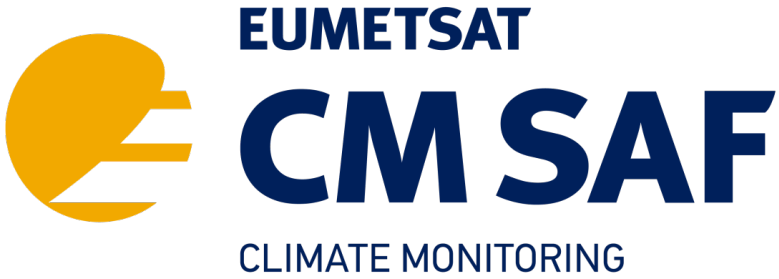
EUMETSAT Satellite Application Facility on Climate Monitoring



Towards a blended in situ-satellite sunshine duration dataset for UK climate monitoring

Reference Number:	SAF/CM/UKMO/CDOP4/REP/SD_UC
Issue/Revision Index:	1.0
Date:	22.02.2024

Document Signature Table

	Name	Function	Signature	Date
Author	Elizabeth Good	Author		22 Feb 2024
	Josh Blannin	Author		13 Feb 2024
	Phil Whybra	Author		13 Feb 2024
	James Frost	Contributor		
	Molly James	Contributor		
	Charlotte Norris	Contributor		
	Dimitrios Theodorakis	Contributor		
Editor	Marc Schröder	CM SAF Science Coordinator		22.02.2024
Approval	CM SAF Steering Group			
Release	Rainer Hollmann	CM SAF Project Manager		

Distribution List

Internal Distribution	
Name	**No. Copies**
DWD / Archive	1

External Distribution		
Company	**Name**	**No. Copies**
Public		0

Document Change Record

Issue/ Revision	Date	DCN No.	Changed Pages/Paragraphs
1.0, draft	13/02/24	SAF/CM/UKMO/CDOP4/REP/SD_UC	First version
1.0	22/02/24	SAF/CM/UKMO/CDOP4/REP/SD_UC	Edits made following review by Marc Schröder

Applicable Documents

Reference	Title	Code
AD 1		

Reference Documents

Reference	Title	Code
RD 1		

Table of Contents

List of Tables

List of Figures

1 Introduction

Sunshine Duration (*SD*) is a key climate variable that has been measured reliably at weather stations since the 1800s. The first official measurement of *SD* in the Met Office station database for the UK is currently recorded on 1 July 1887 (station source identification number 808, Eastbourne) although there may be undigitised data that exist prior to this date. Since then, the UK station network that provides *SD* observations has changed considerably over time, reaching a peak of ~400 stations in the 1970s. However, the number of *SD* sensors has declined since the 1970s and the UK currently has only ~100 stations that measure *SD* (Figure 1-1). These data are used as input into the Met Office 'HadUK-Grid' *SD* dataset, produced by the National Climate Information Centre (NCIC), which provides gridded *SD* data for the UK since 1910 (Met Office, 2018; Perry & Hollis, 2005).

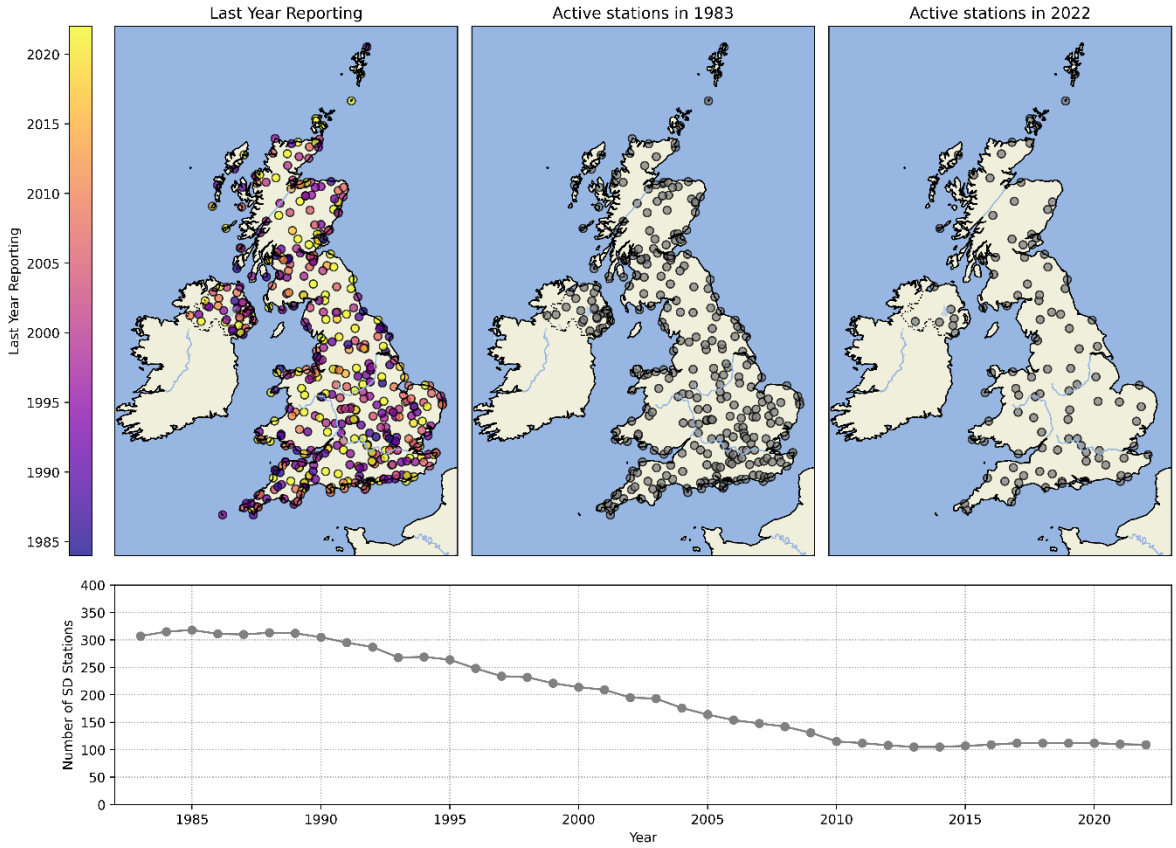


Figure 1-1: Map and time series showing the location and total number of unique stations used in this study between 1 January 1983 and 31 December 2022. The UK station network with *SD*-observing capability includes 'non-standard' stations that are excluded from this study and are therefore not shown in this figure. Only stations with at least 365 valid daily observations are included; further details on the data quality control and screening are presented in Section 2.1. The left-hand map shows all the stations that report *SD* between 1983 and 2022, where the colour indicates the last reporting year (colour bar on the left-hand side of this map).

Due to the increasing sparsity of the station network, uncertainties in the HadUK-Grid *SD* dataset are increasing. This is a problem shared by many other National Meteorological Services (NMS) whose *SD*-observing station networks are also in decline. Consequently, a number of studies have developed satellite-based estimates of *SD*, which can provide

spatially-complete fields. For example, Kandirmaz (2006) proposed a simple model that estimated *SD* from a daily mean cloud cover index derived from geostationary satellite data, which was tested on data from the Meteosat First Generation (MFG). Good (2010) used 15-minute cloud-type observations from Meteosat Second Generation (MSG) to estimate daily *SD* for the UK, which was then evaluated using the same station *SD* observations that are used as input into the HadUK-Grid *SD* dataset. Wu et al. (2016) also used cloud classification from the geostationary FenYung-2D satellite to estimate *SD* over the Heihe River Basin in China. Kothe et al. (2013) extended the approach of Good (2010) to Europe, also using MFG data, and trialed the use of surface incoming direct (SID) radiation in place of cloud type data to improve the satellite *SD* estimates. Later, Kothe et al. (2017) modified the SID approach further to develop the first operational satellite-based SD product for the Satellite Application Facility for Climate Monitoring (CM SAF).

Several studies have also tried to create blended *SD* datasets, which are based on a combination of satellite and station-based observations, in the hope that these datasets may provide better *SD* estimates and with lower uncertainty compared with using sparse station *SD* alone. For example, Frei et al. (2015) used principal component analysis and kriging with external drift to apply spatial patterns in *SD* derived from satellite data to interpolate station *SD* observations to produce improve estimates of *SD* over Switzerland. Journée et al. (2013) also used kriging with external drift to derive a blended satellite-station *SD* dataset for Belgium and Luxembourg, which they found out-performed linear regression and ordinary kriging. Bertrand et al. (2013) also estimated *SD* for Belgium using a combination of an MSG-based clearness index together with station *SD* observations.

The objective of this study is to develop a blended satellite-station dataset for the UK. The study uses the CM SAF *SD* dataset produced by Kothe et al. (2017), which is available since 1983, and in situ station observations from the UK that are maintained by the Met Office between 1983 and 2022; Figure 1-1 shows the location and number of Met Office weather stations used in the study. The aim is to produce blended *SD* data at both monthly and daily temporal resolution, which can be made available to users alongside the monthly, seasonal and annual HadUK-Grid *SD* dataset. A significant component of the study is an assessment of the station *SD* data and how well these data agree with the satellite *SD* estimates. In particular, the station observations from the two different sensor types that are used in the UK are intercompared to determine which sensor agrees more closely with the satellite *SD*. The conversion used by the Met Office to unify the two different types of station observations is also evaluated. The study is presented as follows: Section 2 describes the datasets used in the study. Section 3 presents the methods used to intercompared the different datasets and to produce the blended *SD* estimates. Section 4 presents the results of the intercomparisons and the blended SD dataset, including the justification for the choice of the final station *SD* dataset that is used as input. Section 5 provides the conclusions of the study and looks ahead to the next steps needed in order to provide these new datasets operationally in near-real time.

2 Data

Details of the different *SD* datasets used in this study are provided in the following sub-sections. A summary of these datasets is presented in Table 2-1.

Table 2-1: Summary of the *SD* datasets used in this study.

Dataset acronym	Description	Reference/source
SD	Generic term to describe sunshine duration.	N/A
SD_{stn}	*SD* dataset based on 'point' in situ observation(s) from weather stations.	MIDAS Open (Met Office, 2023)
SD_{cs}	*SD* dataset based on 'point' in situ observation(s) recorded only by Campbell Stokes (CS) sensor(s); no corrections or adjustments applied.	MIDAS Open (Met Office, 2023)
SD_{kz}	*SD* dataset based on 'point' in situ observation(s) recorded only by Kipp & Zonen (KZ) sensor(s); no corrections or adjustments applied.	MIDAS Open (Met Office, 2023). Also see https://www.kippzonen.com/.
SD_{cs_con}	*SD* dataset derived from SD_{cs} converted to the equivalent KZ values only, i.e. no SD_{kz}.	Legg (2014)
SD_{kz_con}	SD dataset derived from SD_{kz} converted to the equivalent CS values, i.e. no SD_{cs}	Legg (2014)
SD_{cs_uni}	*SD* dataset comprising both SD_{cs} and SD_{kz_con} only, i.e. no SD_{kz}.	
SD_{kz_uni}	*SD* dataset comprising both SD_{kz} and SD_{cs_con} only, i.e. no SD_{cs}.	
SD_{sat}	A daily 0.05°x0.05° *SD* dataset for Europe, Africa and parts of S. America based on satellite observation(s) produced by the CM SAF.	Kothe et al. (2017)
SD_{had}	A monthly 5x5 km gridded *SD* dataset for the UK derived from SD_{cs_uni} produced by the Met Office Hadley Centre.	Perry & Hollis, 2005
SD_{bld}	A daily and monthly 0.05°x0.05° *SD* dataset for the UK produced by blending SD_{cs_uni} and SD_{sat} observation(s).	Derived in this study.

2.1 Station SD (SD_{stn})

The in situ station data used in this study have been sourced from the MIDAS (Met Office Integrated Data Archive System) Open database, which is a subset of the full MIDAS record. This study uses MIDAS Open version 202308. MIDAS Open contains only UK mainland land surface observations from weather stations owned and operated by the Met Office. The data set contains approximately 95% of the available weather observations (Met Office, 2019). The UK daily weather observations, including *SD* recorded at stations (SD_{stn}), are recorded on a 24 hour time scale and utilise data transmitted within the DLY3208[1], NCM[2], AWSDLY[3], and SYNOP[4] messages (Met Office, 2023). These observational sources are recorded as the *met domain* in the MIDAS Open database.

Most UK weather stations available in MIDAS Open are used in this analysis; only a few non-standard stations are excluded based on information supplied by the NCIC (NCIC, Personal Communication). These stations either may never be suitable for use or can become available for certain periods of time. For any period that a station is listed as non-standard, it is excluded from this study.

The SD_{stn} in MIDAS Open is recorded in hours per day, and each measurement has a 5-digit code (note: leading zeros may be excluded) conveying the quality control (QC) applied to each measurement. The QC is implemented in this study following recommended practise (MIDAS Open documentation, https://help.ceda.ac.uk/article/4982-midas-open-user-guide; NCIC Personal Communication), where any measurement with a *status flag* equal to 1 is excluded from the data set as this corresponds to the data being "o*bserved and suspect (i.e. has failed the latest QC check), or there are strong grounds for suspecting the accuracy of the observation*". The *status flag* is the third number in the 5-digit code.

In addition to the screening process described above, SD_{stn} observations that do not have a time stamp of 23:59 are excluded; 23:59 is the standard observing time for SD_{stn} but there are a few SD_{stn} data that are recorded in MIDAS Open with other times, e.g. 09:00, which is the standard observation time for precipitation, for example. The post-processing also includes a check for $SD_{stn} \geq$ the day length for each space-time location to ensure no unfeasible SD_{stn} values are included in the final station dataset. Finally, any stations with fewer than 365 daily SD_{stn} values are also excluded to minimise discontinuities in the time series and to ensure there are sufficient data for the additional QC that is applied in this study (see Section 4.3). This minimum number of observations threshold removes 88 stations from the dataset, although only 24 of these removed stations have more than 10 valid measurements.

[1] https://catalogue.ceda.ac.uk/uuid/2fd32ff2da0e4e718fdc23a9b90f791e

[2] https://catalogue.ceda.ac.uk/uuid/dd95b182eead4247a14c5911133e9030

[3] https://catalogue.ceda.ac.uk/uuid/778fc4752688475c8674f555d63bcf14

[4] https://catalogue.ceda.ac.uk/uuid/0cc854e249bf40bb8d47ac7c6f55d682

Where there is more than one SD_{stn} recording from different *met domains* on a single day the Met Office *Data Components Business Rules* are applied. These rules are detailed in an internal Met Office document (Data Components Business Rules v3, May 2007) and assign a priority to each *met domain*, advising the user to select the *met domain* with the highest ranking (ranking = 1). For SD_{stn}, the ranking is: DLY3028=1, NCM=1, AWSDLY=2, SYNOP=3. In the case where both DLY3028 and NCM messages are reported, the *prime_capability_flag* within MIDAS Open is used to determine which message should be used for this date. Where this decision cannot be resolved, the DLY3028 message is used by default; in practise, this only occurs for one station on two days (station ID 56230 in the Western Isles in the Outer Hebrides on 30 and 31 October 2007).

The MIDAS Open data set distinguishes between measurements of SD_{stn} made by Campbell-Stokes (CS) and Kipp & Zonen (KZ) devices. CS is a traditional instrument that consists of a glass ball that burns a trace on a card, while KZ is a more modern sensor that uses photo-diodes to determine when insolation exceeds 120 Wm^{-2}, which is the World Meteorological Organization (WMO) definition of bright sunshine (WMO, 2015). CS instruments are known to suffer from overburn, where the card continues to burn for a short time after a period of bright sunshine and has therefore been found to overestimate *SD* in the UK by up to 7% in winter and 20% in summer (Kerr & Tabony, 2004). Both sets of observations are used in this study, which enables a comparison when stations have measurements from both devices. Additionally, a station time-series can be extended, or infilled when data are missing, by converting between the two measurements. A conversion from KZ to CS was proposed by Legg (2014) where a quadratic fit was applied to the daily fraction of sunshine for the CS vs KZ relationship. In this study, this conversion is applied both ways, where the conversion from CS to KZ requires the inverse of this quadratic equation to be solved. Only one of the solutions returns sensible daily fractions of sunshine, where most values are between 0 and 1, and this is therefore the chosen solution. Due to the original quadratic solution proposed by Legg (2014) not passing directly through the origin, a very small CS measurement can result in a negative solution for the estimated KZ sunshine fraction. Where this occurs, the KZ daily sunshine fraction is set to zero. Additionally, the solution allows for the KZ sunshine fraction to go slightly above 1, but this only occurs when a CS device records a daily sunshine fraction > ~0.9 – an impossible scenario for the UK. The maximum possible *SD* that can be observed in the UK is ~19 hours, which is the number of hours of daylight on the day of the summer solstice at the UK's most northern point, Out Stack, Shetland at 60°51′N 0°52′W (https://gml.noaa.gov/grad/solcalc/). During winter, this maximum reduces to ~8 hours, which corresponds to the number of hours of daylight on the day of the winter solstice at the UK's most southern point, Pednathise Head, Western Rocks, Isles of Scilly at 49°51′N 6°24′W. The daily fractions of sunshine are translated into SD_{stn} by considering the number of daylight hours for the space-time location of the measurement being converted.

This process results in a quality-controlled time-series of SD_{stn} (in hours) for each station. Some stations have two associated time-series if SD_{stn} has been recorded using both CS and KZ devices. This study considers both CS-native (SD_{cs}) and KZ-native (SD_{kz}) time series, i.e. where no conversions have been applied, and CS-unified (SD_{cs_uni}) and KZ-unified (SD_{kz_uni}) time series. The unified time series are generated by applying the Legg (2014) conversion to produce 1) a single time series of CS data for all stations (SD_{cs_uni}), where any KZ data are converted to CS (SD_{kz_con}), and 2) a single time series of KZ data for all stations, where any

CS data are converted to KZ (SD_{cs_con}). For SD_{cs_uni}, SD_{cs} data are always used where available and SD_{kz_con} are only used to infill the time series where there are no SD_{cs} observations. Similarly, the SD_{kz_uni} time series preferentially uses SD_{kz} data. This results in either two or four SD_{stn} time series for each station used in this study, depending on whether a station has data from either CS or KZ sensors, or both CS and KZ sensors, respectively.

2.2 HadUK-Grid SD (SD_{had})

The HadUK-Grid data set is a collection of climate variables which have been gridded through interpolation of UK station observations (Perry & Hollis, 2005). The current suite of variables incudes air temperature (maximum, minimum and mean), precipitation, *SD*, mean sea level pressure, wind speed, relative humidity, vapour pressure, days of snow lying, and days of ground frost. The data are provided at daily, monthly, seasonal and annual temporal resolution, although not all variables are available on all timescales. Data derived for specific climatological reference periods are also available. The HadUK-Grid extends to the present day from 1836, although the exact start date depends on the variable and the temporal resolution. The gridded data are available for the UK at a range of spatial resolutions up to 1 km x 1 km, although again, this also depends on the variable.

The data set used here is sourced from the HadUK-Grid v1.2.0.0 (dated v20230328), which provides monthly *SD* at 5km x 5km resolution over UK land masses for the period 1910 to 2022 (SD_{had}) (Met Office, 2018). The SD_{had} data are not available at higher temporal and spatial resolutions owing to the sparsity of the SD_{stn} network, which has declined over the past few decades and currently consists of around 100 stations (Figure 1-1). The input SD_{stn} observations comprise both SD_{cs} and SD_{kz}, although the SD_{kz} are adjusted to align with observations from the CS instrument using the correction derived by Legg (2014). The choice to use the CS instrument as a reference within SD_{had} is because CS observations comprise the majority of the historical UK SD_{stn} record and are still the dominant sensor today. To produce SD_{had}, the SD_{cs_uni} data are gridded using a three-step process of normalisation, regression on geographic features, and interpolation of the regression residuals. The data are normalised by dividing by the long-term average, and then regressed on the easting, northing, and percentage of water within a 5km radius of the station. A cubed inverse distance weighting method is used to interpolate the residuals (Perry & Hollis, 2005). The SD_{had} data are used as provided and no additional quality screening is applied in this study.

2.3 CM SAF SARAH-3 SD (SD_{sat})

The EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) produces the Surface Solar Radiation Data Set – Heliosat; version three is used in this study (SARAH-3), which is the most recent version at the time of writing (Pfeifroth et al., 2023). The *SD* variable in the SARAH-3 data set (SD_{sat}) is provided over both land and water. SD_{sat} is calculated from the number of 30-minute observations per day where the Direct Normalised Irradiance (DNI) is greater than or equal to 120 W/m^2, which is the WMO definition of bright sunshine (WMO, 2015):

Equation 2-1

$$SD_{sat} = daylength \times \frac{\sum_{i=1}^{\#daylight\ slots} Wi}{\#daylight\ slots}$$

where *#daylight slots* represents the maximum number of 30-minute slots possible during daylight hours for that space-time location where the solar elevation angle exceeds 2.5°. W_i is a weighting function between 0 and 1 that indicates the 'sunniness' of a single slot based on both the DNI of that grid box and the DNI of the 24 surrounding sunny/cloudy grid boxes from both the current and previous slot (Kothe et al., 2017). This accounts for variation in SD that may occur during a single 30-minute slot due to moving, broken cloud.

The SARAH-3 DNI estimates that underpin the SD_{sat} are derived using data from the visible channels of the Meteosat Visible Infra-Red Imager (MVIRI) and Spinning Enhanced Visible Infra-Red Imager (SEVIRI) instruments onboard the geostationary Meteosat First and Second Generation (MFG & MSG) platforms (Pfeifroth et al., 2023). Meteosat is the operational weather satellite for Europe and Africa, although parts of S. America are also captured towards the edge of the western field of view. The SARAH-3 dataset extends from 1 January 1983 to the present day, although the data from 1 January 2021 are provided as an Interim Climate Data Record (ICDR), which are available within 5 days of acquisition. Data are available for the region ±65° longitude and ±65° latitude at a spatial resolution of 0.05° x 0.05° latitude-longitude. The SD_{sat} variable, termed SDU in the SARAH-3 dataset/documentation, is provided as both daily and monthly sums, although only the daily sums are utilised in this work (Pfeifroth et al., 2023).

Kothe et al. (2017) reported good agreement between the CM SAF *SD* and in situ station observations, although they found that the SD_{sat} tends to slightly overestimate *SD* compared to the station *SD* observations. They obtained a mean daily satellite-station difference over Europe of 0.41 hours, a mean absolute difference (MAD) of 1.31 hours and a correlation (*r*) of 0.91. For monthly SD_{sat} data over the Meteosat disc (Europe, Africa and S. America), they obtained a mean satellite-station difference of 7.5 hours, a MAD of 18.4 hours and an *r* value of 0.96. Kothe et al. (2017) also reported that the largest-magnitude satellite-station differences were found in mountainous regions and on islands.

The CM SAF SARAH-3 *SD* product provides an in-built quality flag, the *record status*. The record status flag is a single value applied to all grid-boxes on a given day i.e., a rejection of the data due to the flag results in losing the entire day. The record status can be one of three possible labels: "ok", "void", and "bad quality". For this study, data are only retained if the flag is labelled as "ok" (a numerical value of 0 in the data set). The SD_{sat} data are used as provided and no additional quality screening is applied in this study.

3 Methods

The following sub-sections describe the study approach, which will ultimately result in the production and evaluation of a blended satellite-station SD dataset for the UK (SD_{bld}). The first step in developing SD_{bld} is to characterise the relationship between the input SD_{sat} and SD_{stn} datasets, for example:

- What are the differences between SD_{sat} and SD_{stn} and does this vary between SD_{cs} and SD_{kz}?
- How well does the SD_{kz} to SD_{cs} correction (and vice-versa) perform?
- Is there any spatial and/or temporal variation in the SD_{sat}-SD_{stn} differences?
- Are there any non-climatic discontinuities in either the SD_{stn} and/or SD_{sat}?
- What SD_{stn} and/or SD_{sat} data, if any, should be excluded from SD_{bld}?
- Where is SD_{bld} likely to have larger errors that may result from the input SD_{stn} and/or SD_{sat}?

This characterisation is achieved by comparing SD_{sat} and the 'point' SD_{stn} data directly; the approach used here is described in Section 3.1.1.

The overall goal for this work is to enable production of both SD_{bld} and SD_{sat} for the UK that complement the existing HadUK-Grid monthly SD dataset (SD_{had}), which is based only on SD_{cs_uni}. Therefore, a comparison between SD_{sat} and SD_{had} is also performed to explore the differences and similarities between these datasets so that they can be communicated to users; this method is outlined in Section 3.1.2.

For both sets of comparisons, a suite of standard statistical metrics is used, which are defined in Section 3.1.3.

Finally, the methods used to create the SD_{bld} data are described in Section 3.2. However, it should be noted that both the approach and the results presented for this component of the study are preliminary and the SD_{bld} dataset will be developed and refined further in future work.

3.1 SD dataset intercomparison

3.1.1 Comparison of SD_{sat} and 'point' station SD_{stn}

This analysis is based on spatially and temporally colocated daily observations. Each 'matchup' pair consists of an SD_{stn} observation and the corresponding 0.05°x0.05° SD_{sat} grid cell observation that nominally contains the specified latitude and longitude of SD_{stn} on the same date. Separate comparison results are presented for SD_{sat} vs SD_{cs} and SD_{sat} vs SD_{kz} to establish whether the satellite data are more closely aligned with the CS or KZ *SD* observations. As part of this analysis, temporally colocated SD_{cs} and SD_{kz} are compared for stations with both sensor types (an 'overlap' analysis). The performance of the Legg (2014) KZ→CS *SD* correction is also assessed. (It should be noted that the number of *SD* matchup pairs for the overlap analysis may be fewer than the minimum threshold of 365 valid daily observations required to include a given station in the study database (see Section 2.1). No additional threshold screening is applied in the overlap analysis.) In addition to characterising

the SD_{sat} vs SD_{stn} agreement, the primary objective for this analysis is to inform the choice of SD_{stn} data that is used to generate SD_{bld}.

3.1.2 Comparison of SD_{sat} and station SD_{had}

This analysis is based on spatially and temporally colocated SD_{sat} and SD_{had} observations. To compare the monthly SD_{had} with daily SD_{sat} requires equivalent monthly totals of SD_{sat}, which are produced by summing the daily SD_{sat} over all available days. Following the approach of Kothe et al. (2013), a month is discarded if there are more than three missing days of data; in the event of one to three missing days, these are in-filled with the mean daily SD_{sat} for that month.

The HadUK-Grid SD_{had} is on a 5kmx5km grid while the CM SAF SD_{sat} product on a 0.05°x0.05° grid. In order to make these data sets compatible for comparison, both are re-gridded on to a coarser 0.1°x0.1° grid. This is performed utilising the *regrid* function from the Iris Python package, which offers a number of geostatistical options. The re-gridding method chosen for this study uses a linear interpolation[5] where "*points [are] calculated by extending the gradient of the closest two points*".

3.1.3 Comparison statistics

The agreement between the different datasets is quantified through the use of several standard statistical metrics. The mean difference, μ, (also frequently reported in the literature as 'mean error' or 'bias') is the arithmetic mean of the colocated satellite-minus-station/HadUK-Grid differences and is defined as:

Equation 3-1

$$\boldsymbol{\mu = \frac{1}{n}\sum_{j=1}^{n}\left(SD1_j - SD2_j\right)}$$

Where *n* is the total number of matchups (i.e. *SD* pairs), and *SD1* and *SD2* are the colocated *SD*s (i.e. SD_{sat} – SD_{stn}). The sample standard deviation, σ, (also frequently reported in literature as 'precision' or 'bias-corrected root mean square error') is a measure of the spread around the mean value of the distribution of differences. It is defined as:

Equation 3-2

$$\boldsymbol{\sigma = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(\left(SD1_j - SD2_j\right) - (\mu)\right)^2}}$$

A deficiency of the standard deviation is that it assumes that the sample distribution is Gaussian, which is often not the case. Therefore, the percentiles of the distribution of differences are also reported in this study. The 50th percentile, i.e. the value above which (and below which) exactly 50% of the samples occur, is equivalent to the median difference.

[5] https://scitools-iris.readthedocs.io/en/latest/generated/api/iris.analysis.html#iris.analysis.Linear

3.2 Producing the SD_{bld} dataset

The approach used to produce a blended satellite-station *SD* data set in this study follows the work of Walawender (2017), who used regression-kriging to produce a 1 km blended satellite-station *SD* dataset for Germany using the CM SAF SARAH-2 dataset. This study uses a very similar approach but replaces the kriging component of the blending with a Gaussian Process (GP) model. A GP model is chosen over kriging as it has the flexibility to include point-by-point variance, for example, a variation in the measurement uncertainty at each station. While this is not currently implemented in the blending process, it is likely to be added in the next developmental phase of this work. The Scikit-learn Python tool *gaussian_process.GaussianProcessRegressor* is used in this study (Pedregosa et al., 2011).

The SD_{bld} dataset developed here is based on the inputs SD_{sat} and SD_{stn}. To blend these datasets, two distinct operations are performed. First, a regression model is defined for the relationship between SD_{sat} and SD_{stn}. In this preliminary analysis, three regression models are tested:

Equation 3-3 $$SD_{stn_1} = c_1.SD_{sat} + d_1 + \varepsilon_1$$

Equation 3-4 $$SD_{stn_2} = b_2.lat + c_2.SD_{sat} + d_2 + \varepsilon_2$$

Equation 3-5 $$SD_{stn_3} = a_3.SD_{sat}^{\,2} + b_3.lat + c_3.SD_{sat} + d_3 + \varepsilon_3$$

where *a*, *b*, *c* and *d* are the regression coefficients, *lat* is the latitude of the grid cell corresponding to SD_{sat}, and ε is the error on the estimated SD_{stn}. The regression models are then applied to the SD_{sat} dataset to predict SD_{stn} for all grid cells with valid observations. The model fit can include multiple additional variables as predictors (i.e. further variables on the right hand side of Equation 3-3 to Equation 3-5), such as latitude, elevation, and distance to coast. However, in this initial work, only SD_{sat} and latitude are included as predictors, but additional variables will be trialled in the next developmental phase of this study. Additionally, the flexibility of the models is adapted by fitting a Generalised Linear Model (GLM) in which the error structure can be defined with a different distribution. In this case, a Gamma error distribution with an identity link function is used to prevent the *SD* estimate uncertainty from going below zero. In order for this model to be fitted successfully, data points where zero hours of *SD* were measured for both SD_{stn} and SD_{sat} were removed. Future work will investigate alternative methods i.e. transformations, to help retain data. Separate models are produced for each day of the year using a moving 5-day window, e.g. the model for 3rd January is defined using *SD* observations from 1st to 5th January from all years, the model for 4th January is defined using *SD* observations from 2nd to 6th January from all years, etc. The given model estimates the SD_{stn} for the 0.05° grid cell containing each station location on a daily timescale, provided the necessary explanatory variables (or predictors) are available. From these estimates the residuals are calculated, i.e. actual SD_{stn} – estimated SD_{stn} for each station on each date. An initial assessment of model performance is performed using the residuals based on statistics such as the root mean square error (RMSE) and (pseudo-) R^2. The pseudo R^2 is a 'goodness of fit' statistic that can be used to compare output from GLMs; in this study the Cox-Snell method is used which calculates the likelihood ratio of a model with only a constant (i.e. an intercept term only) and the model with additional predictors. This first step concludes with an estimated SD_{stn} for each valid grid cell for each day in the SD_{sat} dataset, together with residuals at each station location which can be used in the second step in the blending process.

The second step utilises a GP model. The GP is a model where functions, drawn from a multivariate Gaussian distribution, are conditioned on the available data. A type of covariance matrix can be derived through the use of a kernel, which the model interprets as how similar the data are over varying length scales. The result is a distribution of functions where the mean function is used to estimate the variable of interest and the distribution parameters provide confidence intervals.

In this study, the residuals are used in a GP model to estimate a residual field and update the GLM estimates daily. For each day the GP model is provided with the latitude, longitude and *SD* residual (actual SD_{stn} – estimated SD_{stn}) at each available station grid cell, and a radial basis function kernel (RBF) kernel. A correlation length scale (CLS) hyperparameter of 2.0 (Euclidean Distance) is currently used for the GP model, although this will be re-estimated during the next SD_{bld} developmental phase from the correlation between each pair of UK stations. Additionally, a fixed standard deviation hyperparameter of 1.50 hours is applied to the distribution of functions, which aids in finding a balance between under- and over-fitting the model. This value was derived from the data and is roughly the standard deviation of the GLM-estimated *SD* residuals but will also be tuned during the next developmental stage of this work. Lastly, a spatially- and temporally-uniform variance is applied to all station residuals in the GP model, which is defined using the GLM fit statistics. This avoids forcing the GP model to fit to the exact values of the residuals, which can result in spurious output and a smoother fit. In the next SD_{bld} developmental phase, it is likely that a station-by-station variance will be defined, which will account for the different measurement uncertainties at each station. This will enable stations with a lower estimated uncertainty to have a greater influence on the GP model output.

To evaluate the accuracy of the output SD_{bld} dataset, a K-fold cross validation is performed to provide some independent output statistics to assess the goodness of fit. The stations are split into *K* subsets (or folds) where for *K* model runs each subset of stations are excluded from the GLM and GP model training and retained for validation. This study uses a 15-fold cross-validation, providing 28 or 29 stations for validation in each fold. Validation statistics are calculated for each fold (e.g. RMSE, mean, percentiles), which are averaged across all folds to create the performance metrics for each regression-GP model combination. A final GLM-GP model is then defined using data from all stations (i.e. none are excluded), which is used to generate the SD_{bld} output data for each day.

The set-up of the GP model is that on a given day the ‘correction’ field tends to zero for locations where there are no station residuals, and the surrounding station density is such that the inter-station distance is significantly greater than the CLS. In the extreme case, where there are no valid station observations on a given day at any location, the GP correction field will be zero everywhere and therefore the output from the GP model, i.e. the second stage in the blending process, will be identical to the output from the GLM stage of the blending process. It is possible that ‘bullseye’ artifacts (i.e. circular features that originate from a sizeable single station residual) can occur in the SD_{bld} if a station with an outlying residual is relatively isolated, however these can be managed by choice of hyperparameter.

It should be noted that even for days where there are no valid station observations, there will always be an output (i.e. predicted SD_{stn}) from the GLM stage of the blending provided the necessary predictor variables are available on that day. This is because the GLM is trained on data for all available years for a 5-day moving window and therefore the GLM coefficients

exist for every calendar date. With this in mind, the data fields specified in Table 3-1 are proposed for the final blended *SD* dataset. Users can then choose which output is best suited to their application based on the dataset documentation and guidance.

Table 3-1: Data fields proposed for the new UK SD_{bld} dataset. All fields will be provided at daily and monthly temporal resolution and at 0.05°x0.05° latitude-longitude except for the 'SD_{bld}: number of stations' data field, which will contain a single value per day.

Variable	Description
SD_{sat}	*SD* provided by the CM SAF SARAH-3 dataset but just for the UK land (to match HadUK-Grid).
SD_{bld}: GLM only	*SD* output by the (final-selected) GLM model used in the GLM-GP blending process, e.g. in situ-equivalent predicted *SD* for each grid cell based on SD_{sat} and other predictors.
SD_{bld}: GLM-GP	*SD* output by the (final-selected) GLM model with the GP correction applied, i.e. the SD_{bld} from the GLM only data field above + GP correction.
SD_{bld}: nearest GP station	The linear distance to the nearest grid cell with a valid station observation for the given day, expressed as a fraction of the CLS. For grid cells that contain a valid station observation on that day, this value will be 0 for that grid cell. For grid cells that are a distance of >CLS from the nearest valid station observation this value will be >1, and the GP correction field value will tend to 0 at large fractions. This provides an overall QC indicator for the grid cell.
SD_{bld}: number of stations	The number of stations with a valid SD_{stn} observation on the given day for the whole of the UK. This provides an overall QC indicator for the 'SD_{bld} : GLM-GP' data field for the given day. If this field value is 0 the fields 'SD_{bld}: GLM only' and 'SD_{bld}: GLM-GP' will be identical. If the field value is 20, for example, the GP correction field is likely to contain many zeros and the 'SD_{bld}: GLM-GP' may suffer from 'bullseye' artifacts (see pg 17).
SD_{bld}: uncertainty	A field, or combination of fields, representing the estimated uncertainty in SD_{bld} for each grid cell.

4 Results

4.1 Examples of *SD*

Figure 4-1 shows examples of each of the datasets used in this study for 1 June 2003 (SD_{stn} and SD_{sat}) and summer 2003 (SD_{had} and SD_{sat}). In general, there is good visual agreement between the datasets. In the daily data (top row), for example, the areas with low *SD* in the southeast, Northern Ireland and western Scotland are captured in both the SD_{stn} and SD_{sat} datasets. However, there are also some differences, for example, the SD_{stn} values in eastern Scotland seem to be lower than the SD_{sat}.

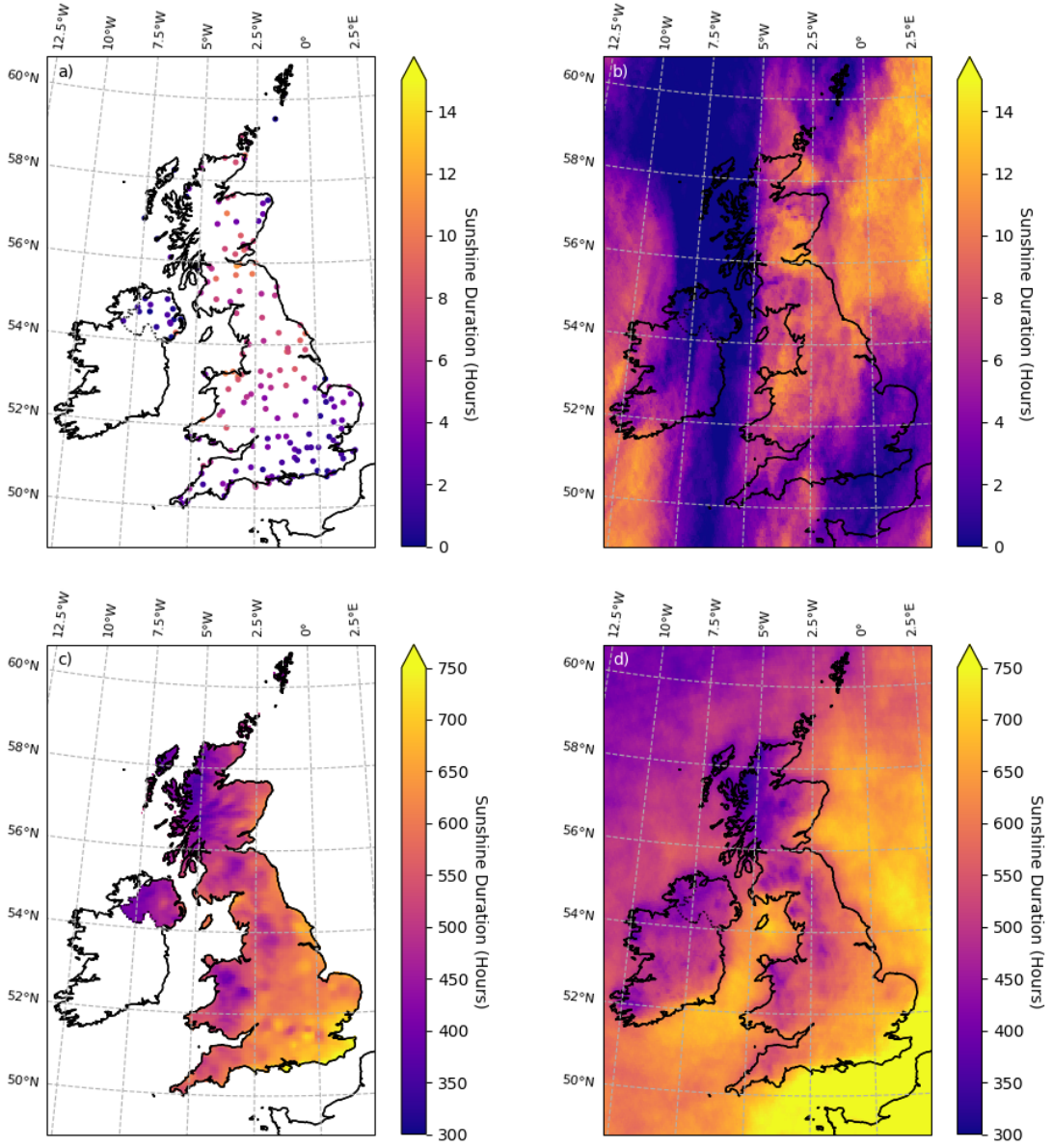


Figure 4-1: Examples of each dataset used in the study a) daily SD_{cs_uni} and b) daily SD_{sat} on 1 June 2003, c) SD_{had} and d) SD_{sat} for June/July/August. See Section 2 for details on each dataset.

For the summer 2003 example (bottom row), the SD_{had} data appear to show more spatial structure than the SD_{sat}. In particular, there are some circular features in southern and eastern parts of the UK in the SD_{had} data that do not appear in the SD_{sat} data. These circular features are known artifacts in the SD_{had} dataset and result from differences between neighbouring stations/regions that the interpolation process cannot reconcile (NCIC, Personal Communication). Of particular note is a circular region of higher *SD* in central Northern Ireland in the HadUK-Grid data (panel c), which appears to correspond to the freshwater lake, Lough Neagh. In the satellite *SD* data there is a corresponding less-well defined circular region of lower *SD* (compared to the surrounding *SD* values). Therefore, it seems likely that the *SD* estimates from both the satellite and gridded station data may be affected by the presence of a large water body. In general, the SD_{sat} data provide a smoother field compared with the SD_{had}.

The SD_{had} and SD_{sat} summer data show many comparable features that are consistent with the expectations for the natural variability in *SD*. For example, the higher *SD* along the south coast, a general northwest-southeast gradient in the *SD* and consistently lower *SD* over areas with higher orography, for example over the Pennines in northern England and the Grampians in Scotland. The magnitude of the *SD* values is also comparable: the highest values in both datasets exceeds ~750 hours in the southeast, while the lowest *SD* values fall below ~400 hours in the northwest of the UK and over areas of higher orography.

4.2 Comparison between daily CS-native (SD_{cs}) and KZ-native (SD_{kz}) *SD*

Both the SD_{stn} and SD_{had} data shown in Figure 4-1 are aligned with observations from CS sensors, where any data from KZ sensors are converted to the equivalent CS data to ensure all UK SD_{stn} data are unified (see Section 2). As unified SD_{stn} data are also required as input for SD_{bld}, the first step in generating this new dataset is to establish:

1. Which SD_{stn} dataset agrees better with SD_{sat}, therefore informing the choice of 'baseline' SD_{stn} data to use in generating the blended satellite-station *SD* dataset, and
2. The performance of the Legg (2014) correction for converting between the two, i.e. KZ→CS and CS→KZ, therefore enabling both types of observations to be used in SD_{bld}.

Figure 4-2 shows the comparison between SD_{cs} and SD_{kz} observations for stations that have periods where both types of observations are reported during an instrumental 'overlap' period. The distribution for all data (top left in Figure 4-2 and Table 4-1) implies that SD_{kz} is on average ~0.4 hours less sunny than SD_{cs}, which is consistent with previous studies that compared these types of sensors and derive a correction between the two (Kerr & Tabony, 2004; Legg, 2014). As reported in Section 2.1, the CS instrument tends to suffer from overburn, which leads to overestimates of *SD*. However, when the SD_{kz} - SD_{cs} differences are partitioned by SD_{cs}, it is clear these differences vary with *SD*. The mean difference (μ) becomes increasingly negative ranging between μ=0.0 for $0 \leq SD_{cs} < 2$ hours and μ=-0.9 for $8 \leq SD_{cs} < 10$ hours, before becoming increasingly more positive up to the maximum *SD* ($16 \leq SD_{cs} < 18$ hours), where μ=-0.2 hours. The partitioned distributions are reasonably symmetrical for $SD_{cs} < 6$ hours but also become increasingly left-skewed with increasing SD_{cs}. This is reflected in the overall distribution, which is also left-skewed.

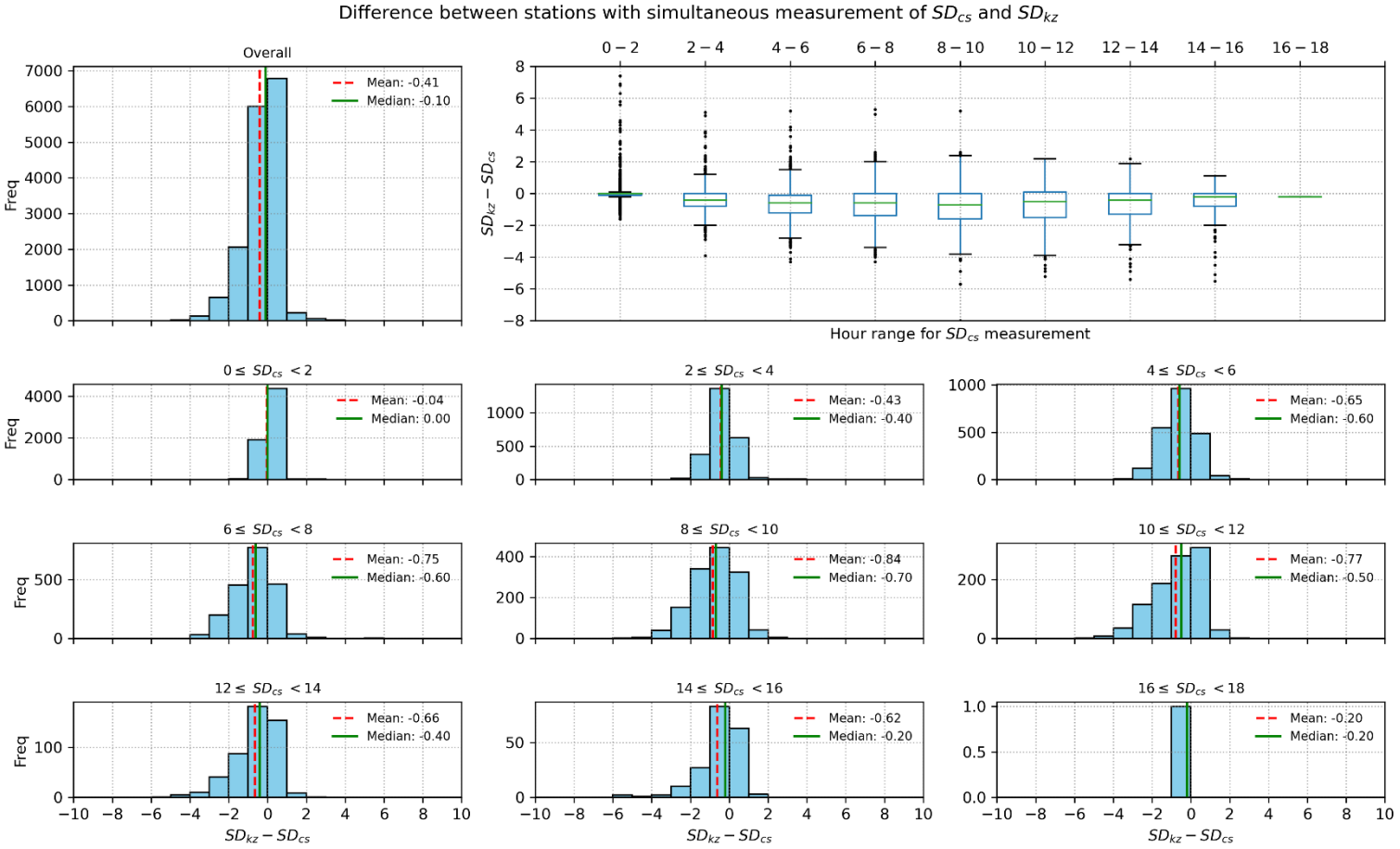


Figure 4-2: Distributions of daily $SD_{kz} - SD_{cs}$ differences (i.e. KZ-native minus CS-native *SD*). The top left panel shows the distribution for all machups, while the other panels show the distributions of differences partitioned by SD_{cs} values. The top-right panel shows the distributions for all partitioned ranges where the central green line corresponds the 50th percentile (median) value, the box shows the 25th and 75th percentiles (or inter-quartile range), while the whiskers extend to the farthest data point within 1.5x the inter-quartile range from the box. The data points that extend past the end of the whiskers are the most extreme values.

Table 4-1: Statistics showing the agreement between SD_{kz} and SD_{cs} (top, i.e. KZ-native and CS-native) and SD_{kz_con} and SD_{cs} (i.e. KZ-converted to CS and CS-native) for all ('All'), winter (DJF), spring (MAM), summer (JJA) and autumn (SON) daily colocated matchups. The results below correspond to the data shown in Figure 4-2 and Figure 4-3.

Data subset	No. Data	*r*	*μ* (hours)	*σ* (hours)	Percentiles (hours) 1st	5th	25th	50th	75th	95th	99th
					SD_{kz} - SD_{cs}						
ALL	15961	0.98	-0.4	0.9	-3.0	-2.1	-0.7	-0.1	0.0	0.5	1.4
DJF	3945	0.97	-0.1	0.6	-1.7	-1.0	-0.3	0.0	0.0	0.5	1.6
MAM	4040	0.97	-0.5	1.0	-3.3	-2.3	-1.0	-0.2	0.0	0.6	1.6
JJA	3894	0.97	-0.8	1.0	-3.4	-2.5	-1.3	-0.5	0.0	0.3	1.0
SON	4082	0.98	-0.3	0.7	-2.3	-1.4	-0.5	-0.1	0.0	0.6	1.4
					SD_{kz_con} - SD_{cs}						
ALL	15961	0.98	0.0	0.8	-2.0	-1.2	-0.2	0.0	0.3	1.2	2.0
DJF	3945	0.98	0.0	0.5	-1.4	-0.7	-0.1	0.0	0.1	0.7	1.8
MAM	4040	0.98	0.0	0.9	-2.3	-1.4	-0.4	0.1	0.4	1.3	2.1
JJA	3894	0.98	0.1	0.9	-2.2	-1.4	-0.4	0.2	0.6	1.3	2.0
SON	4082	0.98	0.0	0.6	-1.6	-0.9	-0.2	0.0	0.3	1.0	1.8

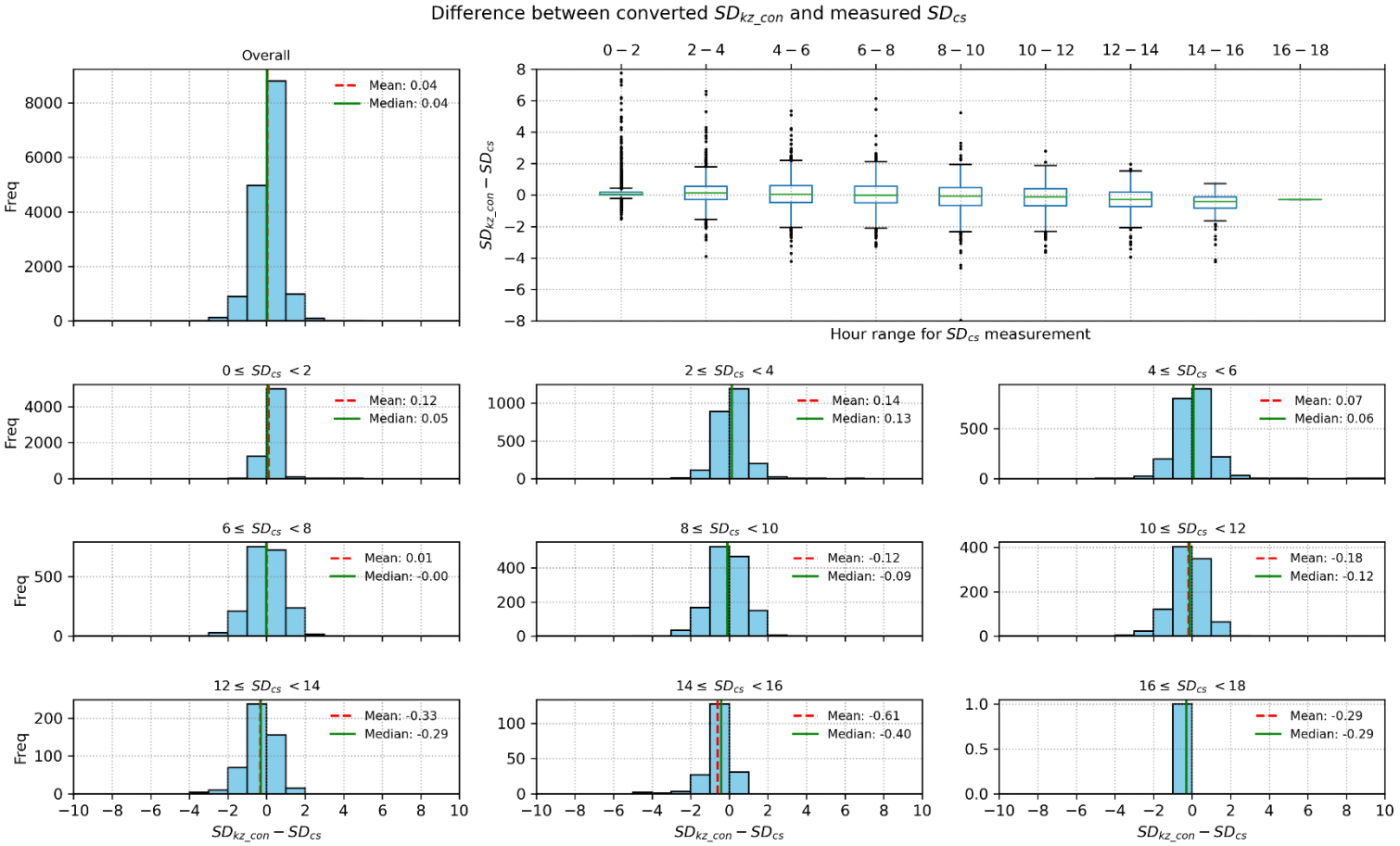


Figure 4-3: Distributions of daily $SD_{kz\_con} - SD_{cs}$ differences (i.e. corrected KZ minus CS-native *SD*). The top left panel shows the distribution for all machups, while the other panels show the distributions of differences partitioned by SD_{cs} values. The top-right panel shows the distributions for all partitioned ranges where the central green line corresponds the 50th percentile (median) value, the box shows the 25th and 75th percentiles (or inter-quartile range), while the whiskers extend to the farthest data point within 1.5x the inter-quartile range from the box. The data points that extend past the end of the whiskers are the most extreme values.

Figure 4-3 shows the comparison between SD_{cs} and SD_{kz_con} (i.e. KZ *SD* that has been corrected to CS *SD*) for the same observations that are shown in Figure 4-2. Across all observations, μ has reduced to ~0.0 hours and the standard deviation (σ) has also reduced by ~0.1 hours (Table 4-1). However, there is still some variation in μ with SD_{cs}, although this is closer to zero and reduced in magnitude compared to the SD_{cs} vs SD_{kz} comparison. The values of μ now range between -0.2 and 0.1 hours for $0 \leq SD_{cs} < 12$, which encompasses the vast majority of the observations. For higher values of SD_{cs} ($12 \leq SD_{cs} < 18$ hours), μ varies between -0.3 and -0.6 hours, suggesting that the Legg (2014) correction is less accurate for this *SD* range. These results demonstrate that the $SD_{kz} \rightarrow SD_{kz\_con}$ conversion (and vice-versa) proposed by Legg (2014) is effective and can be used to unify the SD_{stn} observations for the UK before blending with SD_{sat}. However, there are likely to be larger errors in the conversion for values of *SD* above 12 hours.

4.3 Comparison between daily station (SD_{stn}) and satellite (SD_{sat}) *SD*

4.3.1 CS-native (SD_{cs}) and KZ-native (SD_{kz})

Having characterised the differences between SD_{cs} and SD_{kz} and verified that the Legg (2014) KZ→CS (and vice-versa) conversion can be used to generate a unified SD_{stn} dataset, the next step is to ascertain which sensor should be adopted as the baseline as input to SD_{bld}. Figure 4-4 and Figure 4-5 show the distributions of daily SD_{sat} - SD_{stn} differences for CS-native and KZ-native station observations, respectively (i.e. *SD* observations that have not yet been converted from CS to KZ or vice versa). Notably, there are considerably fewer matchups for SD_{kz} (n=350,639) compared with SD_{cs} (n=2,367,058), which is expected as the UK station network is dominated by CS sensors (Section 2.1). The mean SD_{sat} - SD_{stn} difference (μ) and standard deviation (σ) for the overall distributions (i.e. all data) is 0.1 and 1.6 hours for the SD_{cs}, and 0.5 and 1.2 hours for SD_{kz}. However, it should be noted that both distributions are more Laplacian in shape than Gaussian, and therefore the standard deviation should only be regarded as an indicative measure of the spread. This analysis suggests that the satellite data are generally better aligned with the CS observations, although the SD_{sat} vs SD_{stn} correlation (r) is slightly higher for the SD_{kz} (r = 0.95) compared with SD_{cs} (r = 0.92). This is the opposite result to that obtained by Good (2010), who found that SD_{kz} agreed better than SD_{cs} with daily SD_{sat} for the UK derived from sub-daily Meteosat cloud type data. CS observations are known to overestimate *SD* as they suffer from overburn, when the card continues to burn even after the sun is obscured by cloud (Kerr & Tabony, 2004; Legg, 2014). Therefore, the better agreement between the CM SAF SD_{sat} data and SD_{cs} obtained in this study may be fortuatous as Kothe et al. (2017) found that these satellite data also slightly overestimated *SD* compared with SD_{stn} over Europe.

In addition to the distributions for all data, Figure 4-4 and Figure 4-5 also show the distributions of differences partioned by the SD_{stn} observations in 2-hour intervals. As expected based on the high mid-latitude climate in the UK, the largest number of observations fall within the 0-2 hour range. The mean SD_{sat}-SD_{cs} difference for this *SD* range is 0.3 hours, indicating that SD_{sat} tends to be higher than SD_{cs} for days with no or very low *SD*. By contrast, the mean differences for the two highest *SD* categories ($14 \leq SD_{cs} < 18$) are quite strongly negative (μ = -0.6 and -1.2 hours) indicating that SD_{sat} tends to be lower than SD_{cs} for the sunniest days in

the UK. The results for these most extreme *SD* categories are comparible for the SD_{sat}-SD_{kz} differences therefore demonstrating that the performance of the CS and KZ instruments is similar under these conditions (μ = 0.3 hours for $0 \leq SD_{kz} < 2$ hours, and μ = -0.4 and -1.0 hours for the $14 \leq SD_{kz} < 16$ and $16 \leq SD_{kz} < 18$ hours, respectively). However, it should be noted that the number of matchups for these highest *SD* categories is small.

The mean SD_{sat}-SD_{cs} differences for the partitioned distributions within the 2-14 hours *SD* window are consistently smaller (μ = -0.1 to 0.1 hours) than those for the $0 \leq SD_{cs} < 2$ hours and $14 \leq SD_{cs}$ bins, indicating a closer agreement between the CS and satellite observations in this *SD* range. By contrast, the equivalent mean SD_{sat}-SD_{kz} differences are typically around 0.6 hours (μ = 0.2 to 0.8 hours), i.e. SD_{sat} tends to be higher than SD_{kz} for these mid-to-high *SD* days. This curved response in the variation of μ with *SD* for the SD_{sat}-SD_{kz} comparisons can be seen clearly in the top right-hand panel of Figure 4-5. This mirrors the equivalent curved response that is apparent in Figure 4-2, which shows the SD_{cs} – SD_{kz} differences (i.e. the uncorrected KZ observations).

As observed for the SD_{kz} vs SD_{cs} comparisons, the skew of the partitioned distributions for SD_{sat} - SD_{cs} varies with SD_{cs}. The distributions evolve from having right-skew for $0 \leq SD < 4$ hours, through to left-skew for $8 \leq SD < 14$ hours, with remarkably symmetrical distributions for $4 \leq SD < 8$. By contrast, the SD_{sat} - SD_{kz} distributions are right-skewed for all two-hour *SD* ranges below $12 \leq SD$ hours.

The curved response in the variation of μ with *SD* also manifests as a seasonal variation in the agreement between SD_{sat} and SD_{stn}, where the mean SD_{sat} – SD_{kz} daily differences are smaller in magnitude during the winter (μ = 0.3 hours) and largest in the summer (μ = 0.7 hours). The standard deviations (σ) follow the same pattern, varying between 1.0 hours in winter and 1.5 hours in summer. For the SD_{sat} vs SD_{cs} comparison, σ is also smallest in the winter (σ = 1.3 hours) and largest in the summer (σ = 1.8 hours). However, μ = 0.1 hours for all seasons. This seasonal variation in the SD_{sat} – SD_{kz} and SD_{kz} – SD_{cs} differences is a direct result of the increased tendency of SD_{cs} and SD_{sat} to overestimate the *SD* for mid-to-high range values of *SD*, which are more likely to occur during the UK summer.

Based on the overall better agreement between SD_{sat} and SD_{cs} compared with SD_{kz}, the CS sensor is selected as the reference for the SD_{stn} data that are input into the SD_{bld} dataset. Any valid SD_{kz} data are converted to CS-equivalent SD (Section 4.2), and therefore SD_{cs_uni} data (uncorrected CS and corrected KZ, i.e. SD_{cs} & SD_{kz_corr}) are used with SD_{sat} to generate the SD_{bld} dataset. Selection of CS as a reference is also advantageous as this is the baseline that is also used for the SD_{had} dataset, therefore ensuring that the SD_{bld} and SD_{sat} datasets are more consistent with the current operational Met Office *SD* product for the UK.

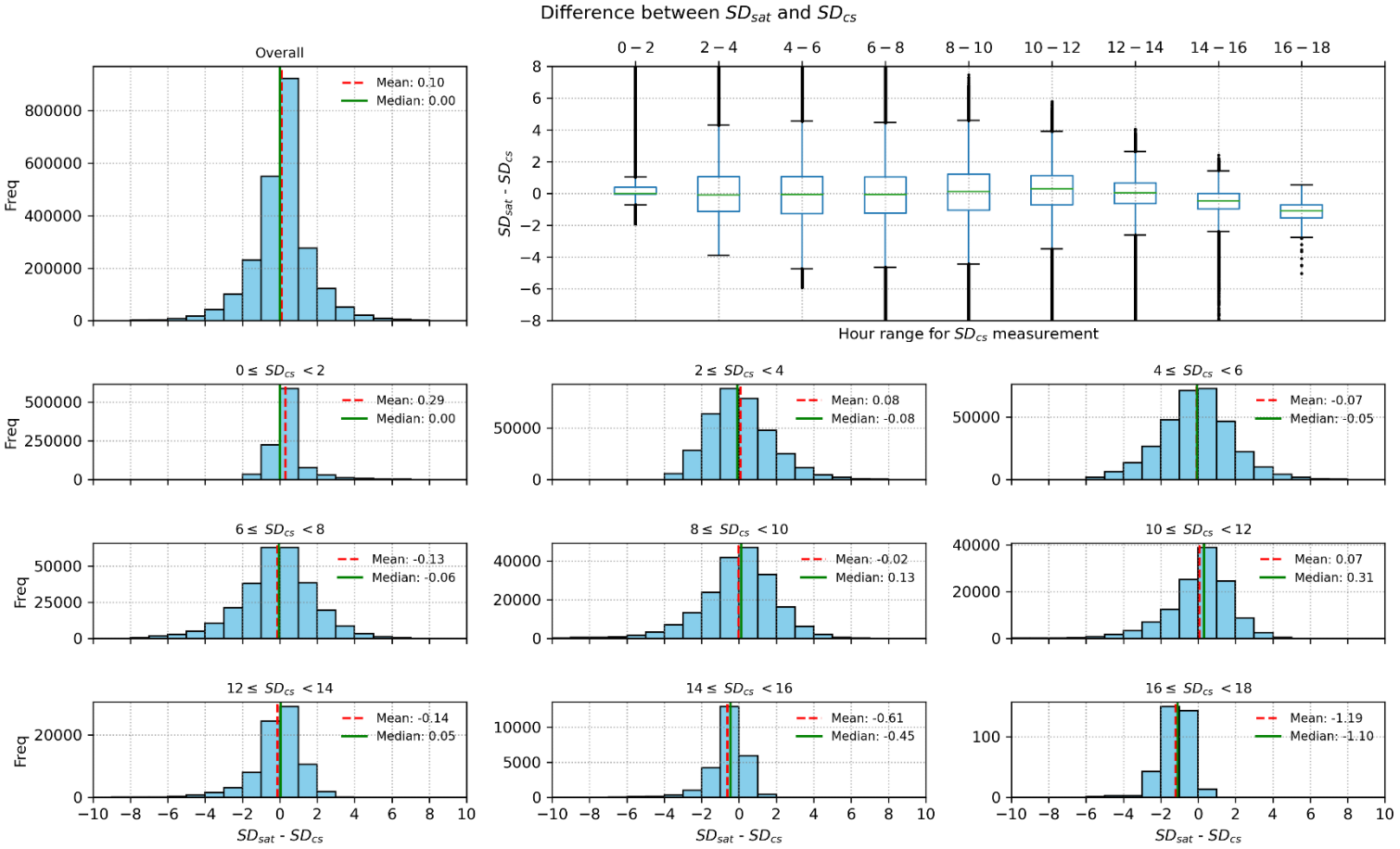


Figure 4-4: Distributions of daily SD_{sat} – SD_{cs} differences (CS-native). The top left panel shows the distribution for all machups, while the other panels show the distributions of differences partitioned by SD_{cs} values. The top-right panel shows the distributions for all partitioned ranges where the central green line corresponds the 50th percentile (median) value, the box shows the 25th and 75th percentiles (or inter-quartile range), while the whiskers extend to the farthest data point within 1.5x the inter-quartile range from the box. The data points that extend past the end of the whiskers are the most extreme values.

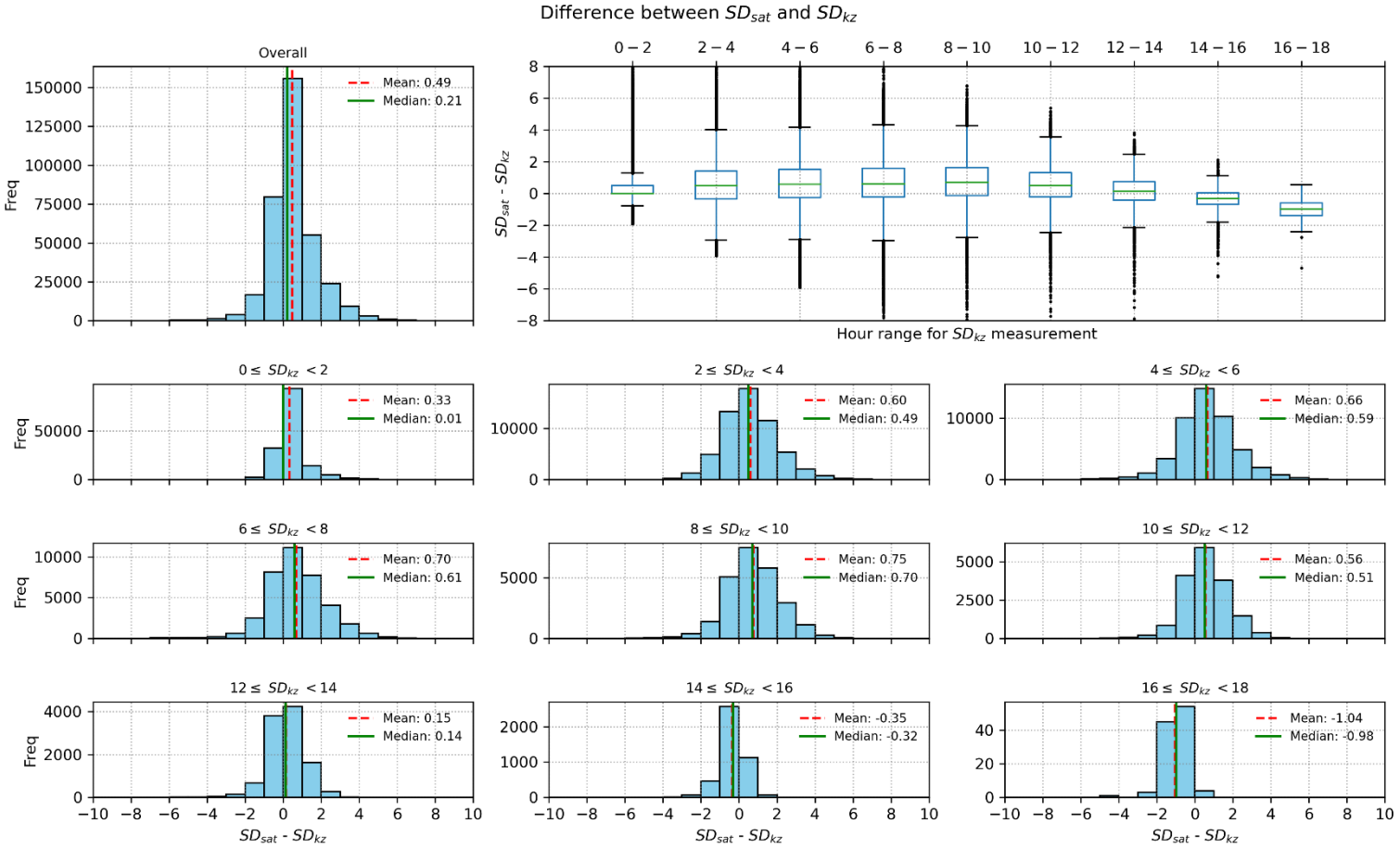


Figure 4-5: Distributions of daily $SD_{sat} - SD_{kz}$ differences (KZ-native). The top left panel shows the distribution for all machups, while the other panels show the distributions of differences partitioned by SD_{kz} values. The top-right panel shows the distributions for all partitioned ranges where the central green line corresponds the 50th percentile (median) value, the box shows the 25th and 75th percentiles (or inter-quartile range), while the whiskers extend to the farthest data point within 1.5x the inter-quartile range from the box. The data points that extend past the end of the whiskers are the most extreme values.

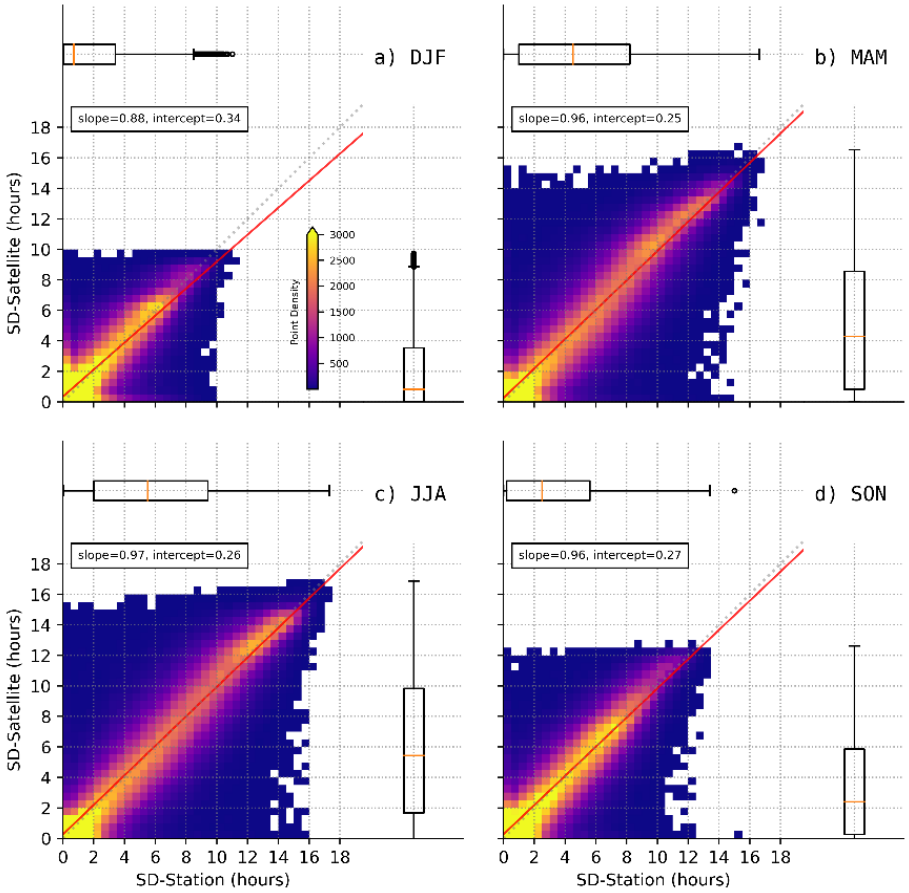


Figure 4-6: Scatter-density plot showing the relationship between colocated daily CS-native and satellite observations over the UK for a) December/January/February, b) March/April/May, c) June/July/August and d) September/October/November. The dotted line is the 1:1 line and the red line is the linear regression for the data shown. The distributions are also shown on each panel by box and whiskers: the central orange line corresponds the 50th percentile (median) value, the box shows the 25th and 75th percentiles (or inter-quartile range), and the whiskers extend to the farthest data point within 1.5x the inter-quartile range from the box. The data points that extend past the end of the whiskers are the most extreme values.

4.3.1.1 Analysis of Outliers

Figure 4-6 and Figure 4-7 are scatter-density plots showing the agreement between the CS-native (SD_{cs}) and KZ-native (SD_{kz}) and satellite (SD_{sat}) observations, partitioned by season. For both station datasets and for all seasons, there is a strong, near-linear relationship between the satellite and in situ data (r = 0.86 to 0.95). However, there are many outliers, where the station data are measuring much higher *SD* compared with the satellite data, and vice versa, particularly for the SD_{cs} data. A closer inspection of the daily data reveals that there are isolated stations that appear to have some large SD_{sat}-SD_{stn} differences on some days (Appendix Figure 8-1). In addition, the SD_{sat} vs SD_{stn} relationship deviates slightly from linear at about 10 hours *SD*, where the SD_{stn} is sunnier than SD_{sat}. This matches the results from the analysis of the SD_{sat} - SD_{stn} distributions (Figure 4-4 and Figure 4-5), which also show the SD_{sat} is less sunny than SD_{stn} at higher values of *SD*.

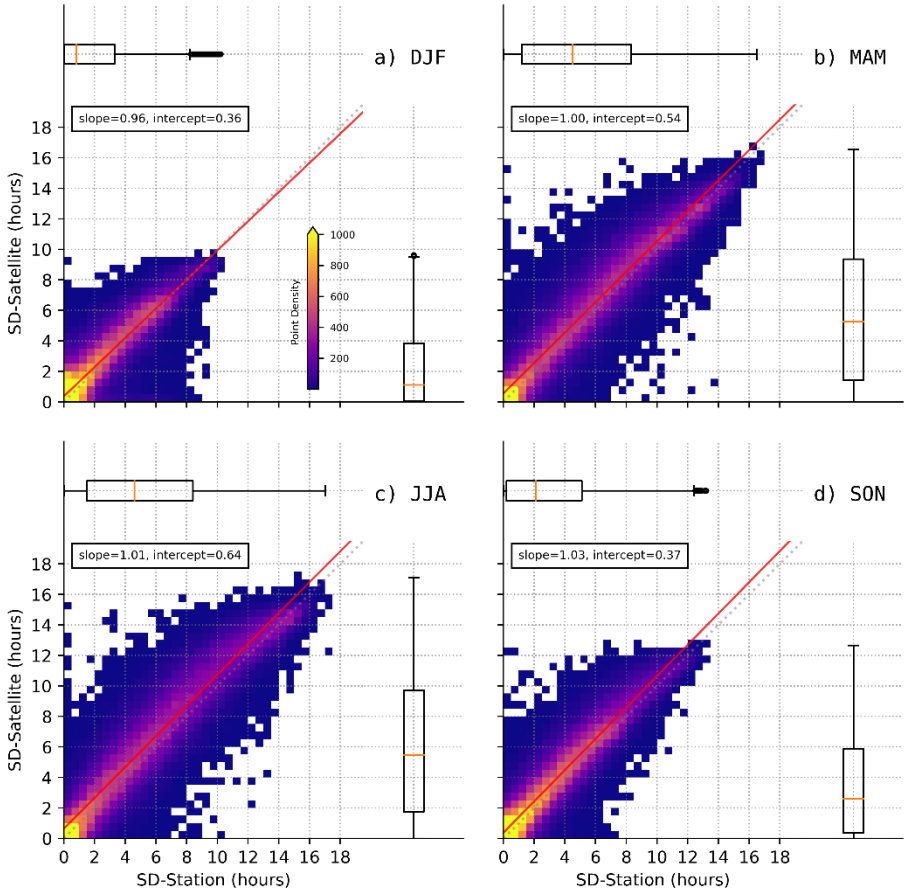


Figure 4-7: As for Figure 4-6 but for KZ-native, although note the different colour scale on each of these plots.

To investigate the outlying *SD* values further, the SD_{sat} vs SD_{stn} correlation (r), mean SD_{sat} - SD_{stn} difference (μ) and standard deviation (σ) at each station are calculated and shown as a function of station elevation in Figure 4-8 and Figure 4-9 for CS-native and KZ native, respectively. These results suggest that there are specific stations where the SD_{sat} vs SD_{stn} relationship may be weaker, which are characterised by r <0.6, μ<|2| hours and/or σ<2.5 hours (shown as dashed lines in the Figures). As the SD_{sat} data are estimated from a single sensor, using a uniform approach, it seems more likely that the SD_{stn} observations are spurious for these stations. Therefore, these stations are flagged as 'suspect' using these statistical thresholds, which have been selected through expert judgement in order to automate the identification of outlying stations. The tests are applied on a seasonal basis, where a station is flagged as an outlier if it fails one (yellow), two (orange) or three (red) of these tests in any season (the colours refer to the data points in Figure 4-8 and Figure 4-9). In addition, this outlier screening process includes a check for any stations where >20% of the daily SD_{sat} - SD_{stn} difference across all seasons and years is >5 hours (not shown in the figures). Notably, for the station data used in this study, there are no stations that fail this latter test that have not also failed at least one of the other three statistical tests. However, this test may remove additional stations when the dataset is extended in future.

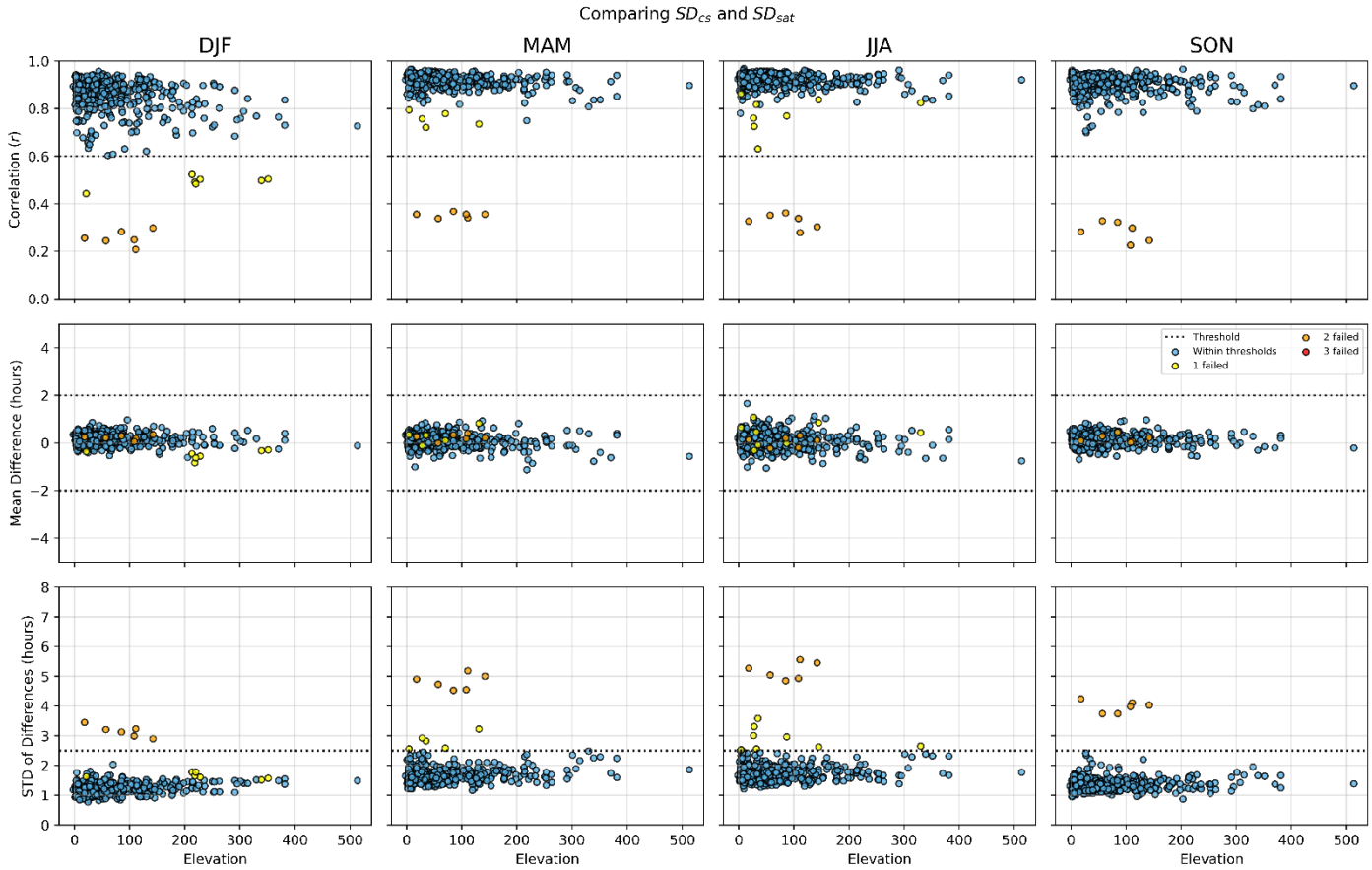


Figure 4-8: The satellite vs CS-native station correlations (top row), mean differences (middle row) and standard deviations of the differences (bottom row) shown as a function of the station elevation for December/January/February (first column), March/April/May (second column), June/July/August (third column) and September/October/November (fourth column). Each dot represents a different station. Stations that fail the statistical threshold tests (dotted line(s) on each panel) for each season are indicated in yellow (fails one test), orange (fails two tests) and red (fails three tests).

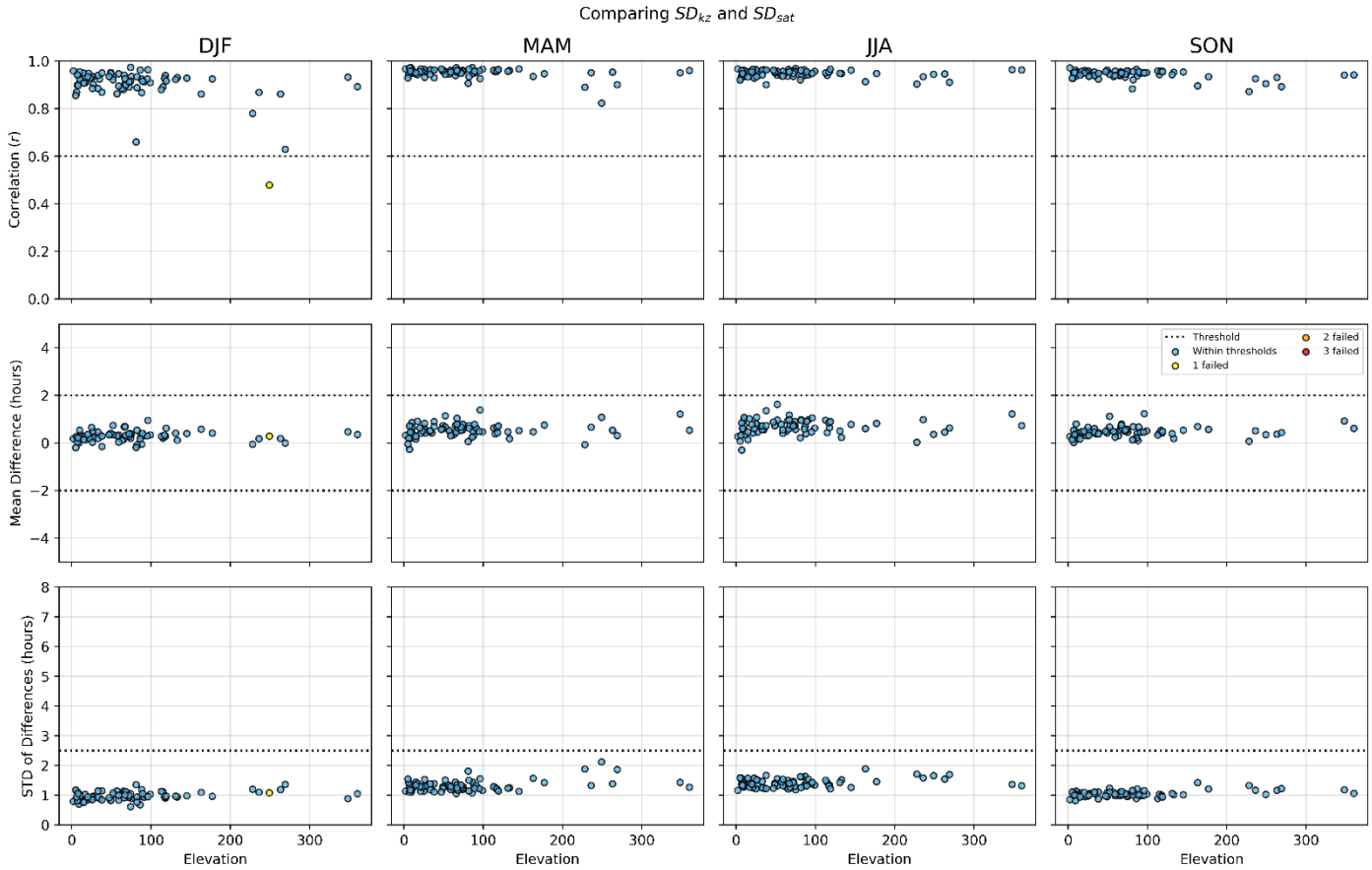


Figure 4-9: The satellite vs KZ-native station correlations (top row), mean differences (middle row) and standard deviations of the differences (bottom row) shown as a function of the station elevation for December/January/February (first column), March/April/May (second column), June/July/August (third column) and September/October/November (fourth column). Each dot represents a different station. Stations that fail the statistical threshold tests (dotted line(s) on each panel) for each season are indicated in yellow (fails one test), orange (fails two tests) and red (fails three tests).

Table 4-2: List of stations removed from the analysis through the outlier screening process. Columns 'Stn ID', 'Lat', 'Long' and 'H' provide the station identification number, latitude, longitude and height, respectively. A red-coloured box in columns '*r*', '*μ*' and '*σ*' indicates the failed threshold-based test(s) within each season ('DJF' = December/January/February, 'MAM' = March/April/May, 'JJA' = June/July/August, 'SON' = September/October/November). A red-coloured box in the final column (>20% >5 hours) indicates stations that have >20% of the daily SD_{sat} – SD_{stn} differences that are >5 hours. See text for further details.

Stn ID	Lat (°N)	Long (°E)	H (m)	DJF			MAM			JJA			SON			>20% >5 h
				r	*μ*	*σ*	*r*	*μ*	*σ*	*r*	*μ*	*σ*	*r*	*μ*	*σ*	
Campbell-Stokes																
105	58.867	-4.708	249													
107	57.144	-4.677	21													
111	56.94	-4.238	351													
128	57.334	-3.600	218													
130	57.335	-3.606	220													
140	57.354	-3.345	213													
147	57.006	-3.398	339													
178	56.969	-2.207	4													
394	53.445	-0.202	108													
432	52.572	1.738	4													
461	52.227	-0.465	85													

Stn ID	Lat (°N)	Long (°E)	H (m)	DJF			MAM			JJA			SON			>20% >5 h
				r	*μ*	*σ*	*r*	*μ*	*σ*	*r*	*μ*	*σ*	*r*	*μ*	*σ*	
564	53.064	-0.963	70													
608	52.073	-1.334	87													
613	51.62	-1.099	57													
715	51.68	-0.056	32													
810	50.865	0.336	18													
1070	54.934	-2.964	28													
1216	51.88	-5.124	111													
1233	51.964	-3.629	330													
1333	50.991	-4.462	142													
1341	50.686	-4.234	131													
1347	51.2	-4.113	145													
13343	54.741	-4.958	27													
30435	58.635	-3.06	35													

The list of stations that fail any of these tests within the SD_{kz}, SD_{cs}, and SD_{cs_uni} datasets is given in Table 8-1 (SD_{kz} and $SD_{cs;}$ Appendix) and Table 4-2 (SD_{cs_uni}). The majority of SD_{stn} data that are classified as outliers are from CS sensors, which is expected given that the number of KZ sensors across the UK station network is considerably lower. Only one KZ station (station ID = 105) fails the outlier screening process (Appendix, Table 8-1). Notably, the list of stations that fail the statistics tests for the SD_{cs_uni} is almost identical to the list for SD_{cs} + SD_{kz}, with the exception of station 113 (north-central Scotland), which fails the correlation test for DJF in the SD_{cs} dataset but passes all tests for all seasons for SD_{cs_uni}. These stations are excluded from the input SD_{cs_uni} data used to generate SD_{bld} dataset (Section 3.2 and Section 4.5). For the stations that are not flagged as outliers, the results in Figure 4-8 and Figure 4-9 imply there is a slight decline in r and a slight increase in σ with increasing elevation for both the CS-native and KZ-native results, which is slightly more marked in winter (December/January/February). This result suggests that elevation should be investigated as an additional predictor variable in the next developmental phase of the regression model used to generate the SD_{bld} dataset (Section 3.2 and Section 4.5).

Analysis of the time series of daily SD_{sat} - SD_{stn} differences for individual stations shows that there are many negative and positive spikes that are several hours in magnitudes, which occasionally exceed 10 hours. These are the extreme outliers that are visible in Figure 4-6 and Figure 4-7. However, the source of these spikes is currently unknown, and they could result from the satellite and/or station *SD* data. Therefore, these data have not been removed from the datasets and may affect the resulting SD_{bld} production. There are also specific time periods when the satellite-station *SD* agreement is clearly worse, which are seen in the comparisons between SD_{sat} and all SD_{stn} variants. For example, station 1007 shows a generally stable time series of SD_{sat} – SD_{cs_uni} differences with the exception of the period between the beginning of 2014 and the end of 2016, where the difference becomes more positive by up to a few hours (Appendix Figure 8-2). The individual SD_{cs_uni} and SD_{sat} time series suggest that the SD_{cs_uni} data are less sunny than adjacent years for this station, whereas the SD_{sat} data for these years are more typical. As this anomalous period seems to be limited to a single station, it seems most likely there was an issue with the SD_{stn} observations for this period, as problems in the SD_{sat} data (e.g. calibration, instrumental error, etc) are likely to manifest quite consistently and simultaneously across all stations/grid cells. However, there are periods where the SD_{sat} – SD_{cs_uni} difference is anomalously large that are consistent across many stations. For many UK stations, the SD_{sat} – SD_{cs_uni} difference is more positive in 1991 and 1992, particularly in this latter year. The SD_{cs_uni} for this period appears to be slightly lower than usual, while the SD_{sat} data appear more 'normal'. An example for station 43 is shown in Figure 8-3 (Appendix). This particular *SD* discrepancy that seems to affect many of the SD_{sat} – SD_{cs_uni} station time series may be a result of the eruption of Mount Pinatubo on 3 June 1991, which affected global climate through the ejection of large quantities of gas and ash into the atmosphere. A reduction in observed SD_{stn} following major volcanic eruptions, including Pinatubo, has been reported in several studies and is therefore expected (Magee et al., 2014; Obregón et al., 2020; Sanchez-Lorenzo & Wild, 2012; Stanhill et al., 2005). Therefore, the observed reduction in $SD_{cs\_unified}$ is likely to be a real effect. Pfeifroth et al. (2018) found a similar positive satellite-minus-station anomaly difference for solar radiation from the SARAH-2 dataset after the 1991 Pinatubo eruption. Post-eruption lofted volcanic aerosols and dust are also known to affect satellite observations at visible and thermal wavelengths, and consequently these data are used to detect and track volcanic emissions (Francis et al., 2012;

Naeger & Christopher, 2014). Therefore, it seems likely that the SD_{sat} data could also be affected by the eruption of Mount Pinatubo, but perhaps to a lesser degree than the SD_{cs_uni} data. Time-dependent differences in the SD_{sat} and SD_{stn} data are examined further in Section 4.4 using the HadUK-Grid *SD* dataset.

4.3.2 CS-unified (SD_{cs_uni}) results

Based on the results of the analysis of SD_{sat} and SD_{stn} presented in the previous sections, SD_{cs_uni} has been selected as the input SD_{stn} for SD_{bld}, although 24 stations that are considered to be 'outliers' have been removed from this dataset. Table 4-2 presents the statistics for the SD_{sat} vs SD_{cs_uni} comparisons for stations that have passed the outlier screening process outlined in Section 4.3.1. The SD_{sat} - SD_{cs_uni} distributions (Figure 8-4, Appendix) are almost identical to those presented in Figure 4-4 for SD_{sat} - SD_{cs} differences. The statistics confirm the close overall agreement between SD_{sat} and SD_{cs_uni}, where μ ~0.1 hour with no seasonal variation. The correlation and standard deviation are both lowest in the winter (r=0.87 and σ=1.2 hours) and highest in the summer (r=0.92 and σ=1.8 hours). However, as noted earlier, there are some large differences of several hours, which are reflected in the extremes of the distributions (1st and 99th percentiles are ~-4.2 hours and ~4.5 hours respectively). These data need to be handled with care when blending the SD_{sat} and SD_{stn} data to create SD_{bld}., which will be addressed further in the next developmental phase for this product.

Table 4-3: Statistics for the SD_{sat} vs SD_{cs_uni} relationship for all ('All'), winter (DJF), spring (MAM), summer (JJA) and autumn (SON) daily colocated matchups. The 24 stations that fail the outlier screening process described in Section 4.3.1 are excluded from these statistics.

Data subset	No. Data	*r*	μ (hours)	σ (hours)	Percentiles (hours)						
					1st	5th	25th	50th	75th	95th	99th
ALL	2700426	0.93	0.1	1.5	-4.2	-2.4	-0.5	0.0	0.7	2.6	4.5
DJF	661871	0.87	0.1	1.2	-3.7	-1.7	-0.2	0.0	0.5	2.2	3.8
MAM	685839	0.92	0.1	1.7	-4.6	-2.7	-0.7	0.0	0.9	2.8	4.8
JJA	683402	0.92	0.0	1.8	-4.6	-2.8	-0.9	0.0	0.9	3.0	5.0
SON	669314	0.91	0.1	1.4	-3.6	-2.0	-0.4	0.0	0.7	2.4	4.1

4.4 Comparison between HadUK-Grid (SD_{had}) and satellite (SD_{sat}) *SD*

In addition to producing the SD_{bld} dataset that is based on both SD_{sat} and SD_{cs_uni}, a satellite-only daily and monthly *SD* product for the UK is also envisaged to complement the HadUK-Grid *SD* dataset (SD_{had}). Therefore, it is necessary to also quantify the agreement between SD_{sat} and SD_{had} to inform the users of any differences.

Figure 4-11 is a scatter-density plot showing the relationship between the monthly totals of SD_{sat} and SD_{had} within each season for all UK grid cells between 1983 and 2022. As seen for the SD_{sat} vs SD_{stn} comparisons, the SD_{sat} and SD_{had} datasets are highly correlated and linear, indicating good overall agreement between the two datasets. The small non-linearity observed at higher *SD* values in the SD_{sat} vs SD_{stn} comparisons is not evident in the SD_{sat} vs SD_{had} relationship. The overall scatter is also smaller for the SD_{sat} and SD_{had} comparison, although there are a few outliers in spring (MAM) where SD_{had} is around twice that of SD_{sat}, although this is at least partly expected given the difference in scale (daily vs. monthly)

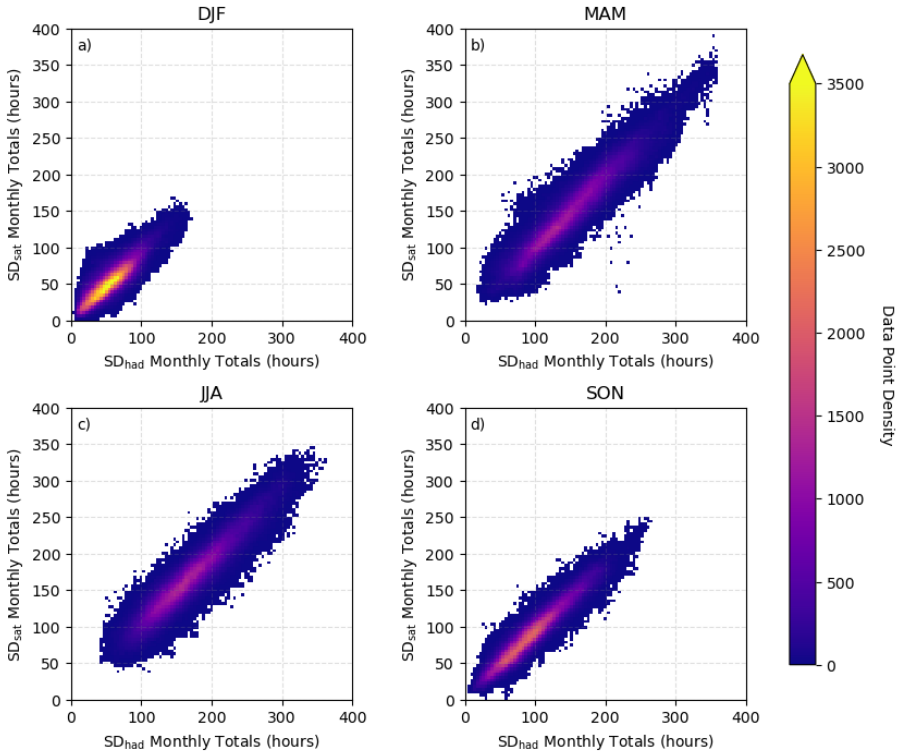


Figure 4-10: Scatter-density plot showing the relationship between monthly totals of SD_{sat} and SD_{had} for all UK grid cells within the period 1983 to 2022 for (a) December/January/February (DJF), (b) March/April/May (MAM), (c) June/July/August (JJA), and (d) September/October/November (SON).

Figure 4-12 shows maps of the mean monthly SD_{sat} – SD_{had} differences for each season, together with the distributions of those differences. Overall, the distributions suggest that SD_{sat} tends to sunnier than SD_{had}, with majority of the monthly mean differences falling in the -10 to +15 hours range. The maps shows that there is some spatial structure in the SD_{sat} – SD_{had} differences, where the more negative differences (i.e. satellite less sunny than HadUK-Grid) correspond to areas of higher orography, for example, the Pennines in Northern England and the Scottish Highlands. The most positive differences (i.e. satellite sunnier than HadUK-Grid)

are located in the south and east of the UK. In all seasonal difference maps, there is a prominent circular outlying area with high *SD* difference at ~51.0°N, ~2.3°W that is particularly evident in the autumn. This is similar to the circular features seen in the seasonal SD_{had} example shown in Figure 4-1 (Section 3.1), which are likely to be a influencing effect from one station that results of the HadUK-Grid interpolation process. In addition, the circular feature colocated with Loch Neagh in Northern Ireland that was identified in Figure 4-1 in both SD_{sat} and SD_{had} is also clearly visible in all seasons but is particularly strong in the spring and summer. The $SD_{sat} - SD_{had}$ differences also tend to be positive along the south-facing coasts of the UK, for example in southern England and Wales, particularly in summer (JJA). This could be a result of the interaction between Meteosat viewing geometry, land-sea cloud contrast and related cloud parallax effects, whereby the satellite is observing the surface at an angle through a cloud-free atmosphere, but the skies above the ground point are cloud-covered. In this situation, the resulting SD_{sat} could conceivably be higher than observed in situ, and therefore in SD_{had}. The variation in $SD_{sat} - SD_{had}$ differences as a function of elevation, satellite view angle and distance from the coast will be explored in future work. In the meantime, this analysis further supports exploring some of these variables as additional predictors when producing an SD_{bld} dataset.

Figure 4-13 shows the time series of the UK-wide mean SD_{sat} and SD_{had} for each month between January 1983 and December 2022. The figure shows both the actual SD values and the anomalies, as well as the time series of SD_{sat} - SD_{had} differences. On the panels showing the *SD* differences (b and d), the dates of several volcanic eruptions are marked, together with the transition date from MVIRI to SEVIRI observations in the SD_{sat} record (1 January 2006). It appears that SD_{sat} is sunnier than SD_{had} during the MVIRI portion of the record (1983-2005) immediately after the two major volcanic eruptions that occurred in this period, which is consistent with the individual SD_{sat} - SD_{stn} time series discussed in Section 4.3.1.1 In the case of the Mount Pinatubo (3 June 1993), the anomaly differences in particular are more positive for almost 2 years post-eruption. As reported earlier in Section 4.3.1.1, SD_{stn} (and therefore SD_{had}) seem to report lower *SD* during this period, while the SD_{sat} observations seem to be less affected. The transition from MVIRI to SEVIRI in 2006 also appears to cause a discontinuity in the differenced time series, where the SD_{sat} are sunnier than SD_{had}. These likely non-climatic discontinuities in SD_{sat} need to be considered carefully when producing SD_{bld}, as they will propagate through to the final dataset where they may cause considerable errors when using the SD_{bld} data for any time series analysis.

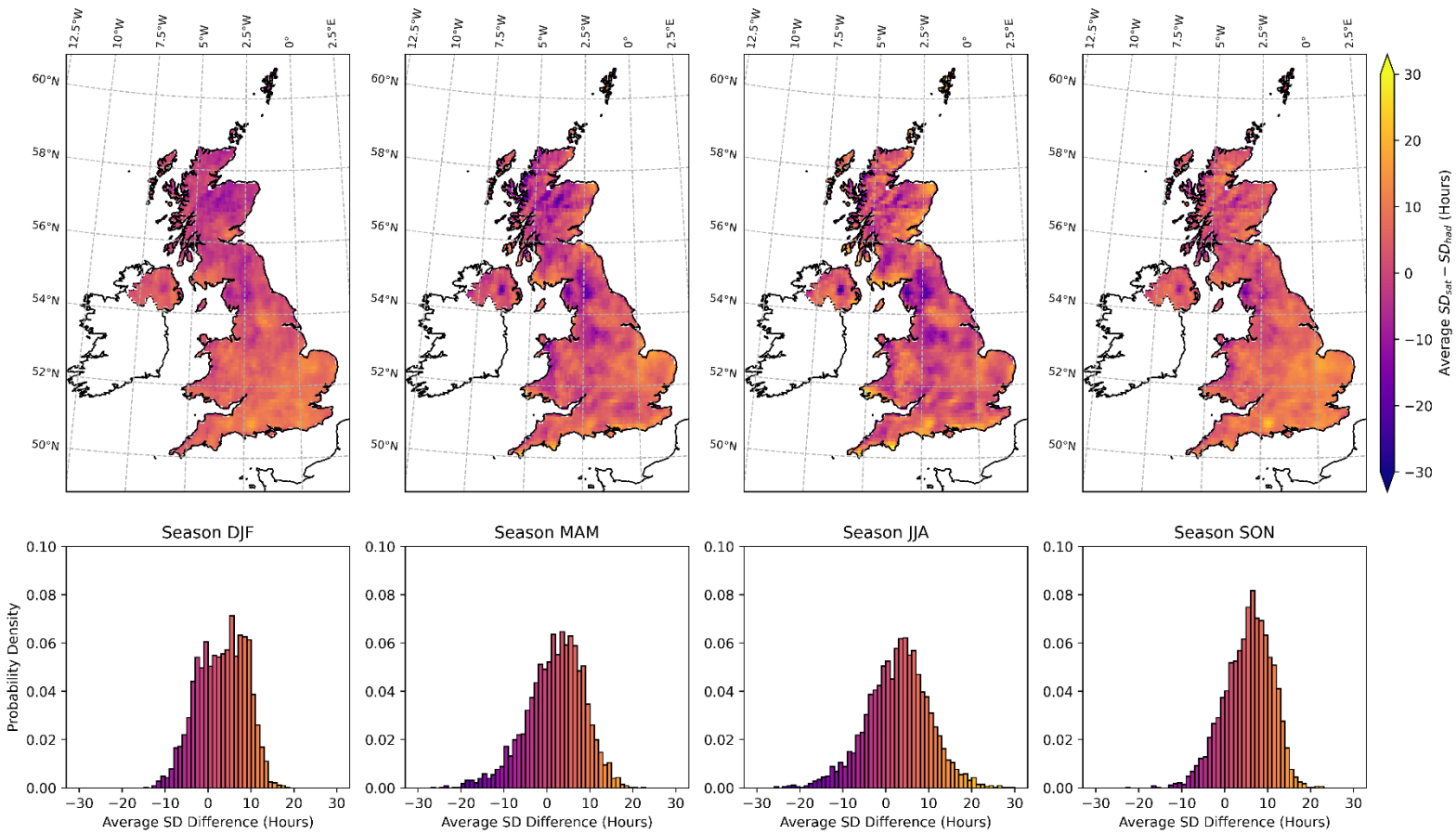


Figure 4-11: Maps of mean monthly SD_{sat}-SD_{had} differences for the period 1983-2022 (top row) for December/January/February (DJF), March/April/May (MAM), June/July/August (JJA) and September/October/November (SON). The distributions of those differences are also shown (bottom row).

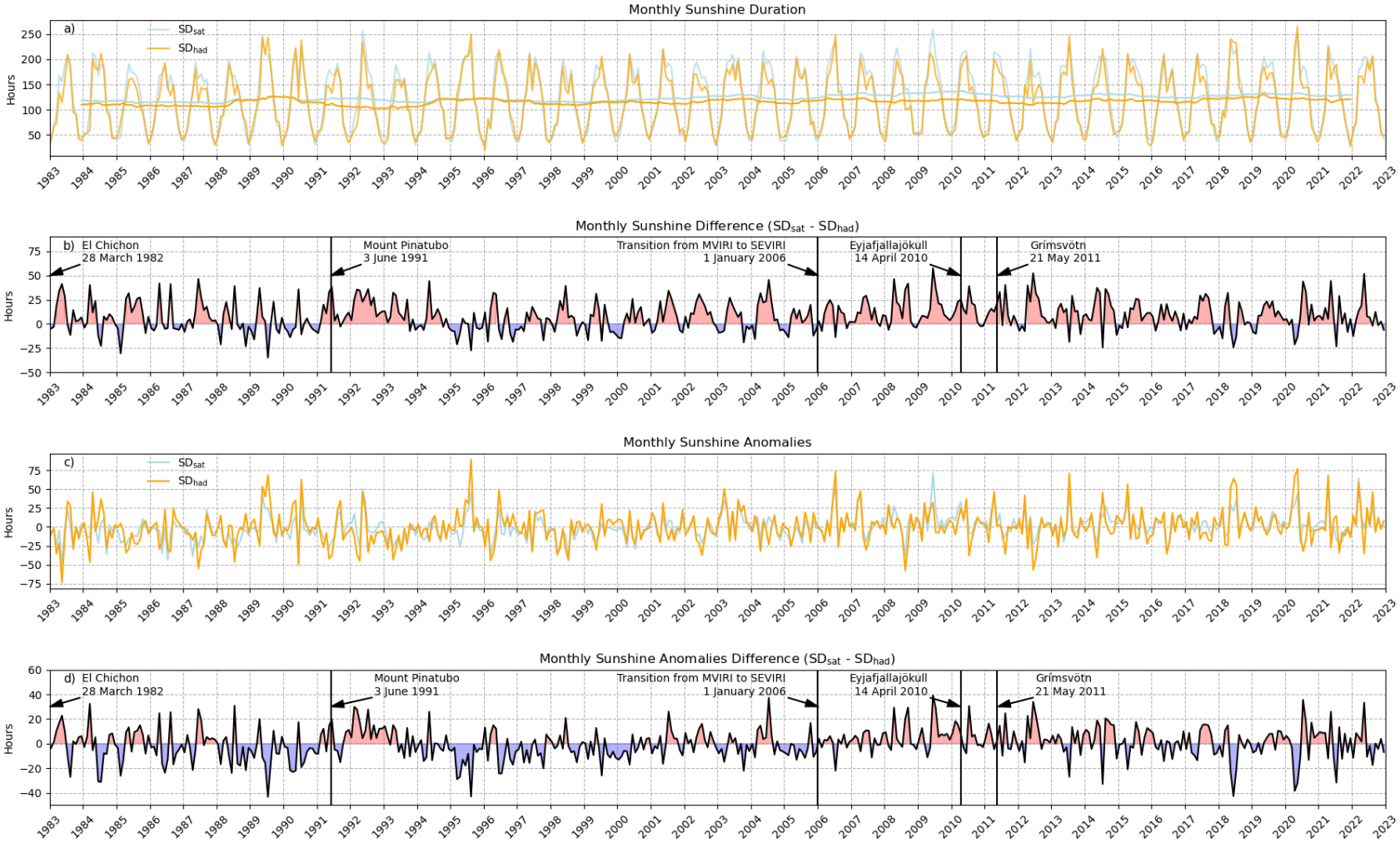


Figure 4-12: Time series of (a) SD_{sat} and SD_{had}, where a 24-month rolling mean is also shown for clarity, (b) SD_{sat} - SD_{had} differences, (c) SD_{sat} and SD_{had} anomalies (with respect to the 1983-2022 baseline period) and (d) SD_{sat} - SD_{had} anomalies. Major volcanic eruptions are marked on panels (b) and (d).

4.5 Blended satellite-station SD (SD_{bld})

In this section, the output from a preliminary blended satellite-station *SD* dataset, SD_{bld}, is presented. This initial SD_{bld} dataset will be developed further in future work, with the intention of delivering a new Met Office operational *SD* product for the UK in the next 1-2 years. As stated in Sections 4.2 and 4.3, the SD_{cs_uni} station dataset is used as input for the SD_{bld} dataset, together with SD_{sat}. This SD_{cs_uni} dataset comprises SD_{cs} data that have been extended and/or infilled with adjusted SD_{kz} data (SD_{kz_corr}) using the Legg (2014) correction.

Figure 4-14 shows the daily regression model coefficients for the three GLMs trialled in this preliminary analysis (see Equation 3-3, Equation 3-4, Equation 3-5), together with the two goodness-of-fit statistics, RMSE and pseudo-R^2 (Section 3.2). All the time series are noisy (also see Figure 8-5, Appendix), which suggests that the 5-day window trialled here to generate the GLM models is too short and a longer window, e.g. 15 or 31 days is required. This will be explored in the next developmental stage for this dataset. The RMSE and pseudo-R^2 are virtually identical for all three models; there are small differences, but these are not visible to the naked eye. Both measures have a clear annual cycle, which peaks in the summer, ranging between ~1.2 and ~1.8 hours for RMSE, and ~0.4 and ~0.9 hours for the pseudo-R^2. This suggests a lower confidence in the SD_{bld} data in winter compared with summer.

The model coefficients also show annual cycles, although the intercept (*d*) coefficient is highly variable/noisy (Figure 4-13 panels g, h & i). (Note that the same data are also shown in Figure 8-5 in the Appendix, but with different *y*-ranges that are tailored to the values for each type of coefficient; the panels showing the coefficients in Figure 4-14 are plotted on the same *y* axes.) Both the *c* and *d* coefficients peak in the summer, whilst the *a* and *b* coefficients peak in the winter. For GLM 1, the contribution of the *d*, or intercept, coefficient ranges between ~0.5 and ~1.0 hours, and for GLM 2 and 3, between ~-4.0 and ~2.0 hours, although there is an outlying trough around day 60 that extends to almost -6.0 hours for both these models. The *c* coefficients (the coefficient for SD_{sat}) for GLM 1 and 2 are almost identical and range between ~0.7 and ~0.9, while the range for GLM 3 is slightly larger, ranging between ~0.5 and ~1.0. Therefore, the maximum possible contribution from this model component is up to ~17 hours for GLM 1 & 2 and ~19 hours for GLM 3 for the UK summer, where the maximum possible *SD* is 19 hours (Section 2.1). The *b* coefficients (the coefficient for *lat*; GLMs 2 & 3 only) range between ~-0.02 and 0.12. For the latitude of the UK, which varies between ~50 to ~60°N, the contribution of these coefficients to the GLM-calculated *SD* therefore ranges between ~-1.2 and 7.2 hours. The *a* coefficient for GLM 3 ranges between ~-0.03 and ~0.06. The contribution of this SD_{sat}^2 coefficient to the GLM-calculated *SD* therefore ranges between ~-0.6 and ~0.5 hours, assuming a maximum possible *SD* of 8 hours and 19 hours in UK winter and summer, respectively (Section 2.1). The largest contribution to the GLM-calculated *SD* in all models is therefore from the non-squared SD_{sat} term, followed by the latitude term (GLM 2 & 3 only).

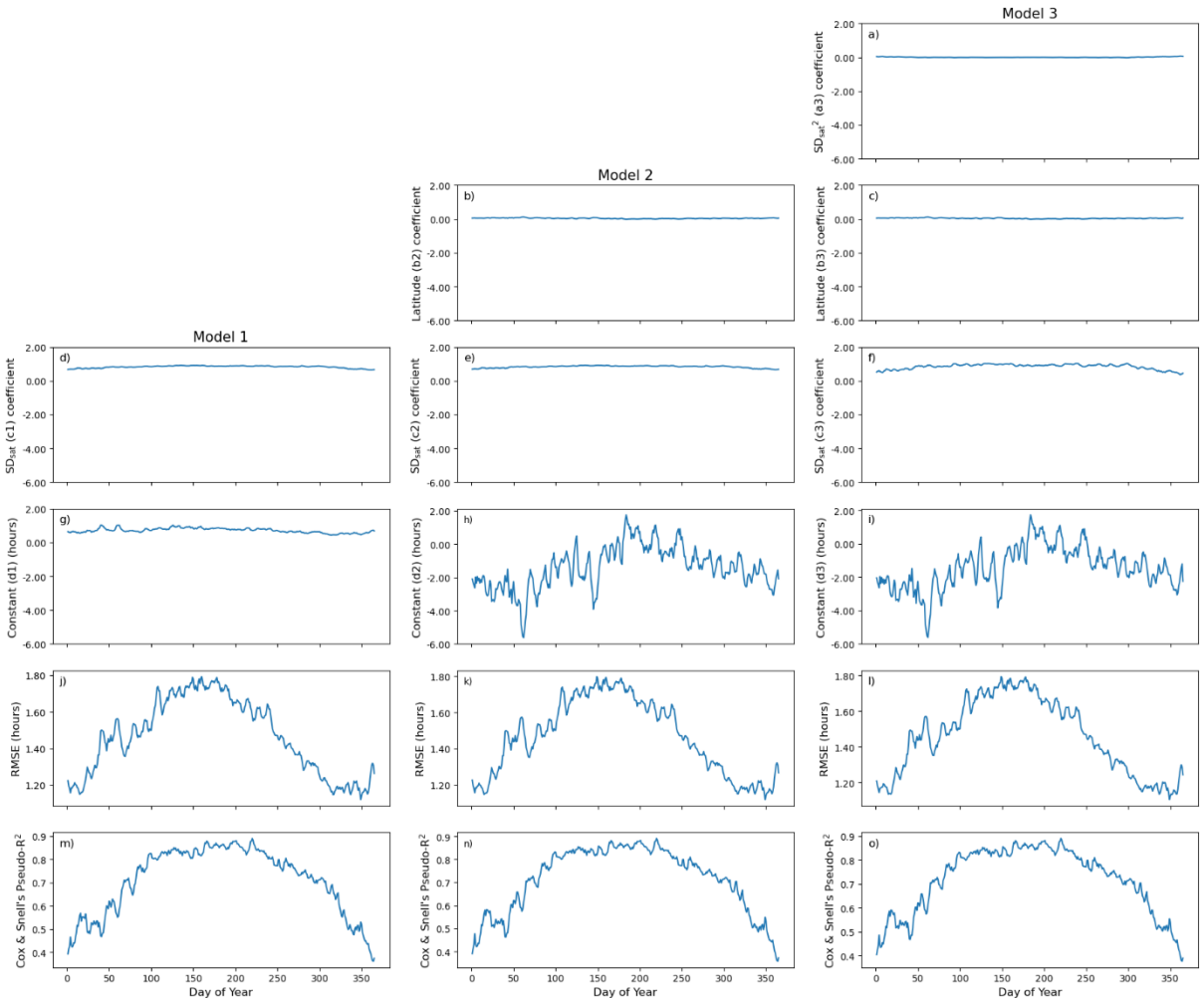


Figure 4-13: Generalised Linear Model coefficients used to blend SD_{stn} and SD_{sat} and the goodness of fit statistics for the three models trialled in this study (see Section 3.2). The top row shows the *a* coefficients (panel a), second row the *b* coefficients (panels b & c), third row the *c* coefficients (panels d, e & f), and the fourth row the *d* coefficients, or *y*-intercept, for each model (panels g, h & i). The fifth row shows the Root Mean Square Error (RMSE) for each model (panels j, k & l), while the bottom row shows the Psuedo-R_2 (panels m, n & o; see Section 3.2). The results for Model 1 (GLM 1; Equation 3-3) are shown in the first column, Model 2 (GLM 2; Equation 3-4) in the second column and Model 3 (GLM 3; Equation 3-5) in the third column. Note that the same *y*-axis range is used for all model coefficients; Figure 8-5 (Appendix) shows the same data where the panels showing the coefficients have *y*-axis ranges that are specific to each group of coefficients.

Figure 4-14 shows an example of the final daily SD_{bld} dataset, which is based on GLM 1 (Equation 3-3) that uses only SD_{sat} as input. The example is for the same day shown in Figure 4-1, which provides the corresponding SD_{sat} and SD_{cs_uni} data (1 June 2003). As expected, the GLM-estimated *SD* has the same overall pattern as the SD_{sat} on this day, with most of the SD_{sat} – GLM *SD* differences falling within ±1 hour *SD* (Figure 4-14b). The differences between the GLM 1 output and the SD_{cs_uni} are typically within ±2 hours, although there are some stations with much larger differences. For example, in southeastern Northern Ireland there is a single coastal station that is ~8 hours sunnier than the GLM 1 output, denoted by a bright yellow filled circle. Notably, the three most westerly mainland coastal stations in West Wales and West Scotland also show some of the most positive SD_{cs_uni} – GLM 1 *SD* differences

(~4-6 hours) across the UK. However, the satellite data do not support very high *SD* at this coastal station location in eastern Northern Ireland but do confirm higher *SD* to the East of the 5°W meridian, which is consistent with the three aforementioned coastal station observations on the UK mainland. This suggests that the high-*SD* observation from the coastal station in eastern Northern Ireland may be erroneous.

A further example of station locations with large SD_{cs_uni} – GLM 1 *SD* differences is the cluster of four stations in northeast Scotland where the SD_{cs_uni} observations are ~4 hours less sunny than the GLM 1 output (~57°N, ~2°W). Figure 4-1 confirms that SD_{sat} is ~3-4 hours sunnier than SD_{cs_uni} for these stations. In this case, it seems more likely that the SD_{sat} data may be overestimating the *SD* in this region, given the consistency in the SD_{cs_uni} – GLM 1 *SD* across the four stations on this date. The effect of these four stations is clear in the GP *SD* correction field, which is based on the SD_{cs_uni} – GLM 1 residuals: there is a regional ~-3 hour *SD* correction in this field in the northeast of Scotland (Figure 4-14d). Consequently, the combined GLM 1 + GP SD_{bld} (Figure 4-14e) has lower *SD* compared with the GLM 1-only *SD* (Figure 4-14a) in this region, and the SD_{cs_uni} – GLM *SD* differences (Figure 4-14f) for the four stations are much closer to zero. The benefit of the GP component of the satellite-station *SD* blending process is clear in this example, where it has aligned the overestimated *SD* from the satellite observations with the SD_{stn} data to produce (what appears to be) a more accurate *SD* representation.

However, the GP component of the *SD* blending process can be similarly detrimental. For example, the influence of the likely-erroneous high-SD_{cs_uni} outlier in southeastern Northern Ireland can also be seen in Figure 4-14d. In this case, it appears that this station is incorrectly 'brightening' the *SD* to the North and the East (Figure 4-14d). Such errors are more likely to occur where the station density is low, which needs to be managed carefully when finalising an operational SD_{bld} dataset. The spatial patterns in Figure 4-14d also indicate that the geographical influence of the GP corrections is extensive and suggests that the CLS used in the GP model may be too large. As noted in Section 3.2, the CLS will be revisited in the next developmental stage for the SD_{bld} dataset. Nevertheless, the resulting combined GLM 1 + GP *SD* field shown in Figure 4-15e is very promising. The overall pattern of the SD_{sat} data is well preserved but the influence of the SD_{cs_uni} data can clearly be seen locally.

The results for the two other models trialled in this study, GLM 2 + GP (Figure 8-6) and GLM 3 + GP (Figure 8-7), are very similar to those obtained for GLM 1 + GP, although there are some small differences. The most notable are visible in the SD_{sat} – GLM *SD* differences, which are shown in panel b in each figure. There is a slight brightening North-to-South gradient in the GLM 2 *SD* (Figure 8-6b) and GLM 3 *SD* (Figure 8-7b) with respect to the GLM 1 *SD* (Figure 4-15b), which results from the additional *lat* term present in both GLM 2 & 3. Perhaps the most noticeable difference between the three GLMs is that GLM 3 appears to be slightly 'sunnier' overall compared with GLM 1 & 2. However, after the GP model correction has been applied, the final SD_{bld} maps, which are shown in panel e in each of the figures, are almost identical.

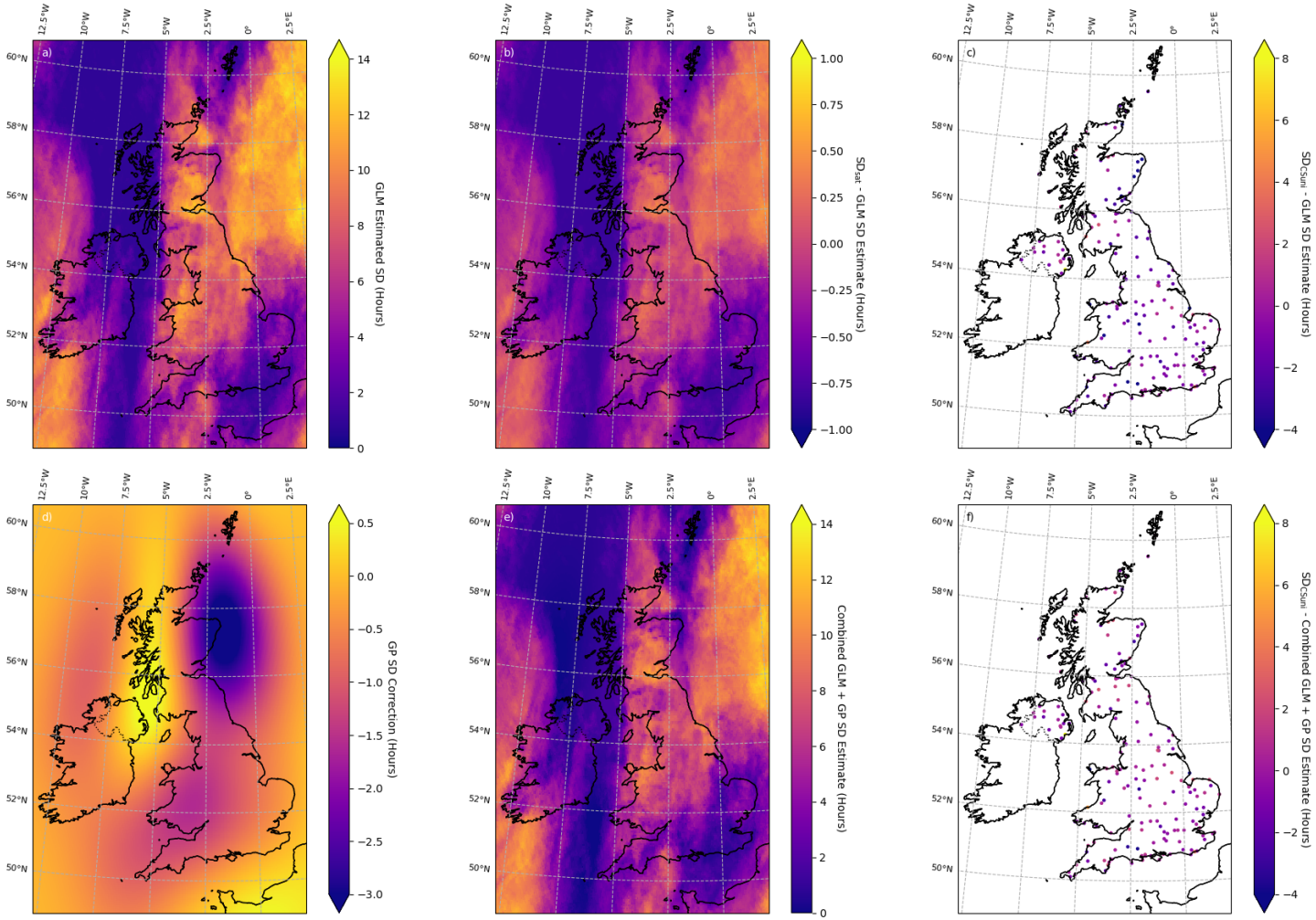


Figure 4-14: Example model output for 1 June 2003 showing the a) Generalised Linear Model 1 *SD* (GLM 1, Equation 3-3), b) SD_{sat} – GLM 1 *SD*, c) SD_{cs_uni} – GLM 1 *SD*, d) the GP *SD* correction, e) the combined GLM 1 + GP corrected *SD* output, and f) SD_{cs_uni} – combined GLM 1 + GP modelled *SD*.

Table 4-4: Pooled K-fold validation statistics for each model tested in this study. Model 1 uses GLM 1 (Equation 3-3), Model 2 uses GLM 2 (Equation 3-4) and Model 3 uses GLM 3 (Equation 3-4). In each case, the validation results represent the $SD_{bld} - SD_{cs\_uni}$ statistics. Results are shown for the GLM-only and combined (GLM + GP) model output. The statistics presented include the Root Mean Squared Error (RMSE), mean difference (μ), standard deviation (σ) and percentiles.

Statistic	Model 1		Model 2		Model 3	
	GLM only (hours)	with GP (hours)	GLM only (hours)	with GP (hours)	GLM only (hours)	with GP (hours)
RMSE	1.4	1.2	1.4	1.2	1.4	1.2
μ	-0.2	-0.0	-0.2	-0.0	-0.2	0.0
σ	1.4	1.2	1.4	1.2	1.4	1.2
Minimum	-14.5	-14.7	-14.9	-14.7	-14.2	-14.7
1st	-4.0	-3.4	-4.0	-3.4	-3.9	-3.4
5th	-2.3	-1.9	-2.3	-1.9	-2.3	-1.9
10th	-1.6	-1.3	-1.6	-1.3	-1.6	-1.3
25th	-0.8	-0.5	-0.9	-0.5	-0.8	-0.5
50th	-0.4	0.0	-0.4	0.0	-0.4	0.0
75th	0.6	0.5	0.6	0.5	0.6	0.5
90th	1.6	1.3	1.6	1.3	1.6	1.3
95th	2.2	1.9	2.2	1.9	2.2	1.9
99th	3.7	3.2	3.7	3.2	3.7	3.2
Maximum	14.4	15.0	14.4	15.0	14.5	15.0

This high degree of similarity between the three models trialled in this study is also evident in the K-fold validation statistics (Table 4-2). Validation results are presented separately for each of the three GLMs (i.e. before the GP model is added), and for each of the GLMs combined with the GP model. The results illustrate there is very little difference between the performance of each of the GLM models. The RMSE, μ and σ values for all three GLMs are identical at 1.4 hours, -0.2 hours and 1.4 hours, respectively. However, there are some small differences in the far extremes. In particular, the 0th percentile for GLM 3 is up to 0.7 hours less negative than for GLM 1 & 2. However, when the GP model is added, the performance of SD_{bld} is improved and there is no statistical difference between the different SD_{bld} variants, confirming this conclusion drawn previously based on the example SD_{bld} maps (Figure 4-14e, Figure 8-6e and Figure 8-7e). For the combined model (i.e. GLM + GP), the RMSE, μ and σ are all reduced

by 0.2 hours, indicating that there is only a small overall effect from the GP model at a national scale. However, as evident from the example SD_{bld} maps (e.g. Figure 4-14d, e and f) the local effect can be substantial; additional work will be carried out in the next phase of this product development to assess the impact of the GP and improve on this further.

5 Conclusions

The aim of this study is to produce a new blended *SD* dataset (SD_{bld}) for the UK that is based on a combination of satellite and station observations of *SD*. The CM SAF *SD* dataset provided by the SARAH-3 product is used as the input satellite dataset (SD_{sat}), whilst station data from the Met Office MIDAS database are used as the input station data (SD_{stn}). An SD_{sat}-only product for the UK is also planned to complement this new blended dataset and the existing HadUK-Grid *SD* product (SD_{had}), which is based on station data alone.

As a first step in creating SD_{bld}, an in-depth comparison is performed between *SD* recorded by the two different sensor types used within the UK: Campbell Stokes (CS), a traditional *SD* sensor invented over 150 years ago, and Kipp & Zonen (KZ), a more modern sensor that uses the WMO definition of 120 Wm^{-2} to define bright sunshine. As reported by other studies, this work finds that CS records higher *SD* compared with KZ, particularly for mid-range *SD* values (~4-12 hours *SD*). However, the model proposed by Legg (2014) is found to be very effective in converting between the two types of measurements, except for at very high *SD* values (≥12 hours). Based on all available observation pairs across all values of *SD*, the correlation (r) between the CS and converted KZ observations is 0.98, the mean difference (μ) is 0.0 hours and the standard deviation (σ) is 0.8 hours.

The second step required to generate SD_{bld} is to perform a comparison between daily SD_{sat} and each type of daily station *SD* measurement. The satellite data are found to be more closely aligned with the CS observations and therefore CS is chosen as the baseline for the SD_{stn} data. This choice is advantageous as CS is also the baseline for SD_{had}, so the new SD_{sat} and SD_{bld} datasets should also be consistent with SD_{had}. The Legg (2014) conversion is therefore used to create a unified station dataset that is aligned with the CS sensor (SD_{cs_uni}) as input into SD_{bld}. However, 24 UK stations where the SD_{sat} vs SD_{cs_uni} agreement is spurious are excluded from this final dataset based on an automated outlier screening process that uses statistical thresholds defined through expert judgement. Based on all remaining valid data from the stations that pass these statistical tests, the SD_{sat} vs SD_{cs_uni} agreement is found to be excellent, where r is 0.93, μ (SD_{sat} - SD_{cs_uni}) is 0.1 hours and σ is 1.5 hours. The SD_{sat} vs SD_{cs_uni} correlation and standard deviation are both lowest in the winter (r=0.87 and σ=1.2 hours) and highest in the summer (r=0.92 and σ=1.8 hours), whilst μ ~0.1 hour with no seasonal variation.

The analysis of the SD_{sat} vs SD_{cs_uni} time series at individual stations indicates there that are many isolated days where the SD_{sat} - SD_{cs_uni} differences exceed several hours. The cause of these apparently random discrepancies is unknown and so these data are not removed from the final satellite/station datasets. Discontinuities in the SD_{sat} - SD_{cs_uni} time series are also evident. In particular, SD_{cs_uni} is observed to be simultaneously and systematically lower across the UK *SD* station network following major volcanic eruptions, such as Mount Pinatubo in 1991. However, SD_{sat} does not seem to be (as) affected, leading to positive SD_{sat} - SD_{cs_uni} differences for up to 1-2 years following an eruption event. Isolated discontinuities in the SD_{sat} - SD_{cs_uni} time series at some individual stations are also present. As the timing and nature of these discontinuities vary between stations, they are more likely to reflect problems with the station observations as satellite-related issues are expected to affect the SD_{sat} vs SD_{cs_uni} agreement more consistently across the station network. These observational

outliers and discontinuities need to be considered carefully when producing the final operational SD_{bld} dataset.

Additionally, a comparison between monthly SD_{had} and monthly SD_{sat} is performed so that the differences between the two datasets can be communicated to users. These independent *SD* datasets are in very good agreement, although SD_{sat} provides slightly higher estimates of *SD* compared with $SD_{had.}$ The largest-magnitude differences are seen over areas of high orography, such as the Pennines and the Scottish Highlands. In addition, there are some circular artifacts in the SD_{had} dataset that are attributed to the interpolation process used to generate this dataset and are not present in the SD_{sat}, which generally provides a smoother field. Analysis of the complete SD_{sat} - SD_{had} time series between 1983 and 2022 confirms the earlier result that the satellite-station differences become more positive for up to 1-2 years after major volcanic eruptions. In addition, a discontinuity in the SD_{sat} - SD_{had} time series is observed in January 2006 where this *SD* difference becomes more positive. This is attributed to the transition between MVIRI onboard MFG and SEVIRI onboard MSG, which indicates that the SD_{sat} timeseries is not perfectly homogeneous, which may have implications for users who wish to use this dataset for timeseries analysis.

In the final part of this study, a preliminary, experimental blended *SD* dataset is generated using the CM SAF SD_{sat} dataset and the SD_{cs_uni} dataset produced and verified as part of this study. Firstly, a Generalised Linear Model (GLM) is used to predict SD_{cs_uni} for all UK grid cells in the SD_{sat} dataset. Three GLMs are trialled that use different combinations of predictors: SD_{sat} only, SD_{sat} & latitude, and SD_{sat} & latitude & SD_{sat}^2. Secondly, a Gaussian Process (GP) model is used to estimate a 'correction' field based on the residuals (observed SD_{cs_uni} - GLM-predicted *SD* at actual station locations), which is then added to the *SD* field output by the GLM. The resulting SD_{bld} datasets appear realistic and retain the overall pattern in *SD* provided by the satellite observations, but with 'localised' adjustments based on real station observations. Based on K-fold validation statistics, there is negligible difference in performance between the three GLMs trialled in this study. The addition of the GP model improves the national validation statistics slightly and RMSE, μ and σ all improve by ~0.2 hours. However, improvements (with respect to station data) can be more substantial at a local level.

Further work is required to deliver an operational blended satellite-station *SD* dataset for the UK, which is planned in the next 1-2 years. Based on the analysis presented in this study, this additional work will need to include the following (see Sections 3.2 and 4.5 for further information):

- Investigate the application of additional station QC flags that are available in MIDAS-Open to improve the station data quality and reduce outliers/spurious observations.
- Investigate the use of an adjusted day-length calculation based on the solar zenith angle (SZA) exceeding e.g. 2.5° (other studies suggest that bright sunshine cannot occur for SZA<2.5°).
- Assess the use of elevation and distance from coast, and possibly principal components, as additional predictors within the GLM.
- Investigate the use of a data transformation before applying the GLM.

- Investigate using moving temporal windows of different lengths, e.g. 15 days or 31 days, to define the training data for the GLM in order to generate smoother daily regression coefficients.
- Include point-by-point variance within the GP model, e.g. variation in the measurement uncertainty at each station.
- Define the correlation length scale (CLS) to apply in the GP model.
- Tune the standard deviation hyperparameter applied to the distribution of functions within the GP model.
- Calculate per-grid cell uncertainties for the GLM + GP modelled *SD*; ideally including observational uncertainties from the station and satellite data, uncertainty due to the KZ→CS conversion, observational coverage, observational inhomogeneity

In addition, delivery of an operational product will require pre-processing and ingestion of real-time station and satellite *SD* data, which may have different characteristics and artifacts from the data utilised in this study.

6 References

Bertrand, C., Demain, C. and Journée, M. (2013), Estimating daily sunshine duration over Belgium by combination of station and satellite data, Remote Sensing Letters, 4(8), DOI:10.1080/2150704X.2013.789569

Francis, P. N., M. C. Cooke, and R. W. Saunders (2012), Retrieval of physical properties of volcanic ash using Meteosat: A case study from the 2010 Eyjafjallajökull eruption, *J. Geophys. Res.*, 117, D00U09, doi:10.1029/2011JD016788.

Frei, C., Willi, M., Stöckli, R. and Dürr, B. (2015), Spatial analysis of sunshine duration in complex terrain by non-contemporaneous combination of station and satellite data. Int. J. Climatol., 35: 4771-4790. https://doi.org/10.1002/joc.4322

Good, E. (2010) Estimating daily sunshine duration over the UK from geostationary satellite data, *Weather*, https://doi.org/10.1002/wea.619.

Journée, M., Demain, C., Bertrand, C. (2013) Sunshine duration climate maps of Belgium and Luxembourg based on Meteosat and in-situ observations. Adv. Sci. Res., 10, 15–19.

Kandirmaz, H.M. (2006) A model for the estimation of the daily global sunshine duration from meteorological geostationary satellite data. Int. J. Remote Sens. 27, 5061–5071.

Kerr, A. & Tabony, R. (2004) Comparison of sunshine recorded by Campbell–Stokes and automatic sensors, *Weather*, Volume 59, Issue 4, pp86-112, i-iv, https://doi.org/10.1256/wea.99.03.

Kothe, S., Good, E., Obregón, A., Ahrens, B., Nitsche, H. (2017) Satellite-Based Sunshine Duration for Europe. *Remote Sens.* **2013**, *5*, 2943-2972. https://doi.org/10.3390/rs5062943

Kothe S, Pfeifroth U, Cremer R, Trentmann J, Hollmann R. (2017) A Satellite-Based Sunshine Duration Climate Data Record for Europe and Africa. *Remote Sensing*. 9(5):429. https://doi.org/10.3390/rs9050429

Legg, T. (2014), Comparison of daily sunshine duration recorded by Campbell–Stokes and Kipp and Zonen sensors. Weather, 69: 264-267. https://doi.org/10.1002/wea.2288

Magee, N. B., Melaas, E., Finocchio, P. M., Jardel, M., Noonan, A. and Iacono, M. J. (2014) Blue Hill Observatory Sunshine: Assessment of Climate Signals in the Longest Continuous Meteorological Record in North America. *Bull. Amer. Meteor. Soc.*, **95**, 1741–1751, https://doi.org/10.1175/BAMS-D-12-00206.1.

Met Office; Hollis, D.; McCarthy, M.; Kendon, M.; Legg, T.; Simpson, I. (2018): HadUK-Grid gridded and regional average climate observations for the UK. Centre for Environmental Data Analysis, *31 January 2024*. http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb

Met Office (2019): Met Office MIDAS Open: UK Land Surface Stations Data (1853-current). Centre for Environmental Data Analysis, 17 January 2024. http://catalogue.ceda.ac.uk/uuid/dbd451271eb04662beade68da43546e1

Met Office (2023): MIDAS Open: UK daily weather observation data, v202308. NERC EDS Centre for Environmental Data Analysis, 03 October 2023. doi:10.5285/1ce37461affc43bbbd78beaaacf5911d

Naeger, A. R. and Christopher, S. A. (2014), The identification and tracking of volcanic ash using the Meteosat Second Generation (MSG) Spinning Enhanced Visible and Infrared Imager (SEVIRI), Atmos. Meas. Tech., 7, 581–597, 2014, doi:10.5194/amt-7-581-2014.

Obregón, M. A, Gallego, M.C., Antón, M., Vaquero J.M. (2020) Sunshine duration data in San Fernando (South of Spain) during 1880s: The impact of Krakatoa volcanic eruption. Geosci. Data J.. 2020; 7: 185–191. https://doi.org/10.1002/gdj3.101

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B. and Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011) Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, v12, pp2825-2830.

Perry, M. and Hollis, D. (2005), The generation of monthly gridded datasets for a range of climatic variables over the UK. Int. J. Climatol., 25: 1041-1054. https://doi.org/10.1002/joc.1161

Pfeifroth, U., Sanchez-Lorenzo, A., Manara, V., Trentmann, J., & Hollmann, R. (2018). Trends and variability of surface solar radiation in Europe based on surface- and satellite-based data records. Journal of Geophysical Research: Atmospheres, 123, 1735–1754. https://doi.org/10.1002/2017JD027418

Pfeifroth, U, Kothe, S., Drücke, J., Trentmann, J., Schröder, M., Selbach, N. and Hollmann, R., (2023): Surface Radiation Data Set - Heliosat (SARAH) - Edition 3, Satellite Application Facility on Climate Monitoring, DOI:10.5676/EUM_SAF_CM/SARAH/V003, https://doi.org/10.5676/EUM_SAF_CM/SARAH/V003.

Sanchez-Lorenzo, A. and Wild, M. (2012) Decadal variations in estimated surface solar radiation over Switzerland since the late 19th century, Atmos. Chem. Phys., 12, 8635–8644, https://doi.org/10.5194/acp-12-8635-2012, 2012.

Stanhill, G., and Cohen, S. (2005) Solar Radiation Changes in the United States during the Twentieth Century: Evidence from Sunshine Duration Measurements. *J. Climate*, **18**, 1503–1512, https://doi.org/10.1175/JCLI3354.1.

Walawender J. (2017), Development and Application of Tools for geostatistical integration of CM SAF data and surface measurements (geoIntSM), Visiting Scientist Project at DWD for the CM SAF.

Wu, B., Liu, S., Zhu, W., Yu, M., Yan, N., Xing, Q. (2016) A Method to Estimate Sunshine Duration Using Cloud Classification Data from a Geostationary Meteorological Satellite (FY-2D) over the Heihe River Basin. *Sensors* 2016, *16*, 1859. https://doi.org/10.3390/s16111859

WMO (2015), Manual on the Global Observing System. WMO-No. 544, https://library.wmo.int/records/item/58672-manual-on-the-global-observing-system-volume-i-global-aspects, updated in 2017.

7 Glossary – List of Acronyms in alphabetical order

AWSDLY	Automatic Weather Station Daily values station network
CDR	Climate Data Record
CM SAF	Satellite Application Facility on Climate Monitoring
CLS	Correlation Length Scale
CS	Campbell Stokes SD sensor
DJF	December/January/February
DNI	Direct Normalised Irradiance
DLY3208	Daily observations from Metform 3208 station network
ECV	Essential Climate Variable
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FCDR	Fundamental Climate Data Record
FMI	Finnish Meteorological Institute
GC-Net	Greenland Climate Network
GCOS	Global Climate Observing System
GP	Gaussian Process
HadUK-Grid	Met Office Hadley Centre UK Gridded datasets
ICDR	Interim Climate Data Record
ID	Identification
JJA	June/July/August
KZ	Kipp & Zonen SD sensor
MAB	Mean of the Absolute Bias
MAD	Mean Absolute Difference
MAM	March/April/May
MFG	Meteosat First Generation
MIDAS	Met Office Integrated Data Archive System
MSG	Meteosat Second Generation

NCIC	National Climate Information Centre (based at the Met Office)
NCM	National Climate Message station network
QC	Quality Control
RBF	Radial Basis Function
RMS	Root Mean Square
RMSE	Root Mean Square Error
SARAH-3	Surface Solar Radiation Data Set – Heliosat 3rd version
SD	Sunshine Duration
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SID	Surface Incoming Direct Radiation
SON	September/October/November
SYNOP	surface SYNOPtic observations station network
SZA	Solar Zenith Angle
UK	United Kingdom
WMO	World Meteorological Organization

8 Appendix

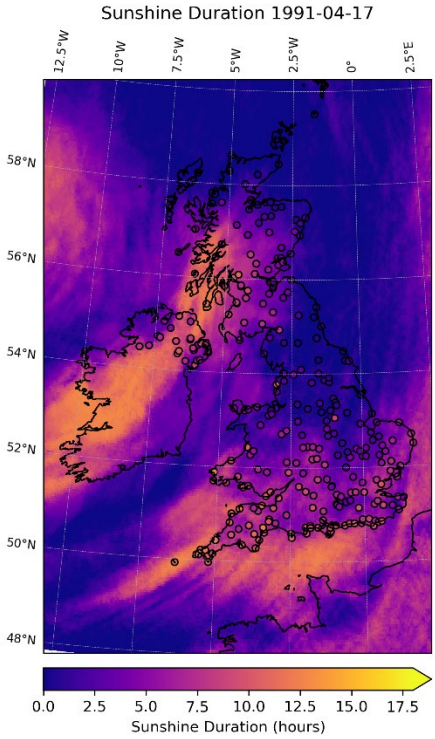


Figure 8-1: Map of SD_{sat} (background) overlayed by the colocated SD_{cs_uni} observations, which are shown by the coloured circles. For most station/satellite matchups, the agreement is quite good, but there are clear outliers. For example, the high station *SD* (bright spot) and corresponding low satellite *SD* that occurs in the Midlands at about 1.1°W, 53.1°N. The data shown are from 17 April 1991.

Table 8-1: List of stations removed from SD_{cs} and SD_{kz} through the outlier screening process. Columns 'Stn ID', 'Lat', 'Long' and 'H' provide the station identification number, latitude, longitude and height, respectively. A red-coloured box in columns '*r*', 'μ' and 'σ' indicates the failed threshold-based test(s) within each season ('DJF' = December/January/February, 'MAM' = March/April/May, 'JJA' = June/July/August, 'SON' = September/October/November). A red-coloured box in the final column (>20% >5 hours) indicates stations that have >20% of the daily $SD_{sat} - SD_{stn}$ differences that are >5 hours. See text for further details.

Stn ID	Lat (°N)	Long (°E)	H (m)	DJF			MAM			JJA			SON			>20% >5 h
				r	μ	σ	*r*	μ	σ	*r*	μ	σ	*r*	μ	σ	
			Campbell-Stokes													
107	57.144	-4.677	21													
111	56.94	-4.238	351													
113	57.206	-3.828	228													
128	57.334	-3.6	218													
130	57.335	-3.606	220													
140	57.354	-3.345	213													
147	57.006	-3.398	339													
178	56.969	-2.207	4													
394	53.445	-0.202	108													
432	52.572	1.738	4													
461	52.227	-0.465	85													
564	53.064	-0.963	70													

Stn ID	Lat (°N)	Long (°E)	H (m)	DJF			MAM			JJA			SON			>20% >5 h
				r	*μ*	*σ*	*r*	*μ*	*σ*	*r*	*μ*	*σ*	*r*	*μ*	*σ*	
608	52.073	-1.334	87													
613	51.62	-1.099	57													
715	51.68	-0.056	32													
810	50.865	0.336	18													
1070	54.934	-2.964	28													
1216	51.88	-5.124	111													
1233	51.964	-3.629	330													
1333	50.991	-4.462	142													
1341	50.686	-4.234	131													
1347	51.2	-4.113	145													
13343	54.741	-4.958	27													
30435	58.635	-3.06	35													
Kipp & Zonen																
105	56.867	-4.708	249													

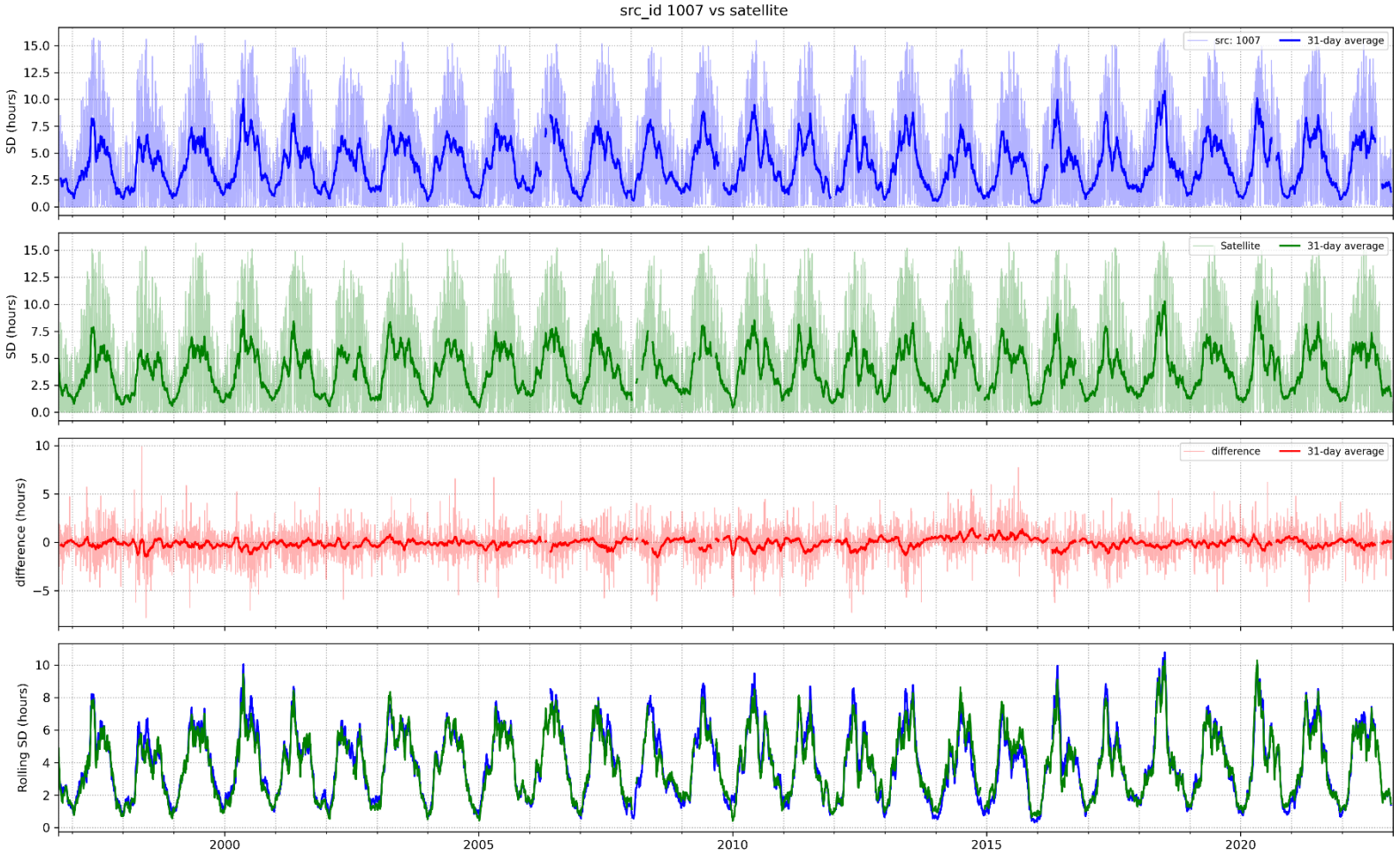


Figure 8-2: Time series of daily SD_{cs_uni} (top row), SD_{sat} (second row), $SD_{sat} - SD_{cs\_uni}$ (third row) and the 31-day average time series from SD_{cs_uni} (blue) *and* SD_{sat} (green) from station 1007.

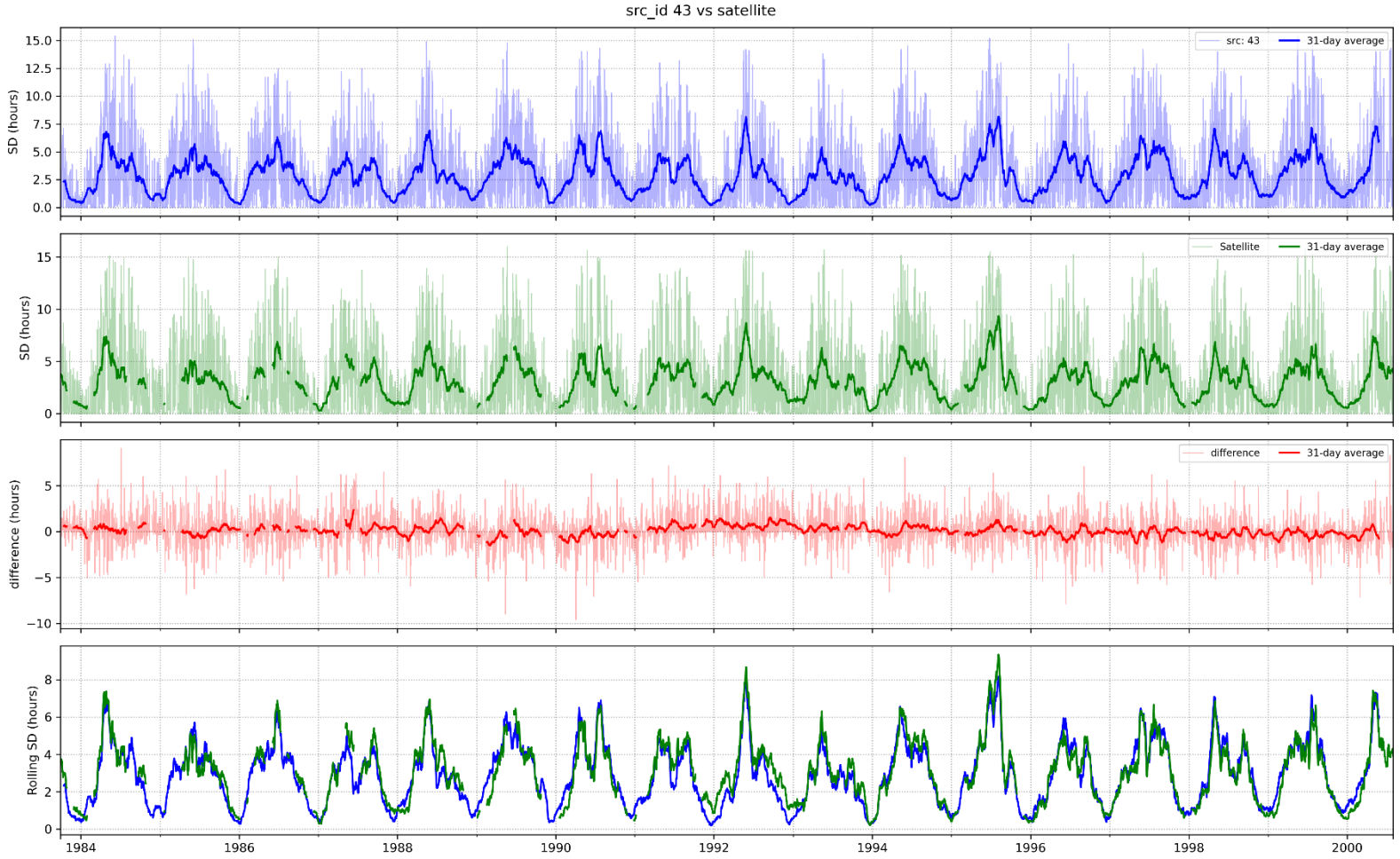


Figure 8-3: Time series of daily SD_{cs_uni} (top), SD_{sat} (second row), $SD_{sat} - SD_{cs\_uni}$ (third row) and the 31-day average time series from SD_{cs_uni} (blue) *and* SD_{sat} (green) from station 43.

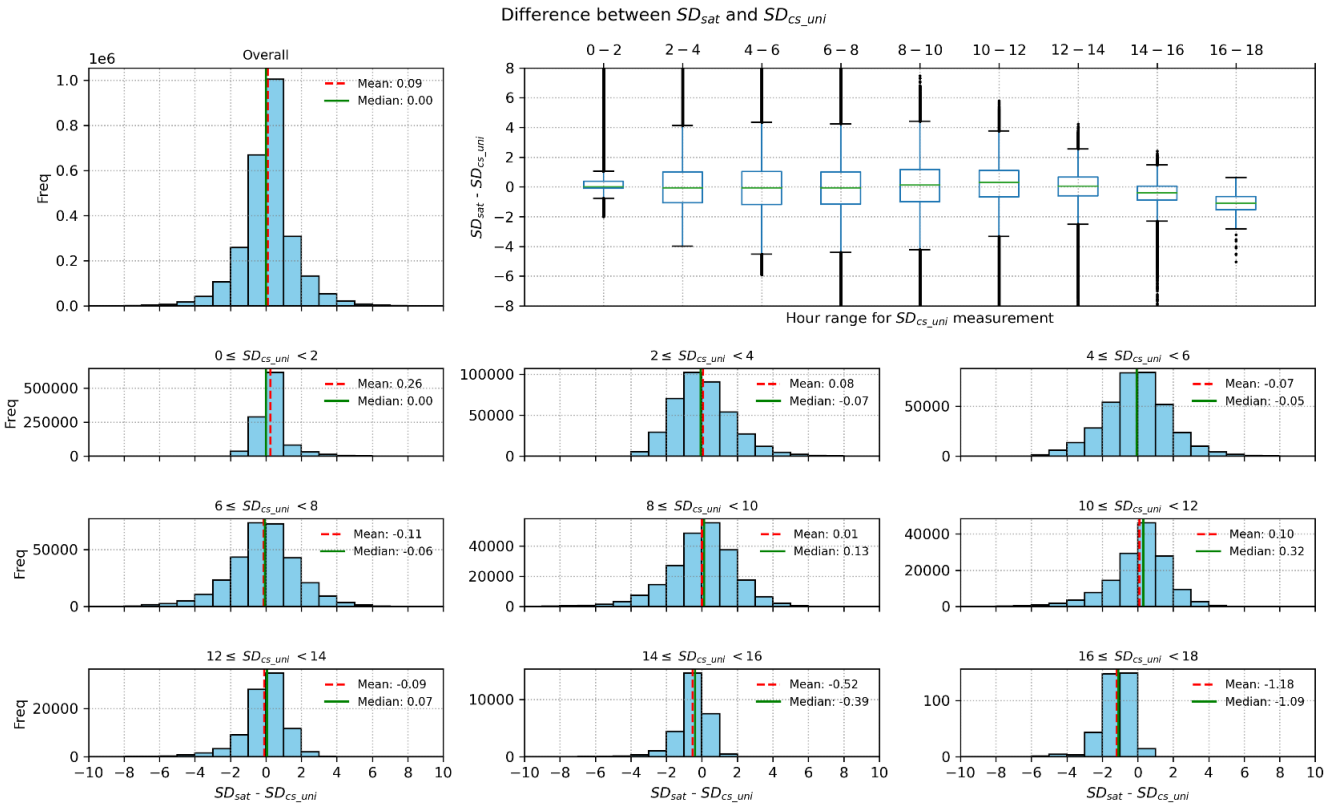


Figure 8-4: Distributions of daily SD_{sat} – SD_{cs_uni} differences (CS-unified); stations that fail the outlier screening process, listed in Table 4-2, are excluded. The top left panel shows the distribution for all machups, while the other panels show the distributions of differences partitioned by SD_{cs_uni} values. The top-right panel shows the distributions for all partitioned ranges where the central green line corresponds the 50th percentile (median) value, the box shows the 25th and 75th percentiles (or inter-quartile range), while the whiskers extend to the farthest data point within 1.5x the inter-quartile range from the box. The data points that extend past the end of the whiskers are the most extreme values.

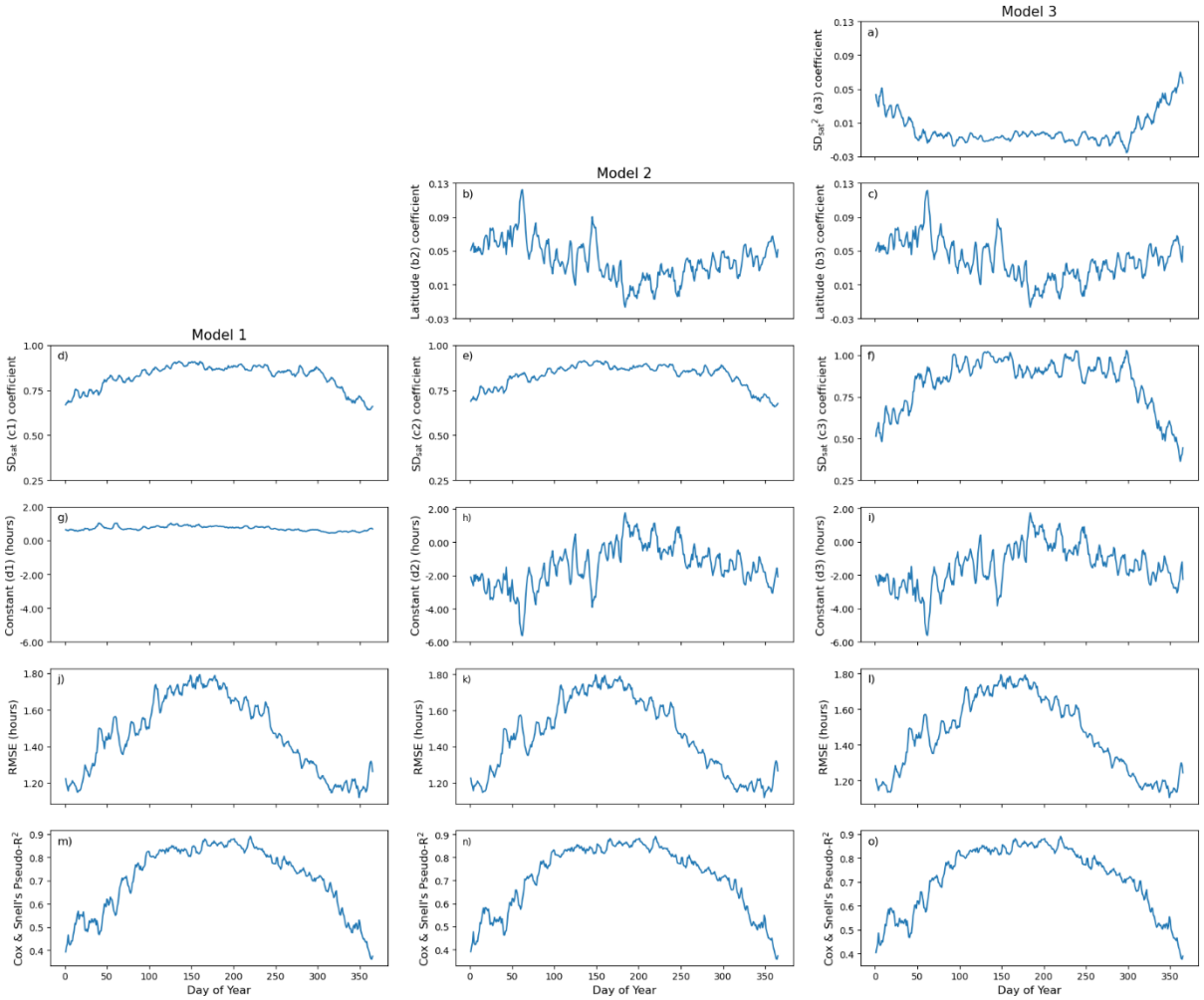


Figure 8-5: Generalised Linear Model coefficients used to blend SD_{stn} and SD_{sat} and the goodness of fit statistics for the three models trialled in this study (see Section 3.2). The top row shows the *a* coefficients (panel a), second row the *b* coefficients (panels b & c), third row the *c* coefficients (panels d, e & f), and the fourth row the *d* coefficients, or *y*-intercept, for each model (panels g, h & i). The fifth row shows the Root Mean Square Error (RMSE) for each model (panels j, k & l), while the bottom row shows the Psuedo-R_2 (panels m, n & o; see Section 3.2). The results for Model 1 (GLM 1; Equation 3-3) are shown in the first column, Model 2 (GLM 2; Equation 3-4) in the second column and Model 3 (GLM 3; Equation 3-5) in the third column. Note the different *y*-axis range for each category of model coefficient; Figure 4-13 shows the same data where the panels showing the coefficients have the same *y*-axis range.

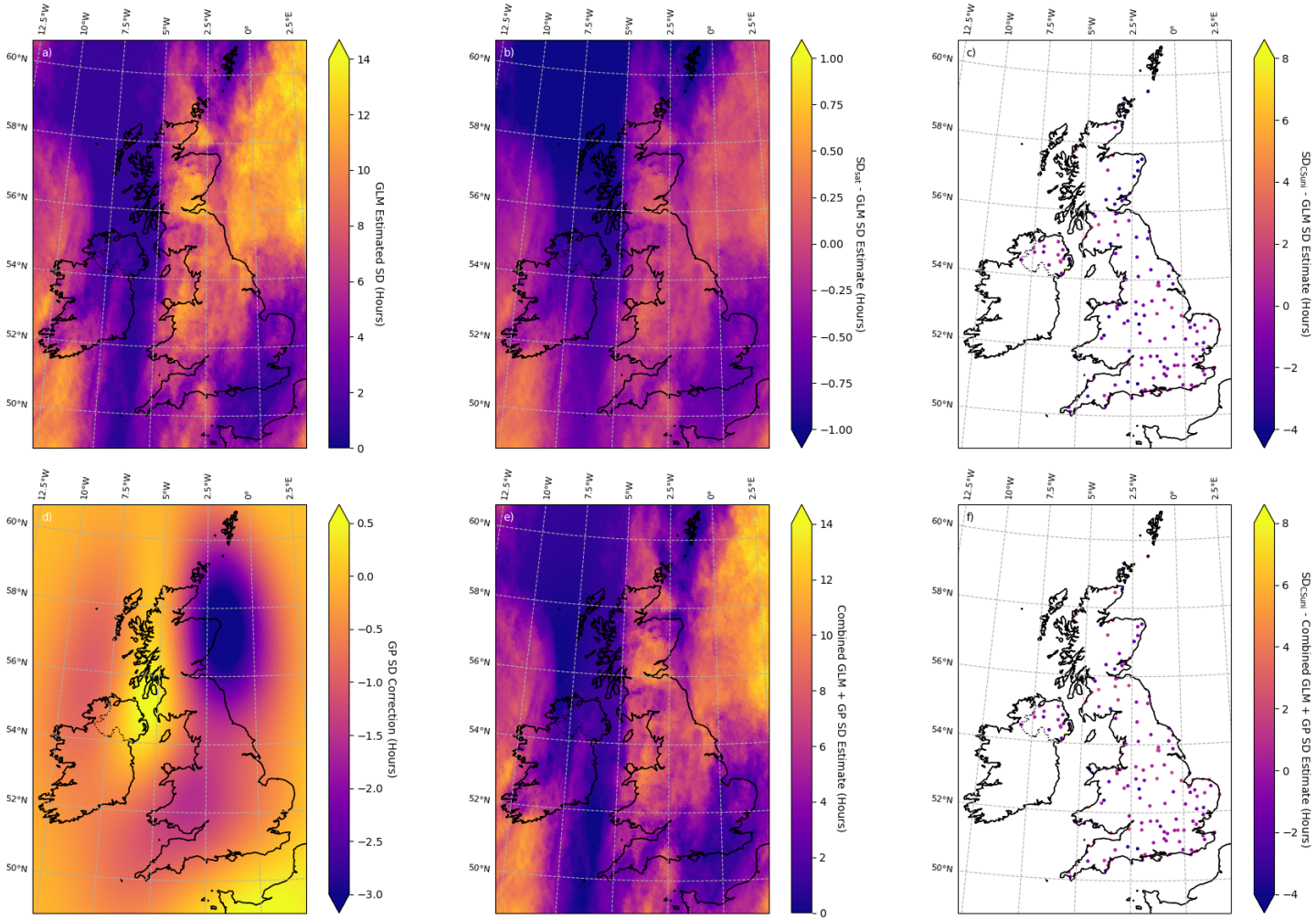


Figure 8-6: Example model output for 1 June 2003 showing the a) Generalised Linear Model 2 *SD* (GLM 2, Equation 3-4), b) SD_{sat} – GLM 2 *SD*, c) SD_{cs_uni} – GLM 2 *SD*, d) the GP *SD* correction, e) the combined GLM 2 + GP corrected *SD* output, and f) SD_{cs_uni} – combined GLM 2 + GP modelled *SD*.

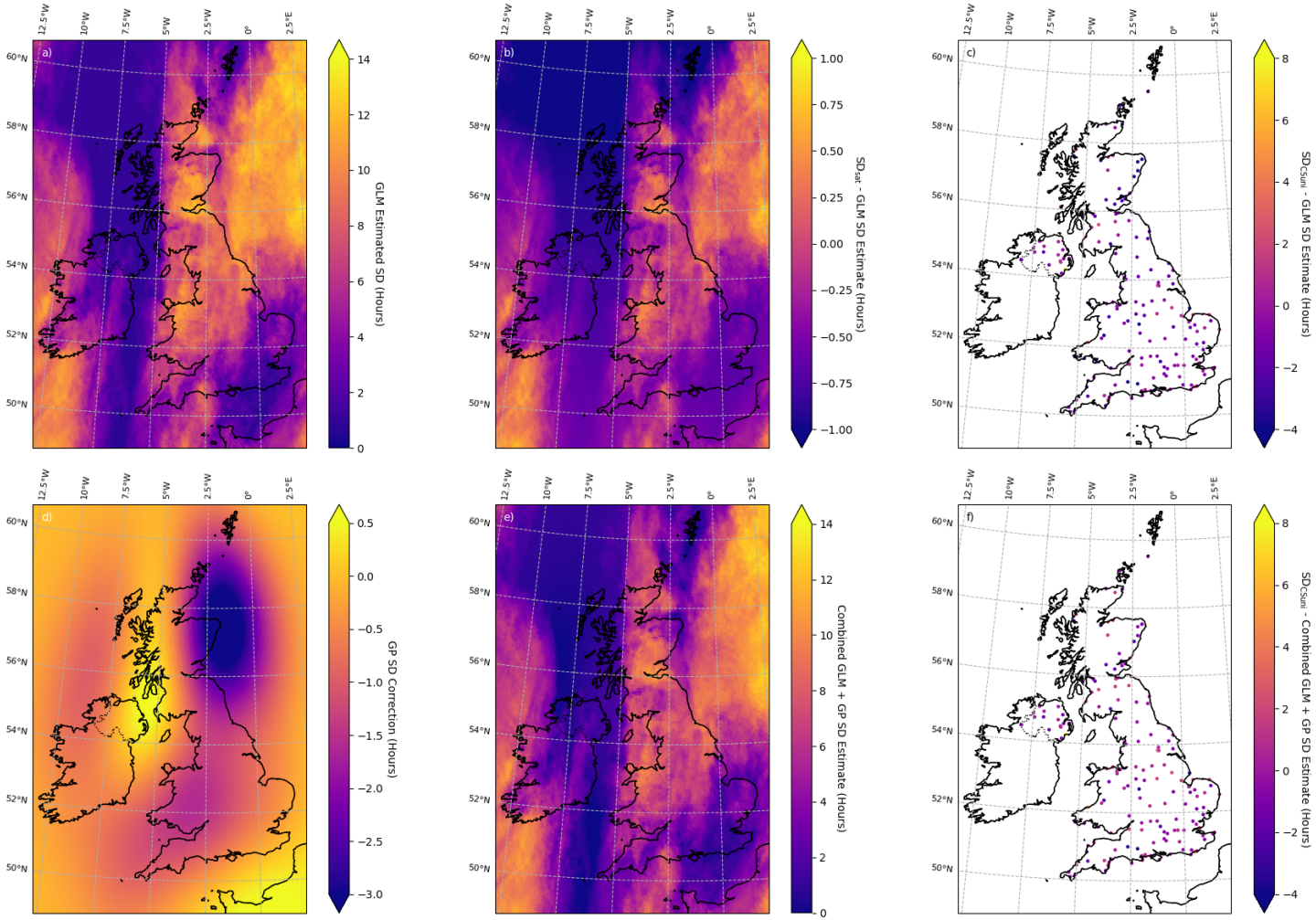


Figure 8-7: Example model output for 1 June 2003 showing the a) Generalised Linear Model 3 *SD* (GLM 1, Equation 3-5), b) SD_{sat} – GLM 3 *SD*, c) SD_{cs_uni} – GLM 3 *SD*, d) the GP *SD* correction, e) the combined GLM 3 + GP corrected *SD* output, and f) SD_{cs_uni} – combined GLM 3 + GP modelled *SD*.