

EDUCE- flagging report for spectral data from Jokioinen, Finland flag_FIJ

Authors/evaluators: JE Williams, PN den Outer and H Slaper (RIVM)

FP1 : Flagging results for Jokioinen, Finland:

General Introduction

Within the context of the EU-project EDUCE a database has been developed for ground based solar UV-spectra. This database includes UV-spectra from measurement stations covering a wide range of instruments and locations over the European continent. The use of these spectral measurements in UV-transfer modelling studies, and trend, climatology and effect assessments usually requires high quality data sets. Therefore, as part of the EDUCE project a quality assurance program has been set up to assess several quality aspects of the measured spectra. Within this context RIVM has developed a quality control/ quality assurance package for spectral UV-measurements called SHICrvm, which will be implemented at the EDUCE database for quality assurance checks and data quality flagging. The SHICrvm package has been distributed to all operators in the EDUCE-project and can be used as QC-tool at the home sites. The package could also be used by end-users of the data to ensure a high degree of standardisation of the data-analysis, a correction of wavelength scale errors and spike identification/correction. The methods are being used to analyse all spectral data in the EDUCE-database, but no data-corrections will be applied at the database. It should be stressed that the operators have full responsibility for the data-quality at their sites and the quality-control procedures conducted. The quality assurance at the database can only be seen as a second line of protection against problematic or erroneous data to warn both, users and operators.

Introduction to flagging

To check the performance of operational spectral instruments and to analyse availability and reliability of the spectral irradiance data at the European UV database an analysis by the SHICrvm code (version 3_093) has been performed on the UV-measurements which can currently be retrieved from the EDUCE-web site. The aim is to identify corrupt spectra and periods of time when the data measured by a particular instrument may have been erroneous due to difficulties introduced by such problems as calibration errors, ageing of components or extreme operational conditions. This information is a matter of importance for the associated instrumental operators as a means of improving/maintaining the performance of a specific instrument, but also gives an overall information on some quality aspects of the data in the database.

The quality flags identified in this report are:

- wavelength scale errors (summarised in two wavelength regions: shift_1 and shift_2)
- identification of the lowest irradiance level which is accurately measured by the instrument (start_irradiance_flag)
- the identification of spikes and significant deviation in the local shape of the measured spectrum (spike+local shape flag)
- irradiance scale errors (rough indication of potential errors in the irradiance calibration in the UVA range of the spectrum above 325 nm up to 400 nm); (transmission_2)

The SHICrvm package provides additional flagging information, which is not reflected in this report (see SHIC website for detailed user documentation and package download: <http://www.rivm.nl/shicrvm>).

All quality flags are given by means of a coloured indicator:

GREEN	refers to spectra that meet the EDUCE-quality standard
YELLOW	refers to spectra where some (minor) deviation occurs (problematic for specific applications only)
RED	refers to spectra that have errors which might be problematic for certain applications,
BLACK	refers to spectra that do not meet the minimum quality requirements, and that should normally not be used in data-analysis. Normal data-retrieval will not include black flagged spectra
GREY	identifies suspicious spectra, for which the evaluation is not regarded reliable, this could be due to noise in low-irradiance spectra, or spectral shape errors due to highly variable conditions during a scan, or measurement errors

It should be noted that in normal operation of instruments some yellow and even some red flags might occur occasionally in relation to extreme conditions. For certain applications the inclusion of such spectra in the

database is still highly relevant, and some of the possible small errors can be corrected for. Thus, the flags should not be interpreted as just GREEN is OK and everything else is not!
A brief description for the flagging-criteria is given in the Flagging definitions section below.

Flagging definitions :

All flagging statistics for SZA $\geq 85^\circ$ are used in the presented analysis. The analysed spectra for SZA $> 85^\circ$ are not used to avoid an over representation of data with very low irradiance levels close to the noise level of instruments, which will lead to an increase in GREY-flagged data. For the specific details as to how each of the quality checks are performed the reader is referred to the second annual report for the EDUCE project and the SHIC users manual. The definition of the thresholds for especially the irradiance scale flag (transmission) is still the subject of some discussion between the partners involved in the EDUCE project. This flag is therefore likely to change in the final implementation at the database. To clarify this point, the thresholds for each flag associated with each of the criteria listed in the flagging tables are briefly described below.

Shift1/Shift2 : GREEN $< 0.1 \text{ nm}$ < YELLOW $< 0.2 \text{ nm}$ < RED $< 0.4 \text{ nm}$ < BLACK.

GREY flag associated with very noisy measurements for which the shift result is not reliable.

Spike+local shape : spikes identified if the ratio of two measured neighbouring spectral irradiances deviates from the modelled ratio of two neighbouring irradiances:

If modelled and measured ratios above the 'starting irradiance' over the full spectral range deviate: GREEN $< 25\%$ < YELLOW $< 50\%$ < RED and spikes are identified as BLACK if $> 200\%$ deviation occurs for a measured ratio compared to a modelled ratio; here the spectral irradiance at a certain wavelength is divided by the median irradiance of 10 consecutive readings around the analysed wavelength. Black spikes are reported here as COR (corrected), which implies that SHIC corrected these spikes. The correction will not be applied at the database and the COR spectra show up as BLACK at the database, and thus should be regarded as BLACK.

Start irradiance : lowest irradiance where 5 consecutive ratios of two subsequent spectral readings are within 25% of the modelled ratios:

GREEN 0.0005 Wm^{-2} < YELLOW < 0.0015 < RED $< 0.005 \text{ Wm}^{-2}$ < BLACK.

Transmission 2 (325-400nm) irradiance ratios between clear-sky and measured : 1.25-0.1 (GREEN), 1.5-1.25 or 0.1-0.05 (YELLOW), 2.0-1.5 or 0.05-0.01 (RED) and > 0.2 or < 0.01 (BLACK). It should be noted that the transmission flagging, which is not included in the graphical representation of the results (see below), is still in discussion and at the database YELLOW and RED flags will not be issued, and criteria for BLACK flagging might change for the data-analysis at the database. Results in the reports will however, for matters of consistency, remain unchanged.

Results are given for each measurement year in a graph and a table:

- the graph provides an overview of the fraction of flags in each colour category for each day in the year; and the number of spectra measured for each day (black line, right Y-axis); white areas indicate that no measurements are available at the database for certain days in the year; for each spectrum four flags are included in the colour graph: shift in region 1, shift in region 2 (if spectra include this wavelength range), spike and local shape flag, and start irradiance flag. The irradiance flags are reported for the UVA region above 325 nm, and are not included in the statistics of the colour graph.
- the table gives the yearly sums of spectra in each of the flagging categories for each of the colours; Spikes appear in the column Cor (corrected) and are not included in the BLACK flags for the spike and local shape flag in the presented analysis. At the database the spikes will be included in the BLACK flags in the spike and local shape flag!

Overall comments on Flagging results

The result of the analysis is briefly commented for each year of data. To allow some comparison of results for various instruments we use standard comments regarding the potential usefulness of the data set for climatological studies and the overall impression of the quality flagging. It should be noted that data sets which are apparently referred to as high quality, can still include considerable errors in the irradiation scale since the checks for the irradiance scale can only provide a rough indication! Furthermore, it should be noted that for climatological studies data sets with limited quality single spectra could still be very useful if a large measurement frequency occurs in combination with random errors in single spectra and/or if data-correction techniques are applied.

Data coverage comments:

<20% of days covered	low annual coverage; limited potential for use in climatological studies
20-40% of days covered	limited annual coverage; some potential for use in climatological studies
40-80% of days covered:	moderate annual coverage; medium potential for use in climatological studies
80-95% of days covered:	high annual coverage; high potential for use in climatological studies
>95% of days covered	Extensive annual coverage; excellent potential for use in climatological studies

Data quality comments:

If overall fraction RED, BLACK , GREY and Corrected (reflecting spikes):

< 2%	very high fraction of potential high quality spectra
2-5%	high fraction of potential high quality spectra
5-10%	useful fraction of potential high quality spectra
10-20%	a fraction of spectra is of questionable quality
20-30%	a significant part of the dataset is of questionable quality
30-50%	large part of the data is of questionable quality
> 50%	data of questionable quality

Measurements details :

Location : Jokioinen, Finland

Elevation (m) : 107

Instrument name : bre107

Instrument type : Brewer #107 Double Monochromator MK III

Wavelength range (nm) : 280-363 nm, 0.5 nm step

Lat, Long : 60.814 N, 23.499 EL (wrong in database)

Years of submitted data : 4, 2 additional submissions are sparse

No spectra submitted (per year) : 3362 (1995), 192 (1996), 3825 (1997), 5858 (1998), 4 (2000), 25 (2001)

No spectra (total submitted) : 13266

Slit width (FWHM) (nm) : 0.55

SHIC version for analysis : 3_093

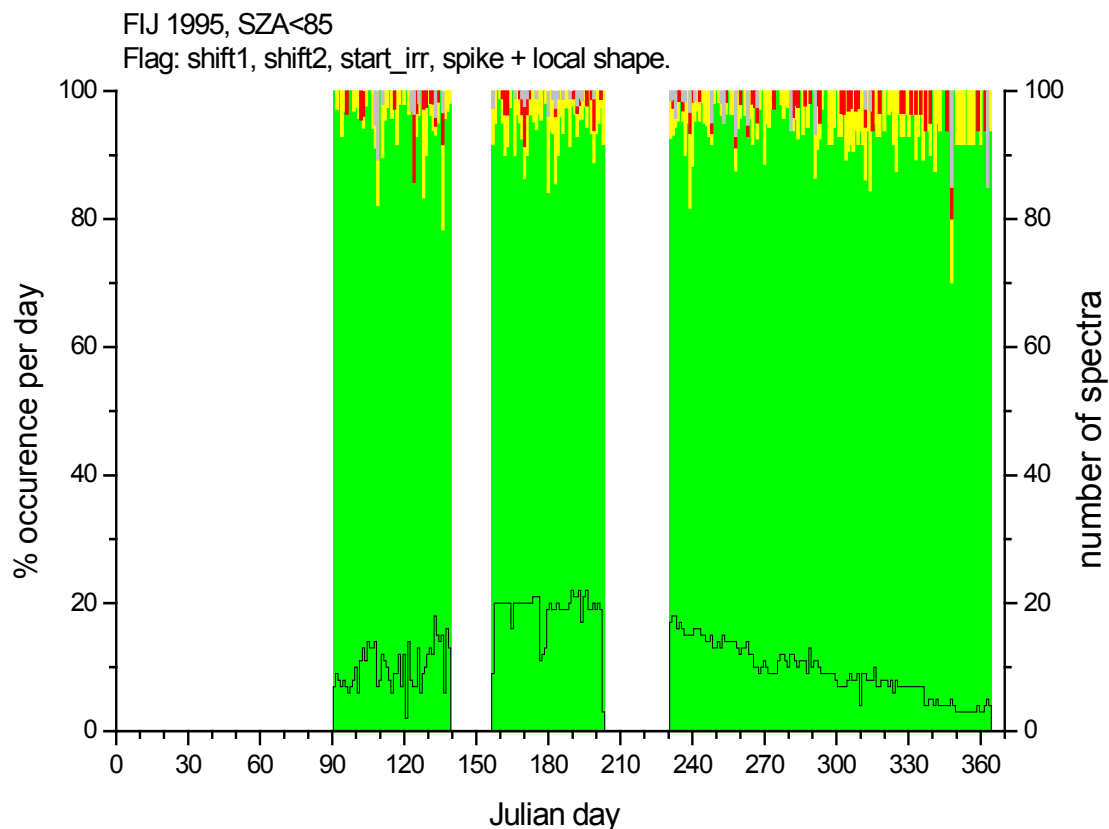
Special comments: 2000 and 2001 data not analysed due to total no. spectra per year < 50.

The authors are unsure whether the data sets were analysed before submission.

Responsible operator/PI: Tapani Koskela (FMI); tapani.koskela@fmi.fi

Operator comments:

The data has been corrected for spikes before submission to the database. More data to be submitted in the near future. A different slit-function should be used before and after the JD113 of 1997.

Results of flagging statistics:**1995:**

flag	Green %	Yellow %	Red %	Black %	Grey %	Cor. %	Green	Yellow	Red	Black	Grey	Cor.	Num
shift1_flagging	98.4	0	0	0	1.6	0	2574	1	0	0	42	0	2617
shift2_flagging	99.2	0	0	0	0.8	0	2596	0	0	0	21	0	2617
start_irradiance_flag	99.6	0.2	0.2	0	0	0	2607	4	6	0	0	0	2617
Spike+local_shape flag	99.1	0.5	0	0	0.5	0	2593	12	0	0	12	0	2617
Transmission_2	98.5	1	0.1	0	0.5	0	2577	26	2	0	12	0	2617

Comments :

Moderate annual coverage (approximately 60%); medium potential value for climatological studies

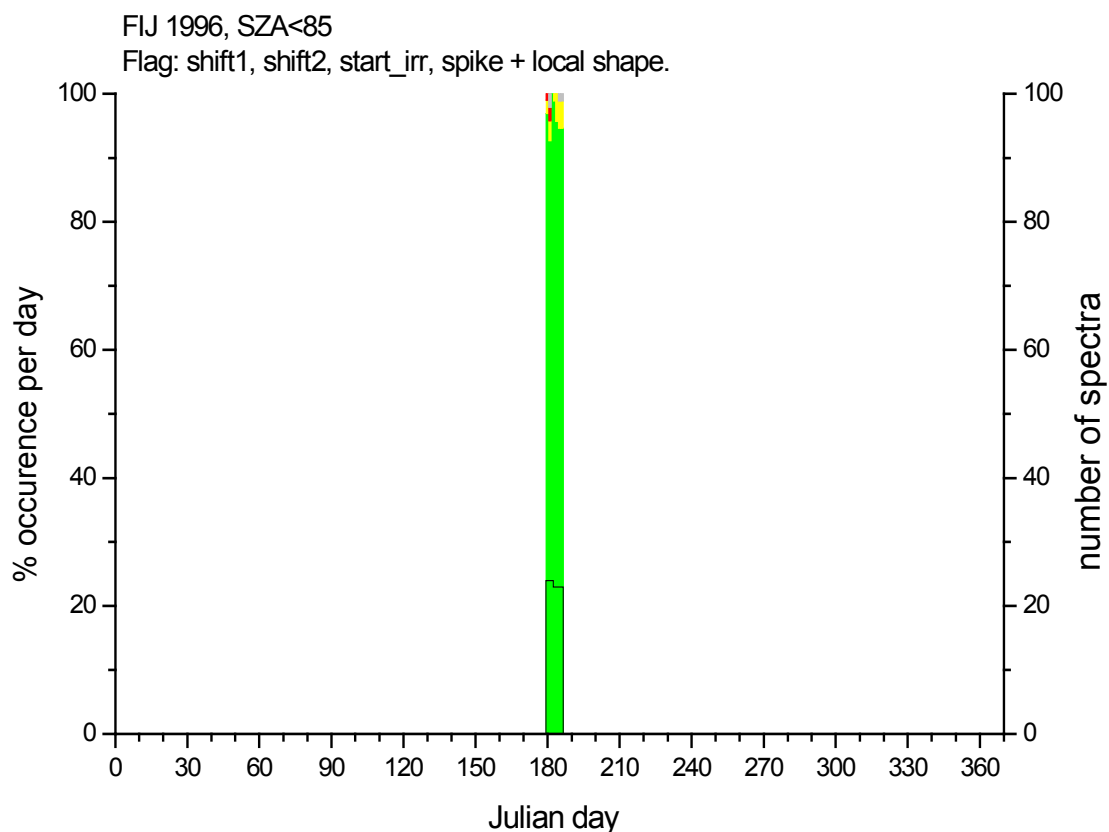
Overall data-quality impression: very high fraction of potential high quality spectra

No black flags occur for any of the chosen flagging categories (with red flags < 0.25%).

Both the shift1 and shift2 flags indicate that the wavelength scale is well calibrated compared with an extraterrestrial spectrum across the entire wavelength region.

No spectra with spikes are reported.

The distribution of errors is fairly uniform throughout the dataset.

1996:

Flag	Green %	Yellow %	Red %	Black %	Grey %	Cor. %	Green	Yellow	Red	Black	Grey	Cor.	Num
Shift1_flagging	98.8	0	0	0	1.2	0	162	0	0	0	2	0	164
Shift2_flagging	98.8	0	0	0	1.2	0	162	0	0	0	2	0	164
start_irradiance_flag	100	0	0	0	0	0	164	0	0	0	0	0	164
Spike+local_shape flag	84.1	10.6	1.8	0	0	3.5	143	18	3	0	0	6	170
Transmission_2	98.8	1.2	0	0	0	0	162	2	0	0	0	0	164

Comments :

Low annual coverage (approximately 5%). Limited potential value for climatological studies

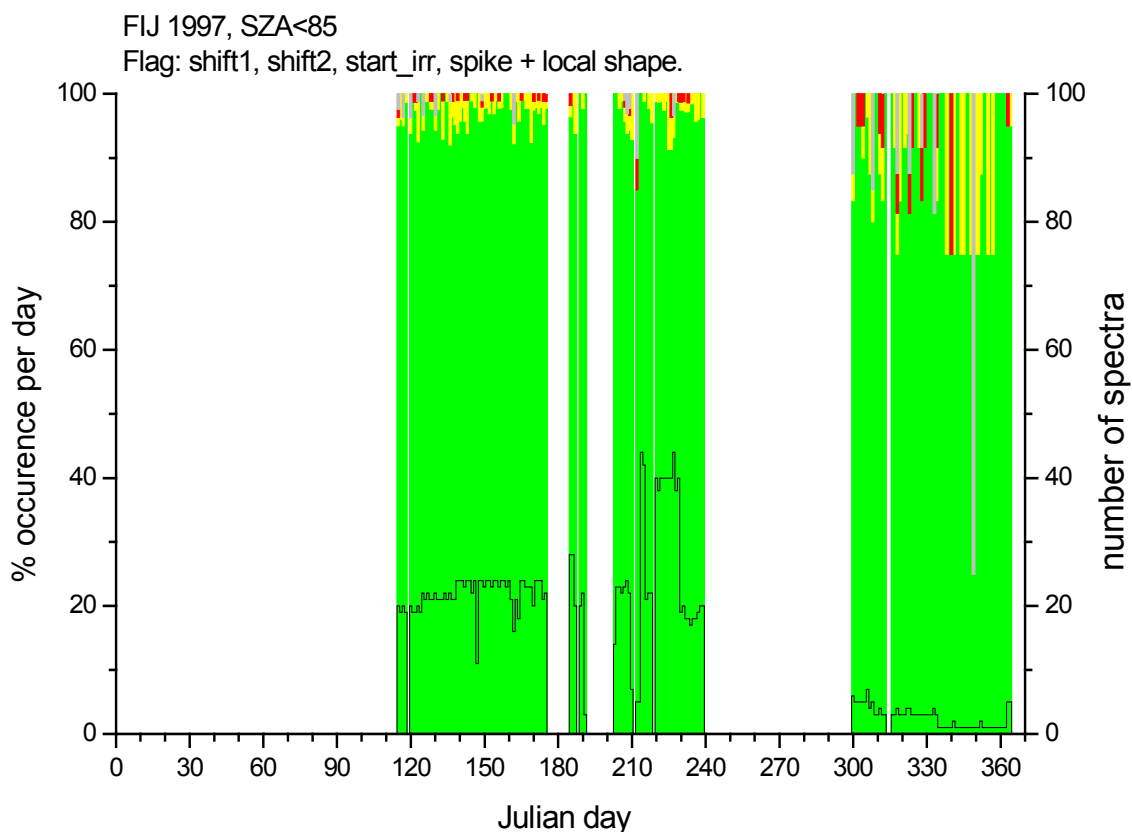
Overall data-quality impression: high fraction of potential high quality spectra

No black flags occur for any of the chosen flagging categories (with red flags < 2%). A very small dataset means that the few grey flagged spectra contribute >1%.

Both the shift1 and shift2 flags indicate that the instrument is well calibrated compared with an extraterrestrial spectrum across the entire wavelength region.

6 spectra (3.5%) with spikes are reported.

The distribution of errors is uniform throughout the dataset.

1997:

	Green %	Yellow %	Red %	Black %	Grey %	Cor. %	Green	Yellow	Red	Black	Grey	Cor.	Num
Flag													
Shift1_flagging	99.2	0	0	0	0.8	0	2480	0	0	0	21	0	2501
Shift2_flagging	99.3	0	0	0	0.7	0	2483	0	0	0	18	0	2501
Start_irradiance_flag	99.8	0.2	0	0	0	0	2497	4	0	0	0	0	2501
Spike+local_shape	88.4	9.1	1.9	0	0.5	0.1	2214	228	47	0	12	3	2504
Transmission_2	97.1	2	0.3	0	0.6	0	2427	50	7	0	15	0	2499

Comments :

Moderate annual coverage (approximately 50%). medium potential for use in climatological studies.

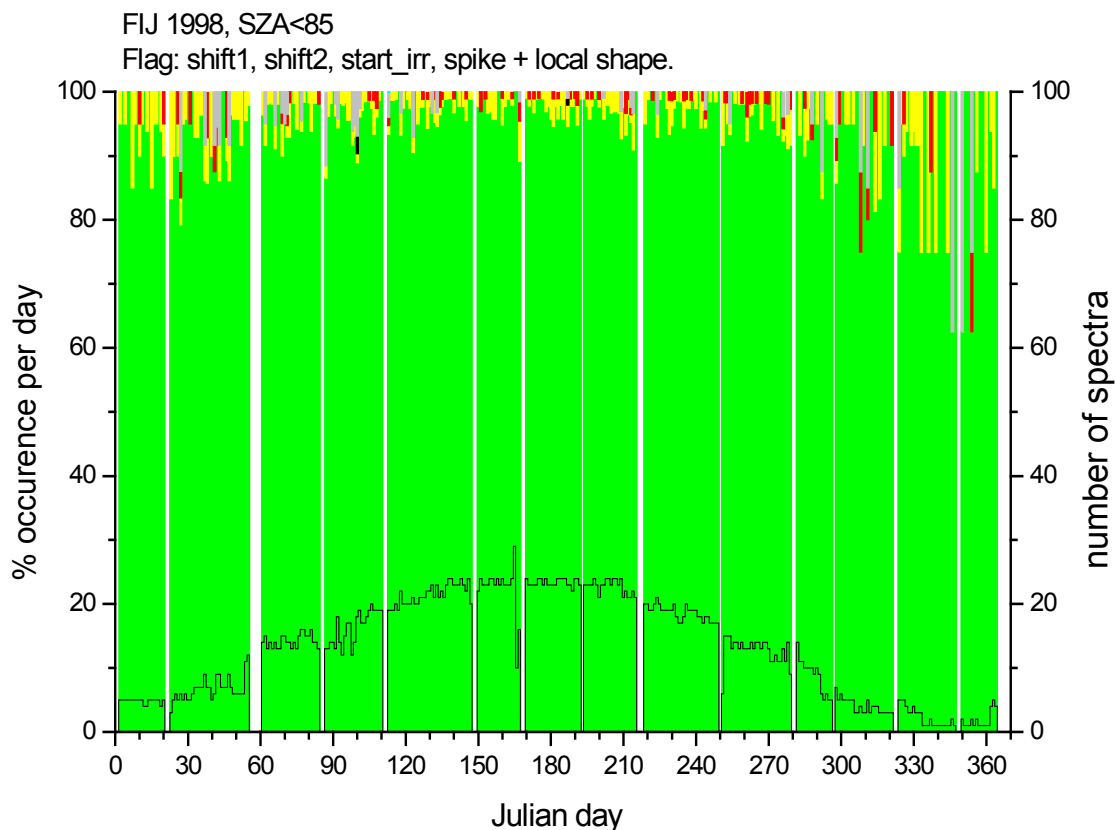
Overall data-quality impression: a high fraction of potential high quality spectra.

No black flags occur for any of the chosen flagging categories (with red flags < 2%). The few grey flagged spectra contribute <1%.

Both the shift1 and shift2 flags indicate that the instrument is well calibrated compared with an extraterrestrial spectrum across the entire wavelength region.

3 (0.1%) spectra with spikes are reported.

The distribution of errors is non-uniform across the dataset, with more red and yellow flags occurring after Julian day 300.

1998:

Flag	Green %	Yellow %	Red %	Black %	Grey %	Cor. %	Green	Yellow	Red	Black	Grey	Cor.	Num
Shift1_flagging	98.9	0.1	0	0	1	0	4446	3	0	0	46	0	4495
Shift2_flagging	99	0	0	0	1	0	4450	1	0	0	44	0	4495
start_irradiance_flag	99.9	0.1	0	0	0	0	4489	3	1	2	0	0	4495
Spike+local_shape	88.5	8.7	1.9	0	0.6	0.3	3990	390	85	1	29	13	4508
Transmission_2	97.9	1.1	0.1	0	0.9	0	4401	49	4	0	41	0	4495

Comments :

Extensive annual coverage (approximately 95%); excellent potential for use in climatological studies

Overall data-quality impression: a high fraction of potential high quality spectra.

A small number of black flags occur but, considering the size of the dataset, this indicates an excellent set of irradiance measurements (with red flags < 2%). No real difference between the previous year.

Both the shift1 and shift2 flags indicate that the instrument is well calibrated compared with an extraterrestrial spectrum across the entire wavelength region.

13 (0.3%) spectra with spikes are reported.

The distribution of errors is non-uniform across the dataset with more yellow and red flags occurring towards the end of the dataset.