Structure of the Lindenberg, Germany spectral radiation data

Josh Peterson[1], Dale Northcutt[1], Frank Vignola[1], Kevin Van Den Wymelenberg[1], Lionel Doppler[2]

1 University of Oregon, Eugene (United States)

2 Deutscher Wetterdienst, Meteorologisches Observatorium Lindenberg – Richard Assmann Observatorium(DWD, MOL–RAO), Lindenberg (Tauche), Germany

Abstract

The German Meteorological Service (Deutscher Wetterdienst, DWD) site in Lindenberg, Germany has been making solar spectral measurements since the middle of 2014 using a PSR Direct Normal spectroradiometer and provided the data. The spectral data set is archived and presented in a comprehensive file format. This article describes the format of the spectral files and the complementary information used to give context to the data. The format utilizes month blocks and data is reported in one-minute time intervals. The spectral data is a direct normal measurement from 300 nm to 1020 nm in roughly 0.7 nm increments. A variety of time stamps are included in the data file to facilitate the use of the data. The files contain broadband metrological data as well as other meteorological information gathered at the site. The files also contain detailed header rows about the site location, instruments used, calibration values utilized, and uncertainties in the calibration values.

Keywords: *solar radiation, direct normal spectrum, PSR spectroradiometer*

Introduction: DNI Spectral Data File Format

Information on the spectral distribution of incident radiation has become increasingly important for a variety of uses; from forecasting solar system performance, to daylighting, to photovoltaic arrays. The Lindenberg Meteorological Observatory [1] has been making spectral measurements since June 2014, using a Precision Solar Spectroradiometer (PSR) [2, 3, 4] manufactured by PMOD. The PSR spectroradiometer makes direct normal measurements. The wavelength range of the instrument is from 300 nm to 1020 nm in 0.7 nm increments.

The spectral data described by this document runs from June 2014 through September 2018. During this time there were several breaks in the data set. Most of these gaps are the result of maintenance and recalibration. Data are presented every minute over the entire 24-hour period of each day.

The PSR takes 10 samples over an approximate 10 second time interval. From these samples the average and standard deviation of the spectral data are calculated. This is done for each specified wavelength. The process is repeated at the next time interval. The time interval between measurements varies drastically depending on the date; from 70 seconds in 2014 and 2015 up to 10 minutes in later years. None of the time intervals is exactly 1-minute. Because the time interval of the PSR was not 1- minute, the spectral data is aligned with the 1-minute output files as best as possible.

The spectral data is presented in a file format that provides the user with supplemental auxiliary information that gives context for the spectral data gathered by the spectroradiometer. Key features of the file format include:

- General information about the station.
- Information on the specific instruments used to make each measurement, including the model number, calibration values, and the uncertainty in the measurement value.
- Various formats of date and time.
- The solar position (SZA and AZM) and extraterrestrial radiation (ETR and ETRn).

- Various supplementary broadband irradiance measurements (Global horizontal, Direct normal, Diffuse horizontal) as well as longwave irradiance, air temperature, and relative humidity.
- An abridged version of the standard deviation of the spectral data.
- Spectral irradiance data at each wavelength.

This combination of measured and calculated values offers a more comprehensive context for evaluation of the data. The purpose of this document is to discuss the format of the data files and how each value was obtained.

The files are csv files separated into month blocks that typically range in size between 250 and 450 MB. The large size of the files is due to the number of spectral values. A schematic diagram of the file format is shown in Figure 1. This article will discuss each of the areas shown in the figure in the following order.

- Header Section 1 contains general information about the station.
- Header Section 2 contains information about each column.
- Data Section 3 contains non-measured quantiles such as: date, time, solar position, and extraterrestrial radiation.
- Data Section 4 contains measured irradiance quantities such as: GHI, DNI, DHI broadband information along with other metrological data.
- Data Section 5 contains the spectral data set. There are 1024 columns of spectral irradiance data.
- Appendix A is a glossary of commonly used terms.
- Appendix B gives a list of the data in each column and the column numbers for quick access.
- Appendix C discusses the file containing the standard deviation of each spectral measurement.

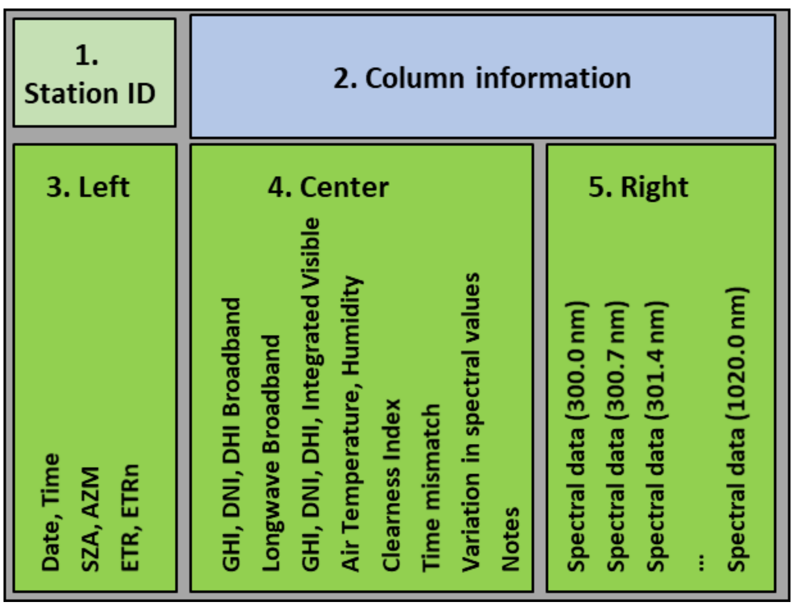

Figure 1: Schematic diagram of the file structure. The different sections of the file are labeled 1 - 5. The figure is not drawn to scale. The wavelength values listed in region 5 are approximate.

In addition to the spectral irradiance file, a file containing the standard deviation of the spectral irradiance during each measurement period for each wavelength is provided. The operational procedure of the PSR instrument makes several measurements during each time interval. These measurements are averaged to produce the spectral irradiance file. The standard deviation of the spectral measurements is also calculated. The file containing the standard deviation of the spectral irradiance is discussed in Appendix C.

File structure header section 1.
Station ID information

The upper left corner of each file contains two columns with useful information about the file. An example is shown in Figure 2. The sample shown is of columns 1-3 and rows 1-7.

	1	2	3
1	Station_Location	Lindenberg_Tauche_Germany	-
2	Latitude_(+N)	52.209	-
3	Longitude_(+E)	14.121	-
4	Altitude_(m)	750	-
5	TimeZone_(+E)	1	-
6	Year//Month	2018//01	-
7	-	-	-

Figure 2: A sample data set of the data contained in header section 1 of the file structure. columns 1-3, rows 1-7.

- **Station location** is the city, region of country, and country name of the station. The three names are separated by an underscore "_".

- **Latitude, longitude, and altitude** of the station. The latitude and longitude are reported in degrees with a decimal point representing fractions of a degree. The latitude and longitude are given to an accuracy of the ±1000 meters. The longitude of the station is given as a positive number as East is defined to be positive. The altitude of the station is given in meters above sea level.

- **The time zone** of the station. The time zone is useful for calculating the sun's position in the sky. Because the site is located east of the Prime Meridian, the time zone is a positive number as is conventionally written.

- **The year and month** of the file block are separated by double forward slash marks "//". This technique prevents some programs, such as Excel, from auto formatting dates and times into their predetermined format. By using the double forward slash, the information will not be recognized as a date and the format of the file will be preserved.

File structure header section 2.
Column header information

The header rows in the column information section contain information about each column. There are 9 header rows, with 6 rows of predefined values and 3 empty rows to allow space for notes.

A sample data set highlighting the header rows is shown in Figure 3. The screen shot is of columns 12-33 and rows 1-9 and. In Figure 3, columns 15 - 31 have been condensed to allow for viewing of the spectral data header rows.

	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	Type_of_measurment	GHI	DNI	DH	Lo	GH	DN	DH	Te	Re	Cle	Sp	std	std	std	std	std	std	std	std	-	DNI_Spectral
2	Instrument	CMP22(020074)	CHP1(030340)	CM	PIF	Cal	Cal	Cal	-	-	Cal	UT	PSI	PSI	PSI	PSI	PSI	PSI	PSI	PSI	Wavelength(nm)	302.06
3	Responsivity_(microV/W/m^2)	9.33_microV/(W/m^2)	9.83_microV/(W/m^2)	9.6	4.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Calibration_Factor((W/m^2/nm)/counts)	-
4	Responsivity_Uncertainty(U95%)	3	1.5	3	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Uncertainty(U95%)	6.6
5	Units	W/m^2	W/m^2	W/	W/	W/	W/	W/	De	%	Un	Se	W/	W/	W/	W/	W/	W/	W/	W/	Units	W/m^2/nm
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	W/m^2	-	-	-	40	-	36	-	-	-	-	se	-	-	-	-	-	-	-	-	-	-
9	ETRn	GHI	DNI	DH	Lo	GH	DN	DH	Te	Re	Cle	Sp	std	std	std	std	std	std	std	std	Notes	302.06

Figure 3: Sample data set of header rows. The sample shows columns 12 - 33 and rows 1-9. Columns 15 - 31 have been condensed to allow for easier viewing.

- **Row 1. Type of measurement:** The type of measurement that is made in this column. The labels are self-explanatory. Please refer to this document for a description of the various columns.

- **Row 2. Instrument:** For broadband and general metrological data, the instrument making the measurement is listed and the instrument serial number is given inside the parentheses. Columns that are calculated are specified as such. For the spectral data, row 2 is the wavelength of light being measured given in nanometers. The serial number of the PSR spectroradiometer was PSR(006).

- **Row 3. Responsivity (Calibration Factor):** The responsivity (or calibration factor) that was used to convert the measured voltage signal to a broadband irradiance values. The formula relating voltage to broadband irradiance is given by Equation 1.

$$\text{Broadband Irradiance} = \frac{\text{Voltage}}{\text{Responsivity}} \qquad \text{(Eq. 1)}$$

The voltage is measured by the instrument and internally changed to irradiance by dividing by the responsivity. The voltage of each measurement is not recorded, only the corresponding irradiance and responsivity are recorded. The responsivity of the spectral instrument was not supplied by the station operator.

For the broadband instruments, the responsivities are computed at an angle of incidence of 45°. For the spectral measurements the calibration factors are determined at an angle of incidence of 0°.

- **Row 4. Estimated uncertainty:** The estimated uncertainty is the percent uncertainty in the measured value. The uncertainty is reported for the 95% level of confidence. Specific details about the uncertainty of the spectroradiometer will be given during the discussion of the instrument.

- **Row 5. Units of each measurement:** Standard units are used for each measurement. Typical units for irradiance are W/m^2. Note the carrot symbol ^ is used to describe a number raised to a power. Typical units for spectral irradiance are W/m^2/nm.

- **Rows 6 – 8.** These three rows allow notes about each column. These columns are not as strictly defined and are a place for the user/editor to make notes about the various columns as they see fit.

- **Row 9.** To avoid confusion, the column labels are repeated in row 9. The date/time information labels are only included in row nine to allow room for the station ID information.

File structure data section 3.
Date/Time, SZA/AZM, ETR/ETRn

The data presented from the Lindenberg monitoring station has a time interval of one minute. The data of the file is separated into three data sections (Section 3, 4, and 5 in Figure 1). The left most section contains date and time information, solar position information, and extraterrestrial irradiance information.

A sample data set of data section 3 (Columns 1 -12) is shown in Figure 4. The sample shown highlights the time stamps near noon on July 3, 2014. Note that the header rows 8 and 9 are included to give the column labels.

	1	2	3	4	5	6	7	8	9	10	11	12
8	LST	LST	LST	LST	LST	UTC	LST	hours	degrees	degrees	W/m^2	W/m^2
9	Year.FOY	DOY.FOD	YYYY-MM-DD--hh:mm	YYYY-MM-DD	DOY	FOD	Hour.FOH	SolarTime	SZA	AZM	ETR	ETRn
3606	2014.5027340183	184.4979167	2014-07-03--11:57	7/3/2014	184	0.456250	11.9500	11.81261	29.34	174.71	1146.62	1315.34
3607	2014.5027359208	184.4986111	2014-07-03--11:58	7/3/2014	184	0.456944	11.9667	11.82927	29.32	175.18	1146.84	1315.34
3608	2014.5027378235	184.4993056	2014-07-03--11:59	7/3/2014	184	0.457639	11.9833	11.84594	29.31	175.65	1146.96	1315.34
3609	2014.5027397260	184.5	2014-07-03--12:00	7/3/2014	184	0.458333	12.0000	11.8626	29.3	176.12	1147.07	1315.34
3610	2014.5027416286	184.5006944	2014-07-03--12:01	7/3/2014	184	0.459028	12.0167	11.87927	29.29	176.59	1147.18	1315.34
3611	2014.5027435312	184.5013889	2014-07-03--12:02	7/3/2014	184	0.459722	12.0333	11.89593	29.28	177.06	1147.29	1315.34

Figure 4: Sample data set of Section 3. The sample shown highlights the various date and time stamps. Header rows 8 and 9 are included for column labels.

The date and time of each row are written in eight different date/time formats. Columns 1, 2, 3, 4, 5, 7 are in local standard time as denoted in row 8. Column 6 is in UTC time as requested by the user. All time intervals represent the date and time at the end of the time interval. The files are separated into month blocks according to the local standard time (not universal time).

- **Column 1. Year.FOY:** The first column is the year with a decimal point representing the fraction of a year (FOY) calculated using Equation 2.

$$\mathrm{Year.FOY} = \mathrm{year} + \frac{(\mathrm{dayofyear.fractionofday}-1)}{\mathrm{days\ in\ year}} \quad \text{(Eq.2)}$$

For example: 2014, July 3rd at 12 noon would be 2014.5027397260.

- **Column 2. DOY.FOD:** The second column is the day of the year (DOY) with the decimal point representing the fraction of a day (FOD) calculated using Equation 3.

$$\mathrm{DOY.FOD} = \mathrm{dayofyear} + \frac{(\mathrm{minuteofday}-1)}{1440} \quad \text{(Eq. 3)}$$

For example: 2014, July 3rd at 12 noon would be 184.5. The year is not included in this column.

- **Column 3. YYYY-MM-DD--hh:mm** The first column is the traditional view of dates and times, in order from largest to smallest, year-month-day--hour:minute (YYYY-MM-DD--hh:mm). Note the double dash marks "--", separate the date and the time. This is done to maintain the date and time format that are often altered when files are imported into spreadsheets. For example: 2014, July 3rd at 12 noon would be 2014-07-03--12:00.

- **Column 4. YYYY-MM-DD:** The fourth column is the date written in the format YYYY-MM-DD. Note that the hour, minute, second are not included in this format.

- **Column 5. DOY:** The second column is the integer value of the day of the year (DOY). The time of day is not included in this column. For example, 2014, July 3rd at 12 noon would be DOY = 184.

- **Column 6. FOD:** The sixth column is the fraction of the day (FOD) computed using Equation 4. Column 6 is given in UTC time as specified by the label in row 8.

$$\mathrm{FOD} = \frac{\mathrm{LST}-1}{24} \quad \text{(Eq. 4)}$$

Where LST is the local standard time written in hours (and fraction of hours), and the value 1 is determined from the time zone. The universal time is given values between 0 and 1. For example, 2014, July 3rd at 12 noon local standard time would be FOD = 0.458333. The date is not included in this column.

- **Column 7. Hour.FOH:** The seventh column is the hour and fraction of hour (FOH). The year, month, day, are not included in this value. The values in this column range from 0 - 23.9833 (1 minute before midnight). For example, 12 noon local standard time would be 12.0.

- **Column 8. Solar time:** The solar time (ST) is calculated using the SOLPOS algorithm [5] available from the NREL website. The details of the solar time calculation are beyond the scope of this document. When the sun is directly south in the northern hemisphere the solar time = 12. When the sun is east of south, the solar time is less than 12. When the sun is west of south the solar time is greater than 12.

- **Columns 9 – 10. SZA and AZM:** The solar zenith angle (SZA) and solar azimuthal angle (AZM) are calculated using the SOLPOS algorithm. An explanation of the zenith angle and azimuthal angle calculation are beyond the scope of this document. The SZA and the AZM are reported in degrees. The solar zenith angle is computed using refraction through the atmosphere. The calculation is done for the middle of time interval. Unlike the SOLPOS code the SZA is also given when the sun is below the horizon.

- **Columns 8 – 9. ETR and ETRn:** The extraterrestrial irradiance (ETR) on a horizontal surface and extraterrestrial normal irradiance (ETRn) are calculated using the SOLPOS algorithm. The units of ETR and ETRn are in W/m^2. The ETRn is calculated using Equation 5.

$$\text{ETRn} = 1360.8 * (1.000110 + 0.034221 * \text{Cos}[\text{DA}] + 0.001280 * \text{Sin}[\text{DA}] + 0.000719 * \text{Cos}[2\ \text{DA}] + 0.000077 * \text{Sin}[2\ \text{DA}]) \qquad \text{(Eq. 5)}$$

where DA is the day angle in degrees given by Equation 6. In Equation 6, the day of the year is the value given in Column 2. The denominator of equation 6 is the days in the year 365 (366 for a leap year)

$$\text{DA} = (\text{day of year} - 1) * \frac{360}{365\ (366)} \qquad \text{(Eq. 6)}$$

The ETR is computed from the ETRn using Equation 7.

$$\text{ETR} = \text{ETRn} * \text{Cos}(\text{SZA}) \qquad \text{(Eq. 7)}$$

In Equations 5 and 7, the solar constant is defined as 1360.8 W/m^2 instead of the previous value of 1367 W/m^2. The ETR and ETRn are set to zero when the entire disk of the sun is below the horizon (SZA > 90.267°). The angular radius of the sun is 0.267°. During the time intervals of sunrise and sunset, when the sun crosses the SZA = 90.267° boundary, the ETR and ETRn are decreased by a scale factor dependent on the fraction of time the sun is visible.

File structure data section 4.
Metrological data

Along with the spectral irradiance, each data file has an extensive set of supplemental metrological data that was simultaneously gathered at the site. A sample data set of section 4 is shown in Figure 5. The sample shown is from July 3, 2014. The 9 header rows are shown as well as several sample data points from around noon. The sample shown in Figure 5 only includes Columns 13 through 19. The time stamp of Column 3 is included for reference, as well as the row labels given in Column 12.

The values of the broadband and metrological measurements are reported at the end of the minute. For example, the measurement listed at 12:00, would be an average of measurements taken from 11:59:01 - 12:00:00.

	3	12	13	14	15	16	17	18	19
1	-	Type_of_measurment	GHI	DNI	DHI	Longwave	GHI_Visible	DNI_Visible	DHI_Visible
2	-	Instrument	CMP22(020074)	CHP1(030340)	CMP22(020073)	PIR(32802)	Calculated	Calculated	Calculated
3	-	Responsivity_(microV/W/m^2)	9.33_microV/(W/m^2)	9.83_microV/(W/m^2)	9.67_microV/(W/m^2)	4.00_microV/(W/m	-	-	-
4	-	Responsivity_Uncertainty(U95%)	3	1.5	3	3	-	-	-
5	-	Units	W/m^2	W/m^2	W/m^2	W/m^2	W/m^2	W/m^2	W/m^2
6	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-
8	LST	W/m^2	-	-	-	4000-50000_nm	-	360.48-829.45_nm	-
9	YYYY-MM-DD--hh:mm	ETRn	GHI	DNI	DHI	Longwave	GHI_Visible	DNI_Visible	DHI_Visible
3606	2014-07-03--11:57	1315.34	1058	849	309	354	NA	478.5404	NA
3607	2014-07-03--11:58	1315.34	1072	858	315	354	NA	474.266	NA
3608	2014-07-03--11:59	1315.34	1078	861	319	353	NA	NA	NA
3609	2014-07-03--12:00	1315.34	1005	778	321	353	NA	341.0884	NA
3610	2014-07-03--12:01	1315.34	877	641	303	349	NA	444.6969	NA
3611	2014-07-03--12:02	1315.34	1002	792	300	349	NA	440.1768	NA

Figure 5: Sample data set of Section 4. The sample shown highlights the various date and time stamps. The header rows 1 - 9 are included for column labels. The column and row numbers are visible for reference.

- **Column 13. GHI:** The broadband global horizontal irradiance (GHI) was measured using a CMP22 pyranometer manufactured by Kipp and Zonen. The responsivity and uncertainty are reported in the data files in rows 3 and 4. The uncertainty value reported in row 4 was obtained from the manufacturers specifications of the instrument. The units of the GHI are in W/m^2.

- **Column 14. DNI:** The direct normal irradiance (DNI) was measured using a CHP1 pyrheliometer manufactured by Kipp and Zonen. The responsivity and uncertainty are reported in the data files in rows 3 and 4. The uncertainty value reported in row 4 was obtained from the manufacturers specifications of the instrument. The units of the DNI are in W/m^2.

- **Column 15. DHI:** The diffuse horizontal irradiance (DHI) was measured using a CMP22 pyranometer with a shade ball blocking the sun. The responsivity and uncertainty are reported in the data files in rows 3 and 4. The responsivity and uncertainty are reported in the data files in rows 3 and 4. The uncertainty value reported in row 4 was obtained from the manufacturers specifications of the instrument. The units of the DHI are in W/m^2.

- **Column 16. Longwave:** The integrated longwave diffuse horizontal irradiance was measured using a Precision Infrared Radiometer (PIR) manufactured by Eppley. The responsivity and uncertainty are reported in the data files in rows 3 and 4. The units of the data are in W/m^2. The wavelength range of the PIR is 4,000 – 50,000 nm as specified in row 8. The longwave irradiance is typically used to determine the sky temperature. The PIR measurement takes into account the irradiance entering the detector from the sky as well as the light leaving the detector (IE. the detector is giving off infrared light).

- **Column 17. GHI_Visible:** There is not a spectral GHI detector located on site. This column is included to standardize the formatting from other stations.

- **Column 18. DNI_Visible:** The irradiance in the visible wavelength range was calculated using Equation 8. The visible wavelength range is λ = 360 – 830 nm. DNI_Spectral_Irr(λ) is the spectral irradiance reported in the file at a particular wavelength. The exact wavelengths used in the summation are listed in Row 8 and may change slightly in time as the instrument is recalibrated. The summed spectral irradiance must be multiplied by the average width of a spectral bin in the wavelength range 360 <λ 830. For the 2014 file used as an example this value is 0.7051 nm. The value changes slightly due to recalibration of the instrument.

$$\text{DNI\_Visible} = \left(\sum_{\lambda=360}^{830} DNI\_Spectral\_Irr(\lambda)\right) * 0.7051 \qquad \text{(Eq 8.)}$$

- **Column 19. DHI_Visible:** There is not a spectral DHI detector located on site. This column is included to standardize the formatting from other stations.

- **Column 20. Temperature:** The air temperature is given in degrees Celsius.

- **Column 21. Relative Humidity:** The relative humidity is given in percent.

- **Column 22. Clearness Index:** As a measure of the clearness of the sky, a sky clearness index was calculated using Equation 9. The clearness index is a unitless quantity. A clearness index of 0.5 means the half the extraterrestrial GHI light is reaching the surface of the earth. A clearness index of 1 means that all the extraterrestrial GHI light is reaching the surface of the earth. The clearness index can have values from zero to larger than one. Values larger than one are possible when there is significant DHI or at dawn or dusk when the GHI values are near zero and the uncertainty of the measurement is significant.

$$\text{Clearness\_index} = \frac{GHI}{ETR} \qquad \text{(Eq 9.)}$$

- **Column 23. Spectral_Time_Mismatch:** The PSR instrument gathered spectral data at irregular (non-minute) time intervals. This causes the PSR data set to not align with the one-minute output files. Moreover, the time interval of the PSR instrument is not consistent and varies over the day and also changes significantly from year to year. Table 1 gives the approximate date range and time intervals of the spectral data set. Also included in Table 1 is a list of what parts of the day the instrument was operated. In the beginning, the PSR was only operated during daylight hours. For latter time periods, the instrument was operated 24 hours. Periodically there are gaps in the data due to calibration, and maintenance.

Table 1. Information about the data collection periods. Note the time interval is an approximate value and will change slightly from day to day. Consult the actual data files for complete information about each file. Gaps and breaks in the data exist and only significant features are listed here.

Start date (YYYY-MM-DD)	End date (YYYY-MM-DD)	Approximate time interval (minutes)	Daylight hours or 24 hours
2014-06-04	2015-04-29	1.3	Daylight hours
2015-04-29	2015-09-14	1.7	24 hours
2015-09-15	2015-11-26	No data (Maintenance)	No data (Maintenance)
2015-11-27	2015-12-10	3.2	Daylight hours
2015-12-11	201-12-17	1.3	24 hours
2015-12-17	2016-01-26	2.5	24 hours
2016-01-26	2016-02-29	4.2	24 hours
2016-02-29	2016-04-05	1.25	24 hours
2016-04-06	2016-11-30	No data (Maintenance)	No data (Maintenance)
2016-12-01	2016-12-09	5.85	24 hours
2016-12-09	2017-02-15	12.3	24 hours
2017-02-16	2017-08-03	8.6	24 hours
2017-08-03	2017-08-28	12.3	24 hours
2017-08-29	2018-01-11	8.6	24 hours
2018-01-11	2018-01-25	12.3	24 hours
2018-01-25	2018-02-08	8.6	24 hours
2018-02-08	2018-04-20	12.3	24 hours
2018-04-20	2018-05-22	8.6	24 hours
2018-05-22	2018-07-03	12.3	24 hours
2018-07-04	2018-07-31	No data (Maintenance)	No data (Maintenance)
2018-08-01	2018-09-19	8.6	24 hours

The PSR instrument makes several measurements in a short period of time (approximately 10 seconds) and can be thought of as an instantaneous measurement. The spectral data was aligned with the appropriate minute of the output file with the middle of the 10 second data collection period used to place the spectral data into the appropriate minute bin.

In merging the spectral data with the one-minute bin a time mismatch will exist between the spectral data timestamp and one-minute time interval of the output file. As a reminder the time stamp of the output file is for the end of the time interval. The difference in time between the spectral data and the one-minute time stamp in the output file are a time mismatch reported in Column 23. The units of the mismatch are in seconds. The time mismatch can have values that range from 0 to 59. The mismatch is calculated using Equation 10.

Spectral time mismatch = UTC – Spectral time(UTC) (Eq 10.)

Where UTC is the Universal Time Coordinated reported in the output file and the spectral time is the middle of the 10 second time interval the spectral data was actually gathered (in UTC time).

- **Column 24 - 31. Standard deviation of spectral data:** The PSR instrument takes 10 samples during each measurement period. Each sample includes a measurement for each of the 1024 wavelengths. The average of these samples is reported as the spectral irradiance. The standard deviation of these samples at each wavelength is also calculated. The standard deviation is a measure of how much the spectral irradiance was changing during that time interval and does not include uncertainties associated with the calibration of the instrument.

A complete list of standard deviations of each measurement is given in the separate standard deviation file. (See Appendix C). An abridged version of the standard deviation is given in Columns 24 – 31. With the various columns corresponding to the following wavelengths: 305, 400, 500, 600, 700, 800, 900, 1020 nm. Spectral measurements begin at approximately 302 nm and end at 1022 nm. The abridged version of the standard deviation information is intended to give the user a brief summary of the standard deviation curve without the need to reference a secondary file.

For each wavelength bin, all wavelengths that are within ±2.5 nm of the bin were selected. This corresponds to 6 or 7 wavelength samples depending on the wavelength. From these standard deviation samples, the median is calculated using Equation 10. Several wavelengths were selected to avoid outlying standard deviation values at a particular wavelength. This calculation was done for each time interval that spectral data was collected.

$$\text{stdev\_305nm} = \text{Median}(StDev\ 302.5 - StDev\ 307.5) \quad \text{(Eq 10A.)}$$
$$\text{stdev\_400nm} = \text{Median}(StDev\ 397.5 - StDev\ 402.5) \quad \text{(Eq 10B.)}$$
$$\text{stdev\_500nm} = \text{Median}(StDev\ 497.5 - StDev\ 502.5) \quad \text{(Eq 10C.)}$$
$$\text{stdev\_600nm} = \text{Median}(StDev\ 597.5 - StDev\ 602.5) \quad \text{(Eq 10D.)}$$
$$\text{stdev\_700nm} = \text{Median}(StDev\ 697.5 - StDev\ 702.5) \quad \text{(Eq 10E.)}$$
$$\text{stdev\_800nm} = \text{Median}(StDev\ 797.5 - StDev\ 802.5) \quad \text{(Eq 10F.)}$$
$$\text{stdev\_900nm} = \text{Median}(StDev\ 897.5 - StDev\ 902.5) \quad \text{(Eq 10G.)}$$
$$\text{stdev\_1020nm} = \text{Median}(StDev\ 1017.5 - StDev\ 1022.5) \quad \text{(Eq 10H.)}$$

Where the phrase "*StDev 302.5 - StDev 307.5*" means the standard deviation of wavelengths from 302.5nm to 307.5 nm are selected. Specifically, this includes wavelengths: 302.76, 303.47, 304.17, 304.87, 305.57, 306.27, 306.97 nm (for a 2014 data set). Similarly, for the other wavelength ranges.

A comparison of the standard deviation at the selected wavelengths vs the standard deviation at all wavelengths is shown in Figure 6 and 7. Figure 6 is for clear sky conditions. Figure 7 is a sunny period on a day with partly cloudy sky conditions (IE the sky conditions are changing throughout the day). The standard deviation of all wavelengths is shown in the red curve. The standard deviation at the selected wavelengths is shown with the blue curve. The sample data points are connected using straight lines to guide the eye. The spectral irradiance value is shown as the black curve and is plotted on the right axis. In Figure 6 the standard deviation rises at large wavelengths. The magnitude of the standard deviation in Figure 6 is significantly smaller than in Figure 7. This is an indication that the sky conditions are changing, as one would expect during a day with partly cloudy skies. In Figure 7, the standard deviation of the spectral irradiance has the same shape as the spectral irradiance (The black and red curves align).

In both figures, the abridged version of the standard deviation does an adequate job of describing overall shape of the complete standard deviation curve, although wavelength specific features are neglected. This is clearly evident in Figure 7 at the wavelength range of around 950 nm.

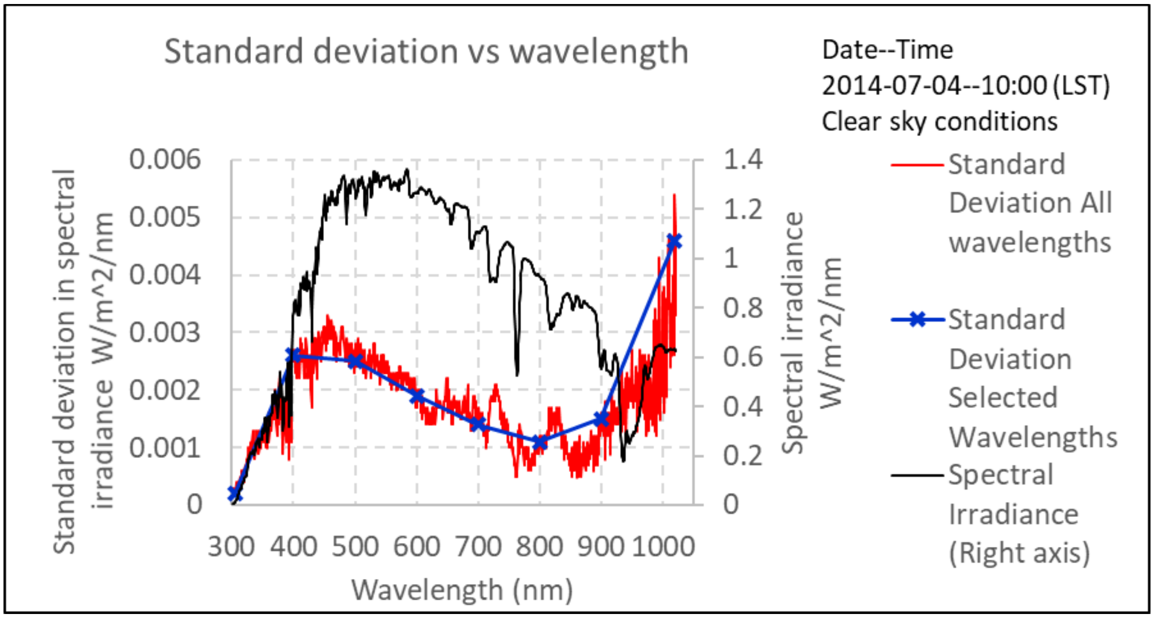

Figure 6. Standard deviation vs wavelength for clear sky conditions. The magnitude of the standard deviation is significantly less than partly cloudy skies. The standard deviation increases at large wavelengths.

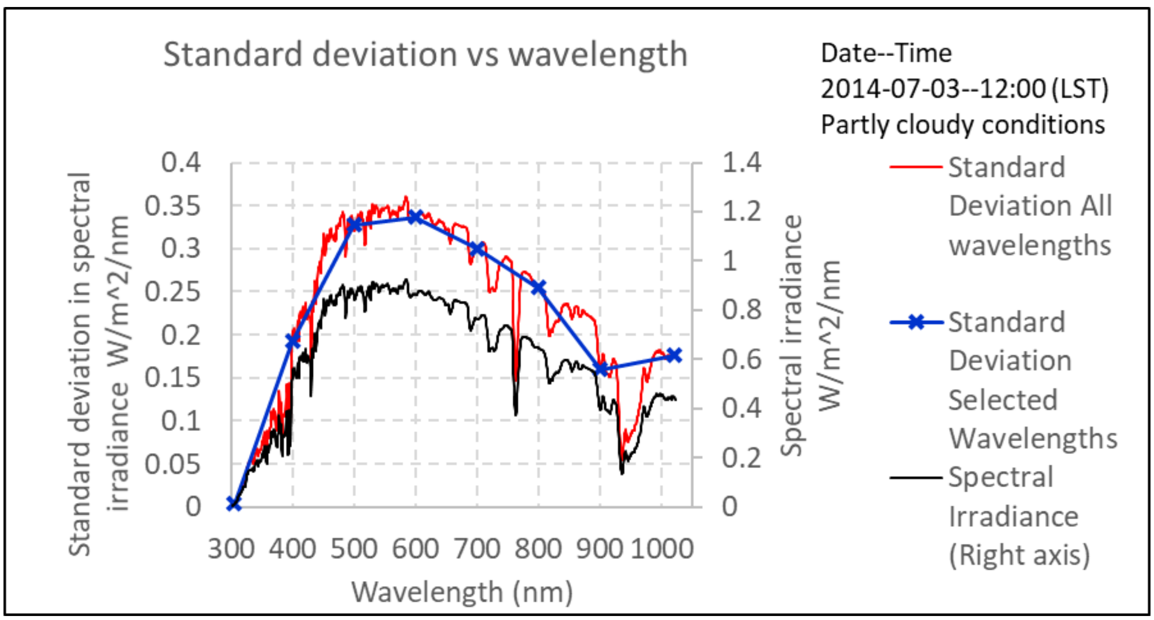

Figure 7. Standard deviation vs wavelength for partly cloudy sky conditions. The data was collected during a sunny period under partly cloudy skies. The standard deviation in all wavelengths is shown in red. The standard deviation at selected wavelengths is shown in blue. The spectral irradiance is shown in black and is plotted on a separate axis on the right. Note that the standard deviation in the spectral irradiance has the same shape as the spectral irradiance.

- **Column 32. Notes:** A column is allotted to allow the user to make notes in the data set. The notes column separates the measured standard deviation in the spectral data and the spectral data contained in Section 5.

File structure data section 5.

Spectral data

Direct normal spectral irradiance data was gathered from June 2014 through September 2018 using a PSR spectroradiometer manufactured by PMOD. Data was gathered at a variety of time intervals, none of which were 1 minute. The spectral data was aligned with the one-minute output file as best as possible. See the time mismatch section for more details. Data points that are missing are given a value of "NA". The units of the spectral irradiance are in $W/m^2/nm$ as listed in Row 5.

The PSR instrument takes 10 samples in a short period of time (10 seconds). Thus the spectral irradiance data can be thought of as an instantaneous measurement. The average of all samples is the irradiance reported in the output file. The standard deviation of all samples is the standard deviation of the data (see Appendix C for a description of the standard deviation output file).

The spectral irradiance data set is contained in columns 33 through 1056 in the data files. The wavelength range of the spectral data is approximately 300 nm through 1020 nm. These wavelengths vary slightly in time as the instruments calibration values change. There are 1024 columns of spectral data. Each spectral data bin is roughly 0.7 nm apart. The wavelength of each column is listed in row 2 and also in row 9 of the data files.

A sample of the spectral data set is shown in Figure 8. The 9 header rows are included as well as several data points from around noon on July 3, 2014. There is a large portion of the middle wavelengths (304 – 1020 nm) that are not shown for brevity. At the time this data set was collected, the spectroradiometer was gathering data at roughly a 75 second time interval. This is why the 11:59 and 12:04 data points are labeled "NA". The instrument did not collect data during these minutes. The spectral time mismatch is also included to illustrate how the instrument progresses in time. Notice how the time mismatch starts at large values (close to 59 seconds), decreases to small values (close to 0 seconds), and then resets, as one would expect with a 75 second time interval.

	3	23	32	33	34	35	1055	1056
1	-	Spectral_Time_Mismatch	-	DNI_Spectral	DNI_Spectral	DNI_Spectral	DNI_Spectral	DNI_Spectral
2	-	UTC-Spectral_Time(UTC)	Wavelength(nm)	302.06	302.76	303.47	1021.06	1021.76
3	-	-	Calibration_Factor((W/m^2/nm)/counts)	-	-	-	-	-
4	-	-	Uncertainty(U95%)	6.6	6.6	6.6	3.2	3.2
5	-	Seconds	Units	W/m^2/nm	W/m^2/nm	W/m^2/nm	W/m^2/nm	W/m^2/nm
6	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-
8	LST	seconds_before_end_of_minute	-	-	-	-	-	-
9	YYYY-MM-DD--hh:mm	Spectral_Time_Mismatch	Notes	302.06	302.76	303.47	1021.06	1021.76
3606	2014-07-03--11:57	31	NA	0.0065	0.0082	0.01	0.6177	0.6142
3607	2014-07-03--11:58	15	NA	0.0062	0.0083	0.01	0.6151	0.603
3608	2014-07-03--11:59	NA	NA	NA	NA	NA	NA	NA
3609	2014-07-03--12:00	58	NA	0.0041	0.0059	0.0072	0.4439	0.4379
3610	2014-07-03--12:01	41	NA	0.006	0.0077	0.0092	0.5825	0.5742
3611	2014-07-03--12:02	24	NA	0.0056	0.0076	0.0091	0.5615	0.5556
3612	2014-07-03--12:03	7	NA	0.0062	0.0071	0.0089	0.5277	0.5255
3613	2014-07-03--12:04	NA	NA	NA	NA	NA	NA	NA

Figure 8. Sample of the spectral data set. Wavelengths from 304 - 1020 are not shown for brevity.

The accuracy of the wavelength measurement is ± 0.1 nm and the optical resolution (full width at half max) of each bin is 1.5 – 6 nm depending on the wavelength. For more information regarding the specifications of the PSR see the manufacturer's specifications sheet [2, 3, 4].

The spectroradiometer was calibrated by the manufacturer before installation in 2014, during the fall of 2015, and the summer/fall of 2016. The operators of the instrument did not supply the calibration history of the instrument. The uncertainty of the spectroradiometer is given in row 4 and is dependent on the wavelength. At wavelengths greater than 350 nm the uncertainty is 3.2%. At wavelengths less than 350 nm the uncertainty in calibration increases to up to 6%.

According to the GUM model [6], the uncertainty associated with the calibration is a Type-B uncertainty. The standard deviation of each data point is a Type-A uncertainty. The standard deviation of each data point is also calculated. See the preceding section for a description of the standard deviation or Appendix C for more details.

Combining these two uncertainties through the sum of squares method produces an overall uncertainty in each measurement (at each wavelength). To compute the uncertainty in a single measurement the standard deviation should be combined with the uncertainty associated with the calibration according to Equation 11. In Equation 11B the calibration uncertainty must first be converted to irradiance units and also converted to standard (not expanded) form.

$$Combined\ expanded\ uncertainty\ = (TypeA^{\ 2}\ +\ TypeB^{\ 2})^{1/2} \quad \text{(Eq. 11A)}$$

$$U95\ = \left(StDev^2 + \left(\frac{Cal\_u95\%}{100*1.96} * spec\_irr\right)^2\right)^{1/2} * 1.96 \quad \text{(Eq. 11B)}$$

$$U95\%\ = \left(\frac{\mathrm{U95}}{\mathrm{spec\_irr}}\right) * 100 \quad \text{(Eq. 11C)}$$

Where *StDev* corresponds to the standard deviation associated with each data point. *Cal_u95%* corresponds to the percentage uncertainty in the calibration (Row 4 in the output file). The *spec_irr* corresponds to the spectral irradiance value during the minute in question. The value *1.96* is needed to convert between expanded uncertainty and standard uncertainty. The value *U95* corresponds to the combined expanded uncertainty. The units of U95 are in $W/m^2/nm$. To compute the expanded percent uncertainty Equation 11C should be used. Each term in Equation 11 is wavelength dependent. The wavelength subscript λ has been omitted for clarity. The combined uncertainty values are not included in the data output files due to large file sizes and different users may have different uses of the data.

Using Equation 11C, the combined expanded uncertainty vs wavelength is shown in Figures 9 and 10. Figure 9 shows the uncertainty for a clear sky period. Figure 10 shows the uncertainty for a sunny period on a partly cloudy day (approximately 50% cloud cover). In Figures 9 and 10, the standard deviation, calibration uncertainty, and combined expanded uncertainty are all plotted separately. All values are given in percentage expanded form. (i.e. the standard deviation has been converted to a percentage and multiplied by a value of 1.96)

During clear sky periods (Figure 9), the standard deviation in irradiance values becomes negligible, i.e. the sky is not changing duing clear sky periods so the standard deviaion in the irradiance is small. During partly cloudy sky conditions (Figure 10), the standard deviation in irradiance can change significantly and is the pedominate uncetainty. During partly cloudy skies conditions the calibration uncerrtinty is negligleble. As seen in Figure 10 during partly cloudy conditions, the uncertainty in the calibration of the instrument is significantly less than the standard deviation in the data.

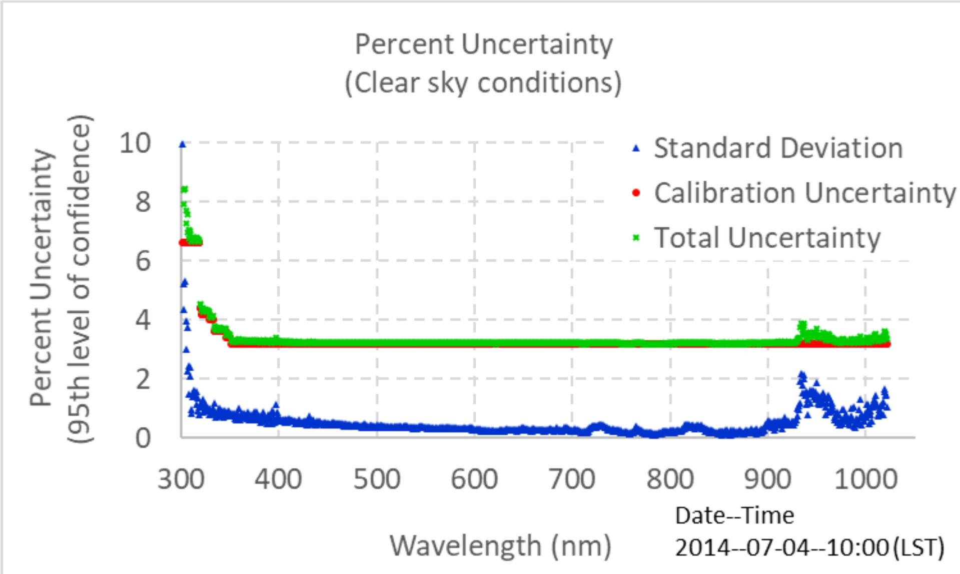

Figure 9. Percent uncertainty vs wavelength for clear sky conditions. The calibration uncertainty is the predominate uncertainty.

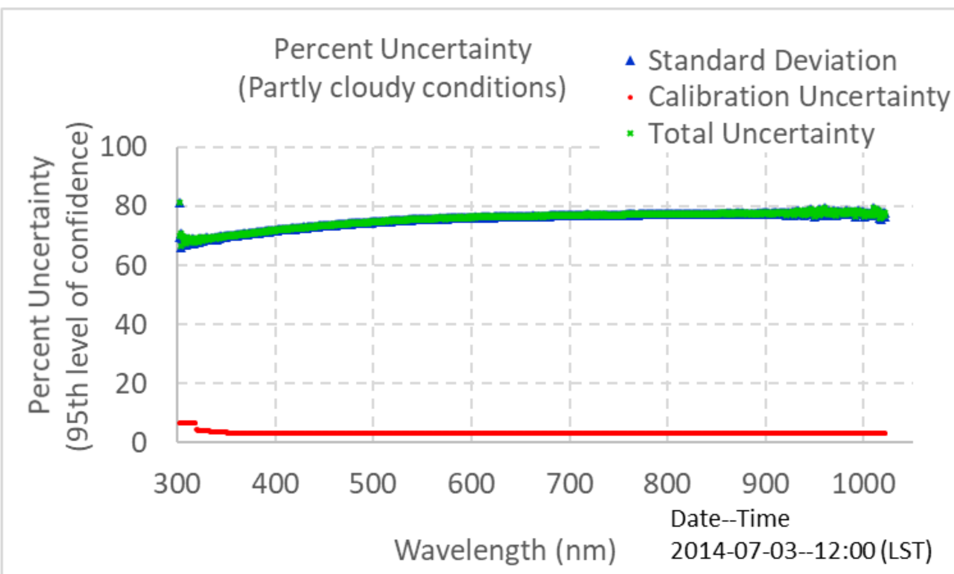

Figure 10. Percent uncertainty vs wavelength for partly cloudy sky conditions. The time period plotted was during a break in the clouds, when dirct sunlight was present.

Samples of various spectrum and Conclusions

In an effort to give a full understanding of the spectral nature of sunlight, several spectrum examples are shown below. These plots are intended to illustrate various features of the information contained in the spectral files. Note that this is a small subset of the entire data set and is only intended to give the reader a better understanding of the data.

As an illustration of the data set, values for a clear day are presented in Figures 11 – 15. A plot of the broadband DNI on 2014-07-04 is shown in Figure 11. The horizontal axis of the plot is of the solar time which is available in the data set. The morning had clear cloudless skies from dawn to midday. In the afternoon the sky conditions were partly cloudy.

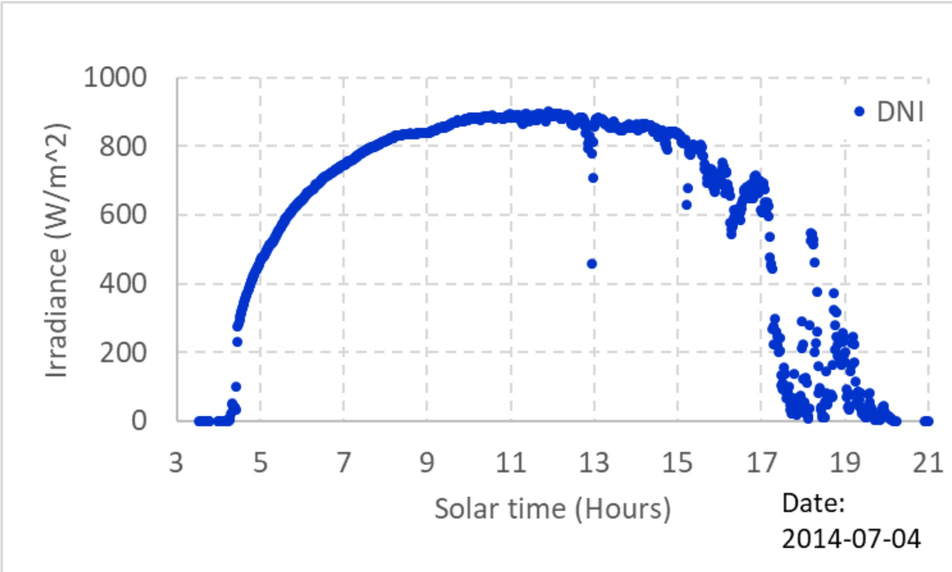

Figure 11. DNI vs time. This entire morning was clear. The afternoon had partly cloudy sky conditions.

Selecting 4 time intervals throughout the morning, the direct normal spectral irradiance is plotted vs wavelength in Figure 12. The plot legend gives the solar time as well as the solar zenith angle. When the sun is low in the sky there is spectral irradiance as one would expect. As the sun gets higher in the sky, the irradiance increases. The large dip that occurs from 900 -1000 nm is due to water absorption. During clear sky periods, the standard deviation the spectral irradiance is less than .02 $W/m^2/nm$ and is not visible in the plot.

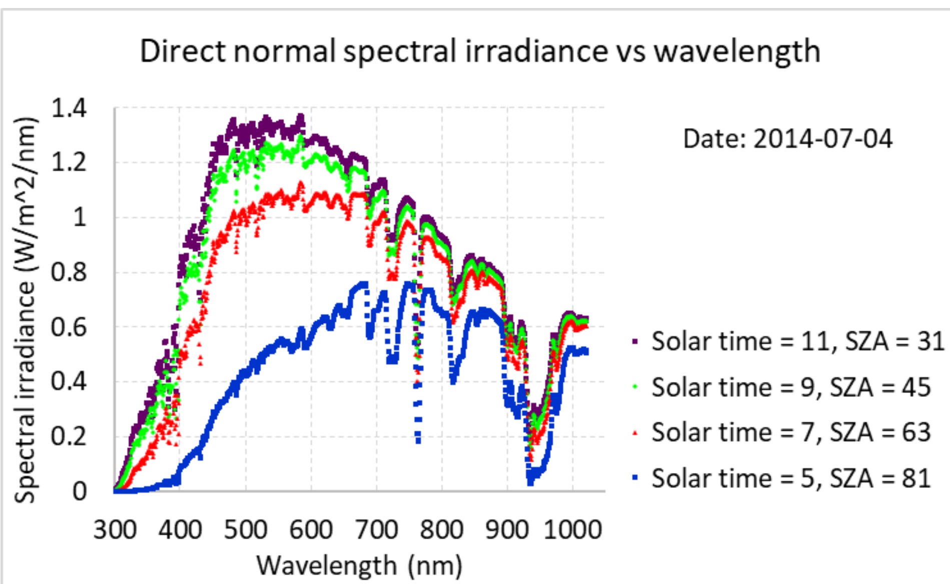

Figure 12. Spectral irradiance vs wavelength. Four different times are shown for comparison of how the spectrum changes over the day. The four different times have four different solar zenith angles. Both the time and the zenith angle are listed in the figure label. The shape of the plot is determined by the radiation profile of the sun (blackbody radiation) and the absorption spectrum of the atmosphere.

To compare the various spectrum shown in Figure 12, each spectrum is normalized to one at its maximum wavelength value. The normalized direct normal spectrum at these four time periods is shown in Figure 13. The most striking feature of Figure 13 is that after the sun reaches a zenith angle of 60 degrees the spectrum does not change significantly. This is noticeable in that the Solar times, 7, 9, and 11 are nearly indistinguishable. The maximum wavelength for all zenith angles less than 60 is between 500 and 600 nm. At large sun angles the direct normal spectrum is shifted to the larger wavelengths. (more red, less blues). This is caused because the atmosphere preferentially scatters the smaller wavelengths.

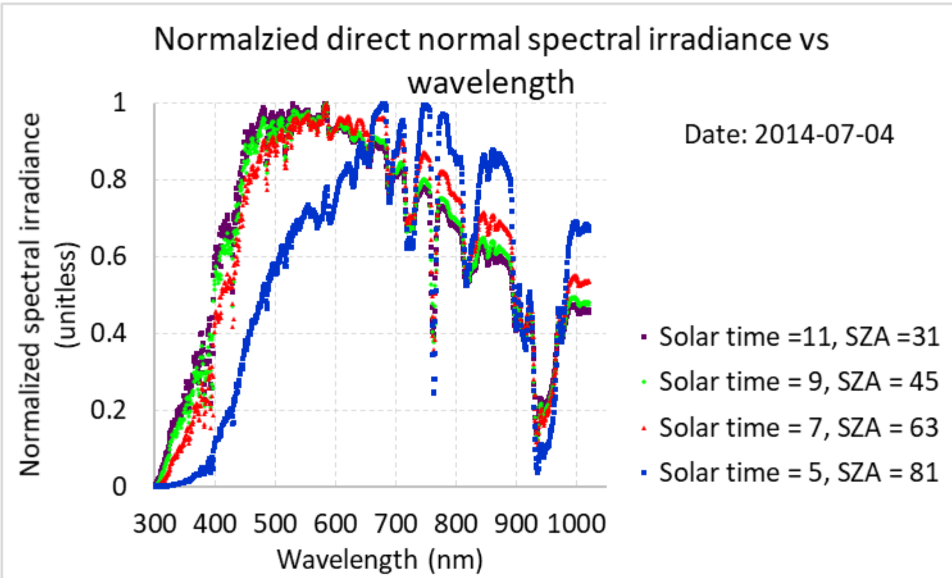

Figure 13. Normalized spectral irradiance vs wavelength. The spectrum shown in Figure 12, are divided by the maximum value during that time. For example, the spectrum of Time 11:00 was divided by a value of 1.35 $W/m^2/nm$, and the data of Time 5:00 was divided by a value of 0.75.

The examples shown in Figure 12 and 13 are intended to offer a snapshot of some of the findings in the dataset.

The German Meteorological Service (Deutscher Wetterdienst, DWD) site located in Lindenberg, Germany has been making solar spectral measurements since the middle of 2014 using an PSR Direct Normal spectroradiometer. The DWD provided the data used in this analysis. The spectral data set is archived and presented in a comprehensive file format. This article describes the format of the spectral files and the complementary information used to give context to the data. The format utilizes month blocks and data is reported in one-minute time intervals. The spectral data is a direct normal measurement from 300 nm to 1020 nm in 0.7 nm increments. A variety of time stamps are included in the data file to facilitate the use of the data. The files contain broadband metrological data as well as other meteorological information gathered at the site. The files also contain detailed header rows about the site location, instruments used, calibration values utilized, and uncertainties in the calibration values. Also included in the archived file are standard deviation values for each time interval. A second file containing a complete listing of the standard deviation for each spectral measurement is also provided for users that want a complete description of the standard deviation.

3. Acknowledgements

The UO Solar Radiation Monitoring Laboratory would like to thank the Lindenberg Meteorological Observatory and Dyson Technology Ltd. for their support that makes this work possible.

Appendix A

Glossary of commonly used terms

AZM: Solar azimuthal angle. Measured in degrees. North = 0, East = 90, South = 180, West = 270

Broadband: The total irradiance value of all wavelengths of light including UV, visible, and infrared.

Calibration Factor: The calibration factor that converts an electrical signal (or counts) into an irradiance value. The calibration factor for each instrument is known by performing a calibration of each instrument.

DHI: Diffuse horizontal irradiance

DNI: Direct normal irradiance

DOY: Day of year, January 1 = 1, February 1 = 32, March 1 (Non leap year) = 60 etc.

ETR: Extraterrestrial radiation on a horizontal surface.

ETRn: Extraterrestrial radiation on a normal surface.

GHI: Global horizontal irradiance

LST: Local standard time

NREL: National Renewable Energy Laboratory

PSR: Precision Spectroradiometer manufactured by PMOD

SOLPOS: Solar position calculator

SZA: Solar zenith angle. Measured in degrees. Straight up = 0, Horizontal = 90

Appendix B
Condensed column labels and locations

A condensed form of the information contained in the data files are listed in Table 2. The information contained in each column is listed along with the column numbers. For ease of use in spreadsheet programs, the alphabetical number of each column is also listed. The following table is intended to be a quick reference. For a complete description of each column see the main article. The wavelength values given in Columns 33 – 1056 vary slightly in time. The values given in the table are values for the year 2014.

Table 2: Condensed form of the data set. Column numbers are listed along with the information contained in each column. The wavelength values listed in Column 33 through 1056 are approximate values.

Column number	Alphabetical number	Column Label	Description
1	A	Year.FOY	Year . Fraction of day (LST)
2	B	DOY.FOD	Day of year. Fraction of day (LST)
3	C	YYYY-MM-DD--hh:mm	Date/time (LST)
4	D	YYYY-MM-DD	Year-Month-Day (LST)
5	E	DOY	Day of year (LST)
6	F	FOD	Fraction of day (UTC)
7	G	Hour.FOH	Hour . Fraction of Hour (LST)
8	H	Solar time	Solar time (hours)
9	I	SZA	Sun position
10	J	AZM	Sun position
11	K	ETR (W/m^2)	Calculated extraterrestrial radiation (horizontal surface)
12	L	ETRn (W/m^2)	Calculated extraterrestrial radiation (normal surface)
13	M	GHI	Metrological irradiance data
14	N	DNI	Metrological irradiance data
15	O	DHI	Metrological irradiance data
16	P	Longwave	Metrological irradiance data
17	Q	GHI_Visible	Does not exist
18	R	DNI_Visible	Calculated from DNI spectral
19	S	DHI_Visible	Does not exist
20	T	Air_Temperature	Atmospheric metrological data
21	U	Relative_Humidity	Atmospheric metrological data
22	V	Clearness index	Calculated from GHI and ETR
23	W	Spectral_Time_Mismatch	Spectral_Time_Mismatch
24 - 31	X – AE	Stdev 305, …, 1020	Standard deviation in Spectral data
32	AF	Notes	Blank column (Room for notes)
33	AG	302.06 nm	Spectral irradiance
34	AH	302.76 nm	Spectral irradiance
35 – 1055	AI – ANO	303.47 – 1021.06 nm	Spectral irradiance
1056	ANP	1021.76 nm	Spectral irradiance

Appendix C
Standard Deviation file

The PSR instrument takes ten measurement samples during each time interval. The average of the multiple measurements is taken. These average values are reported in the main irradiance output file that has already been discussed. The standard deviation of the multiple measurements is also taken. The standard deviation output file is discussed here in Appendix C.

The format of Appendix C is the same as the format of the main output file. Both files have the same dimensions. Columns 1 – 23 are identical to the main output file. Columns 24 - 31 are the abridged standard deviation columns are included only as place holders. There is no information in these columns. Column 32 is a notes column, just as it is in the main output file. Columns 33 – 1056 are the standard deviation at each wavelength. There is a standard deviation for each data point. The location in a file of a particular data point (time and wavelength) in the standard deviation file is the same as that in the main output file. To compute the total uncertainty of a particular wavelength and time interval, the user should follow the procedure outlined in Equation 11.

14. References

[1] https://www.dwd.de/EN/aboutus/locations/observatories/mol/mol.html

[2] Gröbner, J., Kouremeti, N., Nevas, S., Blattner, P., Characterisation Studies of Precision Solar Spectroradiometer, PMOD-WRC Annual Report 2014, p26.
http://www-old.pmodwrc.ch/annual_report/2014_PMODWRC_Annual_Report.pdf

[3] Gröbner, J., Kouremeti, N., Coulon, E., Durig, F., Gyo, M., Soder, R., Wasser, D., Spectroradiometer for Spectral Aerosol Optical Depth and Solar Irradiance Measurements, annual report PMOD/WRC, page 13, http://pmodwrc.ch/annual_report/2012_PMODWRC_Annual_Report.pdf, 2012.

[4]: https://projects.pmodwrc.ch/env03/images/documents_workshop/Groebner_et_al.pdf

[5] SOLPOS Website: http://rredc.nrel.gov/solar/codesandalgorithms/solpos/solpos.c (December 2018)

[6] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML. (1995). Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement. Geneva: ISO TAG 4.