

## RESEARCH ARTICLE

# Representative identification of spectra and environments (RISE) using k-means

Erin E. Looney<sup>1</sup> | Zhe Liu<sup>1</sup> | Andrej Classen<sup>2</sup> | Haohui Liu<sup>3</sup> | Nicholas Riedel<sup>4</sup> |  
 Marília Braga<sup>5</sup> | Pradeep Balaji<sup>6</sup> | André Augusto<sup>6</sup> | Tonio Buonassisi<sup>1</sup> |  
 Ian Marius Peters<sup>2</sup>

<sup>1</sup>Department of Mechanical Engineering,  
 Massachusetts Institute of Technology,  
 77 Massachusetts Ave, Cambridge, MA,  
 02139, USA

<sup>2</sup>High Throughput Methods in Photovoltaics  
 Research Department, HI-ERN,  
 Forschungszentrum Jülich, Helmholtz Institute  
 Erlangen-Nürnberg for Renewable Energy,  
 Immerwahrstraße 2, Erlangen, 91058,  
 Germany

<sup>3</sup>Tandem Solar Cells Group, Solar Energy  
 Institute of Singapore, 7 Engineering Drive  
 1, 06-01 Block E3A, Singapore, 117574,  
 Singapore

<sup>4</sup>Department of Photonics Engineering,  
 Technical University of Denmark,  
 Frederiksborgvej 399, Building 130, Roskilde,  
 4000, Denmark

<sup>5</sup>Laboratório Fotovoltaica da UFSC,  
 Fotovoltaica-UFSC—Centro de Pesquisa e  
 Capacitação em Energia Solar da UFSC, Av.  
 Luiz Boiteux Piazza, 1302, Lotes 114/115,  
 Sapiens Parque, Florianópolis, Brazil

<sup>6</sup>School of Electrical, Computing and Energy  
 Engineering, Arizona State University,  
 TempeAZ, 85281, USA

## Correspondence

Ian Marius Peters, HI-ERN,  
 Forschungszentrum Jülich, Helmholtz Institute  
 Erlangen-Nürnberg for Renewable Energy,  
 Immerwahrstraße 2, 91058 Erlangen,  
 Germany.  
 Email: i.peters@fz-juelich.de

## Funding information

U.S. Department of Energy, Grant/Award  
 Number: DE-EE0007535; National Science  
 Foundation; MISTI, Grant/Award Numbers:  
 EEC-1041895, DE-EE0008975; Bavarian State  
 Government, Grant/Award Number:  
 44-6521a/20/5

## Abstract

Spectral differences affect solar cell performance, an effect that is especially visible when comparing different solar cell technologies. To reproduce the impact of varying spectra on solar cell performance in the lab, a unique classification of spectra is needed, which is currently missing in literature. The most commonly used classification, average photon energy (APE), is not unique, and a single APE value may represent various spectra depending on location. In this work, we propose a classification method based on an iterative use of the k-means clustering algorithm. We call this method RISE (Representative Identification of Spectra and the Environment). We define a set of 18 spectra using RISE and reproduce the spectral impact on energy yield for various solar cell technologies and locations. We explore effects on yield for commercially available solar cell technologies (Si and CdTe) in four locations: Singapore (fully humid equatorial climate), Colorado (cold arid), Brazil (warm, humid, and subtropical), and Denmark (fully humid warm temperature). We then reduce our findings to practice by implementing the spectrum set into an LED current-voltage (IV) tester. We verify our performance predictions using our set of representative spectra to reproduce energy yield differences between Si solar cells and CdTe solar cells with an average error of less than  $1.5 \pm 0.5\%$  as compared to over 5% when using standard testing conditions.

## KEYWORDS

energy yield, machine learning, photovoltaics, solar spectra classification

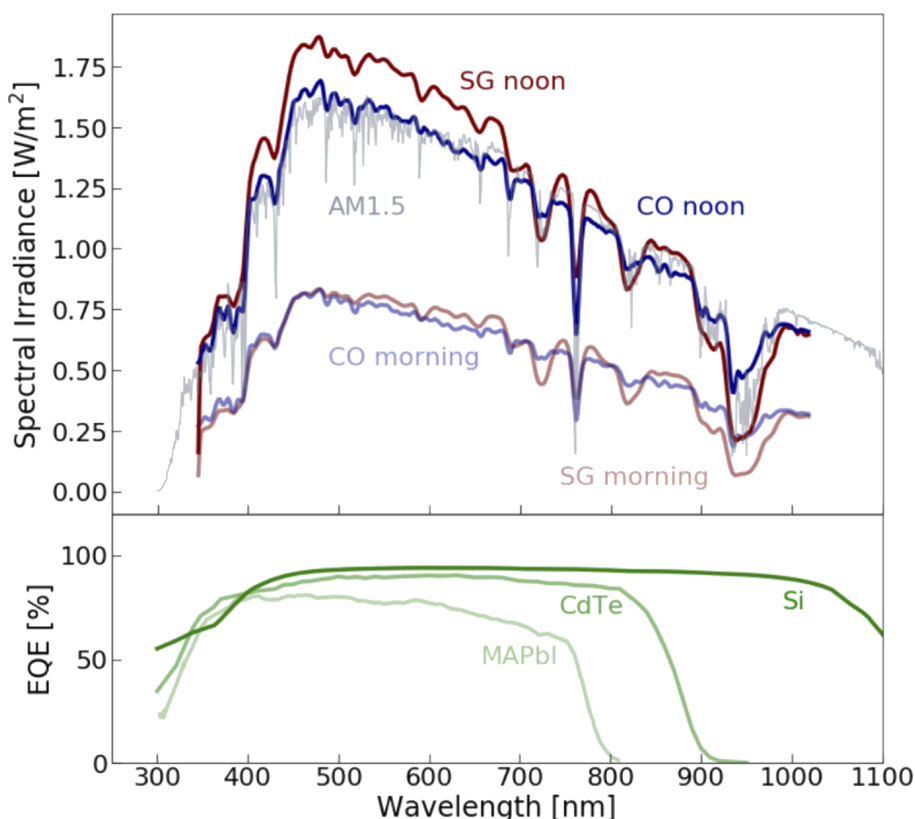
## 1 | INTRODUCTION

The performance of photovoltaic (PV) cells in outdoor operating conditions is affected by several environmental parameters. Two key parameters are solar irradiance and temperature, which have been studied in depth.<sup>1–3</sup> Irradiance and temperature have been taken into account in solar testing in a variety of ways. For example, linear temperature coefficients can be used to account for temperature changes, and the number of sun hours per location or tabulated intensity in  $\text{W}/\text{m}^2$  over time is used for varying intensities. A somewhat less studied effect is the variation in spectral composition of the available sunlight and its impact on PV performance.<sup>4,5</sup> Spectral composition differs depending on time and location due to various environmental factors, including humidity, cloud cover, the surrounding albedo, and aerosols.<sup>6</sup> An example of the different spectral shapes can be seen in Figure 1. Spectral variations also impact the performance of solar cells, an effect that is especially noticeable if solar cell technologies with different band gaps are compared.<sup>7</sup> These performance effects on different solar cell technologies can be inferred from Figure 1, where the external quantum efficiencies (EQEs) are mapped against various spectral shapes. The ratio of available to total solar radiation varies seasonally by about 5%.<sup>8</sup> These variations have been shown to account for annual spectral impact of up to 3.4% when compared to the standard spectra used in testing (AM 1.5).<sup>9</sup> The amount of impact spectra have on performance is highly dependent on the type of technology. For example, larger band gap cells are affected more than smaller band gap cells.<sup>10</sup> Yet many tools for PV energy yield (EY) predictions still neglect spectral effects. Moreover, while there

are many proposed methods for capturing the spectral characteristics of a solar cell such as average photon energy (APE), mismatch factor, and other device independent parameters, there is still no standard method to reliably model or test for these spectral impacts.

The PV industry is always looking to improve EY predictions and to maximize return on investment for solar developers. Several industry-developed methods now exist to account for solar cell performance differences due to spectral variations. For example, First Solar is promoting the spectral characteristics of CdTe solar cells, which are less sensitive to light absorption of precipitable water than silicon due to a larger band gap. Lower spectral sensitivity gives CdTe a relative performance advantage in hot and humid climates<sup>7,15</sup> when compared with other commercial solar cell technologies. Similar to the temperature coefficient, First Solar created a spectral correction factor in 2016, taking precipitable water and air mass into account for module performance predictions.<sup>16</sup> This calculation can be used publicly in First Solar's PlantPredict tool,<sup>17</sup> and a similar method was adopted by PVSyst<sup>18</sup> in October 2018. With the advent of LED-based IV-testers, it is now possible to establish spectral characteristics of different solar cell technologies in smaller labs also. One missing aspect here is a unique and unambiguous classification of solar spectra that is independent of time and location. Spectrum classification greatly simplifies the definition of a small and representative set of spectra for testing and comparing spectral conditions.

The most commonly used spectrum classification is the APE, which is calculated by integrating over the energy of all photons in the spectrum and dividing by the photon flux. APE is typically calculated in a certain wavelength range<sup>19,20</sup> as seen in Equations (1) and (2).



**FIGURE 1** Several spectra from different locations (humid Singapore and drier Colorado), and different times of day (morning and midday) demonstrate the various shapes that the spectra can take. EQE of several technologies are overlaid on the spectra to show how there are differences in how well these materials absorb and therefore how the different spectra can impact these technologies.<sup>11–14</sup> [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

$$APE = \frac{\int_{\lambda_1}^{\lambda_2} I(\lambda) d\lambda}{\int_{\lambda_1}^{\lambda_2} \phi(\lambda) d\lambda}, \quad (1)$$

$$\int_{\lambda_1}^{\lambda_2} \phi(\lambda) d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{I(\lambda)}{(hc/\lambda)} d\lambda, \quad (2)$$

where  $I(\lambda)$  is the intensity and  $\phi(\lambda)$  is the photon flux, both as a function of wavelength  $\lambda$ ,  $h$  is Planck's constant, and  $c$  is the speed of light. APE describes red and blue shifts for a given location well but is insensitive to certain other types of variation. Relevant for this work is that APE is not unique, having an approximate uniqueness only in a limited wavelength range of approximately 450 to 900 nm.<sup>9,21</sup> Furthermore, a single APE value represents spectra of different shapes in different locations. Nevertheless, APE can be used for EY calculations if these limitations are considered, as was demonstrated, for example, by Liu et al. in 2016.<sup>22</sup> An alternative to using APE is to model the spectrum directly. The most commonly used tool for this is the Simple Model of the Atmospheric Radiative Transfer of Sunshine (SMARTS).<sup>23–25</sup> With this tool, full spectra can be simulated if the relevant meteorological parameters are known. This approach can be used to model PV EY, as was shown on numerous occasions.<sup>26–28</sup> Spectrum simulations, however, do not resolve the issue of a missing spectrum classification or answer the question of how to pick a representative set for spectral characterization of solar cells in the lab.

In this paper, we propose a unique method for classification of measured outdoor spectra to overcome the limitations of the described alternative classifications. Our method utilizes an iterative application of the k-means clustering algorithm. We use a two-step process to find representative spectra from measured data for accelerated solar testing and EY prediction for any PV technology. First, we describe the method and then define a representative set of 18 spectra out of millions of measured spectra that describe the spectral conditions of four locations in different climate zones. We then project the EY of different solar cell technologies based on these 18 spectra to compare with the “ground truth,” which is found using the full data set. To conclude, we reduce our method to practice by implementing the spectra into an LED IV-tester and compare our theoretical predictions with measured results.

## 2 | METHOD

### 2.1 | Spectrum classification

The Representative Identification of Spectra and the Environment (RISE) uses a clustering algorithm called k-means<sup>29</sup> from sci-kit-learn<sup>30</sup> in the open-source programming language of Python. The k-means algorithm takes measured data points from all the spectra and optimizes over the position of the means using a distance metric. In this work, the distance metric used is Euclidean distance, and different numbers of means (k-means parameters) are explored. Ripalda et al. proposed a spectra-clustering routine with similar features,<sup>31</sup> focusing on tandem solar cell applications.

The first step of the RISE method shown in Figure 2 is aligning all the meteorological and spectral data to have the same time steps and wavelength intervals. Step 2 is the clustering using k-means. The clustering is done in two parts. In Step 2a (k1-clustering), the raw spectral data are clustered. The leading effect in the resulting classification is a sorting by intensity. Due to the prevalence of spectra with low intensity, this classification will produce many bins with low intensities and few with high ones. As the latter are mainly contributing to yield, we have introduced an additional weighting. To implement this weighting, the k1-clustering step was modified so that the spectra are first binned by irradiance with half as many spectra allotted to each successively higher irradiance bin (k1-clustering-solar). Next, in Step 2b (k2-clustering), the spectra are normalized by total irradiance within each k1-clustering bin and reclustered. The leading effect in the second clustering step is a sorting by redshift and blueshift. For example, for the three k2 means of one k1 irradiance bin, the algorithm typically finds the average spectrum, one that is redshifted and one that is blueshifted. The average, redshifted, and blueshifted spectrum can be seen visually in the different shapes of the spectra within each irradiance cluster in Figure 4. Lastly, temperatures corresponding to the measured spectra are binned using the existing spectra clusters.

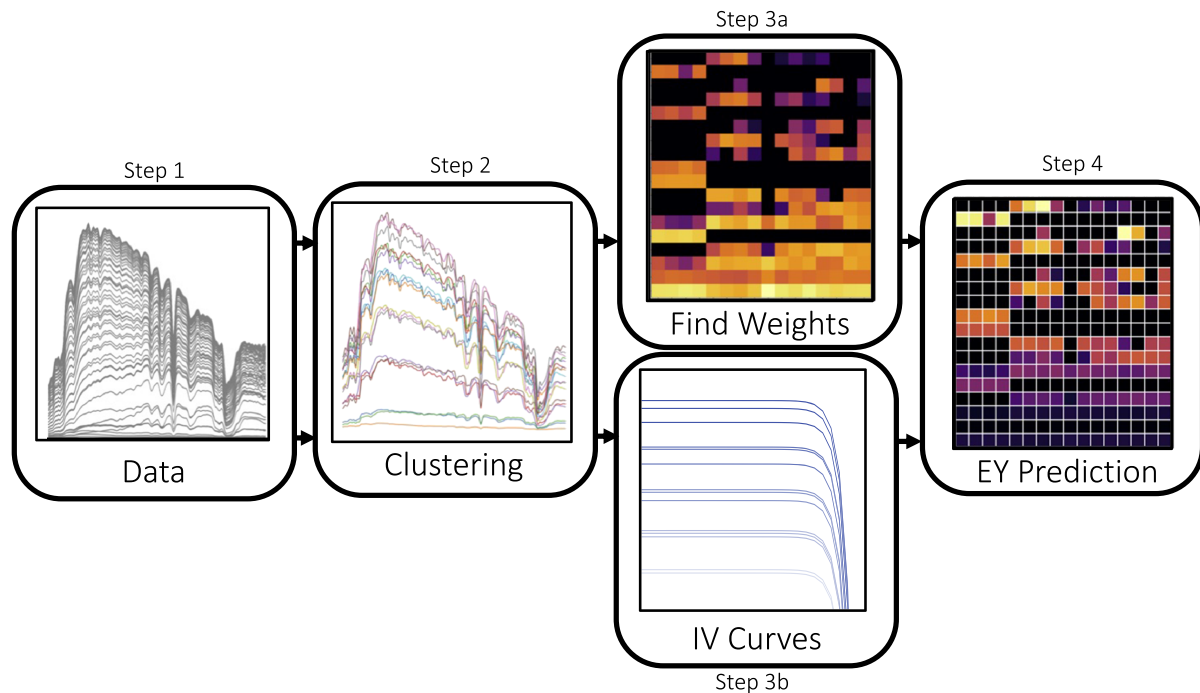
Step 3 also has two parts. In Step 3a, we find weights for each representative spectra. The weights are the number of spectra from the full data sets that are binned in that cluster. Then, the spectra and temperature clusters found in Step 2 are used to model PV device IV curves. Step 4 then uses these IV curves to predict EY and weighs these predictions by the weights for different locations and times to predict differences between EY for different technologies in, for example, Singaporean winter and summer in Colorado.

The workflow of this method is shown in Figure 2.

### 2.2 | Used spectrum data

To find a representative set, the used spectrum data should cover as many situations as possible. To test our method, we used four spectral data sets for four different climate zones worldwide: cold arid (BSk, Golden, Colorado), fully humid equatorial (Af, Singapore), humid subtropical oceanic climate with hot summer, (Cfa, Santa Catarina, Brazil), and fully humid warm temperate (Cfb, Denmark). Each data set covers 1 year and had a temporal resolution of one spectrum per 30 min or better. Comprehensive data sets of measured spectra worldwide for PVs are still scarce, and most research is done with simulated spectra. We believe that this data set is one of the most comprehensive in the scientific literature to date. This is one of the major differences between this work and the spectral clustering work done by Ripalda et al.<sup>31</sup>

The cold arid (BSk) data set was measured by the National Renewable Energy Laboratory (NREL) in Golden, Colorado. Spectra from the entire year of 2018 were measured every 5 min by an EKO WISER system with an MS-711 spectroradiometer that is ISO 17025 accredited. The wavelength interval is 0.73 nm that has been interpolated to 1 nm, and the wavelength range is 290 to 1650 nm.<sup>32</sup>



**FIGURE 2** Flow chart of the RISE method used in this work. The majority of this paper focuses on the first three steps: data alignment, k1 and k2 clustering, and weight finding. EY prediction for photovoltaics (Step 4) is given as an example of a potential application of the RISE method [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

The fully humid equatorial (Af) data set was measured by the Solar Energy Research Institute of Singapore (SERIS) in Singapore. Spectra from the entire year of 2018 were measured every minute. The wavelength interval is around 3.3 nm, and the wavelength range is around 303 to 1145 nm.

The humid subtropical oceanic climate without dry season and with hot summer (Cfa) data set was measured by the Solar Energy Research Laboratory at the Universidade Federal de Santa Catarina in Florianópolis, Brazil. Spectra from the entire year of 2018 were measured every minute by an EKO WISER system with an MS-711 spectroradiometer. The wavelength interval is around 0.43 nm, and the wavelength range is around 285 to 1120 nm.

The fully humid warm temperate (Cfb) data set was measured by the Technical University of Denmark's Department of Photonics Engineering. The spectra used in this work are from 2017, except a few months that were missing. The missing months are filled in by data from 2018, including 1 January through 14 March and 6 June to 7 July. Spectra were measured every 30 min with a wavelength interval of around 0.43 nm, and the wavelength range is around 282 to 1119 nm. The instruments used include an EKO MS-711 (direct normal irradiance) mounted on a dual axis tracker and an MS-711 that is continuously shaded by a 5° field of view shadow ball. The global horizontal spectrum is thus calculated by the summation of the diffuse and cosine-corrected beam spectral components as described in Riedel.<sup>33</sup>

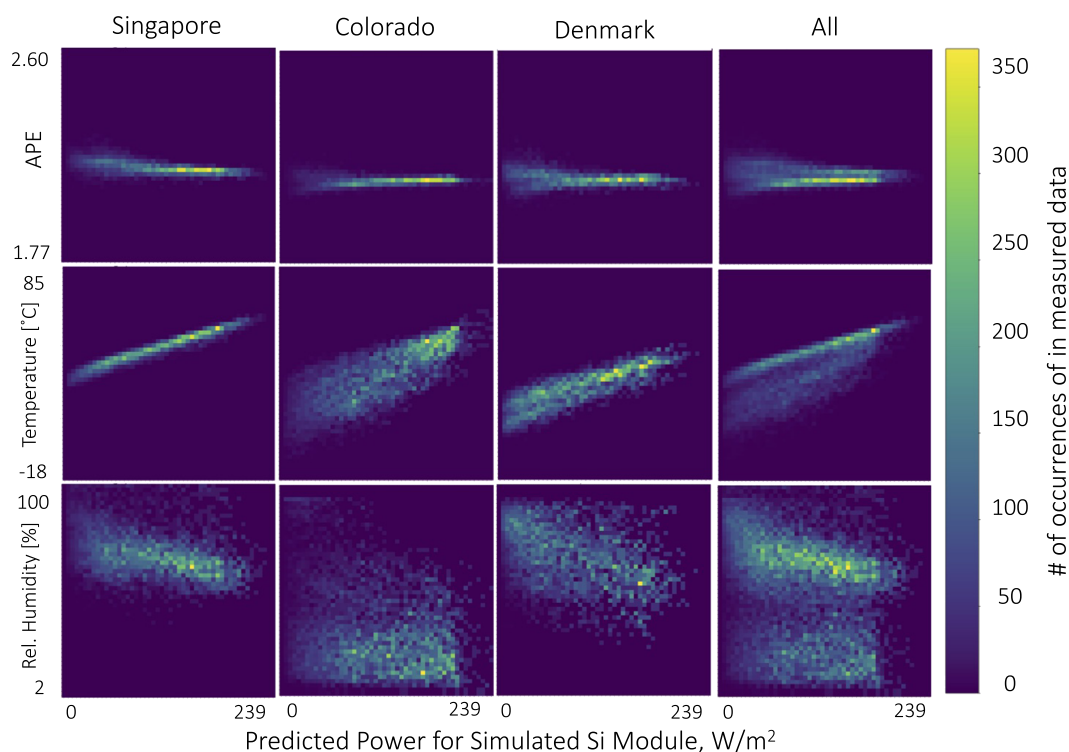
To perform the classification method described in this paper, the data sets need to be matched in wavelength interval and range. The interval matching is done through linear interpolation using an interval that is approximately the average of the intervals of all the data: 1.5

nm. The data used in this work are shown in Figure 3. To further generalize this method, more spectra from more locations should be used. We hope that these will become available in the future.

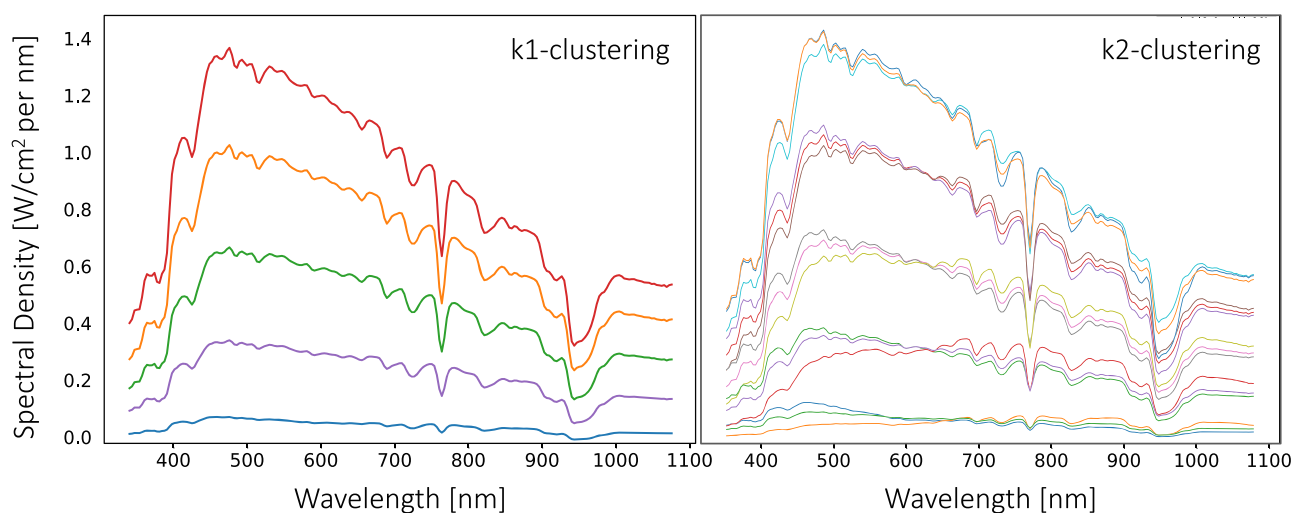
### 3 | SPECTRUM SET AND FEATURES

The RISE method with a nonweighted Euclidean distance was first tried on a data set from Colorado. In Figure 4, k1-clustering and k2-clustering are shown for five k1-cluster and three k2-cluster centers. The k1-clustering step creates clusters according to irradiance, which is plausible because irradiance is the dominant feature in the spectral functions. The k2-step clusters according to slope differences. This is also plausible as spectra will have different slopes according to how much light of certain wavelengths are transmitted or absorbed by the atmosphere. As a consequence, k2-clustering sorts by a physically interpretable feature—a blueshifted spectrum (spectrum that has relatively more blue light) which has a more negative slope than a redshifted spectrum. The blueshift and redshift of the k2 clusters are seen clearly in Figure S1. Some of the features shown here have also been discussed in the work of Ripalda et al.<sup>31</sup> Our findings largely agree with results there, even considering differences in the used data sets (we used measured spectra, whereas in Ripalda et al.,<sup>31</sup> exclusively simulated spectra are considered). Discussion also differs as the focus of Ripalda et al.<sup>31</sup> centers on the customization of band gaps for advanced solar cell technologies, whereas we concentrate on a method for EY prediction and measurement for commercially available single junction solar cells.





**FIGURE 3** Data from three of the climate zones (Singapore, Colorado, Denmark) used in this work showing average photon energy, temperature, and relative humidity versus predicted power for simulated silicon module [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

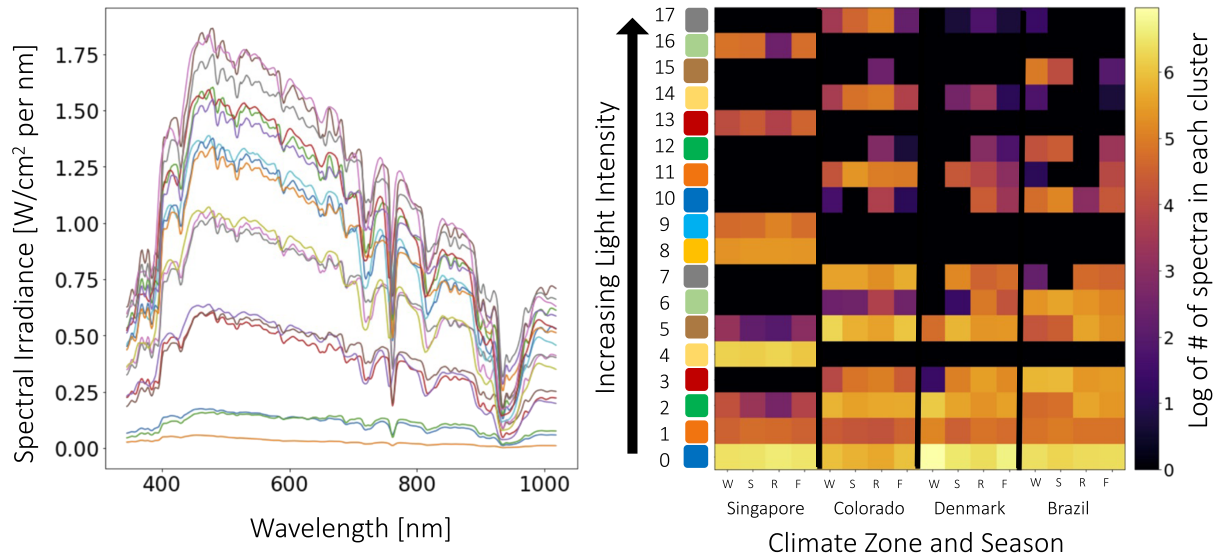


**FIGURE 4** k1-clustering and k2-clustering steps in RISE method [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

The ultimate goal of the RISE method is to find one set of representative spectra that can uniquely describe conditions anywhere in the world. With the available data set, we can find a representative set that describes conditions in the four climate zones used in this clustering. All four data sets from Singapore, Colorado, Denmark, and Brazil are aligned into one data set and clustered according to the procedure outlined in Figure 2. The results are shown in Figure 5. This clustering routine is done for k1-clustering parameter  $k1 = 6$  (six means used in the first clustering) and k2-clustering parameter  $k2 = 3$

(three means used in the second clustering). To visualize how these clusters capture the climatic conditions for different locations and times of year, a heat map or “fingerprint” is constructed as seen in Figure 5.

One feature of this binning method shown in Figure 5 is segmentation of clusters by location. In this segmentation, we expect to see the humid Singapore spectra separate from a dry climate like Colorado due to a shifted spectrum. For example, this expectation proves valid for Bins 9–11 (light blue, dark blue, and orange). Bin 9 has a majority



**FIGURE 5** Left: spectra clusters for  $k_1 = 6$ ,  $k_2 = 3$ . Right: fingerprint of these clusters and the location and seasons that they fall into. Intensity and seasonal patterns emerge. Note that each of the bin numbers 0 through 17 correspond to one of the spectra at left. Here, bin 0 is highlighted in orange and bin 16 is highlighted in pink [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

of spectra from Singapore, and as can be seen, it has lower irradiance values after 700 nm in the water absorption bands when compared to Bins 10 and 11. This segmentation demonstrates the differences that can arise between spectra in different locations both in intensity and shape.

This heat map can be used by weighting the performance prediction found for each representative condition by the number of spectra that are placed into that bin. For example, when calculating EY for Colorado, Bin 7 would be given a higher weight than Bins 6 or 8 as this shape is more likely to occur in Colorado. In this way, these weights will become important when calculating EY of a specific PV technology.

## 4 | APPLICATION TO PV TECHNOLOGIES

### 4.1 | PV modeling

The RISE method can be used for research, development, and testing of PV cells and modules. To use RISE for PV cells, we must consider more than just spectra as modeling inputs and will include temperature data. We use meteorological data that correspond to the spectral data in each cluster, resulting in a range of temperatures for each bin using parameter set  $k_1 = 6$ ,  $k_2 = 3$  as seen in Figure 6 with the 18 bins corresponding to those from Figure 5.

Like with the spectra, patterns emerge in the temperature binning, with higher temperatures corresponding to higher intensity spectra. Also, there is a segmentation of the temperature by location, for example, with differences between Clusters 6, 7, and 8. Also, at lower intensities, there is the largest variation of temperature, which is plausible as the most data points for morning, evening, and cloudy days of all seasons are within these bins.

The temperature used in EY calculations needs to be module temperature rather than ambient temperature. Therefore, module temperatures have been calculated from the ambient temperature using the following equation.<sup>34</sup> In this case,  $k$  is the Ross coefficient, which is between 0.02 and 0.04°C m<sup>2</sup>/W. In this work, we use the original Ross coefficient of  $k = 0.03$ .<sup>35</sup>

$$T_{\text{module}} [^\circ\text{C}] = T_{\text{ambient}} + k * \int (Irr(\lambda)). \quad (3)$$

There are several ways temperature information can be used in EY calculations. The simplest is just using the average of each temperature cluster. This was implemented as a first approximation.

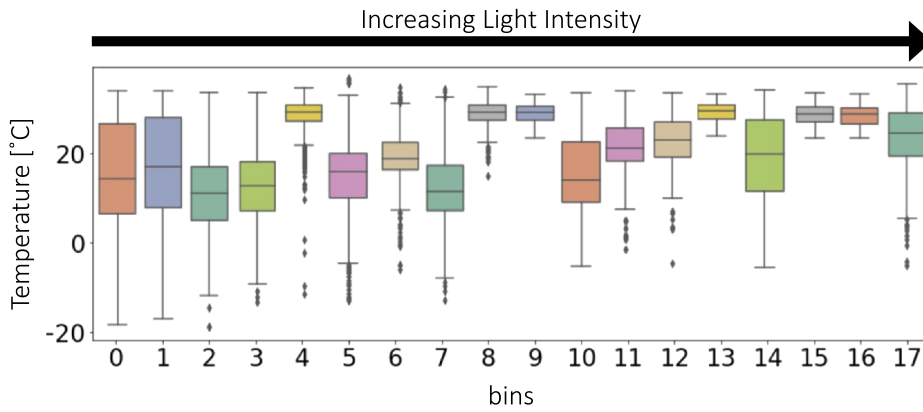
Now that we have the input spectrum and temperature sets, we model solar cell performance using a one-diode model shown below in Equation (4).

$$I = I_{ph} - I_0 (e^{-q(V + IR_s)/(nk_b T)} - 1) - (V + IR_s)/R_{sh}, \quad (4)$$

with  $I_{ph}$  as the photocurrent found using Equation (5),  $I_0$  as the dark saturation current found using Equation (7),  $q$  as the elementary charge ( $1.602 \times 10^{-19}$  C),  $V$  as voltage,  $I$  as current,  $R_s$  as series resistance,  $n$  as ideality factor,  $k_b$  as Boltzmann constant ( $1.380 \times 10^{-23}$  J/K),  $T$  as temperature, and  $R_{sh}$  as shunt resistance.

$$I_{ph} = \sum ((EQE(\lambda) * Irr(\lambda)) / E_{ph}(\lambda)), \quad (5)$$

with  $EQE$  as external quantum efficiency resolved by wavelength,  $\lambda$ ,  $Irr$  is the irradiance falling on the cell resolved by wavelength, and  $E_{ph}$  as the photon energy resolved by wavelength as found in Equation (6).



**FIGURE 6** Ambient temperatures that correspond to spectra bins [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

$$E_{ph} = (hc * 10^9) / \lambda_{EQE}, \quad (6)$$

with  $h$  as Planck's constant ( $6.626 * 10^{-34} Js$ ),  $c$  as the speed of light ( $3 * 10^8 m/s$ ), and  $\lambda_{EQE}$  as the wavelength.

$$I_0 = I_{0,STC} (T/T_{STC})^3 * e^{((qE_g/nk_b T) * (1/T_{STC} - 1/T))}, \quad (7)$$

with  $I_{0,STC}$  as the dark saturation current at standard testing conditions (STC),  $T$  as temperature,  $T_{STC}$  as the temperature at STC,  $q$  as the elementary charge ( $1.602 * 10^{-19} C$ ),  $E_g$  as the band gap of the material,  $n$  as the ideality factor, and  $k_b$  as Boltzmann constant ( $1.380 * 10^{-23} J/K$ ).

From these models in Equations (4)–(7), current–voltage (IV) curves are found with weather and spectral data of the representative clusters. These IV curves are used to determine power for both Si and CdTe solar cells. Then, a temperature coefficient is applied to the power output for each technology. A temperature coefficient of  $-0.41\%$  per degree Celsius<sup>36</sup> for silicon and  $-0.34\%$  per degree Celsius<sup>37</sup> for cadmium telluride were used. With the temperature coefficients taken into account, these new powers are used to calculate efficiency and then EY for each of the representative spectra using the following equation:

$$EY [Wh/m^2] = \sum_{i=1}^N \eta(\Delta n, T, P(\lambda), \dots) P_{in} t, \quad (8)$$

with  $\eta$  as efficiency of the solar cell, which is dependent on minority charge carrier injection,  $\Delta n$ , (number of electrons or holes injected due to excitation of electrons from photons hitting the material), temperature,  $T$ , power per wavelength,  $P(\lambda)$ , and more.  $P_{in}$  is the total power into the cell in  $W/m^2$ , which can be found by integrating  $Irr(\lambda)$ , where  $t$  is time in hours and  $N$  is total number of measured spectra evaluated over.

The Si and CdTe cells were modeled to have the same efficiency for direct comparison. A list of the parameters for these models is in the Table 1.

## 4.2 | EY predictions

To find a generalized EY predication for a specific technology, we use the EY calculated for each of the representative spectrum

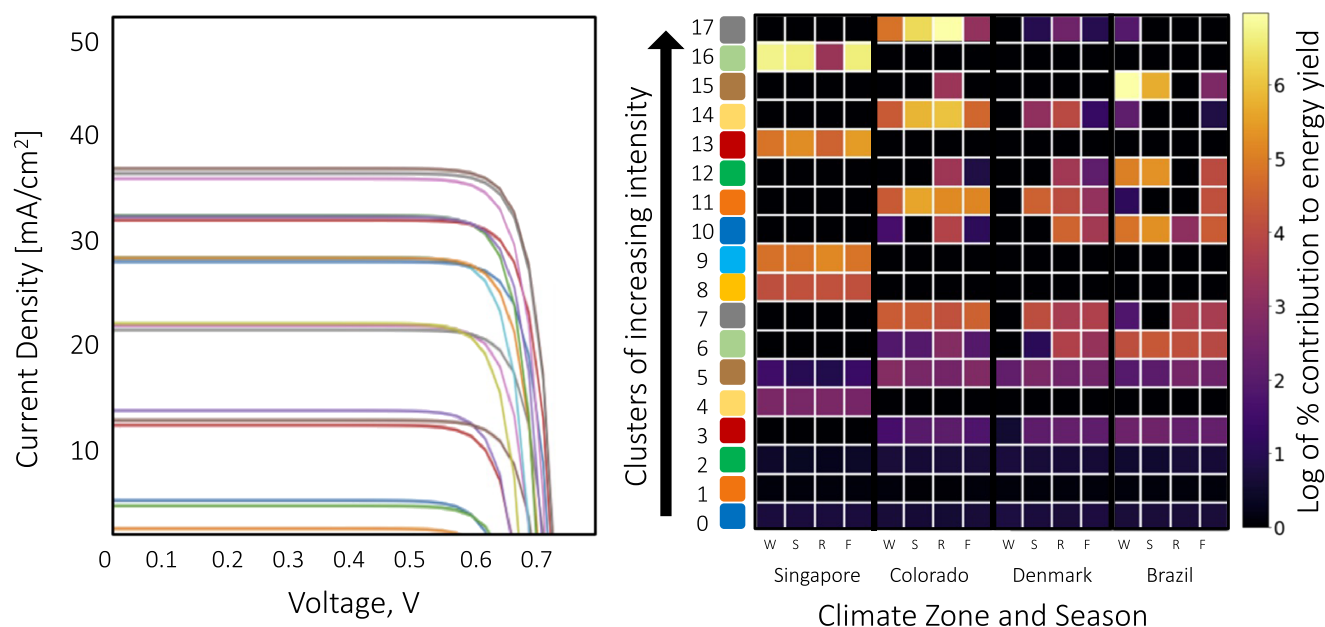
**TABLE 1** Modeling parameters and STC results

Parameter	Silicon	CdTe
Efficiency, %	18.2	18.2
Fill factor, %	80.6	81.3
Open circuit voltage, V	0.565	0.815
Short circuit current, mA/cm <sup>2</sup>	40.1	27.6
Temperature coefficient, % $\eta$ lost per degree K	−0.41	−0.32
Series resistance, $\Omega$ cm <sup>2</sup>	1.3	1.215
Shunt resistance, $\Omega$ cm <sup>2</sup>	500	500
Band gap	1.12	1.5

above and weight them using the weight matrices (shown by the heat maps) to calculate a weighted average. For example, in Figure 5, using the column for summer months in Singapore, we take a weighted average to find the total EY for the summer season. The formula for this weighted average is given in (9) with the number of spectra ( $w_i$ ) in each square multiplied by the EY of that cluster ( $P_{out}$ ) and time ( $t$ ) and summed over all clusters. This can be done for any unit of time, and in this work, we use the common unit of  $kWh/m^2$  per year.

$$EY [kWh/m^2] = \sum_{i=1}^N w_i P_{out} t. \quad (9)$$

We calculated the EY for all four locations, using the clustering steps with parameters  $k1=6$ ,  $k2=3$  to find representative conditions, model PV devices, create IV curves for each scenario, and develop the EY heat map seen in Figure 7. The temperatures input into the model are averages of the temperatures within each individual cluster. With  $k1$ -clustering implemented favoring cluster centers at higher irradiances, there is less of a discrepancy between lower and higher intensity spectra in the heat map. The segregation of the locations and distinctions between seasons are again clearly evident. Figure 7 demonstrates the ultimate goal of the RISE method: a measurement procedure that enables EY prediction in any location (encompassed by the input data) with a limited number of measurements.



**FIGURE 7** Left: IV curves for all climate zones using average temperature within bins. Right: fingerprint of the EY contributions of each climate zone and season [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

## 5 | EVALUATION OF METHOD FOR PV TESTING AND EY PREDICTIONS

### 5.1 | PV modeling results

We modeled the difference between CdTe and Si solar cells in different environments, Denmark, Colorado, and Singapore, showing that this representative set of 18 RISE IV curves accurately describes over half a million calculated IV curves (“ground truth”) effectively with an average of less than 1 kWh/m<sup>2</sup> error, or  $0.52 \pm 0.2\%$  when compared to ground truth. This is the accuracy of the method when no temperature coefficient is introduced. The average difference between RISE EY predictions and ground truth when temperature coefficients are taken into account is less than 5.5 kWh/m<sup>2</sup> or below  $1.5 \pm 0.5\%$  of overall EY. When temperature coefficients are introduced, the errors increase due to the use of only one average temperature for each of the 18 binned RISE spectra as compared to the tens of thousands of temperatures used for the ground truth calculation. This error can be further reduced by using several temperatures for each spectrum rather than just the average. The RISE method is compared to STC results, which are calculated by using the simulated STC power multiplied by the number of cumulative sun hours in the year calculated from the measurements. The RISE errors are all much less than that of EY predictions using STC, which has an average error of over 17 kWh/m<sup>2</sup> per year, translating to over 5% error in overall EY as seen below in Figure 8. Furthermore, STC predicts the same EY for both CdTe and Si, whereas the RISE predictions capture the different effects of spectra and temperature on the different technologies.

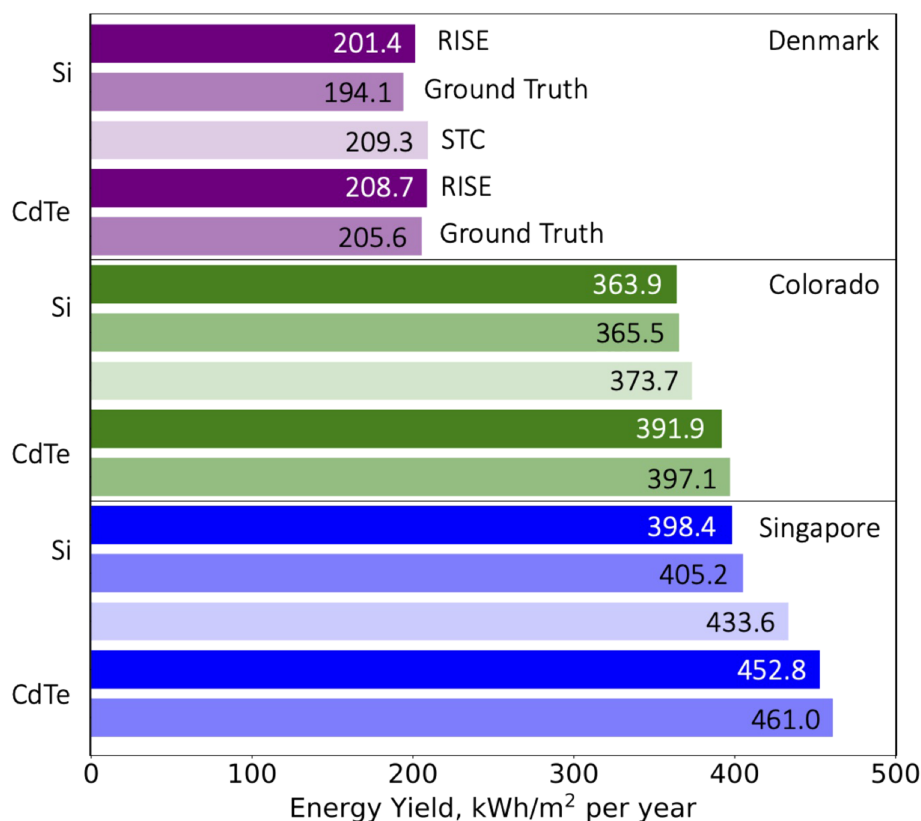
We also find that, according to expectations, the RISE method captures that the largest differences in performance between CdTe and Si technologies occur in Singapore, smaller differences in

Colorado, and no clear trend in Denmark. It should be noted that the temperature model assumptions do not include radiative cooling giving a general overestimate of module temperature, which is advantageous for CdTe. Several studies have compared CdTe and Si around the world. Hedegus et al. reviewed EY differences in literature between C-Si and thin film technologies including CdTe. The results in Figure 8 are in line with the reviewed literature.<sup>26,38–40</sup>

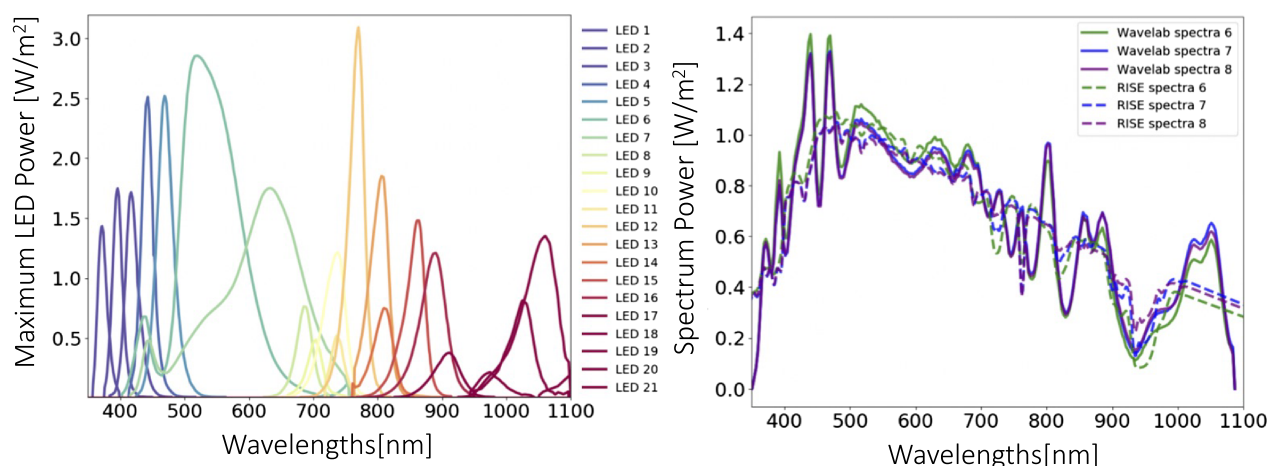
### 5.2 | Experimental verification

To verify that our approach can be used for indoor EY estimates, the 18 representative spectra from Figure 7 were programmed manually into a Wavelab Sinus 70 LED solar simulator. The Wavelab instrument has 21 different LEDs, the powers of which can be controlled to reproduce a certain solar spectra. The distribution of the LEDs over the spectral range is shown in Figure 9 alongside a fit of these LEDs to the representative spectrum of Bins 6, 7, and 8. To fit the spectra appropriately, the settings of each LED from the AM1.5G default spectra provided by Wavelab were used to calculate the appropriate settings for the 18 representative spectra. To do this, the power settings for the AM1.5G spectra were scaled given the ratio between the AM1.5G spectra and the RISE spectra that we were fitting. The fit of the LEDs to the spectra is much better than widely used Xenon or Halogen lamps and allows for fine adjustment to closely approximate a variety of spectra.

Two types of PV materials were tested: Si and CdTe. The Si cell used was a reference cell calibrated by CalLab at the Fraunhofer Institute for Solar Energy (FhISE), and the CdTe cells were fabricated at First Solar, Inc. To test each cell, electrical contact was made using an aluminum alloy wire and silver paste. Up to three IV curves were



**FIGURE 8** In this figure, we show the differences in performance of CdTe and Si cells due to spectral and temperature differences for different climate zones. The “ground truth” is the EY found when using every spectra and temperature measurement, and the RISE method is using only the 18 representative measurements and averaged temperatures [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 9** Left: 21 LEDs used by Wavelab Sinus 70. Right: fit of LED lights versus a desired representative spectra from RISE method [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

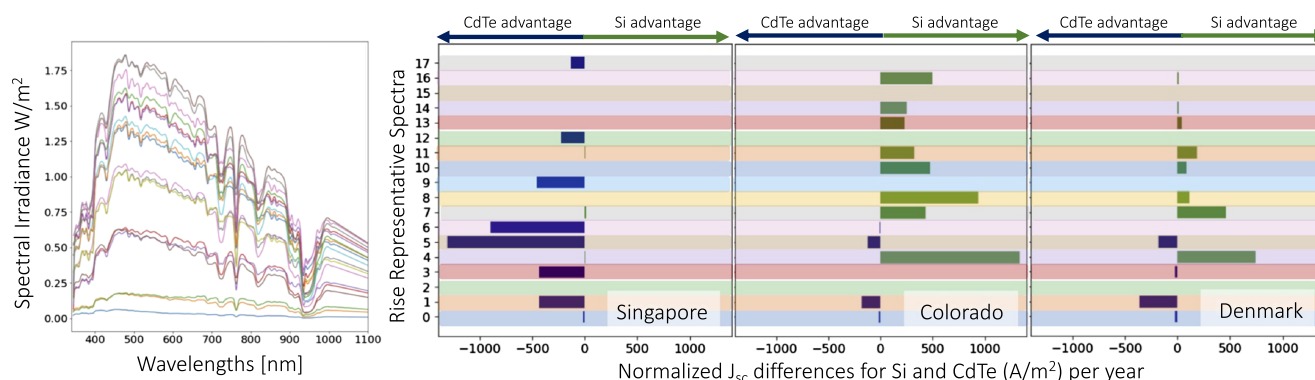
measured for each spectra and for each solar cell technology. The different shapes of the spectra, shown in Figure 9, are captured visibly by the LED simulator.

From these IV measurements, we extract the short circuit current ( $J_{sc}$ ).  $J_{sc}$  is the parameter of interest as it captures the response of different solar technologies to variations in spectrum. Different spectra have a different spectral photon flux that, combined with variations in quantum efficiency, results in differences in  $J_{sc}$ . In Figure 10, a comparison of  $J_{sc}$  normalized by STC AM1.5G for CdTe and Si technologies is shown for each of the 18 used spectra. In this figure, if the

number on the x axis is positive, then Si has a higher normalized  $J_{sc}$  for this particular spectrum than CdTe and vice versa. The performance advantages of CdTe in Singapore become apparent also from this figure; negative values are dominant over all spectra. For Denmark, there is more of a split between CdTe and Si  $J_{sc}$ s, matching the modeling results with no temperature coefficient.

It can also be seen that Si is shown to be the best in Colorado when only spectral effects through  $J_{sc}$  are taken into account, which can be understood due to its relatively dry climate. However, these  $J_{sc}$  differences do not exactly correspond to the power out of the cells as





**FIGURE 10** Left: RISE representative spectra, Right: normalized short circuit current. The colors correspond to the spectra at left so spectra shapes can be matched by eye to advantages in certain locations.  $J_{sc}$  differences between Si and CdTe [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

temperature effects on voltage must also be taken into account. We can expect that in a hotter place like Singapore, the CdTe advantage would be even greater, and for places like Colorado, with many days where module temperatures are above STC of 25°C, CdTe would also have a voltage advantage over Si while still having a  $J_{sc}$  disadvantage. For a climate like in Denmark with relatively colder temperatures, Si voltages would not be hurt and might even be helped in aggregate by temperature effects. As can be seen in Figure 8, these simulated trends demonstrate the combined spectral and temperature effects in different climate zones as CdTe performs better in Singapore and even surpasses Si in Colorado. A next step for this work is to measure these voltage effects during testing with temperature changes implemented in the experimental setup.

Capturing spectral difference in the  $J_{sc}$  of CdTe and Si with indoor testing of solar modules demonstrates that we can realize and predict differences in EY due to spectra variations. This demonstrates the potential for these methods to be used within industry or academia in the future to test PV technologies for different climates zones.

## 6 | CONCLUSIONS AND FUTURE POSSIBILITIES

In this work, we present the RISE method for classifying solar spectra that overcomes the limitations of past spectral classifiers used in industry and academia. RISE is technology agnostic, and the two key parameters of RISE,  $k_1$  and  $k_2$ , can uniquely classify all spectra worldwide, unlike the commonly used APE. Using RISE classification, we define a representative set of spectra and corresponding temperatures that describes conditions of all the input data from locations from the four climate zones used in this work. With this representative set, PV devices can be tested with a quick and simple procedure to produce EY estimates in the lab. For each condition in the representative set, one IV curve is recorded. Weighing factors are found for a given location as seen in the heat map in Figure 5, and yield is obtained through a weighted average shown in Equation (9). The advantage of these EY calculations is that each additional location or

seasonal prediction can be done with the same set of measurements, only weighted differently per location or time. This enables a more efficient testing approach for research and development encompassing different climate zones and creating more data per unit testing time. We find that our 18 representative spectra correctly reproduce EY differences between Si and CdTe solar cells with an error of less than  $1.5 \pm 0.5\%$  when compared to the ground truth. Also, the RISE method captures the differences between CdTe and Si technologies not captured by STC, increasing accuracy by over 5%.

We also demonstrate a proof of concept of EY testing using LED-based solar simulators that can capture relative spectral and intensity differences between different outdoor spectra. With this experimental setup, we capture the relative  $J_{sc}$  differences between different band gap materials for the various spectra used. This is an improvement over the current STC that uses just one measurement and extrapolates to other conditions. The new LED-RISE method is a promising advance for PV cell and module testing for laboratory or industrial settings.

The next step in this work is the addition of adequate temperature controls to the testing method to emulate the EY modeling predictions in their entirety. Once the temperature ramping is in place, we plan to optimize the testing procedure to decrease overall testing time without losing quality results. This testing routine could then be used in laboratories tuning material designs for local environmental conditions or in industry to predict EY for different climate zones, two applications important for the advancement of PV technologies.

## ACKNOWLEDGEMENTS

The authors would like to thank Patrick Komiske for helpful discussions of ML algorithms and data management techniques. The authors would like to acknowledge Bill Huber at First Solar Inc. for providing CdTe experimental samples. The authors would like to acknowledge Ian Mathews for sourcing GaAs solar cells from FhISE for testing verification. This material is based upon work partially supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number DE-EE0007535. This material is based upon work

partially supported by the U.S. Department of Energy and National Science Foundation (NSF) joint QESST grant. MISTI funding also supported this work. Erin Looney acknowledges support by the NSF Graduate Research Fellowship. André Augusto acknowledges support from NSF and the Department of Energy (DOE) under NSF CA No. EEC-1041895 and DE-EE0008975. This work was supported by the Bavarian State Government (project "PV-Tera—Reliable and cost efficient photovoltaic power generation on the Terawatt scale"; No. 44-6521a/20/5). Open access funding enabled and organized by Projekt DEAL.

## REFERENCES

1. Dupre O, Vaillon R, Green MA. Thermal behavior of photovoltaic devices: physics and engineering. *Solar Energy Mater Solar Cells*. 2015; 140:92-100.
2. Green MA. General temperature dependence of solar cell performance and implications for device modelling. *Prog Photovoltaic Res Appl*. 2003;11:333-340.
3. Singh P, Ravindra NM. Temperature dependence of solar cell performance—an analysis. *Solar Energy Mater Solar Cells*. 2012;101: 36-45.
4. Faine P, Kurtz SR, Riordan C, Olson JM. The influence of spectral solar irradiance variations on the performance of selected single-junction and multijunction solar cells. *Solar Cells*. 1991;31(3):259-278. [https://doi.org/10.1016/0379-6787\(91\)90027-M](https://doi.org/10.1016/0379-6787(91)90027-M)
5. Alonso-Abella M, Chenlo M, Nofuentes G, Torres-Ramirez M. Analysis of spectral effects on the energy yield of different PV (photovoltaic) technologies: the case of four specific sites. *Energy*; 67: 435-443.
6. Peters IM, Karthik S, Liu H, Buonassisi T, Nobre A. Urban haze and photovoltaics. *Energy Environ Sci*. 2018;11(10):3043-3054.
7. Peters IM. Global comparison of the impact of temperature and precipitable water on CdTe and silicon solar cells. In: 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC); 2017; Washington, DC, USA:1140-1142.
8. Eke R, Betts TR, Gottschalg R. Spectral irradiance effects on the outdoor performance of photovoltaic modules. *Renew Sustain Energy Rev*. 2017;69:429-434. <https://doi.org/10.1016/j.rser.2016.10.062>
9. Dimberger D, Blackburn G, Müller B, Reise C. On the impact of solar spectral irradiance on the yield of different PV technologies. *Solar Energy Mater Solar Cells*. 2014;132:431-442. <https://doi.org/10.1016/j.solmat.2014.09.034>
10. Louwen A, de Waal AC, Schropp REI, Faaij AP, van Sark W. Comprehensive characterisation and analysis of PV module performance under real operating conditions. *Prog Photovoltaic Res Appl*. 2017;25: 218-232. <https://doi.org/10.1002/pip>
11. Moon S. Highly efficient single-junction GaAs thin-film solar cell on flexible substrate. *Sci Rep*. 2016;6:30107.
12. Rix AJ, Steyl JDT, Rudman J, Terblanche U, van Niekerk JL. First Solar's CdTe module technology—performance, life cycle, health and safety impact assessment. *Indep Rev Technol*. 2015;1-32.
13. Kato M, Fujiseki T, Miyadera T, Fuyiwara H. Universal rules for the visible-light absorption in hybrid perovskite materials. *J Appl Phys*. 2016;121(11):115501.
14. Wuerz R, Hempel W, Jackson P. Diffusion of Rb in polycrystalline Cu(In,Ga)Se 2 layers and effect of Rb on solar cell parameters of Cu(In,Ga)Se 2 thin-film solar cells. *J Appl Phys*. 2018;124(16):165305.
15. Nelson L, Fritschl M, Panchula A. Changes in cadmium telluride photovoltaic system performance due to spectrum. *IEEE J Photovoltaic*. 2013;3(1):488-493.
16. Lee M, Panchula A. Spectral correction for photovoltaic module performance based on air mass and precipitable water. In: 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC); 2016; Portland, OR, USA:1351-1356. <https://doi.org/10.1109/PVSC.2016.7749836>
17. PlantPredict, version 8. First Solar, Inc.; 2020.
18. Mermoud A. PVSyst Photovoltaic Software; 2020.
19. Jardine CN, Betts TR, Gottschalg R, Infield DG, Lane K. Influence of spectral effects on the performance of multijunction amorphous. In: *Proc of PV in Europe—From PV Technology to Energy Solutions*. Munich, Germany; 2002:2-5.
20. Williams SR, Betts TR, Helf T, Gottschalg R, Beyer HG, Infield DG. Modelling long-term module performance based on realistic reporting conditions with consideration to spectral effects. In: *Proceedings of the 3rd World Conference on Photovoltaic Energy Conversion*; 2003; Osaka, Japan:1908-1911.
21. Nofuentes G, Gueymard CA, Aguilera J, Pérez-Godoy MD, Charte F. Is the average photon energy a unique characteristic of the spectral distribution of global irradiance? *Solar Energy*. 2017;149:32-43. <https://doi.org/10.1016/j.solener.2017.03.086>
22. Liu H, Aberle AG, Buonassisi T, Peters IM. On the methodology of energy yield assessment for one-Sun tandem solar cells. *Solar Energy*. 2016;135:598-604. <https://doi.org/10.1016/j.solener.2016.06.028>
23. Gueymard CA. SMARTS2: a simple model of the atmospheric radiative transfer of sunshine: algorithms and performance assessment. Report No. FSEC-PF-270-95; 1995. <http://institre.org/GCCE/SMARTS2.pdf>
24. Gueymard CA. Parameterized transmittance model for direct beam and circumsolar spectral irradiance. *Solar Energy*. 2001;71(5):325-346. [https://doi.org/10.1016/S0038-092X\(01\)00054-8](https://doi.org/10.1016/S0038-092X(01)00054-8)
25. Gueymard CA. The Sun's total and spectral irradiance for solar energy applications and solar radiation models. *Solar Energy*. 2004;76(4):423-453. <https://doi.org/10.1016/j.solener.2003.08.039>
26. Peters IM, Liu H, Reindl T, Buonassisi T. Global prediction of photovoltaic field performance differences using open-source satellite data. *Joule*. 2018;2(2):307-322. <https://doi.org/10.1016/j.joule.2017.11.012>
27. Peters IM, Buonassisi T. Energy yield limits for single-junction solar cells. *Joule*. 2018;2(6):1160-1170. <https://doi.org/10.1016/j.joule.2018.03.009>
28. Schmager R, Langenhorst M, Lehr J, Lemmer U, Richards B, Paetzold U. Methodology of energy yield modelling of perovskite-based multi-junction photovoltaics. *Opt Express*. 2019;27:507-523.
29. Lloyd SP. Least squares quantization in PCM. *IEEE Trans Inform Theory*. 1982;28(2):129-137. <https://doi.org/10.1109/TIT.1982.1056489>
30. Pedregosa F, Varoquaux G, Gramfort A, others. Scikit-learn: machine learning in Python. *J Machine Learn Res*. 2011;12:2825-2830.
31. Ripalda JM, Buencuerpo J, García I. Solar cell designs by maximizing energy production based on machine learning clustering of spectral variations. *Nature Commun*. 2018;9(1):5126. <http://www.nature.com/articles/s41467-018-07431-3>
32. Andreas A, Stoffel T. NREL Solar Radiation Research Laboratory (SRRL): baseline measurement system. NREL Report No DA-5500-56488, Golden, Colorado; 1981.
33. Riedel N. Direct beam and diffuse spectral irradiance measurements in a nordic country analyzed with the average photon energy parameter. *Proceedings of 7th World Conference on Photovoltaic Energy Conversion IEEE*. Waikoloa, Hawaii; 2018.
34. Ross RG. Interface design considerations for terrestrial solar cell modules. In: *Photovoltaic Specialists Conference, 12th, Baton Rouge, La., November 15-18, 1976, Conference Record (A78-10902 01-44)*. New York: Institute of Electrical and Electronics Engineers, Inc., ERDA-sponsored research; 1976:801-806.
35. Jakhiani AQ, Othman AK, Rigit ARH, Samo SR. Comparison of solar photovoltaic module temperature models. *World Appl Sci J*. 2011; 14:1-8.

36. The Honey Module: DataSheet. Trina Solar Limited; 2015.
37. First Solar Series 4 PV Module. First Solar, Inc.; 2015.
38. Hegedus S. Review of photovoltaic module energy yield (kWh/kW): comparison of crystalline Si and thin film technologies. *Energy Environ Sci.* 2011;2:218-233.
39. Strevel N, Trippel L, Gloeckler M. Performance characterization and superior energy yield of First Solar PV power plants in high temperature conditions. *Photovoltaic Int.* 2012;17(148-154):1-6.
40. Braga M, do Mascimento LR, Rüther R. Spectral modeling and spectral impacts on the performance of mc-Si and new generation CdTe photovoltaics in warm and sunny climates. *Solar Energy.* 2019; 188:976-988.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Looney EE, Liu Z, Classen A, et al. Representative identification of spectra and environments (RISE) using k-means. *Prog Photovolt Res Appl.* 2021;29: 200–211. <https://doi.org/10.1002/pip.3358>