Estimating Hourly Solar Near-infrared Irradiance Using Meteorological Data for Sustainable Building Design and Engineering

Qiuhua Duan[1], Yanxiao Feng[1] and Julian Wang[1,*]

[1] Department of Architectural Engineering, Pennsylvania State University, State College (United States)

[*] Corresponding Author Julian.wang@psu.edu

Abstract

Solar radiation is a key factor influencing sustainable building engineering, in terms of both optical and thermal properties of building envelopes. Solar irradiance data in a conventional weather data file are broadband, representing the total of ultraviolet (UV), visible light (VIS), and near-infrared radiation (NIR), three components of the solar spectrum; however, these three components play different roles in sustainable building design and engineering. For instance, solar VIS always provides benefits to indoor building energy savings (e.g., electrical lighting), while solar NIR is beneficial to building energy savings in winter but undesirable in summer. As a consequence, there is a need for reliable separate analyses focusing on individual solar radiation components. In this work, we explore and test classification-based modeling methods for decomposing hourly broadband global horizontal solar irradiance data in conventional weather files into hourly global horizontal solar NIR component. This model can then be conveniently implemented for sustainable building design and engineering purposes.

Keywords: Solar Radiation, Solar Building, Classification Trees, Prediction Model, Solar Energy, Building Energy

1. Introduction

Solar architecture is one of major category in the field of sustainable buildings, which makes the best possible use of locally available solar energy by employing both passive and active measures to achieve building sustainability and energy efficiency goals (Schittich 2012). The first solar building in America was proposed by Tod Neubauer in the 1950s.(Perlin 2013) Research in this field has addressed the theoretical background, simulation techniques, and experimental testing. Computational analysis in solar building design and engineering has been described and discussed widely in recent decades (Kisilewicz 2007). Usually an entire year's weather data are imported in a conventional format (e.g., Typical Meteorological Year 2 (TMY2), Typical Meteorological Year 3 (TMY3), Weather Year for Energy Calculations Version 2 (WYEC2)) into an energy simulation program to calculate the energy consumption of a building. Solar irradiance data in a complete weather file also include global horizontal irradiation (GHI), diffuse horizontal irradiation (DHI), and direct normal irradiation (DNI). Regardless of the three solar irradiance types noted above, the solar irradiance data are broadband and represent the total of ultraviolet (UV), visible light (VIS), and near-infrared radiation (NIR), three components of the solar spectrum.

With two known solar data components, the other component can be calculated via the mathematical relations among them. However, these three components play different roles in sustainable building engineering and design. Of these three major components, solar VIS always provides benefits to indoor building energy savings (e.g., electrical lighting), while solar NIR is strongly correlated to solar heat gains that are beneficial to building energy savings in winter but undesirable in summer (Eicker 2006). Similarly, the COVID-19 pandemic has heightened interest in the solar UV component and its potential impact on the spread and seasonality of disease. Therefore, in some in-depth building environment performance analyses, especially building energy simulation work, separate analyses focusing on each solar radiation component are desirable. With recent discoveries and engineering solutions emerging related to nanomaterials and nanostructures, independent band modulation of solar radiation on building envelopes (including glazing systems) has become increasingly viable as a potential means of improving building energy savings and indoor visual comfort (Wang et al. 2017). Meanwhile,

doi:10.18086/solar.2020.01.12 Available at http://proceedings.ises.org

separated indicators or parameters related to solar light and heat have been integrated into the sustainable building design and engineering processes. For example, in our previous works solar heat gain coefficient and visible transmittance, two major building window properties, have been comprehensively studied in terms of their different impacts on building energy, potential conflicting effects, and measurement methods. (Feng et al. 2020; Wang et al. 2016). However, the meteorological data in conventional weather files do not normally include the spectral power distribution data of incident solar light because measuring the narrowband spectral distribution of sunlight is much more difficult and expensive than measuring broadband radiation (e.g., using pyranometers) (Duan et al. 2020). As a consequence, there is a pressing need for reliable performance estimations of spectral solar radiation control and response on a building scale. To assess this, we need band solar irradiance data as input.

To address this research gap and the practical need for sustainable building design and engineering, this work has developed an estimation model for the NIR component that can be captured efficiently from readily available datasets without the addition of new measurements and associated sensors; this can then be conveniently implemented into current practices and research activities related to solar building design and engineering. The methodology established in this work presents a new, efficient, and accurate method based on readily available weather data documented in conventional weather files, enabling more comprehensive and precise building energy and performance-related analyses, especially with respect to building elements and products that have NIR selectivity features.

2. Methodology

2.1 Data collection

Two major datasets, meteorological measurements (MM, in the TMY3 data format) and outdoor solar spectra data (WISER), were selected from the Baseline Meteorological System (BMS) database of the National Renewable Energy Laboratory (NREL) Solar Radiation Research Laboratory (SRRL) for the modeling done in this study (Andreas and Stoffel 1981). The MM dataset was used to retrieve and process the independent variables, including GHI, DNI, DHI, cloud coverage, dry-bulb temperature, dew-point, relative humidity, and wind speed, while the key dependent variables (i.e., solar NIR irradiance) were calculated from the WISER dataset (Andreas and Stoffel 1981).

The MM dataset for 2018 and 2019 was used in this project. It describes the basic solar radiation and meteorological elements with hourly timestamps. Note that the average value of all measured points each hour is defined as the MM value for the timestamp at the end of the one-hour interval (SOLARGIS, 2020). For example, the value at timestamp 08:00 in the MM dataset equals the average value of all measurements taken from 07:00 to 08:00. This dataset is well-organized and has been used widely to simulate the solar radiation and building energy performance in the architecture, engineering, and construction industries.

The WISER measurement database is formed from two spectroradiometers (i.e., MS-711 and MS-712) that are combined to measure global horizontal spectral solar irradiance data (Andreas and Stoffel 1981). MS-711 spectroradiometer covers the measurement range from 300 nm to 1,100 nm and MS-712 spectroradiometer focuses on the NIR range from 900 nm to 1,700 nm (Andreas and Stoffel 1981). We selected data from the same period: 2018 and 2019. The WISER database has a higher resolution measurement for both wavelengths (0.41 nm and 1.6 nm resolutions for the MS-711 and MS-712, respectively) and time intervals (typically 5 minutes, but occasionally 1 minute). To coordinate these two solar datasets from different sources, the 5-minute interval data were processed using the statistical computing software R. The hourly spectrum data were calculated by averaging the 5-minute interval data for each hour, following the criterion of timestamp calculation regulated in the MM dataset. The day-of-year time format was also modified to fit the time format of UTC (Coordinated Universal Time), as it was the same format used in the MM.

2.2 Data processing

First, to obtain the solar NIR component, we summed the spectral data for the corresponding wavelength ranges of 781 nm to 1,650 nm for NIR. We obtained the fraction of NIR/GHI by using the NIR values calculated from the WISER dataset and GHI values calculated from the MM dataset. Second, to potentially enhance modeling accuracy, we generated two additional predictor variables including Clearness index K_t and Cloud

transmittance T_{cld}. The mathematical expressions of these two parameters are as Equations (1) and (2):

- Clearness index K_t

$$K_t = \frac{GHI}{I_0 \cos(\zeta)} \quad (1)$$

where GHI is the horizontal global irradiance, I_0 is extraterrestrial solar radiation on the horizontal surface, and ζ is the solar zenith angle.

- Cloud transmittance T_{cld}

We formed a new parametric cloud transmittance T_{cld} based on our understanding of the physical behavior of solar irradiance transmission. T_{cld}, defined as:

$$T_{cld} = \frac{(1-0.1T_{opq})(1-0.1T_{tot}+0.1T_{opq})}{1-0.05T_{tot}} = \frac{(1-0.1T_{opq})(1-0.1T_{trn})}{1-0.05T_{tot}} \quad (2)$$

where T_{opq} is the opaque sky cover transmittance, T_{tot} is the total sky cover transmittance, and T_{trn} is the translucent sky cover transmittance $T_{trn} = T_{tot} - T_{opq}$.

3. Results and Discussion

Classification and regression trees (CART) are a simple but powerful technique for modeling. In this study, we used the $rpart$ package of CART method in R software to build regression trees for NIR/GHI. We split the entire dataset D into a training dataset (90% of D) and a test dataset (10% of D). The $rpart$ implementation first fit a fully grown tree onto the training dataset with N terminal nodes. Then, it pruned the fully grown tree by k-fold cross-validation (default k =10).

3.1 CART results for the NIR/GHI fraction

1) Cross-validation error plot

Fig. 1 shows the cross-validation error plot for the NIR/GHI tree. From this figure, we can see that when cp = 0.01, the Size 10 regression tree has the minimum cross-validation error. This tree model is shown in **Fig. 2**.

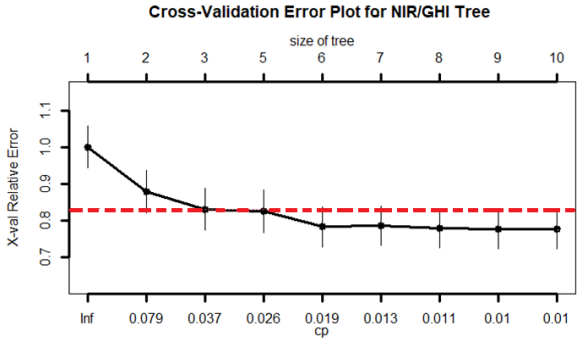

Fig. 1 Cross-validation error plot for the NIR/GHI tree.

(The red dotted line refers to the simplest tree, following the 1-SE rule).

2) Regression tree with minimum cross-validation error

The CART procedure generated a tree containing 10 terminal nodes for NIR/GHI (see **Fig. 2**), ranging from 0.1% to 33.5%. The first variable selected for splitting in this resulting tree was the clearness index K_t. If $K_t <$ 0.415, the group was further split according to $RH \geq 82\%$ or $RH < 82\%$. If $K_t < 0.415$ and $RH \geq 82\%$, the group was further split according to Dry -9.55℃ into two groups: the NIR/GHI values are 0.358 and 0.437. In another branch, if $K_t < 0.415$ and $RH < 82\%$, the group was further split according to the dewpoint temperature 15.3℃ and yielded three more groups in which the NIR/GHI values are 0.391, 0.418, and 0.577, respectively. In the other major branch of this regression tree, if $K_t \geq 0.415$, the parameters of dewpoint

temperature, total cloudiness, and relative humidity were used to further form the groups, including 0.414, 0.42, 0.439, 0.456, and 0.743 for NIR/GHI.

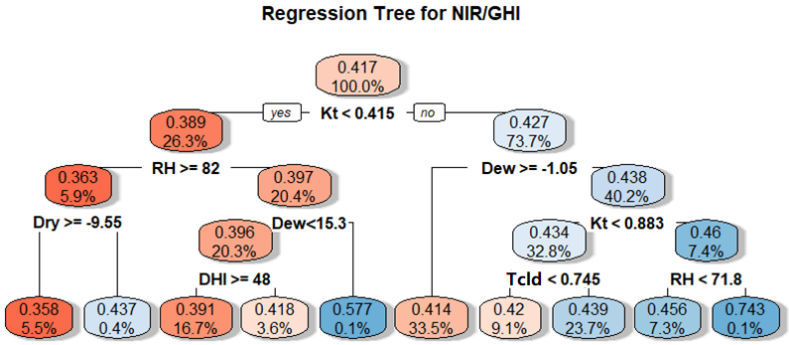

Fig. 2 Regression tree model for $\boldsymbol{NIR/GHI}$.

3) Regression tree with the 1-SE rule

The dashed red line in **Fig. 1** shows the position of the 1-SE rule with the minimum $xerror + xstd$; **Fig. 3** shows that the pruned tree using the 1-SE rule for NIR/GHI contained three terminal nodes. The percentage of NIR/GHI ranged from 26.3% to 40.2% in these three groups. The first variable selected for splitting was the clearness index K_t. If $K_t < 0.415$, no further split was observed for Group 1: 26.3% of NIR/GHI, with a mean value of 0.389. If $K_t \geq 0.415$, the group was further split according to $Dew \geq$ -1.05℃ (Group 2: 33.5% of NIR/GHI, with a mean value of 0.414) or $Dew <$ -1.05℃ (Group 3: 40.2% of NIR/GHI, with a mean value of 0.438).

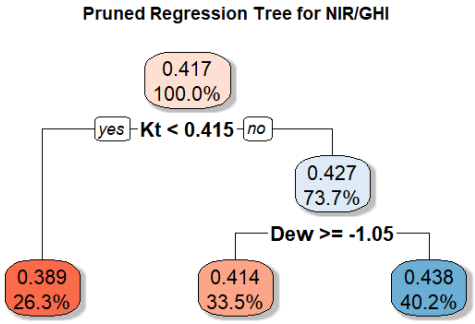

Fig. 3 Pruned regression tree model for NIR/GHI.

3.2 Estimation performance evaluation

The resultant tree models in **Figs. 2** and **3** are named Model 1 and Model 2, respectively. To further understand each model's estimated performance, we calculated the root mean squared error (RMSE) and the mean absolute error (MAE) of these two tree models on the test dataset with 758 observations. The $\hat{y}_j$ variable was the prediction.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2} \quad (3)$$

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - \hat{y}_j\right| \quad (4)$$

Table 1 Comparison of RMSE and MAE by Models

	NIR/GHI	
Regression Tree	**Model 1**	**Model 2**
Tree Size	10	3
RMSE	0.0391	0.0388
MAE	0.0213	0.0216

From **Table 1**, we can see that the RMSE decreased as the tree size decreased, but the MAE increased as the tree size decreased. Comparing Models 1 and 2, the RMSE decreased by 0.77% and the MAE increased by 1.4%. Regarding the changes in RMSE, since the errors were squared before they were averaged, larger errors receive a relatively higher weight. This means that the RMSE is more useful when significant errors are particularly undesirable. However, the RMSE did not necessarily increase with the variance of the errors. The RMSE increased with the variance of the frequency distribution of error magnitudes. Based on the information shown in **Table 1**, we can find the accuracy level differences among the models were negligible in this work. Both Models 1 and 2 had excellent prediction performances for NIR/GHI. This offers the opportunity to simplify the computation process if the weather data are insufficient.

4. Conclusion

This work demonstrated the feasibility and excellent prediction performance of regression tree models for hourly NIR/GHI. The solar spectra and conventional hourly weather data obtained from the BMS database of NREL's SRRL were utilized for model development. This research yielded models capable of converting the broadband solar irradiance data in weather files into NIR solar component, for building energy and performance-related studies in which independent solar spectra products are examined, such as analyses of spectrally selective glazing, transparent photovoltaic panels, etc. Solar components, especially NIR, are significantly affected by atmospheric parameters (e.g., water vapor levels), but these parameters are not very well documented observationally and dependent on local geographic and climatic features.

5. Acknowledgments

We acknowledge the financial supports provided by the National Science Foundation CMMI-1847024 and National Science Foundation CMMI-1953004.

6. References

Andreas, Afshin, and Tom Stoffel. 1981. *NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS); Golden, Colorado (Data); NREL Report No. DA-5500-56488*. Golden, Colorado (Data); NREL Report No. DA-5500-56488. http://dx.doi.org/10.5439/1052221 (May 20, 2020).

Song, Yuhui et al. 2020. "Solar Infrared Radiation towards Building Energy Efficiency: Measurement, Data, and Modeling." *Environmental Reviews*.

Eicker, Ursula. 2006. *Solar Technologies for Buildings*. John Wiley & Sons.

Kisilewicz, Tomasz. 2007. "Computer Simulation in Solar Architecture Design." *Architectural Engineering and Design Management* 3(2): 106–23.

Perlin, John. 2013. *Let It Shine: The 6,000-Year Story of Solar Energy*. New World Library.

Schittich, Christian. 2012. *Solar Architecture: Strategies, Visions, Concepts*. Walter de Gruyter.

Solargis. 2019. “Data Format.” https://solargis.com/docs/product-guides/time-series-and-tmy-data/data-format (May 20, 2020).

Wang, Julian, and Donglu Shi. 2017. “Spectral Selective and Photothermal Nano Structured Thin Films for Energy Efficient Windows.” *Applied Energy* (208): 83–96.