



SolarDecomp: A Web App for Decomposing Solar Data for Spectrally Selective Building Simulation

Chenshun Chen^(✉), Qiuhua Duan, Yanxiao Feng, and Julian Wang^(✉)

Department of Architectural Engineering, Penn State University, State College, USA
{cjc7032, jqw5965, cjc7032, jqw5965}@psu.edu

Abstract. Solar radiation plays an important role in solar architecture design. It not only determines the optical regime of the building envelope, but also influences the heating and cooling loads of the building. To understand the impacts of solar radiation on building thermal and lighting performance, computational analysis is essential. However, conventional modeling tools only take broadband solar radiation into consideration, which limits the modeling accuracy since window glazing is spectrally dependent. In this sense, we developed a new solar decomposing tool that can separate major solar irradiation components such as VIS, NIR in GHI, DNI and DHI, by using machine learning algorithms such as extreme boosting regression tree (*XGBoost*). The predictors are meteorological parameters that already exist or can be easily derived from traditional weather files (such as TMY3). Model performance is validated by NREL Solar Radiation Research Laboratory's spectral solar dataset. Integrating these decomposing models, a web app (SolarDecomp) has been developed to enable researchers and designers import solar spectral data into existing building simulation programs for specific simulations in terms of spectrally selective design components.

Keywords: Narrowband solar irradiation · Solar architecture design and simulation · TMY weather files · Spectral solar dataset · Solar architecture

1 Introduction

Solar radiation plays a crucial role in solar architectural design, which strives to optimize the use of the sun's energy for the creation of energy-efficient structures. Several computational tools are employed to examine the thermal and optical attributes of such designs. Typically, weather data for an entire year is loaded into a simulation program to model conditions like a building's daylight environment and energy consumption. A comprehensive solar-radiation dataset is indispensable during this process, ensuring reliable forecasting for solar initiatives. The consistency of the solar resource, as reflected in historical solar data, coupled with the accuracy of the dataset, significantly impacts the accuracy of the estimates [1].

Several comprehensive datasets are readily accessible: 1) The National Solar Radiation Data Base (NSRDB), a creation of the National Renewable Energy Laboratory

(NREL) and Sandia National Laboratory. 2) The Surface Meteorology and Solar Energy (SSE) by NASA, offering global coverage of monthly averages, annual data, and data every three hours from 1983–2005. 3) SolarGIS, which merges ground observations, satellite data, and an atmospheric patterns database to deliver highly accurate global solar radiation data. The Typical Meteorological Year (TMY) data files were derived from these datasets. Each TMY file incorporates a complete year's data, created from 12 representative months chosen from the database's total years. The TMY data includes three key solar components: Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI), and Diffuse Horizontal Irradiance (DHI). These are crucial for solar analysis or energy simulation. Importantly, all three types of solar irradiance (GHI, DNI, DHI) in the typical weather data files are broadband, signifying the total quantum of Ultraviolet (UV), Visible Light (VIS), and Near-Infrared Radiation (NIR), which are the main constituents of the solar spectrum.

In recent times, growing evidence from the building and solar simulation has underscored the importance of partitioning broadband solar radiation into narrowband or even design and analytics focused on specific spectrum. For instance, among these three solar components, solar VIS radiation is advantageous for conserving electrical lighting and enhancing indoor health due to its ability to regulate circadian rhythm [2]. However, it may also have adverse effects, leading to visual problems such as glare. While solar NIR transmission into the building could aid in reducing the total heating load in cold climates, which is undesirable in hot climates. Recent studies also show that solar VIS and NIR that penetrate through window have different impacts on thermal comfort near window zones [3]. Regarding the design of solar devices or systems, with the advent of spectral selective and nano-structured thin films for energy-efficient envelopes, the independent modulation of solar radiation in different spectral bands has become increasingly viable as a practical solution for improving building energy savings and maintaining indoor visual comfort. The example studies include reversible photothermal windows [4], silver nanorods-embedded smart windows [5], and other light/heat splitting materials [6–9].

However, the full potential of the materials/structures mentioned earlier in building envelopes may not be comprehended thoroughly if corresponding simulations only factor in broadband solar radiation. Even worse, it may lead to misleading or even erroneous outcomes. Traditional weather files usually do not contain the spectral distribution or narrowband details of solar irradiation due to the complexity and high cost associated with spectral data measurement. To bridge this research gap, it is necessary to develop a reconstruction algorithm that is able to decompose solar visible and infrared irradiance from broadband solar radiation and weather data for building simulations. The solar radiation data used for such simulations include GHI, DHI, and DNI, so all these three solar components need to be decomposed to narrowband data. In our previous work, we have already developed predicting models using the machine learning algorithms such as CART for decomposing VIS and NIR components in GHI, DNI, and DHI [10]. ***This work presents our preliminary development of a data conversion portal*** built upon these predicting models so that broadband solar irradiance inside traditional TMY datasets can be replaced by individual solar components (i.e., VIS and NIR components in GHI, DNI, and DHI), then these modified TMY files can be used for spectrum-related building simulations.

2 Methodology

Firstly, we built two estimation models for decomposing solar components (VIS and NIR) from the broadband GHI and DNI, respectively. The GHI decomposing model and the framework of the data portal have been presented in prior work [10]. Based on these two elements (GHI and DNI), it is possible to calculate the VIS and NIR components in DHI via the transposition equation. Subsequently, one can import their original weather files (such as TMY), and then export processed weather files with individual solar components in replace of broadband GHI, DNI, and DHI information.

In this study, all the datasets were obtained from the BMS database of the NREL Solar Radiation Research Laboratory. All the datasets were based in Golden, Colorado with a latitude 39.742° N, a longitude of 105.18° W, and an elevation of 1828.8 m AMSL [12]. In general, two major datasets were curated: weather datasets (including broadband solar irradiance and other typical meteorological data) and spectral solar irradiance data from multiple sources.

1) Weather Dataset

This dataset includes three components. The major component was based on hourly meteorological measurements (HMM) and included the most independent variables for modeling, such as GHI, DNI, DHI, cloud coverage, dry-bulb temperature, dewpoint, relative humidity, and wind speed. The HMM is hourly data by averaging the value of all measurements taken from the 1-h interval and was collected from Jan 1, 2016, to December 31, 2019. In addition, based on the prior studies, both aerosol optical depth and precipitable water vapor parameters are often found in typical weather stations' data collection and are important to the solar spectra, while they are not included in HMM datasets. As such, in this work, we obtained these two atmospheric data from the NREL BMS AOD and PWV (GPS-based PWV) database. As a result, the curated weather dataset has identical variables and formats to the TMY weather file that has been widely applied to solar radiation and building energy performance simulations. All these weather data including broadband solar irradiance were used as the predicting variable candidates in this work.

2) Spectral Solar Dataset

The corresponding spectral solar components (VIS and NIR irradiance) in GHI were extracted from the outdoor solar spectra data (WISER), and the components in DNI were extracted from the outdoor solar spectra data (PGS-100). The WISER dataset was measured by EKO WISER spectroradiometer MS-710, MS-711, and MS-712 from 2016 to 2019. The MS-710, MS-711, and MS-712 instruments have 4nm, 5nm, and 6.5nm spectral bandwidth respectively. Their measurement range is 300 nm–1100 nm, 300 nm–1100 nm, and 900 nm–1700 nm, respectively [13]. The WISER dataset has a higher resolution measurement for both wavelengths and time intervals (typically 5 min, but occasionally 1 min). The hourly spectrum data were calculated by averaging the 5-min interval data for each hour. The average value of all measured points each hour is defined as the value for the time-stamp at the end of the 1-h interval. The PGS-100 dataset was measured through a LICOR LI-1800 spectroradiometer. The instrument has 3.6 nm spectral bandwidth and the usable spectral range of the instrument is 350 to 1050 nm. Data was taken at approximately 0.7 nm intervals (slightly variable, differs for each serial number) every 5 min [32].

3 Results and Discussion

3.1 VIS and NIR Components in GHI and DNI

In our previous work, we demonstrated the feasibility of decomposing broadband GHI into VIS and NIR components by using the CART algorithm, a simple logistic diagram was developed to cluster VIS and NIR into different fractions of GHI, based on meteorological parameters. The regression tree models for VIS/GHI and NIR/GHI were shown in Fig. 1.

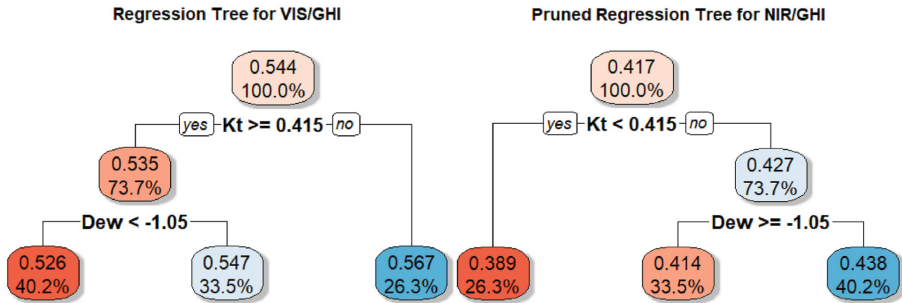


Fig. 1. Regression trees for VIS/GHI (left) and NIR/GHI (right)

Table 1. Model comparison for VIS, NIR in DNI

Model	VIS			NIR		
	RMSE	MAE	R ²	RMSE	MAE	R ²
Cost complexity tree	25.1084	12.9560	0.9650	26.1609	16.9067	0.9791
M5' tree	19.9164	9.6899	0.9780	21.3417	10.6917	0.9861
Random forest	19.1448	9.3176	0.9797	19.9415	9.7087	0.9879
XGBoost	18.2803	7.9894	0.9814	18.3894	8.0108	0.9897

In this study, the general process of predicting VIS and NIR from broadband DNI is shown as an example illustrating the steps of constructing the embedded machine learning algorithm in our SolarDecomp tool. Firstly, by using *sklearn*, *m5py*, and *XGBoost* packages in Python, we built various regression tree models, including Cost-complexity tree, M5' tree, Random forest and Extreme boosting tree. The entire dataset D was split into a training dataset (80% of D) and a testing dataset (20% of D). Different models were tuned by using cross-validation to find the best training parameters so that each individual model was optimized to fit the training set. After the parameter tuning process, a performance evaluation was conducted to find the best model. As shown in Table 1, XGBoost was found as the best-fit model for predicting both VIS and NIR components in DNI, as it provided the lowest RMSE, MAE, and R^2 results when fitting the testing datasets. Tuned parameters for this optimized XGBoost model are shown in Table 2.

Table 2. XGBoost Parameter Tuning

Parameter	Model type	
	VIS	NIR
max_depth	6	4
min_child_weight	5	4
gamma	0	0.4
subsample	0.8	1.0
colsample_bytree	0.8	1.0
reg_alpha	100	100

3.2 VIS and NIR Components in DHI

Solar transposition rule has been frequently used to describe the mathematical relationship among GHI, DHI, and DNI on a horizontal surface: GHI is the sum of incident DHI plus the DNI projected onto the horizontal surface.

$$GHI = DNI \times \cos(sza) + DHI \quad (1)$$

Theoretically, one can use this equation to compute VIS and NIR components in DHI if one takes VIS and NIR components in DNI and GHI as input. However, it is unknown whether this relationship is still valid for each spectral solar component. To verify the transposition model, two datasets (collected between May 12th and June 1st 2021) in the BMS database of the NREL Solar Radiation Research Laboratory were selected. The measured spectral DHI data were obtained by EKO MS-711 and MS-712 sensors, and then VIS and NIR of DHI were prepared. On the other hand, the RG 780 dataset (Schott Glass filter with a cut-on wavelength at 780nm) was used to extract VIS and NIR components from GHI and DNI, which were then input into the transposition equation to compute the VIS and NIR components of DHI. As such, both measured and calculated VIS and NIR irradiance of DHI were obtained and then compared.

Figure 2 shows the comparison results of VIS in DHI, the comparison of NIR in DHI has similar results. The average relative error between the measured VIS and NIR data and the data calculated by the transposition model is 7.62% and 6.89% for VIS and NIR, respectively. This comparison demonstrates that the diffuse, direct, and global spectral solar irradiance values are also complying with the relationship defined by the transposition model. After the solar transposition model was validated in the broadband solar spectrum, VIS and NIR components in DHI can be calculated by using previously predicted VIS and NIR in GHI and DNI as inputs to the solar transposition model. In general, NIR occupies a small portion of DHI since the most diffused solar irradiance is from the sky and tends to be in short wavelengths. There are potential errors with this prediction and computation method as the calculated spectral component of DHI might accumulate prediction errors from both DNI and GHI. As such, in the actual application, a set of calibrations will be processed before the final spectral data of solar are reported to users.

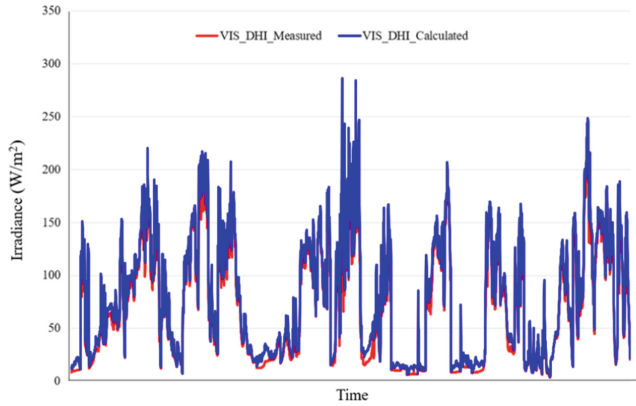


Fig. 2. Comparison of measured/calculated VIS results in DHI

4 Web Portal Development of Solar Decomposing in Solar Architecture Design

The GHI, DNI, and DHI solar decomposing models were integrated into an executable web app, SolarDecomp, via Python *streamlit* package. After a user chooses their preferable solar component, they will be asked to upload a weather file (e.g., TMY file) and the location of that weather file, then SolarDecomp would automatically execute previously trained XGBoost model and generate a different weather file in which the original broadband GHI, DNI, and DHI are replaced with the narrowband solar components (VIS or NIR). The final outputs would be new weather files labeled TMY_VIS and TMY_NIR. Note that the input weather file should be pre-processed to ensure that the input file is in.csv format (it is recommended to use EnergyPlus™ Weather Converter to convert the original.epw file to.csv file), all features used for predicting exists (except for newly added parameters, such as K_b and sza , which will be automatically computed based on the existing parameters) and all features' names and units should be in the same format. Table 3 summarizes all the required input variables and their name formats. General instruction of using SolarDecomp is shown in Fig. 3.

It's also worth mentioning that the predicting values will be automatically calibrated before exporting the final output files. The calibration logic is shown in Fig. 4. While feeding the model with weather files, if any features were missing or their numbers were meaningless (e.g., some stations do not provide Albedo, liquid precipitation depth, snow depth, AOD, or PWV data), a simplified model will be used to replace of the full model. This simplified model could only use part of the features to make predictions: for example, GHI, DNI, DHI, SKC, Opqcl, Dry, Dew, RH, AM, I_0 , K_b , and Tcld. Normally, the simplified model can be used at most of the locations, but this generalizability is a consequence of losing accuracy (e.g., the simplified model has a testing RMSE = 20.6103, MAE = 10.2091, $R^2 = 0.9764$ for VIS/DNI decomposing).

With such new weather files, designers and researchers are now able to integrate separable solar irradiance when simulating a building solar environment by using simulation tools such as ClimateStudio, EnergyPlus, etc. Figure 5 demonstrates a simple

Table 3. Model required input variables and their unified formats

Feature	Name format	Units format	Source
<i>Global Horizontal Irradiance</i>	GHI	W/m ²	TMY
<i>Direct Normal Irradiance</i>	DNI	W/m ²	TMY
<i>Diffuse Horizontal Irradiance</i>	DHI	W/m ²	TMY
<i>Total sky cover</i>	SKC	.1	TMY
<i>Opaque sky cover</i>	Opqclد	.1	TMY
<i>Dry-bulb temperature</i>	Dry	°C	TMY
<i>Dew-point temperature</i>	Dew	°C	TMY
<i>Relative Humidity</i>	RH	%	TMY
<i>Atmospheric pressure</i>	Pressure	Mbar	TMY
<i>Zenith angle</i>	Zenith	°	Calculation
<i>Air mass</i>	AM		Calculation
<i>Extraterrestrial Direct Normal Radiation</i>	I0	W/m ²	Calculation
<i>Normal clearness index</i>	Kb		Calculation
<i>Cloud transmittance</i>	Tcld		Calculation
<i>Azimuth angle</i>	Azimuth	°	Calculation
<i>Wind direction</i>	Wdr	° from N	TMY
<i>Wind speed</i>	Wspd	m/s	TMY
<i>Albedo</i>	Albedo	.01	TMY
<i>Liquid precipitation depth</i>	Precip	mm	TMY
<i>Snow depth</i>	Snow	cm	TMY
<i>Aerosols Optical Depth</i>	AOD	.001	TMY
<i>Precipitable Water Vapor</i>	PWV	mm	TMY

example of investigating the surface irradiance in a high-dense urban area. The urban model contains all-glass buildings, on which Double Low-E windows were applied. These Low-E windows had relatively high VIS transmittance (51%) but extremely low NIR transmittance (0.1%) (glazing materials' spectral properties modeled in LBNL Optics software, and averaged across the spectrum). The original weather file (Philadelphia_International_Ap:: 724080:: TMY3) was modified by the data conversion tool and then imported to simulation software (ClimateStudio in Rhino). With the original weather file as the input, ground solar irradiance in July was around 138 kWh/m², however, after separating the simulation into VIS and NIR parts, the total summation of VIS and NIR irradiance was around 176 kWh/m², much larger than the original simulation. Although this demonstration is simple and inaccurate, it still reveals the necessity of separating solar components for building solar simulations, especially when considering spectral-selective windows.

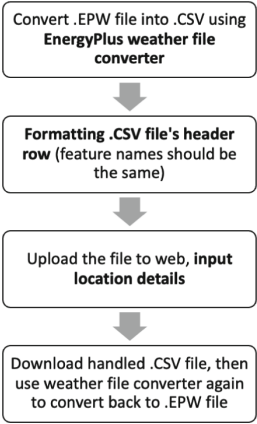


Fig. 3. Instruction for converting traditional weather file to VIS/NIR weather file by using SolarDecomp

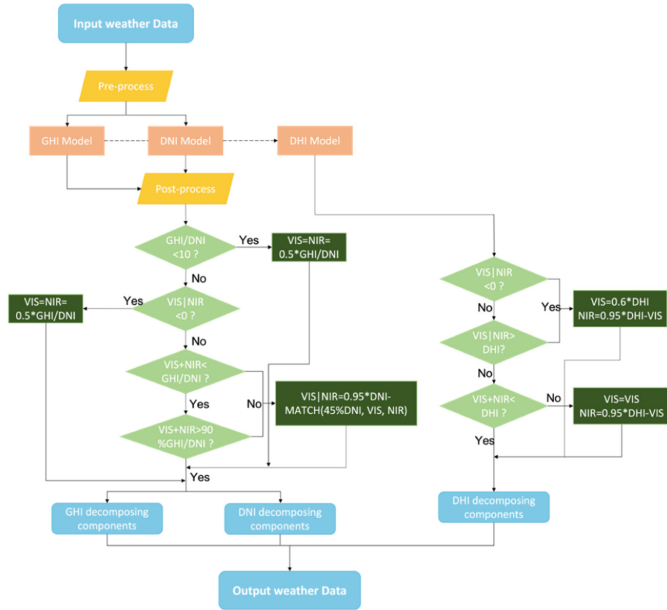


Fig. 4. Logistic diagram of model calibration

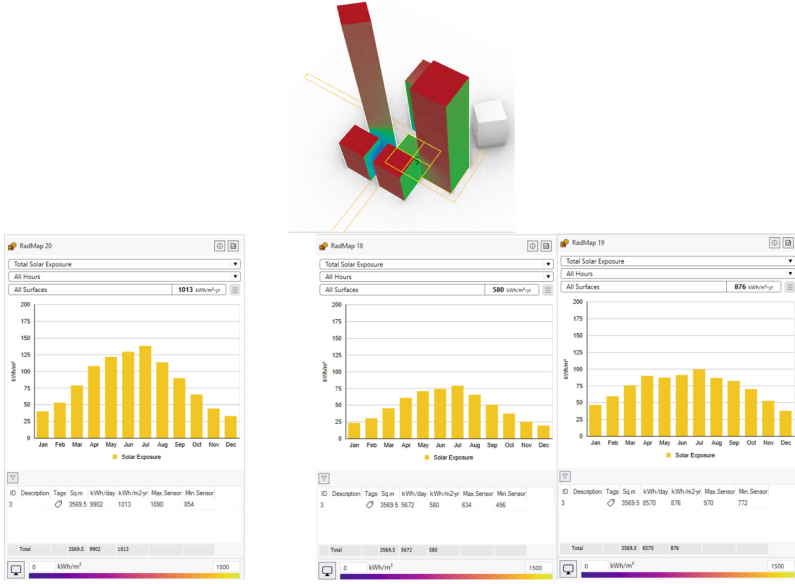


Fig. 5. Urban solar irradiance simulation demo: Bottom left: Ground solar irradiance with original weather file; Bottom right: VIS and NIR solar irradiance with modified weather file

5 Conclusion

In conclusion, we have successfully developed an executable web app, SolarDecomp, leveraging the Python streamlit package, which integrates models for decomposing GHI, DNI, and DHI solar data. Users can upload a weather file and its location, and the app will use the pre-trained XGBoost model to generate a modified weather file with distinct narrowband solar components (VIS or NIR). The result is new weather files, TMY_VIS and TMY_NIR. The input file should be pre-processed into a.csv format, with all necessary predictive features present and correctly formatted. This research highlights the effective use of machine learning to convert broadband solar data into separate components, utilizing readily available meteorological data. Despite these advancements, there remain research gaps such as the need for solar analyses on tilted and vertical surfaces and model validation under different climate conditions, the generalizability of trained model by using solar spectral information from different locations, which will be addressed in future work.

Acknowledgements. This project is supported by the NSF award: # 2001207: CAREER: Understanding the Thermal and Optical Behaviors of the Near Infrared (NIR)-Selective Dynamic Glazing Structures.

References

1. Vignola, F., Grover, C., Lemon, N., McMahan, A.: Building a bankable solar radiation dataset. *Sol. Energy* **86**(8), 2218–2229 (2012). <https://doi.org/10.1016/j.solener.2012.05.013>
2. Roberts, J.E.: Visible light induced changes in the immune response through an eye-brain mechanism (photoneuroimmunology). *J. Photochem. Photobiol. B* **29**(1), 3–15 (1995). [https://doi.org/10.1016/1011-1344\(95\)90241-4](https://doi.org/10.1016/1011-1344(95)90241-4)
3. Wang, N., Wang, J.: A spectrally-resolved method for evaluating the solar effect on user thermal comfort in the near-window zone. *Build. Environ.* **202**, 108044 (2021). <https://doi.org/10.1016/j.buildenv.2021.108044>
4. Anwar Jahid, M., Wang, J., Zhang, E., Duan, Q., Feng, Y.: Energy savings potential of reversible photothermal windows with near infrared-selective plasmonic nanofilms. *Energy Convers. Manag.* **263**, 115705 (2022). <https://doi.org/10.1016/j.enconman.2022.115705>
5. Pu, J., Shen, C., Yang, S., Zhang, C., Chwieduk, D., Kalogirou, S.A.: Feasibility investigation on using silver nanorods in energy saving windows for light/heat decoupling. *Energy* **245**, 123289 (2022). <https://doi.org/10.1016/j.energy.2022.123289>
6. Du, W.-C., Xie, J., Xia, L., Liu, Y.-J., Yang, H.-W., Zhang, Y.: Study of new solar film based on near-infrared shielding. *J. Photochem. Photobiol. Chem.* **418**, 113410 (2021). <https://doi.org/10.1016/j.jphotochem.2021.113410>
7. Besteiro, L.V., Kong, X.-T., Wang, Z., Rosei, F., Govorov, A.O.: Plasmonic glasses and films based on alternative inexpensive materials for blocking infrared radiation. *Nano Lett.* **18**(5), 3147–3156 (2018). <https://doi.org/10.1021/acs.nanolett.8b00764>
8. Gao, Q., Wu, X., Huang, T.: Novel energy efficient window coatings based on In doped CuS nanocrystals with enhanced NIR shielding performance. *Sol. Energy* **220**, 1–7 (2021). <https://doi.org/10.1016/j.solener.2021.02.045>
9. Xu, Q., et al.: Cs_{0.33}WO₃ as a high-performance transparent solar radiation shielding material for windows. *J. Appl. Phys.* **124**(19), 193102 (2018). <https://doi.org/10.1063/1.5050041>
10. Duan, Q., Feng, Y., Wang, J.: Clustering of visible and infrared solar irradiance for solar architecture design and analysis. *Renew. Energy* **165**, 668–677 (2021). <https://doi.org/10.1016/j.renene.2020.11.080>
11. Myers, D.R.: *Solar Radiation: Practical Modeling for Renewable Energy Applications*, 1st ed. CRC Press (2017). <https://doi.org/10.1201/b13898>
12. Achleitner, S., Kamthe, A., Liu, T., Cerpa, A.E.: SIPS: Solar irradiance prediction system. In: *IPSN-2014 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, pp. 225–236, April 2014. <https://doi.org/10.1109/IPSN.2014.6846755>
13. MIDC: SRRL BMS Instruments. https://midcdmz.nrel.gov/srrl_bms/instruments.html. Accessed 16 Dec 2022