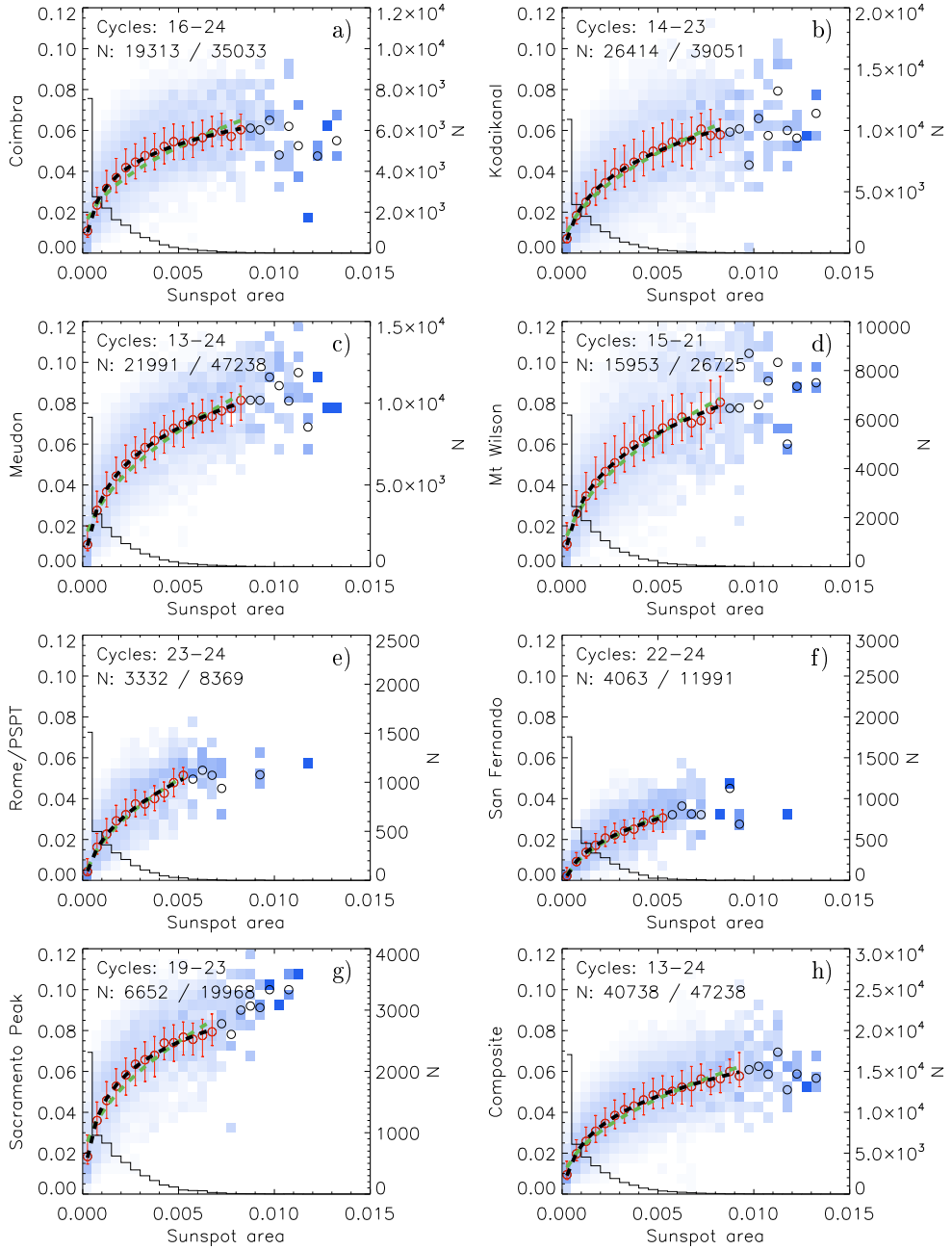

Fig. 1. Probability distribution functions (PDFs) for plage fractional areas as a function of the sunspot fractional areas from MEA20. Both plage and spot areas were taken on the same days. The PDFs are shown in bins of 0.005 and 0.0005 fractional areas for plage and sunspots, respectively. Each panel shows the PDF matrix for a given Ca II K archive; Co (*a*)), Ko (*b*)), MD (*c*)), MW (*d*)), RP (*e*)), SF2 (*f*)), SP (*g*)), and the CEA20 plage area composite (*h*)). The PDFs are colour-coded between white for 0 and bright blue for 1. Circles denote the average plage area value within each column. For columns with less than 20 days of data these circles are shown in black, otherwise in red. Also shown is the asymmetric 1σ interval for each column. Two different fits to the average values are overplotted: a power-law fit (black) and a square-root function (green). The number of days included in each column is also shown as a solid black histogram (see right-hand axis). The period of overlap between the two archives shown in a given panel (expressed in solar cycles), as well as the total number of overlapping days used to construct each matrix (N) and the total number of days within this time interval are listed in the top part of each panel.

direction is to obtain a reliable relationship between faculae and sunspots spanning more than the last couple of sunspot cycles.

Multiple studies have endeavoured to determine the relationship between faculae and sunspots (e.g. Foukal & Vernazza 1979; Foukal 1996, 1998; Kuriyan et al. 1982; Lawrence 1987b; Chapman et al. 1997; Tlatov et al. 2009; Solanki & Unruh 2013; Shapiro et al. 2014; Bertello et al. 2016; Criscuoli 2016; Yeo et al. 2020; de Paula & Curto 2020; Berrilli et al. 2020; Nèmec et al. 2022) using different data and methodologies. Table 1 summarises the findings of some previous works, emphasising the analyses that used Ca II K plage areas as the facular data. The majority of the previous studies supported a linear and a quadratic relation (considering sunspot data as a function of plage areas) when annual and daily areas were considered, respectively. Lawrence (1987b,a) also found that the ratio of the annual plage to sunspot areas varies over the solar cycle (SC) being lowest during SC maxima, hence hinting at a non-linear relation. Furthermore, Lawrence (1987b),

Table 4. Best fit parameters between the daily values of CEA20 plage area composite and various sunspot series.

	$p_a = a_1 + a_2 \times s^{a_3}$			RSS/d.o.f.
	a_1	a_2	a_3	
		Power law		
MEA20	−0.004 ± 0.008	0.342 ± 0.093	0.35 ± 0.08	0.066
ISNv1	0.002 ± 0.002	0.009 ± 0.001	0.66 ± 0.05	0.116
ISNv2	0.002 ± 0.002	0.009 ± 0.001	0.69 ± 0.04	0.074
CEA17	0.004 ± 0.001	0.005 ± 0.001	0.86 ± 0.06	0.078
		Squared root		
MEA20	0.006 ± 0.002	0.606 ± 0.028	0.5	0.110
		Linear		
MEA20	0.020 ± 0.001	5.476 ± 0.261	1	0.604
ISNv1	0.011 ± 0.001	0.0027 ± 0.0001	1	0.349
ISNv2	0.009 ± 0.001	0.0033 ± 0.0001	1	0.304
CEA17	0.007 ± 0.001	0.0034 ± 0.0001	1	0.114

Notes. RSS/d.o.f. are the sum of the squared residuals per degree of freedom. The ISNv1 and ISNv2 sunspot number values were divided by 12.08 (ISNv2 values are also multiplied by 0.6) prior to the fit to bring them roughly to the same level as the group sunspot number series.

Chapman et al. (1997), and Bertello et al. (2016) reported that the relation depends on the SC strength. Most of these studies used plage areas derived from uncalibrated historical Ca II K data (except for Chapman et al. 1997, 2001, 2011, who used CCD-based data) and mainly either data from a single archive or simply appending data from different observatories without cross-calibrating them, for example in the Solar Geophysical data series (SGD, hereafter; see Sect. 2). At the same time, relations derived from historical archives carry significant uncertainties due to intrinsic inconsistencies of the various datasets and their processing (see Sect. 3.3). Therefore, any study based on these data (e.g. reconstruction of solar irradiance variations) is also affected by the intrinsic inconsistency of the various archives. These limitations can be overcome by exploiting the results of the analysis of Ca II K observations by Chatzistergos et al. (2018b, 2019b, 2020c).

Here we use the consistent record of plage areas derived by Chatzistergos et al. (2020c) from the analysis of 38 Ca II K archives together with several available series of sunspot observations to study the relationship between plage areas and sunspot data. We examine the functional form of the relationship for various sunspot datasets, and also analyse the dependence of this relationship on the activity level and the bandwidth of the Ca II K observations.

In Sect. 2 we give an overview of the various full-disc Ca II K observations, and the published plage and sunspot time series considered in this study. In Sect. 3 we study the relationship between the plage areas and the various sunspot records. We then use these relationships to reconstruct plage time series based on sunspot observations, which we test by comparing with the actual plage datasets. We also discuss our findings and compare them to results presented in the literature. Finally, we summarise our results and draw our conclusions in Sect. 4.

2. Data

2.1. Ca II K plage series

We use the time series of projected plage areas produced and made available by Chatzistergos et al. (2020c). These plage area series were produced after accurately processing over 290 000 images from 43 historical and modern Ca II K archives (Chatzistergos et al. 2016, 2018a, 2019a,b,c, 2020a,b,c) spanning the period 1892–2019. A composite of plage areas was produced by combining results from 38 archives (Chatzistergos et al. 2020c,d, hereafter CEA20)[1] covering the period 1892–2019. The composite includes separately projected and corrected areas for foreshortening; however, here we used only the projected areas. The five archives considered by Chatzistergos et al. (2020c), but finally not used for the composite, included observations taken off-band. While we actually considered all individual series from Chatzistergos et al. (2020c), some of them are quite short and their analysis does not fortify the results or conclusions presented in the following. Therefore, for the sake of simplicity and brevity, we do not show the results for these series here. In the following we present results only for the 15 archives summarised in Table 2, where we also give the acronyms used for them in this paper. We refer the reader to Table 1 in Chatzistergos et al. (2020c) for the main characteristics of these archives. One exception is the plage area record from Meudon (Malherbe et al. 2022). Instead of using the series by Chatzistergos et al. (2020c), we produced a new plage area series by applying exactly the same process as Chatzistergos et al. (2020c). This was done to include newly digitised data covering mainly the period 1948–1961. The new dataset includes 5865 new images, which were missing in the series produced by Chatzistergos et al. (2020c).

Details of the processing procedure and the derivation of plage areas can be found in Chatzistergos et al. (2018b, 2019b, 2020c). Briefly, images from the photographic archives were photometrically calibrated and compensated for limb-darkening as described by Chatzistergos et al. (2018b). Images taken with a CCD were also processed to compensate for the limb-darkening with the same method. All images were segmented to identify plage areas with a multiplicative factor to the standard deviation of the quiet Sun intensity values (Chatzistergos et al. 2019b). Chatzistergos et al. (2018b, 2019b) showed that the developed method significantly reduces errors in the estimates of plage areas from the various historical Ca II K archives.

We also considered the previously published Ca II K plage area series by Kuriyan et al. (1983), Foukal (1996, 2002), Ermolli et al. (2009b), Tlatov et al. (2009), Chatterjee et al. (2016),

[1] The series is available at http://www2.mps.mpg.de/projects/sun-climate/data.html

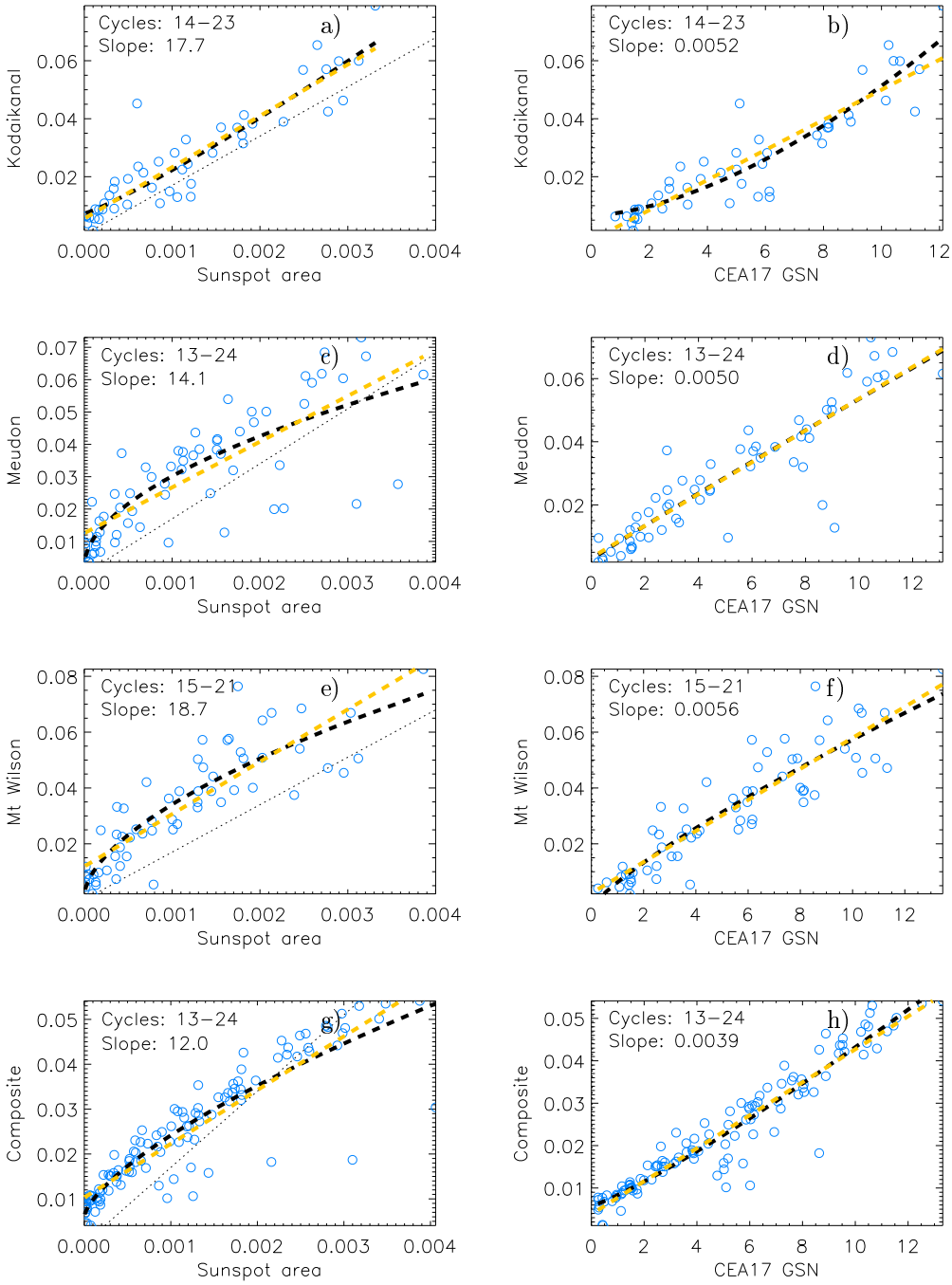

Fig. 2. Scatter plots between annual plage fractional areas and the annual MEA20 sunspot fractional areas (*left columns*) and the annual CEA17 GSN series (*right columns*). Two different fits are overplotted, a power-law (black) and a linear (yellow) fit. The dotted line has a slope of 17, which corresponds to the mean ratio between the CEA20 plage area series and the MEA19 sunspot area series (see Sect. 3.1). The period of overlap between the two archives (expressed in solar cycles) as well as the slope of the linear fit are shown in each panel.

Priyal et al. (2017), Singh et al. (2018), SGD[2], as well as the Ca II K emission index by Bertello et al. (2010).

The processing and data selection vary considerably among the above published series (see Chatzistergos et al. 2019b, for more details). For example, the series by Kuriyan et al. (1983) was created by manually selecting the plage regions from the actual Ko photographs. Similarly, the SGD series includes areas determined manually from the physical photographs of the MM (06/1942–09/1979), MW (10/1979–09/1981), and BB (10/1981–11/1987) observatories. Foukal (1996) manually selected plage areas from an earlier version of the MW dataset digitised with an 8bit device. Foukal (2002) derived plage (including the enhanced network) areas from the 8bit MW dataset and extended it with SP data up to 1999. Ermolli et al. (2009b) processed the 8bit Ko and 16bit MW datasets. Tlatov et al. (2009) processed the 8bit Ko data and the 16bit MW data. Bertello et al. (2010) analysed the 16bit MW data and derived a Ca II K index, which they found to be related to the plage areas by Tlatov et al. (2009). Chatterjee et al. (2016), Priyal et al. (2017), and Singh et al. (2018) processed the 16bit digitised Ko archive. Ermolli et al. (2009b), Tlatov et al. (2009), Priyal et al. (2017),

[2] Available at https://www.ngdc.noaa.gov/stp/solar/calciumplages.html

Table 5. Best fit parameters between the annual median values of CEA20 plage area composite and annual median values of various sunspot series.

	$p_a = a_1 + a_2 \times s^{a_3}$			RSS/d.o.f.
	a_1	a_2	a_3	
		Power law		
MEA20	0.007 ± 0.001	0.0176 ± 0.0013	0.78 ± 0.06	0.68
ISNv1	0.005 ± 0.001	0.0049 ± 0.0009	0.87 ± 0.06	0.58
ISNv2	0.005 ± 0.001	0.0048 ± 0.0008	0.94 ± 0.07	0.69
CEA17	0.005 ± 0.001	0.0029 ± 0.0007	1.12 ± 0.10	0.66
SvSc16	0.005 ± 0.001	0.0038 ± 0.0009	1.03 ± 0.09	0.67
		Squared root		
MEA20	0.0018 ± 0.0008	0.0247 ± 0.0008	0.5	0.96
		Linear		
MEA20	0.0090 ± 0.0006	0.0137 ± 0.0004	1	0.77
ISNv1	0.000 ± 0.000	0.0070 ± 0.0007	1	0.61
ISNv2	0.000 ± 0.000	0.0062 ± 0.0006	1	0.70
CEA17	0.000 ± 0.000	0.0040 ± 0.0007	1	0.67
SvSc16	0.000 ± 0.000	0.0044 ± 0.0007	1	0.67

Notes. RSS/d.o.f. are the sum of the squared residuals per degree of freedom. The ISNv1 and ISNv2 sunspot number values were divided by 12.08 (ISNv2 values are also multiplied by 0.6) prior to the fit to bring them roughly to the same level as the group sunspot number series.

and Singh et al. (2018) applied simple photometric calibration to the data, while SGD, Foukal (1996, 2002), Bertello et al. (2010), and Chatterjee et al. (2016) used photometrically uncalibrated observations. All the above series are given at a daily cadence except those by Kuriyan et al. (1983), Tlatov et al. (2009), and Bertello et al. (2010), which are available as annual values, and the series by Foukal (2002), for which 12-month running mean values are provided. In this work we use the series by Bertello et al. (2010) after applying a linear scaling to match our derived plage area series from the 16bit MW data. We note that all these series include areas corrected for foreshortening in fractions of the hemisphere, while the ones by Chatterjee et al. (2016), Priyal et al. (2017), and Singh et al. (2018) include projected areas.

2.2. Sunspot series

For our study we used the sunspot area record compiled by Mandal et al. (2020b, MEA20, hereafter)[1]. We used the projected areas from this dataset as we did for the plage areas. This series is a composite of the RGO (1874–1976), Debrecen (1976–2019), and Kislovodsk (1977–2019) sunspot area records.

We also used the various available sunspot number and group number series. These series and their acronyms used here are summarised in Table 3. Briefly, ISNv1, ISNv2, and SvSc16 used simple linear scaling to calibrate the records by the various observers. CEA17 used a non-linear and non-parametric approach to cross-calibrate the counts by different observers, thus taking their diverse observing capabilities into account.

We note that the values from ISNv1 and ISNv2 were divided by 12.08 and 20.13 (12.08/0.6), respectively, to bring them to the level of the GSN series. Owing to the corrections introduced to ISNv1 to produce ISNv2 (see e.g. Clette & Lefèvre 2018), this scaling results in a small discontinuity between ISNv1 and ISNv2 over 1947 (Clette & Lefèvre 2016), with ISNv2 being consistently lower than ISNv1 afterwards. We further note that the series by SvSc16 has annual cadence, while the other series have daily values.

3. Results

3.1. Relationship between plage and sunspot series

We first analysed the relationship between the daily plage areas derived from the analysis of each individual Ca II K archive and the sunspot area series from MEA20 by considering their entire respective overlap periods. Following Chatzistergos et al. (2017), we did this by using probability distribution function (PDF) matrices between the daily plage and sunspot areas. Briefly, for a given archive, the PDF matrices are created by using another archive as the reference. For this, we first sort the values of the considered archive in bins of specific width. For each bin, we compute the histogram of the co-temporal observations from the reference archive. Each histogram is then normalised to the total number of data points included, thus resulting in a PDF. Examples of such PDF matrices when considering the Ca II K series as the references (ordinate) are shown in Fig. 1. We note that the PDF matrices were computed with bin sizes of 10^{-3}, 10^{-4}, and 0.1, for plage area, sunspot area, and sunspot number series, respectively; however, to aid the visualisation of the PDF matrices in the following they are shown with bins of 5×10^{-3}, 5×10^{-4}, and 1. We then fit the plage area values averaged over every sunspot-area bin with a square-root and a power-law function. Since our purpose is to use sunspot data to reconstruct plage areas, we expressed plage areas as a function of sunspot data, which is the opposite of how previous results are listed in Table 1. The results are qualitatively similar for all archives, and the overall shape of the relationship is in general in agreement with previous studies. However, we find that the power law fits the data even better than the square-root function reported in previous studies, although at the cost of an additional free parameter (the exponent). This also holds when the sunspot areas are taken as reference. Considering the CEA20 plage area composite and the MEA20 sunspot area record, we find the relationship between them to be

$$p_a = (0.34 \pm 0.09) \times s_a^{(0.35 \pm 0.08)} - (0.004 \pm 0.008), \quad (1)$$

where p_a and s_a are the plage and sunspot areas, respectively.

Table 4 lists the best fit parameters for the fits between the CEA20 plage area and MEA20 sunspot area composite series

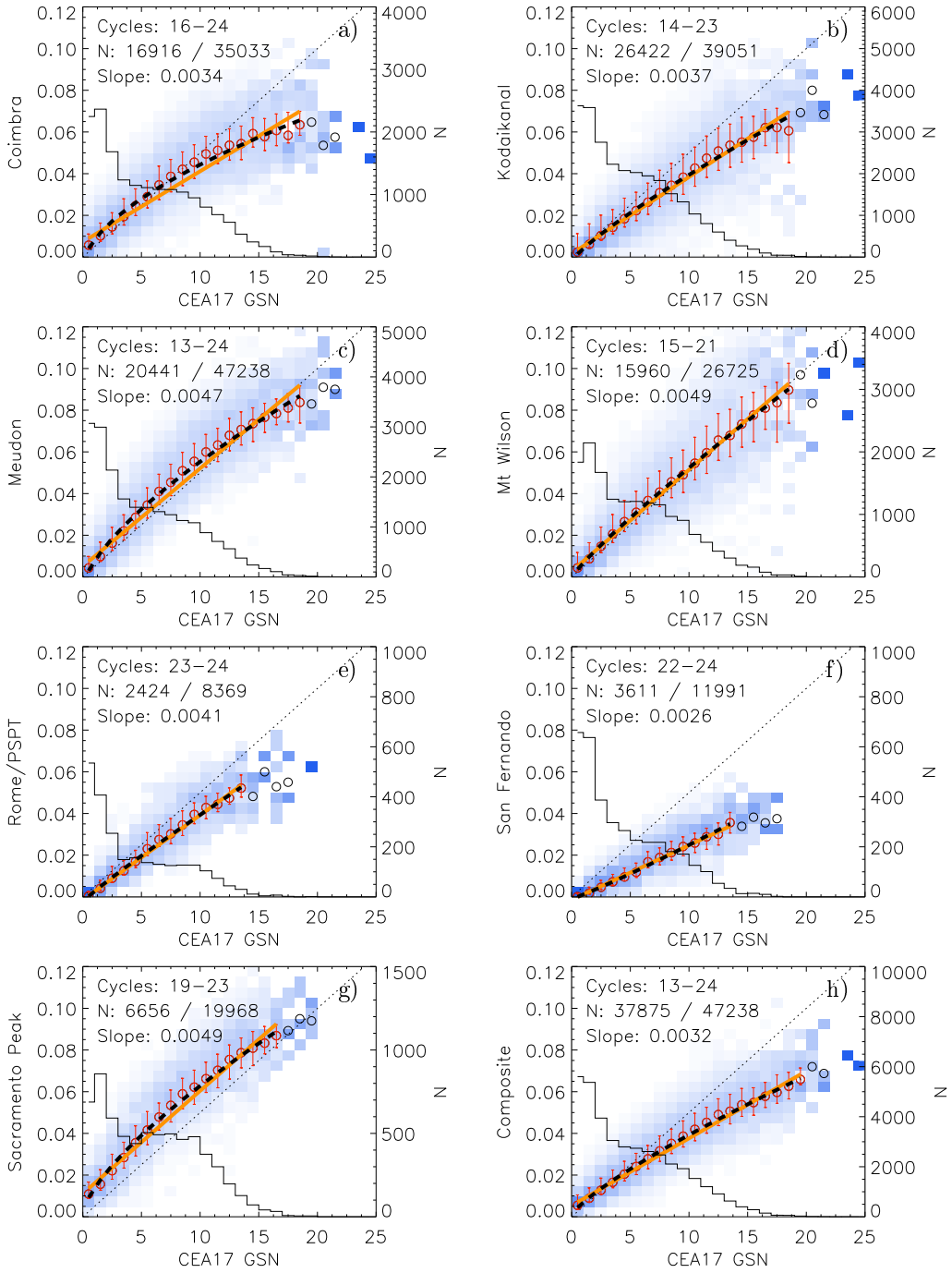

Fig. 3. Same as Fig. 1, but for the CEA17 GSN series. A linear fit (orange) and a power-law fit (black dashed) were performed to the mean values of each column. The slope of the linear fit is also listed in each panel. The black dotted line has a slope of 0.005.

along with the sums of squared residuals per degree of freedom (RSS/d.o.f.). Interestingly, even when considering annual values (seen in Fig. 2), we see a clear tendency for a non-linearity in the relationship, although there is a considerable scatter for some series. The parameters of the fits for the annual values of the CEA20 plage area composite and the MEA20 sunspot area series are listed in Table 5.

These results are in agreement with those by Chapman et al. (1997), but in contrast to those of Foukal (1996, 1998) and Chapman et al. (2011), who reported a linear relationship between sunspot and plage areas when annual values were considered.

Figure 3 shows the relationship between the various Ca II K plage area series and the CEA17 GSN series. To a good first approximation, the relation can be considered linear for all archives, although some archives hint at a weak non-linearity (see e.g. the results for SP, MD, or Co data). The parameters of the power-law fit between the CEA20 plage area composite and the various sunspot series are given in Table 4 along with the parameters of a linear fit and the resulting RSS/d.o.f. For the CEA20 plage area composite and the CEA17 GSN series, we find that the power-law function fits the data better than the linear relation.

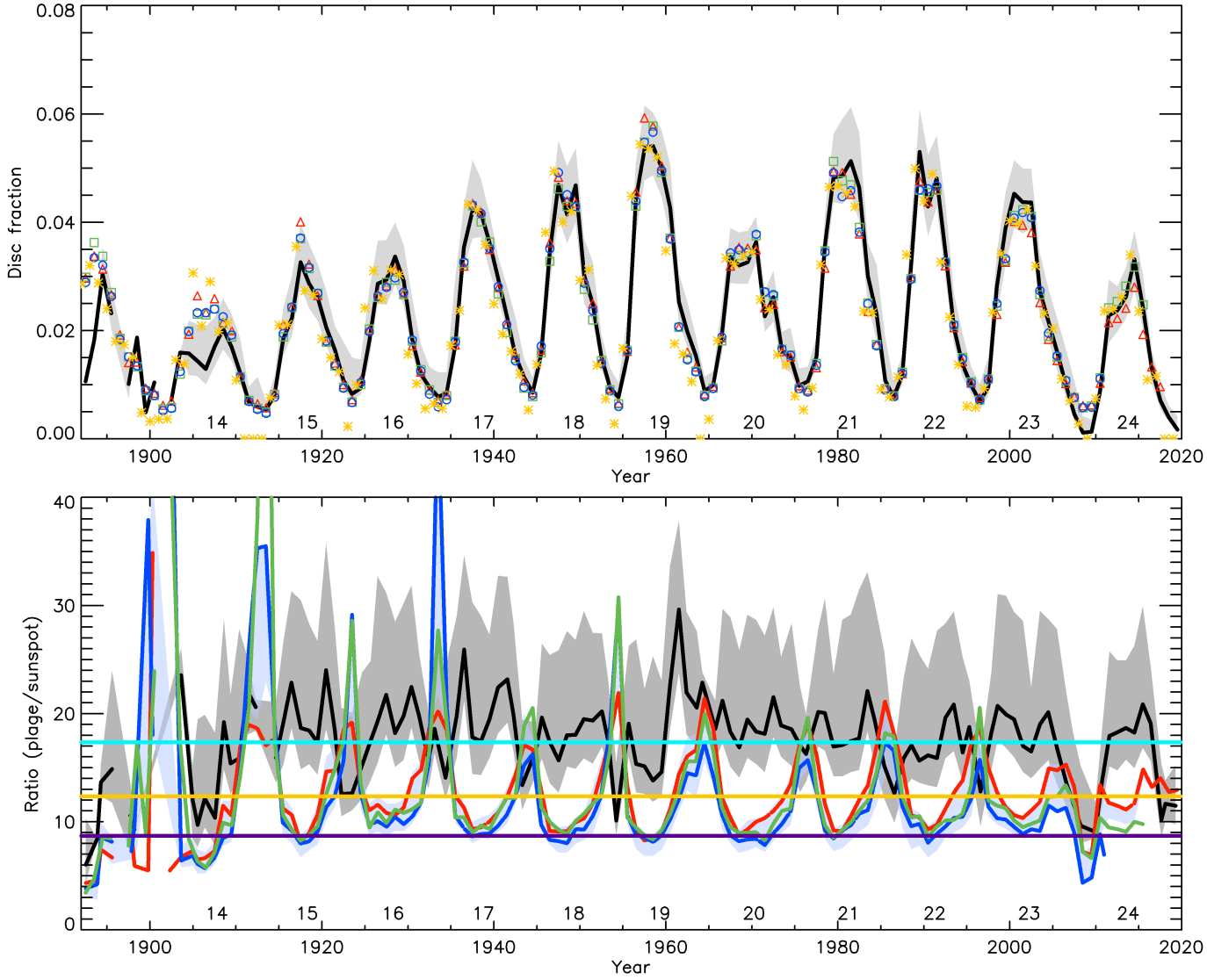

Fig. 4. *Upper panel*: CEA20 plage area composite (black line) along with scaled sunspot series by MEA20 (yellow asterisks), CEA17 (blue circles), ISNv2 (red triangles), and SvSc16 (green squares). All series are scaled linearly to the plage areas except MEA20, which is scaled with a square-root function. The solid lines are annual median values, while the grey shaded surface is the asymmetric 1σ interval. *Lower panel*: ratio of the CEA20 plage areas composite to the sunspot area series by MEA20 (black) and sunspot number series by CEA17 (blue), SvSc16 (green), and ISNv2 (red) as a function of time. The ratios to the GSN series have been multiplied by 2000 to allow a direct comparison with the ratios to the sunspot areas. The blue shaded surface gives the range of annual values for the plage areas to sunspot number ratios from all sunspot number series used in this study. The ratios are calculated only for the days on which sunspot areas are greater than 0.0005 of the disc area or the sunspot number is greater than 0. The yellow (light blue) horizontal line indicates the mean ratio of plage areas to sunspot numbers (sunspot areas), while the purple horizontal line is the mean ratio for the plage areas to sunspot number series over four-year intervals around cycle maxima. The numbers below the curves in each panel denote the conventional SC numbering and are placed at times of SC maxima.

Table 4 shows that the relation is only mildly affected by the choice of the sunspot series. For the CEA17 GSN series we adopt the relationship

$$p_a = (0.005 \pm 0.001) \times s_{gn}^{(0.86\pm0.06)} + (0.004 \pm 0.001), \quad (2)$$

where s_{gn} is the group count.

The relation is, to a good approximation, linear for the annual values (see Fig. 2h). However, the slope of the fit to the annual values tends to be slightly higher than that to the daily values. This is most likely due to the greater lifetime of plage ensemble regions compared to sunspots. The parameters of the fits for the annual values of the CEA20 plage area composite and the various sunspot series are listed in Table 5. This is in agreement with the results by Kuriyan et al. (1982), Foukal (1996), and Bertello et al. (2016), who found a linear relation between annual ISNv1 and Ca II K plage areas. We note that the shape of the relationships we find for plage-sunspot areas and plage-sunspot number series imply a non-linear relation between sunspot areas and sunspot number series, which is in agreement with previous results (e.g. Fligge & Solanki 1997; Balmaceda et al. 2009; Carrasco et al. 2016; Nagovitsyn & Osipova 2021; Mandal et al. 2021).

Figure 4 (top panel) shows the annual values of the plage area composite along with the scaled series by MEA20, CEA17, SvSc16, and ISNv2. We find good agreement between all the series and the CEA20 plage area composite. An exception is the period before 1905; however, the coverage of the Ca II K plage

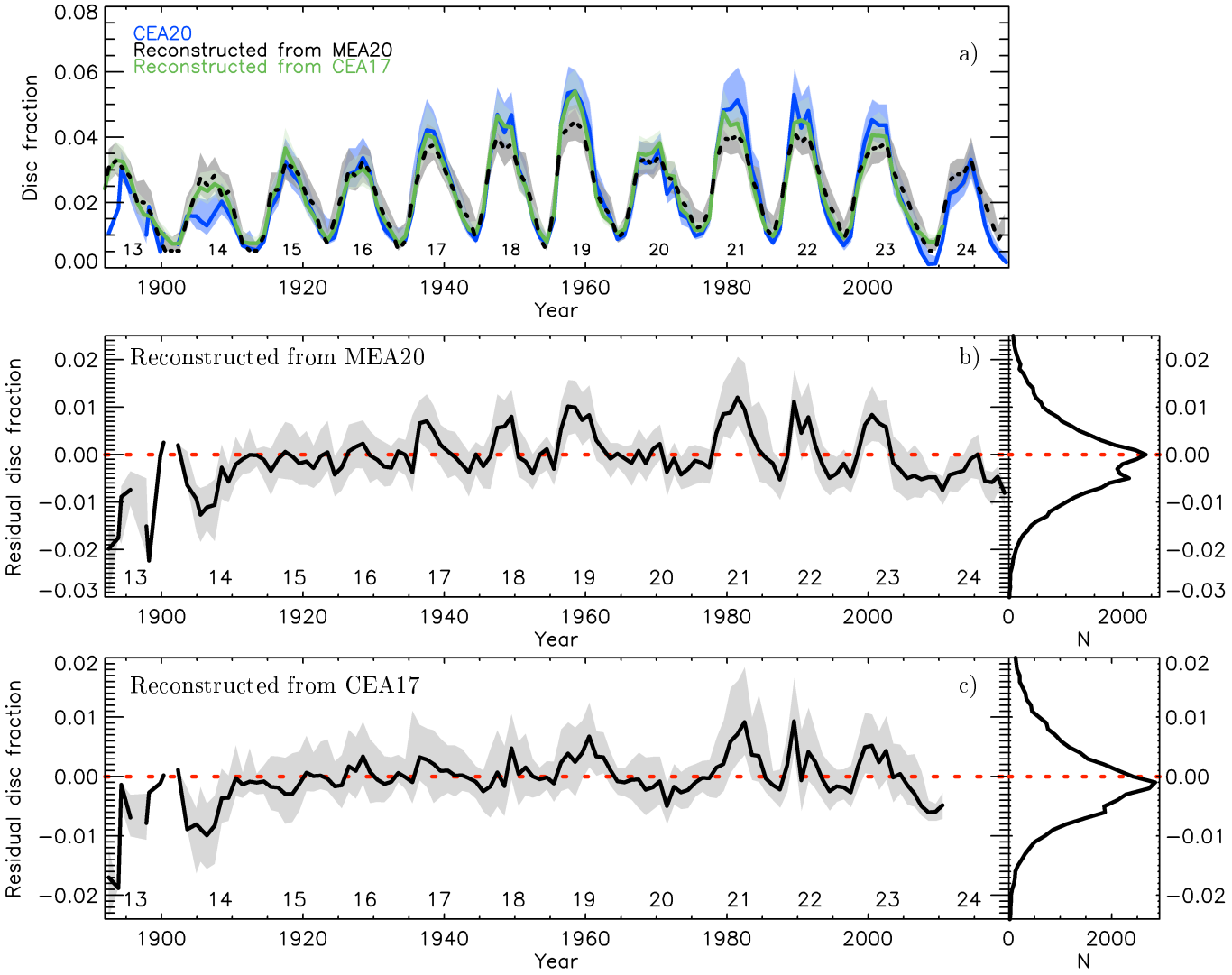

Fig. 5. Comparison of plage areas reconstructed from sunspot areas and sunspot group number series. *Panel a*: reconstructed plage areas from MEA20 sunspot area series (black) and the CEA17 GSN series (green) along with the CEA20 plage area composite (blue). Shown are annual median values (solid lines) along with the asymmetric 1σ intervals (shaded surfaces) only for the period covered by CEA20 plage area composite. The numbers below the curves denote the conventional SC numbering and are placed at SC maximum periods. *Panels b–c left*: residual areas in fractions of the disc between the CEA20 plage area composite and the plage areas reconstructed from the MEA20 sunspot areas (*panel b*) and the CEA17 GSN series (*panel c*). The red dashed line denotes 0 plage area differences. *Panels b–c right*: distributions of the differences in bins of 0.001 in fractional areas.

area composite during this period is poorer than at other times, which leads to poorer statistics in comparison to sunspot observations.

Figure 4 (bottom panel) shows the ratio of the plage area composite to the various sunspot series. To compute the ratios we ignored the days when the sunspot number was 0 or sunspot areas were lower than 0.0005 of the disc area. We note that a lower threshold for the sunspot areas would leave the ratio over activity maximum periods unaffected, but would increase the ratio during activity minima. The ratio of the plage to sunspot areas is in the range [6,30] with an average of 17 ± 4 when annual values are considered. The range is [0.6,147] for the daily values with an average of 21 ± 11. These values are consistent with those from Schatten et al. (1985), Lawrence (1987a), and Chapman et al. (2001), but they are lower than those reported by Chapman et al. (2011). It is noteworthy that the ratio increases during the descending phase of SC 19. Also interesting is the ratio of the plage areas to the sunspot number, which was roughly the same (~0.0043) during all activity maxima over SCs 15–23.

A careful comparison of different panels of Figs. 1 and 3 seems to hint at changes in the relationship between faculae and spots with time. For example, panels e and f, which are limited to cycles 22–24 exhibit weaker slopes than relationships covering earlier periods. In Appendix A we take a closer look at this, and actually do find a weak dependence of the relationships on cycle strength.

3.2. Reconstructing plage areas from sunspot series

We now use the relationships derived in Sect. 3.1 to reconstruct plage areas from sunspot data and to analyse the performance of these reconstructions. We use the power-law function on sunspot area (Eq. (1)) and GSN (Eq. (2)), respectively. Figure 5a shows the reconstructed plage areas from the MEA20 sunspot area and the CEA17 GSN series, along with the CEA20

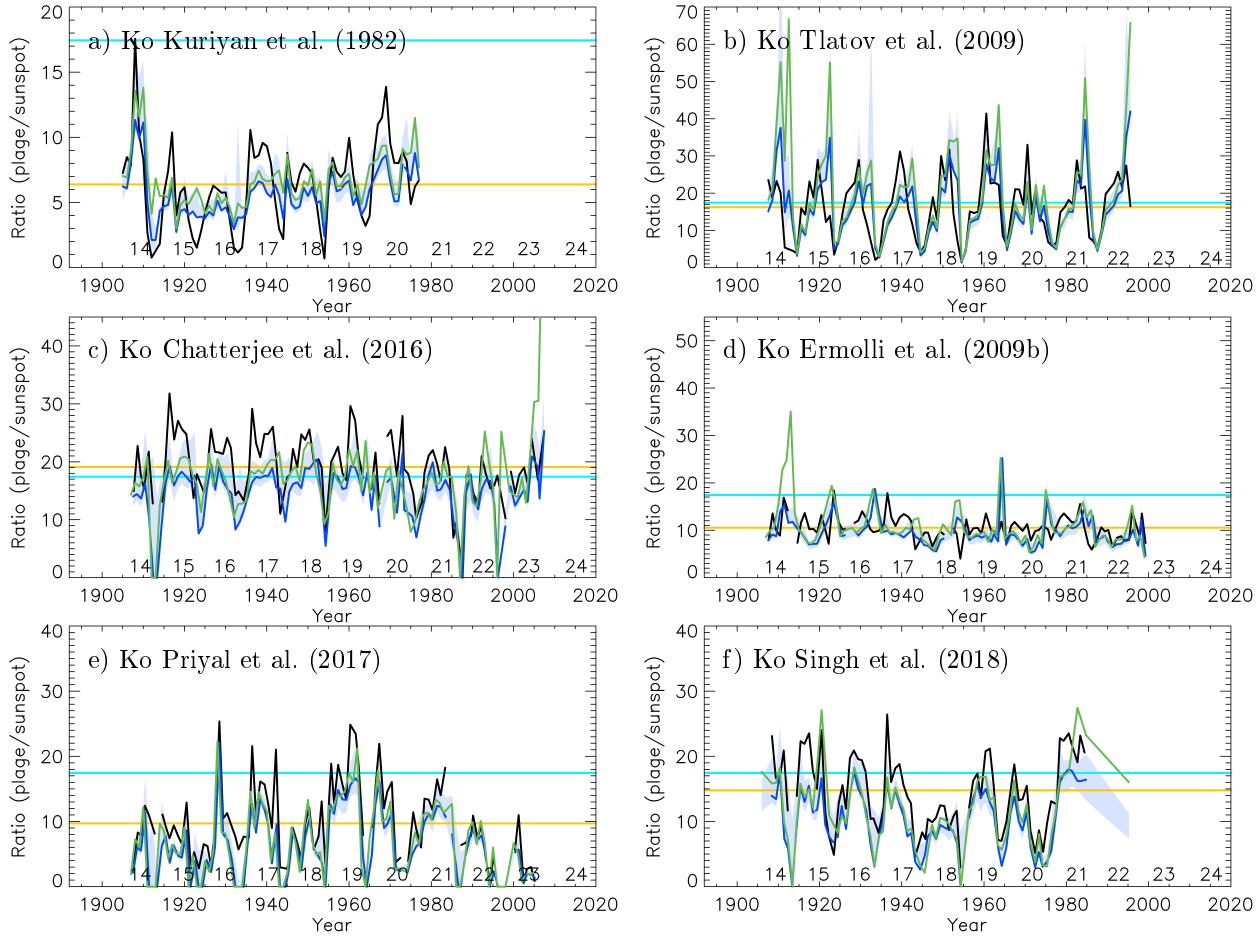

Fig. 6. Ratio of published plage area series to the sunspot area composite by MEA20 (black), and the group sunspot number series by CEA17 (blue) and SvSc16 (green). The plage area series shown are those for the Ko data by Kuriyan et al. (1982, *a*), Tlatov et al. (2009, *b*), Chatterjee et al. (2016, *c*), Ermolli et al. (2009b, *d*), Priyal et al. (2017, *e*), and Singh et al. (2018, *f*). The first two series provide only annual mean values, while the others have daily cadence. The ratios to the group sunspot number series have been multiplied by 2000 to be plotted alongside the ratios to the sunspot areas. The solid lines are annual median values. The blue shaded area denotes the range of annual values for the ratios from all the sunspot number series used in this study (see Sect. 2). The ratios for the series with daily values are calculated only for the days when the plage areas are greater than 0.0005 in disc fraction. The yellow (light blue) horizontal line represents the average ratio value between the MEA20 sunspot areas and the corresponding plage series (the CEA20 composite plage area series). The numbers at the bottom of each panel denote the conventional SC numbering, and are placed at SC maximum periods.

plage area composite. The residual plage areas between the CEA20 composite and the reconstructions from the MEA20 sunspot area and CEA17 GSN series are shown in Figs. 5b and c, respectively. Both reconstructions perform well, with RMS differences (linear correlation coefficients) between the CEA20 plage area composite and the reconstructions from the MEA sunspot area and the CEA17 GSN series being 0.009 and 0.0079 (0.83 and 0.88), respectively. Appendix C presents reconstructions using a linear, quadratic, or power-law relation with SC-dependent parameters, and also uses ISNv2 to reconstruct plage areas.

3.3. Comparison to other published plage areas

To compare our results with those based on earlier published plage area series in a consistent way, we repeated our analysis by applying the same procedure to such records. Figures 6 and 7 show the ratios of the previously published Ko and MW Ca II K plage area series to the MEA20 sunspot area series. The figure clearly demonstrates that even when based on the same Ca II K archive, but processed by different authors using different techniques and digitisation versions, the plage-to-spot ratio exhibits quite different behaviour.

The divergence is even more substantial when different archives are considered. In this respect, we note that five of the considered plage series shown in Fig. 6 have only annual values and the results derived from them have very different statistics and are thus less informative than those derived from the daily values, while three of the series are in fractional projected areas and the rest in hemispheric areas corrected for foreshortening. The ratios for the Ko series by Chatterjee et al. (2016) and Ermolli et al. (2009b) slightly decrease with time, while for the record by Kuriyan et al. (1982) a slight increase is seen. Furthermore, the ratio for the same archive derived from the data by Priyal et al. (2017) first increases towards SC 19 and then decreases again. All MW series show an increase in the ratio over SC 19 and then an abrupt drop over SC 21, consistent with the report of instrumental issues over that period (Chatzistergos et al. 2019b). The ratio for the MW series by Foukal (1996) shows a significant increase over SC 19, suggesting that the plage areas over this period are probably overestimated in this series, as also suggested by Ermolli et al. (2009b)

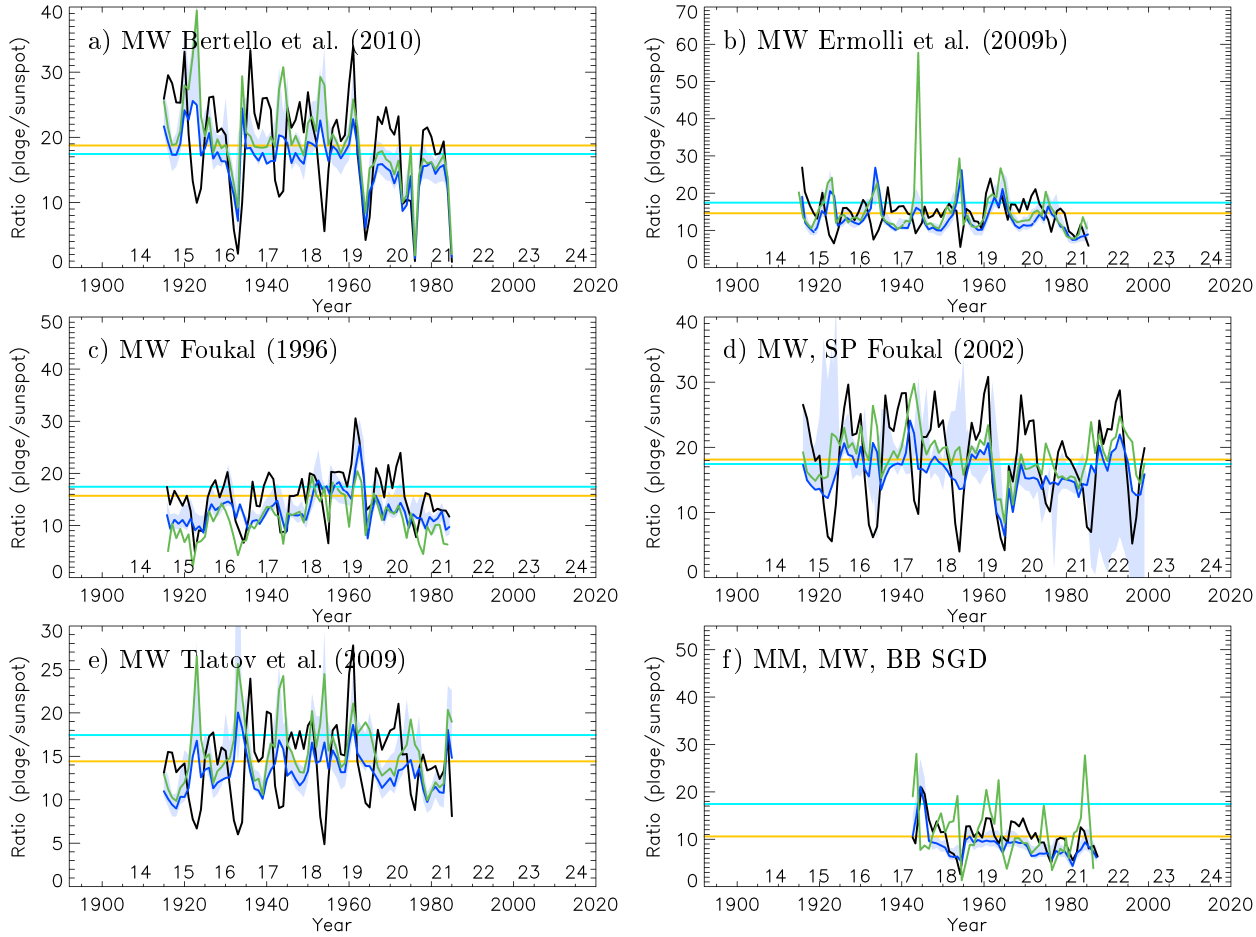

Fig. 7. Same as Fig. 6, but for the plage area series for MW data by Bertello et al. (2010, *a*); MW data by Ermolli et al.(2009b, *b*); MW data by Foukal(1996, *c*); MW and SP data by Foukal(2002, *d*); MW data by Tlatov et al. (2009, *e*); and for MM, MW, and BB data by SGD (*f*);. The series in *panels b*, *c*, and *f* have daily values, while the others only have annual mean values. The Bertello et al. (2010) series, which is a Ca II K emission index series, was linearly scaled to match the plage areas from our MW data (see Sect. 2).

and Chatzistergos et al. (2019b). The SGD series returns an average plage-to-sunspot area ratio of ∼11 compared to the value of 18 obtained from the CEA20 composite. Furthermore, the ratio based on SGD also slightly decreases with time. The series by Foukal (2002) results in a value for the plage-to-sunspot area ratio similar to ours. However, in contrast to our result, this ratio decreases over SC 20 and 21 and increases again over SC 22. This is consistent with the conjecture of a potential problem in the MW data over SC 21 (Chatzistergos et al. 2019b), while the following increase is mainly due to the use of SP data after 1985, when MW stopped operation. These findings are in agreement with the studies by Ermolli et al. (2018) and Chatzistergos et al. (2020c) who reported inconsistencies within the various published plage area series discussed here.

To our knowledge, only three of the previously published plage area records had been used for irradiance reconstructions: the series by SGD (used e.g. by Oster et al. 1982; Schatten et al. 1985; Foukal & Lean 1986, 1988; Lean & Skumanich 1983; Lean et al. 2001), Foukal (2002, used e.g. by Foukal 2012), and Bertello et al. (2010, used by Ambelu et al. 2011). All three of these series show a slight decreasing trend with time (Fig. 6), which is in contrast to our results here. This suggests that the competing contributions from sunspots and plage to irradiance variations might not have been accounted for correctly in those irradiance reconstructions. Furthermore, the decreasing trend in the plage-to-spot area ratio with time can potentially disguise any long-term trend in solar irradiance.

4. Summary and conclusions

We have studied the relation of plage areas versus sunspot areas or numbers. To do so, we used the plage area composite created by Chatzistergos et al. (2020c) by combining the data derived from 38 Ca II K archives. For comparison, we have also analysed the relationship using the underlying individual archives. The sunspot records considered are the sunspot area composite by Mandal et al. (2020a) and the various available sunspot number and group number series. The relationship between the daily plage and sunspot areas is best described by a power-law function and remains slightly non-linear also when annual values are considered. Recently, Nèmec et al. (2022) used a surface flux transport model to show that a more efficient cancellation of the diffuse magnetic flux in faculae leads to a slower increase in facular coverage with activity compared to spot coverage, eventually resulting in a non-linear relationship between them. The relationship between the plage areas and the various sunspot number records considered here is also best represented with a power-law function; however, the linear relation is a good approximation for the annual values.

We also find that the relation between the plage and sunspot areas depends on the bandwidth employed for the observations.

Furthermore, the relation shows a dependence on the solar cycle strength. However, accounting for this dependence when reconstructing plage areas from sunspot areas (numbers) results in only small (minute) improvements.

Furthermore, the relationship between plage and sunspot is affected by the accuracy of the processing of the Ca II K images. We showed that employing various published plage area series from the literature resulted in rather diverging trends in their ratios to sunspot areas and numbers. This stands when considering plage areas derived from data from the same archive (but treated differently) and when considering data from different archives. The advantage of our study is that we have considered not only numerous individual archives, but also their composite, which builds on a careful and consistent processing, analysis, and cross-calibration of the individual records (see Chatzistergos et al. 2020c).

Thus, the results of our study have significant implications for reconstructions of past irradiance variations (e.g. Domingo et al. 2009; Yeo et al. 2015; Krivova 2018, and references therein) and stellar activity studies (e.g. Lanza et al. 2003, 2019; Gondoin 2008; Reinhold et al. 2019). Such studies are often carried out assuming a constant ratio of faculae to sunspot areas, which we showed to not be very accurate.

Acknowledgements. We thank the observers at the Arcetri, Baikal, Big Bear, Brussels, Catania, Coimbra, Kanzelhöhe, Kharkiv, Kodaikanal, Kyoto, Mauna Loa, McMath-Hulbert, Mees, Meudon, Mitaka, Mt Wilson, Pic du Midi, Rome, Sacramento Peak, San Fernando, Schauinsland, Teide, Upice, Valašské Meziříčí, and Wendelstein sites. We thank Sami K. Solanki for helpful discussions during this work. We thank the anonymous reviewer for the helpful comments that helped improve this paper. We also thank Isabelle Buale for all her efforts to digitise the Meudon archive, as well Subhamoy Chatterjee, Muthu Priyal, and Andrey Tlatov for providing their published time series of plage areas. T.C. and I.E. thank ISSI for supporting the International Teams 417 "Recalibration of the Sunspot Number Series" and 475 "Modeling Space Weather and Total Solar Irradiance over the Past Century", respectively. This work was partly supported by the German Federal Ministry of Education and Research (Project No. 01LG1909C). This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 824135 (SOLARNET). ChroTel is operated by the Kiepenheuer-Institute for Solar Physics in Freiburg, Germany, at the Spanish Observatorio del Teide, Tenerife, Canary Islands. The ChroTel filtergraph has been developed by the Kiepenheuer-Institute in cooperation with the High Altitude Observatory in Boulder, CO, USA. This research has made use of NASA's Astrophysics Data System.

References

Ambelu, T., Falayi, E. O., Elemo, E. O., & Oladosu, O. 2011, Latin-Am. J. Phys. Educ., 5, 741
Arlt, R., Leussu, R., Giese, N., Mursula, K., & Usoskin, I. G. 2013, MNRAS, 433, 3165
Arlt, R., & Vaquero, J. M. 2020, Liv. Rev. Sol. Phys., 17, 1
Babcock, H. W., & Babcock, H. D. 1955, ApJ, 121, 349
Balmaceda, L. A., Solanki, S. K., Krivova, N. A., & Foster, S. 2009, J. Geophys. Res.: Space Phys., 114, A07104
Belkina, I. L., Beletskij, S. A., Gretskij, A. M., & Marchenko, G. P. 1996, Kinematics Phys. Celestial Bodies, 12, 55
Berrilli, F., Criscuoli, S., Penza, V., & Lovric, M. 2020, Sol. Phys., 295, 38
Bertello, L., Ulrich, R. K., & Boyden, J. E. 2010, Sol. Phys., 264, 31
Bertello, L., Pevtsov, A., Tlatov, A., & Singh, J. 2016, Sol. Phys., 291, 2967
Bhowmik, P., & Nandy, D. 2018, Nat. Commun., 9, 5209
Borgniet, S., Meunier, N., & Lagrange, A.-M. 2015, A&A, 581, A133
Cameron, R. H., Jiang, J., Schmitt, D., & Schüssler, M. 2010, ApJ, 719, 264
Carrasco, V. M. S., Vaquero, J. M., Gallego, M. C., & Sánchez-Bajo, F. 2016, Sol. Phys., 291, 2931
Carrasco, V. M. S., Vaquero, J. M., Gallego, M. C., et al. 2018, Sol. Phys., 293, 153
Carrasco, V. M. S., Bertolin, C., Domínguez-Castro, F., et al. 2020, Sol. Phys., 295, 112
Carrasco, V. M. S., Nogales, J. M., Vaquero, J. M., Chatzistergos, T., & Ermolli, I. 2021, J. Space Weather Space Clim., 11, 51
Chapman, G. A., Cookson, A. M., & Dobias, J. J. 1997, ApJ, 482, 541
Chapman, G. A., Cookson, A. M., Dobias, J. J., & Walton, S. R. 2001, ApJ, 555, 462
Chapman, G. A., Dobias, J. J., & Arias, T. 2011, ApJ, 728, 150
Chatterjee, S., Banerjee, D., & Ravindra, B. 2016, ApJ, 827, 87
Chatzistergos, T. 2017, Analysis of historical solar observations and long-term changes in solar irradiance, Ph.D. Thesis (Uni-edition), Germany https://ui.adsabs.harvard.edu/abs/2017PhDT.....259C/abstract
Chatzistergos, T., Ermolli, I., Solanki, S. K., & Krivova, N. A. 2016, in Coimbra Solar Physics Meeting: Ground-based Solar Observations in the Space Instrumentation Era, eds. I. Dorotovic, C. E. Fischer, & M. Temmer (San Francisco: Astronomical Society of the Pacific Conference Series), ASP Conf. Ser., 504
Chatzistergos, T., Usoskin, I. G., Kovaltsov, G. A., Krivova, N. A., & Solanki, S. K. 2017, A&A, 602, A69
Chatzistergos, T., Ermolli, I., Krivova, N. A., & Solanki, S. K. 2018a, in Long-term Datasets for the Understanding of Solar and Stellar Magnetic Cycles, eds. D. Banerjee, J. Jiang, K. Kusano, & S. Solanki (Cambridge, UK: Cambridge University Press), IAU Symp., 340, 125
Chatzistergos, T., Ermolli, I., Solanki, S. K., & Krivova, N. A. 2018b, A&A, 609, A92
Chatzistergos, T., Ermolli, I., Falco, M., et al. 2019a, in Il Nuovo Cimento, 42C, 5
Chatzistergos, T., Ermolli, I., Krivova, N. A., & Solanki, S. K. 2019b, A&A, 625, A69
Chatzistergos, T., Ermolli, I., Solanki, S. K., et al. 2019c, Sol. Phys., 294, 145
Chatzistergos, T., Ermolli, I., Solanki, S. K., et al. 2019d, A&A, 626, A114
Chatzistergos, T., Ermolli, I., Giorgi, F., Krivova, N. A., & Puiu, C. C. 2020a, J. Space Weather Space Clim., 10, 45
Chatzistergos, T., Ermolli, I., Krivova, N. A., & Solanki, S. K. 2020b, J. Phys.: Conf. Ser., 1548
Chatzistergos, T., Ermolli, I., Krivova, N. A., et al. 2020c, A&A, 639, A88
Chatzistergos, T., Ermolli, I., Krivova, N. A., et al. 2020d, VizieR Online Data Catalog: J/A+A/639/A88
Chatzistergos, T., Krivova, N. A., Ermolli, I., et al. 2021, A&A, 656, A104
Chowdhury, P., Belur, R., Bertello, L., & Pevtsov, A. A. 2022, ApJ, 925, 81
Clette, F., & Lefèvre, L. 2016, Sol. Phys., 291, 2629
Clette, F., & Lefèvre, L. 2018, Proc. Int. Astron. Union, 13, 17
Clette, F., Berghmans, D., Vanlommel, P., et al. 2007, Adv. Space Res., 40, 919
Cliver, E. W., & Ling, A. G. 2016, Sol. Phys., 291, 2763
Criscuoli, S. 2016, Sol. Phys., 291, 1957
Dasi-Espuig, M., Jiang, J., Krivova, N. A., et al. 2016, A&A, 590, A63
de Paula, V., & Curto, J. J. 2020, Sol. Phys., 295, 99
Domingo, V., Ermolli, I., Fox, P., et al. 2009, Space Sci. Rev., 145, 337
Dorotović, I., Rybàk, J., Garcia, A., & Journoud, P. 2010, Proc. 20th Nat. Sol. Phys. Meeting, 20, 58
Ermolli, I., Criscuoli, S., Centrone, M., Giorgi, F., & Penza, V. 2007, A&A, 465, 305
Ermolli, I., Chatzistergos, T., Krivova, N. A., & Solanki, S. K. 2018, in Long-term Datasets for the Understanding of Solar and Stellar Magnetic Cycles, eds. D. Banerjee, J. Jiang, K. Kusano, & S. Solanki (Cambridge, UK: Cambridge University Press), IAU Symp., 340, 115
Ermolli, I., Marchei, E., Centrone, M., et al. 2009a, A&A, 499, 627
Ermolli, I., Solanki, S. K., Tlatov, A. G., et al. 2009b, ApJ, 698, 1000
Fligge, M., & Solanki, S. K. 1997, Sol. Phys., 173, 427
Fligge, M., & Solanki, S. K. 1998, A&A, 332, 1082
Fligge, M., & Solanki, S. K. 2000, Geophys. Res. Lett., 27, 2157
Foukal, P. 1993, Sol. Phys., 148, 219
Foukal, P. 1996, Geophys. Res. Lett., 23, 2169
Foukal, P. 1998, ApJ, 500, 958
Foukal, P. 2002, Geophys. Res. Lett., 29, 2089
Foukal, P. 2012, Sol. Phys., 279, 365
Foukal, P., & Lean, J. 1986, ApJ, 302, 826
Foukal, P., & Lean, J. 1988, ApJ, 328, 347
Foukal, P., & Lean, J. 1990, Science, 247, 556
Foukal, P., & Vernazza, J. 1979, ApJ, 234, 707
Gondoin, P. 2008, A&A, 478, 883
Hayakawa, H., Besser, B. P., Iju, T., et al. 2020a, ApJ, 890, 98
Hayakawa, H., Clette, F., Horaguchi, T., et al. 2020b, MNRAS, 492, 4513
Hayakawa, H., Iju, T., Murata, K., & Besser, B. P. 2021, ApJ, 909, 194
Hoyt, D. V., & Schatten, K. H. 1998, Sol. Phys., 179, 189
Jiang, J., Cameron, R. H., Schmitt, D., & Schüssler, M. 2011, A&A, 528, A83
Jiang, J., Cameron, R. H., Schmitt, D., & Işık, E. 2013, A&A, 553, A128
Klimeš, J. J., Bělik, M., Klimeš, J., & Marková, E. 1999, in 8th SOHO Workshop: Plasma Dynamics and Diagnostics in the Solar Transition Region and Corona, eds. J. C. Vial, & B. Kaldeich-Schü , ESA Spec. Publ., 446, 375

Krivova, N. A. 2018, Climate Changes in the Holocene (CRC Press), 107
Krivova, N. A., Solanki, S. K., Hofer, B., et al. 2021, A&A, 650, A70
Krivova, N. A., Vieira, L. E. A., & Solanki, S. K. 2010, J. Geophys. Res. (Space Phys.), 115, 12112
Kuriyan, P. P., Muralidharan, V., & Sampath, S. 1982, Indian J. Radio Space Phys., 11, 229
Kuriyan, P. P., Muralidharan, V., & Sampath, S. 1983, J. Atmos. Terr. Phys., 45, 285
Lanza, A. F., Rodonó, M., Pagano, I., Barge, P., & Llebaria, A. 2003, A&A, 403, 1135
Lanza, A. F., Netto, Y., Bonomo, A. S., et al. 2019, A&A, 626, A38
Lawrence, J. K. 1987a, J. Geophys. Res.: Atmos., 92, 813
Lawrence, J. K. 1987b, Sol. Phys., 110, 73
Lean, J. L. 2018, Earth Space Sci., 5, 133
Lean, J. L., & Skumanich, A. 1983, J. Geophys. Res., 88, 5751
Lean, J. L., White, O. R., Livingston, W. C., & Picone, J. M. 2001, J. Geophys. Res.: Space Phys., 106, 10645
Lefebvre, S., Ulrich, R. K., Webster, L. S., et al. 2005, Mem. Soc. Astron. It., 76, 862
Lockwood, M., Owens, M. J., & Barnard, L. 2014, J. Geophys. Res. (Space Phys.), 119, 5172
Lockwood, M., Owens, M. J., Barnard, L., & Usoskin, I. G. 2016, ApJ, 824, 54
Lourenço, A., Carvalho, S., Barata, T., et al. 2019, Open Astron., 28, 165
Malherbe, J.-M., & Dalmasse, K. 2019, Sol. Phys., 294, 52
Malherbe, J. M., Corbard, T., Barbary, G., et al. 2022, Exp. Astron., 53, 1127
Mandal, S., Chatterjee, S., & Banerjee, D. 2017, ApJ, 835, 158
Mandal, S., Krivova, N. A., Solanki, S. K., Sinha, N., & Banerjee, D. 2020a, A&A, 640, A78
Mandal, S., Krivova, N. A., Solanki, S. K., Sinha, N., & Banerjee, D. 2020b, VizieR Online Data Catalog: J/A+A/640/A78
Mandal, S., Krivova, N. A., Cameron, R., & Solanki, S. K. 2021, A&A, 652, A9
Mohler, O. C., & Dodson, H. W. 1968, Sol. Phys., 5, 417
Nagovitsyn, Y. A., & Osipova, A. A. 2021, MNRAS, 505, 1206
Nagovitsyn, Y. A., Tlatov, A. G., & Nagovitsyna, E. Y. 2016, Astron. Rep., 60, 831
Naqvi, M. F., Marquette, W. H., Tritschler, A., & Denker, C. 2010, Astron. Nachr., 331, 696
Nèmec, N.-E., Shapiro, A. I., Işk, E., et al. 2022, ApJ, 934, L23
Oster, L., Schatten, K. H., & Sofia, S. 1982, ApJ, 256, 768
Priyal, M., Singh, J., Belur, R., & Rathina, S. K. 2017, Sol. Phys., 292, 85
Rast, M. P., Ortiz, A., & Meisner, R. W. 2008, ApJ, 673, 1209
Reinhold, T., Bell, K. J., Kuszlewicz, J., Hekker, S., & Shapiro, A. I. 2019, A&A, 621, A21
Rincon, F., & Rieutord, M. 2018, Liv. Rev. Sol. Phys., 15, 6
Schatten, K. H., Miller, N., Sofia, S., Endal, A. S., & Chapman, G. 1985, ApJ, 294, 689
Shapiro, A. I., Solanki, S. K., Krivova, N. A., et al. 2014, A&A, 569, A38
Singh, J., Priyal, M., & Ravindra, B. 2021, ApJ, 908, 210
Singh, J., Priyal, M., Ravindra, B., Bertello, L., & Pevtsov, A. A. 2022, ApJ, 927, 154
Singh, J., Priyal, M., Sindhuja, G., & Ravindra, B. 2018, Proc. Int. Astron. Union, 13, 23
Solanki, S. K. 1993, Space Sci. Rev., 63, 1
Solanki, S. K. 2003, A&ARv., 11, 153
Solanki, S. K., & Fligge, M. 1998, Geophys. Res. Lett., 25, 341
Solanki, S. K., & Unruh, Y. C. 2013, Astron. Nachr., 334, 145
Solanki, S. K., Schüssler, M., & Fligge, M. 2000, Nature, 408, 445
Solanki, S. K., Inhester, B., & Schüssler, M. 2006, Rep. Progress Phys., 69, 563
Solanki, S. K., Krivova, N. A., Schüssler, M., & Fligge, M. 2002, A&A, 396, 1029
Steinegger, M., Brandt, P. N., & Haupt, H. F. 1996, A&A, 310, 635
Svalgaard, L., & Schatten, K. H. 2016, Sol. Phys., 291, 2653
Tlatov, A. G., Pevtsov, A. A., & Singh, J. 2009, Sol. Phys., 255, 239
Usoskin, I. G., Kovaltsov, G. A., & Chatzistergos, T. 2016a, Sol. Phys., 291, 3793
Usoskin, I. G., Kovaltsov, G. A., Lockwood, M., et al. 2016b, Sol. Phys., 291, 2685
Usoskin, I., Kovaltsov, G., & Kiviaho, W. 2021, Sol. Phys., 296, 13
Vaquero, J. M., Svalgaard, L., Carrasco, V. M. S., et al. 2016, Sol. Phys., 291, 3061
Vaquero, J. M., & Vázquez, M. 2009, in The Sun Recorded Through History: Scientific Data Extracted from Historical Documents (New York, NY: Springer), Astrophys. Space Sci. Library, 361
Vokhmyanin, M., Arlt, R., & Zolotova, N. 2020, Sol. Phys., 295, 39
Wang, H., Deng, N., & Liu, C. 2012, ApJ, 748, 76
Willamo, T., Usoskin, I. G., & Kovaltsov, G. A. 2017, A&A, 601, A109
Wolf, R. 1850, Astron. Mitt. Eidgenössischen Sternwarte Zurich, 1, 27
Wu, C. J., Usoskin, I. G., Krivova, N., et al. 2018, A&A, 615, A93
Yeo, K. L., Krivova, N. A., & Solanki, S. K. 2015, in The Solar Activity Cycle, (New York: Springer), Space Sci. Ser. ISSI, 53
Yeo, K. L., Solanki, S. K., & Krivova, N. A. 2020, A&A, 639, A139

Appendix A: Dependence on cycle strength

Table A.1. Cycle-averaged values for the various sunspot series.

SC	Sunspot series			
	$\langle s_a \rangle$	$\langle s_n \rangle$		$\langle s_{gn} \rangle$
	MEA20	ISNv1	ISNv2	CEA17
1	-	-	-	3.04
2	-	-	-	5.23
3	-	-	-	3.32
4	-	-	-	3.42
5	-	-	-	2.62
6	-	1.06	1.06	2.50
7	-	3.24	3.23	3.84
8	-	4.78	4.78	4.79
9	-	4.43	4.80	4.71
10	-	4.04	4.61	4.72
11	0.28	4.43	4.44	4.34
12	0.69	2.84	2.84	3.29
13	0.88	3.25	3.25	3.84
14	0.70	2.69	2.69	3.12
15	0.89	3.67	3.67	4.09
16	0.95	3.41	3.41	3.95
17	1.24	4.80	4.81	5.38
18	1.60	6.22	5.45	6.06
19	1.98	7.60	6.47	6.85
20	1.11	5.10	4.34	5.16
21	1.73	6.63	5.56	6.24
22	1.66	6.58	5.32	6.04
23	1.32	4.57	4.12	5.01
24	0.88	-	2.57	-

Notes. The values for MEA20 are given in fractional areas multiplied with 1000. The ISNv1 and ISNv2 sunspot number values were divided by 12.08 (ISNv2 values are also multiplied by 0.6) to bring them roughly to the same level as the group sunspot number series.

Comparing the different panels of Figs. 1 and 3, it is evident that the relationship between the plage areas and the sunspot observations shows differences among the datasets. While some of the differences are clearly due to specifics of the individual plage series, there also seems to be a hint of changes in the relationship with time. Thus, for example, the relationships limited to cycles 22 to 24 (see panels e and f) have a weaker slope than relationships covering earlier periods. Here we take a closer look at the temporal behaviour of the relationship.

In Fig. A.1 we compare the relationship between plage and sunspot areas for individual SCs for the Ko, MD, and MW series, which are the longest series considered in this study, and for the CEA20 plage area composite. We defined the strength of the SC as the median value of sunspot areas over each SC and use different colours to represent this in Fig. A.1, with dark blue for the weakest SCs through green and yellow for intermediate SCs to dark red for the strongest ones. Figure A.1 shows a clear dependence of the plage versus spot area relationship on the SC strength, with plage areas generally rising faster with sunspot areas for stronger cycles. This is probably because during stronger cycles, more active regions emerge overall, which in turn leads to an increased number of decaying regions, eventually seen as faculae. As a result, at a given instant in time there are more facular regions on the surface than there would be during weaker cycles for the same spot coverage. However, some deviations from strong cycles rising faster are also evident in our results. Thus, interestingly, the two strongest cycles 19 and 21 show a rather similar relationship (except for MW), but in Ko the values for SC 21 are somewhat higher than those for SC19, while the opposite is seen for MD. We note, however, that a significant scatter at higher activity levels might affect the fits. In MW, plage areas for SC 19 reach notably higher values than in other series, while values for SC21 are significantly lower in other records than would be expected for such a strong cycle. This hints at potential problems and inconsistencies in the MW archive over these periods of time. Figure A.1d shows the relation between the composite plage areas from CEA20 and MEA20 sunspot areas for each SC between 13 and 24. The general form of the studied relation and its dependence on the cycle strength is overall the same as when individual series are considered (panels a–c). The change with the cycle strength is, however, significantly clearer, with flatter relations for weaker cycles. Again, SC 19 shows a very similar relationship to the next two cycles in strength, SCs 21 and 22.

Figure A.2 is similar to Fig. A.1, except that it shows the Ca II K plage area series versus the CEA17 GSN series (instead of the MEA20 sunspot area composite shown in Fig. A.1). The dependence of the relation between plage areas and the GSN on the SC strength is evident, although less pronounced than between the plage and sunspot areas. For most series, it manifests for higher sunspot group number, while it appears to be absent from the MD series. The results from the archives showing anomalous behaviour over certain SCs when compared to the sunspot areas exhibit the same behaviour when compared to the GSN series. This is unsurprising, considering that the various GSN and sunspot area series are rather consistent with each other over the 20th century.

To highlight the differences between the archives and issues affecting them, Fig. A.3 shows the same relationships for the Co, Ko, MD, and MW series, and for the CEA20 composite, but now for individual cycles 15–22 in each panel. We note that these four archives employ different nominal bandwidths for their observations, with Co and MD using 0.15Å, MW 0.2Å, and Ko 0.5Å. Furthermore, the CEA20 composite series was created with RP as the reference dataset, which has a nominal bandwidth of 2.5Å. A narrower bandwidth generally results in higher plage areas compared to a broader one (Chatzistergos et al. 2020c). The effect of the bandwidth on the relation between plage areas and sunspot data is discussed in Appendix B. The differences between these four archives are generally higher for earlier cycles and are highest for SCs 18 and 19. In SC 20 the Ko, MD, and MW values lie comparatively close to each other, whereas the Co values are a factor of about 2–3 lower. The plage areas from the Ko and Co archives are generally close to the composite values and lower than MD and MW areas. However, in SC 21 MW plage areas are the lowest, while they are highest in SC 19. In addition, in SC19 the Co values are lower than all others, while in SC 22 the Ko values are clearly the lowest. Unfortunately, such a comparison for most of the other plage series is not possible due to their insufficient coverage. Figure A.3 highlights once again that care is needed when using historic observations for past activity studies and that employment of multiple overlapping archives is of great advantage (see also Chatzistergos et al. 2021).

To further study the variation of the derived relationship with the SC strength we relate the fit parameters to the SC-averaged sunspot area. Figure A.4 shows the exponents of the power-law fits to the relation between plage and sunspot areas as a function of the SC-averaged sunspot area. Also shown are the linear fits to the values for individual cycles (solid red lines). Thus, SCs 13 and 14 were excluded from the fits due to poor data

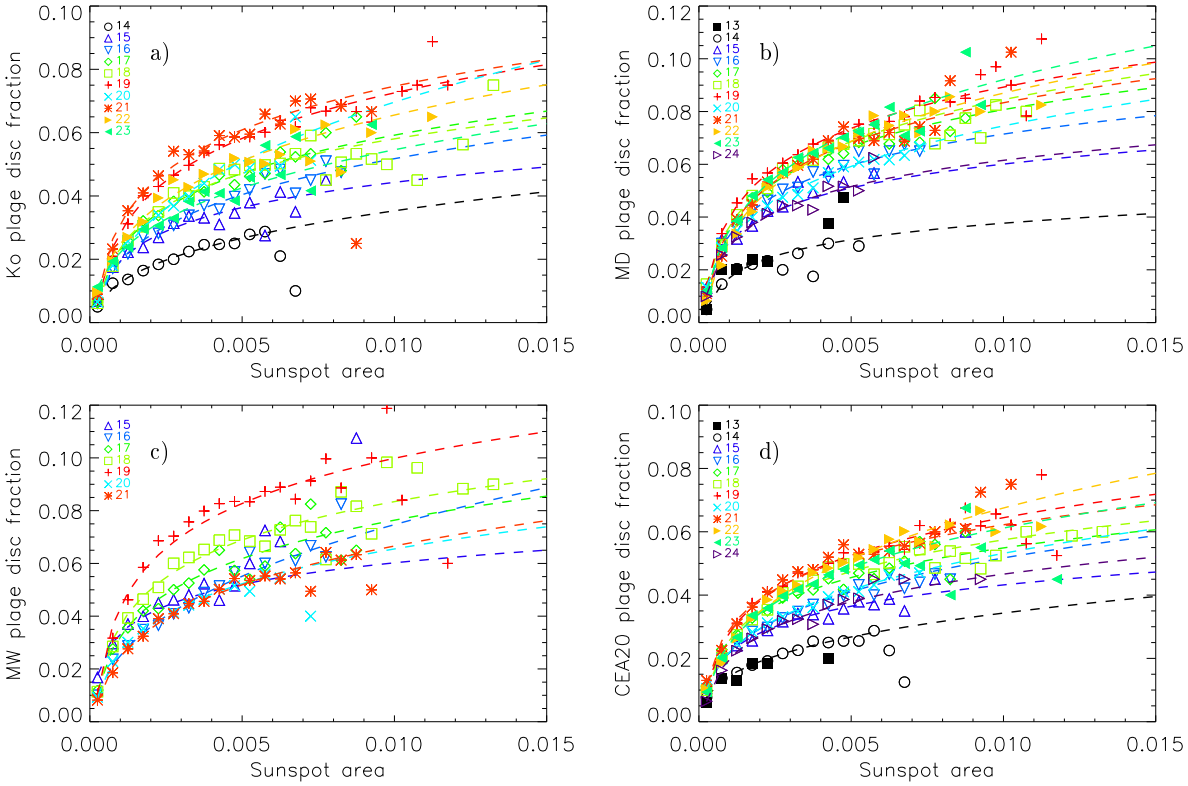

Fig. A.1. Relation between plage and sunspot areas for individual SC based on the data from the Ko (a), MD (b), and MW (c) archives, and from the CEA20 plage area composite (d). The symbols denote the mean values of the PDF matrices (as shown in Fig. 3). The dashed lines are power-law fits and are coloured according to the strength of the SC with the strength increasing from black for the weakest cycle, via dark blue, turquise, green, lime, orange, bright red to dark red for the strongest SC.

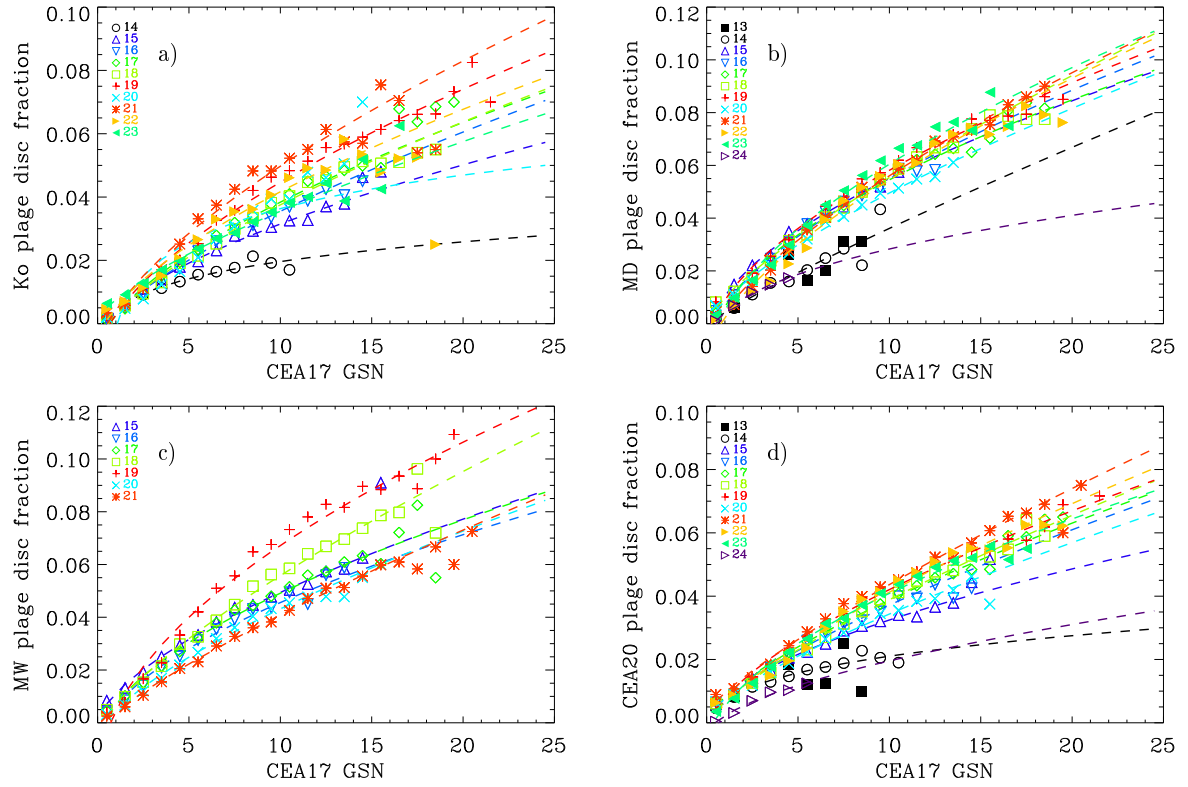

Fig. A.2. Relation between plage areas and CEA17 GSN for individual SC based on the data from the Ko (a), MD (b), and MW (c) archives, and from the CEA20 plage area composite (d). Shown are the binned values along with power-law fits. The lines are coloured according to the strength of the SC with the strength increasing from black for the weakest cycle, via dark blue, turquise, green, lime, orange, bright red to dark red for the strongest SC.

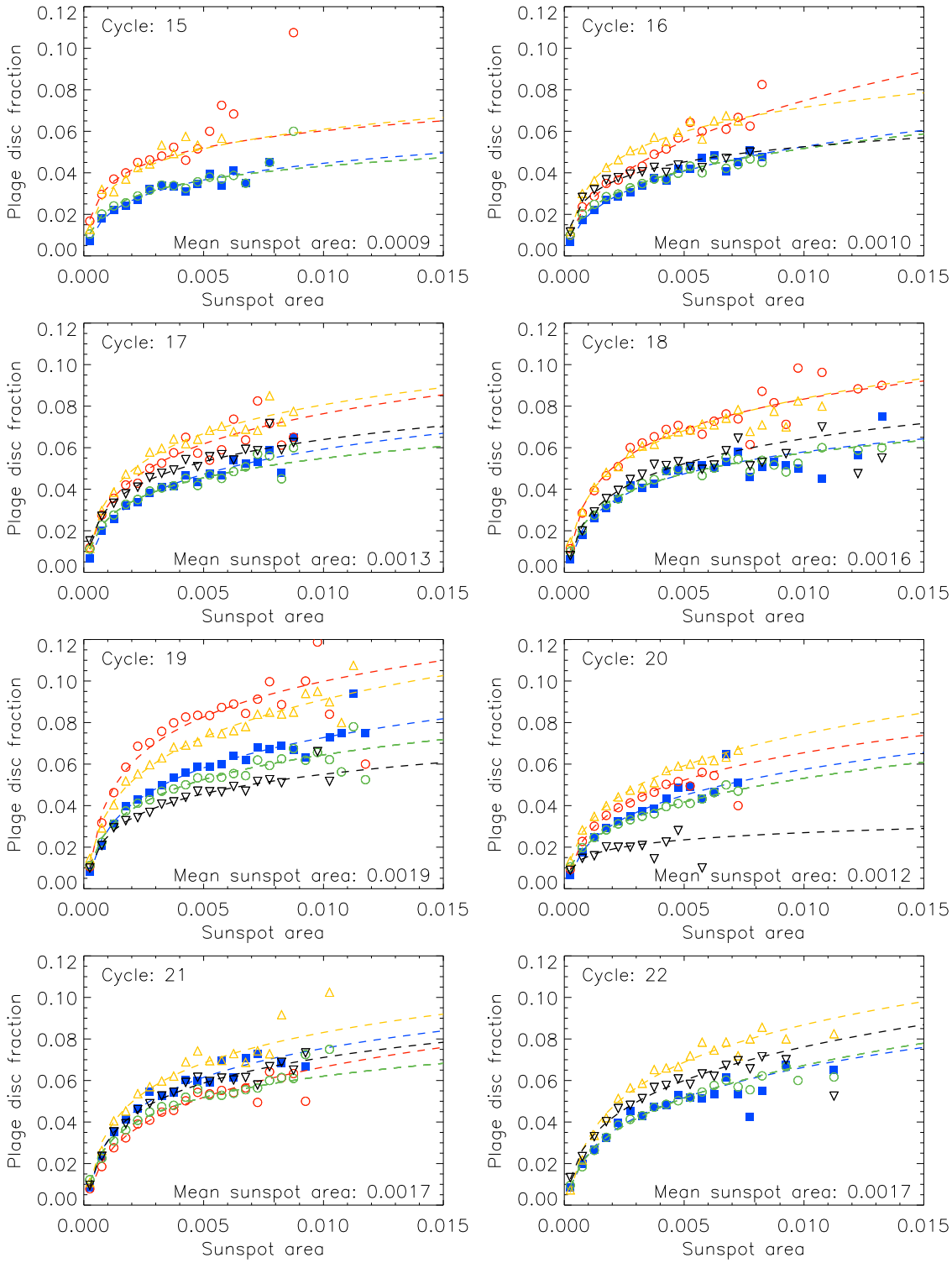

Fig. A.3. Similar to Fig. A.1, but here each panel shows the relationships for Co (black downward triangles), Ko (blue squares), MD (orange upward triangles), MW (red circles), and the CEA20 composite (green circles) series for individual SCs 15 to 22, as specified in each panel. Also listed in each panel is the cycle-averaged sunspot fractional area.

coverage. The fits for the CEA17 GSN series also exclude SC 24 since these records do not cover the entire SC 24. The exponent of the power-law fit decreases consistently with increasing SC-averaged sunspot areas with a slope of ~ -53. The slope of the linear relation between plage areas and the CEA17 GSN increases weakly with SC-averaged GSN with a coefficient of 0.02. The opposite behaviour of the exponents for the sunspot areas and sunspot number series occurs because sunspot areas are measured in fractions and are always smaller than one, whereas sunspot numbers are typically greater than 1. We note that SC 21 has the lowest (highest) value of the exponent of the fits to sunspot areas (CEA17 GSN). SC 19 is trailing SC 21, even though SC 19 has a stronger amplitude in all the analysed series. We find similar results for ISNv2 to those reported for the

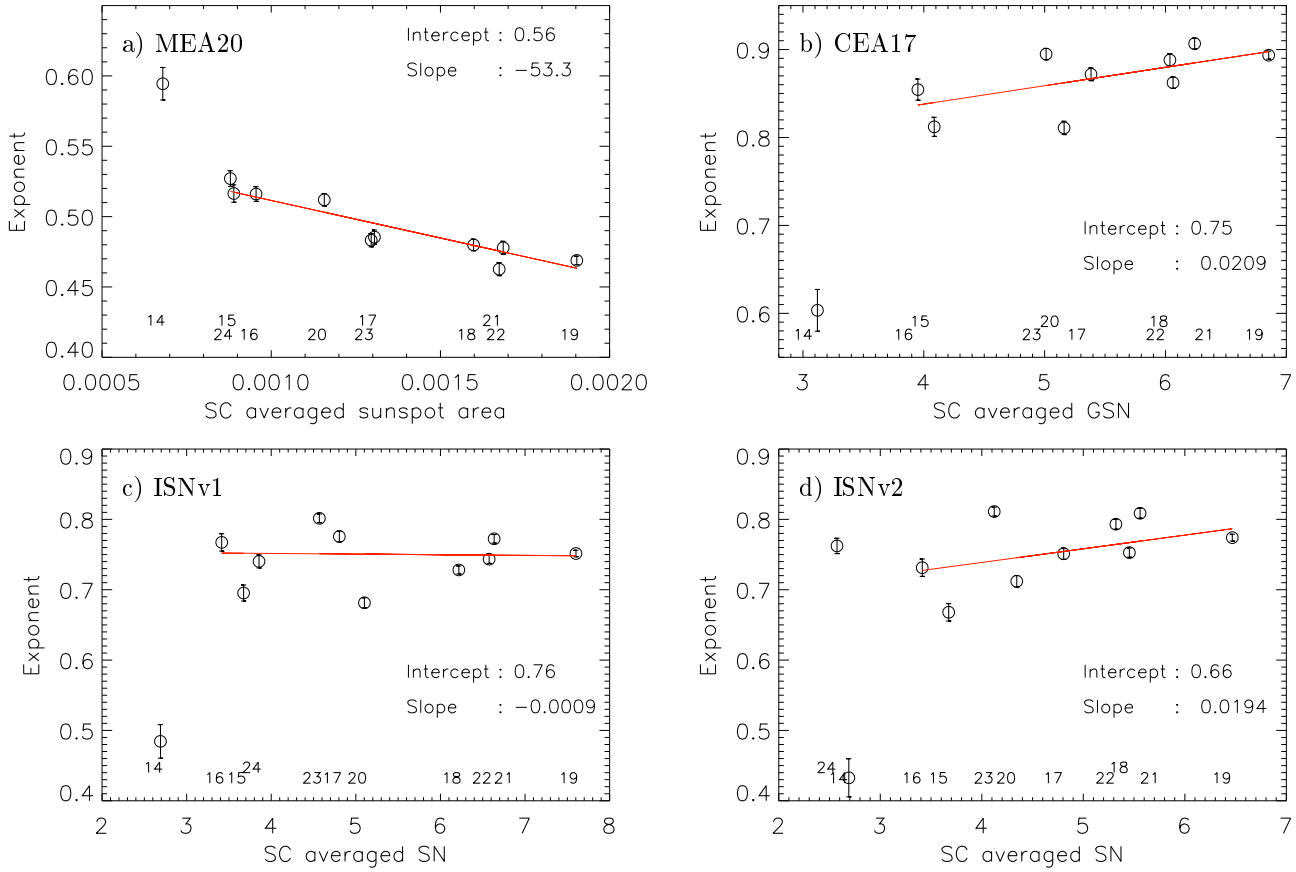

Fig. A.4. Exponents of power-law fits to the relation between the CEA20 plage area composite and MEA20 sunspot areas (panel a), and the CEA17 GSN (panel b), ISNv1 (panel c), and ISNv2 (panel d) series plotted as a function of SC average sunspot number. Solid red lines are linear fits to the individual SC values excluding SCs 13 and 14, which do not have enough statistics in the plage area composite. For the comparison to the sunspot number series SC 24 was also excluded since the CEA17 GSN series ends in 2010, while ISNv1 in 2014. The parameters of the linear fits are given in each panel. The numbers within the panels denote the conventional SC numbering.

CEA17 GSN series, with a slightly lower slope of 0.019 instead of 0.02. The results for ISNv1, however, show a much lower dependence of the exponents on the cycle strength, with a slope of -9 × 10^{-4}.

Taking the dependence of the relationship on the strength of the solar cycle into account, Eqs. 1 and 2 can be written as

$$p_a = (0.34 \pm 0.09) \times s_a^{(0.565\pm0.006-(53.3\pm4.0)\times\langle s_a\rangle)} - (0.004 \pm 0.008), \quad \text{(A.1)}$$

$$p_a = (0.005\pm0.001)\times s_{gn}^{(0.75\pm0.03+(0.021\pm0.003)\times\langle s_{gn}\rangle)} + (0.004\pm0.001), \quad \text{(A.2)}$$

where $\langle s_a \rangle$ and $\langle s_{gn} \rangle$ are the cycle-averaged sunspot area and group number, respectively. The values of $\langle s_a \rangle$ and $\langle s_{gn} \rangle$ for the various sunspot series mentioned in this study are given in Table A.1.

Appendix B: Dependence on bandwidth of observation

Here, we discuss how the bandwidth used for the observations affects the relationship between plage areas and sunspot data. Some systematic differences in the relationship between the plage and sunspot areas among the Ca II K observations taken with different bandwidths were already hinted at in Figs. 1 and 3, where the relationship for archives taken with a narrow bandwidth (e.g. MD, MW, or SP) is steeper compared to archives taken with a broader bandwidth (e.g. RP or SF2).

To illustrate this in a more systematic way, Fig. B.1 (panels a and b) shows the exponent of the power-law fits to the average values of the PDF bins (as shown in Fig. 1 and 3, respectively) as a function of the nominal bandwidth of each archive. To avoid the potential effects of the activity variations with time, we only show the data covering the SC 23. We chose this cycle because it is covered by the largest number of available long-running archives, including photographic archives. There is a slight tendency for the exponent to increase (decrease) with the nominal bandwidth for the sunspot areas (GSN series), although with a high scatter. We recall that the opposite behaviour of the exponent is due to the sunspot areas always being lower than one, while sunspot number series are in general greater than 1. We note, however, that the actual bandwidth of the observations does not always correspond to the nominal value, and considerable ambiguities and temporal variations have been noticed for some of the archives (Chatzistergos et al. 2019b, 2020c), hindering a more accurate evaluation of the bandwidth effect. These archives are shown by triangles in the figure, and it is evident that these archives (see Chatzistergos et al. 2019b, 2020c) are exactly the ones that show the highest scatter with respect to the expected relationships. For instance, Fig. B.1 suggests that SP and Kh used a narrower bandwidth, while BB, Ko, and UP a broader bandwidth than their nominal ones. Furthermore, inconsistencies within the individual archives (Chatzistergos et al. 2019b) also affect this comparison, for example the changes in the instrument used to record the BB and ML archives.

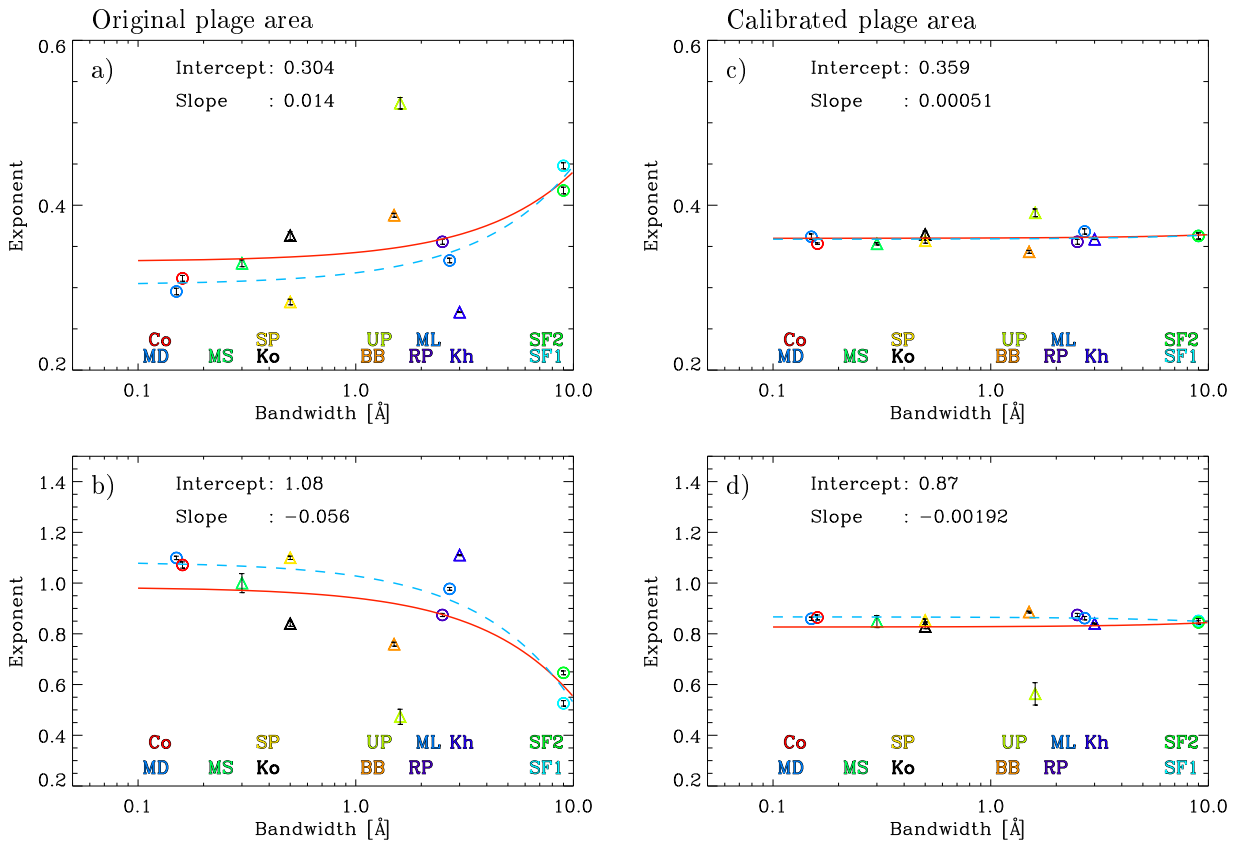

Fig. B.1. Dependence on bandwidth of the exponent of a power-law function and the slope of a linear fit between the plage areas and MEA20 sunspot area (a, c) and the CEA17 GSN (b, d), respectively. Shown are results over SC 23. The archives used to derive the plage areas are BB, Co, Kh, Ko, MD, ML, MS, RP, SF1, SF2, SP, and UP. The results shown in panels a) and b) are for the original plage area series from each archive, while those in panels c and d are for the plage areas after the scaling applied by Chatzistergos et al. (2020c) to cross-calibrate all series and merge them into the composite series. The symbols denote the fit result to the mean values of the PDF matrices (as shown in Fig. 1) along with the uncertainty of the fit parameter. Triangles are used for the archives for which there are inconsistencies rendering the bandwidth value uncertain, while circles are used for all other archives. The lines are linear fits for all points (solid red) and excluding the archives for which the nominal bandwidth appears inconsistent (dashed blue). The parameters of the latter fit are also written in each panel. Archive abbreviations at the bottom of each panel follow the order of the respective symbols in the plots.

We further note that the dependence of the exponent of the fits on the bandwidth of the observations is lifted when using the series after their cross-calibration to RP, as done by Chatzistergos et al. (2020c) in order to merge them into the CEA20 plage area composite series. This is illustrated in panels c and d of Fig. B.1, which are the same as panels a and b, but for the plage area series after applying the scaling by Chatzistergos et al. (2020c). We find that the exponents for all cross-calibrated series lie closer to that of RP compared to the original series (without the cross-calibration). The slope of the exponent is significantly reduced compared to the values for the original plage area series. These results support the corrections done by Chatzistergos et al. (2020c).

Appendix C: Reconstructing plage areas from sunspot data and different relations

Here we use the relationships derived in Sects. 3.1 and Appendix A to reconstruct plage areas from the various sunspot series and analyse the performance of these reconstructions. For comparison, we use three different relations: 1) a power law (Eq. 1) and 2) a square root (linear for CEA17 and ISNv2), both taken to be constant over the entire period analysed in this study (see Table 4), and 3) a power-law function with a linear dependence of the exponent on the SC-averaged sunspot area (Eq. A.1) and GSN (Eq. A.2), respectively.

Figure C.1a shows the reconstructed plage areas from the MEA20 sunspot area series with the three above mentioned relations along with the CEA20 plage area composite. Figure C.1b–d show the residuals between the CEA20 plage area composite and the plage areas reconstructed as described above with all three relations from the MEA20 sunspot areas, the CEA17 GSN series, and ISNv2, respectively.

In all cases we find that the power-law relation with the SC-dependent parameters performs better; however, the improvement is rather small. In particular, the RMS differences between the CEA20 composite series and the series reconstructed with a square root, a power law, and a power law with SC-dependent exponents used on the MEA20 sunspot area series are 0.0096, 0.0090, and 0.0086, respectively, while the linear correlation coefficients are 0.83, 0.83, 0.85, respectively. Hence, the overall improvement when using an SC-dependent power-law function is rather small. Nevertheless, this reduces the activity dependent effect on the reconstructed plage areas and improves the agreement over cycle maxima, which tend to be slightly underestimated for strong cycles for the reconstructions with a power law with time-independent parameters.

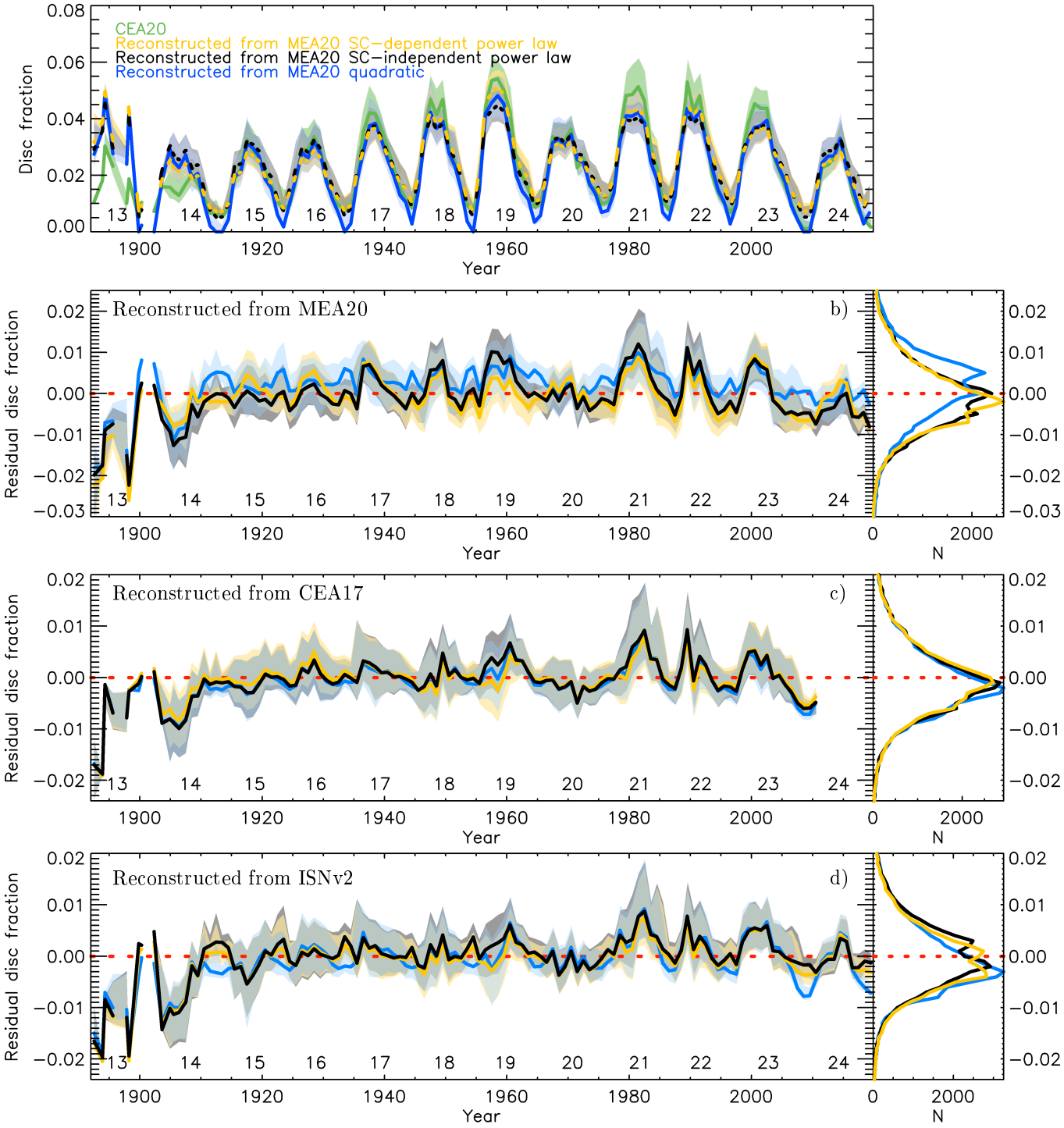

Fig. C.1. Comparison of reconstructed plage areas from sunspot areas using different relationships. *Panel a:* Reconstructed plage areas from sunspot data and from the CEA20 plage area composite. Shown are annual median values (solid lines) along with the asymmetric 1σ intervals (shaded surfaces). The numbers below the curves denote the conventional SC numbering and are placed at SC maximum periods. *Panels b–d Left:* Residual areas in fractions of the disc between the CEA20 plage area composite and the plage areas reconstructed from the MEA20 sunspot areas (top panels), the CEA17 GSN series (middle panels), and the ISNv2 series (bottom panels). The reconstructions were produced with a simple linear scaling (blue, only for CEA17 GSN and ISNv2), with a square-root function (blue, only for the MEA20 sunspot areas), with a power-law function (black), and a power-law function with SC-dependent exponents (orange). The red dashed line denotes 0 plage area differences. *Panels b–d Right:* Distributions of the differences in bins of 0.001 in fractional areas.

Figure C.1 (panel c) shows the residuals between the CEA20 plage area composite and the reconstructed plage areas derived from the CEA17 GSN record. Again, for comparison, we also show the plage areas reconstructed using a linear, a power law with time-independent parameters, and a power-law function with SC-dependent exponents. All three reconstructions perform equally well, with RMS difference to the plage area composite for the linear, power law, and power law with SC-dependent exponents being 0.0079, 0.0079, and 0.0078, respectively, while the linear correlation coefficients are 0.87, 0.88, and 0.88. We

note that using the ISNv2 series for the plage area reconstruction returns a marginally better agreement with the composite than when CEA17 is employed (Figure C.1 panel c; RMS differences of 0.0076 and 0.0078 for power-law function with SC-dependent and SC-independent exponents, and 0.0078 for the linear relationship, all when considering the overlapping periods between ISNv2 and CEA17). We note that the distribution of residuals of the reconstructed plage areas from ISNv2 to the CEA20 plage area composite exhibits two peaks, for instance in the series reconstructed with a power-law function one peak is close to 0 and the other at about 0.0055. The latter arises due to days with sunspot number in ISNv2 of 0 for which the plage areas are closer to 0.01 (see Fig. 4). Similar double peaks are seen in the residuals for the reconstructions with MEA20 and CEA17, though they are less evident for CEA17.