

ORIGINAL RESEARCH

Open access

Estimation of Broadband Atmospheric Turbidity from Global Irradiance or Photovoltaic Power Data Using Deep Neural Networks

Isabella Garcia^{1*}, Diego Herrera¹

Abstract

Photovoltaic systems contribute substantially to electrical grid capacity, yet the intermittent nature of solar resources and uncertainties in output make grid balancing and reserve allocation challenging. Accurately forecasting solar production begins with reliably estimating clear-sky irradiance levels. While broadband atmospheric turbidity parameters can capture reductions in clear-sky conditions, the precision of these models hinges on data capturing current and future aerosol concentrations and water vapor levels. Obtaining such information requires observations that are both temporally and spatially proximate, given the high variability in turbidity. Unfortunately, these kinds of measurements are available in near real time only at a small—and shrinking—set of locations. This study introduces an innovative approach to infer time-dependent, site-specific turbidity levels using more widely accessible data from pyranometers or monitored PV system outputs. The technique leverages a long short-term memory (LSTM) recurrent neural network to isolate the turbidity-related component embedded within global irradiance (or PV power-derived) records, overcoming the natural suppression of this signal. Designed for near real-time implementation, the method supports solar forecasting needs. Evaluation focuses on two aspects: (1) its capacity to replicate turbidity values calculated historically from beam irradiance observations during clear-sky periods; and (2) its effectiveness when supplying turbidity estimates (via persistence forecasts based on daily averages) as inputs to clear-sky models.

Keywords Solar power forecasting, Broadband atmospheric turbidity, Clear-sky modeling, Long short-term memory (LSTM)

*Correspondence:

Isabella Garcia
isabella.garcia@gmail.com

¹ Department of Smart Grid Technologies, Faculty of Engineering, University of Buenos Aires, Buenos Aires, Argentina

Introduction

Background

According to the Solar Energy Industries Association (SEIA) 2020 Solar Market Insight Report [1], solar technologies accounted for the largest share of new grid capacity additions in the United States during both 2019 and 2020. In 2020 alone, photovoltaic (PV) installations represented a record 43% of all new capacity. Projections from SEIA indicate that over the following decade, solar capacity connected to the grid will grow fourfold, adding 324 GW.

This expansion holds particular relevance for Hawaii, where energy consumption greatly exceeds in-state production—roughly by a factor of 11—and imported petroleum fuels about four-fifths of electricity needs. Thanks to abundant sunshine, Hawaii emerged as a prime location for solar deployment and became the first U.S. state to achieve grid parity

for PV. By 2019, solar accounted for more than half of the state's renewable energy output, and in 2020 it supplied over 17% of total electricity generation.

PV generation is non-dispatchable because operators cannot directly control output; instead, production tracks irradiance—the incident radiant power per unit area incident on the panels. At elevated penetration levels, variability and unpredictability in irradiance create difficulties for maintaining grid balance, ensuring stability, and managing reserves. These challenges intensify in Hawaii, where each island relies on its own compact, islanded power system.

Irradiance fluctuations stem from both extraterrestrial and atmospheric influences. Outside the atmosphere, irradiance follows predictable patterns determined by Earth's orbit and the solar cycle. Within the atmosphere, however, incoming radiation undergoes extensive modification through scattering, absorption, and re-emission processes involving air molecules, aerosols, and other particulates [2]. The core challenge in solar forecasting lies in characterizing atmospheric conditions [3, 4]. Many prediction techniques separate baseline clear-sky irradiance—computed via a clear-sky model (CSM)—from cloud-induced reductions. The overall effectiveness of these approaches depends heavily on the quality of the CSM.

In this context, clear-sky irradiance refers to the radiation level under cloud-free atmospheric conditions, where clouds exert no influence on the solar resource or resulting PV output. Under clear skies, attenuation primarily arises from aerosols (suspended particles sized roughly 1 nm to 5 μm [5]) and water vapor [6, 7]. Typical aerosols include sea salt from ocean-air exchange, smoke and soot from wildfires, volcanic ash and dust, sulfates from fossil fuel combustion, and black carbon from heavy oil burning.

Clear-sky models fall into two broad groups: detailed radiative transfer models and simpler empirical formulations [6, 8, 9]. Radiative transfer approaches generally yield superior attenuation estimates but require multispectral observations that are seldom available. Empirical models, better suited to real-time use, typically rely on nearby broadband beam irradiance data from specialized instruments [7, 10–14]. Deployment of such tracking or shadowing devices remains restricted—and is declining—owing to their high cost (around \$30K USD) and ongoing maintenance/calibration demands for moving components [15, 16].

By contrast, pyranometers cost far less (\$1–5K USD) and have no moving parts, reducing maintenance requirements. As a result, standardized global horizontal irradiance (GHI) data from pyranometers are abundant, including many real-time streams. Yet pyranometer readings capture global irradiance, which responds less sensitively to changes in turbidity than beam measurements do. In global data, the turbidity signal weakens because the direct and diffuse components react in opposite directions: direct irradiance decreases with increasing turbidity, while diffuse irradiance increases. This opposing behavior has historically restricted the utility of pyranometer records for turbidity assessment. Earlier efforts using archived GHI data include Polo *et al.* [17]'s approach to derive monthly turbidity from solar-noon GHI, Hove *et al.* [18]'s minimization technique comparing observed GHI to CSM predictions in Zimbabwe, and a comparable strategy by Riveros-Rosas *et al.* [19] for monthly estimates in Mexico.

Aim, hypothesis, and validation

This work seeks to develop a novel technique for estimating near-real-time turbidity using deep neural network (DNN) methods and more commonly available inputs, such as pyranometer readings, weather station data, and PV system monitoring records. DNNs excel at capturing complex nonlinear relationships among multiple variables. The core hypothesis is that a DNN can learn to recover the suppressed turbidity signal from global irradiance measurements, enabling broadband turbidity quantification during clear-sky periods. Broadband turbidity is expressed here via Linke's turbidity coefficient, TL [20, 21], which indicates the equivalent number of clean, dry atmospheres needed to match the radiative extinction of the actual atmosphere.

The DNN serves as a decomposition tool, predicting diffuse irradiance from inputs from weather stations, pyranometers, or PV outputs, along with supplementary model-derived information. These inputs encompass atmospheric attenuation metrics, meteorological variables, solar positioning, and clear-sky flags. Direct irradiance is then inferred from the predicted diffuse component using an energy balance, followed by inversion of a direct irradiance model to derive turbidity. Should the approach prove effective, it would enable broader, more detailed turbidity monitoring. Pyranometer and weather station networks far outnumber beam- or diffuse-measurement stations, and distributed PV data is increasingly accessible via platforms like pvoutput.org. Applications include retrospective analysis of turbidity trends and spatial patterns, as well as near real-time provision of Linke turbidity values to CSMs that use this parameter to account for attenuation beyond Rayleigh scattering in operational solar forecasting.

Method validity is discussed in terms of resulting error, with an individual error, e_i , defined as the difference between a predicted value, p_i , and a corresponding “true” value, y_i ,

$$e_i = p_i - y_i, \quad (1)$$

Obtained from diffuse irradiance measurements (or turbidity derived therefrom). Note that measured values approximate true conditions with inherent uncertainties; “error” here denotes only the difference between method output and reference, not measurement uncertainty.

Summary error metrics include Mean Bias Error (MBE) and Root Mean Square Error (RMSE), calculated for N observations as:

$$MBE = \frac{1}{N} \sum_{i=1}^N (e_i), \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (e_i)^2}. \quad (2)$$

Normalized versions express these as percentages of the mean observed (or derived) value. The coefficient of determination (R^2) quantifies explained variance:

$$R^2 = 1 - \frac{\sum_{i=1}^N (p_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (3)$$

Effective models exhibit low RMSE, MBE near zero, and high R^2 . Evaluations are conducted against references that include observational uncertainties, which affect reported metrics.

Testing and validation use four datasets from three sites on Oahu, Lanai, and Maui (**Figure 1**). To maximize independence, the DNN is trained on only the Kala dataset, with the remaining three datasets reserved for validation.

Each dataset covers at least 500 days and includes diffuse irradiance (for DNN training/validation) and derived turbidity estimates (for overall method validation), along with inputs such as temperature, humidity, wind speed, global irradiance or PV power, solar position, clear-sky irradiance, and clear-sky Boolean flags. Datasets, sites, and instrumentation details follow (**Table 1**).

The Kala dataset originates from instruments at Kalaeloa Airport on southeast Oahu, provided by the National Renewable Energy Laboratory’s (NREL) Measurement and Instrument Data Center (MIDC) under the Solar Resource and Meteorological Assessment Project (SOLRMAP) [22]. Covering July 2009 to November 2012, it features diffuse irradiance from a Multi-Filter Rotating Shadowband Radiometer (RSR) Model MFR-7, global irradiance from a LICOR LI-200 pyranometer, and temperature/humidity from a Campbell Scientific CS215 sensor. Raw 3-second samples were averaged

to 1-minute intervals before release [22]. Both irradiance sensors rely on unfiltered silicon photodiodes with diffusers; post-processing corrects for spectral response limitations. The resulting uncertainty approximates that of first-class thermopile pyranometers, around $\pm 5\%$.

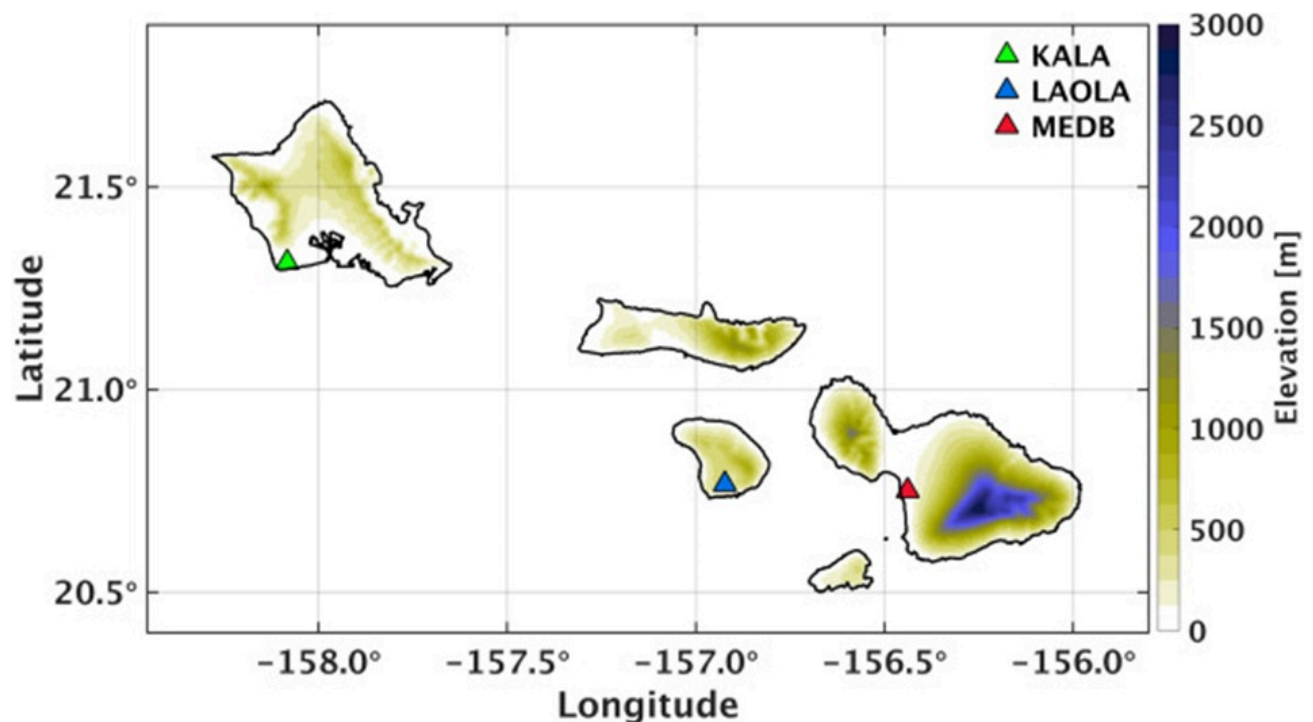


Figure 1. Study site locations.

Another set of observations, designated as Laola, consists of solar irradiance records collected from equipment installed at the 1.5 MW La Ola photovoltaic power station on Lanai Island. This information is distributed by the National Renewable Energy Laboratory (NREL) through its Measurement and Instrumentation Data Center (MIDC) within the SOLRMAP project [22]. Data collection for Laola took place between July 2009 and November 2012. The measurement devices matched those deployed for the Kala campaign; raw samples were acquired every 3 seconds and later averaged over 1 minute before release.

The final pair of datasets comes from monitoring instruments placed at the Maui Economic Development Board (MEDB) facility in Kihei, situated along Maui's southwestern shoreline. There, researchers from the Hawaii Natural Energy Institute (HNEI) developed a dedicated photovoltaic testing platform outfitted with precision sensors to capture fine-scale solar and PV performance data. Both datasets span April 2016 to September 2018. They include diffuse horizontal irradiance obtained with a Delta-T Devices SPN1 Sunshine Pyranometer, together with air temperature and relative humidity logged by a Gill MetPak Pro meteorological station.

The Medb dataset includes total (global) irradiance readings from a Kipp & Zonen SMP21-A class pyranometer. In contrast, the Medb-PV dataset records direct-current electrical output from a dedicated measurement unit monitoring a 1.96 kW array comprising eight premium n-type monocrystalline silicon modules. The SMP21-A qualifies as a secondary standard instrument, with measurement uncertainty typically below $\pm 3\%$. This pyranometer was deliberately aligned to record irradiance in the plane of array (POA) consistent with the PV modules themselves, which face an azimuth of [specific value] $^{\circ}$ (measured clockwise from geographic north) and are inclined at [specific value] $^{\circ}$.

The SPN1 instrument determines diffuse irradiance using a combination of seven thermopile detectors and proprietary software that applies dynamic shading patterns; its uncertainty is typically kept to $\pm 8\%$.

Table 1. Summary of the four time series analyzed. Each series includes observations of environmental conditions, diffuse irradiance data, and estimates of diffuse irradiance generated by a DNN ensemble model.

Empty Cell	LAOLA	KALA	MEDB	MEDB-PV
Location	La Ola, Lanai	Kalaeloa, Oahu	Kihei, Maui	Kihei, Maui
Start Date	July 09'	March 10'	April 16'	April 16'
End Date	December 12'	October 11'	September 18'	September 18'
Latitude	20.8°	21.3°	20.8°	20.8°
Longitude	-156.9°	-158.1°	-156.4°	-156.4°
Altitude	381 m	4 m	59 m	59 m
Tilt	0°	0°	20°	20°
Azimuth	180°	180°	194°	194°

Materials and Methods

General description and illustrative example

An example day illustrating the inputs and outputs of the proposed approach is presented in [Figure 2](#).

In Panel A, the global horizontal irradiance recorded by a pyranometer is plotted alongside the corresponding clear-sky model (CSM) values, assuming a fixed turbidity coefficient of 3. Periods identified as clear-sky conditions—based on the procedure—are also marked on the pyranometer measurements.

Panel B displays the clear-sky attenuation factor, computed as the ratio of the observed pyranometer irradiance (under clear-sky periods) to the CSM-predicted irradiance. This panel also includes the clearness index values, which serve as a quantitative indicator of overall atmospheric attenuation and form one of the primary inputs to the deep neural network (DNN) models .

In Panel C, diffuse irradiance estimates produced by the DNN-based separation model are compared against the actual diffuse measurements. These measured diffuse values were used for training, testing, and validation of the DNN.

Panel D presents both the direct irradiance derived from the DNN outputs and the direct irradiance calculated directly from observations. These two direct irradiance datasets are then utilized in Panel E to derive estimates of atmospheric turbidity.

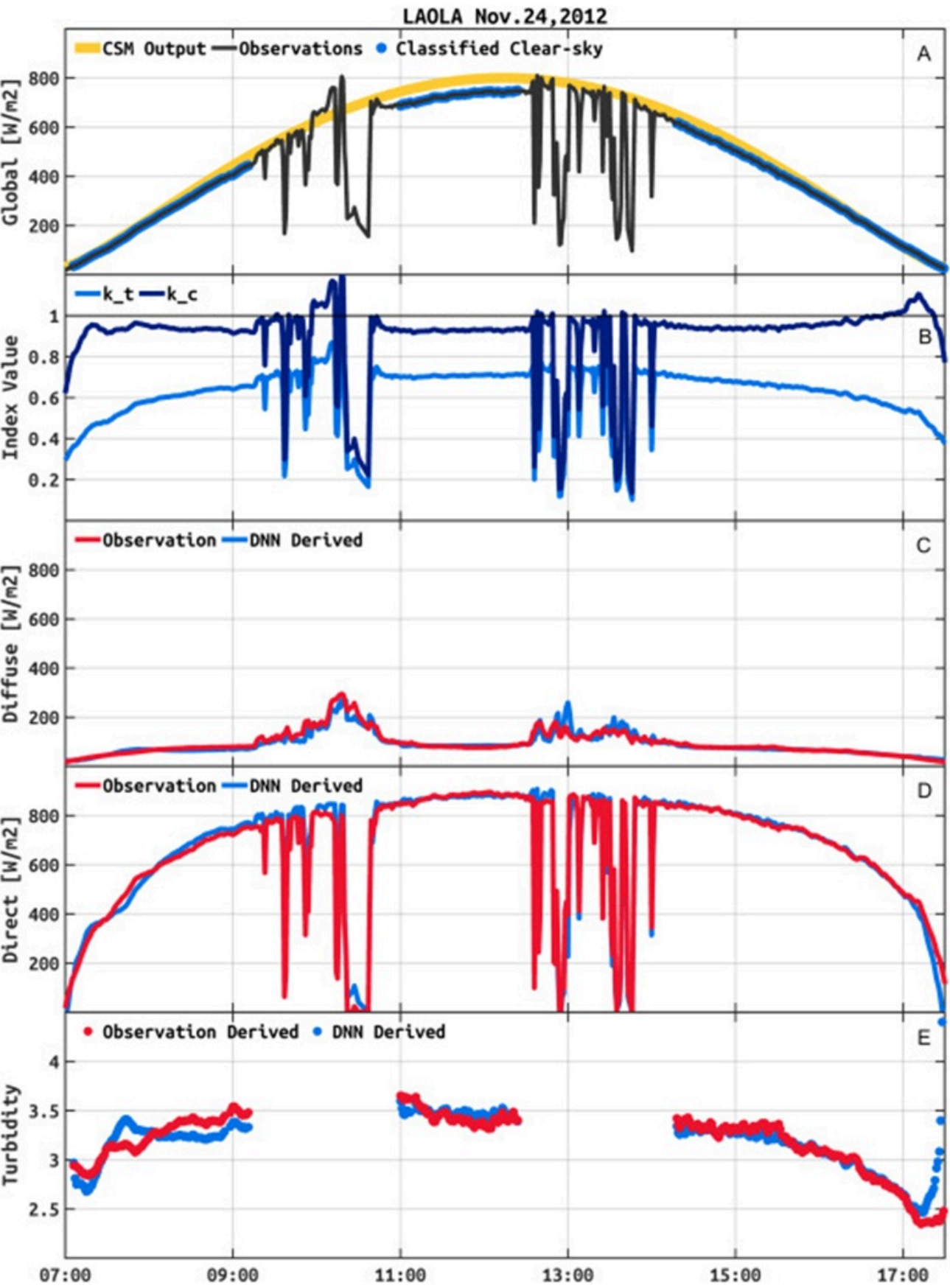


Figure 2. Example day illustrating: (A) observed method inputs, clear-sky model (CSM) predictions, and periods classified as clear-sky, (B) calculated indices of atmospheric attenuation, (C) diffuse irradiance estimates from the DNN alongside actual measured diffuse values, (D) direct irradiance values derived from the DNN and from direct observations, and (E) turbidity estimates produced by the method and those based on observations, shown only for clear-sky periods.

Clearness index and clear-sky index

Atmospheric attenuation of solar radiation is assessed by normalizing measurements from pyranometers or PV system output using standardized index values. The foundational idea of dividing observed solar radiation by its theoretical clear-sky maximum under a completely transparent atmosphere was introduced by Black *et al.* [23]. This evolved into the “clearness index” (k_t), defined as the ratio of measured global horizontal irradiance to extraterrestrial horizontal irradiance [24], and later the “clear-sky index” (k_c), which compares measured horizontal irradiance to the irradiance expected under cloud-free conditions as simulated by a clear-sky model (CSM) [25]. More recently, Engerer *et al.* [26] proposed a “modified clear-sky index” that normalizes measured PV power output against the expected output under clear-sky conditions, computed using a CSM coupled with a PV performance model.

Building on these concepts, the present study adopts the term “clearness index” (k_t) to denote either the ratio of measured irradiance (G) to extraterrestrial irradiance (G_0) or the ratio of measured DC PV power output (P_{dc}) to the corresponding extraterrestrial power output (P_0). The term “clear-sky index” (k_c) refers to the ratio of measured irradiance to modeled clear-sky irradiance (G_{cs}), or the ratio of measured PV power to modeled clear-sky PV power output (P_{cs}), or:

$$k_t = \begin{cases} \frac{G}{G_0}, & \text{if irradiance measurement} \\ \frac{P_{dc}}{P_0}, & \text{if PV power measurement} \end{cases} \quad (4)$$

$$k_c = \begin{cases} \frac{G}{G_{cs}}, & \text{if irradiance measurement} \\ \frac{P_{dc}}{P_{cs}}, & \text{if PV power measurement.} \end{cases} \quad (5)$$

Extraterrestrial irradiance is calculated by applying Fourier series adjustments to the solar constant—taken here as 1361.2 W/m^2 [6]—to account for variations caused by the eccentricity of Earth’s orbit. Solar position is determined using a fast and accurate solar geometry model (SGM) with a maximum error of 0.0027° [27], providing the solar zenith angle (θ_z), declination angle (δ), and hour angle (ω).

Clear-sky direct and global horizontal irradiance values are generated using the Ineichen–Perez clear-sky model [10], which incorporates the Linke turbidity factor (TL) to account for attenuation beyond simple Rayleigh scattering. TL expresses the equivalent number of clean, dry atmospheres needed to match the observed attenuation, based on the optical depth of a theoretical clean dry atmosphere [20, 21]. The Ineichen–Perez version corrects for spurious diurnal fluctuations in turbidity that appear in the original formulation under constant aerosol conditions, ensuring more stable daily turbidity estimates. For tilted surfaces, global clear-sky irradiance is computed by scaling the clear-sky diffuse component with a sky-view factor (assuming isotropic diffuse sky), adding ground-reflected irradiance using an albedo value [28], and enforcing closure. Albedo values for typical land surfaces fall between 0.1 and 0.4; an albedo of 0.1 is adopted in this analysis.

Clear-sky and extraterrestrial PV power outputs are estimated for a given system—defined by module size, type, mounting configuration, and orientation—by combining clear-sky irradiance inputs with a photovoltaic modeling framework (PVM) based on NREL’s PVWatts Version 5 [29]. The PVM incorporates temperature-induced efficiency

reductions via an empirical thermal model and corrects for angle-of-incidence reflection losses using a polynomial fit [30]. Ambient weather data from on-site instruments drives the thermal calculations.

To ensure that changes in broadband turbidity produce consistent changes in the quantified atmospheric attenuation captured by the clear-sky index (kc), both G_{cs} and P_{cs} are computed under a fixed turbidity value ($TL = 3$).

For convenience in subsequent sections, the specific normalization model applied to irradiance and PV power data is denoted CSMre.

Detection of clear-sky periods

Under cloud-free skies, solar irradiance exhibits smooth temporal variation. Clear-sky intervals are detected by applying a sequence of threshold-based criteria that evaluate the smoothness, profile shape, and overall magnitude of observations relative to their corresponding clear-sky references within a sliding time window. This procedure adapts and reformulates the clear-sky detection algorithm originally developed by Reno *et al.* [31] for use with either pyranometer irradiance or PV power time series.

The initial criterion checks whether the clear-sky index falls within an acceptable range, confirming that the observed value is comparable in scale to the modeled clear-sky value. Next, slope and fluctuation statistics are computed over an 11-minute moving window. The second test examines the average slope of the observations against the average slope of the clear-sky reference over the same window. The third test compares a Variability Index (VI) [32] calculated from the measurements to the VI derived from clear-sky values. The fourth and final test compares the standard deviation of the observations in the window to that of the clear-sky counterparts. An observation is classified as clear-sky only if it satisfies all four threshold conditions. Any data point lacking a sufficient number of adjacent valid observations is automatically treated as cloud-influenced.

Starting threshold values were drawn from recommendations in Reno *et al.* [31] and subsequently refined through visual evaluation of the resulting variability metrics and classification outputs. Because quantitative validation of clear-sky detection accuracy is challenging, a qualitative assessment is provided in **Figure 3**, which displays twelve consecutive days with diverse sky conditions extracted from the Laola dataset. Observations flagged as clear-sky should exhibit markedly lower short-term variability while closely matching the magnitude and slope of the CSMre predictions. This behavior is especially evident on December 8 and 9, 2009, when predominantly clear conditions prevailed.

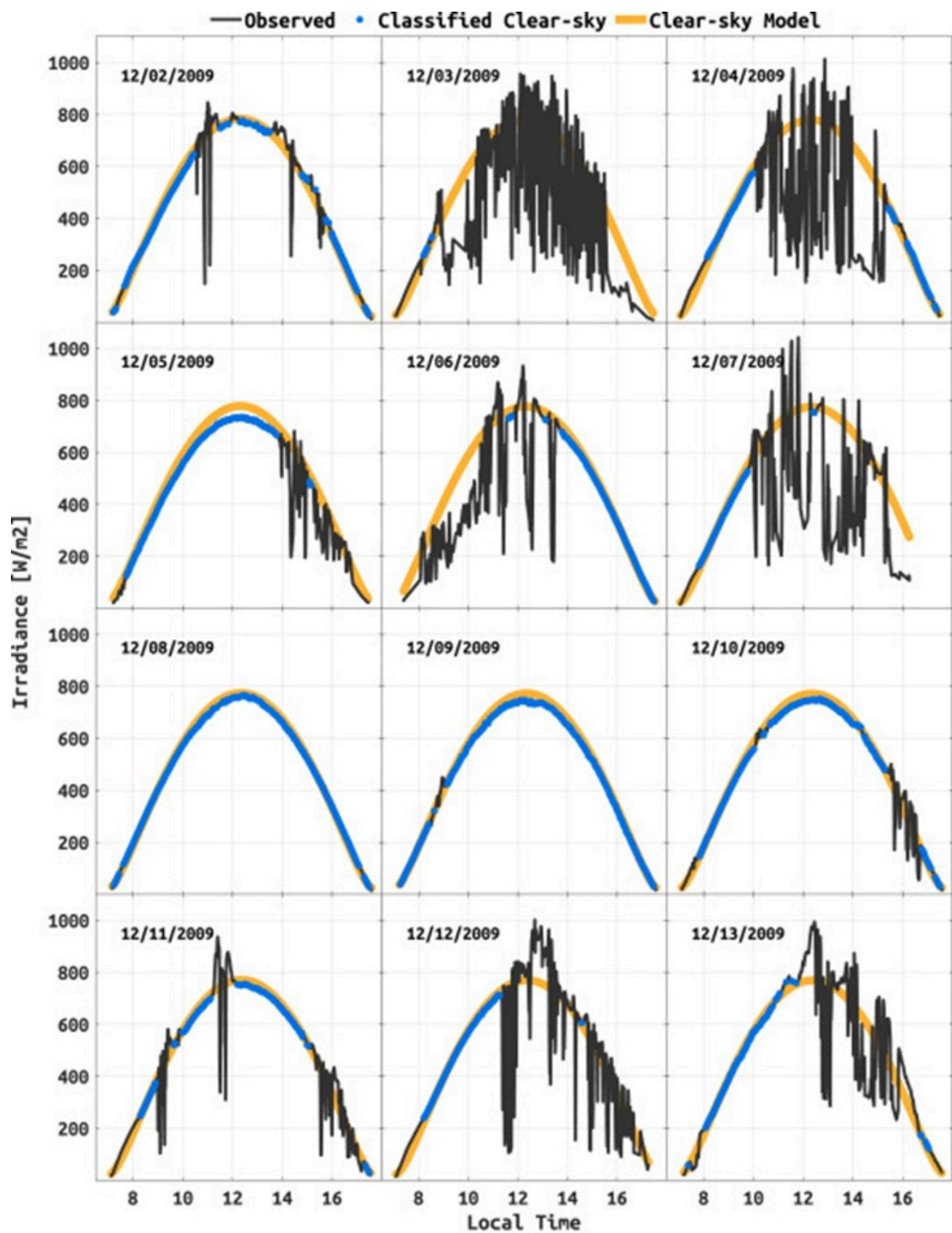


Figure 3. Example time-series of global horizontal irradiance from the LAOLA dataset, with clear-sky conditions highlighted (in blue).

(For detailed interpretation of color references in this figure, refer to the online version of the article.)

DNN separation model

Early separation techniques were primarily used for low-resolution data (hourly or longer) and relied on empirical correlations among global, diffuse, and extraterrestrial irradiance. As time progressed, these models became more intricate by including additional predictor variables. Even though many hourly models can be adapted to higher resolutions [9], recent research has focused on improving 1-minute-resolution predictions. Even though machine learning methods have shown progress at this scale [33–35], separating diffuse and direct irradiance components from global radiation measurements remains challenging. More recent models report nRMSE values of 20%–35% for the diffuse component [36–39]. The approach proposed in this study uses a DNN to separate diffuse irradiance at 1-minute intervals. The model incorporates long short-term memory (LSTM), a specific type of recurrent neural network. LSTM networks are well-suited for learning long-term dependencies in time-series data, using both long-term and short-term memory feedback loops [40]. This ability allows the model to adapt to processes that change over different time scales. The DNN models were developed using MATLAB's Deep Learning Toolbox.

Architecture of the model

The DNN model begins with an input layer followed by six LSTM layers, each containing 100 hidden units. The final layer is a fully connected regression output layer. Additionally, each LSTM layer is followed by a dropout layer with a 10% dropout rate. Dropout is a simple yet effective regularization technique that reduces overfitting by randomly omitting nodes during training, improving the model's ability to generalize [41–43].

Training the model

Training is performed using the Adam optimizer, an adaptation of stochastic gradient descent with adaptive learning rates for each parameter [44]. The input data are standardized to address differences in scale between the input and target variables. The DNN's structure and hyperparameters were optimized through trial-and-error. The training dataset consists of the first half of the KALA dataset, totaling approximately 140,000 data points collected over 8 months. The model training process concluded after 200 epochs. The model's performance was then tested on the full 3-year LAOLA dataset (over 570,000 points) and a subset of LAOLA data from clear-sky conditions (over 128,000 points).

Feature selection for input data

During model development, a variety of input variables were explored, and those that did not contribute meaningfully were excluded. The remaining input features included: relative humidity, ambient temperature, clearness index, clear-sky index, clear-sky classification (as a Boolean value), and solar zenith angle (SZA). The clearness and clear-sky indices are derived from the ratio of measured to theoretical extraterrestrial irradiance. To identify the most effective combination of input variables, 13 DNN models were trained and evaluated based on their performance in predicting diffuse irradiance. Details of the input features used in each model, along with their naming conventions, are provided in Table 2. Performance was assessed by comparing the predicted and observed irradiance, with results presented in Table 3.

Across both cloud-affected and clear-sky conditions in the LAOLA dataset, all models performed reasonably well, with RMSE values ranging from 43.2 to 62 W/m² and normalized values ranging from 23.3% to 33.5%. The majority of models exhibited negative bias, indicating they tended to overestimate irradiance values. This bias was most significant in models that included environmental factors, which reduced model accuracy, particularly under clear-sky conditions. Nevertheless, all 13 models explained a significant portion of the observed variance, with R² values ranging from 0.79 to 0.9. Under clear-sky conditions, performance varied more between models. Two models, S14z and S014z, stood out with notably lower error rates and higher R² values, with rRMSE under 30%. These two models also exhibited the least bias and the highest R² scores. They were among only four models to produce positive R² values, and the only ones with R² values

above 0.5. S014z slightly outperformed S14z in both the full dataset and under clear-sky conditions, making it the chosen configuration for the DNN separation model. The input features for this model include the clearness index, the clear-sky index, the clear-sky classification, and SZA. No additional training was performed; all further discussions about the DNN separation model refer to the S014z configuration.

Table 2. Test configurations of the DNN separation model.

Input ID	Clearness index 0	Clear-sky index 1	Temperature 2	Humidity 3	Clear-sky class 4	SZA z
S0z	X					X
S1z		X				X
S01z	X	X				X
S02z	X		X			X
S12z		X	X			X
S012z	X	X	X			X
S13z	X			X		X
S013z		X		X		X
S0123z	X	X	X	X		X
S01234z	X	X	X	X	X	X
S04z	X				X	X
S14z		X			X	X
S014z	X	X			X	X

Table 3. DNN model performance metrics during input feature testing. Full dataset metrics are calculated using 57,082 data points, and normalized values are calculated using a mean of 185.06 W/m²; clear-sky metrics are calculated using 12,880 points and normalized using a mean of 67.62 W/m².

DNN model	Full dataset					Clear-sky conditions				
	RMSE [W/m ²]	rRMSE [%]	MBE [W/m ²]	rMBE [%]	R ²	RMSE [W/m ²]	rRMSE [%]	MBE [W/m ²]	rMBE [%]	R ²
S0z	56.82	30.7	−14.4	−7.78	0.82	25.86	38.16	−13.66	−20.19	−0.11
S1z	49.02	26.49	−3.5	−1.89	0.87	21.65	31.53	−5.31	−7.78	−0.12
S01z	53.12	28.71	−12.23	−6.61	0.85	24.91	36.58	−8.04	−11.9	−0.03
S02z	57.33	30.98	−15.51	−8.38	0.81	33.56	49.35	−15.19	−22.34	−1.48
S12z	54.62	29.52	−13.21	−7.14	0.84	32.97	48.49	−13.22	−19.44	−1.86
S012z	49.21	26.59	−9.51	−5.14	0.87	28.52	41.84	−12.39	−18.31	−1.35

S13z	51.97	28.08	−15.6	−8.43	0.85	42.79	63.22	−26.79	−39.64	−2.03
S013z	58.72	31.78	−16.06	−8.68	0.81	50.43	74.53	−25.58	−37.83	−3.21
S0123z	61.97	33.49	−23.58	−12.74	0.79	62.62	92.45	−33.45	−49.44	−5.49
S01234z	52.11	28.16	−12.7	−6.06	0.85	26.83	39.47	−14.24	−21.06	−0.19
S04z	48.05	25.96	7.43	4.02	0.87	21.63	31.85	11.03	16.31	0.23
S14z	44.57	24.09	3.39	1.83	0.89	16.01	23.5	−3.61	−5.28	0.58
S014z	43.18	23.33	2.9	1.57	0.9	15.43	22.27	−1.51	−2.24	0.61

Comparative effectiveness

The effectiveness of the deep neural network (DNN) separation approach is assessed by benchmarking it against six widely used separation techniques (Table 4). Among these, the Erbs, Louche, OH, Reindl1, and Reindl2 models rely solely on a single input—the clearness index. In contrast, the Disc model incorporates two variables: the clearness index and air mass. The Dirint model extends the Disc framework by additionally including dew-point temperature and a measure of short-term fluctuations, captured via a so-called “variability index.”

Each of these models determines the direct irradiance component, after which the diffuse irradiance is derived through the closure relation using the measured global irradiance:

$$G = G_n \cdot \cos z + G_d, \tag{6}$$

Where G represents global horizontal irradiance, G_n is direct normal irradiance, z denotes the solar zenith angle (SZA), and G_d stands for diffuse irradiance.

Following the methodology of the previous section, performance is evaluated across the entire dataset (all sky conditions) and only under clear-sky conditions. The corresponding error metrics appear in Table 5. Among the conventional models examined, the four-predictor formulation (Dirint) delivered the strongest performance in both the complete dataset evaluation and the clear-sky subset.

Table 4. Overview of separation models utilized to assess DNN performance. These models rely on the empirical correlation between global and direct clearness indices.

Author	Acronym	Predictor
Erbs et al. [45]	ERBS	Clearness index
Louche et al. [46]	LOUCHE	Clearness index
Orgill and Hollands [47]	OH	Clearness index
Reindl et al. [48]	REINDL1	Clearness index
Reindl et al. [48]	REINDL2	Clearness index
Maxwell [49]	DISC	Clearness index and air-mass

Perez et al. [50]	DIRINT	Clearness index, dew point temperature, variability index, and solar zenith angle
-------------------	--------	---

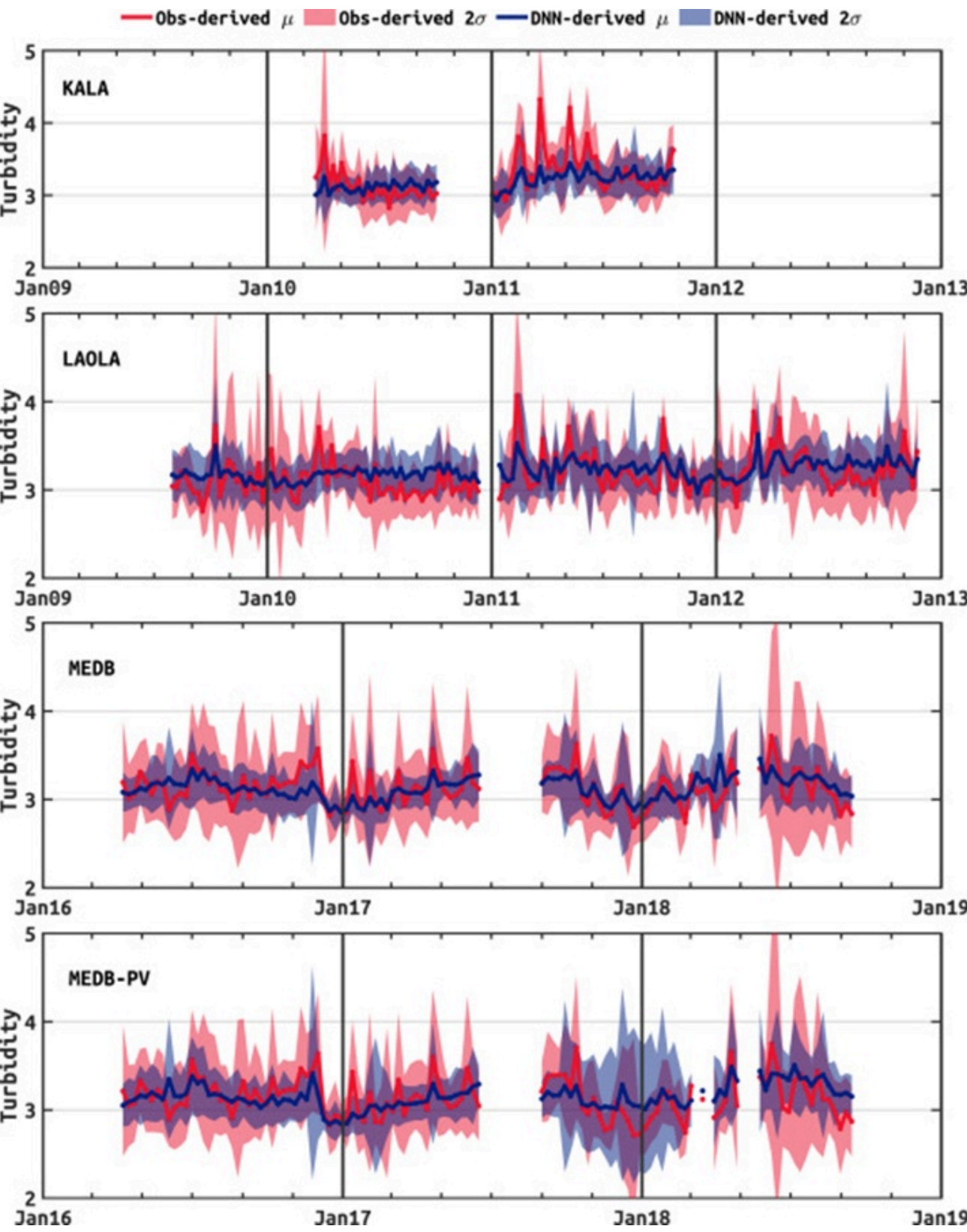


Figure 4. Weekly average and 2D turbidity distributions comparing values derived from observations and those predicted by the DNN model.

Across all sky conditions, the DNN-based separation approach outperforms all conventional methods, even though those standard techniques deliver reasonably strong results overall. Specifically, the DNN model predicts diffuse irradiance with a root mean square error (RMSE) of 73.64 W/m² (equivalent to a relative RMSE of 39.79%) and a mean bias error (MBE) of 12.8 W/m² (corresponding to a relative MBE of 6.91%). The six benchmark models successfully capture a substantial portion of the observed variability in diffuse irradiance, achieving coefficient of determination (R²) values between 0.41 and 0.70. In comparison, the DNN model reduces RMSE by 44.4% to 58.4% relative to the standard approaches, lowers absolute MBE by 77.3% to 90.5%, and achieves R² values that are 0.20 to 0.49 higher.

When restricted to clear-sky conditions, the superiority of the DNN separation model becomes even more pronounced, while the conventional methods exhibit markedly weaker performance. Here, the DNN achieves RMSE reductions of 54.4% to 79% and MBE reductions of 53.7% to 94.4% compared with the standard models. Notably, none of the six conventional models yield a positive R² under clear skies, meaning they fail to explain any of the observed variance in diffuse irradiance (a similar lack of skill was previously observed in 8 out of the 11 DNN configurations evaluated in the earlier section).

Table 5. Performance metrics of the DNN separation model in comparison with the six established separation methods.

The number of data points and mean normalization values used for relative metrics are reported in [Table 3](#).

Model	Full time-series					Clear-sky conditions				
	RMSE [W/m ²]	rRMSE [%]	MBE [W/m ²]	rMBE [%]	R2	RMSE [W/m ²]	rRMSE [%]	MBE [W/m ²]	rMBE [%]	R2
Erbs	90.64	48.98	20.95	11.32	0.55	48.06	71.06	-42.49	-62.86	-2.82
Louche	103.81	56.09	30.69	16.58	0.41	33.63	49.72	-27.94	-41.34	-0.87
OH	88.34	47.74	11.89	6.43	0.57	62.25	92.06	-55.46	-82.04	-5.42
Reindl1	95.52	51.62	22.12	11.95	0.50	59.05	87.32	-51.65	-76.4	-4.77
Reindl2	85.58	46.24	-12.05	-6.51	0.60	71.72	106.08	-64.36	-95.23	-7.52
DISC	82.88	44.79	17.23	9.31	0.62	50.86	75.23	-21.28	-31.42	-3.30
DIRINT	73.64	39.79	12.8	6.91	0.7	33.05	48.87	-7.73	-11.4	-0.81
DNN	43.18	23.33	2.9	1.57	0.9	15.43	22.27	-1.51	-2.24	0.61

Estimating turbidity

The suggested technique is designed to process inputs from either pyranometer measurements or photovoltaic (PV) power output. It estimates direct irradiance through an adapted closure relation expressed as

$$G_n = \begin{cases} \frac{G - G_d}{\cos z}, & \text{if irradiance measurement} \\ \frac{k_c \cdot G_{cc} - G_d}{\cos z}, & \text{if PV power measurement} \end{cases} \quad (7)$$

This estimation depends on the clear-sky index and outputs from the reference clear-sky model (CSM ref).

Subsequently, atmospheric turbidity is derived from an inversion of the Ineichen–Perez clear-sky model, using approaches described by Nou *et al.* [51] and Inman *et al.* [52]. The procedure assumes that the turbidity parameterization in the Ineichen–Perez framework remains consistent under uniform aerosol conditions, enabling the extraction of a characteristic turbidity value applicable across cloud-free conditions regardless of air mass. The inversion takes the following form:

$$T_L = \frac{11.1}{AM} \cdot \ln \frac{b \cdot G_0}{G_n} + 1. \tag{8}$$

Values for air mass (AM) and extraterrestrial irradiance (G0) are obtained directly from the CSM_ref, while the parameter b is computed based on the site’s elevation z according to:

$$b = 0.664 + \frac{0.163}{e^{-h/8000}} \tag{9}$$

This methodology facilitates turbidity estimation either from diffuse irradiance generated by the deep neural network (DNN) or from direct observational data (used for validation).

Results and Discussion

Attention now turns to evaluating the reliability of the turbidity estimates generated by this approach. Performance is measured by comparing discrepancies between turbidity values derived from the DNN and those calculated from observations.

High-frequency validation

To evaluate how well the technique matches turbidity values obtained from observations, various performance indicators are computed at the native temporal resolution of the input data (1-minute intervals). The influence of different input sources on estimation accuracy is examined by analyzing errors across multiple datasets. Findings are summarized in [Table 6](#).

As anticipated, the DNN’s turbidity values show the greatest accuracy when applied to the dataset used for model training (which includes more than 47,000 clear-sky observations). In this case, the DNN-based turbidity closely matches observation-derived values, achieving an RMSE of 0.26 and explaining 55% of the variance ($R^2 = 0.55$). For the second dataset (with over 128,000 clear-sky points), the DNN estimates exhibit marginally larger errors, with an RMSE of 0.28 and an explained variance (R^2) of 0.44. Results for the Medb dataset (containing over [number] clear-sky points) show error levels comparable to those of the Kala dataset. However, its R^2 value is the lowest among pyranometer-based datasets at 0.36. In contrast, the [fourth] dataset, which comprises more than 97,000 clear-sky observations, yields the largest errors among the four, with an RMSE of 0.35—representing roughly a 35% increase compared to the Kala results—and an extremely low R^2 of 0.04, indicating minimal capture of the observed turbidity variability. Across all datasets, the normalized RMSE remains under 11%, while the normalized MBE stays less than 2%

Table 6. DNN-derived turbidity performance metrics at 1-min resolution.

Dataset	RMSE	rRMSE %	MBE	rMBE %	R2
Kala	0.26	8.48	0.01	0.2	0.55
Laola	0.28	9.19	−0.06	−1.88	0.44

Medb	0.26	8.37	0.02	0.64	0.36
Medb-PV	0.35	10.98	0	−0.01	0.04

Validation at the weekly aggregation level

Key factors influencing variations in atmospheric turbidity include nearby pollution sources, wind-driven dispersion, air temperature patterns, and larger-scale atmospheric flows [53]. The analysis here examines how effectively the proposed technique captures changes in turbidity linked to mesoscale and regional circulation dynamics. This averaging timescale aligns with the primary periods of tropical atmospheric circulation, typically spanning 5–10 days [54], while also incorporating seasonal influences and providing sufficient samples per distribution to support robust statistical analysis.

In broad terms, increasing the averaging period reduces RMSE and increases R^2 . Relative to the 1-minute results, daily aggregations lower RMSE by 3% to 17% and boost R^2 by 0.12 to 0.22.

Figure 4 presents a time series of weekly mean turbidity estimates, along with their 2σ spreads (defined here as the 5th–95th percentile interval). Across most weeks, turbidity values range mainly from 2 to 4, though occasional episodes show elevated levels. Both observation-based (Medb and Medb-PV) and DNN-based estimates reveal clear seasonal cycles, which appear more distinct in the Laola and Kala datasets, featuring higher turbidity during summer periods.

The weekly means derived from observations and from the DNN show strong qualitative correspondence overall. However, the 2σ ranges for observation-based estimates are systematically wider than those produced by the DNN.

Figure 5 illustrates a weekly time series of the differences between observation-derived and DNN-derived turbidity values. The mean differences stay predominantly within ± 0.4 , with no evidence of consistent directional bias. The 2σ spreads remain mostly within ± 1 across the datasets, though discrepancies are notably larger in the Medb-PV case than in the others.

Many clear-sky models that incorporate the Linke turbidity parameter default to climatological turbidity datasets when real-time observations are unavailable, an approach that has been demonstrated to degrade model precision [52]. A widely adopted resource is the Solar Radiation Data (SoDa) Service, which provides monthly global turbidity maps at approximately 10 km resolution [55].

For context, **Figure 5** also plots the differences between SoDa climatology-derived turbidity and the observation-derived values at weekly resolution, alongside the observation/DNN differences. The SoDa product exhibits a persistent positive bias relative to the site-specific observation-derived estimates at all three locations. This offset is relatively modest at the more eastern island sites on Lanai and Maui (Laola and Medb), where SoDa values average 0.21 above observations for Laola and 0.3 above for Medb. Yet the 2σ spreads for Laola tend to be broader than those for Medb, occasionally exceeding ± 2 . In contrast, the bias is substantially larger for the Kala dataset on Oahu: SoDa turbidity averages 1.4 higher than observation-derived values, with weekly 2σ difference distributions frequently extending beyond -3 .

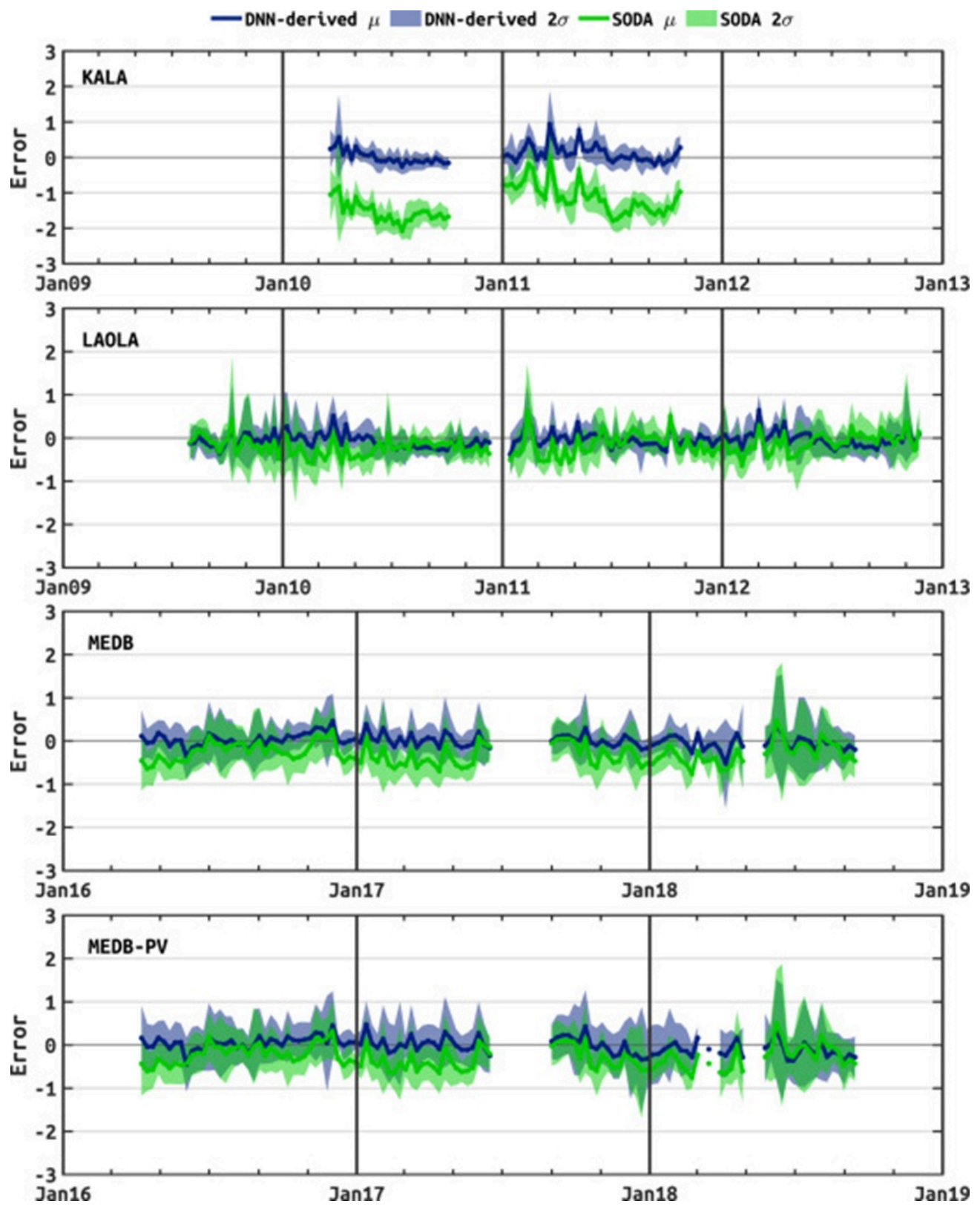


Figure 5. Weekly averaged turbidity values along with the 2σ error distributions for estimates produced by the DNN approach and those derived from the SoDa climatological dataset.

Turbidity estimation using PV power output

Several factors may explain the noticeable drop in performance when the method relies on photovoltaic (PV) power measurements as input. The most significant issues likely stem from shading caused by construction activity on an adjacent property that started in late 2016, as well as soiling of the PV panels. Soiling tends to reduce PV output more severely than pyranometer readings do because the surface area of solar panels is much larger than that of a pyranometer dome, making dust or debris accumulation more impactful on power generation.

To detect periods when PV-specific problems may have degraded the turbidity estimation, differences in the clear-sky index calculated from pyranometer data versus PV power data are examined. The upper panel of **Figure 6** displays the mean bias error (MBE) of these clear-sky index differences, organized into bins by calendar date and time of day (using 5-day and 30-minute bin widths). Comparably, the middle panel of **Figure 6** presents the MBE based on differences between observation-derived and DNN-derived diffuse irradiance values for the Medb-PV dataset.

Shading influences become visible in the clear-sky index differences only when one instrument (pyranometer or PV array) is shadowed while the other remains exposed. Given that the pyranometer is positioned roughly 15 meters from the PV panels, such mismatches can occur. The distinct pattern and timing of the observed bias—an increase in afternoon clear-sky index MBE starting in late 2016—are attributed to shading from the nearby building construction. The strongly positive bias indicates that the PV panels were obstructed while the pyranometer continued to receive full sunlight. A marked escalation in this bias during spring 2017 corresponds to further progress in the construction work. By the following winter, once the building was completed, persistent afternoon shading of the PV system became apparent and persisted for more than 4 months. At its most severe point in December 2017, shading of the PV array began shortly after solar noon.

These same time periods and hours show correspondingly larger discrepancies between observation-based and DNN-based diffuse irradiance estimates. The largest errors in the separation model approach exceed 25 W/m^2 during peak PV shading.

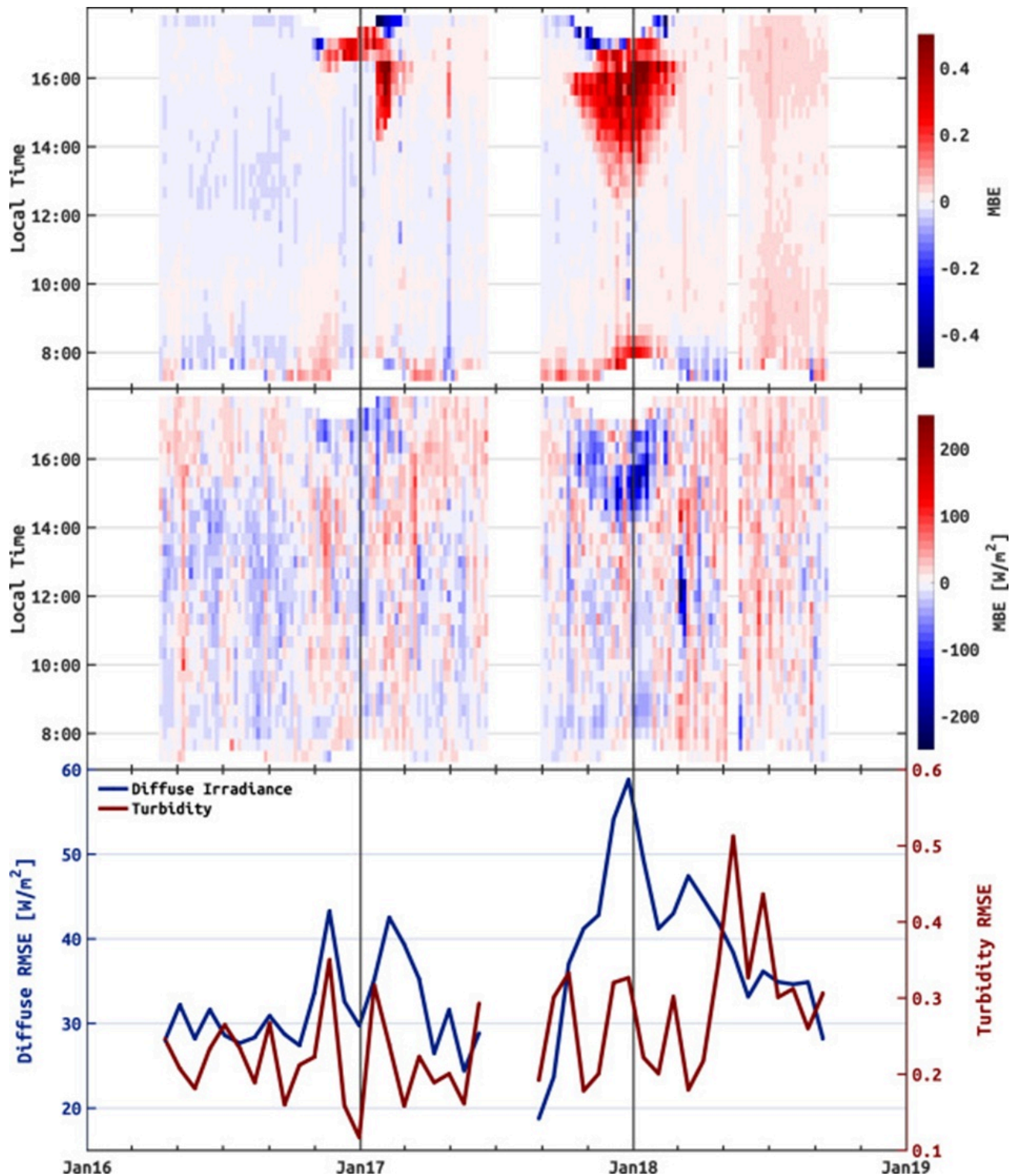


Figure 6. Illustration of issues linked to PV power data: (upper panel) mean bias error (MBE) of the clear-sky index derived from differences between pyranometer-based and PV power-based values; (middle panel) MBE of diffuse irradiance based on differences between DNN-derived (from PV power) and observation-based values; (lower panel) RMSE values for diffuse irradiance and turbidity, computed from differences between DNN-derived (PV power input) and observed values, or from observation-based references, shown at 20-day resolution.

When shading effects are absent, a consistent pattern emerges in the differences in the clear-sky index. Throughout 2016, a prevailing negative bias was observed across most date and time bins. After April 2018, however, this shifts to a positive bias that appears in nearly all bins. This change points to a gradual decline in PV power generation relative to expectations. The drop in output is likely due to the aging of system components and the progressive buildup of dust, dirt, and other deposits on the panels. The particularly rapid rise in bias during 2018 coincides with major volcanic activity on the Big Island of Hawaii, which commenced on April 18, 2018. Prevailing trade winds transported volcanic emissions, including gases and ash, across the island chain, degrading air quality on the affected islands during that period.

The bottom panel of **Figure 6** displays RMSE values for diffuse irradiance and turbidity aggregated over 20-day intervals. The diffuse irradiance RMSE is determined from differences between observation-based and DNN-based (PV-powered) estimates across the entire Medb-PV dataset. Turbidity RMSE is computed similarly, but only for clear-sky conditions, using differences between observation-derived and DNN-derived turbidity. Periods affected by PV array shading exhibit elevated errors in diffuse irradiance separation, yet these do not translate into corresponding increases in turbidity estimation error—shaded measurements are flagged and treated as cloud-contaminated. In contrast, periods of heightened turbidity estimation error align with times when PV soiling is believed to be the dominant factor.

Relevance to solar power forecasting

The method presented here is intended for near-real-time operation, delivering turbidity information that can support improved solar forecasting. This section assesses the performance of the technique when its turbidity estimates serve as inputs to the Ineichen–Perez clear-sky model in a forecasting context (denoted CSM_fore).

To illustrate the influence of turbidity on photovoltaic output, synthetic clear-sky PV generation profiles were created for the MEDB site. These profiles use observed clear-sky irradiance and meteorological conditions as drivers within a PV simulation model. Each annual profile assumes a fixed turbidity level while maintaining otherwise identical inputs. Turbidity values are varied from 2 to 6 in steps of 0.5, spanning the typical range observed across the three measurement sites. Daily energy yields from these profiles are presented in **Figure 7**. Compared to a baseline turbidity of $TL = 2$, energy production at $TL = 6$ is reduced by 14–31%. For each 0.5 increment in turbidity, daily output changes by approximately 1.5–5%, resulting in an overall shift of roughly 3% in total annual energy production.

The evaluation compares CSM_fore performance when driven by three different turbidity sources:

- Day-ahead forecasts based on observation-derived turbidity (CSM_foreOBS),
- Day-ahead forecasts based on DNN-derived turbidity (CSM_foreDNN),
- Day-ahead forecasts using the SoDa climatological turbidity (CSM_foreSoDa).

Day-ahead turbidity forecasts assume simple persistence: the daily mean turbidity from the previous day is carried forward (**Figure 8**). Each daily mean is computed from 1-minute irradiance or PV power observations collected during daylight hours on that day.

Forecast accuracy is evaluated by comparing predicted global horizontal irradiance from CSM_fore against observed clear-sky values. Key performance statistics appear in **Table 7**. For the Medb dataset, metrics are derived exclusively from time steps classified as clear-sky in both the Medb (pyranometer) and Medb-PV (power-based) records. This filtering excludes about 12% of the data but ensures a fairer comparison of the impact of switching to PV power as the input source. Across all CSM_fore configurations, R^2 values exceed 0.99 and are therefore omitted from the table.

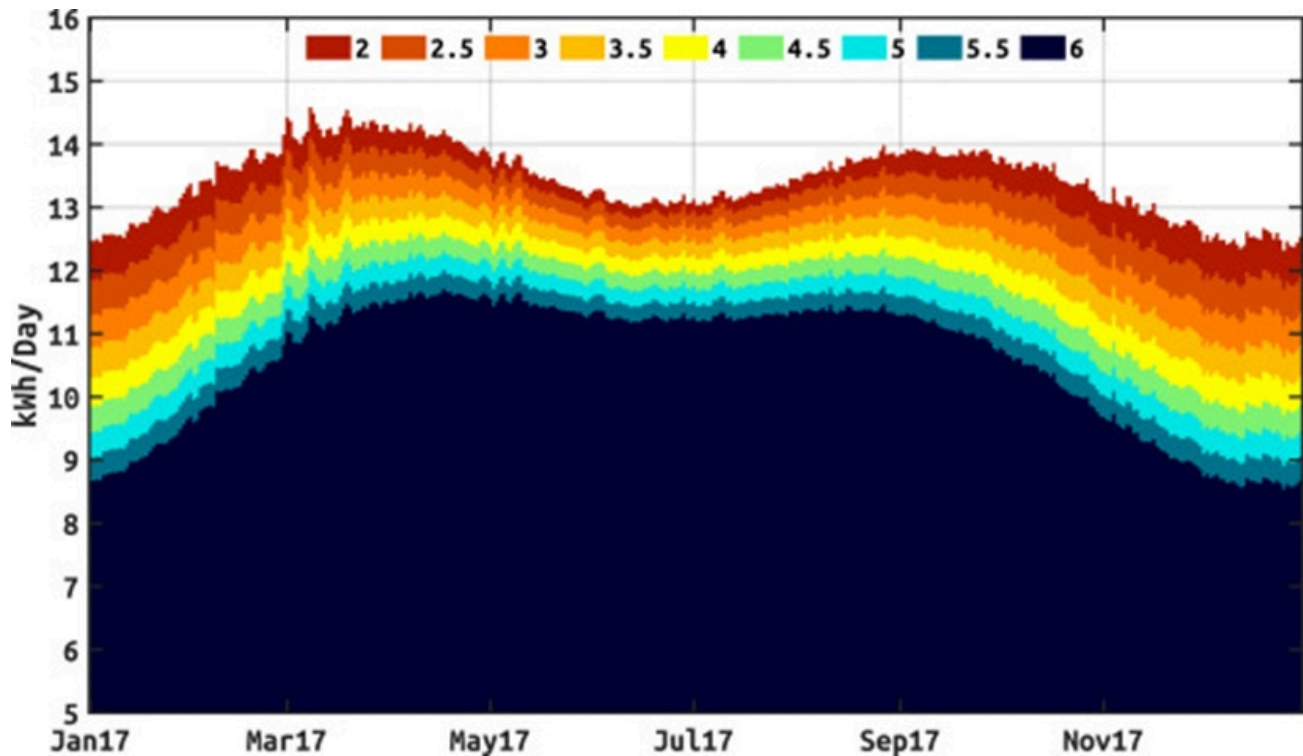


Figure 7. Daily energy output from simulated clear-sky photovoltaic profiles generated across a range of fixed turbidity levels.

Among the evaluated configurations, CSM_foreOBS (using day-ahead persistence forecasts of observation-derived turbidity) delivers the highest accuracy. Root mean square error (RMSE) values fall between 9.13 and 12.49 W/m², which translates to normalized errors of approximately 2%. Mean bias error (MBE) remains small in magnitude, ranging from 0.7 to 1.52 W/m² and corresponding to relative mean bias errors (|rMBE|) below 0.4%.

When turbidity inputs come from the DNN applied to pyranometer measurements (CSM_foreDNN pyranometer-based), a modest reduction in precision is observed. RMSE increases by 1.26 to 2.15 W/m² compared to CSM_foreOBS, although the MBE values show virtually no change and remain effectively unbiased.

Performance declines further when the DNN relies on PV power data as input (CSM_foreDNN PV-based). In this case, RMSE rises by approximately 3.85 W/m² relative to the observation-based reference (CSM_foreOBS), reflecting the cumulative impact of the additional uncertainties introduced by the PV-derived clear-sky index and subsequent turbidity estimation steps.

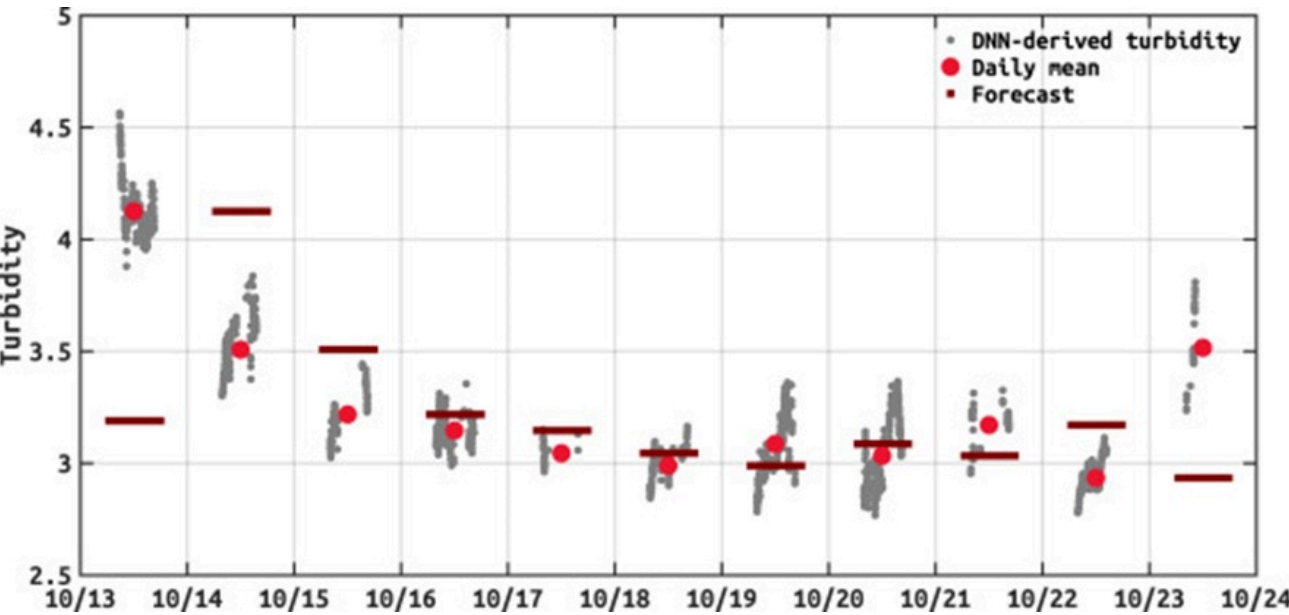


Figure 8. Sample time-series of turbidity predictions made using the daily mean value from the previous day. The daily mean is derived from 1-minute resolution turbidity estimates produced by the DNN.

Table 7. Comparison of CSM_{fore} performance statistics using different turbidity products as input.

Dataset	input	RMSE (W/m2)	rRMSE %	MBE (W/m2)	rMBE%
Kala 452.9 W/m ²	Obs-derived	9.13	2.01	−1.52	−0.34
	DNN-derived	12.39	2.73	−1.05	−0.23
	SoDa	30.88	6.82	27.38	6.04
Laola 459.3 W/m ²	Obs-derived	9.58	2.09	0.84	0.18
	DNN-derived	11.46	2.5	−1.07	−0.23
	SoDa	15.97	3.48	−2.44	−0.53
Medb 690.4 W/m ²	Obs-derived	12.49	1.81	−0.7	−0.1
	DNN-derived	14.64	2.12	−1.87	−0.27
	PV DNN-derived	16.34	2.37	0.72	0.11
	SoDa	20.92	3.03	5.63	0.81

In summary, employing simple persistence-based day-ahead turbidity forecasts in CSM_foreOBS and CSM_foreDNN yields highly accurate clear-sky irradiance predictions, with relative root-mean-square errors (rRMSE) below 3%. These results markedly outperform the configuration that relies on the SoDa climatological turbidity input (CSM_foreSoDa). Specifically, the method-derived forecasts (CSM_foreDNN) reduce RMSE by 28%–66% and MBE by 66%–96% compared to the climatology-driven case.

Conclusion

This study introduces, implements, and rigorously evaluates a novel approach for estimating time-varying, site-specific atmospheric turbidity based solely on global horizontal irradiance measurements from pyranometers. The technique is designed to make optimal use of widely available instrumentation and to enable frequent, near-real-time turbidity retrieval across large areas.

Pyranometers for global irradiance are far more commonly installed than specialized direct or diffuse sensors, owing to their lower cost, simpler deployment and maintenance requirements, absence of moving parts, and longer operational lifespan. The core premise of the method is that a deep neural network (DNN) can be trained to isolate the turbidity-related signal embedded within global irradiance—or even PV system power output—despite the opposing responses of the direct and diffuse components to changes in aerosol loading. By functioning as an effective irradiance separation model, the DNN estimates the beam irradiance component, which is then inverted using a clear-sky model to derive turbidity.

The DNN receives a set of inputs that encode atmospheric attenuation under the assumption of constant turbidity within the reference clear-sky model (Ineichen–Perez): modified clearness and clear-sky indices computed from pyranometer or PV measurements, solar position data from a solar geometry model (SGM), and a binary clear-sky flag produced by a detection algorithm applicable to either pyranometer or PV data. In the case of PV systems, inputs are supplemented by outputs from a PV performance simulation tool driven by clear-sky irradiance and observed environmental conditions.

An extensive search for the best input combination involved training and testing numerous DNN architectures. Models were initially trained on approximately nine months of data collected on Oahu and subsequently validated using more than three years of independent measurements from Lanai. Model quality was assessed across the entire dataset and under clear-sky conditions using RMSE, MBE, and R^2 metrics, with particular emphasis on R^2 during cloud-free periods, as this metric best reflects the model's ability to capture turbidity-driven variations in diffuse irradiance.

Early experiments with both DNNs and conventional separation models yielded persistently negative R^2 values under clear skies. Including temperature and relative humidity as additional inputs actually degraded performance, especially during clear conditions. The decisive improvement came from incorporating a simple Boolean clear-sky indicator as an input feature. With this addition, three DNN configurations achieved positive R^2 under clear skies. The best-performing model used a combination of the clearness index, the clear-sky index, the clear-sky flag, and the solar zenith angle (SZA).

Compared with six widely used empirical separation models, the optimized DNN delivered substantial gains: RMSE reductions of 44.4%–58.4% across all sky conditions and 54.4%–79% under clear skies. Across the full dataset and under clear conditions, the DNN reproduced observed diffuse irradiance with rRMSE below 24%, rMBE within $\pm 3\%$, and R^2 values of 0.90 (all conditions) and 0.61 (clear skies).

Subsequent validation targeted the accuracy of the resulting turbidity estimates across four independent datasets collected at three Hawaiian sites. At the native 1-minute resolution, DNN-derived turbidity closely matched reference values derived from rotating shadowband radiometer (RSR) observations, exhibiting very small bias (rMBE magnitude $< 2\%$) and moderate error (rRMSE $< 11\%$). Explained variance (R^2) reached 0.44–0.55 when using horizontal pyranometer inputs, fell to 0.36 for a tilted pyranometer, and approached zero when relying on PV power-derived indices. Detailed analysis revealed that shading from nearby construction (beginning late 2016) and progressive soiling of PV panels were the primary factors degrading PV performance. Aggregating to coarser timescales markedly improved results: RMSE decreased from 0.26–0.35 (1-min) to 0.24–0.3 (daily) and 0.15–0.23 (weekly).

The final evaluation assessed the utility of the method for supplying turbidity inputs to Linke-turbidity-based clear-sky models in a day-ahead forecasting context. The highest irradiance prediction accuracy was achieved with persistence forecasts based on directly observed turbidity. Forecasts based on the proposed DNN method produced only a marginal

increase in rRMSE (less than 1 percentage point) relative to observation-based forecasts, yet delivered major improvements over forecasts that used the SoDa monthly climatology.

Collectively, these findings demonstrate that the DNN successfully extracts a meaningful turbidity signal from global irradiance measurements, overcoming the natural dampening effect present in pyranometer data. Although performance declines when switching to PV power as the input source, even PV-derived turbidity forecasts still yield substantially better clear-sky model accuracy than climatological defaults. Moreover, the growing availability of pyranometer networks combined with public PV data-sharing platforms (e.g., pvoutput.org) offers the potential for dense regional sampling. Such networks could enable detailed mapping of spatiotemporal turbidity patterns and provide avenues for further enhancing estimation accuracy [56, 57].

It should be emphasized that the present results are specific to the Hawaiian environment, as both training and validation relied exclusively on local measurements. While the method shows strong promise for other tropical regions, broader geographic applicability would require retraining the DNN on representative data from those areas and conducting additional independent validation.

Acknowledgements

The author also wishes to thank HNEI's Severine Busquet, Jonathan Kobayashi, and Richard Rocheleau for their efforts related to the MEDB testbed deployment and data collection, which ONR, United States, also funded.

Conflict of interest

None

Financial support

This work was funded by the Office of Naval Research (ONR), United States, under the Asia Pacific Research Initiative for Sustainable Energy Systems (APRISSES) project, Grant Award Number [N00014-16-1-2116](#), and by the U.S. Department of Energy, United States, under the U.S.-India Collaborative for Smart Distribution System with Storage (UI-ASSIST) project, Award Number [DE-IA0000025](#).

Ethics statement

None

Received: 19 Oct 2024

Revised: 25 Nov 2024

Accepted: 04 Jan 2025

Published online: 20 March 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Perea A, Smith C, Davis M, Sun X, White B, Cox M, et al. Solar market insight report 2020 year in review. 2020 [cited 2021 Nov 15]. Available from: <https://www.seia.org/research-resources/solar-market-insight-report-2020-year-review>
- Goody R, Yang Y. Atmospheric radiation: theoretical basis. 2nd ed. New York: Oxford University Press; 1995.
- Davies JA, McKay DC. Estimating solar irradiance and components. *Sol Energy*. 1982;29(1):55–64.
[https://doi.org/10.1016/0038-092X\(82\)90280-8](https://doi.org/10.1016/0038-092X(82)90280-8).
- Gueymard C, Myers D. Validation and ranking methodologies for solar radiation models. In: Badescu V, editor. Modeling solar radiation at the earth's surface. Berlin: Springer; 2008. p. 479–510.
https://doi.org/10.1007/978-3-540-77455-6_20.
- Wen G, Marshak A, Cahalan RF. Impact of 3-D clouds on clear-sky reflectance and aerosol retrieval in a biomass burning region of Brazil. *IEEE Geosci Remote Sens Lett*. 2006;3(1):169–72.
<https://doi.org/10.1109/LGRS.2005.861932>.
- Gueymard CA. Temporal variability in direct and global irradiance at various time scales as affected by aerosols. *Sol Energy*. 2012;86(12):3544–53.
<https://doi.org/10.1016/j.solener.2012.09.013>.
- Eltbaakh Y, Ruslan M, Alghoul M, Othman M, Sopian K. Issues concerning atmospheric turbidity indices. *Renew Sustain Energy Rev*. 2012;16(8):6285–94.
<https://doi.org/10.1016/j.rser.2012.07.007>.
- Engerer N, Mills FP. Validating nine clear sky radiation models in Australia. *Sol Energy*. 2015;120:9–24.
<https://doi.org/10.1016/j.solener.2015.06.044>.
- Gueymard C, Ruiz-Arias J. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. *Sol Energy*. 2016;128:1–30.
<https://doi.org/10.1016/j.solener.2015.10.010>.
- Ineichen P, Perez R. A new airmass independent formulation for the Linke turbidity coefficient. *Sol Energy*. 2002;73(3):151–7.
[https://doi.org/10.1016/S0038-092X\(02\)00110-0](https://doi.org/10.1016/S0038-092X(02)00110-0).
- Chow CW, Urquhart B, Lave M, Dominguez A, Kleissl J, Shields J, et al. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Sol Energy*. 2011;85(11):2881–93.
<https://doi.org/10.1016/j.solener.2011.08.025>.
- Yang H, Kurtz B, Nguyen D, Urquhart B, Chow CW, Ghonima M, et al. Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. *Sol Energy*. 2014;103:502–24.
<https://doi.org/10.1016/j.solener.2014.02.044>.
- Louche A, Peri G, Iqbal M. An analysis of Linke turbidity factor. *Sol Energy*. 1986;37(6):393–6.
[https://doi.org/10.1016/0038-092X\(86\)90028-9](https://doi.org/10.1016/0038-092X(86)90028-9).
- Gueymard C, Vignola F. Determination of atmospheric turbidity from the diffuse-beam broadband irradiance ratio. *Sol Energy*. 1998;63(3):135–46.
[https://doi.org/10.1016/S0038-092X\(98\)00054-2](https://doi.org/10.1016/S0038-092X(98)00054-2).
- Boland J, Ridley B. Models of diffuse solar fraction. In: Badescu V, editor. Modeling solar radiation at the earth's surface. Berlin: Springer; 2008. p. 193–219.
https://doi.org/10.1007/978-3-540-77455-6_8.

Stoffel T, Reda I. NREL pyrhelimeter comparisons. Golden (CO): National Renewable Energy Laboratory; 2008. Report No.: NREL/TP-3B0-40926.

Polo J, Zarzalejo L, Martin L, Navarro A, Marchante R. Estimation of daily Linke turbidity factor by using global irradiance measurements at solar noon. *Sol Energy*. 2009;83(8):1177–85.
<https://doi.org/10.1016/j.solener.2009.01.016>.

Hove T, Manyumbu E. Estimates of the Linke turbidity factor over Zimbabwe using ground-measured clear-sky global solar radiation and sunshine records based on a modified ESRA clear-sky model approach. *Renew Energy*. 2013;52:190–6.
<https://doi.org/10.1016/j.renene.2012.11.005>.

Riveros-Rosas D, Gonzalez-Cabrera AE, Valdes-Barrón MG, Zarzalejo LF, Ramirez L. Analysis of Linke turbidity index from solar measurements in Mexico. *AIP Conf Proc*. 2018;2033(1):190008.
<https://doi.org/10.1063/1.5067201>.

Linke F. Transmissions-Koeffizient und Trübungsfaktor. *Beitr Phys Atmos*. 1922;10:91–103.

Linke F. Messungen der Sonnenstrahlung bei vier Freiballonfahrten. *Beitr Phys Atmos*. 1929;15:176–82.

Wilcox S, Andreas A. Solar resource & meteorological assessment project (SOLRMAP): rotating shadowband radiometer (RSR). Golden (CO): National Renewable Energy Laboratory; 2009. Report No.: NREL/DA-5500-56495.

Black JN, Bonython CW, Prescott JA. Solar radiation and the duration of sunshine. *Q J R Meteorol Soc*. 1954;80(344):231–5.
<https://doi.org/10.1002/qj.49708034411>.

Liu B, Jordan R. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Sol Energy*. 1960;4(3):1–19.
[https://doi.org/10.1016/0038-092X\(60\)90062-1](https://doi.org/10.1016/0038-092X(60)90062-1).

Bird R, Hulstrom R. Simplified clear sky model for direct and diffuse insolation on horizontal surfaces. Golden (CO): Solar Energy Research Institute; 1981. Report No.: SERI/TR-642-761.

Engerer N, Mills F. KPV: a clear-sky index for photovoltaics. *Sol Energy*. 2014;105:679–93.
<https://doi.org/10.1016/j.solener.2014.04.019>.

Grena R. An algorithm for the computation of the solar position. *Sol Energy*. 2008;82(5):462–70.
<https://doi.org/10.1016/j.solener.2007.10.001>.

Loutzenhiser P, Manz H, Felsmann C, Strachan P, Frank T, Maxwell G. Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation. *Sol Energy*. 2007;81(2):254–67.
<https://doi.org/10.1016/j.solener.2006.03.009>.

Dobos AP. PVWatts version 5 manual. Golden (CO): National Renewable Energy Laboratory; 2014.
<https://doi.org/10.2172/1158421>.

King D, Boyson W, Kratochvill J. Photovoltaic array performance model. Albuquerque (NM): Sandia National Laboratories; 2004. Report No.: SAND2004-3535.

Reno M. Using cloud classification to model solar variability. In: *Proceedings of the ASES National Solar Conference*; 2013 Apr 16–20; Baltimore, MD. 2013. p. 1–10.

Stein J, Hansen C, Reno M. The variability index: a new and novel metric for quantifying irradiance and PV output variability. In: *Proceedings of the World Renewable Energy Forum*; 2012 May 13–17; Denver, CO. 2012. p. 1–8.

Alam S, Kaushik S, Garg S. Assessment of diffuse solar energy under general sky condition using artificial neural network. *Appl Energy*. 2009;86(4):554–64.

<https://doi.org/10.1016/j.apenergy.2008.09.004>.

Elminir HK, Azzam YA, Younes FI. Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models. *Energy*. 2007;32(8):1513–23.

<https://doi.org/10.1016/j.energy.2006.10.010>.

Tapakis R, Michaelides S, Charalambides A. Computations of diffuse fraction of global irradiance: Part 2 – neural networks. *Sol Energy*. 2016;139:723–32.

<https://doi.org/10.1016/j.solener.2015.12.042>.

Engerer N. Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia. *Sol Energy*. 2015;116:215–37.

<https://doi.org/10.1016/j.solener.2015.04.012>.

Hofmann M, Seckmeyer G. Influence of various irradiance models and their combination on simulation results of photovoltaic systems. *Energies*. 2017;10(10):1495.

<https://doi.org/10.3390/en10101495>.

Aler R, Galván IM, Ruiz-Arias JA, Gueymard CA. Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting. *Sol Energy*. 2017;150:558–69.

<https://doi.org/10.1016/j.solener.2017.04.066>.

Yang D, Boland J. Satellite-augmented diffuse solar radiation separation models. *J Renew Sustain Energy*. 2019;11(2):023705.

<https://doi.org/10.1063/1.5087463>.

Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735–80.

<https://doi.org/10.1162/neco.1997.9.8.1735>.

Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580 [cs.NE]*. 2012.

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929–58.

Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems 25 (NIPS 2012)*. 2012. p. 1097–105.

Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv:1412.6980 [cs.LG]*. 2014.

Erbs D, Klein S, Duffie J. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. *Sol Energy*. 1982;28(4):293–302.

[https://doi.org/10.1016/0038-092X\(82\)90302-4](https://doi.org/10.1016/0038-092X(82)90302-4).

Louche A, Notton G, Poggi P, Simonnot G. Correlations for direct normal and global horizontal irradiation on a French Mediterranean site. *Sol Energy*. 1991;46(4):261–6.

[https://doi.org/10.1016/0038-092X\(91\)90069-6](https://doi.org/10.1016/0038-092X(91)90069-6).

Orgill J, Hollands K. Correlation equation for hourly diffuse radiation on a horizontal surface. *Sol Energy*. 1977;19(4):357–9.

[https://doi.org/10.1016/0038-092X\(77\)90006-8](https://doi.org/10.1016/0038-092X(77)90006-8).

Reindl D, Beckman W, Duffie J. Diffuse fraction correlations. *Sol Energy*. 1990;45(1):1–7.

[https://doi.org/10.1016/0038-092X\(90\)90060-P](https://doi.org/10.1016/0038-092X(90)90060-P).

Maxwell E. A quasi-physical model for converting hourly global horizontal to direct normal insolation. Golden (CO): Solar Energy Research Institute; 1987. Report No.: SERI/TR-215-3087.

Perez R, Ineichen P, Maxwell EL, Seals RD, Zelenka A. Dynamic global-to-direct irradiance conversion models. *ASHRAE Trans*. 1992;98(1):354–69.

Nou J, Chauvin R, Thil S, Grieu S. A new approach to the real-time assessment of the clear-sky direct normal irradiance. *Appl Math Model*. 2016;40(15-16):7245–64.
<https://doi.org/10.1016/j.apm.2016.03.022>.

Inman RH, Edson JG, Coimbra CF. Impact of local broadband turbidity estimation on forecasting of clear sky direct normal irradiance. *Sol Energy*. 2015;117:125–38.
<https://doi.org/10.1016/j.solener.2015.04.028>.

Bilbao J, Román R, Miguel A. Turbidity coefficients from normal direct irradiance in central Spain. *Atmos Res*. 2014;143:73–84.
<https://doi.org/10.1016/j.atmosres.2014.02.007>.

Stefanick M. Space and time scales of atmospheric variability. *J Atmos Sci*. 1981;38(5):988–1002. <0988:SATSOA>2.0.CO;2
[https://doi.org/10.1175/1520-0469\(1981\)038](https://doi.org/10.1175/1520-0469(1981)038).

Remund J, Wald L, Lefèvre M, Ranchin T, Page J. Worldwide Linke turbidity information. In: *Proceedings of the ISES Solar World Congress*; 2003 Jun 14-19; Göteborg, Sweden. 2003. p. 1–6.

Webb R, Smith P, Firag A. On the probability of improved accuracy with increased sample size. *Am Stat*. 2010;64(3):257–62.
<https://doi.org/10.1198/tast.2010.09204>.

Dolnicar S, Grün B, Leisch F. Increasing sample size compensates for data problems in segmentation studies. *J Bus Res*. 2016;69(2):992–9.
<https://doi.org/10.1016/j.jbusres.2015.09.004>.