Renewable Energy 206 (2023) 635–644

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

A transferable turbidity estimation method for estimating clear-sky solar irradiance

Shanlin Chen [a], Zhaojian Liang [a], Peixin Dong [a], Su Guo [b,c], Mengying Li [a,d,*]

[a] Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region
[b] College of Energy and Electrical Engineering, Hohai University, Nanjing, China
[c] Nanjing Jurun Information Technology Co., Ltd, Nanjing, China
[d] Research Institute for Smart Energy, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

ARTICLE INFO

Keywords:
Solar resourcing and forecasting
Turbidity estimation
Transferable model
Clear-sky irradiance

ABSTRACT

A transferable turbidity estimation method is proposed for estimating the turbidity and clear-sky solar irradiance. Instead of using on-site irradiance measurements (i.e., the local model), a transferable model is developed involving stations with sufficient information, and then applied at locations with limited data availability. Compared with the local method, the transferable model yields results with slightly higher discrepancies regrading normalized root mean squared error (nRMSE, 2.80% vs 2.75%). When compared with the Ineichen–Perez (PVLIB) model, the nRMSE of clear-sky global horizontal irradiance (GHIcs) estimation is reduced from 4.99% to 2.44%, and the normalized mean bias error (nMBE) is improved from -3.37% to 0.57%. The GHIcs estimation is comparable with physical models (i.e., McClear and REST2), where the McClear produces a nRMSE of 3.32% and the nMBE is 2.10%, while the REST2 generates results with an nRMSE of 2.55% and an nMBE of 1.30%. We further compare aforementioned models for day-ahead GHIcs forecasts using a day persistent way. GHIcs forecast from the transferable method has slightly lower discrepancies of nRMSE and nMBE than the physical models. Considering the complexity of physical models, the transferable turbidity estimation method with comparable performance demonstrates valuable potential for solar resourcing and forecasting applications.

1. Introduction

Clear-sky models, which estimate ground-level solar irradiance under clear-sky (cloudless) conditions, are an important part in solar resourcing and forecasting applications to support solar energy projects [1–4]. In solar resourcing, the ground solar irradiance, e.g., global horizontal irradiance (GHI), can be retrieved from satellite images using either physical or semi-empirical satellite methods based on a clear-sky model [1]. The retrieved solar irradiance data can help with the project feasibility study and optimal system design when there is no on-site ground irradiance measurement available [1,5,6]; Moreover, clear-sky models are also essential in solar forecasting to reduce the negative impact on the system operation caused by the intermittency and variability [1,3,7]. The solar forecasts usually rely on the clear-sky index (CSI), which has different definitions depending on the forecasting method. CSI is the ratio of measured GHI and clear-sky GHI (GHIcs) in a time series forecasting [3]. Meanwhile, CSI can also be calculated from the cloud index (CI) based on the satellite images [1,4], which is particularly applied for locations without sufficient solar irradiance data.

In physical-based solar resourcing methods, radiative transfer simulation is applied through various layers in the atmosphere taking the advantage of modern satellite remote sensing technologies [2,8]. Where the physical clear-sky models, e.g., McClear [9] and REST2 [10], are used to quantify the surface solar irradiance when the sky is free from clouds. The essential inputs such as aerosol properties, atmospheric profiles, and surface albedo can be obtained from a number of reanalysis products including the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Moderate Resolution Imaging Spectroradiometer (MODIS), and Copernicus Atmosphere Monitoring Service (CAMS) reanalysis [2,8]. As a major modulator attenuating the solar irradiance in the atmosphere, cloud properties can be derived from the geostationary satellites, for example, Geostationary Operational Environmental Satellite (GOES) and Meteosat Second Generation (MSG) satellites [2,8]. Although physical clear-sky models generally have better performance as they technically require more detailed atmospheric inputs [1], the inputs acquisition and model implementation are typically associated with difficulties and uncertainties [3,11].

* Corresponding author at: Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region.
E-mail address: mengying.li@polyu.edu.hk (M. Li).

https://doi.org/10.1016/j.renene.2023.02.096
Received 16 June 2022; Received in revised form 17 December 2022; Accepted 18 February 2023
Available online 21 February 2023
0960-1481/

In semi-empirical satellite methods for solar resourcing, a clear-sky model is used to determine the clear-sky irradiance, i.e., GHIcs, which accounts for the irradiance attenuation of aerosol, water vapor content, and other gaseous atmospheric constituents [1]. While the effect of clouds relies on the CI and CSI derived from a set of satellite images within a period [12,13]. Semi-empirical satellite methods are much easier and faster to apply, and the combination with a physical clear-sky model can improve the performance [14]. Chen et al. [15] compared four clear-sky models, namely, REST2, McClear, Ineichen–Perez [16], and Ineichen–Perez TL [15], in GHI estimation using optimized semi-empirical satellite method and GOES-16 imagery. The results show the performance of semi-empirical satellite method is comparable with the physical method, i.e., National Solar Radiation Database (NSRDB) [2]. The Ineichen–Perez TL even produces better GHIcs estimations than McClear and REST2, therefore, has potential in supporting solar resourcing applications. The Ineichen–Perez TL model is based on the Ineichen–Perez model using Linke turbidity (T_L) estimated from common local meteorological measurements [17].

When it comes to solar forecasting, the clear-sky model is suggested to be used to deseasonalize the variations of solar irradiance by using CSI in the development of solar forecasting models [18]. Generally, local sensing is particularly suitable for intra-hour forecasting with adequate on-site instrument [19,20]. Numeric Weather Prediction (NWP) performs better in day-ahead forecasting, but the computation is demanding and the initial boundary conditions may inherit biases [1]. Satellite-based method is popular for intra-day forecasting where the satellite images are used to identify and forecast the cloud distribution [1,18]. Yang [3] discussed the choice of clear-sky models in time series solar forecasting, and it concluded that a better clear-sky model does not yield better forecasts, so McClear model is recommended due to its global availability. However, in practical forecasting applications, the clear-sky irradiance of McClear model is not readily available as an online service that can only be downloaded since 2004-01-01 up to two days ago [21]. Moreover, the required atmospheric inputs of physical clear-sky models (e.g., REST2 and McClear) are difficult to obtain [3,11], their forecasts are therefore associated with more difficulties. In the development of solar forecasting models, the clear-sky irradiance of the applied clear-sky model (e.g., REST2 or McClear) is assumed to be available, which may not be known as a priori for real-time applications.

The Ineichen–Perez model in PVLIB [22] based on T_L from SoDa database [23] is extensively applied in solar forecasting due to its simplicity in implementation [3]. Recently, a new T_L estimation method is proposed to improve the accuracy of T_L estimation using meteorological measurements [17]. The improved T_L estimation is then used as the input of the Ineichen–Perez model with noticeable accuracy improvement in estimating clear-sky irradiance. Considering that the forecasts of meteorological information such as temperature and humidity are far more accurate than solar irradiance forecasts [18], the Ineichen–Perez TL model based on T_L estimated via meteorological information has broad potential in supporting solar forecasting applications.

Despite the potentials of the aforementioned T_L estimation method, it has one limitation that on-site solar irradiance data is required for model development. However, solar irradiance measurements might not be always available due to technical and financial constraints [6,24]. To further expand the applicability of the T_L estimation method, in this work, we herein propose a transferable T_L estimation model based on the methodology presented in [17]. Instead of using local solar irradiance measurements for model development, we first train the model involving the locations with sufficient data, and then apply the developed model at the location of interest for T_L estimation and then clear-sky irradiance estimation. The main meteorological inputs are ambient temperature, relative humidity, wind speed, and atmospheric pressure, which are available at most of the weather stations or can be easily obtained using low-cost instrumentation. The major contributions of this work are summarized as follows:

- Develops a transferable T_L estimation method using common meteorological measurements.
- Compares and evaluates the performance of the transferable T_L estimation method with SoDa interpolated (the default T_L used in PVLIB) and locally estimated T_L.
- Further compares the performance of GHIcs estimation with high-performance physical models including McClear and REST2.
- Evaluates the performance of the transferable T_L estimation model for solar forecasting applications by comparing with physical models.

The remainder of this paper is structured as follows: Section 2 describes the used data, the T_L estimation method, and details of clear-sky models. The performance of the T_L estimation method, and clear-sky models for GHI estimation and discussion are presented in Section 3. Finally, the key findings of this study and recommendations are summarized in Section 4.

2. Data and methods

This section describes the used data and the transferred T_L estimation method. As shown in Fig. 1, we first develop the T_L estimation model involving stations with sufficient solar irradiance data, and then apply the trained model at locations of interest where solar irradiance measurement is not available. The meteorological inputs are ambient temperature, relative humidity, wind speed, and atmospheric pressure, which are easy to obtain and available at most of the weather stations. There are mainly four steps involved in the T_L estimation, namely, clear-sky detection, T_L derivation, model development, and model transfer for T_L and clear-sky irradiance estimations. The details of the used method for each step are presented in the following subsections.

2.1. Data

The data used in this work is from the Surface Radiation Budget Network (SURFRAD) stations [25], namely, Bondville (BON), Desert Rock (DRA), Fort Peck (FPK), Goodwin Creek (GWN), Pennsylvania State University (PSU), Sioux Falls (SXF), and Table Mountain (TBL). The detailed information of the SURFRAD stations is presented in Fig. 2. Data including GHI, diffuse horizontal irradiance (DHI), solar zenith angle, and meteorological measurements over 2010–2020 of all the stations are downloaded and quality controlled. The used meteorological data includes ambient temperature, relative humidity, wind speed, and atmospheric pressure. All the aforementioned measurements are in the time resolution of 1-min and indexed using coordinated universal time (UTC). Note that the GHI and other data at solar zenith angles over 85° are removed, since the GHI value is very low, and the derived T_L is unrealistic due to the high airmass effect [17].

2.2. Clear-sky models

The clear-sky models used for comparison in this study are REST2 [10], McClear [9], Ineichen–Perez [16] model using default T_L (available in PVLIB [22]). The physical REST2 model has repeatedly been verified as one of the high-performance models [3,26], and many of the required input parameters, such as aerosol optical depth (AOD) at 550 nm, amount of ozone, and precipitable water need to be locally measured or remotely sensed [3,10]. The clear-sky irradiance of REST2 used in this work is from the NSRDB [2] with a time resolution of 5-min. The McClear is also a fully physical model requiring atmospheric inputs including AOD at 550 nm, ozone amount, water vapor content, and the aerosol type [9]. McClear applies a lookup table to speed up the RTMs calculations, and the clear-sky irradiance of McClear is available from CAMS [21], with the best time resolution of 1-min up to two days ago since 2004-01-01. PVLIB [22] estimates clear-sky irradiance based on the interpolated T_L coefficient of SoDa monthly means, and the time resolution used in this work is 1-min.

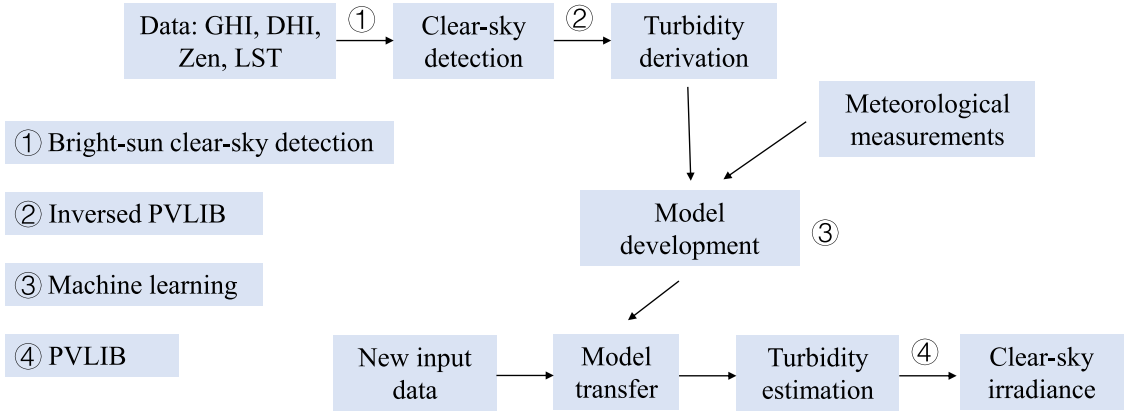

Fig. 1. The flowchart of transferred T_L estimation model using machine learning and meteorological measurements.

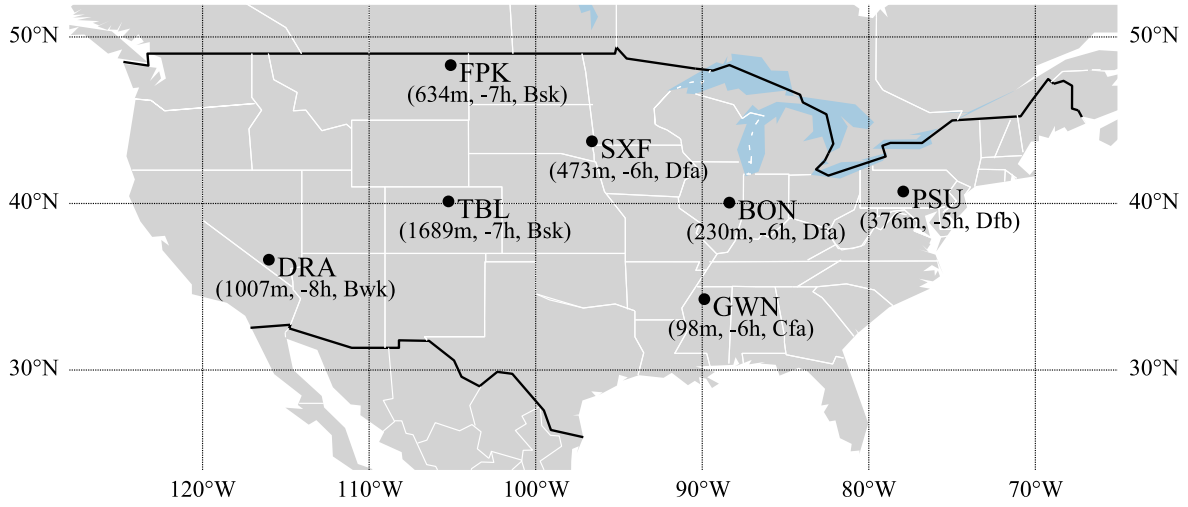

Fig. 2. Summary of the seven SURFRAD stations. The information in the brackets is (altitude [m], time difference from UTC [hours], climate classifications), where Köppen climate classifications are: Bsk (arid, steppe, cold), Bwk (arid, desert, cold), Cfa (temperate, without dry season, hot summer), Dfa (continental, without dry season, hot summer), Dfb (continental, without dry season, warm summer).

2.3. Clear-sky detection and turbidity derivation

The clear-sky instants for all the SURFRAD stations are detected by the Bright–Sun clear-sky detection algorithm [27], which is a globally applicable and freely available clear-sky detection model. The inputs of Bright–Sun model are GHI [$W\,m^{-2}$], GHIcs [$W\,m^{-2}$], DHI [$W\,m^{-2}$], clear-sky DHI (DHIcs) [$W\,m^{-2}$], solar zenith angle [°], horizontal projection of extraterrestrial irradiance [$W\,m^{-2}$], and local standard time (LST). Where GHI, DHI, and solar zenith angle are already available, LST is calculated based on UTC and the timezones detailed in Fig. 2. GHIcs, DHIcs, and extraterrestrial irradiance are derived using PVLIB, horizontal projection of extraterrestrial irradiance is then calculated referring the solar zenith angle. The Bright–Sun model consists of three steps, namely, clear-sky irradiance optimization, tri-component analysis of GHI, DHI, and direct normal irradiance (DNI), and duration filter [27]. The optimization of clear-sky irradiance is to remove the excessive dependence on clear-sky models, the tri-component analysis is to identify the 'clear' periods of all the irradiance components (i.e., GHI, DNI, DHI), and the duration filter is to further improve the accuracy of clear-sky detection by removing the cloud ramp events. More details could be reached in [27]. After detecting the clear-sky instants, the ground truth T_L is derived based on measured GHIcs using inversed Ineichen–Perez model [16] and PVLIB [22] (see the adopted equations from [17] in the Appendix).

2.4. Turbidity estimation model development

In our previous work [17], the results show that the T_L estimation model developed on a daily basis yields comparable GHIcs estimations with the models developed on the time basis of hourly and 5-min, but with much less complexity. The daily T_L estimation model can also be applied in partially clear days [17]. That said, both data samples of clear-sky days and partially clear days can be included in the model development following the same methodology. In specific, to better represent the GHIcs-derived T_L on a daily basis, only the days with more than one third detected clear-sky periods of the daytime are involved in the model training (e.g., if the daytime of day with the solar zenith angle less than 85° is 8 h, only when the detected clear-sky instants are more than 2.4 h, the day is included). Note that the derived T_L should be averaged in the clear-sky periods to represent the daily T_L value, while the meteorological measurements need to be averaged on the daily basis (when the solar zenith angle is less than 85°).

The daily T_L estimation model can be trained locally if the location has adequate data, especially the solar irradiance measurements. In

this work, local T_L estimation model is trained, validated, and tested independently for all the SURFRAD stations for comparison. Data in the year range of 2010–2017 is used for training (20% of which is used for validation), data in 2018–2020 is used for testing. The transferable T_L estimation model is firstly trained and validated at stations with sufficient instrumentation, and then the developed model is applied at another location of interest where the common meteorological measurements are available. For instance, the T_L estimation model can be trained and validated using data from BON, DRA, FPK, PSU, SXF, TBL, and then be applied at GWN for estimating T_L and then GHIcs estimation. The transferred T_L model for each SURFRAD station is developed likewise, where the model is first trained and validated with the other six stations (e.g., excluding GWN) and then applied at the target location (e.g., GWN). Data in 2010–2017 at other six stations is used for training (80%) and validation (20%), data in 2018–2020 at the target location is used for testing and comparison.

The used machine learning algorithm in this study is multilayer perceptron (MLP). MLP is known as feed-forward neural network consisting of the input layer, output layer, and one or more hidden layers based on the applications [28]. The parameters in MLP networks are obtained through back propagation [28]. MLP has a high flexibility in approximation and is widely applied in real applications. The hyperparameters of MLP are tuned using tenfold cross-validation method. For more details on MLP algorithm and the cross-validation method, the reader is referred to Scikit-learn [29], which is the used tool for the model development in this work.

2.5. Turbidity and clear-sky irradiance estimations

Two T_L estimation models, namely, local and transferred T_L estimation models are developed for comparison. The daily T_L is then estimated by the local and transferred models, separately. The meteorological inputs are daily averaged ambient temperature, relative humidity, wind speed, and atmospheric pressure. The 1-min GHIcs at all the SURFRAD stations in 2018–2020 are then derived using Ineichen–Perez (PVLIB) model with the estimated T_L factor, for both the local and transferred T_L estimation models. Since the measured GHI at solar zenith over 85° are not included, the corresponding GHIcs estimations are also removed.

The comparison of GHIcs estimated by different models is in two time resolutions. The GHIcs estimated by PVLIB using default T_L (referred as T_L default), T_L estimated by the local model (referred as T_L local), and T_L estimated by the transferred model (referred as T_L transfer) are in the time resolution of 1-min. Therefore, the comparison of T_L default, T_L local, and T_L transfer is also in 1-min resolution. Since the clear-sky irradiance of REST2 is in a time resolution of 5-min, then the comparison involving REST2 should have the same time resolution. Note that the aggregating measured and estimated GHIcs to 5-min resolution should be the round way (i.e., data points from 13:58, 13:59, 14:00, 14:01, 14:02 are aggregated and indexed as 14:00) [30].

The error evaluation metrics are root mean squared error (RMSE), mean bias error (MBE), and their normalized counterparts (nRMSE, nMBE) defined by the following equations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum(e_i - o_i)^2}$$

$$\mathrm{nRMSE} = \frac{\sqrt{\frac{1}{N}\sum(e_i - o_i)^2}}{\frac{1}{N}\sum o_i}$$

$$\mathrm{MBE} = \frac{1}{N}\sum(e_i - o_i)$$

$$\mathrm{nMBE} = \frac{\sum(e_i - o_i)}{\sum o_i}$$

where e_i and o_i are the pair of GHI estimation and ground observation, N is the total number of compared data points.

Table 1
The RMSE and MBE between derived T_L and T_L estimations from interpolation, local and transferred models at seven SURFRAD stations for both clear-sky and partially clear days in 2019.

	T_L default[a]		T_L local[b]		T_L transfer[c]	
	RMSE	MBE	RMSE	MBE	RMSE	MBE
BON	1.02	0.85	0.34	−0.06	0.34	0.01
DRA	0.62	0.37	0.35	−0.17	0.32	−0.04
FPK	0.69	0.46	0.33	−0.02	0.33	−0.12
GWN	0.75	0.57	0.41	−0.18	0.43	−0.20
PSU	0.91	0.77	0.27	−0.06	0.28	−0.03
SXF	0.92	0.73	0.44	−0.17	0.44	−0.05
TBL	0.73	0.05	0.64	−0.32	0.95	−0.67

[a] 'T_L default' means the T_L interpolation of SoDa monthly means, which is used in the default PVLIB calculations.
[b] 'T_L local' means the T_L estimated by the local model.
[c] 'T_L transfer' means the T_L estimated by the transferred model.

3. Results and discussion

In this section, we evaluate the local and transferable T_L estimation models described above through the comparison of default T_L, derived T_L, and estimated T_L in Section 3.1, the performance of GHIcs estimation using different T_L factors is presented in Section 3.2. We then quantitatively compare the local and transferable T_L estimation models for estimating and forecasting GHIcs with physical clear-sky models of REST2 and McClear in Section 3.3, as physical models are generally considered as the most accurate models. Where the comparison of GHIcs estimation is presented in Section 3.3.1, and the result of GHIcs forecasting is elaborated in Section 3.3.2.

3.1. Evaluation of turbidity estimation models

The invariant T_L interpolations based SoDa monthly climatology means cannot account for the short-term and long-term variations of the aerosols and water vapor content in the atmosphere. As shown in Fig. 3, the derived and estimated T_L factors generally exhibit high fluctuations during the year of 2019, while the default T_L has a comparatively smoother profile. Meanwhile, all the T_L curves show a similar trend that the T_L tends to increase from the beginning of year to some peak point around the middle of the year, then follows a drop till the year end. Apart from higher variations, the derived and estimated T_L coefficients at most stations are generally lower than SoDa interpolations with some exceptions that are likely to happen around the third quarter of the year. The diurnal and monthly variations of T_L factor were also observed in the study of Chaâbane et al. [31], and the typically higher T_L value of the monthly climatology means than the derivations was confirmed by Hove and Manyumbu [32].

The detailed comparison between the derived T_L and T_L estimations is presented in Table 1. It is shown that the default T_L based on interpolations generally shows larger discrepancy in terms of RMSE and MBE than the T_L estimations regardless of local or transferred method is applied (excluding TBL, the possible reason is explained later on). For instance, at BON, default T_L shows a RMSE of 1.02 and a MBE of 0.85, the RMSE of T_L estimation from the local method is 0.34 and the corresponding MBE is −0.06, while T_L estimation of the transferred method also has smaller RMSE and MBE values of 0.34 and 0.01, respectively (see Table 1). When comparing the derived T_L with T_L estimations, both local and transferred T_L estimations can follow the fluctuations of the derived T_L along the year with different discrepancies (see Fig. 3), which means the local and transferred T_L estimation methods can account for the variations of the aerosol and water vapor concentration in the atmosphere. Generally, the transferred method tends to produce T_L estimation with larger values of RMSE and MBE since the local T_L estimation method is developed based on the on-site sensed data, which is more likely to generate better results. For

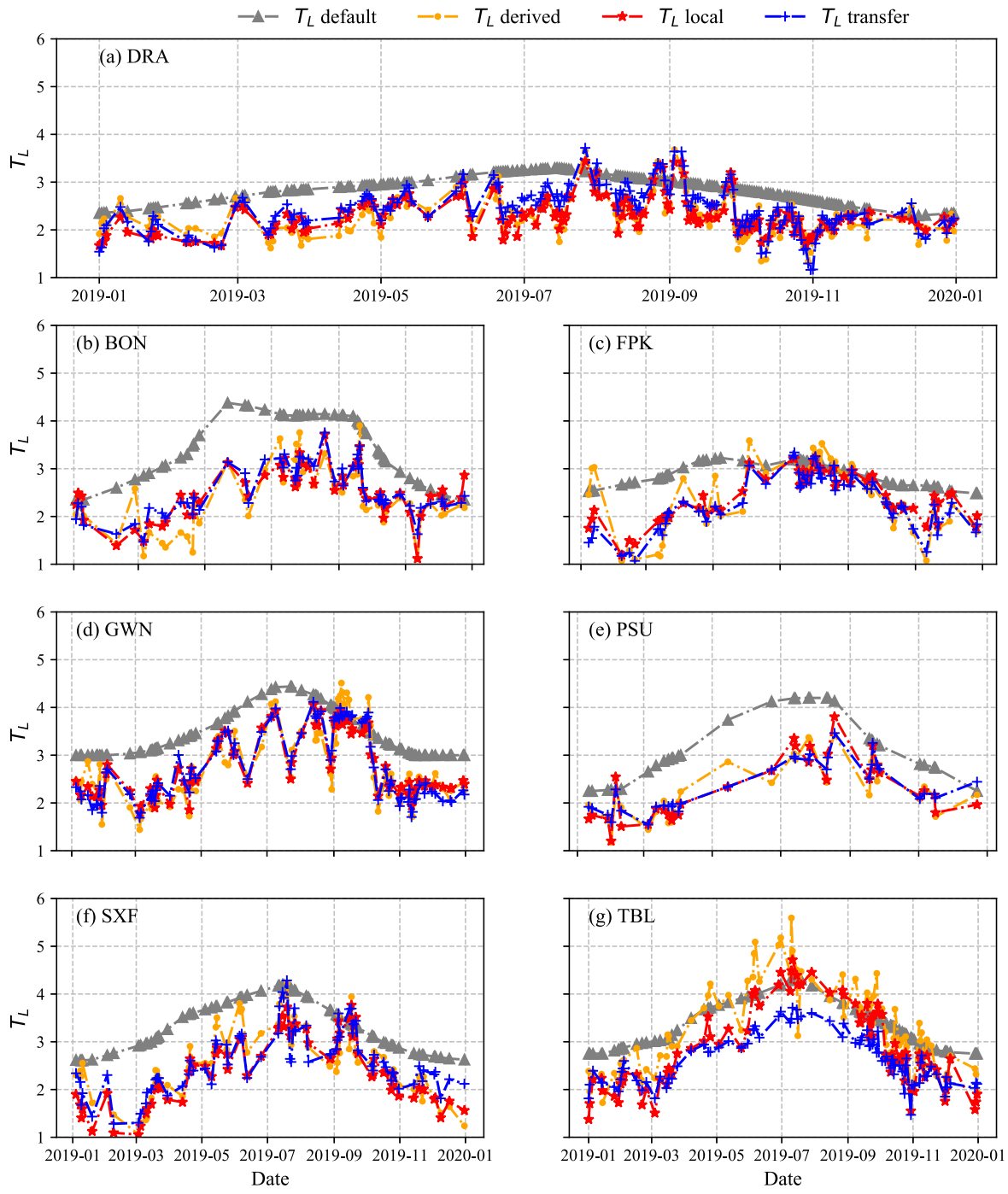

Fig. 3. The comparison of ground truth T_L (T_L derived), PVLIB T_L (T_L default), and T_L estimations from local (T_L local) and transferred methods (T_L transfer) for clear-sky and partially clear days in 2019, at seven SURFRAD stations: (a) DRA, (b) BON, (c) FPK, (d) GWN, (e) PSU, (f) SXF, and (g) TBL. Due to the high occurrence of clear-sky conditions, DRA has more clear-sky and partially clear days detected than the other six stations. Only the results of 2019 are presented here to show the trends, and the T_L profiles in 2018 and 2020 are similar.

example, at GWN, the locally estimated T_L has a RMSE of 0.41 and a MBE of −0.18, while the T_L estimation of the transferred method yields a result with comparatively larger RMSE and MBE of 0.43 and −0.20, respectively. Note that the transferred method sometimes may also yield comparable or even better T_L estimations (see the results of DRA in Table 1). Although the transferred T_L estimation method may lead to larger uncertainties when compared with local estimation, it is a potential way to provide more accurate T_L coefficient than the default PVLIB interpolations.

The generally higher T_L estimations based on the SoDa monthly means, the variations of derived T_L followed by the T_L estimations, and the comparable T_L estimation results of local and transferred models are observed at most of the SURFRAD stations in Fig. 3 and Table 1. However, when it comes to TBL, the observations change noticeably. The default T_L shows relatively lower errors with a RMSE of 0.73 and a MBE of 0.05 when compared with most of the other stations. As illustrated in Fig. 3(g), the derived T_L is more likely to be higher than the default interpolations of SoDa monthly means during the year. The locally estimated T_L exhibits a similar profile with the T_L derivation, but the overall result is inferior to the results at other stations, and the improvement of the T_L estimation performance is also limited. The transferred T_L estimation model even fails to estimate the T_L during

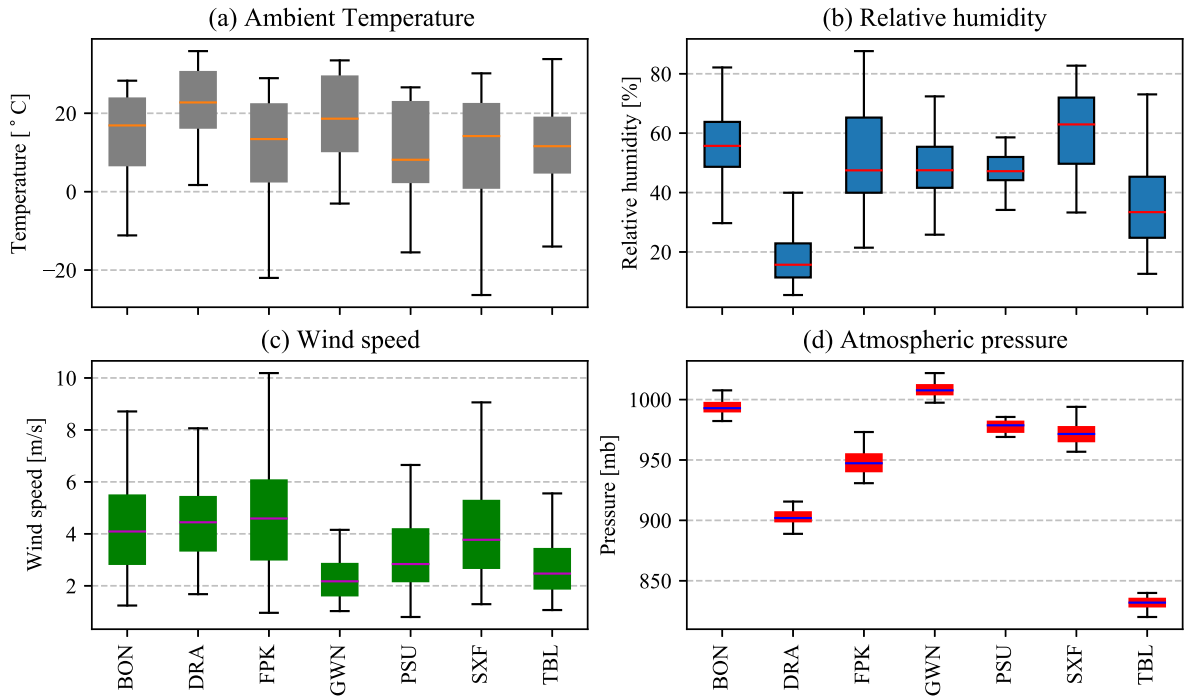

Fig. 4. The statistic properties of the meteorological inputs in the year of 2019. (a) Ambient temperature, (b) Relative humidity, (c) Wind speed, and (d) Atmospheric pressure.

some periods, where a clear underestimation is observed and thus the intra-year T_L variations are also not accounted for. It is most likely that the unique climate condition at TBL leads to the under-performance of the transferred T_L estimation model. As presented in Fig. 4, the meteorological measurements show different statistic properties, such as the maximum, the minimum, and the mean, at each station. Therefore, a better T_L estimation model could be developed when local climate features are accounted for. When there is no sufficient on-site data for local model development, the transferred model could be an option. Recall that the transferred T_L estimation model is first trained and validated at stations with adequate information, and then applied at the location of interest. The transferred T_L estimation method can generally work for most locations but TBL. The possible reason could be the lower atmospheric pressure than the other stations as shown in Fig. 4(d). When developing the transferable T_L estimation model for TBL using data from other six stations, the atmospheric pressure, unlike ambient temperature, relative humidity, and wind speed, stays far away from the range of used inputs. The use of out-of-range local pressure as an input for the transferred T_L estimation method might produce a result with large discrepancy, this is why the transferable T_L estimation model underperforms. That said, it is suggested to use data from the locations with similar climate characteristics when developing the transferable T_L estimation model for practical applications.

The sensitivity analysis in the study of Chen and Li [17] shows that ambient temperature and relative humidity have comparatively larger influence on the T_L estimation than wind speed. The estimated T_L goes up with the increases in ambient temperature and relative humidity, while the increases in wind speed and atmospheric pressure result in a drop of T_L estimation. Wind speed has the least importance in T_L estimation. As for pressure, the sensitivity analysis presents relatively larger variations when changing the pressure by 10%. However, this have limited implications as the piratical atmospheric pressure for a certain place does not has high variations compared with other meteorological parameters. Since the transferred T_L estimation follows the same methodology of the local methods, the sensitivity analysis of the transferred T_L estimation model should have similar results.

Table 2
The comparison of T_L and GHIcs estimations at TBL in 2018–2020, including the T_L estimation models developed with and without atmospheric pressure (P_a).

	T_L estimation		GHIcs estimation	
	RMSE	MBE	nRMSE [%]	nMBE [%]
T_L local with P_a	0.64	−0.32	3.23	1.55
T_L local without P_a	0.70	−0.46	3.40	1.84
T_L transfer with P_a	0.95	−0.67	4.43	2.84
T_L transfer without P_a	1.48	−1.20	6.70	5.23

To further investigate the influence of atmospheric pressure in the T_L estimation models, new local and transferred T_L estimation models for TBL are developed without using atmospheric pressure. The results of T_L and related GHIcs estimations are presented in Table 2. When the atmospheric pressure is excluded in the development of T_L estimation models, both local and transferred models produce T_L estimations with larger errors of RMSE and MBE as shown in Table 2. Similarly, the GHIcs estimations based on T_L estimated by models without using atmospheric pressure also have comparatively larger discrepancies of nRMSE and nMBE. Therefore, it is suggested to use the atmospheric pressure as an input parameter in the development of T_L estimation models.

3.2. Comparison of GHIcs estimations using different turbidity factors

Since the default T_L interpolations generally have larger RMSE and MBE values and do not show many variations along the year (see Table 1 and Fig. 3), the related GHIcs estimations are then also associated with higher discrepancies of nRMSE and nMBE as shown in Table 3. It tends to underestimate the GHIcs using default PVLIB as the interpolated T_L from SoDa monthly means are typically higher than the T_L derivations. When compared with default T_L interpolations, the estimated T_L from both local and transferred methods are more likely to yield better GHIcs estimations (see Table 3).

As discussed in Section 3.1, the local T_L estimation models tends to generate better T_L estimations via the direct inferring to the on-site meteorological features. Therefore, the GHIcs estimation using the

Table 3
The nRMSE [%] and nMBE [%] of 1-min averaged GHIcs estimations using different T_L factors at SURFRAD stations in 2018–2020. The used T_L coefficients are default interpolations (T_L default), estimations of the local model (T_L local) and transferred model (T_L transfer).

Station	T_L default		T_L local		T_L transfer	
	nRMSE [%]	nMBE [%]	nRMSE [%]	nMBE [%]	nRMSE [%]	nMBE [%]
BON	8.04	−6.32	3.25	0.39	3.31	−0.18
DRA	3.40	−1.96	2.15	0.89	1.97	0.14
FPK	4.88	−2.78	2.56	0.24	2.54	0.80
GWN	5.91	−3.83	3.58	1.32	3.67	1.37
PSU	6.71	−5.29	2.86	0.25	2.78	0.36
SXF	6.76	−4.88	3.69	1.48	3.68	0.67
TBL	3.55	−0.23	3.23	1.55	4.43	2.84
ALL[a]	5.24	−3.41	2.75	0.45	2.80	0.84

[a]TBL is not included due to the under-performance of the transferred T_L estimation model.

Table 4
Overall results of GHIcs estimations and day-ahead persistent GHIcs forecasts with the time resolution of 5-min at SURFRAD stations excluding TBL in 2018–2020. Used models are the Ineichen–Perez model with three T_L inputs, the McClear model, and the REST2 model.

Model	GHIcs estimation		GHIcs forecasts[a]	
	nRMSE [%]	nMBE [%]	nRMSE [%]	nMBE [%]
T_L default	4.99	−3.37	4.64	−3.02
T_L local	2.38	0.16	2.85	0.32
T_L transfer	2.44	0.57	2.99	0.74
McClear	3.32	2.10	4.11	1.55
REST2	2.55	1.30	3.51	0.79

[a]The forecast is based on a day persistent method, where the day-ahead GHIcs is assumed as the same as the present day.

locally estimated T_L might also have smaller divergences in terms of nRMSE and nMBE. Since the transferred models are developed using data from other locations, there could be more uncertainties associated in the related T_L estimations. This also introduces comparatively larger discrepancies in the GHIcs estimations using T_L estimated by the transferable models. However, there is no significant difference between the local and transferred T_L estimation methods regarding the GHIcs estimations. Consequently, the transferred T_L estimation model at TBL shows inferior performance in T_L estimations, which also leads to larger errors in estimating GHIcs.

The overall performance of GHIcs estimation at SURFRAD stations (TBL is not included due to the under-performance) is also presented in Table 3. Using the default T_L to generate GHIcs estimation has comparatively larger errors with an nRMSE of 5.24% and an nMBE of −3.41%, while using the estimated T_L from local and transferred methods show noticeable improvements. The nRMSE of GHIcs estimation based on estimated T_L is reduce to 2.75% and 2.80% for the local and transferred methods, respectively. The nMBE is improved to 0.45% for the local method, and to 0.84% for the transferred method. Although the overall performance of locally estimated T_L is slightly better than the transferred T_L estimations, the transferable model is still viable to estimate the T_L and GHIcs for locations without sufficient data.

3.3. Comparison of GHIcs estimations and forecasts with physical models

To further evaluate the results of GHIcs estimation using improved T_L estimations, we herein compare the performance with physical models, namely, the McClear model and the REST2 model. The comparison is in two folds: one is the real-time GHIcs estimation and the other one is persistent day-ahead GHIcs forecasts.

3.3.1. Comparison of GHIcs estimations with physical models

Table 4 details the overall performance of 5-min GHIcs estimations and forecasts using T_L based model and physical at SURFRAD stations, TBL is not included due to the under-performance in T_L estimations using the transferred method. The default T_L produces a GHIcs estimation with the largest nRMSE of 4.99% and the nMBE is −3.37%, while the local T_L estimation generates the best GHIcs estimation with the nRMSE of 2.38% and the nMBE of 0.16%. As expected, the transferred T_L estimation yields a result with relatively larger discrepancies compared with the locally estimated T_L, the nRMSE and nMBE are 2.44% and 0.57%, respectively. Note that the transferred T_L estimation even produces better results than the McClear model and the REST2 model in terms of nRMSE and nMBE (see Table 4), while the REST2 model outperforms McClear with the nRMSE of 2.55% and nMBE of 1.30%. However, there is no significant difference between estimations from the transferred model and the physical models. The detailed comparison of GHIcs estimations at each SURFRAD station is presented in Fig. 5. Physical models are more likely to produce over-estimations, and the local and transferred T_L estimations also tend to overestimate GHIcs but with smaller bias.

The possible reason why the empirical model based on improved T_L estimation yields comparable results with physical clear-sky models could be the avoidance of uncertainty accumulation. Physical models require detailed atmospheric inputs such as aerosol, water vapor and ozone, these inputs are usually based on reanalysis products such as MERRA-2. The MERRA-2 reanalysis products are derived from satellite measurements and therefore are associated with uncertainties [33]. This means the use of reanalysis products for clear-sky irradiance estimation in physical models would have accumulated uncertainties. However, the T_L derivation is a one-step-through process based on quality-controlled irradiance data, which includes imbedded T_L information. Therefore, it is more likely to avoid the accumulation of multi-step uncertainties. Although physical clear-sky models are proved to have higher accuracy in GHIcs estimation, the comparable results with less complexity of the transferred T_L estimation model demonstrates its potential usage for locations without sufficient on-site information.

3.3.2. Comparison of GHIcs forecasts with physical models

Since clear-sky irradiance (clear-sky model) is also essential in solar forecasting, we also evaluate and compare the performance of T_L based forecasts with the McClear model and the REST2 model. Considering that McClear and REST2 are physical models requiring detailed atmospheric inputs, which are difficult to obtain and forecast, we therefore apply a day persistent method to predict the GHIcs in the coming day. In specific, the profile of GHIcs for the coming day is assumed as the same as the present day [34,35]. Note that McClear is available as a web service from 2004-01-01 up to two days ago, and is recommended for solar forecasting applications in [3]. However, in real time forecasting applications, e.g., the present day and the coming day, the clear-sky irradiance of McClear is not available. This means the atmospheric inputs for both McClear and REST2 should be obtained at present day, which introduces even more difficulties in retrieving and measuring the atmospheric optical properties. To perform a fair comparison, we assume the clear-sky irradiance (i.e., GHIcs) of REST2 and McClear could be obtained for the present day, and the meteorological measurements are available.

The overall result of GHIcs forecasts using the day persistent method is presented in Table 4. The largest nRMSE (4.64%) and nMBE (−3.02%) are generated using the default T_L, while the other two T_L based GHIcs predictions show lower nRMSE than the physical models. The local and transferred T_L estimations yield GHIcs forecasts with the nRMSE of 2.85% and 2.99%, respectively. The nRMSE of McClear based forecasts is 4.11%, while REST2 produces a result with the nRMSE of 3.51%. GHIcs forecasts based on estimated T_L also have smaller biases, the transferred method generates a comparatively larger nMBE of 0.74%, while the local model produces a result with the nMBE of 0.32%. The physical models are likely to produce relatively larger over-estimations in forecasting GHIcs, where the McClear shows an nMBE of 1.55 and

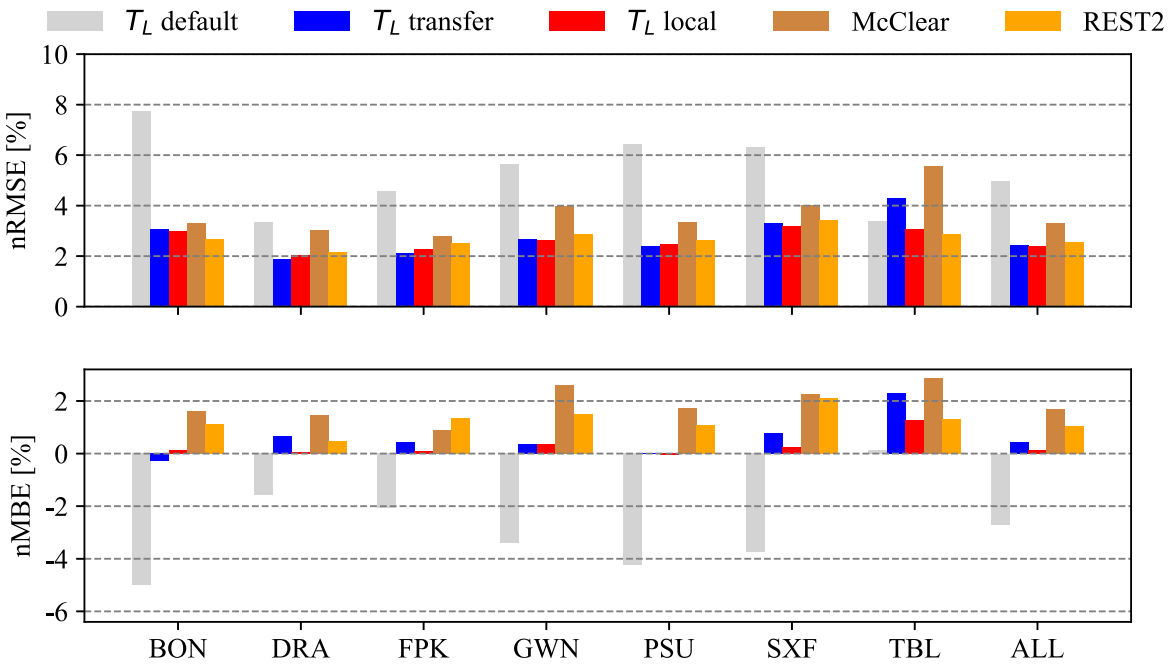

Fig. 5. The nRMSE and nMBE between the 5-min averaged GHIcs measurements and estimations using different models at SURFRAD stations in 2018–2020. GHIcs is estimated by the Ineichen–Perez model with three different T_L inputs, the McClear model, and the REST2 model.

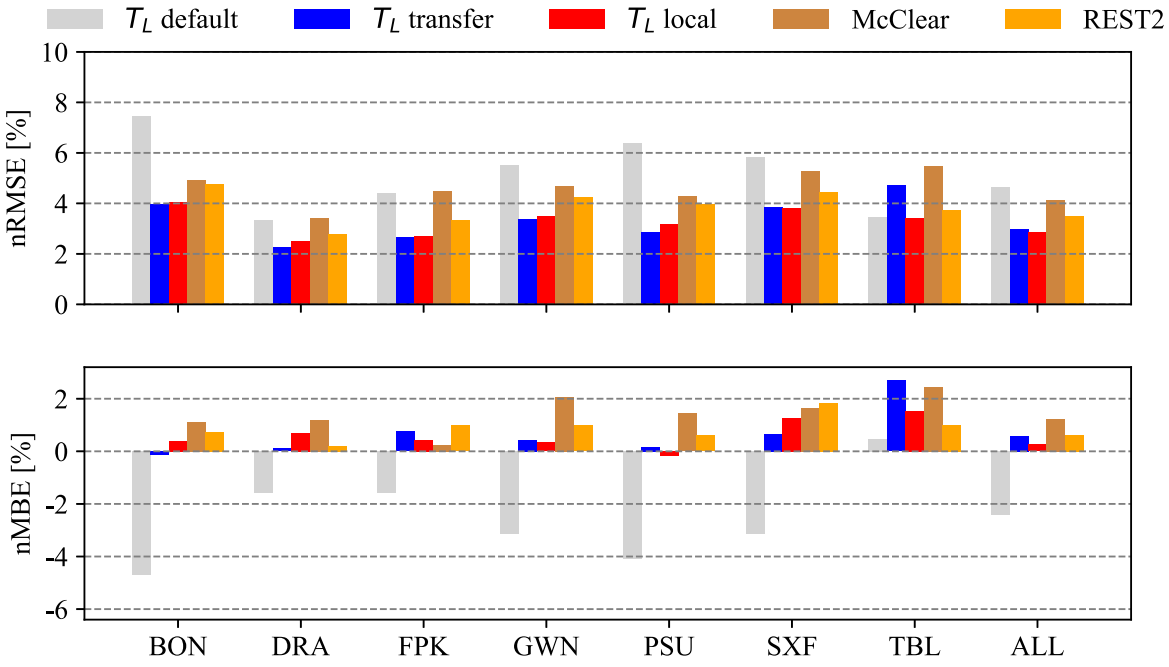

Fig. 6. The nRMSE and nMBE between the 5-min averaged GHIcs measurements and day-ahead forecasts using the day persistent method at SURFRAD stations in 2018–2020.

the REST2 has an nMBE of 0.79%. Compared with GHIcs estimations (except the default T_L method), GHIcs forecasts are generally associated with larger uncertainties of nRMSE as shown in Table 4. The comparison of day-ahead GHIcs forecasts at each SURFRAD station is illustrated in Fig. 6. The results of local and transferred T_L estimations produce comparable day-ahead GHIcs forecasts with physical models in terms of nRMSE and nMBE. Similarly, considering the comparable results for day-ahead GHIcs forecasts with less complexity, the uncertainty and time lag in obtaining atmospheric inputs for physical models, the methods of improved T_L estimations show potential in supporting solar forecasting applications.

4. Conclusions

In this work, we propose a transferable T_L estimation method for estimating GHIcs. The transferred T_L estimation model follows a similar methodology with the local T_L estimation model presented in [17]. Instead of using on-site solar irradiance data for model development in the local model, the transferable model is first trained and validated involving stations with sufficient data, and then applied at the location of interest for T_L estimation and thus the clear-sky irradiance. The main meteorological inputs of the T_L estimation model are ambient temperature, relative humidity, wind speed, and local atmospheric

pressure. As common meteorological information, they are easy to obtain and available at most of the weather stations. Both local and transferred T_L estimation models are applied at the SURFRAD stations, the performance of GHIcs estimation is evaluated with the on-site measurements and also compared with physical McClear and REST2 models.

The local T_L estimation method shows a high performance in GHIcs estimation with the nRMSE of 2.38% and the nMBE of 0.16%, the nRMSE and nMBE of GHIcs forecasts are 2.85% and 0.32%, respectively. The transferred T_L estimation model yields results with slightly larger divergences for both GHIcs estimations and forecasts. When applying the method at all the SURFRAD stations (excluding TBL), the overall nRMSE of GHIcs estimation is reduced from 4.99% to 2.44%, and the overall nMBE is decreased from −3.37% to 0.57% compared with the default PVLIB calculations. The result of GHIcs estimation based on the estimated T_L is also comparable with the physical clear-sky models, where the McClear yields an overall nRMSE of 3.32% and the nMBE is 2.10%, while the REST2 produces the overall result with an nRMSE of 2.55% and an nMBE of 1.30%. Considering the difficulties and uncertainties in forecasting the atmospheric inputs and meteorological data, we further compare the aforementioned methods for estimating the day-ahead GHIcs using a persistent way, where the day-ahead GHIcs is assumed as the same as the present day. The results show that the local T_L estimation has an overall forecasting with the nRMSE of 2.85% and nMBE of 0.32%, the transferred method for T_L estimation generates the GHIcs forecasts with an overall nRMSE of 2.99% and an nMBE of 0.74%, the McClear produces an nRMSE of 4.11% and an nMBE of 1.55%, while the REST2 yields a result with the nRMSE of 3.51% and the nMBE of 0.79%.

The transferred T_L estimation model does not yield similar results at TBL due to its unique climate features, so one recommendation for developing the transferable T_L estimation model is to use data from locations with similar climate conditions. Considering the improved GHIcs estimations and day-ahead forecasts, the comparable results with physical clear-sky models, and the complexity and difficulty in obtaining atmospheric inputs, both the local and transferred T_L estimation methods show a potential to support the solar resourcing and forecasting applications. The local T_L estimation method is suggested for stations with sufficient data, and the transferred T_L estimation model is therefore recommended for locations without adequate information.

CRediT authorship contribution statement

Shanlin Chen: Methodology, Software, Validation, Data curation, Visualization, Writing – original draft. **Zhaojian Liang:** Methodology, Validation, Writing – original draft. **Peixin Dong:** Methodology, Validation, Writing – original draft. **Su Guo:** Conceptualization, Resources, Funding acquisition. **Mengying Li:** Conceptualization, Resources, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the partial support from The Hong Kong Polytechnic University Grants P0035016 and P0038816, and the partial support from Jiangsu Province Science and Technology Department Grant BZ2021057.

Appendix

The equations used to derive ground truth T_L are adopted from [17].

$$T_L = \left[\ln\left(\frac{\mathrm{GHI_{cs}}}{c_1 \cdot I_0 \cdot \cos(\theta)}\right)/(-c_2 \cdot AM) - f_1\right]/f_2 + 1$$

with:

$$AM = \left(\frac{1}{\cos(\theta) + 0.50572 \cdot \left(6.07995 + (90-\theta)^{-1.6364}\right)}\right) \cdot \frac{P_a}{101325}$$

$$c_1 = 5.09 \cdot 10^{-5} \cdot h + 0.868$$

$$c_2 = 3.92 \cdot 10^{-5} \cdot h + 0.0387$$

$$f_1 = \exp(-h/8000)$$

$$f_2 = \exp(-h/1250)$$

where T_L is the Linke turbidity, $\mathrm{GHI_{cs}}$ [$\mathrm{W\,m^{-2}}$] is the measured GHIcs. c_1, c_2, f_1, f_2 are altitude-dependent coefficients, I_0 [$\mathrm{W\,m^{-2}}$] is the solar constant, θ [°] represents the solar zenith angle, AM is the absolute airmass, T_L is the Linke Turbidity factor, P_a [Pa] is the local atmospheric pressure, and h [m] is local altitude.

References

[1] J. Kleissl, Solar Energy Forecasting and Resource Assessment, Academic Press, 2013.
[2] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, J. Shelby, The national solar radiation database (NSRDB), Renew. Sustain. Energy Rev. 89 (2018) 51–60.
[3] D. Yang, Choice of clear-sky model in solar forecasting, J. Renew. Sustain. Energy 12 (2) (2020) 026101.
[4] V. Kallio-Myers, A. Riihelä, P. Lahtinen, A. Lindfors, Global horizontal irradiance forecast for Finland based on geostationary weather satellite data, Sol. Energy 198 (2020) 68–80.
[5] L.M. Ayompe, A. Duffy, An assessment of the energy generation potential of photovoltaic systems in Cameroon using satellite-derived solar radiation datasets, Sustain. Energy Technol. Assess. 7 (2014) 257–264.
[6] G.M. Yagli, D. Yang, O. Gandhi, D. Srinivasan, Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance? Appl. Energy 259 (2020) 114122.
[7] Y. Chu, M. Li, C.F.M. Coimbra, D. Feng, H. Wang, Intra-hour irradiance forecasting techniques for solar power integration: A review, iScience (2021) 103136.
[8] Z. Qu, A. Oumbe, P. Blanc, B. Espinar, G. Gesell, B. Gschwind, L. Klüser, M. Lefèvre, L. Saboret, M. Schroedter-Homscheidt, Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method, Meteorol. Z. 26 (1) (2017) 33–57.
[9] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, et al., McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions, Atmos. Meas. Tech. 6 (9) (2013) 2403–2418.
[10] C.A. Gueymard, REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation–Validation with a benchmark dataset, Sol. Energy 82 (3) (2008) 272–285.
[11] X. Zhong, J. Kleissl, Clear sky irradiances using REST2 and MODIS, Sol. Energy 116 (2015) 144–164.
[12] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, F. Vignola, A new operational model for satellite-derived irradiances: description and validation, Sol. Energy 73 (5) (2002) 307–317.
[13] C. Rigollier, M. Lefèvre, L. Wald, The method Heliosat-2 for deriving shortwave solar radiation from satellite images, Sol. Energy 77 (2) (2004) 159–169.
[14] Z. Qu, B. Gschwind, M. Lefèvre, L. Wald, Improving HelioClim-3 estimates of surface solar irradiance using the McClear clear-sky model and recent advances in atmosphere composition, Atmos. Meas. Tech. 7 (11) (2014) 3927–3933.
[15] S. Chen, Z. Liang, S. Guo, M. Li, Estimation of high-resolution solar irradiance data using optimized semi-empirical satellite method and GOES-16 imagery, Sol. Energy 241 (2022) 404–415.
[16] P. Ineichen, R. Perez, A new airmass independent formulation for the Linke turbidity coefficient, Sol. Energy 73 (3) (2002) 151–157.
[17] S. Chen, M. Li, Improved turbidity estimation from local meteorological data for solar resourcing and forecasting applications, Renew. Energy 189 (2022) 259–272.

[18] D. Yang, W. Wang, C.A. Gueymard, T. Hong, J. Kleissl, J. Huang, M.J. Perez, R. Perez, J.M. Bright, X. Xia, et al., A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality, Renew. Sustain. Energy Rev. 161 (2022) 112348.
[19] Y. Chu, M. Li, H.T. Pedro, C.F.M. Coimbra, Real-time prediction intervals for intra-hour DNI forecasts, Renew. Energy 83 (2015) 234–244.
[20] Y. Chu, M. Li, C.F.M. Coimbra, Sun-tracking imaging system for intra-hour DNI forecasts, Renew. Energy 96 (2016) 792–799.
[21] M. Schroedter-Homscheidt, F. Azam, J. Betcke, C. Hoyer-Klick, M. Lefèvre, L. Wald, E. Wey, L. Saboret, User's Guide to the CAMS Radiation Service (CRS): Status December 2020, Copernicus Atmosphere Monitoring Service, 2021.
[22] W.F. Holmgren, C.W. Hansen, M.A. Mikofski, pvlib python: A python package for modeling solar energy systems, J. Open Source Softw. 3 (29) (2018) 884.
[23] J. Remund, L. Wald, M. Lefèvre, T. Ranchin, J. Page, Worldwide Linke turbidity information, in: ISES Solar World Congress 2003, Vol. 400, International Solar Energy Society (ISES), 2003, pp. 13–p.
[24] A.F. Zambrano, L.F. Giraldo, Solar irradiance forecasting models without on-site training measurements, Renew. Energy 152 (2020) 557–566.
[25] J.A. Augustine, J.J. DeLuisi, C.N. Long, SURFRAD–A national surface radiation budget network for atmospheric research, Bull. Am. Meteorol. Soc. 81 (10) (2000) 2341–2358.
[26] X. Sun, J.M. Bright, C.A. Gueymard, B. Acord, P. Wang, N.A. Engerer, Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis, Renew. Sustain. Energy Rev. 111 (2019) 550–570.
[27] J.M. Bright, X. Sun, C.A. Gueymard, B. Acord, P. Wang, N.A. Engerer, Bright-Sun: A globally applicable 1-min irradiance clear-sky detection model, Renew. Sustain. Energy Rev. 121 (2020) 109706.
[28] C.M. Bishop, Pattern recognition, Mach. Learn. 128 (9) (2006).
[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[30] D. Yang, Validation of the 5-min irradiance from the National Solar Radiation Database (NSRDB), J. Renew. Sustain. Energy 13 (1) (2021) 016101.
[31] M. Chaâbane, M. Masmoudi, K. Medhioub, Determination of Linke turbidity factor from solar radiation measurement in northern Tunisia, Renew. Energy 29 (13) (2004) 2065–2076.
[32] T. Hove, E. Manyumbu, 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 52 (2013) 190–196.
[33] C.A. Gueymard, D. Yang, Worldwide validation of CAMS and MERRA-2 reanalysis aerosol optical depth products using 15 years of AERONET observations, Atmos. Environ. 225 (2020) 117216.
[34] M. Sengupta, A. Habte, S. Wilbert, C. Gueymard, J. Remund, Best practices handbook for the collection and use of solar resource data for solar energy applications, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2021.
[35] Y. Eissa, S.N. Beegum, I. Gherboudj, N. Chaouch, J. Al Sudairi, R.K. Jones, N. Al Dobayan, H. Ghedira, Prediction of the day-ahead clear-sky downwelling surface solar irradiances using the REST2 model and WRF-CHIMERE simulations over the Arabian Peninsula, Sol. Energy 162 (2018) 36–44.