

Review Article

Open-Source Satellite-Derived Solar Resource Databases Comparison and Validation for Indonesia

Anindio P. Harsarapama ¹, Dwi Riana Aryani,² and Dedy Rachmansyah³

¹PT Tractebel Engineering, Jakarta 12160, Indonesia

²Department of Electrical Engineering, Universitas Indonesia, Depok 16424, Indonesia

³PT Energia Prima Nusantara, Jakarta Timur 13930, Indonesia

Correspondence should be addressed to Anindio P. Harsarapama; anindio.prabu@gmail.com

Received 13 June 2019; Revised 16 May 2020; Accepted 7 September 2020; Published 22 September 2020

Academic Editor: Jędrzej Szmytkowski

Copyright © 2020 Anindio P. Harsarapama et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Solar resource data derived from satellite imagery are widely available nowadays, either as an open-source or paid database. This article is intended to assess open-source databases, which cover the region of Indonesia. Here, four known solar resource databases, which spatially cover the Indonesian archipelago, have been used, namely, Prediction of Worldwide Energy Resource (POWER), Surface Solar Radiation–Heliosat-East (SARAH-E), CM SAF Cloud, Albedo, Radiation edition 2 (CLARA-A2), and SolarGIS. In addition, a minor portion of the Meteororm database by Meteotest, around five sample points across Indonesia, has been assessed in terms of coherency to the four mentioned databases. Correlation coefficient and relative bias of the multiyear monthly mean annual cycle global horizontal irradiation (GHI) between pairs of databases are inspected. Three out of four databases are then validated through the available irradiation ground measurement data provided by the World Radiation Data Centre (WRDC). The correlation between each pair varies mostly between 0.7 and 1, which shows that the four databases to a certain extent agree on how the intermonthly variation would behave throughout the year. On the other hand, the validation result reveals that the three databases, i.e., POWER, CLARA-A2, and SARAH-E, are suffering from positive bias error ranging from 3% to 7%. Despite that fact, the correlation between measured and estimated values is still acceptable with SARAH-E showing the best performance among the three. Careful selections and adjustment enable the possibility of these databases to be utilized as a tool for depicting interannual and intermonthly variations of solar irradiation throughout the Indonesian archipelago.

1. Introduction

In the modern energy society, photovoltaic (PV) has emerged as one of the leading technologies contributing more than 20% of the worldwide total installed renewable energy power plant by the end of 2018. Around 57% of this PV capacity was installed within the region of Asia where China, Japan, and India are the three top contributors. Meanwhile, Indonesia is only able to contribute for less than 0.03% (around 60 MW) of the global total installed PV capacity [1]. For the last 10 years, PV installed capacity annual growth in Indonesia only counts at 5.2 MW on average. This rate is the fourth slowest growth among Southeast Asia countries despite being the

largest economic power within the region. Although the price of solar panels has dropped significantly over the last decade, the growth of the PV industry in Indonesia does not seem to get much impact from it. Other issues are frequently addressed regarding this slow development such as ineffective support regulation, opaque framework, low IRR for the developer, bank reluctance on financing solar developers, low electricity tariff, and many more [2, 3].

Nevertheless, the fact that the solar PV project requires high initial investment cost remains. Employing reliable solar resource data with a long-term historical database enables a more detailed solar resource analysis such as inquiring how solar resources on potential power plants would

behave throughout the investment years. This kind of analysis will help investors to understand more about the uncertainty involved in solar project investment. Likewise, the developer would also be able to refine this analysis into a P50/P90 uncertainty report, which is required to obtain funding from the investors. The existence of reliable solar resource data without a doubt is an important aspect of stimulating the growth of the solar PV project within a region.

Several studies have been conducted to estimate local solar irradiation in some regions in Indonesia via ground measurements [4–10]; unfortunately, most of the measurements were only conducted for a short period. Some works present a thorough summary of solar resources data acquired from ground measurements throughout the Indonesian archipelago between the mid-70s and mid-90s periods [11, 12]. A couple of models also have been proposed as tools to estimate local solar irradiation values, e.g., artificial neural network (ANN) [13], Weather Research and Forecasting (WRF) [7], the stochastic model [5, 6], and the physics model [10]. Among those prior studies, one work has tried to build a complete map of Indonesian solar resources based on the surface and solar energy (SSE) database whose data are mainly derived from images of atmosphere captured by satellite [14].

Large area coverage of satellite images makes them superior in acquiring data over certain regions instantaneously, while the ground-based sensor only acquires data within a perimeter of the sensor's location. Moreover, the captured images are well-archived, which enables the possibility of building a proper solar resources database from archived historical images. Due to these pronounced advantages, most of the solar resources databases that exist today incorporate satellite images into their algorithms either as a primary or secondary input variable.

Unlike developed countries who put some interest in developing their solar resource database and provide the service for free, Indonesia currently is only able to rely on the data that have been provided by the third party. As a support to the development of renewable energy worldwide, some research institutions funded by developed countries have been providing worldwide open-source solar resource databases service for quite sometimes, while some private institutions charge some fees for the solar resource data services. Despite having their limitations, such as temporal and spatial coverage ranges and resolutions, the open-source databases are more preferable for solar PV players in Indonesia since they do not incur any additional cost.

The aforementioned limitations of several open-source solar resources databases will be thoroughly discussed within this article, and finally, a common ground map where all open-source solar resources databases agree on several statistical measures limit is drawn out. Hopefully, the map can act as a guide for solar PV players in Indonesia to utilize the open-source databases and adjust them appropriately to any specific needs.

2. Methods

2.1. Satellite-Derived Solar Resources Databases. As mentioned previously, open-source solar databases addressed within this article are all satellite-derived. In general, each database only differs in terms of satellite selection as images producer and surface irradiation derivation algorithm. The basic derivation algorithm usually starts with predicting the total amount of solar radiation that would reach the Earth's surface if the sky is clear (commonly known as clear-sky irradiation). The prediction applies some physics modeling that normally requires the estimated value of solar radiation on the top of the atmosphere (known as TOA irradiation) and some atmospheric parameters such as aerosols optical depth (AOD), precipitable water vapor (PWV), and others [15].

Upon predicting clear-sky irradiation, the algorithm continues by deciding the inclusion of additional effects caused by cloud conditions. The cloud condition is often described by several cloud properties retrieved from satellite images. The cloud properties retrieval algorithm inspects each pixel (normally in form of a square) of the image and determines the most appropriate cloud properties for that particular pixel. The number of pixels within an image determines the level of detail on the captured 2D space that can be derived or simply known as spatial resolution. While the number of captured images over a period of time determines the level of 1D time detail that can be inferred from a group of images or often known as temporal resolution. The quality of the solar resources databases also depends on these resolution values, higher resolutions mean closer the database to resemble continuous behavior of true irradiation values.

The first solar resource database is Prediction of Worldwide Energy Resource (POWER) developed by NASA Langley Research Centre. This database is an outgrowth of the previously mentioned SSE database with a similar spatial resolution. Unlike most solar resources databases, POWER historical data are built from 3 different sources for each respective period, GEWEX SRB 3.0 (1983–2007), FLASHFlux v2 (2008–2012), and FLASHFlux v3 (2013-present) [16]. Due to this discontinuity, it is not recommended to use this database for analysis that encompasses a change of sources. Validation of daily global irradiance value from the aforementioned sources relative to ground measurements stations (BSRN, ARM, and Ocean Buoy) shows that FLASHFlux product is slightly better than GEWEX SRB product in terms of the root mean square error (RMSE) by about 0.5%, while in average worldwide, the relative root mean square error (rRMSE) value is around 18% [17].

The second database is Surface Solar Radiation–Heliosat-East (SARAH-E) provided by the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) [18]. Since the easternmost view of METEOSAT-East Satellite is limited at 128 E; therefore, only around 70% of the Indonesian archipelago is spatially covered by this database. Despite having this limitation, SARAH-E offers high spatial resolution and a long

historical dataset, which is beneficial for conducting thorough solar resources analysis. Unlike the POWER database, global scale validation of SARA-H-E was only conducted on 8 ground measurement stations and included a single year instantaneous global irradiance dataset. The result reveals that the average rRMSE value is around 33%, while the relative mean bias error (rMBE) shows inconsistent patterns across the eight stations with some high positive and negative bias [19]. Some local validations within the region of Asia are also available, such as India [20] and China [21].

The third database is Cloud, Albedo, and Surface Radiation edition 2 (CLARA-A2), also developed by CM SAF [22]. Unlike SARA-H-E who derived the data from the MVIRI sensor of METEOSAT-E, CLARA-A2 utilized the AVHRR sensor mounted on polar-orbiting satellites operated by NOAA and MetOp polar orbiter satellites operated by EUMETSAT. CLARA-A2 daily mean global irradiance dataset has been validated against the data provided by the BSRN ground measurement stations network. The result shows that the database on average is suffering from negative bias for about 1.7 W/m^2 [23], while a local validation against Chinese ground measurement stations reveals that the database overestimates the global irradiance value by about 10 W/m^2 [21].

The fourth solar resources database is provided by SolarGIS, which comes with paid and free versions. An open-source version of the database is presented in form of high spatial resolution local solar resource map product assisting the request of the World Bank's Energy Sector Management Assistance Program (ESMAP) as the funding source [24]. The map itself is presented as a multiyear monthly mean annual cycle of irradiation values (as shown in Figure 1); therefore, it does not hold any information regarding the historical datasets as the proper solar resources database should do. In exchange, it offers a very fine spatial resolution dataset, which might be a good alternative for solar players in Indonesia, e.g., government, NGO, solar PV developer, consultant, R&D entity (university and research centre), and manufacturer, who wish to use Typical Meteorological Year (TMY) solar radiation datasets. Local validation of SolarGIS global irradiance database of Indonesia shows that the rMBE and rRMSE values are 0.6% and 2.5% for Bukit Kototabang Station, while for Palangkaraya Station, the values are around -4.6% and 8% , respectively. Since Indonesia is located in the humid tropical climate region, the rMBE value could go as high as 8% outside the measurement sites [25]. Important information regarding these databases is summarized in Table 1.

Nowadays, the initial solar resource database is usually included within common PV design software, e.g., PVsyst with Meteonorm and HOMER with NASA-SSE. Since many Indonesian solar developers are already familiar with both software types, the coupled solar resource databases within will be assessed in terms of coherency with the four aforementioned satellite-derived databases; NASA-SSE is excluded since it has been represented by POWER database. Similar to SolarGIS, Meteonorm 7.2 brought by Meteotest also offers free and paid versions. The open-source version

covers a multiyear monthly mean annual cycle irradiation dataset, limited at a certain maximum number of sites. The algorithm of Meteonorm is a mixed version of ground measurement data interpolation and satellite-derived process, which roughly follows this rule [26]:

- (1) If the closest station is within 30 km (for Europe 10 km) radius, the data will be either similar to the station's measured data or derived from the interpolation of nearby stations
- (2) If the closest station is within a radius of 30–300 km, the data will be a mixture of interpolation and satellite-derived data
- (3) If there is no station within a 300 km radius, the data will be purely derived from a satellite

To stand in equal terms with the other databases in this article, only data generated from the last term of the aforementioned rule is considered. So far, Meteonorm Satellite Irradiation (MNSI) has only been validated for Europe, Africa, and Middle East regions whose result has found that rRMSE of hourly values lies around 12–25% [27].

2.2. Solar Radiation Ground Measurement Stations.

Besides comparing, the article also intends to check these databases locally whose result might be able to reveal some specific characteristics that the user needs to be aware of. The local validation process of the solar resource database requires a reliable ground measurement dataset, and in this article, daily sum global irradiation data provided by World Radiation Data Centre (WRDC) network stations [28] was selected. As the basic requirement for the validation process, the actual measurement time should coincide with the available timeframe of the satellite database. Even though there are 5 WRDC stations located in Indonesia, only one station satisfies this requirement. Therefore, other nearby stations from surrounding countries that fulfill the timeframe are also included within the validation process. A list of stations that comply with coincidence time terms is presented in Table 2.

2.3. Physics Parameter and Statistical Measures.

Variations of global horizontal irradiation (GHI) value has been proven able to represent the energy yield variations of the flat PV system [29], which is the most frequent PV installation type in Indonesia nowadays. At least for the present, GHI alone is sufficient to be the main parameter for comparison and validation processes. The main interest focuses on daily, monthly, and annual behavior of flat solar PV energy yield potential; therefore, the main unit of the parameter is initially defined as kWh/m^2 . The database usually comes with its output parameter, mostly either as daily total incident irradiation or daily mean of valid instantaneous irradiance, which later needs to be adjusted into the energy unit.

The first thing that has to be verified is the GHI annual intermonth trend of each database, which is a cycle formed by a monthly average of irradiance data within a year. The

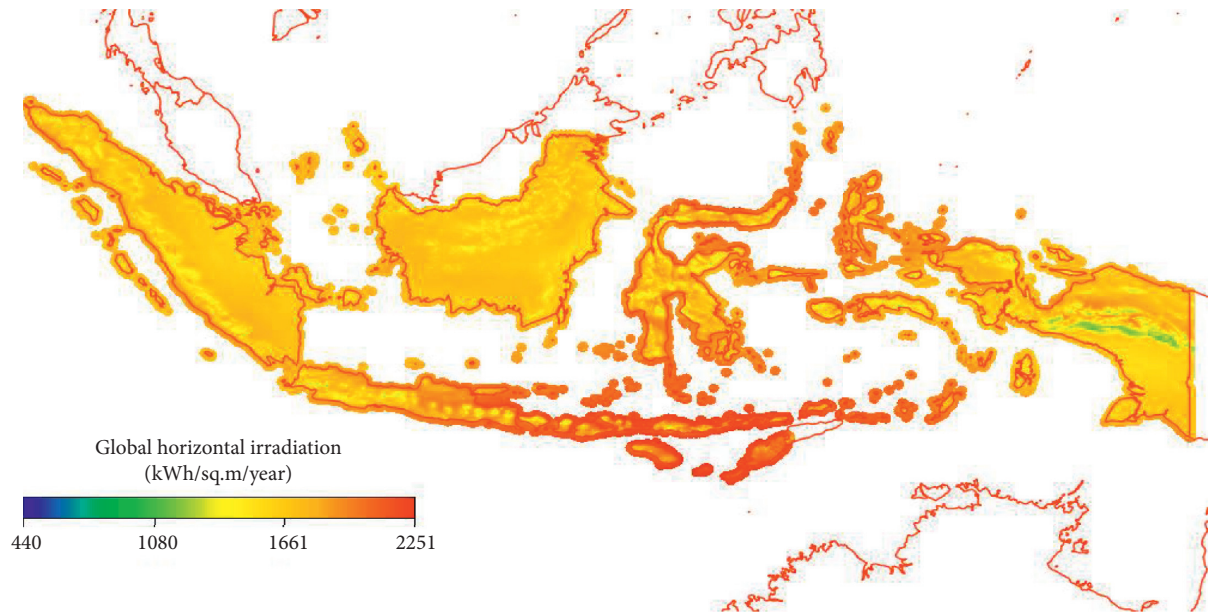


FIGURE 1: SolarGIS long-term mean surface radiation in Indonesia.

TABLE 1: The general summary of open-source satellite-derived solar resource database for Indonesia.

	SARAH-E	SolarGIS
Spatial resolution	0.05	0.0025
Temporal resolution	Hourly	Monthly average (free edition)
Spatial coverage	40 S-62 30' N, 65 W-128 E	Global
Temporal coverage	1999–2016	1999–2016/2007–2016 (average)
Derivation algorithm	MAGICSOL	SolarGIS algorithm
Satellite	METEOSAT by EUMETSAT	METEOSAT, GOES, and Himawari
Database access	https://wui.cmsaf.eu/	https://solargis.com/maps-and-gis-data/download/indonesia
	POWER	CLARA-A2
Spatial resolution	1	0.25
Temporal resolution	Hourly	Daily
Spatial coverage	Global	Global
Temporal coverage	1983–present	1982–2015
Derivation algorithm	LPSA, CERES-2 cloud property retrieval MAGIC, and SAFNWC	
Satellite	Aqua and terra by NASA	POES by NOAA and EUMETSAT
Database access	https://power.larc.nasa.gov/	https://wui.cmsaf.eu/

TABLE 2: Summary of the WRDC station network used for validation.

Measurement stations	Period	Coordinates	Elevation (m)
Brunei Airport–Brunei Darussalam	1987–1995	4 56'N, 114 56'E	22
Christmas Island–Australia	1996–1997	10 27'S, 105 41'E	262
Changi Airport–Singapore	1984–2004	1 22'N, 103 59'E	15
Bayan Lepas–Malaysia	1984–2009	5 18'N, 100 16'E	3
Pengkalan Chepa–Malaysia	1984–2009	6 10'N, 102 17'E	5
Subang–Malaysia	1984–2009	3 07'N, 101 33'E	17
Kuching–Malaysia	1989–2009	1 29'N, 110 21'E	20
Kota Kinabalu–Malaysia	1989–2007	5 56'N, 116 03'E	2
Bukit Kototabang–Indonesia	1996–2017	0 12'S, 100 19'E	864

trend similarity of two databases is estimated by employing Pearson correlation coefficient (r) to the daily long-term average for each month within a year; this means a multiyear average of typical daily irradiation data for each month, which is then arranged as a year dataset. Since each database has a different spatial and temporal resolution, some adjustments such as data downscaling and timeframe filtering are required to compare both datasets in an equal manner. Spatial downscale via averaging reduces variations within a dataset; therefore, the coefficient of variation (CoV) [30] is employed to estimate how much has been lost; E_i and E_p terms stand for GHI value for each pixel and mean GHI value of pixels after downscaling, respectively. Temporal CoV (CoV_T) is also added as an extra information to analyze some disagreement on GHI annual intermonth trend between the four databases as shown on the later section; E_j and E_q terms stand for multiyear average of typical daily GHI value for each month and mean of E_j in a year period. Since the four databases have different temporal coverage, the 1999–2016 period has been selected to analyze CoV_T for POWER, SARAH-E, and SolarGIS, while for CLARA-A2, the 1999–2015 period has been chosen:

$$r(\text{db1}, \text{db2}) = \frac{\text{CoV}(\text{db1}, \text{db2})}{\sigma_{\text{db1}} \sigma_{\text{db2}}}, \quad (1)$$

$$\text{CoV} = \frac{\left[\sum_{i=1}^{i=n} (E_p - E_i)^2 / n \right]^{1/2}}{E_p}, \quad (2)$$

$$\text{CoV}_T = \frac{\left[\sum_{j=1}^{j=12} (E_q - E_j)^2 / 12 \right]^{1/2}}{E_q}, \quad (3)$$

$$B_m = \frac{\sum_{i=1}^{i=12} (\text{db2}_i - \text{db1}_i)}{12}. \quad (4)$$

Another statistical measure that has been employed is the mean bias (B_m) between one database relative to another. Computation steps of mean bias are similar to mean bias error (MBE). However, instead of comparing the estimated values against true values of the parameter, it compares two estimated values relatively. The measure is computed on a yearly time frame for each month and then mapped across the Indonesian archipelago to explore the major behavior of a particular database compared to another. Unlike r and CoV, B_m is not a dimensionless parameter, and it holds kWh/m²/day as the unit. Besides, a rule where a database with the lowest spatial resolution acts as a reference for B_m computation has been set. The reference data (db1) simply replace “measured data” term on widely known MBE formula, while the db2 term still acts as estimated data. As for validation purposes, common statistical measures such as RMSE and MBE are employed; E_e and E_m terms on equations (4) and (5) stand for estimated and measured solar energy resources, respectively:

$$\text{RMSE} = \left[\frac{\sum_{i=1}^{i=n} (E_e - E_m)^2}{n} \right]^{1/2}, \quad (5)$$

$$\text{MBE} = \frac{\sum_{i=1}^{i=n} (E_e - E_m)}{n}, \quad (6)$$

At the last step, both computed r and B_m are employed as agreement constraints to depict a spatial common ground for the four databases. The values are set to be $r > 0.8$ for “very strong” correlation [31] and $-x < B_m < x$, where $x = 0.4$ kWh/m²/day; selected x value covers roughly 90% of overall database pairs mean bias distribution. Generated common ground map is acting as a reminder for solar players in Indonesia that there are some regions where the databases statistically exhibit weak agreement or even do not agree on each other. In these regions, the user is advised to be careful when using the data provided by one of the aforementioned databases.

3. Results and Discussion

3.1. Databases Validation against WRDC Stations. The validations presented in Table 3 are only conducted on a database, which provides free access to at least a daily basis historical dataset, i.e., POWER, CLARA-A2, and SARAH-E. Overall, POWER and SARAH-E exhibit consistent trends across all validation stations both in terms of time series trend and bias error direction, which consistently overestimate their GHI data. Meanwhile, CLARA-A2 performs relatively well on time series coherency, but it does not have any tendency toward either positive or negative bias for every measurement station. It might be related to strangely high initial spatial variations that will be discussed on the CoV topic. Another interesting fact is related to CLARA-A2 low performance (especially on RMSE and r) compared to POWER, which fundamentally offers a lower spatial resolution. Both SARAH-E and CLARA-A2 similarly utilize the MAGIC algorithm for estimating clear-sky irradiance; therefore, the low performance issue might be related to how the database handles cloudy conditions.

According to the result, it is no doubt that SARAH-E outperforms the other two databases especially on RMSE and time series correlation measures. However, there is a concern regarding sudden performance drops on 2 validation sites, namely Kuching and Kota Kinabalu. Similarly, POWER and CLARA-A2 also suffer from these performance reductions, but somehow, the drops are not as pronounced as the one on SARAH-E. The two stations are located at the Malay side of Kalimantan, close to Indonesian regions where the majority of databases do not agree on how the GHI annual intermonth trend should be. Therefore, the performance drop could also be triggered by low/no-temporal variation dataset as elaborated on the next section or transnational forest fire pollution, which has been proven to be one of the causes of air quality degradation in Malaysia [32]. Databases validations on Christmas Island station seem to exhibit unsatisfactory MBE and RMSE values, most likely due to a limited amount of validated data. Meanwhile, the

TABLE 3: Daily energy resource validation result of POWER, CLARA-A2, and SARAHE against data retrieved from WRDC network stations.

	Number of pairs	AVG (kWh/m ²)	RMSE (kWh/m ²) and rRMSE	MBE (kWh/m ²) and rMBE	r
POWER					
Brunei Airport	2658	5.00	1.20 (24.0%)	0.44 (8.7%)	0.56
Christmas Island	485	4.74	1.50 (31.6%)	1.77 (37.4%)	0.85
Changi Airport	7366	4.55	1.04 (22.8%)	0.03 (0.6%)	0.72
Bayan Lepas	8917	4.96	0.90 (18.2%)	0.19 (3.8%)	0.80
Pengkalan Chepa	8307	4.95	1.01 (20.3%)	0.34 (6.9%)	0.80
Subang	8577	4.38	0.99 (22.6%)	0.53 (12.0%)	0.70
Kuching	7059	4.17	1.06 (25.5%)	0.50 (11.9%)	0.69
Kota Kinabalu	6646	5.02	0.99 (19.7%)	0.09 (1.8%)	0.66
Bukit Kototabang	7107	4.35	0.89 (20.5%)	0.40 (9.2%)	0.73
CLARA-A2					
Brunei Airport	1655	5.03	1.61 (32.1%)	0.42 (8.3%)	0.41
Christmas Island	319	4.81	1.41 (29.3%)	2.30 (47.9%)	0.84
Changi Airport	5183	4.53	1.23 (27.3%)	0.13 (2.8%)	0.72
Bayan Lepas	6969	4.95	1.30 (26.3%)	-0.20 (-4.1%)	0.65
Pengkalan Chepa	6572	4.94	1.16 (23.6%)	0.57 (11.6%)	0.78
Subang	6456	4.40	1.29 (29.4%)	-0.02 (-0.4%)	0.60
Kuching	5710	4.17	1.35 (32.4%)	0.43 (10.3%)	0.58
Kota Kinabalu	5381	5.00	1.39 (27.8%)	-0.42 (-8.5%)	0.57
Bukit Kototabang	6197	4.38	1.08 (24.6%)	0.32 (7.3%)	0.65
SARAHE					
Brunei Airport	x	x	X	x	x
Christmas Island	x	x	X	x	x
Changi Airport	2045	4.49	0.62 (13.7%)	0.26 (5.9%)	0.92
Bayan Lepas	3652	4.95	0.60 (12.1%)	0.34 (6.8%)	0.92
Pengkalan Chepa	3427	5.02	0.53 (10.6%)	0.12 (2.4%)	0.95
Subang	3154	4.44	0.74 (16.6%)	0.44 (10.0%)	0.86
Kuching	3607	4.16	1.06 (25.4%)	0.51 (12.3%)	0.70
Kota Kinabalu	3129	5.05	0.97 (19.3%)	0.19 (3.7%)	0.72
Bukit Kototabang	6010	4.39	0.70 (16.0%)	0.27 (6.2%)	0.84

validations on Brunei Airport station reveals that POWER and CLARA-A2 estimated data do not perform well on predicting local time series trends. Since the last data time stamp was recorded in 1995, it was less likely caused by transnational forest fire effects. The most probable reason should be related to the low/no-temporal variation within the satellite-derived and/or ground-measurement datasets. Bukit Kototabang station is located relatively high on altitude compared to other mentioned stations; therefore, the result of the validations could represent database performance on tropical hilly conditions. Despite being on a different condition, the three databases are still performing quite well with SARAHE exhibiting the best estimation, while CLARA-A2 is still behind POWER on three statistical measures. By excluding Christmas Island result validation, it is clear that all validated databases overestimate the true GHI value by about 0.07–0.32 kWh/m²/day (3–7% rMBE) on average, which is roughly in line with the upcoming statement regarding the position of true GHI value among satellite-estimated GHI dataset.

3.2. Comparison between Databases. As implicitly stated by the previous equation, the correlation coefficient explains the relation of two variables that move together in parallel, or in this case, the two GHI databases intermonth variations.

Two representative results, POWER vs SolarGIS and CLARA-A2 vs SolarGIS, presented in Figure 2 describe that every pairing case exhibits roughly similar GHI annual intermonth trend across the country indicated by r value close or equal to +1. Some disagreements on the GHI annual intermonth trend largely occur at the centre region of Kalimantan and are vaguely visible at the southern and central parts of Sumatera. While the exact cause of the discrepancies has yet to be revealed, some clues from prior works might provide one of the possible answers.

Looking closely at the distribution of low/no correlation regions, there are two possibilities that might be able to explain this situation. The first cause should be related to the low/no temporal variation cells within the dataset; without temporal variation, the correlation itself would be minimum or even do not exist since the covariance between pair of databases should also be minimum. The maps of CoV_T in Figure 3 represent the magnitude of temporal variation within each cell of each database with their own original spatial resolution. As can be seen that regions with low/no temporal variation are roughly similar between the three representative databases with majority of low/no correlation cells coinciding with low/no temporal variation cells. Though one question still remain, only some part of low/no temporal variation cells trigger low/no correlation cells, while the rest still exhibit some correlation between databases.

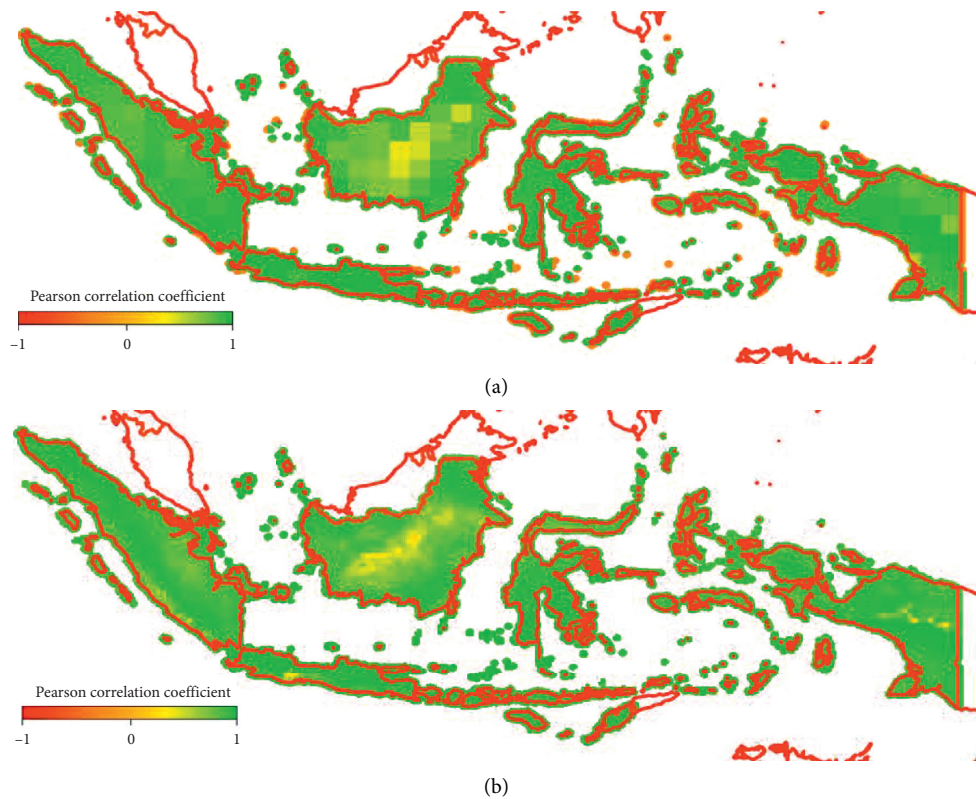


FIGURE 2: Representative figure of correlation coefficient between two databases. (a) POWER vs SolarGIS. (b) CLARA-A2 vs SolarGIS.

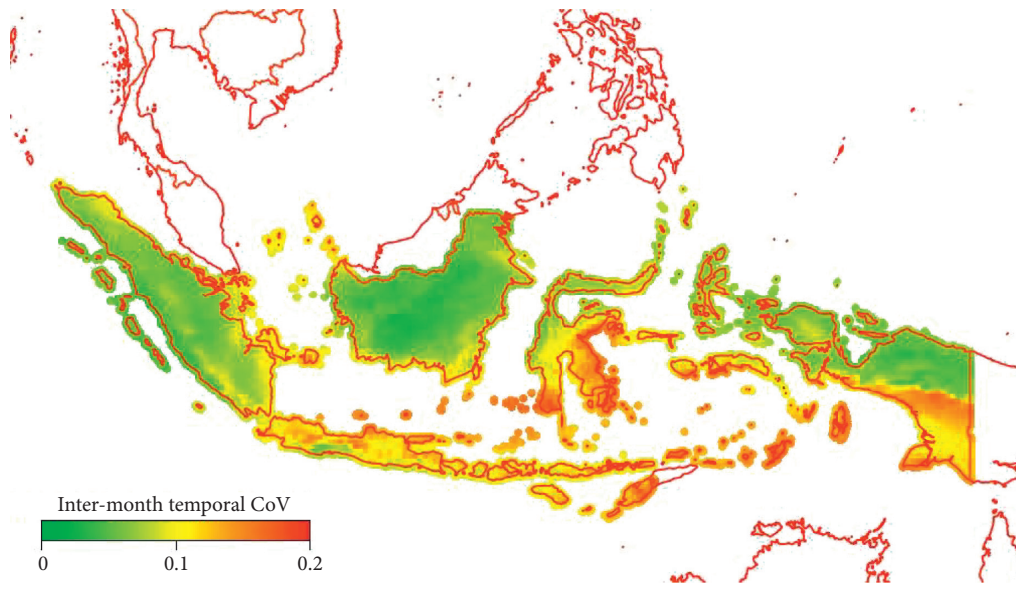
The second most probable cause is the disturbance of aerosol temporal trend within the atmosphere, which is most likely related to a natural or man-made forest fire that occurs frequently during the dry season. Citing a summary data released by Indonesian Ministry of Living Environment and Forestry (MENLHK) within the 2014–2016 period, Sumatera, Kalimantan, and Papua were the three top regions suffered from forest fire phenomenon, with each contributing 38%, 33%, and 17% of total around 3 million hectares of the burnt area across the country [34]. Moreover, locations of specific provinces where large portions of forest fire occurred coincide well with low/no correlation regions, e.g., Central Kalimantan, Riau, South Sumatera, and Papua. It is suspected that the combination of regions with low/no temporal variation existence and the absence of aerosol temporal trends due to the forest fire phenomenon triggers the low/no correlation regions within the result.

CM SAF surface radiation products have been validated on several ground measurement sites around the world; among those sites, there are some cases where the local climate is prone to aerosol load changes such as China [21], India [20], and Eastern Mediterranean [33]. These three cases reported that CM SAF surface radiation products exhibit some discrepancies, and the most probable explanation so far is related to the absence of aerosol temporal trends, which is inherited from algorithm as stated on both databases manual [18, 22]. The reasoning could also be applied to the present issue regarding solar databases

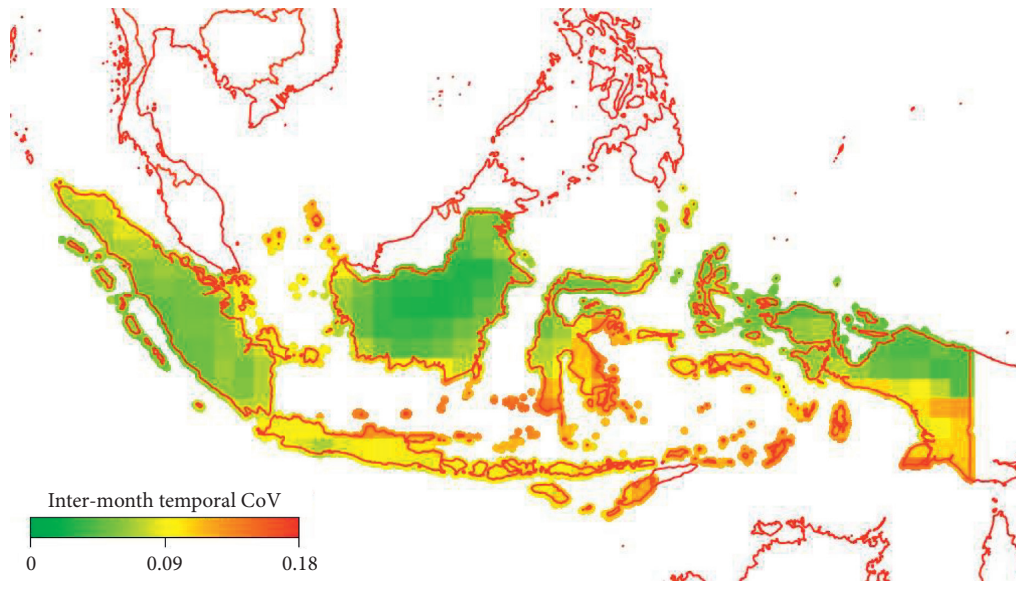
disagreement of the GHI annual intermonth trend for some regions in Indonesia.

The spatial distribution of B_m is presented in the box plot format instead of the map as shown in Figure 4. The box plot explains the upper and lower limits of the dataset through its caps (ends of whiskers), which has been set to be 95% and 5%, respectively, while the upper and lower border of the box informs Q3 and Q1 positions along the axis. Points outside the caps (red diamonds) are usually called outliers, or in this context, they are interpreted as the region with the mean bias value exceeds either the upper or lower limit value. As a reminder, zero axis of the box plot is relative to reference definition, i.e., POWER [1–3], CLARA-A2 [4, 5], and SARAH-E [6].

As depicted by the first three box plots, the POWER database overvalues other databases by about 0.5–0.22 kWh/m²/day on average. More than 90%, 60%, and 55% of POWER data estimations overvalue estimated GHI of SolarGIS, SARAH-E, and CLARA-A2, respectively. Meanwhile, when CLARA-A2 is appointed as reference (box plot 4 and 5), more than 95% and 75% of its data estimations overvalue SolarGIS and SARAH-E by 0.22 and 0.14 kWh/m²/day, respectively. The last box plot reveals that around 70% of SARAH-E data estimations overvalue SolarGIS by 0.09 kWh/m²/day. Sorting the result in the ascending order will lead to following rank configuration, SolarGIS < SARAH-E < CLARA-A2 < POWER, which overall explains the relative position of each database in term of estimated GHI magnitude.



(a)



(b)

FIGURE 3: Continued.

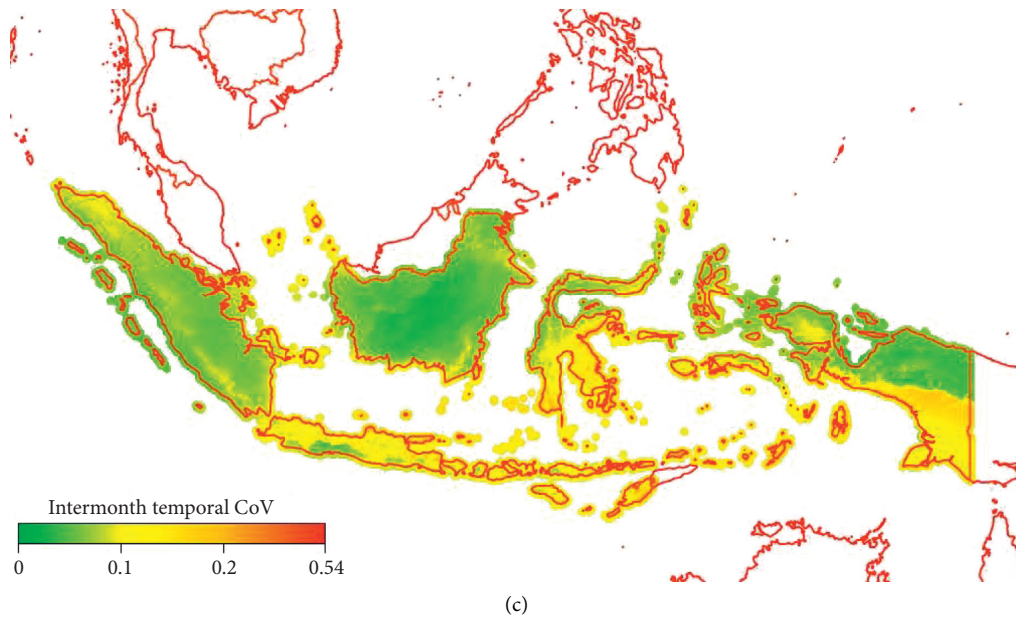


FIGURE 3: CoV_T on the 1999–2016 period. (a) CLARA-A2. (b) POWER. (c) SolarGIS.

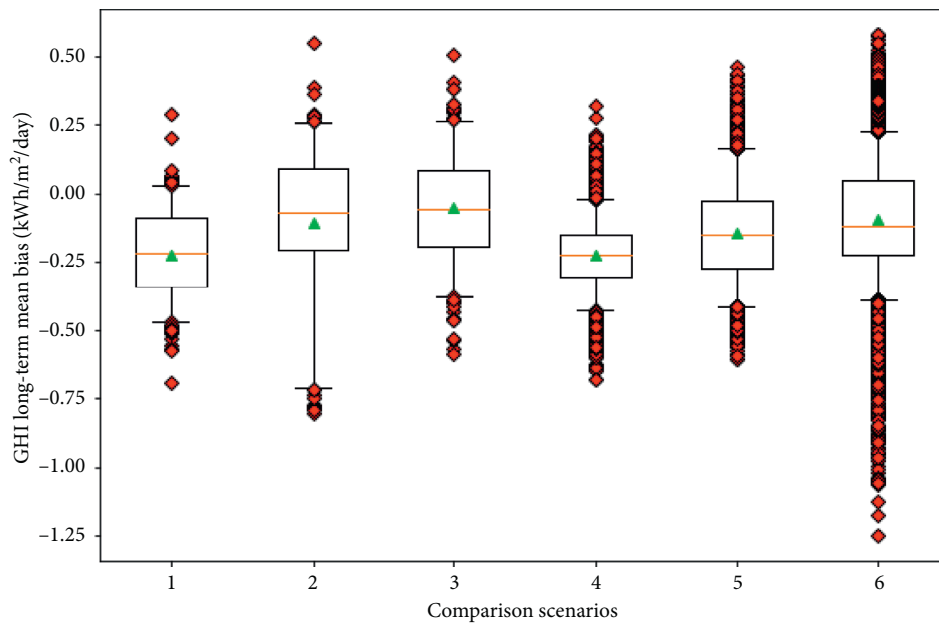


FIGURE 4: GHI mean bias between two databases. (1) POWER vs SolarGIS. (2) POWER vs SARAH-E. (3) POWER vs CLARA-A2. (4) CLARA-A2 vs SolarGIS. (5) CLARA-A2 vs SARAH-E. (6) SARAH-E vs SolarGIS. Orange line, median and green triangle, mean.

The loss of spatial variations during the downscaling process is recorded and presented in Figure 5, which has been sorted to follow the scale of spatial resolution reduction in the descending order. By comparing 95% data of each plot (upper cap), it is clear that the box plots exhibit expected trends where the CoV value decreases as the scale of spatial resolution reduction are getting smaller. The CoV reduction of the downscaling process resembles a linear behavior with the different decreasing rate (gradient) between databases, i.e., SolarGIS (upper caps 1, 2, and 4) and SARAH-E (upper caps 3 and 5). The plots also reveal strange facts regarding

variations in loss of CLARA-A2. Somehow, with a much smaller downscaling factor (upper cap 6), it is suffering roughly similar or even more spatial variations loss compared to other databases. It implies that, initially, the CLARA-A2-estimated GHI dataset already has high spatial variations. In addition, all box plots agree to have right-skewed data distribution, which implies a minor portion (less than 25%) of CoV tends to spread over high values.

In addition, the spatial CoV loss of two representative figures that have undergone the strongest downscaling (i.e., SolarGIS to CLARA-A2 and SolarGIS to POWER) is

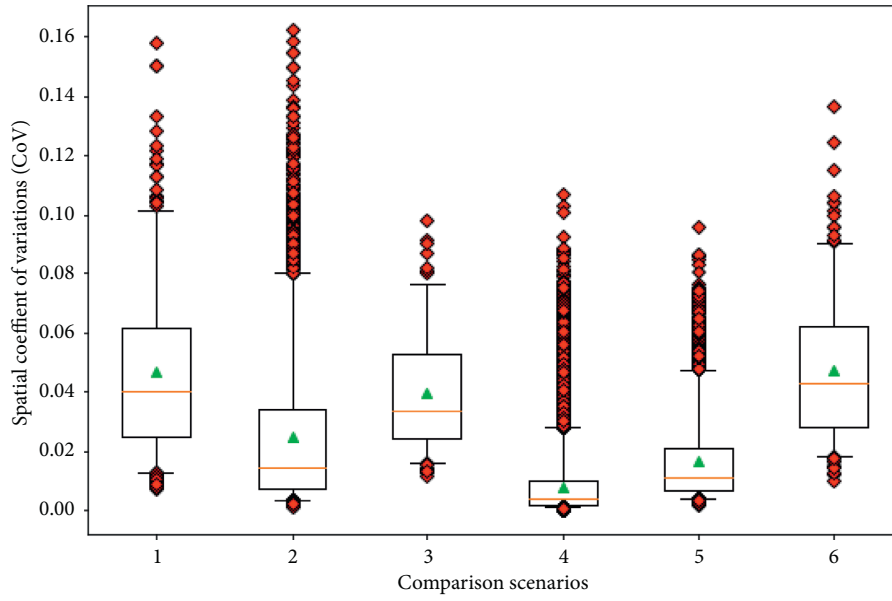


FIGURE 5: Spatial CoV lost due to downscaling. (1) SolarGIS in comparison with POWER. (2) SolarGIS in comparison with CLARA-A2. (3) SARAH-E in comparison with POWER. (4) SolarGIS in comparison with SARAH-E. (5) SARAH-E in comparison with CLARA-A2. (6) CLARA-A2 in comparison with POWER. Orange line, median and green triangle, mean.

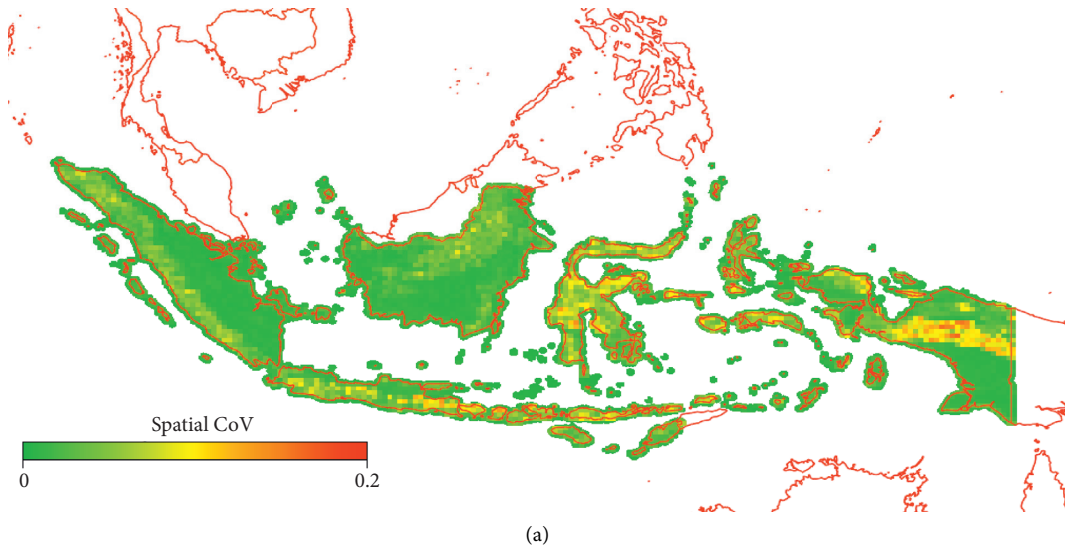


FIGURE 6: Continued.

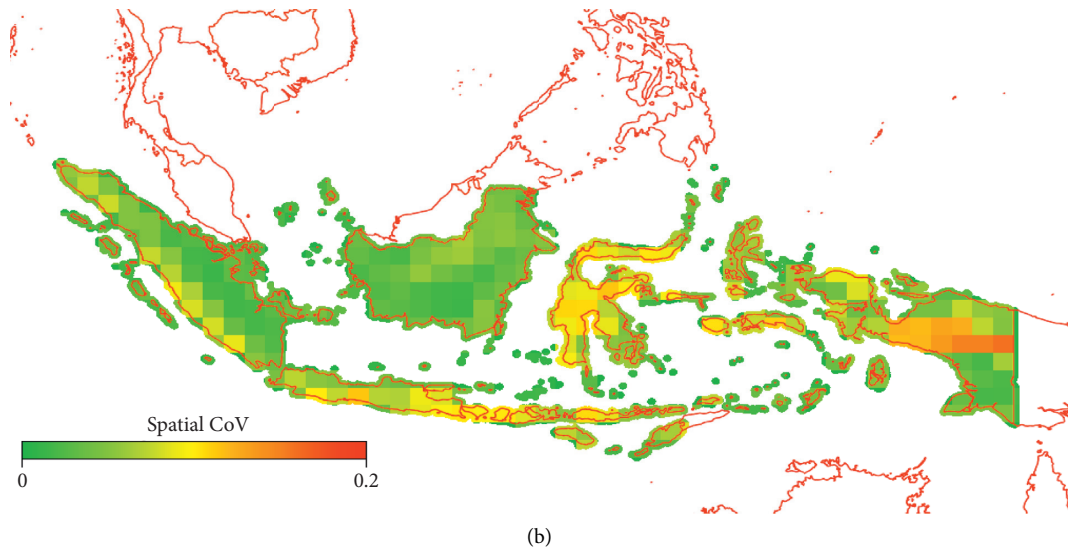


FIGURE 6: The spatial CoV loss figure due to downscaling. (a) SolarGIS to CLARA-A2. (b) SolarGIS to POWER.

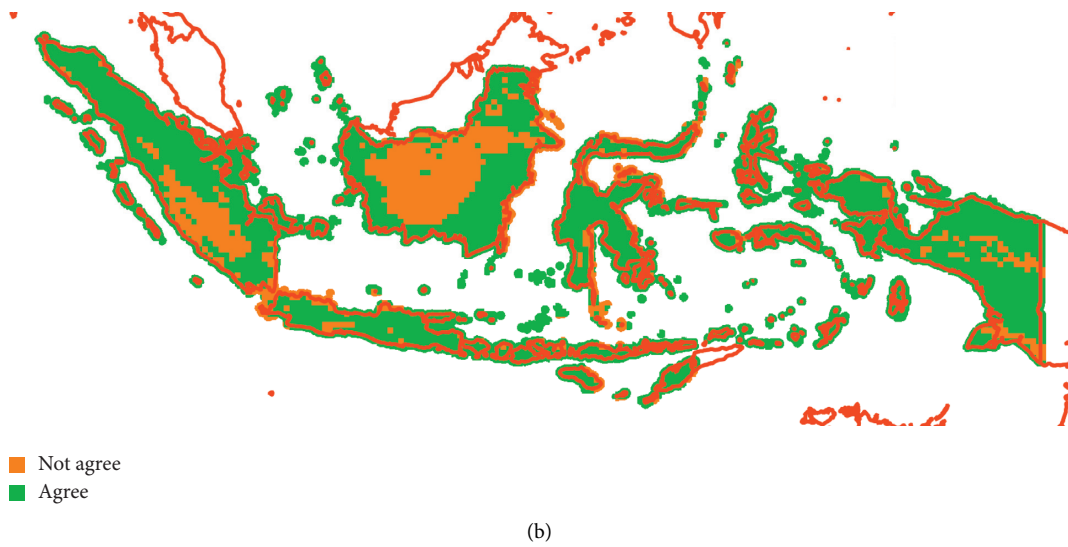
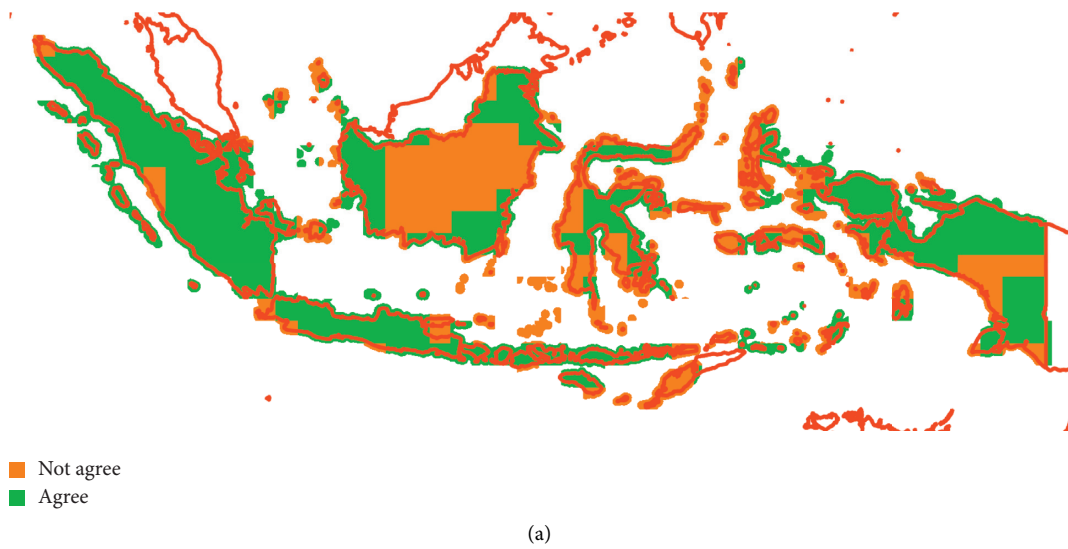


FIGURE 7: Continued.

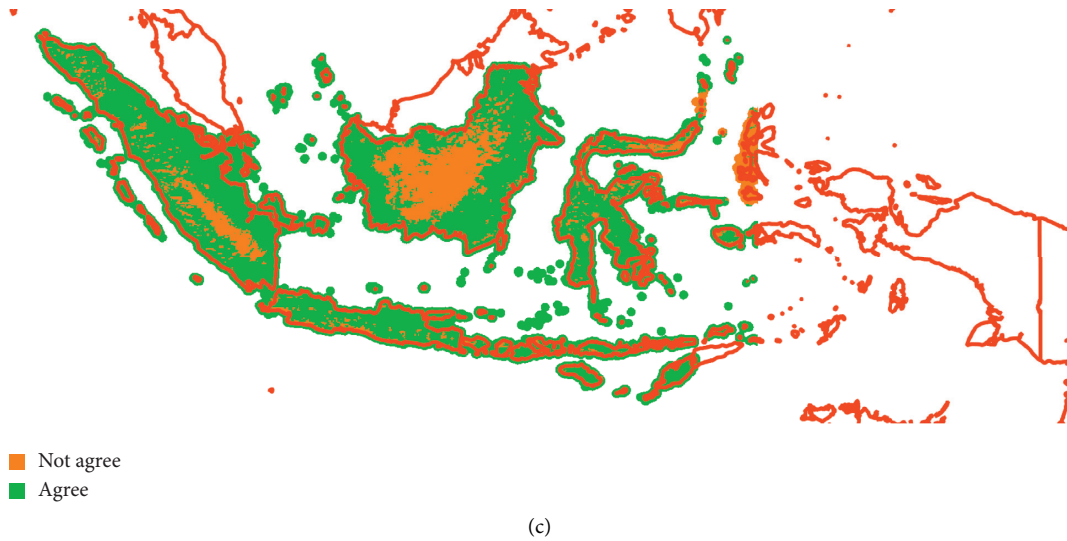


FIGURE 7: The common ground region between databases. (a) POWER, CLARA-A2, SARAH-E, and SolarGIS. (b) CLARA-A2, SARAH-E, and SolarGIS. (c) SARAH-E and SolarGIS.

TABLE 4: Correlation coefficient and mean bias of Meteonorm relative to other satellite-derived databases.

Locations	Correlation coefficient				Mean bias (kWh/m ²)			
	POWER	CLARA-A2	SARAH-E	SolarGIS	POWER	CLARA-A2	SARAH-E	SolarGIS
Bali (8.39 S, 115.56 E)	0.62	0.80	0.77	0.76	-0.01	0.11	-0.46	-0.58
Java (7.39 S, 111.95 E)	0.74	0.80	0.71	0.72	0.30	0.23	0.42	0.24
Kalimantan (2.47 N, 117.41 E)	0.87	0.90	0.81	0.84	0.41	0.31	0.30	0.07
Sulawesi (2.42 S, 120.97 E)	0.62	0.72	0.73	0.71	-0.07	-0.28	-0.09	-0.16
Sumatera (4.80 N, 97.63 E)	0.90	0.87	0.95	0.95	0.14	0.24	0.44	0.12

presented on Figure 6. Both figures generally agree in which region that significant impact from this downscaling process is estimated to be emerged, i.e., northern Kalimantan, most regions in Sulawesi, centre of Papua, southern of Sumatera, and Java. Apart from these regions, the spatial CoV loss is negligible.

By putting some constraints over r (GHI annual inter-month trend) and B_m (magnitude difference range) as requirements, several relative common grounds have been extracted out of the previous result and summarized into maps which are presented in Figure 7. Each map represents the agreement region of databases relative to the selected reference database. Assuming the true GHI data lies within databases estimated values, the pixels labeled as “agree” should give a higher chance of being close to the true value instead of the one with “not agree” label since they have been agreed at least by two databases. Due to the possibility of false signals triggered by one or more databases, three maps with a different number of databases combination and reference databases have been generated.

According to the maps, it is clear that the main highlights of “not agree” regions are highly influenced by previously discussed GHI annual intermonth trend disagreement around Kalimantan, Sumatera, and Papua. Since the GHI annual intermonth trend does not seem to be the cause of some “not agree” pixels on Sulawesi and Maluku region as shown in Figure 7(a), the disagreement must be triggered by

unfit mean bias values, which means the relative bias between two databases does not satisfy the requirement of $-x < B_m < x$, where $x = 0.4 \text{ kWh/m}^2/\text{day}$. According to Figure 4 (1–4), the three databases (i.e., SolarGIS, SARAH-E, and CLARA-A2) hold some pixels whose B_m values are outside the required range. However, among the three, only SolarGIS and SARAH-E, whose caps, extend beyond the limit of defined x value, the higher probability of exhibiting excessive B_m value. Upon confirming B_m distribution for each possible databases pair, it is found that “not agree” pixels around eastern Sulawesi, eastern Nusa Tenggara, and Maluku are indeed caused by the unfit B_m values of SARAH-E and SolarGIS against POWER as the reference database. Both SARAH-E and SolarGIS employ METEOSAT-East products as either primary or secondary inputs [19, 25]. As mentioned previously, METEOSAT-East easternmost viewpoint lies at 128 E, which approximately coincides with questioned regions. As commonly known, the edge of the viewpoint is prone to the parallax effect, which usually triggers misinterpretation of actual cloud conditions, and in the end generates inaccurate irradiance values.

To assess Meteonorm, five sample points across Indonesia were randomly selected. The points have been arranged to be positioned within regions that satisfy “agree” requirements on the three aforementioned common ground maps. The points also have been filtered to only consist of a satellite-derived dataset (no interpolation involved). The assessment is

conducted by simply checking r and B_m values for designated points and confirm whether the values satisfy the minimum requirements of being “agree” pixel. The assessment results presented in Table 4 reveal that only two sample points, i.e., Sumatera and Kalimantan, fulfill the GHI annual intermonth trend, as indicated by a strong correlation coefficient of more than 0.8. And one sample point complies with “agree” cell requirement of “ $-0.4 < B_m < 0.4 \text{ kWh/m}^2/\text{day}$,” i.e., Sulawesi. Overall, none of these sample points passes the minimum requirement of being “agree” pixel. At this point, the Meteonorm satellite-derived dataset seems to be undervalued around $0.12\text{--}0.15 \text{ kWh/m}^2/\text{day}$ by the majority of the database while only overvalues SolarGIS by about $0.06 \text{ kWh/m}^2/\text{day}$. Further investigation using a larger pool of data points is suggested to enhance the reliability of the present conclusion.

4. Conclusions

In this article, four satellite-derived databases, i.e., POWER, SARAH-E, CLARA-A2, and SolarGIS, which cover the region of Indonesia have been assessed in terms of data correlation and mean bias relatively between each other. Most database pairs are showing a strong correlation, which implies they follow roughly similar GHI annual intermonth trend. Meanwhile, some regions such as Central Kalimantan, Riau, South Sumatera, and Papua are exhibiting weak/no correlation, which is suspected to be triggered by the existence of cells with low/no temporal variation within the inspected databases and the absence of aerosol temporal trends due to the forest fire phenomenon. Spatial distributions of mean bias values reveal the relative position of each database in terms of estimated GHI magnitude, which on average follows this order: SolarGIS < SARAH-E < CLARA-A2 < POWER. Both computed correlation coefficient and relative bias are employed as agreement constraints to depict spatial common grounds for the four databases. The majority of “not agree” regions on presented common ground maps are influenced by the GHI annual intermonth trend disagreement between databases. While some minor “not agree” regions triggered by mean bias disagreement are believed to be the consequence of METEOSAT-East’s parallax effect.

In addition, a similar assessment was conducted on Meteonorm satellite-derived dataset relative to the four databases in five selected sample points, Bali, Java, Kalimantan, Sulawesi, and Papua. It reveals that none of these sample points passes the minimum requirement of being “agree” pixel, while the four databases do. Three databases, i.e., POWER, CLARA-A2, and SARAH-E are validated against WRDC stations located in Indonesia and surrounding nations by employing three statistical measures, namely RMSE, MBE, and correlation coefficient. According to the result, SARAH-E outperforms the other two databases especially on RMSE and time series correlation measures, while CLARA-A2 is still behind POWER on three statistical measures. The result also reveals that the three satellite-derived databases tend to overestimate their GHI value by about $0.07\text{--}0.32 \text{ kWh/m}^2/\text{day}$ (3–7% rMBE) on average compared to the true measured value.

Abbreviations

σ_{db1} :	The standard deviation of db1
σ_{db2} :	The standard deviation of db2
B_m :	Mean bias between one database relative to another
CoV:	Coefficient of variation
CoV _T :	Temporal coefficient of variation
Cov:	Covariance
db1:	The GHI reference data
db2:	The GHI estimated data
E_e :	Estimated solar energy resources
E_i :	GHI value for each pixel
E_j :	Multiyear average of typical daily GHI value for each month
E_q :	Mean of multiyear average of typical daily GHI value for each month in a year
E_m :	Measured solar energy resources
E_p :	Mean GHI value of pixels before downscaling
MBE:	Mean bias error
n :	Amount of data
RMSE:	Root mean square error
r :	Pearson correlation coefficient
rMBE:	Relative mean bias error
rRMSE:	Relative root mean square error.

Data Availability

The data used to support this study are obtained from NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program. The work performed was done (i.a.) by using SARAH-E and CLARA-A2 data from EUMETSAT’s Satellite Application Facility on Climate Monitoring (CM SAF), copyright (2019) EUMETSAT. © 2019, The World Bank, Source: Global Solar Atlas 2.0, Solar resource data: Solargis.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] IRENA, *Renewable Capacity Statistics*, International Renewable Energy Agency (IRENA), Abu Dhabi, United Arab Emirates, 2019.
- [2] S. R. Sulaiman, *New Solar Panel Regulation Discourages Public Use*, 2019.
- [3] M. Djakovic, *Rooftop Solar Photovoltaic Market Development in Indonesia—Business Case for Rooftop Solar PV Project Development in Indonesia*, The Jakarta Post, Jakarta, Indonesia, 2018.
- [4] L. D. Rifai, S. H. J. Tongkukut, and S. S. Raharjo, “Analisis intensitas radiasi matahari di Manado dan maros,” *Jurnal MIPA*, vol. 3, no. 1, pp. 49–52, 2014.
- [5] A. Octavianti and A. Muliadi, “Estimasi Intensitas Radiasi Matahari di Wilayah Kota Makassar,” *Prisma Fisika*, vol. 6, pp. 152–159, 2018.
- [6] D. Septiadi, P. Nanlohy, M. Souissa, and F. Y. Rumlawang, “Proyeksi potensi energi surya sebagai energi terbarukan

- (studi wilayah ambon dan sekitarnya),” *Jurnal Meteorologi Dan Geofisika*, vol. 10, pp. 22–28, 2009.
- [7] B. Pranoto, D. G. Cendrawati, N. W. Hesty, and E. Kusriadie, “Evaluasi pemetaan potensi energi surya berbasis model wrf di desa palihan dan desa aikangkung,” *Jurnal Sains Dirgantara*, vol. 15, no. 2, p. 63, 2018.
- [8] I. S. Nuraini, N. Hidayanto, Wandayantolis, Analisis Potensi Energi Matahari di Kalimantan Barat, 2017.
- [9] Y. Mohamad, *Pemetaan potensi energi surya pada lahan potensial agropolitan di propinsi gorontalo*, Jurusan Teknik Elektro Universitas Negeri Gorontalo, Gorontalo, Indonesia, 2016.
- [10] Nasruddin, M. A. Budiyanto, and M. H. Lubis, “Hourly solar radiation in depok, west Java, Indonesia (106.7942 Longitude, –6.4025 Latitude),” *IOP Conference Series: Earth and Environmental Science*, vol. 105, Article ID 012088, 2018.
- [11] G. L. Morrison and Sudjito, “Solar radiation data for Indonesia,” *Solar Energy*, vol. 49, no. 1, pp. 65–76, 1992.
- [12] I. Rahardjo and I. Fitriana, *Analisis Potensi Pembangkit Listrik Tenaga Surya Di Indonesia*, P3TKKE-BPPT, Jakarta, Indonesia, 2005.
- [13] M. Rumbayan and K. Nagasaka, “Estimation of daily global solar irradiation in Indonesia with artificial neural network (ANN) method,” *International Journal on Advanced Science, Engineering and Information Technology*, vol. 1, no. 2, p. 190, 2011.
- [14] V. Nurliyanti, M. Pandin, and B. Pranoto, “Pembuatan peta potensi energi surya,” *M&E*, vol. 10, no. 4, pp. 47–54, 2012.
- [15] A. P. Harsarapama, “Accurate irradiance measurement on large scale photovoltaic system application,” M. S Thesis, Delft University of Technology, Delft, Netherlands, 2017.
- [16] P. W. Stackhouse Jr., T. Zhang, D. Westberg et al., “POWER Release 8.0.1 (with GIS Applications) methodology (data parameters, sources, & validation),” Technical report, NASA Langley Research Center, Hampton, Virginia, 2018.
- [17] P. W. Stackhouse, R. Perez, M. Sengupta et al., “An assessment of new satellite data products for the development of a long-term global solar resource at 10–100 km,” in *Proceedings of the Solar 2016 Conference*, International Solar Energy Society, Abu Dhabi, UAE, October 2016.
- [18] T. Huld, R. Müller, A. Gracia-Amillo, U. Pfeifroth, and J. Trentmann, “Surface solar radiation data set–Heliosat,” *Meteosat-East (SARAH-E)–Edition*, vol. 1, 2017.
- [19] A. Amillo, T. Huld, and R. Müller, “A new database of global and direct solar radiation using the eastern meteosat satellite, models and validation,” *Remote Sensing*, vol. 6, no. 9, pp. 8165–8189, 2014.
- [20] A. Riihelä, V. Kallio, S. Devraj, A. Sharma, and A. Lindfors, “Validation of the SARAH-E satellite-based surface solar radiation estimates over India,” *Remote Sensing*, vol. 10, no. 3, p. 392, 2018.
- [21] Y. Wang, J. Trentmann, W. Yuan, and M. Wild, “Validation of CM SAF CLARA-A2 and SARAH-E surface solar radiation datasets over China,” *Remote Sensing*, vol. 10, no. 12, p. 1977, 2018.
- [22] K. Karl-Göran Anttila, J. Trentmann, M. Stengel et al., “CLARA-A2: CM SAF cCloud, Albedo and surface RAdiation dataset from,” *AVHRR Data–Edition*, vol. 2, 2017.
- [23] K.-G. Karlsson, K. Anttila, J. Trentmann et al., “CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data,” *Atmospheric Chemistry and Physics*, vol. 17, no. 9, pp. 5809–5828, 2017.
- [24] M. Suri, T. Cebecauer, N. Suriova et al., “Solar resource and photovoltaic potential of indonesia,” Technical Report May, ESMAP, The World Bank Group, Washington, DC, USA, 2017.
- [25] S. Solargis, *Solar Resource Database Description and Accuracy*, Solargis S.R.O., Bratislava, Slovakia, 2016.
- [26] J. Remund, S. Müller, S. Kunz, and C. Schilter, *Meteonorm V7 Handbook Part I: Software; Number May*, p. 39, Genossenschaft Meteotest, Bern, Switzerland, 2012.
- [27] S. C. Mueller and J. Remund, “Validation of the meteonorm satellite irradiation dataset,” in *Proceedings of the 35th European Photovoltaic Solar Energy Conference and Exhibition*, pp. 3–6, Brussels, Belgium, September 2018.
- [28] World Radiation Data Centre, 2019, <http://wrdc.mgo.rssi.ru/>.
- [29] A. Dobos, P. Gilman, and M. Kasberg, “P50/P90 analysis for solar energy systems using the system Advisor model,” *World Renewable Energy Forum*, p. 8, 2012.
- [30] C. A. Gueymard and S. M. Wilcox, “Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data,” *Solar Energy*, vol. 85, no. 5, pp. 1068–1084, 2011.
- [31] J. Evans, *Straightforward Statistics for the Behavioral Sciences*, Brooks/Cole Publishing Company, Pacific Grove, CA, USA, 1996.
- [32] M. I. Mead, S. Castruccio, M. T. Latif et al., “Impact of the 2015 wildfires on Malaysian air quality and exposure: a comparative study of observed and modeled data,” *Environmental Research Letters*, vol. 13, Article ID 044023, 2018.
- [33] G. Alexandri, A. K. Georgoulas, C. Meleti et al., “A high resolution satellite view of surface solar radiation over the climatically sensitive region of Eastern Mediterranean,” *Atmospheric Research*, vol. 188, pp. 107–121, 2017.
- [34] Rekapitulasi Luas Kebakaran Hutan dan Lahan (Ha) Per Provinsi Di Indonesia Tahun 2014-2019, 2019, http://sipongi.menlhk.go.id/hotspot/luas_kebakaran.