SKYSPECTRA: AN OPENSOURCE DATA PACKAGE OF WORLDWIDE SPECTRAL DAYLIGHT

Balakrishnan, P.[1], Diakite-Kortlever, A.[1,9], Dumortier, D.[2], Hernández-Andrés, J.[3], Kenny, P.[4], Maskarenj, M.[5], Pierson, C.[6], Thorseth, A. [7], Xue, P.[8], Knoop, M.[1]
[1] Technische Universität Berlin, Chair of Lighting Technology, Berlin, GERMANY, [2] Université de Lyon, LTDS-ENTPE, Vaulx-en-Velin, FRANCE, [3] University of Granada, Departamento de Óptica, Facultad de Ciencias, Granada, SPAIN, [4] University College Dublin, UCD Architecture, Dublin, IRELAND, [5] Université catholique de Louvain, Louvain la Neuve, BELGIUM, [6] Oregon State University, College of Engineering, Corvallis, Oregon, USA, [7]Technical University of Denmark, Department of Electrical and Photonics Engineering, Roskilde, DENMARK, [8] Beijing University of Technology, Beijing Key Laboratory of Green Built Environment & Energy Efficient Technology, Beijing, CHINA, [9] Vrije Universiteit Brussel, Brussels, BELGIUM
priji.balakrishnan@tu-berlin.de

Abstract

This paper introduces SKYSPECTRA, an open-source data package comprising spectral daylight measurements collected from various sources worldwide. The dataset encompasses measurements from both long-term measurement sites and specific periods or experiments. The paper describes the data collection, selection, and organisation process within a standardised framework, ensuring consistency in naming conventions, formats, and units across all datasets. It also describes how the data is structured for operability in programming languages like R or Python for data analysis. SKYSPECTRA aims to facilitate further research and development of daylight applications that leverage spectral sky data by providing a comprehensive and standardised sample spectral daylight dataset that represents geographical locations, seasons, atmospheric conditions, and times of day.

Keywords: daylight, skylight, data package, measurement campaign, spectral characteristics

1 Introduction

In recent years, there has been a growing recognition of the significance of daylight's spectral characteristics in lighting design, application, and simulation. Research has shown that the spectral properties of daylight profoundly impact our non-image-forming responses and overall well-being (Knoop et al. 2020, Münch et al 2020). Consequently, there is a need to incorporate spectral daylight data into daylight prediction tools, such as spectral global horizontal irradiance and spectral radiance. This has led to various approaches for modelling or collecting such data, including standardised spectral distributions, sky models, physics-based radiative transfer models, and the use of RGB sensors or cameras. However, validating these models and data collection methods requires well-documented datasets representing different locations, seasons, atmospheric conditions, and times of day (Inanici 2019, Knoop et al. 2019, Diakite-Kortlever et al. 2022).

While interest in spectral measurements is increasing, the number of sites where such measurements are performed is limited and only a few make the datasets publicly available (Diakite-Kortlever et al. 2021). Examples of exemplary data collections available to the public are listed in Table 1. Apart from publicly available datasets, there are ongoing-long term measurement sites focussing on lighting application, like Berlin, Germany (Diakite-Kortlever et al. 2022), Vaulx-en-Velin, France (Tourasse 2016), and Beijing, China (Luo et al. 2015, Xue et al. 2022) where the data is not public. In addition, spectral measurements are also carried out at weather stations and research sites, e.g. for applications in the fields of atmospheric science and solar energy applications.

All these datasets are collected and archived in different formats. The inconsistency in the measuring and protocoling approaches hampers the comparability between and application of the datasets hence there is the need for a uniform standardised method to collate the datasets.

Some efforts were or are currently undertaken to systematically collect and protocol daylight measurements. The IDMP daylight measuring site initiative (CIE 1994) offered a common set of rules and guidelines to measure daylight availability and to describe the measurement protocol and instrumentation requirements. Knoop et al. (2019) proposed a template for documentation and measurement of the spectral characteristics of daylight, which was later modified and is currently used in the worldwide daylight measurement campaign within the activities of the CIE Technical Committee (TC) 3-60.

This paper is part of the CIE TC 3-60 effort and addresses these challenges by presenting an open-source data package, SKYSPECTRA, that collects and standardises spectral daylight measurement data from specialised sources worldwide.

Table 1 – List of spectral data collections publicly available

Location	Status	Website (last accessed July, 2023)	Publication for details
Granada (Spain)	completed	https://colorimaginglab.ugr.es/pages/data	Hernández-Andrés et al. 2001b
Granada (Spain)	completed	https://colorimaginglab.ugr.es/pages/data	Hernández-Andrés et al. 2001a
Eugene (USA)	ongoing	http://solardat.uoregon.edu/SelectSpectral.html	Vignola et al. 2021
Lindenberg, Tauche (Germany)	completed	http://solardat.uoregon.edu/SelectSpectral.html	Peterson at al. (no date)
Singapore	completed	https://github.com/prijibalakrishnan/equatorialskies	Balakrishnan 2019
Lausanne (Switzerland)	completed	https://zenodo.org/record/5919054	Pierson et al. 2023
Eindhoven (Netherlands)	completed	https://zenodo.org/record/5919054	Pierson et al. 2023
Cherry Springs State Park (USA)	completed	https://dx.doi.org/10.6084/m9.figshare.2009070.v1	Spitschan et al. 2016
Philadelphia (USA)	completed	https://dx.doi.org/10.6084/m9.figshare.2009070.v1	Spitschan et al. 2016
Tokyo (Japan)	completed	https://zenodo.org/record/5217752	Morimoto et al. 2022

2 Data Collection

The CIE TC 3-60, along with a worldwide campaign for the collection of spectral daylight measurements, was initiated at the 29th quadrennial session of the CIE to investigate geographical differences in spectral characteristics of daylight. The availability of structured and consistent datasets is crucial for the CIE TC 3-60 work. Hence, TC members that collect spectral data describe their respective measurement sites using a template introduced by Knoop et al. (2019) and subsequently refined through the work of TC 3-60. This template describes the spectral daylight measuring stations, details about the measuring instruments, specifics of the measurements and data collection procedures, and supplementary illustrative imagery. For easy reference, the complete template is provided in Section A.1 (Appendix A) of this paper.

2.1 Locations with spectral daylight measurements

The CIE TC 3-60 has currently identified a total of 16 locations for which spectral daylight measurements are available or will be made available for the committee’s work. Table 2 provides a summary of these locations. The current version of the SKYSPECTRA data package includes spectral data from 8 of these locations. While measurements from some of these locations are publicly available (Table 1), measurements from Berlin, Germany, Beijing, China and Vaulx-en-Velin, France are made specifically available for this dataset. Further expansion of the data package is planned to include the remaining locations, subject to the availability and feasibility of making the data publicly accessible.

Table 2 – Summary of locations for which spectral daylight data is available within CIE TC 3-60

Current release	Future release[1]
Berlin, Germany	Bialystok, Poland
Beijing, China`	Chilton, UK
Eindhoven, Netherlands	Chongqing, China
Granada, Spain	Dublin, Ireland
Lausanne, Switzerland	Eugene, USA
Roskilde, Denmark	Naples, Italy
Singapore, Singapore	São Paulo, Brazil
Vaulx en Velin, France	Seattle, USA

1 subject to availability and feasibility of making the data public

2.2 Selection of data for SKYSPECTRA

A subset of the data collected is selected for SKYSPECTRA to ensure manageability and consistency. Among the 8 recorded locations, Granada, Singapore, Roskilde, Lausanne, and Eindhoven represent discrete datasets recorded for specific periods. On the other hand, Berlin, Beijing, and Vaulx-en-Velin have long-term or ongoing measurements spanning years, with measurements taken at regular intervals of a few minutes. Hence spectral daylight measurements are selected for two dates in the months of March, June, September, and December, considering data availability and accessibility. Within these dates, timestamps recorded during clear and overcast skies are chosen or indicated when the information is available or can be calculated. For discrete datasets that have limited measurements like Singapore, all dates are included, indicating the sky conditions as clear or overcast when available.

For both the continuous and discrete datasets, measurements are chosen within the 360 nm to 800 nm wavelength range. Additionally, measurements are limited to timestamps with solar elevation angles equal to or above 15º to ensure consistency and relevance. Overall the selection ensures that SKYSPECTRA represents different locations, seasons, atmospheric conditions, and times of day.

3 Standardisation and organisation of selected data

One of the main challenges when working with spectral daylight measurement data collected from different sources is their inherent variability in format and structure. Each dataset employs different naming conventions and storage formats. Furthermore, variations were observed regarding the time representation, ranging from local time to UTC or GMT. The measurements differ, with some stations recording global spectral irradiance while others record patch radiances or global spectral irradiances on tilted planes. Moreover, certain datasets have missing values, gaps or are noisy for certain wavelengths.

Hence, to standardise and organise the collected data, three measures are taken:

1. The data is organised as two categories, the measurement and supplementary categories. The measurement datasets contain spectral daylight measurements, while the supplementary provides additional information about the location of the measurement sites, weather conditions, or descriptions of the measuring setup (Section 3.1).

2. Functions are developed in the R programming language to organise the data in a tidy format. The principles of tidy data, as described by Wickham (2014), are followed, where each variable is represented by a column, each observation forms a row, and each type of observational unit is represented as a separate table. Further details on the organisation of the data into a tidy format are provided in Section (3.1).

3. For quality control, any inconsistencies, missing values or noisiness in the data are clarified through direct communication with the data source. The reason for the gaps or measures taken to have consistent data is included as a brief note in the measurement parameter dataset (Section 3.1.2), enabling relevant interpretation of the dataset as required.

The aforementioned three measures ensure organisation of collected data within a standard framework and consistency in naming conventions, formats, and units across each column and its associated values. Hence facilitating efficient data manipulation, analysis, and visualisation when considering the dataset as a whole.

3.1 Structure of the datasets in SKYSPECTRA

Only a sample of the collected data, as described in Section 2.2, from various locations is included in the SKYSPECTRA data package. The datasets are primarily organised as measurement and supplementary data frames. The measurement dataset comprises four tables or data frames, each representing a distinct type of spectral daylight measurement: spectral horizontal irradiance (global and diffuse), spectral tilted irradiance (global and diffuse), spectral direct irradiance (of the sun), and spectral radiance (patch-wise). The supplementary dataset comprises of four data frames representing specific additional information for each measurement: location, weather, sun positions, and measurement parameters.

The column names of each data frame are variable names. As per the tidy data format, a variable contains values that measure the same underlying attribute (example: ***wavelength***, ***timestamp*** or ***location_code***). The main distinction between tidy data and non-tidy data format is that column names are not values of a variable. For example, a typical non-tidy data frame structure for storing spectral daylight measurements might use values of the variable ***wavelength*** as column names, such as, 380, 385, 390, and so forth.

Sections 3.1.1 and 3.1.2 present descriptions of the four measurement and supplementary data frames. Table 3 and 4 provide an overview of the variable names and value formats used in the measurement and supplementary data frames respectively. Lastly, Figures 1 and 2 provide a visual overview of the data frames, variable names (or column names) and examples of values used in the measurement and supplementary data frames respectively.

3.1.1 Measurement data frames

Following are the descriptions of the four measurement data frames and Table 3 lists all the variable names used in these data frames and their corresponding formats or descriptions.

- **spectral_horizontal_irradiance**: this data frame consists of spectral irradiance measurements of global and diffuse visible radiation recorded on a horizontal plane. The variable ***sun_included*** indicates whether the measurement corresponds to spectral global horizontal measurements (when set to TRUE) or spectral diffuse horizontal measurements (when set to FALSE). The latter refers to measurements taken with the sun blocked, for example, using a shadow ring.
- **spectral_tilt_irradiance**: this data frame contains spectral irradiance measurements of global and diffuse visible radiation recorded on a tilted plane. The variable ***sun_included*** distinguishes between spectral global tilted measurements (when set to TRUE) and spectral diffuse tilted measurements (when set to FALSE). Additionally, two specific variables in this data frame, ***detector_orientation*** and ***detector_inclination***, provide information about the tilt and orientation of the measurement device.
- **spectral_patch_radiance**: this data frame contains spectral radiance measurements of visible radiation from patches of sky. The variable ***sun_included*** indicates whether a measurement includes the sun (when set to TRUE) or excludes it (when set to FALSE).

Furthermore, two specific variables, ***patch_almucantar*** and ***patch_azimuth,*** provide information about the location of the patch.

- **spectral_direct_normal_irradiance**: this data frame consists of spectral irradiance measurements of visible radiation directly from the sun.

Table 3 – Variable names and description used in measurement data frames

Variable name	Description and Unit/Format
location_code	Two letter ISO country abbreviation-three letter user-defined abbreviation for each measuring site or experiment (string: e.g for Berlin, DE-BLN)
timestamp_local	Date, time and time zone (tz) of measurement taken in local time (string: YY-MM-DD HH:MM tz)
measurement_setup	Index of the measurement device (number: 01, 02, 03 …)
sun_included	Indicator whether measurements include or exclude the sun (binary: TRUE/FALSE)
detector_orientation	Deviation from the North in ° (number: clockwise consideration, e.g. East is 90)
detector_inclination	Inclination of detector from horizontal in ° (number: e.g. 30, 45, 60, 75, 90)
patch	Patch number (number: 1, 2, 3, 4, …)
patch_almucantar	Almucantar height for patch measurements in ° (number)
patch_azimuth	Azimuth increment for each patch from the North in ° (number)
wavelength	Wavelength in nm (number)
spectral_horizontal_irradiance	Spectral irradiance measured in $W \cdot m^{-2} \cdot nm^{-1}$ (number)
spectral_tilt_irradiance	Spectral irradiance measured in $W \cdot m^{-2} \cdot nm^{-1}$ (number)
spectral_patch_radiance	Spectral radiance measured in $W \cdot m^{-2} \cdot sr^{-1} \cdot nm^{-1}$ (number)
spectral_direct_irradiance	Spectral irradiance measured in $W \cdot m^{-2} \cdot nm^{-1}$ (number)
cct_horizontal_measured ***cct_horizontal_calculated***	Corelated colour temperature provided or calculated using Robertson (1968) in K (number)
cct_tilt_measured ***cct_tilt_calculated***	Corelated colour temperature provided or calculated using Robertson (1968) in K (number)
cct_patch_measured ***cct_patch_calculated***	Corelated colour temperature provided or calculated using Robertson (1968) in K (number)
cct_direct_measured ***cct_direct_calculated***	Corelated colour temperature provided or calculated using Robertson (1968) in K (number)

Table 4 – Additional variable names and their description used in supplementary data frames

Variable name	Description and Unit/Format
location_name	Name of the geographical location (string)
country	Country of the location (string)
timezone	Local time zone as defined in IANA time zone database (string: e,g Europe/Paris)
latitude	Geographical latitude in ° and with the hemisphere (string: xx.xx N/S)
longitude	Geographical latitude in ° and with the hemisphere (string: xx.xx E/W)
altitude	Altitude above mean sea level in m (number)
environment	Surrounding setting (string: urban/rural/natural)

industrial	Industrial nature of the surroundings (binary: TRUE/FALSE)
sky_type_cie	CIE sky type provided (number: 1,2,3,4,5,6,7,8,9,10 .. 16)
sky_condition	Sky condition provided (string: clear, overcast, other)
global_horizontal_illuminance	Global horizontal illuminance measured in lux (number)
global_horizontal_irradiance	Global horizontal illuminance measured in lux (number)
sun_elevation	Sun elevation angle in (number)
sun_azimuth	Sun azimuth angle in (number)
solar_time	True solar time calculated (string: HH:MM)
measurement_type	The type of spectral measurement conducted (string: spectral_horizontal_irradiance, spectral_tilt_irradiance, spectral_patch_radiance, spectral_direct_irradiance)
device_type	Type of device used for measurement (string: spectroradiometer, colorimeter)
wavelength_resolution	Wavelength resolution in nm (number)
filter_type	Type of filter used (string)
aperture_angle	Aperture angle of measurement device in ° (number)
notes	Notes on gaps, missing data, or corrections to data (string)

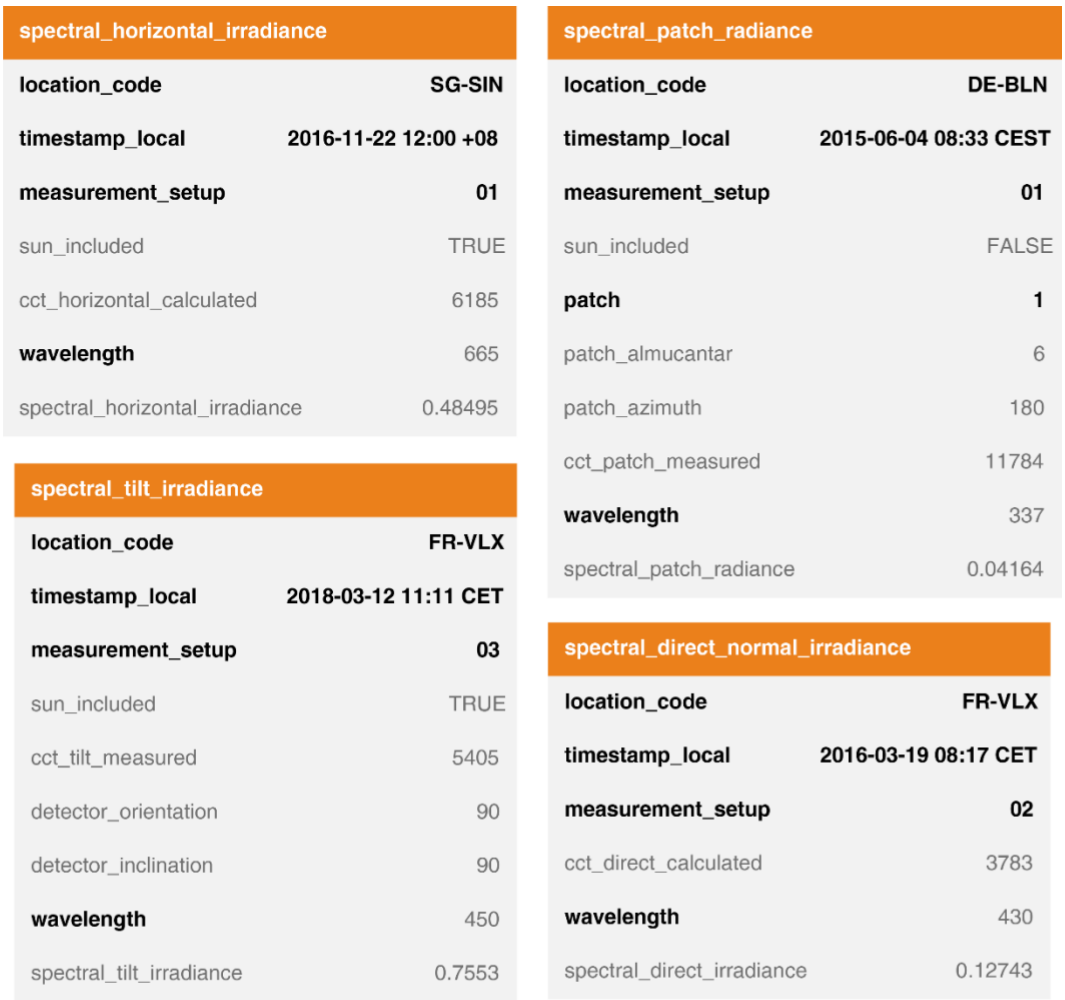


Figure 1 – Overview measurement data frames - column names and example values

3.1.2 Supplementary data frames

Following are the descriptions of the four measurement data frames and Table 4 lists the additional variable names used in these data frames, along with their corresponding formats or descriptions:

- **location**: this data frame contains information about the geographical location and immediate environment of where the measurements were taken.
- **weather**: this data frame provides details on sky conditions when the measurements were taken.
- **sun_positions**: this data frame includes information about sun angles and true solar time when the measurements were taken.
- **measurement_paramaters**: this data frame contains information about specific measurement parameters used for each measurement.

location	
location_code	**FR-VLX**
location_name	Vaulx-en-Velin
country	France
timezone	Europe/Paris
latitude	45.78 N
longitude	4.93 E
altitude	170
environment	urban
industrial	FALSE

measurement_parameters	
location_code	**FR-VLX**
measurement_setup	**02**
measurment_type	spectral_direct_normal_irradiance
device_type	spectroradiometer
wavelength_resolution	1
filter_type	ND OD1
aperture_angle	90
detector_orientation	90
detector_inclination	90
notes	measurements smoothed and interpolated

sun_positions	
location_code	**DE-BLN**
timestamp_local	**2015-03-12 11:38 CET**
measurement_setup	02
sun_elevation	32.9
sun_azimuth	168.5
solar_time	11:21

weather	
location_code	**FR-VLX**
timestamp_local	**2016-06-23 13:55 CEST**
measurement_setup	02
sky_type_cie	NA
sky_condition	clear
global_horizontal_illuminance	101000
global_horizontal_irradiance	988

Figure 2 – Overview of supplementary data frames - column names and example values

3.1.3 Relational aspects of the data package

Each of the measurement data frames is structured to be standalone such that meaningful information can be extracted for the required application. However, when performing data analysis, it may become necessary to combine information between data frames, especially the supplementary data frames. To facilitate this process in programming languages like R or Python, relational attributes are established between data frames using ‘primary’ and ‘foreign’ keys. The concept of primary and foreign keys is common in relational database management systems.

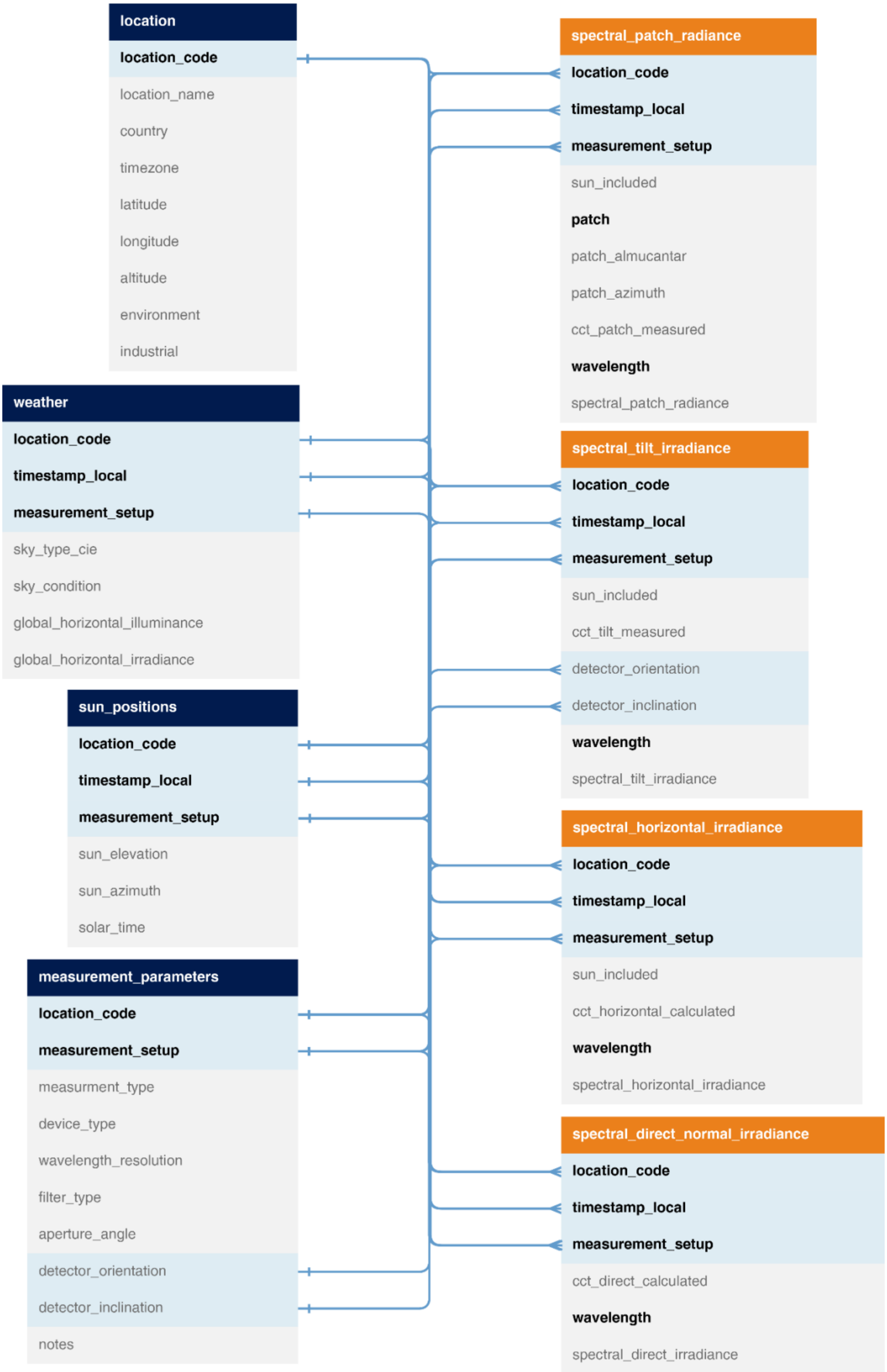


Figure 3 – Overview of supplementary (left) and measurement (right) data frames highlighting primary and foreign keys

The measurement and supplementary data frames consist of rows and columns. Columns represent variables described in Section 3.1 and rows represent individual observations. Each row has a unique identifier, or primary key, to prevent data repetition. The primary keys are bold variable names in Figure 3. For example, in the spectral_patch_radiance data frame, the primary key is a combination of location_code, timestamp_local, measurement_setup, patch, and wavelength.

Foreign keys are used to establish links between the data frames. Variable names highlighted in blue in Figure 3 represent foreign keys. As illustrated in Figure 3, a variable can function as both a primary and foreign key. For example, the location data frame can be linked to the spectral_patch_radiance data frame using the foreign key location_code, which is also a primary key in both the location and spectral_patch_radiance data frames. Similarly, the weather data frame can be linked to any measurement data frame using a combination of location_code, timestamp_local and measurement_setup.

By using primary and foreign keys, data frames can be seamlessly joined for various data manipulations. This approach avoids data duplication when linking the data frames and performing analysis.

4 Application of SKYSPECTRA

The collection and standardisation efforts for SKYSPECTRA provide a valuable resource for fundamental research in the field of lighting. It supports the objectives of the CIE TC 3-60 committee in investigating variations in geographical, seasonal, and diurnal variations in daylight spectral power distribution (SPD). It enables adaptations to existing CIE reconstitution procedures (Judd et al. 1964, CIE 2018), currently used to define Standard Daylight Illuminants, such as D65 and create representative local SPDs for lighting research and planning. The dataset also aids in characterising the spectral non-uniformity of the sky for further development of localised spectral sky models (Balakrishnan et al. 2023).

Moreover, the dataset is useful for technological applications within the lighting industry, supporting the development of simplified measuring devices for spectral daylight measurements using inexpensive colour sensors. The dataset can serve as a validation dataset for such device developments. SKYSPECTRA is of interest to other domains, such as photovoltaic (PV) cell manufacturers. They can use the dataset to design and characterise their systems effectively, as it provides valuable information on the variations in the spectrum of light across seasons and locations, which significantly impact the efficiency of PV cells (Kinsey et al. 2022).

In daylight prediction and simulation programmes, the dataset enhances the accuracy, accessibility and applicability of spectral daylight assessments for urban and indoor environments (Balakrishnan & Jakubiec 2019, Kenny 2021, Pierson et al. 2023). It allows for customising sky models in simulations, validating spectral sky models in different simulation tools, and enabling the development of spectrally dependent daylight evaluation metrics. Thereby also facilitating previously challenging comparative studies considering the location-specific variations in, for example, circadian entrainment due to practical difficulties of obtaining such data.

5 Conclusions

This paper presents an open-source data package, SKYSPECTRA that standardises and structures spectral daylight measurement data collected from world-wide sources as a part of the CIE TC 3-60 initiative. SKYSPECTRA will be available online from September 2023 (DOI:10.5281/zenodo.8147546). By providing a comprehensive and well-documented spectral daylight measurement dataset that represents different geographical locations, seasons, atmospheric conditions, or times of day, this data package aims to further research and development of daylight applications that leverage spectral sky data, ultimately leading to improved lighting solutions and human well-being considerations.

6 Acknowledgements

This project is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Individual Fellowship (grant agreement No. 101032279)

References

Balakrishnan, P. 2019. *Measuring and modelling equatorial light*. (Doctoral dissertation, Singapore University of Technology and Design, Singapore)

Balakrishnan, P., Doppler, L., Knoop, M. 2023. September. *A framework to generate local spectral skies for spectral daylight simulations*. In Journal of Physics: Conference Series. IOP Publishing.

Balakrishnan, P., Jakubiec, J. A. 2019. *Spectral rendering with daylight: a comparison of two spectral daylight simulation platforms*. Proceedings of the BS2019, Rome, Italy, 2-4

CIE 1994. CIE 108:1994 *Guide to recommended practice of daylight measurement.* Vienna. CIE Central Bureau

CIE 2018. CIE 015:2018 *Colorimetry*. Vienna, Austria.

Diakite-Kortlever, A. K, Balakrishnan, P., Darula, S., Geisler-Moroder, D., Jakubiec, A. J., Knoop, M. et al. 2021. *Spectral sky models for advanced daylight simulations*. A Technical Report of International Energy Agency (IEA) SHC Task 61 / EBC Annex 77 Subtask C3.

Diakite-Kortlever, A., Weber, N., & Knoop, M. 2022. *Reconstruction of Daylight Spectral Power Distribution Based on Correlated Color Temperature: A Comparative Study between the CIE Approach and Localized Procedures in Assessing Non-image Forming Effects*. LEUKOS, 19(2), 18–145. DOI:10.1080/15502724.2021.1989310

Hernández-Andrés, J., Romero, J., Lee, R. L. 2001a. *Colorimetric and spectroradiometric characteristics of narrow-field-of-view clear skylight in Granada, Spain*. In: Journal of the Optical Society of America. A, Optics, image science, and vision 18 (2), S. 412–420. DOI: 10.1364/JOSAA.18.000412.

Hernández-Andrés, J., Romero, J., Nieves, J. L., Lee, R. L. 2001b. *Color and spectral analysis of daylight in southern Europe*. In: Journal of the Optical Society of America. A, Optics, image science, and vision 18 (6), S. 1325–1335. DOI: 10.1364/JOSAA.18.001325.

Inanici, M. 2019. *Tri-stimulus color accuracy in image-based sky models: simulating the impact of color distributions throughout the sky dome on daylit interiors with different* … In: F. Cappelletti und I. Ballarini (Hg.): Proceedings of Building simulation 2019: 16th IBPSA international conference and exhibition: Taylor & Francis (27).

Kenny, P., 2021. *Characterising the melanopic/spectral microclimate of indoor spaces.* (Doctoral dissertation, Loughborough University).

Knoop, M., Stefani, O., Bueno, B., et al. 2020. *Daylight: What makes the difference?* Lighting Research & Technology.;52(3):423-442. DOI:10.1177/1477153519869758

Knoop, M., Weber, N., Diakite, A. K. 2019. *Approach to analyse seasonal and geographical variations in daylight illuminants.* In: CIE x046:2019 Proceedings of the 29th Session of the CIE. Washington, USA, July. CIE. Vienna, Austria: International Commission on Illumination, CIE.

Kinsey, Geoffrey S., Riedel-Lyngskær, Nicholas C., et al., 2022. Impact of measured spectrum variation on solar photovoltaic efficiencies worldwide. In: Renewable Energy 196, S. 995–1016.

Luo, T., Lin. R., Yan, D., & Zhang, B. 2015. *Measurement and study on the sky luminance and spectral distribution in Beijing*. CIE 216:2015: Proceedings of the 28th Session of the CIE, 28 June – 4 July 2015, Manchester, United Kingdom.

Morimoto, T., Zhang, C., Fukuda, K., Uchikawa, K., 2022. *Spectral measurement of daylights and surface properties of natural objects in Japan*. In: Opt. Express, OE 30 (3), S. 3183. DOI: 10.1364/OE.441063.

Münch, M., Wirz-Justice, A., Brown, S. A., et al., 2020. *The Role of Daylight for Humans: Gaps in Current Knowledge. Clocks & sleep*, 2(1), 61–85. DOI:10.3390/clockssleep2010008

Peterson, J., Northcutt, D., et al. (no date). *Structure of the Lindenberg, Germany spectral radiation data*. Available: http://solardat.uoregon.edu/cgi-bin/ShowSpectralFiles.cgi (Accessed: 14 July 2023)

Pierson, C., Aarts, M.P. and Andersen, M., 2023. *Validation of spectral simulation tools in the context of ipRGC-influenced light responses of building occupants*. Journal of Building Performance Simulation, 16(2), pp.179-197.

Robertson AR. 1968. *Computation of correlated color temperature and distribution temperature*. J Opt Soc Am. 58(11):1528

Spitschan, M., Aguirre, G. K., Brainard, D. H., Sweeney, A. M. 2016. *Variation of outdoor illumination as a function of solar elevation and light pollution*. In: Sci Rep 6 (1), S. 26756. DOI: 10.1038/srep26756.

Tourasse, G. 2016. *Mesure et Analyse Statistique Tout Temps du Spectre du Rayonnement Solaire. Electromagnétisme*. PhD thesis. University of Lyon, France. NNT: 2016LYSET014. tel-01466710

Vignola, F., Peterson, J., Kessler, R. et al., 2021. *Improved field evaluation of reference cells using spectral measurements*. In: Solar Energy 215, S. 482–491.

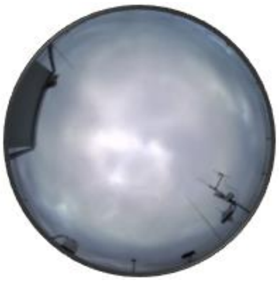
Wickham, H. 2014. “Tidy Data”, *Journal of Statistical Software*, 59(10), pp. 1–23. DOI: 10.18637/jss.v059.i10.

Xue, P., Wang, H., Luo, T., Zhao, Y., Fan, C., & Ma, T. 2022. *Clear sky color modeling based on BP neural network*. Building and Environment, 226, 109715

A.1 Appendix A: Spectral measuring site template for Berlin

Description of site	
name contact person:	Martine Knoop
location code:	DE-BLN
time zone in UTC, without Daylight Saving Time:	+01
geographical latitude:	52.51
hemisphere - latitude:	N
geographical longitude:	13.33
hemisphere - longitude:	E
altitude - Above Mean Sea Level:	35
height of the building on which the measuring site is located:	20
setting:	urban
industrial setting:	no
typical turbidity:	average

number of devices on site included in this template:	01
pictures of measuring site to reveal obstruction:	available
publication on measurement site: Diakite-Kortlever, A., Weber, N., & Knoop, M. (2022). Reconstruction of Daylight Spectral Power Distribution Based on Correlated Color Temperature: A Comparative Study between the CIE Approach and Localized Procedures in Assessing Non-image Forming Effects. LEUKOS, 1-28.	
Measuring device	
device type:	spectroradiometer
brand:	Zeiss Multi-Channel Spectrometer MCS-CCD with a Hamamatsu back-thinned CCD sensor
type:	custom-made
details:	see paper Diakite-Kortlever et al. 2022 3.1.2. Measuring equipment
measured quantity:	spectral patch radiance
type of daylight measured:	skylight
wavelength range calibrated range - start:	380
wavelength range calibrated range - end:	780
optical resolution - FWHM:	0.8
wavelength resolution - step size data:	1
type of filter, if available:	ND
location of the filter, if available:	--
details to calibration or characterisation of the system: Relative and absolute characterization of the measuring equipment was conducted in the laboratory of the Technical University of Berlin, with a FEL lamp for the spectral, relative characterization. Absolute characterization has been performed with a large uniform LED light source, with a prominent short-wavelength component in combination with a calibrated luminance meter (Knoop et al. 2017). Latest characterisation beginning of 2022. As described in Knoop, M., Diakite, A., Weber, N., Luo, T., & Abdelmageed, A. E. (2017). Spatially resolved spectral sky daylight measurement data: Methods of collection, evaluation and representation. In Proceedings of the Conference at the CIE Midterm Meeting (pp. 23-25).	
Measurements	
type of measurement:	patch
patch measurements	
number of patches:	145

location of patches:	according to Tregenza (2004, fig. 2a)
aperture angle:	10
almucantar height for patch measurements:	12
(approximate, maximum) duration of one complete scan under normal daytime lighting conditions in minutes:	1
frequency of scans (per hour):	30
Data format	
time stamp used in data set:	UTC
Pictures	
general picture of measuring site (panoramic or fish eye)	
to reveal obstruction: 360° fish eye picture with horizon line	
a topographic map showing the land within a 5-kilometre radius of the site	
in case of patches: location of patches