

Flagging reports for data submitted to the EDUCE database

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

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 one of the participants (RIVM< The Netherlands) 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, but it cannot be seen as a full proof analysis of the data-quality. In particular the irradiance scale errors can only be identified if the error is very large, in view of the large variability in atmospheric conditions.

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 $\leq 85^\circ$ 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. This flag is only included for datasets which include measurements which cover the UVA region of the spectrum.

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 spectra is of questionable quality
30-50%	large part of the data is of questionable quality
> 50%	data of questionable quality

It should be noted that it might be possible to improve or correct the data of questionable quality by means of additional quality checks, an improved instrument characterisation (slit function), and/or data-correction techniques (such as the use of SHIC to correct wavelength shifts and instrumental spikes). Furthermore, it could be possible that for certain applications the data-quality is still sufficient, depending on the specific requirements of the users and the type of error occurring. Some of the errors might be corrected by using the SHICrvm package. Nevertheless, data of questionable quality should not be used without specific considerations and/or corrections being made by the data-user.

Special comments: These are comments which are specific to the instrument and location relating to certain trend observable in the datasets.

Operator comments: These are comments and suggestions made by the instrument operator concerning the analysis of the datasets. Typical remarks cover instrument stability, possible/planned re-submission of data and changes in the maintenance and/or components of the measuring device.

Relevance of the analysis to the current database

Due to the fact that instrument operators can improve on their own submitted data by updating both the irradiance files and slit functions stored at the database, a time stamp is included for the analysis for each station, which provides details concerning the date at which each dataset was retrieved. In certain instances these reports have provided the impetus for re-submission of the data, due to the detection of a high number of errors with the original dataset. Moreover, an objective of EDUCE is to allow flagging to take place AT the database, so as to provide some indication of the quality of the datasets for any potential users. Therefore, the following reports should be viewed as an initial overview of the data for 2003.