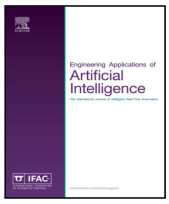
Neural networks for estimating surface solar irradiation from satellite images

Engineering Applications of Artificial Intelligence 144 (2025) 110101

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



Research paper

Neural networks for estimating surface solar irradiation from satellite images

Raimondo Gallo [b,*], Marco Castangia [b], Alberto Macii [b], Edoardo Patti [b,c], Alessandro Aliberti [a]

[a] Interuniversity Dept. of Regional and Urban Studies and Planning, Politecnico di Torino, 10129 Torino, Italy
[b] Department of Control and Computer Engineering, Politecnico di Torino, 10129 Torino, Italy
[c] Energy Center Lab, Politecnico di Torino, 10129 Torino, Italy

ARTICLE INFO

Keywords:
Solar radiation
Neural networks
Nowcasting
Meteosat
Satellite images
Heliosat-4

ABSTRACT

In this work, we present a MultiLayer Perceptron (MLP) model to estimate ground solar radiation, in terms of Global Horizontal Irradiance (GHI), over a given site and for a specific time. The MLP model generates GHI estimates from the Meteosat 12-channel satellite images centered over the target location, and GHI values in clear-sky conditions over the same location. The dataset includes two years of data, covering 2016 and 2017, with a temporal granularity of 15 min, relative to a set of 16 test sites distributed across Europe, Africa and South America. We populate the training- and test-sets with all available data for the 15 and remaining station, respectively. We test all possible combinations of stations to define training and test sets, demonstrating the generalizability of the presented MLP model over potentially any location included in the Meteosat full-disk image. The estimated GHI values are compared to ground-measured GHI data achieving an overall Root Mean Square Error (RMSE) and Coefficient of determination (R^2) of 77.682 W/m^2 and 0.929, respectively, across all locations. Finally, the GHI estimates are set against those generated by the Heliosat-4 method, our benchmark, yielding an overall RMSE improvement of 3 W/m^2. The experiments show that neural networks produce competitive results with fewer and accessible inputs compared to complex physical models for estimating solar radiation. Furthermore, historical and near real-time GHI estimation enabled by the proposed methodology would help photovoltaic (PV) planners determine the irradiance profile of a site where the deployment of a weather station is precluded.

1. Introduction

Solar energy generated by the sun is the source of photovoltaic (PV) systems, representing a renewable and eco-friendly form of energy. In 2021, global electricity demand increased by 6% (IEA, 2022), and to cope with such necessity, it becomes necessary to exploit renewable energy sources due to the limited availability of fossil fuels. In addition, the Earth receives enough solar energy in a single hour to meet the world's energy needs for an entire year (Shaikh et al., 2017). Using such an amount of energy would reduce dependence on fossil fuels, leveraging their impact on climate change. The rapid increase in electricity demand in developing countries requires the rapid deployment of renewable energy sources in these regions to meet their needs and enable gradual growth.

Solar radiation is a flux of energy measured as power per unit area, expressed in W/m^2. The total amount of solar radiation from the entire sky on a horizontal surface is called Global Horizontal Irradiance (GHI). To measure GHI at a specific location, it is necessary to deploy an unshaded pyranometer onto a stable horizontal platform (Besharat et al., 2013). Gathering GHI measurements allows users to quantify the amount of solar radiation received at the site and evaluate the efficiency of PV panels deployed in the same area. Prior solar radiation estimation would prevent decision makers from deploying costly ground sensors and allow them to gain preliminary insights into the quality of the site. However, installing measuring devices can be challenging, especially in developing countries, due to the cost, maintenance and calibration of the instruments.

Nowadays, the availability of remote sensing data enriches data collection, which can be used in many applications such as early warning of thunderstorms, monitoring sea surface temperature, and deforestation (Emery and Camps, 2017b,c,a). Satellite observations can also be used to improve solar radiation estimation and prediction (Müller and Pfeifroth, 2022). In fact, the literature is rich with

* Corresponding author.
E-mail addresses: raimondo.gallo@polito.it (R. Gallo), marco.castangia@polito.it (M. Castangia), alberto.macii@polito.it (A. Macii), edoardo.patti@polito.it (E. Patti), alessandro.aliberti@polito.it (A. Aliberti).

https://doi.org/10.1016/j.engappai.2025.110101
Received 21 December 2022; Received in revised form 24 October 2024; Accepted 14 January 2025

Nomenclature

List of Acronyms

ANN	Artificial Neural Network
$BSRN$	Baseline Surface Radiation Network
$CAMS$	Copernicus Atmosphere Monitoring Service
GHI	Global Horizontal Irradiance
$HRIT$	High Rate Image Transmission
HRV	High Resolution Visible
MLP	MultiLayer Perceptron
MSG	Meteosat Second Generation
PV	Photovoltaic
R^2	Coefficient of Determination
$RMSE$	Root Mean Square Error
$SEVIRI$	Spinning Enhanced Visible and InfraRed Imager

studies that use satellite observations or features extracted from such data to improve solar radiation monitoring. In general, solar radiation estimation methods are divided into empirical methods and artificial intelligence methods (Zhang et al., 2017). The former uses mathematical formulas to empirically model solar radiation. At the same time, the latter uses nonlinear statistical models to determine relationships between a series of inputs and solar irradiance, the prediction output. The state-of-the-art Heliosat method converts observations collected by geostationary meteorological satellites into ground-level global irradiance estimates (Rigollier et al., 2004). Such a method uses Meteosat satellite images to collect information about the state of the atmosphere and cloud cover over the target to compute the radiation reaching the ground. However, the Heliosat method struggles with the sudden appearance of snow cover in a cloud-free atmosphere, which leads to large errors (Rigollier et al., 2004). Additionally, physical models are difficult to reproduce due to their complexity and the difficulty of retrieving input parameters.

Therefore, a MultiLayer Perceptron (MLP) model, belonging to the class of Artificial Neural Network (ANN) methods, is proposed to improve the achievable accuracy of solar irradiance estimation in contrast to the Heliosat method, a sophisticated physical method for solar radiation estimation.

The presented methodology is shown in Fig. 1 and an explanation of the main adopted steps is provided. The dataset used includes data from two years, relating to 2016 and 2017, and is characterized by a temporal resolution of 15 min. The proposed MLP model generates GHI estimates over the target location, leveraging the multi-spectral images from the Meteosat Second Generation (MSG) satellites and clear-sky GHI estimates. The multi-spectral Meteosat images, collected by Spinning Enhanced Visible and InfraRed Imager (SEVIRI) (EUMETSAT, 2022e), are image data acquired within a specific wavelength from 12 different channels, each capturing specific information about the composition of the atmosphere. The multi-spectral images used are centered on the location of interest and are provided by the 0 Degree High Rate SEVIRI Level 1.5 Image Data Service, the primary mission of the MSG satellites (EUMETSAT, 2022a). The satellite images are coupled with another input, the estimated clear-sky GHI values over the same location. To obtain such data, McClear (Lefèvre et al., 2013) is used, a physical model that generates clear-sky GHI estimates based on a set of physical laws and a range of astronomical and atmospheric inputs. The coupling between the input satellite images and the estimated clear-sky GHI data is necessary to establish a relationship between the Meteosat satellite observations and the amount of ground solar radiation reaching the target area.

The satellite images are disseminated by the source satellite with a temporal resolution of 15 min, this imposes a bond over the rest of the data to match the same granularity. Consequently, the GHI estimation will reflect the same constraint; the MLP model generates GHI estimates matching the 15-min resolution of the inputs. The target locations of our analysis include 16 meteorological stations belonging to the Baseline Surface Radiation Network (BSRN) (WRMC-BSRN, 2022). We wanted to define an estimation model that can generalize over any location included in the MSG full-disk images. Therefore, given the input 16 stations, the model is trained independently by selecting all available samples from 2016 and 2017 for a subset of 15 stations and we tested it with the samples relative to the remaining station. Through a location-based K-fold cross validation, all possible combinations of stations are selected to populate training and testing sets. In this way, it is possible to verify the efficiency of our methodology in generalizing over heterogeneous areas. The quality of the proposed model is demonstrated through a performance assessment. This important step involves comparing the GHI estimation results for a given target station and the ground GHI measured from the same station. In addition, the same estimation results are also compared with those of the Heliosat-4 method, the selected reference benchmark, to verify the effectiveness of neural network approaches against complex physical models in estimating irradiance from satellite observations.

The novelty of the proposed methodology lies in the complete independence of surface weather information for producing GHI estimates. In fact, the selected inputs can be accessed anywhere and at any time, allowing the user to easily reproduce the methodology. Furthermore, we show that the MLP model can generalize with good accuracy over unseen locations, enabling GHI estimation potentially over any location of interest as long as it is geographically included in the Meteosat full-disk image.

To further highlight the advantages of the proposed methodology, the estimation results are compared with those of the Heliosat-4 method, a state-of-the-art, fully physical method. Indeed, the selected physical benchmark Heliosat-4 consists of a complex physical model that requires a non-trivial acquisition of atmospheric inputs resulting in a two-day delay for obtaining GHI estimates. On the contrary, the proposed MLP model is solely bounded to the primary input, the satellite images, available in near real-time eliminating the Heliosat-4's delay in retrieving the GHI estimates. The MLP model generally proves to be more accurate than its fully physical counterpart and, also, allows the GHI estimation to be extended to a near real-time scenario. The remaining sections are organized as follows: Section 2 provides an overview of the most effective machine learning method for estimating GHI. Section 3 provides a detailed description of the datasets used. Section 4 presents an overview of the selected data processing and the proposed methodology. Then, Section 5 discusses the estimation results and their comparison with the reference estimates of Heliosat-4 using various analytical indices. Finally, Section 6 presents the relevant findings of this work and provides some directions for future works.

2. Related works

As explained in Section 1, solar irradiance estimation is essential for many applications. In the literature, the methods for estimating solar radiation can be generally divided into empirical and artificial intelligence methods (Zhang et al., 2017). The former uses concise mathematical formulations to relate the target solar radiation with a disparate set of exogenous parameters, i.e., astronomical, and meteorological factors. latter instead demonstrated their efficiency in solving nonlinear problems by identifying hidden relationships between a set of input features and the target variable.

Many empirical methods have been defined to estimate either daily or monthly global solar radiation using various available meteorological data collected from weather stations. The empirical models can

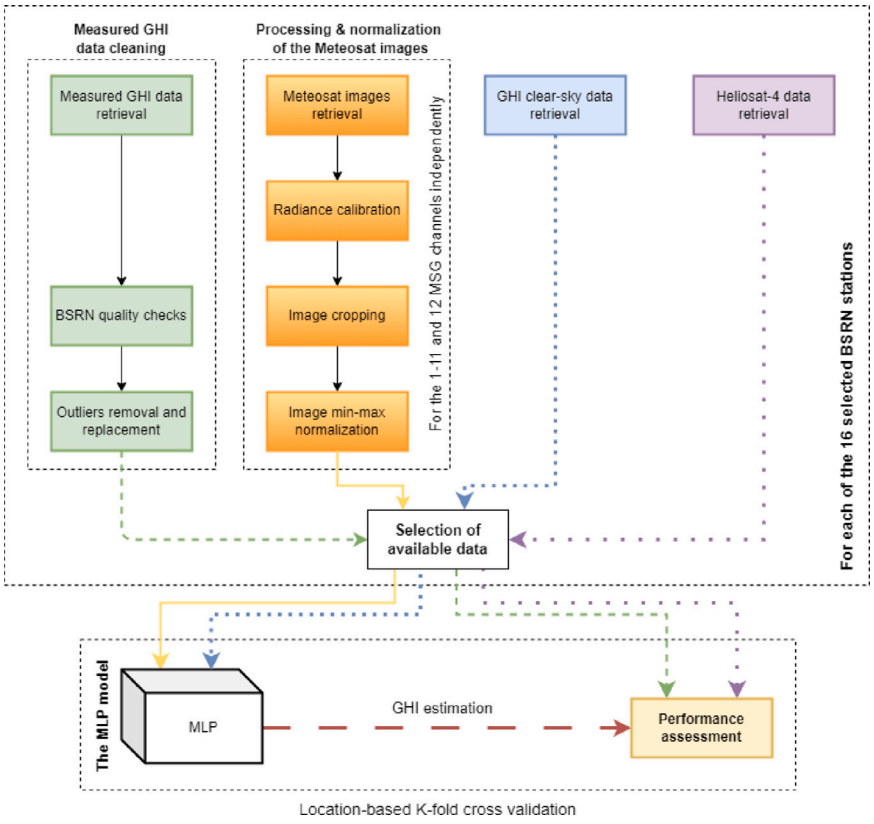


Fig. 1. The whole pipeline of the presented methodology.

be divided into four groups: sunshine-based, cloud-based, temperature-based and other meteorological parameter-based models (Besharat et al., 2013). They differ in how solar radiation is compared with the available measured meteorological parameters. The most commonly used parameter to estimate solar radiation is sunshine duration because it is easy to measure, and has generally been shown to be more accurate (Zhang et al., 2017). Nevertheless, the constants used in the mathematical expression mainly depend on the location being analyzed (Zhang et al., 2017).

Heliosat-4 (Qu et al., 2017; Schroedter-Homscheidt et al., 2022) features a state-of-the-art empirical, fully physical method for estimating the down-welling shortwave ground irradiance in all-sky conditions. Among the outputs, it provides global irradiance on a horizontal plane through a fast and accurate radiative transfer modeling. The method consists of two models: the McClear model, which calculates irradiance values under cloud-free conditions, and the McCloud model that calculates the extinction of irradiance due to clouds (Qu et al., 2017). Inputs to the Heliosat-4 method include aerosol properties, total column water vapor, and ozone content. The cloud properties are derived from satellite images with a temporal resolution of 15 min and distributed by the geostationary MSG satellites. The accuracy of the all-sky irradiances produced by Heliosat-4 and the model's ability to generalize across all areas included in the Meteosat full-disk images make it a highly reliable and efficient method. For these reasons, we choose the Heliosat-4 method as a reference benchmark in our work to prove the quality of our artificial intelligence-based methodology in contrast to complex empirical physical models.

The counterpart to empirical models are artificial intelligence methods, which have become an alternative to traditional techniques and allow researchers to expand the analysis to obtain more accurate results (Yadav and Chandel, 2014). In the literature, ANNs have been shown to be very efficient compared to empirical models in estimating solar radiation by using different exogenous data collected either from ground or remote sensors such as geostationary satellites. Khan et al. (2014) developed an ANN model to estimate daily solar irradiance in Dhaka using multiple daily meteorological data. Their proposed model outperformed an empirical reference and achieved a Root Mean Square Error (RMSE) of 113.6 Wh/m^2 over a year of test data, proving the efficiency of artificial intelligence approaches.

Dorvlo et al. (2002) proposed an MLP model for estimating daily solar radiation using exogenous geographic and meteorological measurements collected from a series of eight stations in Oman, Asia; Six were used to train the proposed MLP model and the remaining two were used to test the methodology. The proposed ANN estimates the clearness index from which solar radiation is derived. The model proved to be very accurate and achieved an RMSE of 1.35 MJ/m^2/day. Woldegiyorgis et al. (2022) investigated the feasibility of using ANN to predict mean daily Global Solar Radiation (GSR). They combined three years of multiple daily meteorological data retrieved from the NASA database and ground-measured sunshine duration over Lalibela, Ethiopia. The predicted GSR is compared with that provided by the NASA database, and the ANN model achieves an RMSE of 0.3310 and 0.0433 KWh/m^2, for mean daily and monthly averaged daily, respectively. Furthermore, they compared the proposed ANN model with three empirical reference models, with the latter far outperforming the competing benchmark methods.

Kurniawan and Harumwidiah (2021) used an ANN model to estimate daily solar radiation in Surabaya, Indonesia. They selected five years of daily average weather ground parameters to train the ANN model, which outputs daily solar radiation. The results are validated with daily average global solar radiation data from the NASA database. Their proposed model achieved a Mean Absolute Percentage Error (MAPE) of less than 20%, illustrating how the estimated solar irradiance values approach the measured ones. Sahan and Yakut (2016) proposed an ANN model to estimate the monthly average global horizontal solar radiation for five locations in Turkey. The model is trained with more than 14 years of monthly mean meteorological and geographic data values. The ANN was tested using data over a

period of three years. The results provide a Coefficient of Determination (R^2) ranging between 0.97 and 0.99 and a minimum RMSE of 0.506 MJ/m^2, demonstrating good agreement between estimated and measured monthly mean global solar irradiation.

Chen et al. (2019) used an ANN model to estimate the hourly GHI over Hong Kong using three years of ground-measured hourly meteorological data. Given the inputs, the ANN provides the clearness index as output, from which the hourly GHI is extracted. The proposed ANN achieves an R^2 of 0.925 on the test set. As a less complex alternative to machine learning, they implemented several empirical models that use different input combinations to predict the clearness index. However, the empirical benchmark models, validated over a year of data, did not improve ANN performances. They also emphasized the importance of the sunshine duration and solar altitude angle predictors for estimating global solar irradiance. Kurniawan and Shintaku (2021) developed a two-step ANN to estimate daily and hourly average solar radiation over the Java Island, Indonesia. The first ANN is trained on two years of meteorological data from the NASA database and outputs daily average solar radiation. The output of the first ANN, along with geographic data, is then fed into a second ANN, which provides daily average estimates of solar radiation. The results of both ANNs are validated over a year, and they yield an R^2 of 0.98 and 0.97 for daily and hourly estimates, respectively. The presented methodology outperforms several previous studies in the literature.

Goncu et al. (2021) proposed an ANN model to estimate hourly GSR in Adana, Turkey. The dataset covers a full year, including hourly ground meteorological measurements and the target GSR. They conducted a fine-tuning of the ANN parameters, and the optimized model achieves an R^2 of 0.87 for the hourly GSR estimates. Choudhary et al. (2020) provided an overview of solar radiation estimation studies and also proposed their own ANN model for hourly solar radiation estimation. They retrieved one year of hourly data from the National Solar Radiation Database for New Delhi, India. The input data includes different geographical and meteorological measurements, and the ANN model achieves an R^2 of 0.94 on the test set. The availability of remote sensing data allows users to improve the quality of proposed solar irradiance estimation solutions.

In fact, Fallahi et al. (2018) aimed to prove the efficiency of ANN compared to other regression models and remote sensing data using ground measurements. They chose an ANN model to estimate monthly solar radiation over Kurdistan, Iran. They used two independent sets of geographic data, including monthly mean temperature, one retrieved from the NOAA satellite's Advanced Very High-Resolution Radiometer (AVHRR) sensor and the other measured on the ground. The ANN trained with remote sensing data achieved an RMSE of 6.4% on the test set compared to an RMSE of 10.4% for the ANN trained with ground data, demonstrating the potential of satellite data compared to expensive ground sensors. Furthermore, the ANN model outperformed the competing regression model, which achieved an RMSE of 14.7% using satellite data.

Koo et al. (2020) proposed an ANN model to estimate the hourly GHI in Korea using two-year GHI data from 25 stations, solar position variables, and five Communication, Ocean and Meteorological Satellite (COMS) Meteorological Imager (MI) channels. After parameterizing the ANN model, they validated the model temporally and spatially to ensure the consistency of the results. Therefore, the final ensemble ANN model was used to estimate the hourly GHI and achieved an overall RMSE and R^2 of 54.44 W/m^2 and 0.975. Finally, they used the ensemble model to generate hourly irradiance maps in Korea.

Guijo-Rubio et al. (2020) developed an ANN model to estimate hourly solar radiation in Toledo, Spain. To validate the estimates, they used two years of hourly GHI ground data, and the selected predictive variables include various astronomical and SEVIRI (EUMETSAT, 2022e) satellite data. They tested different combinations of inputs and several typologies of hidden neurons and output functions to train the ANN. The best configuration achieved an RMSE of 51.82 W/m^2 and an R^2 of 0.97. The methodology proposed by Guijo-Rubio et al. (2020) outperformed state-of-the-art machine learning algorithms such as Support Vector Regressors and Extreme Learning Machines.

Tang et al. (2017) worked on methods to estimate instantaneous surface solar radiation using products from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor. In one of their studies (Tang et al., 2016), they employed an ANN that defines a functional relationship between MODIS and Multifunctional Transport Satellite (MTSAT) geostationary satellite signals. The ANN is used to retrieve the hourly parameters from MODIS, which are instead provided at a low temporal resolution. Their methodology achieves an RMSE of 98.50 W/m^2 for hourly surface solar radiation estimates over three experimental stations in China.

The related works demonstrate the efficiency of artificial intelligence methods as a valid alternative to empirical models for estimating solar radiation. Furthermore, they highlighted the benefits of using remote sensing data; In particular, such data prevents users from having to deploy costly weather stations, allowing them to expand the study to any area of interest, thereby achieving state-of-the-art performances. Nevertheless, only a few studies use raw satellite images as input to estimate or predict solar radiation. Most of these works employ either time series of remote sensing data or features extracted from satellite images.

In addition to ANNs, convolutional neural networks have demonstrated their effectiveness in extrapolating spatiotemporal features from satellite images, achieving remarkable results, especially for forecasting tasks (Yang et al., 2021; Choi et al., 2021). However, the benefits of using more sophisticated models that implement a convolutional architecture arise when utilizing a sufficiently large image that allows the model to extract spatial cloud features from a larger scene (Jiang et al., 2020). Conversely, ANN proved to be more effective than convolutional neural networks in estimating solar radiation whenever the input satellite images were limited to a smaller ground area (Jiang et al., 2020).

The proposed solution uses satellite imagery covering a ground area of about 30 × 30 km at the sub satellite point as explained in Section 3, and for such geographical extent the ANN results are more accurate than a convolutional neural network (Jiang et al., 2020). Furthermore, we believe that the convolution operation would result in pixel values being averaged, thereby losing georeferenced information from the input satellite images. Also, the dimension of the kernel for the convolution operation would not be much larger than input images, which would result in the extraction of a fewer number of features, drastically reducing the amount of input information.

These considerations led us to the adoption of an ANN model, providing each and all image pixels as input, preserving all geographical information in the multispectral images.

In our previous work (Gallo et al., 2022), we have demonstrated the benefits of Meteosat satellite images used as raw input for training deep learning models to forecast solar radiation over Turin, Italy. The results outperformed competing models trained with vast exogenous ground input data. Providing raw satellite imagery as input brings several benefits; They are easily accessible and require very little processing. Such properties overcome the challenging retrieval of specific inputs inherent in empirical models.

In this work, rather than forecasting, we focus on solar radiation punctual estimation, for a specific location and time. Specifically, the proposed MLP model was trained with data from 16 locations in Europe, Africa and South America (all included in the Meteosat full-disk image), rather than a single target city. The dataset includes two years of all twelve SEVIRI channels from the geostationary MSG satellite, coupled with time series of clear-sky GHI. The GHI estimates are then validated with ground GHI measurements from the same two years retrieved from the BSRN.

Finally, the GHI estimates of the proposed ANN are compared with those obtained using the Heliosat-4 method to evaluate the efficiency of

Table 1
The general information of the selected BSRN stations included in our study.

City	Abbreviation	Location	Latitude [°]	Longitude [°]	Altitude [m]
Cabauw	CAB	Netherlands	51.9711	4.9267	0
Camborne	CAM	United Kingdom	50.2167	−5.3167	88.0
Carpentras	CAR	France	44.0830	5.0590	100.0
Cener	CNR	Spain, Navarra	42.8160	−1.6010	471.0
De Aar	DAA	South Africa	−30.6667	23.9930	1287.0
Florianopolis	FLO	Brazil, South Atlantic Ocean	27.6047	−48.5227	11.0
Gobabeb	GOB	Namibia, Namib Desert	−23.5614	15.0420	407.0
Izaña	OZA	Spain, Tenerife	28.3093	−16.4993	2372.9
Lindenberg	LIN	Germany	52.2100	14.1220	125.0
Palaiseau	PAL	France	48.7130	2.2080	156.0
Payerne	PAY	Switzerland	46.8123	6.9422	491.0
Petrolina	PTR	Brazil	−9.0690	−40.3200	387.0
São Martinho da Serra	SMS	Brazil	−29.4428	−53.8231	489.0
Sonnblick	SON	Austria	47.0540	12.9577	3108.9
Tamanrasset	TAM	Algeria	22.7903	5.5292	1385.0
Toravere	TOR	Estonia	58.2540	26.4620	70.0

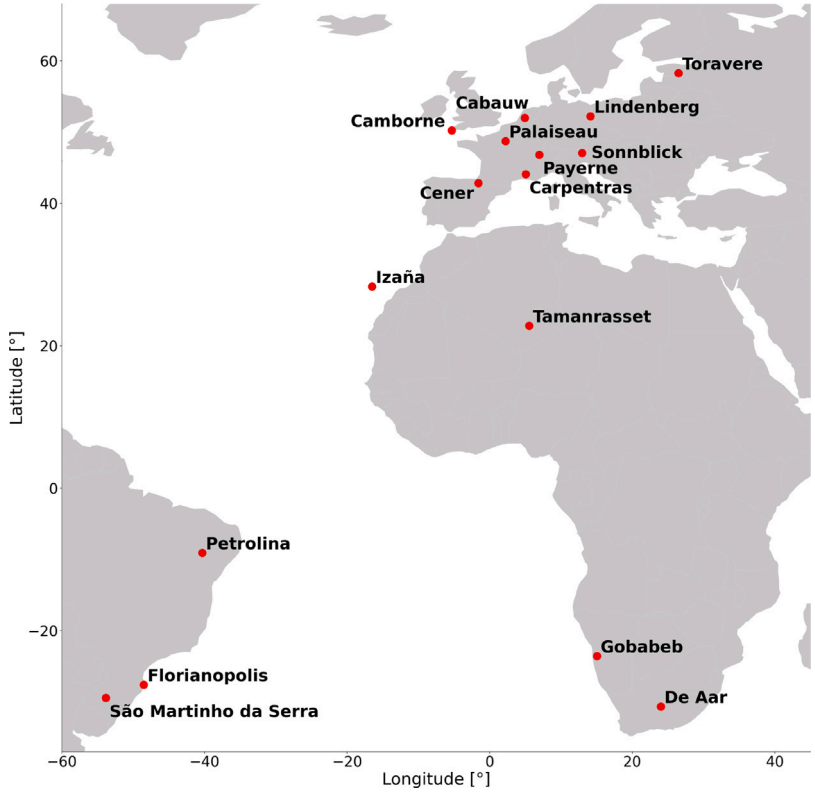


Fig. 2. The geographical position of the selected BSRN stations used in our study to validate the GHI estimates.

artificial intelligence methods compared to physical empirical models. The resulting model is a robust model capable of estimating GHI across heterogeneous areas and of generalizing solar irradiance estimation across unseen locations.

3. Dataset

The dataset employed to estimate solar radiation includes two years of data relating to 2016 and 2017 and is characterized by a temporal granularity of 15 min. All data is provided in UTC Universal Time. The analysis is extended to 16 BSRN stations spread across Europe, Africa and South America - areas all included in Meteosat's nominal full-disk images. Table 1 provides a characterization of the selected BSRN stations that represent the test sites considered in our study. Instead, Fig. 2 shows the geographic location of the selected test BSRN stations.

The dataset includes, for each test station, Meteosat 0 Degree multi-spectral images, estimated GHI data in clear-sky conditions, GHI ground measurements collected from weather stations located in the test areas, and GHI estimates generated using the Heliosat-4 physical method. To ensure the temporal consistency of all employed datasets, we collect all GHI data that match the dissemination times of the Meteosat satellite images. In this way, for each 15-min time step sample relative to 2016 and 2017, we obtain all the necessary information to implement our methodology. The clear-sky and Heliosat-4 GHI time series are created using information from the Copernicus Atmosphere Monitoring Service (CAMS) (Copernicus, 2022). In particular, we use the CAMS radiation service, which performs the calculation of the radiation on-the-fly at the user's desired location. The following sections provide a deeper insight into the selected data, its source, and properties. In addition, Section 4 discusses the procedures for dealing with missing data and outliers and preprocessing the satellite images.

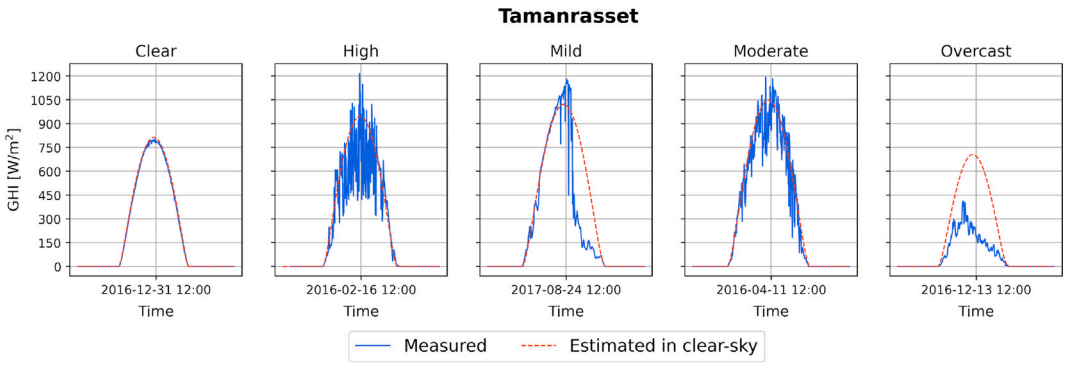


Fig. 3. A sample of five days, each representative of the mentioned variability categories, for the station of Tamanrasset (TAM).

3.1. GHI in clear-sky conditions

The clear-sky GHI values are generated by the McClear model (Lefèvre et al., 2013; Gschwind et al., 2019), which provides irradiation data observed in clear-sky conditions. The McClear model is a fast clear-sky model that implements physical modeling exploiting aerosol properties and total column content in water vapor, ozone and other inputs. The selected clear-sky model provides time series of global solar surface radiation at different temporal resolutions for any location in the world and at every point in time, starting from the year 2004. The McClear radiation model goes through a validation process in which the estimated clear-sky data are compared with ground measurements in clear-sky conditions. Such a comparison gives an RMSE of about 20–30 W/m^2 across different stations, proving the accuracy of the selected clear-sky model (Lefèvre et al., 2013).

The McClear clear-sky data is provided by the CAMS Radiation service and accessed via the *Pvlib* Python library (Holmgren et al., 2018, 2022), which provides open, reliable and benchmark implementations of PV system models. For our use case, Pvlib provides functions for accessing data from the CAMS radiation service data through SoDa (SoDa, 2022), a broker for a list of services and web services that provides access to a large amount of information about solar radiation. For each location analyzed, i.e. the individual BSRN stations, the query includes the geographical coordinates and altitude of the location, providing a temporal resolution of 15 min corresponding to the granularity of the input satellite images. The collected clear-sky GHI time series contain average irradiance values expressed in W/m^2, rather than integrated values expressed in Wh/m^2. In this way, for each BSRN station, we collect a time series of clear-sky GHI for the entire years 2016 and 2017 with a granularity of 15 min.

3.2. Ground GHI data

The GHI ground measurements are used to validate the GHI estimates generated by the proposed MLP model. The GHI measurements are retrieved from the BSRN (WRMC-BSRN, 2022), which provides high quality observations for short- and long-wave surface radiation fluxes at a high sampling rate of 1 min. In addition, such data are well suited to validating and evaluating satellite-based estimates of surface radiative fluxes (McArthur, 2005). The BSRN data of interest belong to the *basic radiation and other radiation measurements* datasets, which include global, direct and diffuse radiation, air temperature, pressure and relative humidity for each station and time of interest. For our study, we only selected data on global solar radiation. The BSRN provides the measured GHI data in monthly datasets, each specific to the year and station of interest. Table 2 lists the two-year datasets monthly availability of the measured GHI data for all selected stations. Few monthly datasets are unavailable for the Camborne and São Martinho da Serra stations. Therefore, we exclude from our analysis all data related to the timestamps of the missing GHI data, as we would not be able to assess the performance of the MLP model for these timestamps.

The GHI measurements for the chosen stations, reported in Table 1, are retrieved through the Pvlib Python library from the BSRN FTP server (BSRN, 2022). Since the BSRN global radiation data are furnished with a temporal resolution of 1 min, we select the measurements relative to the dissemination times of the satellite images, performing a subsetting of the GHI time series. In this way, we obtain a GHI time series for each station with a temporal resolution of 15 min, consistent with that of the input satellite images. This ensures the temporal consistency of the data.

Furthermore, we classified the days included in the two-year dataset into five different variability categories, following the criteria proposed by Trueblood et al. (2013). Such a method relies on the calculation of two indices, namely the daily clearness index and the daily variability index, and uses only ground GHI measurements and GHI data in clear-sky. The clearness index measures the atmosphere clearness, while the variability index quantifies the variability of the measured irradiance. The method distinguishes days into five variability conditions using a set of thresholds for the clearness and variability indices: high variability, moderate variability, mild variability, clear, and overcast days. In Fig. 3 we show a sample of five different days, each representative of the aforementioned variability categories, for the Tamanrasset station. Furthermore, Fig. 4 shows the distribution of daily variability conditions for each selected station. Such a variability condition is used to further describe the MLP performances on days characterized by different meteorological scenarios in Section 5.

3.3. Heliosat-4 GHI data

The GHI estimates provided by our proposed MLP model are compared with those of the Heliosat-4 method (Qu et al., 2017; Schroedter-Homscheidt et al., 2022) to evaluate the efficiency of artificial intelligence approaches over a state-of-the-art physical model. The CAMS radiation service provides the Heliosat-4's GHI data, retrieved using Pvlib via SoDa. Therefore, access to Heliosat-4's GHI data is similar to McClear's data. For each BSRN station, the query includes the geographical information of the site, specifying a temporal resolution of 15 min and irradiance values expressed in W/m^2. In this way, for each location, we retrieve the GHI time series from Heliosat-4 for the entire years 2016 and 2017, with a granularity of 15 min corresponding to that of the input satellite images.

3.4. Meteosat satellite images

The input satellite images to our MLP model are disseminated by the Meteosat 3 satellite, which belongs to the MSG family of geostationary satellites (EUMETSAT, 2022d). The data is transmitted as high-rate transmissions in 12 spectral channels covering a geographical area included between −66° and 66° in latitudes and longitudes. The images

Table 2
The availability of the BSRN ground measured GHI monthly datasets for 2016 and 2017, and all selected stations.

Station	January	February	March	April	May	June	July	August	September	October	November	December
Cabauw	2016, 2017: Knap (2022)											
Camborne	2016, 2017: Hodgetts et al. (2022)							2016: Hodgetts et al. (2022) 2017: –	2016: Hodgetts et al. (2022) 2017: –	2016: Hodgetts et al. (2022) 2017: –	2016: Hodgetts et al. (2022) 2017: –	2016: Hodgetts et al. (2022) 2017: –
Carpentras	2016, 2017: Brunier et al. (2021)											
Cener	2016, 2017: Olano (2021)											
De Aar	2016: Ntsangwane (2017a) 2017: Ntsangwane (2017m)	2016: Ntsangwane (2017b) 2017: Ntsangwane (2017n)	2016: Ntsangwane (2017c) 2017: Ntsangwane (2017o)	2016: Ntsangwane (2017d) 2017: Ntsangwane (2017p)	2016: Ntsangwane (2017e) 2017: Ntsangwane (2017q)	2016: Ntsangwane (2017f) 2017: Ntsangwane (2017r)	2016: Ntsangwane (2017g) 2017: Ntsangwane (2017s)	2016: Ntsangwane (2017h) 2017: Ntsangwane (2017t)	2016: Ntsangwane (2017i) 2017: Ntsangwane (2017u)	2016: Ntsangwane (2017j) 2017: Ntsangwane (2017v)	2016: Ntsangwane (2017k) 2017: Ntsangwane (2017w)	2016: Ntsangwane (2017l) 2017: Ntsangwane (2018)
Florianopolis	2016: Colle (2016a) 2017: Colle (2017e)	2016: Colle (2016b) 2017: Colle (2017f)	2016: Colle (2016c) 2017: Colle (2017g)	2016: Colle (2016d) 2017: Colle (2017h)	2016: Colle (2017a) 2017: Colle (2017i)	2016: Colle (2016e) 2017: Colle (2017j)	2016: Colle (2016f) 2017: Colle (2017k)	2016: Colle (2016g) 2017: Colle (2017l)	2016: Colle (2016h) 2017: Colle (2018a)	2016: Colle (2017b) 2017: Colle (2018b)	2016: Colle (2017c) 2017: Colle (2018c)	2016: Colle (2017d) 2017: Colle (2018d)
Gobabeb	2016, 2017: Al-Abbadi et al. (2017)											
Izaña	2016: Cuevas-Agulló (2016a) 2017: Cuevas-Agulló (2017b)	2016: Cuevas-Agulló (2016b) 2017: Cuevas-Agulló (2017c)	2016: Cuevas-Agulló (2016c) 2017: Cuevas-Agulló (2017d)	2016: Cuevas-Agulló (2016d) 2017: Cuevas-Agulló (2017e)	2016: Cuevas-Agulló (2018b) 2017: Cuevas-Agulló (2017f)	2016: Cuevas-Agulló (2016e) 2017: Cuevas-Agulló (2017g)	2016: Cuevas-Agulló (2016f) 2017: Cuevas-Agulló (2017h)	2016: Cuevas-Agulló (2016g) 2017: Cuevas-Agulló (2017i)	2016: Cuevas-Agulló (2016h) 2017: Cuevas-Agulló (2017j)	2016: Cuevas-Agulló (2016i) 2017: Cuevas-Agulló (2017k)	2016: Cuevas-Agulló (2016j) 2017: Cuevas-Agulló (2017l)	2016: Cuevas-Agulló (2017a) 2017: Cuevas-Agulló (2018a)
Lindenberg	2016, 2017: Wacker and Behrens (2022)											
Palaiseau	2016, 2017: Haeffelin (2022)											
Payerne	2016, 2017: Vuilleumier and Heimo (2022)											
Petrolina	2016: Pereira (2018a) 2017: Pereira (2018b)	2016: Pereira (2018c) 2017: Pereira (2018d)	2016: Pereira (2018e) 2017: Pereira (2018f)	2016: Pereira (2018g) 2017: Pereira (2018h)	2016: Pereira (2018i) 2017: Pereira (2018j)	2016: Pereira (2018k) 2017: Pereira (2018l)	2016: Pereira (2018m) 2017: Pereira (2018n)	2016: Pereira (2018o) 2017: Pereira (2018p)	2016: Pereira (2018q) 2017: Pereira (2018r)	2016: Pereira (2018s) 2017: Pereira (2018t)	2016: Pereira (2018u) 2017: Pereira (2018v)	2016: Pereira (2018w) 2017: Pereira (2018x)
São Martinho da Serra	2016: Pereira (2018y) 2017: Pereira (2018z)	2016: Pereira (2018aa) 2017: Pereira (2018ab)	2016: Pereira (2018ac) 2017: Pereira (2018ad)	2016: Pereira (2018ae) 2017: Pereira (2018af)	2016: Pereira (2018ag) 2017: Pereira (2018ah)	2016: Pereira (2018ai) 2017: Pereira (2018aj)	2016: Pereira (2018ak) 2017: –	2016: Pereira (2018al) 2017: –	2016: Pereira (2018am) 2017: –	2016: Pereira (2018an) 2017: –	2016: Pereira (2018ao) 2017: –	2016: Pereira (2018ap) 2017: –
Sonnblick	2016, 2017: Olefs (2022)											
Tamanrasset	2016: Mimouni (2016a) 2017: Mimouni (2017b)	2016: Mimouni (2016b) 2017: Mimouni (2017c)	2016: Mimouni (2016c) 2017: Mimouni (2017d)	2016: Mimouni (2016d) 2017: Mimouni (2017e)	2016: Mimouni (2016e) 2017: Mimouni (2017f)	2016: Mimouni (2016f) 2017: Mimouni (2017g)	2016: Mimouni (2016g) 2017: Mimouni (2017h)	2016: Mimouni (2016h) 2017: Mimouni (2017i)	2016: Mimouni (2016i) 2017: Mimouni (2017j)	2016: Mimouni (2016j) 2017: Mimouni (2017k)	2016: Mimouni (2016k) 2017: Baika and Abdessadak (2017)	2016: Mimouni (2017a) 2017: Baika and Abdessadak (2018)
Toravere	2016, 2017: Kallis (2022)											

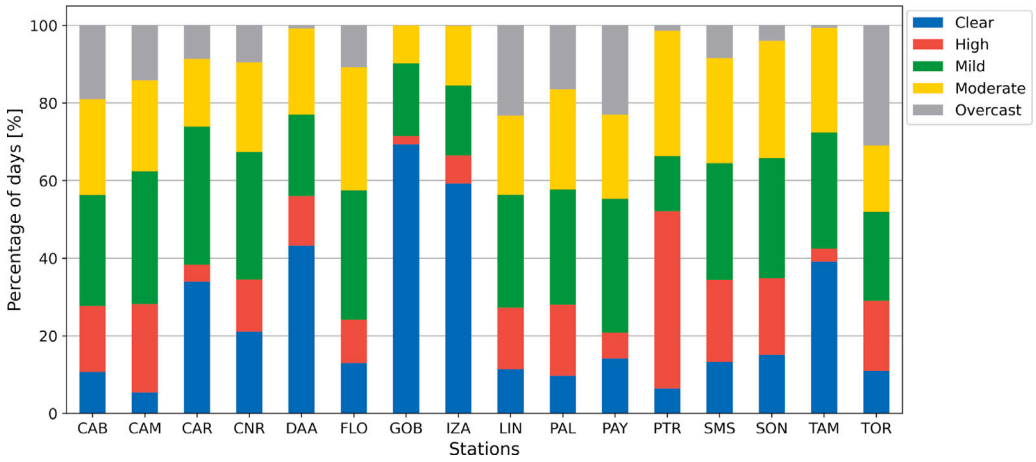


Fig. 4. The distribution of the five variability conditions for each selected station (referred to by its abbreviation, as reported in Table 1).

Table 3
An overview of the 12 channels including central, minimum and maximum wavelength of the SEVIRI channels, their identifiers and their main aim.

Channel identifiers		Characteristics of spectral band [μm]			Additional information
No.	Name	λ_{cen}	λ_{min}	λ_{max}	Aim
1	VIS0.6	0.635	0.56	0.71	Cloud detection and tracking, land surface and vegetation monitoring
2	VIS0.8	0.81	0.74	0.88	Cloud detection and tracking, land surface and vegetation monitoring
3	NIR1.6	1.64	1.50	1.78	Discrimination between snow and cloud, ice and water clouds
4	IR3.9	3.90	3.48	4.36	Low cloud and fog detection
5	WV6.2	6.25	5.35	7.15	Water vapor, winds and semitransparent clouds
6	WV7.3	7.35	6.85	7.85	Water vapor, winds and semitransparent clouds
7	IR8.7	8.70	8.30	9.10	Thin cirrus clouds and discrimination between ice and water clouds
8	IR9.7	9.66	9.38	9.94	Ozone patterns tracking
9	IR10.8	10.80	9.80	11.80	Cirrus cloud and volcanic ash clouds detection
10	IR12.0	12.00	11.00	13.00	Cirrus cloud and volcanic ash clouds detection
11	IR13.4	13.40	12.40	14.40	Cirrus discrimination, cloud top pressure evaluation
12	HRV	Broadband (about 0.4 – 1.1)			Monitoring of convection in high resolution

are collected by the SEVIRI radiometer, the primary instrument on board the MSG satellite. The SEVIRI radiometer provides images on 12 spectral wavelength channels, 8 of which belong to the thermal infrared, 3 to the visible channels and 1 to the near-infrared channel. Table 3 provides a detailed characterization of the 12 channels along with their major scope (Schmetz et al., 2002). Throughout the manuscript, we refer to the channels by using the channel number provided in Table 3.

The multispectral images acquired by the SEVIRI radiometer contain the measured radiance for each pixel. The radiance corresponds to the radiant flux leaving a given surface per unit projected area per unit solid angle and is expressed as $\mathrm{mWm^{-2}sr^{-1}(cm^{-1})^{-1}}$. The pixel radiance can be calibrated to either the bidirectional reflectance [%] for the solar channels or to the brightness temperature [K] for infrared channels. Pixel conversion allows better interpretation of the image information. The images disseminated by the SEVIRI instrument are provided as a pixel grid, the normalized geostationary projection that describes the view from a virtual satellite to an idealized Earth (Combal and Noel, 2009). The nominal Earth full-disk image for channels 1–11 consists of 3172 × 3172 pixels, while channel 12 (HRV) image has a size of 11136 × 5568 pixels. The different image sizes reflect the different spatial resolutions that characterize channels 1–11 and channel 12. In fact, at the sub-satellite point, the images for channels 1–11 have a spatial resolution of 3 km per pixel. Instead, channel 12 has a spatial resolution of 1 km per pixel, about three times higher than the other 11 channels. The full-disk Earth imagining is characterized by a repeat cycle of 15 min, which allows observation of rapidly changing phenomena on the Earth's surface and the up-welling atmosphere (Schmetz et al., 2002). Fig. 5 shows a sample one-hour sequence of channel 12 images over Gobabeb, dated February 16, 2017. In particular, Fig. 5 illustrates the rapidly changing evolution of the clouds, highlighting the temporal variation in pixel intensity, where wither pixels are associated with higher reflectance values, indicating the presence of clouds.

By providing each and all pixels of the multispectral images as input, specific to each timestamp and location, the MLP model can capture tiny variations within the observed scene, thus generating specific and accurate GHI estimates.

The satellite images were downloaded using the *Eumdac* (EUMETSAT, 2022) Python library, which allows users to interact with EUMETSAT's Data Store (EUMETSAT, 2022b) and Data Tailor (EUMETSAT, 2022c) web services. The Data Store provides users with a central access point to EUMETSAT's meteorological, climatic and oceanic data. The Data Tailor allows users to customize the products of interest.

Through the Data Tailor web service, the satellite images were downloaded, selecting the High-Rate Information Transmission (HRIT) format used by EUMETSAT to encode SEVIRI level 1.5 imagery data. For each BSRN station, we extracted 6 × 6 and 16 × 16 pixels images for channels 1–11 and 12, respectively, from the full-disk images, with the BSRN station located in the central pixel. The different sizes of the images reflect the different spatial resolution between channel 12 and the other 11 channels. By selecting these sizes, we consider approximately the same geographical area for all channels, covering approximately an area of 30 × 30 km at the subsatellite point.

For each test BSRN station, the satellite image dataset contains 70 139 multi-channel images of size (6, 6, 11) and 70 139 images of size (16, 16) for channels 1–11 and 12, respectively. Each sample of the available 70 139 refers to each 15-min timestamp for the years 2016 and 2017. The baseline dataset of satellite image would include 70 176 samples for the two selected years. However, 37 samples are not available in the Data Store, and it is not possible to retrieve them, which

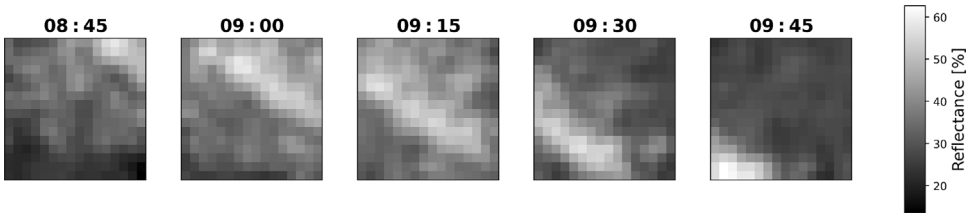


Fig. 5. A sample one-hour sequence of channel 12 images, relative to the 16th of February 2017, representing the evolution of the clouds above Gobabeb. Each pixel of the images includes the calibrated reflectance.

is why they are missing from the image dataset. The resulting satellite image dataset includes 70 139 samples. In Section 4.2, the methods used to further process the satellite images are discussed in detail.

4. Methodology

This work presents an artificial intelligence-based approach to irradiance estimation using multispectral satellite imagery and clear-sky GHI data. A time resolution of 15 min characterizes the inputs. Therefore, instead of estimating hourly or daily solar radiation, the MLP model provides GHI estimates for each 15-min timestamp, starting from the inputs. In this way, it would be possible to infer solar radiation over the target area based on inputs that are easily accessible to all users. Fig. 1 shows the entire pipeline of the proposed methodology. The following sections describe the processing procedures used for the input data, the characteristics of the selected model, and the procedures adopted for the training and testing phases.

4.1. Measured GHI data cleaning

The GHI time series generated by McClear and Heliosat-4 for the selected years 2016 and 2017 do not contain erroneous or missing data, as they were generated by physical models. Therefore, as shown in Fig. 1, GHI data cleaning is applied only to the BSRN ground measurements used to validate the estimation results. Before submission to BSRN servers, quality checks and updates of ground measurements are carried out to ensure high data quality (BSRN, 2022). Nevertheless, BSRN personnel suggests to locally run quality checks to detect physically incorrect data. We used the BSRN Toolbox (Schmithüsen et al., 2012) to perform the quality checks of the measured GHI data and chose the Solpos algorithm without refraction. The measurements that exceed the physically possible and extremely rare limits (conditions marked with quality codes 2 and 8, respectively) were excluded from the measured GHI dataset.

Moreover, the GHI data may have some missing or negative values due to a temporary malfunction of the ground sensor. Such outliers were replaced by two simple procedures: (I) negative solar radiation values, which are physically incorrect, are set to zero; (II) The missing data is filled using a linear regression method. However, if sequences of missing data exceed one hour, the missing data will not be filled as this would produce imprecise GHI values. Consequently, the decision to discard physically incorrect measurements and long sequences of missing data reduces the total number of GHI samples by approximately 1.79%. Table 4 describes the total measured GHI samples before and after data cleaning. In addition, the table also reports the reduction of GHI samples per station due to the violation of BSRN quality checks and the exclusion of missing sequences lasting more than one hour.

4.2. Processing the Meteosat images

According to Fig. 1, the satellite images for our study are manipulated using the *Satpy* Python library (Satpy, 2022), which provides a comprehensive set of operations for processing meteorological remote sensing data.

The input HRIT files, which include the full-disk channel images, are loaded into the Satpy scene, selecting the default calibration for the different channels: brightness temperature and reflectance for infrared and visible channels respectively. No reprojection is performed; Therefore, the images are manipulated using the baseline geostationary projection. The full-disk images are georeferenced, meaning that each image pixel is associated with a pair of geographic coordinates. Using this information and exploiting Satpy's capabilities, we searched for the pixel, including the target BSRN station.

The image was cropped to have a size of 6×6 and 16×16 pixels for channels 1–11 and 12, respectively, with the station located at the central pixel. Given the input HRIT files, relative to a 15-min timestamp, the cropping process is repeated for each analyzed BSRN station and all 12 channels. Fig. 6 shows the adopted cropping process for the VIS 0.6 channel geographically only, just for readability reasons. More in detail, Fig. 6 shows a sample image of Europe, a detail of the full-disk image, before and after the cropping process. For each location, the cropped images are then stored as multidimensional arrays using the Python library *NumPy* (Harris et al., 2020). Such a library performs vectorized operations that quickly perform mathematical operations over a sequence of data, thus efficiently enabling possible manipulations of the input channel images.

Given the 15-min granularity, there are 70 176 timestamps in the two years 2016 and 2017. However, 37 timestamp data are unavailable from the EUMETSAT Data Store. Therefore, after the image processing phase, for each test BSRN station, we collect 70 139 multidimensional arrays of size (16, 16, 11) for channels 1–11 and 70 139 arrays of size (6, 6) for channel 12. The image data for the channels 1–11 and 12 are saved separately due to the different sizes chosen. However, the downloaded and cropped satellite images may contain might include one or more missing pixel values. We will not include incomplete images in either the training and testing phases. Such a choice slightly reduces the size of the satellite image datasets, as explained in Section 4.4.

4.3. Satellite images normalization

Normalization of the input satellite channel images is necessary to give each feature the same magnitude and speed up the training phase of the proposed estimation model. The satellite channel images are extracted using Satpy and each pixel of the images reports either reflectance or brightness temperature values for visible and infrared channels, respectively. The images are normalized using a min–max normalization defined by Eq. (1) so that every pixel value in the entire image dataset falls within the range $[0, 1]$. Scanning all images for each of the 12 channels, the absolute minimum and maximum channel pixel values that are used for the normalization described in Eq. (1).

$$X_{norm,i} = \frac{X_{input,i} - min_i}{max_i - min_i} \tag{1}$$

where, i refers to the channel number, $X_{input,i}$ is the input channel i satellite image, min_i and max_i are the absolute minimum and maximum pixel values for the channel i, respectively. Finally, the normalized channel image is $X_{norm,i}$. In this way, each pixel of the channel image is normalized, producing a fully normalized image. The satellite channel images were processed using the Numpy Python library (Harris et al.,

Table 4
The number of measured GHI samples before and after the data cleaning procedure, divided by station.

Station	Number of measured GHI samples			
	Before data cleaning	After data cleaning	Discarded missing values	Discarded physically incorrect
Cabauw	70 176	69 896	279 (~0.39%)	1 (~0.001%)
Camborne	55 488	55 372	106 (~0.19%)	10 (~0.01%)
Carpentras	70 176	69 394	782 (~1.11%)	–
Cener	70 176	70 121	52 (~0.07%)	3 (~0.004%)
De Aar	70 176	69 083	1089 (~1.55%)	4 (~0.005%)
Florianopolis	70 176	67 628	2507 (~3.57%)	41 (~0.05%)
Gobabeb	70 176	69 127	1463 (~2.08%)	–
Izaña	70 176	69 785	391 (~0.55%)	–
Lindenberg	70 176	70 144	30 (~0.04%)	2 (~0.002%)
Palaiseau	70 176	69 784	392 (~0.55%)	–
Payerne	70 176	69 946	227 (~0.32%)	3 (~0.004%)
Petrolina	70 176	60 884	9292 (~13.24%)	–
São Martinho da Serra	50 932	48 877	2052 (~4.02%)	3 (~0.005%)
Sonnblick	70 176	69 489	590 (~0.84%)	97 (~0.13%)
Tamanrasset	70 176	69 632	542 (~0.77%)	2 (~0.002%)
Toravere	70 176	70 171	2 (~0.002%)	3 (~0.004%)
			19 796 (~1.81%)	169 (~0.02%)
	1 088 884	**1 069 333**	**19551 (~1.79%)**	

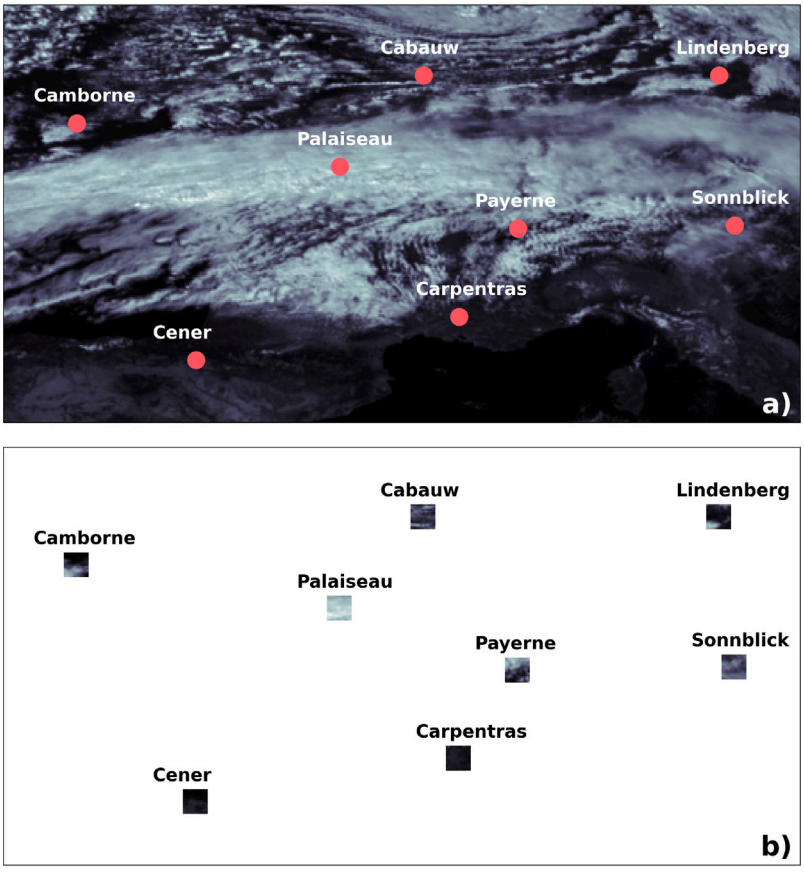


Fig. 6. The adopted cropping procedure depicted geographically. Figure (a) shows a sample image of Europe for channel VIS 0.6 and the relative position of the included BSRN stations; instead, figure (b) shows the cropped images of size 16 × 16 pixels around the areas of interest.

2020), which allows the application of element-wise scalar operations. Thus, the subtraction and division operations are applied to each pixel value of the input channel images, achieving very efficient and fast normalization.

4.4. Selection of the available data

The proposed MLP model estimates solar radiation for a given timestamp and location from the inputs, which are images from all 12 Meteosat channels and the corresponding estimated clear-sky GHI value. Furthermore, our methodology includes the use of measured GHI data in addition to the estimated GHI data from Heliosat-4's, to verify the accuracy of the proposed MLP model during the performance assessment phase. After cleaning the measured GHI data and satellite image processing procedures, the number of available samples inevitably decreases. With the inputs, the MLP model can estimate solar radiation. Using the ground-measured GHI data, it will be possible to evaluate the performance of the proposed model. Therefore, the selection of samples used is strictly tied to the simultaneous availability of satellite images and measured GHI data in order to be able to carry

Table 5
The number of total available samples used during all phases of the proposed methodology, for each location.

Station	Available samples
Cabauw	69 839
Camborne	55 231
Carpentras	69 340
Cener	70 068
De Aar	69 016
Florianopolis	67 567
Gobabeb	68 648
Izaña	69 732
Lindenberg	70 090
Palaiseau	69 729
Payerne	69 889
Petrolina	60 815
São Martinho da Serra	48 774
Sonnblick	69 432
Tamanrasset	69 577
Toravere	70 118

out all methodology's steps. By combining the results of the processing procedures, it is thus possible to precisely define the available data to be used during the training, testing, and performance evaluation phases. Such expedients ensure the consistency of all datasets employed in our methodology. Table 5 lists the total available samples for each station. Each sample refers to a 15-min timestamp of the considered years 2016 and 2017, for which we have the 12-channel satellite images, the estimated GHI in clear-sky value, the measured GHI, and the GHI value estimated by Heliosat-4, all data necessary to fulfill the procedures of the methodology.

4.5. The MLP model

The ANN is a machine learning model inspired by the connection of the neurons in our brain. The perceptron or neuron defines the simplest ANN architecture, a threshold logic unit that maps the vector of n inputs $\boldsymbol{x}$ to a single binary output value. Each input connection is labeled with the weight w_i, and the perceptron calculates the weighted sum of the inputs $z = w_1 x_1 + \cdots + w_n x_n$, which is then given as input to a step function, or activation function, f that outputs the result.

The training procedure for ANNs is based on the back-propagation algorithm, which consists of two passes, the forward pass and backward pass. In the forward pass, the inputs are provided to the input layer, passing through the network to the output layer. The obtained output values $\hat{y}_i$ are compared with the actual outputs y_i, and the error is quantified using a loss function. The algorithm then calculates the error contribution of each output connection, and the error is then propagated back towards the input layer, measuring the error gradients of the network with respect to each parameter of the model.

Finally, all network connection weights are updated by the gradient descent algorithm. By stacking perceptrons, it is possible to define the MLP. It is a class of ANNs characterized by an input layer with several neurons matching the number of inputs, one or more hidden layers, and a final output layer that provides the classification or regression results.

We chose the MLP model to perform the solar radiation estimation, using Meteosat multispectral satellite images and the estimated GHI values in clear-sky conditions as inputs, relative to 16 different BSRN stations. The goal is to build a machine learning model to generalize estimates for different geographically dispersed locations based on the selected inputs.

The MLP model has a fixed structure and to check the quality of the proposed methodology, we perform K-fold cross-validation. Generally, for K-fold cross validation the randomly shuffled dataset is split into K groups that are independent of each other. One group is kept as test set, while the other K-1 groups are used to train the model. In this way, K training procedures are executed changing the composition of test and training sets and imposing that each sample is used only once as a test sample once and K-1 times to train the model. Typically, the composition of the K folds is randomly defined by simply dividing the shuffled dataset into K folds.

For the proposed methodology, the division into K groups is location-based, meaning that for each fold, one out of the 16 available stations defines the test set, while all other 15 stations populate the training set. In this way, it is possible to evaluate the model design and determine its generalization capabilities in estimating solar radiation over an unseen location. Fig. 8 shows all K combinations adopted for the available stations. However, the K groups are not uniform in the number of samples, as shown in Table 5. Therefore, it is possible that the performance of the model under different combinations may vary slightly during the testing phase due to the small variations in the composition of the sets.

The considered location-based K-fold cross validation not only proves the effectiveness of the selected architecture, but also demonstrates the precision of the MLP model in estimating solar radiation over potentially every location included in the Meteosat full-disk images. The limitation to the 16 selected sites is simply due to the availability of ground-measured GHI data to assess the performances of the proposed methodology.

To implement the MLP model, we used the *Keras* Python library (Chollet, 2015), a high-level neural network API running atop *TensorFlow* (Abadi et al., 2015). The structure of the proposed MLP model is shown in Fig. 7. In the same Fig. 7, the model includes three input layers, one for each input, layer 1: 1–11 channel images, layer 2: channel 12 images, and layer 3: GHI values in clear-sky conditions. The images of channel 12 are provided to a different input branch because they are characterized by a different size, larger than those of the other 11 channels. The satellite images are flattened before being fed to the input layers. Each input layer consists of 16 neurons, or units in Fig. 7, and a linear activation function. The output of each input layer is concatenated into a single hidden layer, characterized by 1024 neurons and the ReLu activation function, which provides as output the GHI estimate. The dropout was also included in the model to perform regularization and increase the robustness of the model. Finally, the maximum number of epochs is set to 100, a reasonably high value to improve the generalization of the model.

For training the MLP model, we used the mean squared error as the loss function and we chose the Adam optimizer due to its computational efficiency and small memory requirement (Kingma and Ba, 2017).

We performed a grid search to test different models' hyperparameters values, as shown in Table 6, which helped us determine the most accurate MLP architecture through a trial-and-error approach. For each MLP configuration examined during the grid search, we performed the whole location-based K-fold cross validation discussed previously and we trained the model independently for each combination of stations selected to populate the training and test sets, as shown in Fig. 8.

Once the grid search is complete, for each MLP configuration, we put together the results obtained across the test stations in all combinations to determine the overall optimal MLP configuration that has the lowest error across all stations considered. The grid search's records are reported in Table 7 to ease the discussion of the results.

By changing the composition of training and test sets, following the combinations given in Fig. 8, and training the MLP model from scratch, an independent verification of the test station can be performed. In fact, the test station for each combination completely independent of the other 15 training stations, ensuring the validation of the MLP model at unseen locations.

We also included the Patience as an early stopping condition, which specifies the number of epochs with no improvements, after which the training phase is stopped. Patience is set to 20 epochs so that a sufficient number of neuron combinations generated by the dropout can be examined in the training phase. The tests were carried out with an NVIDIA GeForce RTX 3070 Ti GPU. Depending on the available samples and batch size, each epoch takes approximately 60 s to complete.

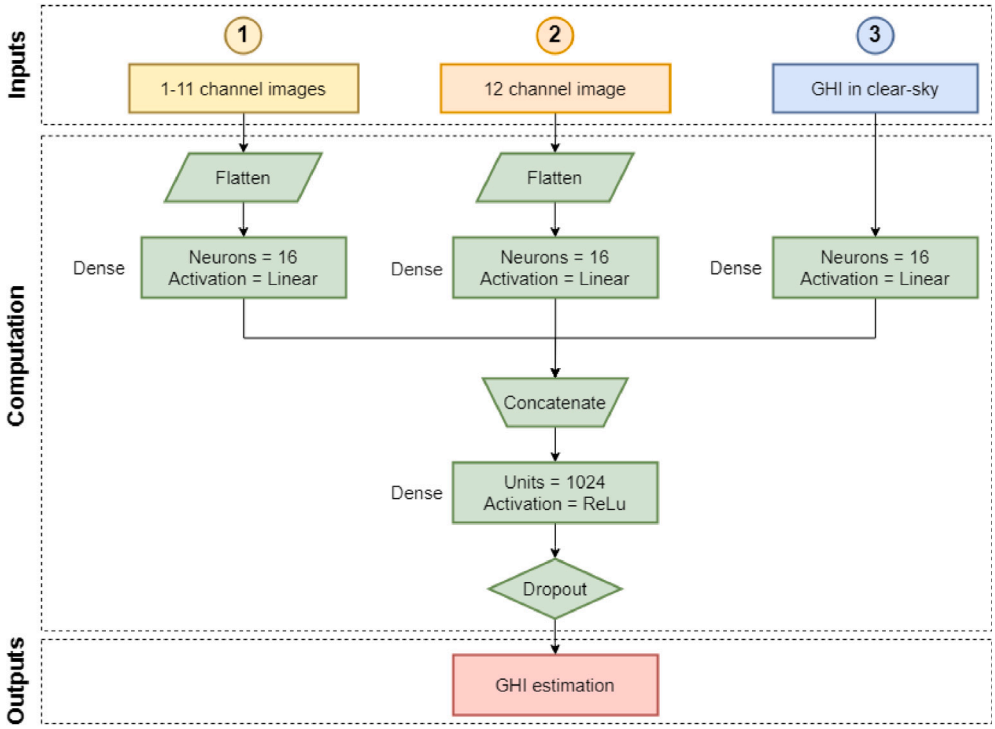


Fig. 7. The structure of the proposed MLP model.

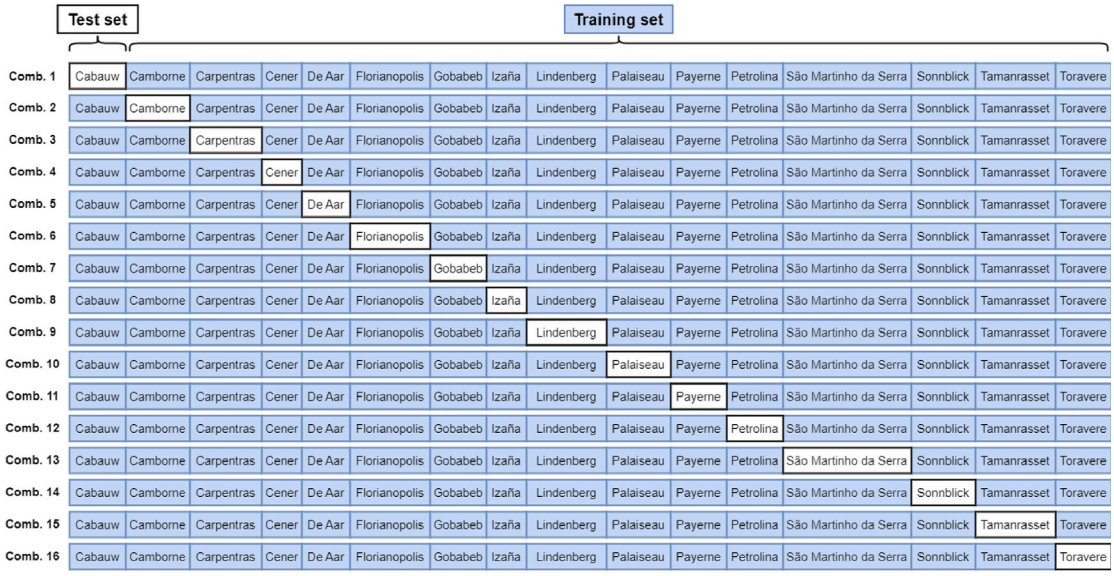


Fig. 8. The possible combinations, across all available stations, used for the location-based K-fold cross validation to populate training and test sets.

Table 6
The hyperparameters investigated in the grid search.

Hyperparameters	Tested values
Batch size	64, 128, 256
Learning rate	0.001, 0.01
Dropout	0.2, 0.5

5. Results

In this section, we present the estimation results of the proposed MLP model. The performance metrics used to quantify the accuracy of the estimates are described. We will unfold the results of the analyzed MLP configurations during the grid search to also select the optimal hyperparameters' configuration for the proposed MLP model, which achieves better results overall. Finally, we compare the estimation results for each test location with those of the Heliosat-4 method to evaluate the benefits of using our methodology.

5.1. Performance metrics

The performances of the proposed MLP model are evaluated using the same metrics used in the literature to assess the accuracy of the Heliosat-4 method, our benchmark method, to enable a fair comparison. The selected performance metrics are the *Root Mean Square Error* (RMSE), the *Relative Root Mean Square Error* (rRMSE), *Coefficient of*

Table 7
The description of the cases investigated in the grid search with Patience set to 200 epochs.

MLP configuration #	Batch size	Learning rate	Dropout
1	64	0.001	0.2
2	128	0.001	0.2
3	256	0.001	0.2
4	64	0.001	0.5
5	128	0.001	0.5
6	256	0.001	0.5
7	64	0.01	0.2
8	128	0.01	0.2
9	256	0.01	0.2
10	64	0.01	0.5
11	128	0.01	0.5
12	256	0.01	0.5

Table 8
The results of the case studies investigated in the grid search in terms of RMSE, rRMSE, R^2, and Bias.

MLP configuration #	RMSE [W/m^2]	rRMSE [%]	R^2	Bias [W/m^2]
1	77.841	23.195	0.929	−1.224
2	78.386	23.374	0.928	−2.082
3	77.804	23.226	0.929	−1.747
4	**77.682**	**23.059**	**0.929**	**−0.033**
5	77.828	23.062	0.929	−1.007
6	78.043	23.081	0.929	−0.311
7	78.591	23.283	0.928	0.331
8	79.319	23.707	0.926	−0.674
9	77.887	23.282	0.929	−2.012
10	78.647	23.405	0.928	−0.932
11	78.558	23.336	0.928	0.125
12	79.032	23.312	0.927	1.054

determination (R^2), and *Bias* (Botchkarev, 2019). The RMSE quantifies the residuals' standard deviation and prediction errors; the rRMSE represents the variation of the RMSE; the R^2 measures the variability in the observed values that can be explained using the predicted values; the Bias measures numerically the trueness between the average value of the estimations and the average of the actual values. In this context, the RMSE and Bias are expressed in terms of W/m^2; the rRMSE is expressed as a percentage, while R^2 does not have unit and it ranges between 0 and 1. In the following, the mathematical formulation of these metrics is reported, where y_{pred} is the predicted value, y_{test} is the observed value, $\bar{y}_{test}$ is the mean value of the observed values, and N is the total number of predictions.

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(y_{pred,n} - y_{test,n})^2}{N}} \tag{2}$$

$$rRMSE = \sqrt{\frac{\frac{1}{N}\sum_{n=1}^{N}(y_{pred,n} - y_{test,n})^2}{\sum_{n=1}^{N}(y_{pred,n})^2}} \times 100 \tag{3}$$

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_{test,n} - y_{pred,n})^2}{\sum_{n=1}^{N}(y_{test,n} - \bar{y}_{test})^2} \tag{4}$$

$$Bias = \frac{\sum_{n=1}^{N} y_{pred,n} - y_{test,n}}{N} \tag{5}$$

5.2. Selection of the optimal MLP configuration

As explained previously, a grid search was performed over the hyperparameters described in Table 6 to determine the optimal configuration for the proposed MLP model. For each grid search MLP configuration, described in Table 7, we performed a full location-based K-fold cross validation, changing the contents of the training and test sets according to the combinations indicated in Fig. 8. In this way, for all MLP configuration, we can assess the robustness and accuracy of the model in generalizing the GHI estimates for different locations.

We used the previously defined performance metrics to quantify the performance of the model. Once all training phases for a given MLP configuration are completed, we calculate the total test errors across all stations. Such a process was repeated for all MLP configurations; Then, the combination of hyperparameters that minimizes such overall errors is selected as the optimal configuration. Table 8 shows the overall RMSE, rRMSE, R^2, and Bias between the measured GHI values and the predicted ones across all test stations of the combinations described in Fig. 8 and for each grid search MLP configuration.

The MLP configuration #4 offers the best performances in terms of all metrics considered. In fact, it achieves an RMSE and R^2 of 77.682 W/m^2 and 0.929, respectively; It also achieves the lowest bias of −0.033 W/m^2 and an rRMSE of 23.059%. The MLP configuration #4 is undoubtedly the best overall, although the errors' deviation from the other configurations remains small. The performances of the MLP for each individual test station, discussed in the following section, are analyzed and compared solely by observing the results relative to the best MLP configuration. The structure of such configuration is shown in Fig. 7, matching the number of neurons in the input and hidden layers and relative activation functions, but with a dropout of 0.5. While for the training phase, we selected a learning rate of 0.001 and a batch size of 64. The complete characterization of the grid search's results is reported in Table 9, Table 10, Table 11, and Table 12 in terms of RMSE, rRMSE, R^2, and Bias, respectively.

5.3. Estimation results and comparison with Heliosat-4

After determining the overall best hyperparameters' configuration for the proposed MLP model, we compare the obtained GHI estimates, characterized by a temporal resolution of 15 min, with those generated using the Heliosat-4 method. Such a comparison is carried out station by station to grasp the improvements locally achieved by the MLP model. Table 13 shows the complete comparison between the MLP estimates and those of Heliosat-4 for each test station considered in our analysis.

Considering the results in Table 13, the proposed artificial intelligence MLP model outperforms the Heliosat-4 method for most of the stations. Indeed, the RMSE improvements range from 0.558 W/m^2 to 10.813 W/m^2 for Carpentras and Tamanrasset stations. The improvements in R^2 are less prominent than those of the RMSE, but they are nevertheless consistent.

Likewise, the improvement in MLP accuracy compared to Heliosat-4 in terms of rRMSE ranges from 0.160% to 3.092% for the Palaiseau and Tamanrasset stations. While the maximum R^2 improvement encountered is 0.014 for the Toravere and São Martinho da Serra stations. Hence, the proposed model can accurately capture the temporal evolution of solar radiation throughout the year. As for the Bias, it is worth noting that the spectral range of the pyranometer instruments used by the BSRN network is between 285 and 2800 nm.

On the other hand, the spectral range in Heliosat-4 ranges between 240 and 4606 nm, following Kato et al. (1999). Such a difference in the spectral range leads to a bias overestimation by Heliosat-4. Therefore, the biases achieved with our methodology are generally lower. Nevertheless, besides cases where the bias improvements are pretty significant, the MLP model improves the accuracy of the GHI estimates overall.

Our proposed methodology is outperformed by Heliosat only at three stations, namely Cener, Izaña, and Lindenberg. Nevertheless, the improvements for Cener and Lindenberg in terms of RMSE and R^2 are negligible with respect to our methodology. In fact, Heliosat provides an RMSE improvement of 0.461 and 0.772 W/m^2, and an R^2 enhancement of 0.001 and 0.003, respectively. For Izaña, Heliosat gives an improvement of 0.008 with respect to R^2, which is negligible. Instead, the RMSE obtained by Heliosat for the same station reaches

Table 9
Grid search results for all MLP configurations and stations in terms of RMSE, expressed in [W/m^2].

Station	MLP configuration #											
	1	2	3	4	5	6	7	8	9	10	11	12
Cabauw	72.825	69.113	68.545	**68.540**	69.747	68.988	70.100	69.200	69.490	70.383	70.247	69.380
Camborne	70.322	69.509	70.357	69.462	70.928	69.348	**69.297**	69.620	69.357	71.200	69.616	71.242
Carpentras	**53.104**	53.246	54.571	53.381	53.623	53.521	53.822	53.960	55.281	54.464	55.499	54.427
Cener	71.161	71.161	71.320	**71.139**	71.657	71.577	71.584	72.103	71.443	72.305	71.889	73.846
De Aar	**71.395**	74.137	72.628	72.198	71.570	71.808	72.094	71.545	71.727	72.439	72.261	71.905
Florianopolis	99.482	94.958	95.776	95.778	95.342	96.215	95.595	95.633	**94.297**	95.925	96.031	94.763
Gobabeb	40.817	47.919	41.845	**40.628**	43.100	41.546	41.889	64.266	43.931	42.382	43.174	42.197
Izaña	84.639	89.071	86.308	82.797	83.769	83.959	85.918	83.161	85.902	83.200	84.188	**82.728**
Lindenberg	69.234	69.213	**68.798**	69.339	69.386	70.299	71.681	69.725	69.899	72.734	73.216	69.922
Palaiseau	**72.474**	73.025	73.137	72.802	74.664	77.689	73.006	73.121	72.992	74.091	73.760	73.850
Payerne	65.698	**64.925**	65.362	66.841	65.321	65.125	71.955	65.995	65.337	66.211	66.043	66.253
Petrolina	103.767	105.514	103.792	104.152	104.221	104.72	104.166	111.144	103.703	**103.494**	105.925	111.758
São Martinho da Serra	89.698	89.040	88.590	89.990	88.199	88.477	88.704	88.840	**87.941**	90.292	89.106	89.281
Sonnblick	**115.508**	117.073	117.011	116.453	117.59	117.843	119.077	117.635	118.223	120.290	118.296	121.256
Tamanrasset	**68.096**	69.357	69.096	68.855	69.121	69.014	69.522	68.191	68.902	68.525	69.042	68.457
Toravere	**69.838**	71.430	70.891	73.751	70.659	70.555	71.439	73.105	71.497	73.311	72.016	73.404

Table 10
Grid search results for all MLP configurations and stations in terms of rRMSE, expressed as a percentage.

Station	MLP configuration #											
	1	2	3	4	5	6	7	8	9	10	11	12
Cabauw	31.212	29.427	29.209	28.844	**28.312**	28.431	29.668	29.352	29.806	29.899	29.491	30.050
Camborne	29.378	28.243	30.466	29.169	28.129	28.748	28.590	28.401	28.855	28.401	29.306	**27.841**
Carpentras	**16.164**	16.302	16.446	16.494	16.677	16.240	16.531	16.501	17.798	16.747	17.806	16.781
Cener	22.853	23.103	22.859	**22.695**	23.412	23.332	23.321	23.151	22.748	23.191	23.134	24.527
De Aar	16.586	16.566	16.594	17.106	**16.431**	16.599	16.533	16.609	16.713	16.912	16.848	16.504
Florianopolis	33.909	29.676	30.137	**28.829**	29.531	30.351	30.688	30.846	29.404	30.286	30.334	29.742
Gobabeb	8.839	10.899	9.237	**8.809**	9.513	8.938	9.158	15.605	9.586	9.255	9.497	9.211
Izaña	19.288	19.943	19.541	17.919	18.328	**17.900**	18.577	18.134	18.966	18.033	18.443	18.205
Lindenberg	29.166	30.107	**29.027**	29.990	30.071	29.574	32.399	30.131	30.026	32.785	34.035	30.573
Palaiseau	**28.605**	29.796	30.106	29.361	31.684	34.411	29.141	29.346	29.411	30.412	28.783	28.699
Payerne	23.005	23.492	23.087	**22.765**	23.006	23.205	23.491	23.623	23.655	24.211	23.563	23.440
Petrolina	25.451	26.013	25.503	25.860	25.290	25.0610	25.388	25.426	25.393	25.493	**24.956**	25.196
São Martinho da Serra	25.031	25.481	25.568	27.018	25.154	25.554	25.496	25.257	**24.988**	25.159	25.225	25.365
Sonnblick	**42.718**	44.502	44.903	44.241	46.017	45.806	45.422	45.390	44.947	45.447	45.633	48.690
Tamanrasset	15.439	15.477	15.582	15.374	15.343	15.696	15.914	15.269	15.905	15.501	15.389	**15.267**
Toravere	**32.489**	34.720	35.040	38.075	33.493	32.902	34.226	36.652	34.169	36.496	33.571	36.659

Table 11
Grid search results for all MLP configurations and stations in terms of R^2.

Station	MLP configuration #											
	1	2	3	4	5	6	7	8	9	10	11	12
Cabauw	0.881	0.893	0.894	**0.894**	0.891	0.893	0.889	0.892	0.891	0.889	0.889	0.892
Camborne	0.894	0.896	0.894	0.896	0.892	**0.897**	0.897	0.896	0.897	0.891	0.896	0.891
Carpentras	0.962	0.962	0.960	**0.962**	0.961	0.961	0.961	0.961	0.959	0.960	0.959	0.960
Cener	0.930	0.930	0.929	**0.93**	0.929	0.929	0.929	0.928	0.929	0.928	0.928	0.924
De Aar	0.959	0.956	0.958	0.958	0.959	**0.959**	0.958	0.959	0.959	0.958	0.958	0.959
Florianopolis	0.879	0.890	0.888	0.888	0.889	0.887	0.889	0.889	**0.892**	0.888	0.888	0.891
Gobabeb	0.988	0.983	0.987	**0.988**	0.986	0.987	0.987	0.970	0.986	0.987	0.986	0.987
Izaña	0.949	0.943	0.947	**0.951**	0.950	0.949	0.947	0.950	0.947	0.950	0.949	0.951
Lindenberg	0.892	0.892	**0.893**	0.891	0.891	0.888	0.884	0.890	0.890	0.880	0.879	0.890
Palaiseau	**0.897**	0.895	0.895	0.896	0.890	0.881	0.895	0.895	0.895	0.892	0.893	0.893
Payerne	0.928	**0.930**	0.929	0.925	0.929	0.929	0.914	0.927	0.929	0.927	0.927	0.927
Petrolina	0.906	0.903	0.906	0.905	0.905	0.904	0.905	0.892	**0.906**	0.906	0.902	0.891
São Martinho da Serra	0.911	0.912	0.913	0.911	0.914	0.914	0.913	0.913	**0.915**	0.910	0.912	0.912
Sonnblick	**0.778**	0.772	0.773	0.775	0.770	0.769	0.764	0.770	0.768	0.760	0.768	0.756
Tamanrasset	**0.965**	0.964	0.964	0.964	0.964	0.964	0.963	0.965	0.964	0.964	0.964	0.965
Toravere	**0.874**	0.868	0.870	0.859	0.871	0.871	0.868	0.862	0.868	0.861	0.866	0.861

76.005 W/m^2, compared to 82.797 W/m^2 obtained by our MLP model. Likewise, Heliosat achieves an rRMSE of 16.707% versus 17.919% of the MLP.

Only in this case does Heliosat significantly outperform our methodology. Such an occurrence is likely due to an underestimation of the clear-sky GHI values provided as input, which often exceed the measured GHI. However, apart from this isolated case, the MLP model can provide accurate GHI estimates for test sites with heterogeneous geographical characteristics. Therefore, the proposed methodology proves to be robust in generalizing across different areas. Such a feature allows the analysis to be extended to any location of interest included in the Meteosat full-disk image, provided the necessary inputs.

A graphical representation of the MLP estimation results can be found in Fig. 9, which shows the scatterplot of the MLP and Heliosat-4 estimated and reference ground GHI values for the Gobabeb, Payerne, and Tamanrasset stations. Fig. 9 also shows the absolute distance between the estimates and the ground truth. Looking at the distribution of points, the MLP's GHI estimates for Gobabeb station are closest to the

Table 12
Grid search results for all MLP configurations and stations in terms of Bias, expressed in [W/m^2].

Station	MLP configuration #											
	1	2	3	4	5	6	7	8	9	10	11	12
Cabauw	−1.218	2.265	−0.568	7.429	2.766	6.188	8.265	4.254	**0.283**	7.592	1.902	−2.791
Camborne	**0.181**	4.083	−3.594	10.08	−1.205	1.231	0.915	5.391	1.788	10.389	0.916	13.057
Carpentras	3.161	−0.355	4.686	**−0.146**	0.746	4.948	2.995	2.188	−7.537	2.889	−8.031	1.597
Cener	**−0.243**	−2.772	1.290	−2.095	2.858	−2.496	−1.350	6.066	4.983	2.243	2.438	4.917
De Aar	**−0.173**	11.568	7.326	2.993	−5.483	0.428	4.434	−0.607	−2.005	−2.393	−1.500	2.349
Florianopolis	−19.123	−4.066	−3.361	**−0.279**	4.680	−3.563	−9.957	−7.594	−2.368	−5.218	−1.546	−4.503
Gobabeb	**−0.621**	−16.956	−6.126	−4.831	1.670	3.458	1.310	−31.762	−4.061	−3.325	−2.409	−1.585
Izaña	−16.166	−7.275	−11.937	−3.961	−1.536	3.484	−2.907	−1.350	−6.274	**−0.926**	−2.259	−3.139
Lindenberg	2.786	−3.280	3.919	−0.639	−0.685	5.187	−3.476	**−0.073**	0.319	−5.632	−8.810	−1.517
Palaiseau	2.482	−2.073	−3.309	−7.764	−2.722	−16.425	−0.625	**0.072**	−0.715	−2.132	3.682	7.084
Payerne	6.583	**−0.001**	4.945	6.461	12.116	2.941	19.770	0.701	1.326	−1.119	4.210	5.074
Petrolina	3.789	**1.576**	1.612	5.038	−1.743	7.139	4.083	36.647	3.584	2.281	13.967	25.390
São Martinho da Serra	6.640	−1.407	−4.441	**−0.011**	−11.291	−1.270	1.138	1.324	0.268	4.860	2.343	−0.794
Sonnblick	**−4.665**	−10.744	−12.886	−13.609	−8.754	−14.235	−11.509	−10.423	−9.519	−10.842	−11.019	−19.648
Tamanrasset	−1.597	5.362	1.031	5.828	4.012	−1.237	−4.046	3.591	−5.008	**0.190**	5.992	4.049
Toravere	1.316	−7.417	−7.365	−2.362	−14.596	**0.119**	−3.217	−13.27	−5.213	−9.419	4.542	−7.857

Table 13
The performance comparison between the proposed MLP model and the Heliosat-4 for the analyzed test stations in terms of the selected metrics.

Stations	MLP				Heliosat-4			
	RMSE [W/m^2]	rRMSE [%]	R^2	Bias [W/m^2]	RMSE [W/m^2]	rRMSE [%]	R^2	Bias [W/m^2]
Cabauw	**68.540**	**28.844**	**0.894**	2.766	70.585	29.825	0.888	**−0.194**
Camborne	**69.462**	29.169	**0.896**	**−1.205**	70.875	**28.555**	0.892	2.992
Carpentras	**53.381**	16.494	**0.962**	**0.746**	53.940	**16.061**	0.961	6.723
Cener	71.139	22.695	0.930	**2.858**	**70.677**	**22.289**	**0.931**	3.425
De Aar	**72.198**	**17.106**	**0.958**	−5.483	76.788	17.712	0.953	**1.531**
Florianopolis	**95.778**	**28.829**	**0.888**	4.680	98.832	29.067	0.881	**3.255**
Gobabeb	**40.628**	**8.809**	**0.988**	**1.670**	47.854	10.436	0.983	−2.753
Izaña	82.797	17.919	0.951	**−1.536**	**76.005**	**16.707**	**0.959**	−8.813
Lindenberg	69.339	29.990	0.891	−0.685	**68.566**	**29.000**	**0.894**	**0.335**
Palaiseau	**72.802**	**29.361**	**0.896**	−2.722	74.946	29.521	0.890	**0.712**
Payerne	**66.841**	**22.765**	**0.925**	12.116	69.131	23.998	0.920	**5.947**
Petrolina	**104.152**	**25.860**	**0.905**	**−1.743**	111.428	26.404	0.892	10.176
São Martinho da Serra	**89.990**	**27.018**	**0.911**	−11.291	96.505	27.539	0.897	**−1.527**
Sonnblick	**116.453**	**44.241**	**0.775**	**−8.754**	118.325	44.787	0.767	−15.582
Tamanrasset	**68.855**	**15.374**	**0.964**	**4.012**	79.668	18.466	0.952	−6.749
Toravere	**73.751**	**38.075**	**0.859**	−14.596	77.318	36.855	0.845	**−3.955**

target values as the points are closer to the perfect prediction line, as also shown in Table 13. Furthermore, when analyzing the distribution of GHI estimates for the Tamanrasset station, it is clear that Heliosat-4 tends to underestimate GHI, while the MLP's predictions are closer to the perfect prediction line. These graphical point distributions are all consistent with the results reported in Table 13.

A clearer insight into the estimation results can be grasped by observing Fig. 10, which puts side-by-side the ground-measured GHI, the MLP's, and Heliosat-4's GHI estimates, respectively.

Specifically, Fig. 10 shows the GHI evolution for five sample days, each belonging to the five daily variability conditions defined in Section 3.2, for three test stations, namely Camborne, São Martinho da Serra, and Tamanrasset, belonging to different continents. In general, overcast days are usually more challenging to handle due to the rapid and unpredictable movement of clouds above, which makes estimating solar radiation quite difficult.

However, looking at Fig. 10, the GHI estimates generated by the MLP model follow the target GHI and capture the temporal evolution of solar radiation for distinct locations as well as for different and adverse meteorological conditions. In fact, the MLP model's GHI estimates for São Martinho da Serra closely follow the evolution of the measured GHI across all observed reference days and daily variability conditions. Over Camborne, the GHI estimates generated by the MLP model appear to be more accurate and precise than Heliosat-4's GHI estimates. Indeed, the MLP model accurately detects solar irradiance peaks during the reference days, especially on cloudy days, in contrast to Heliosat-4, which generates GHI estimates that have a larger deviation from the target measure. A similar trend can be observed in Tamanrasset.

Such results leave room for new solutions and prove the efficiency of the proposed artificial intelligence approach compared to a state-of-the-art physical model.

Finally, for completeness, Fig. 11 compares the estimation errors for the proposed MLP model and Heliosat-4, taking into account the daily variability conditions introduced in Section 3. Fig. 11 shows the deviation from the estimated GHI obtained from the MLP model and Heliosat-4 and the ground-measured GHI with respect to all selected metrics. Specifically, for each station and considering a specific daily variability condition, we selected all estimated GHI values and the reference ground GHI corresponding to the days in the dataset matching those specific conditions. The error between the estimated and actual GHI is then calculated for each station and each variability condition, which distribution is reported in Fig. 4.

Although we provide clear-sky GHI as input to the MLP model, one can see from Fig. 11 that our methodology also produces good results under cloudy skies when the sky is mostly covered by clouds. For example, considering Cambourne station, the RMSE calculated between the MLP's GHI estimates and the relative measured GHI is 16.311 W/m^2 on clear days, while it reaches 19.758 W/m^2 on overcast days. A similar trend can be observed for most stations. It can thus be shown that the input clear-sky GHI is mostly useful to relate the input satellite images to a reference value of solar radiation. Nevertheless, except for a few cases, the lowest RMSE errors occur on clear days. Conversely, the highest rRMSE errors can be observed on cloudy days. As we might expect, the days labeled with high variability are those characterized by higher RMSE errors and lower R^2. Indeed, it is difficult to accurately estimate the GHI when meteorological conditions vary frequently.

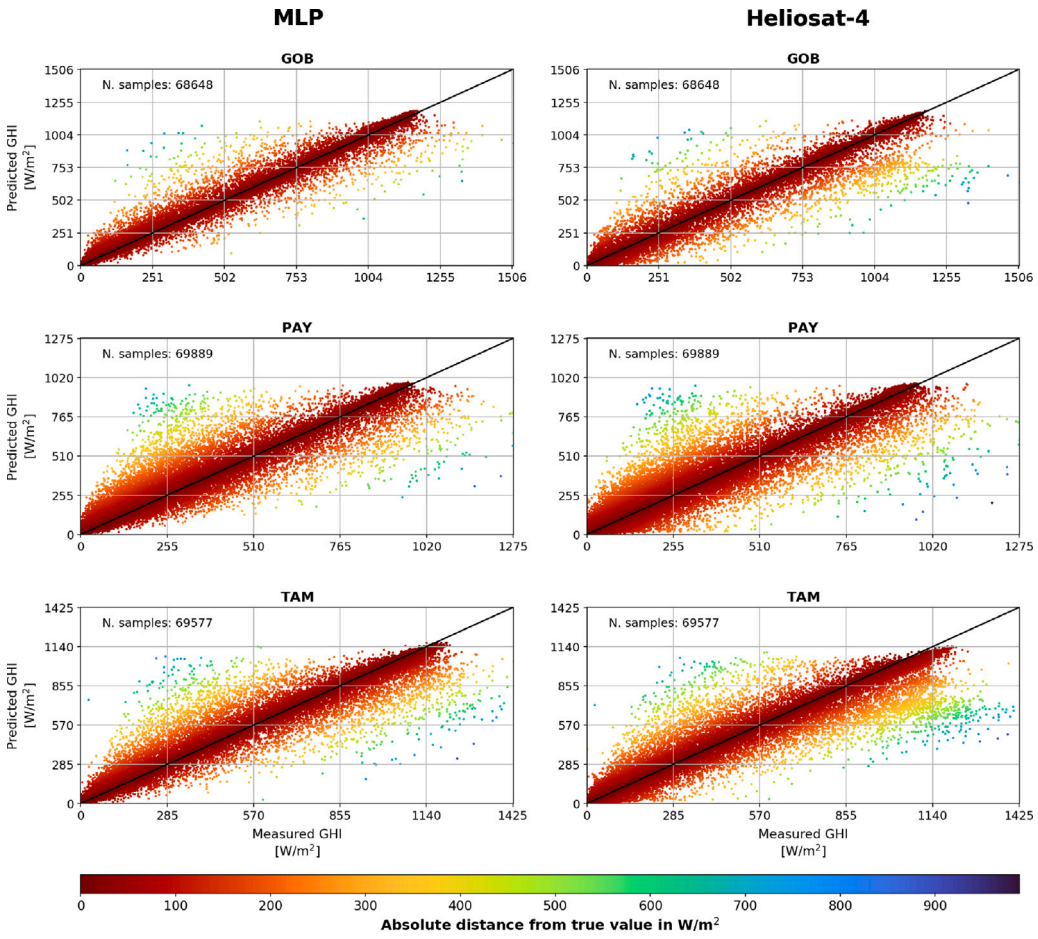


Fig. 9. The scatter plots including the MLP's GHI estimations, relative to the best MLP configuration, and ground-measured GHI values for the Gobabeb, Payerne, and Tamanrasset stations.

Furthermore, observing Fig. 11, it is possible to directly compare the MLP's GHI estimates with those of Heliosat-4 for the same daily conditions. Indeed, it can be observed that Heliosat-4 also has the lowest errors on clear days, and it struggles during highly variable and cloudy meteorological conditions. However, we generally observe higher accuracy of the MLP model when comparing the GHI estimates under the same conditions and the same station.

For example, examining the Petrolina station, the estimates of MLP under clear, mild, moderate, high and overcast conditions in terms of RMSE are 20.571, 58.913, 94.953, 126.989, and 61.12 W/m^2, respectively. On the other hand, for the same station and conditions, Heliosat-4 achieves RMSE values of 27.837, 65.600, 103.159, 134.520 and 69.799 W/m^2. A similar trend can be observed for the rRMSE, R^2, and Bias. However, as already explained in Table 13, Heliosat-4 significantly outperforms the proposed MLP model for the Izaña station alone. Indeed, for such a station, Heliosat-4 achieves an R^2 of 0.988, 0.927, 0.879, 0.816, and 0.344 in clear, mild, moderate, high, and overcast conditions, respectively. While our MLP model achieves an R^2 of 0.981, 0.933, 0.863, 0.772, and 0 under the same conditions. However, apart from one such isolated case, the proposed MLP model generally outperforms the Heliosat-4 benchmark model for the five identified variability conditions and overall.

5.4. Results discussion

Heliosat-4 is a great method to produce solar irradiance estimates and generalize them to any location of interest within the Meteosat full-disk images. Nevertheless, as described in Section 2, the Heliosat-4 method requires aerosol properties, total column water vapor, ozone content, and cloud properties to produce solar radiation estimates. Cloud properties are derived from the Meteosat satellite images using the APOLLO_NG cloud processing scheme, which is a probabilistic interpretation of the original physical APOLLO method for detecting clouds and cloud property (Klüser et al., 2015). Retrieving such inputs requires interacting with specific web services or implementing complex methods.

An advantage of the proposed methodology lies in the selection of inputs. Indeed, the proposed MLP model uses easily accessible inputs, i.e., the Meteosat satellite images and GHI in clear-sky conditions, which require very little processing. Furthermore, the previous analysis of the results highlighted the accuracy of the MLP model over the Heliosat- 4 benchmark method.

The estimates generated by the proposed methodology are generally closer to the measured GHI target. Also, the model can identify solar radiation peaks throughout the day under different meteorological conditions. Such properties demonstrate the efficiency of the MLP model at generalizing heterogeneous locations and conditions with a high confidence level.

Moreover, as mentioned earlier, the acquisition of the required atmospheric inputs for the Heliosat-4 method is challenging and it entails a two-day delay for the estimates' retrieval. In contrast, the proposed MLP model is solely bounded to the primary input, the satellite images, available in near real-time thereby eliminating the Heliosat-4's delay. Therefore, with the proposed solution the analysis can be extended in a near real-time scenario overcoming the structural delay of the selected physic benchmark counterpart.

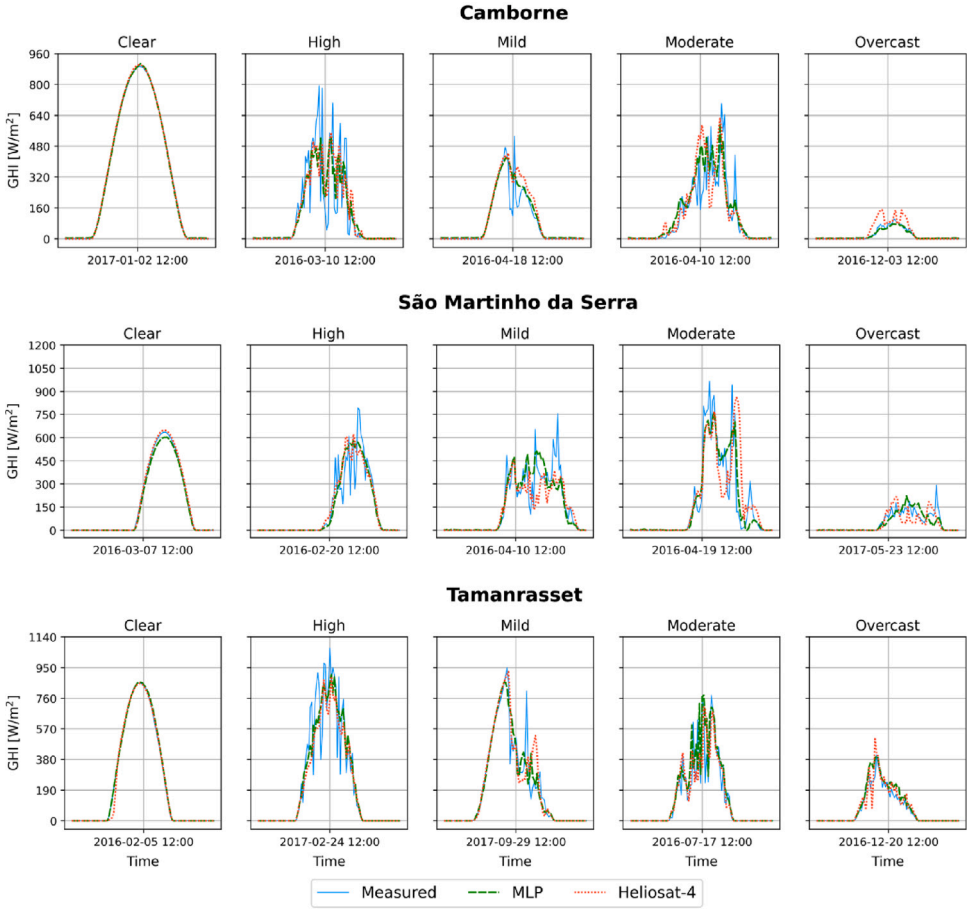


Fig. 10. The MLP's and Heliosat-4's GHI estimates along with the ground-measured GHI for different sample days in UTC time, belonging to each daily variability condition, for the Camborne, São Martinho da Serra, and Tamanrasset stations.

Combining the need for fewer and easily accessible inputs with the accuracy of the results, the MLP model becomes a valuable and promising approach for estimating solar radiation over any area included in the Meteosat full-disk images. Additionally, the model was trained on two years of data, which leaves room for improvement by using larger dataset, including additional years of data.

6. Conclusion

In this work, an MLP model for estimating solar radiation in terms of GHI is presented using Meteosat's multispectral satellite images with GHI values under clear skies conditions. The results showed that the proposed methodology could accurately detect GHI patterns over time for different testing locations scattered over Europe, Africa, and South America.Indeed, through a location-based K-fold cross-validation, the MLP model demonstrated good generalization abilities over unseen locations, suggesting that the model is capable of producing reliable GHI estimates potentially for any location of interest included in the Meteosat images.

To evaluate the efficiency of artificial intelligence models against a physical counterpart, we also compared the estimation results of the MLP with those generated by the state-of-the-art Heliosat-4 model. The proposed methodology significantly outperformed the physical model for most selected test stations for all performance metrics considered and for distinct meteorological conditions. In this way, we have collectively demonstrated the strength of the developed MLP model in generalizing across different locations and its accuracy compared to a complex, reliable physical model. Nevertheless, the MLP model presented was trained with two years of data and temporally extending the dataset would provide even more precise results. For future works, it would be interesting to reproduce the methodology at a zonal level, i.e. to define an MLP model for different climate zones, rather than a global estimation model. Additionally, we would test different input clear-sky models and extend the analysis to forecasting. In fact, the MLP model could be used to estimate the GHI from future satellite images created using optical flow methods, thus achieving reliable and accurate forecast results over potentially any location of interest covered by the Meteosat images.

CRediT authorship contribution statement

Raimondo Gallo: Writing – original draft, Software. **Marco Castangia:** Resources, Data curation. **Alberto Macii:** Validation, Investigation. **Edoardo Patti:** Writing – review & editing, Methodology. **Alessandro Aliberti:** Writing – review & editing, Supervision, Methodology.

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.

Data availability

Data will be made available on request.

Fig. 11. The performance comparison between the proposed MLP model and the Heliosat-4 for the analyzed test stations (referred to by their abbreviations, as reported in Table 1) in terms of RMSE for the five daily variability conditions.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. software available from tensorflow.org, URL https://www.tensorflow.org/.

Al-Abbadi, N., Forgan, B., Lyubansky, V., Vogt, R., 2017. Basic and other measurements of radiation at stations Alice Springs, Gobabeb, Sede Boquer and Solar Village, reference list of 48 datasets. http://dx.doi.org/10.1594/PANGAEA.875769.

Baika, S., Abdessadak, S., 2017. Basic measurements of radiation at station Tamanrasset (2017-11). http://dx.doi.org/10.1594/PANGAEA.884030.

Baika, S., Abdessadak, S., 2018. Basic measurements of radiation at station Tamanrasset (2017-12). http://dx.doi.org/10.1594/PANGAEA.885366.

Besharat, F., Dehghan, A.A., Faghih, A.R., 2013. Empirical models for estimating global solar radiation: A review and case study. Renew. Sustain. Energy Rev. 21, 798–821. http://dx.doi.org/10.1016/j.rser.2012.12.043, URL https://www.sciencedirect.com/science/article/pii/S1364032112007484.

Botchkarev, A., 2019. A new typology design of performance metrics to measure errors in machine learning regression algorithms. Interdiscip. J. Inf. Knowl. Manag. 14, 045–076. http://dx.doi.org/10.28945/4184.

Brunier, L., Duprat, T., Morel, J.-P., Olivieri, J., 2021. Basic measurements of radiation at station Carpentras (1996-09 - 2018-12). http://dx.doi.org/10.1594/PANGAEA.932930.

BSRN, 2022. Bsrn data retrieval from ftp. https://bsrn.awi.de/data/data-retrieval-via-ftp/. (Accessed 15 September 2022).

BSRN, 2022. Bsrn quality checks. https://bsrn.awi.de/data/quality-checks/. (Accessed 15 September 2022).

Chen, W., Li, D.H.W., Li, S., Lam, J.C., 2019. Estimating hourly global solar irradiance using artificial neural networks – a case study of Hong Kong. IOP Conf. Ser. Mater. Sci. Eng. 556 (1), 012043. http://dx.doi.org/10.1088/1757-899x/556/1/012043.

Choi, M., Rachunok, B., Nateghi, R., 2021. Short-term solar irradiance forecasting using convolutional neural networks and cloud imagery. Environ. Res. Lett. 16 (4), 044045. http://dx.doi.org/10.1088/1748-9326/abe06d.

Chollet, F., 2015. keras. https://github.com/fchollet/keras.

Choudhary, A., Pandey, D., Bhardwaj, S., 2020. Overview of solar radiation estimation techniques with development of solar radiation model using artificial neural network. Adv. Sci. Technol. Eng. Syst. J. 5 (4), 589–593. http://dx.doi.org/10.25046/aj050469.

Colle, S., 2016a. Basic measurements of radiation at station Florianopolis (2016-01). http://dx.doi.org/10.1594/PANGAEA.859090.

Colle, S., 2016b. Basic measurements of radiation at station Florianopolis (2016-02). http://dx.doi.org/10.1594/PANGAEA.859826.

Colle, S., 2016c. Basic measurements of radiation at station Florianopolis (2016-03). http://dx.doi.org/10.1594/PANGAEA.859427.

Colle, S., 2016d. Basic measurements of radiation at station Florianopolis (2016-04). http://dx.doi.org/10.1594/PANGAEA.867793.

Colle, S., 2016e. Basic measurements of radiation at station Florianopolis (2016-06). http://dx.doi.org/10.1594/PANGAEA.867794.

Colle, S., 2016f. Basic measurements of radiation at station Florianopolis (2016-07). http://dx.doi.org/10.1594/PANGAEA.867795.

Colle, S., 2016g. Basic measurements of radiation at station Florianopolis (2016-08). http://dx.doi.org/10.1594/PANGAEA.867796.

Colle, S., 2016h. Basic measurements of radiation at station Florianopolis (2016-09). http://dx.doi.org/10.1594/PANGAEA.867797.

Colle, S., 2017a. Basic measurements of radiation at station Florianopolis (2016-05). http://dx.doi.org/10.1594/PANGAEA.871811.

Colle, S., 2017b. Basic measurements of radiation at station Florianopolis (2016-10). http://dx.doi.org/10.1594/PANGAEA.871812.

Colle, S., 2017c. Basic measurements of radiation at station Florianopolis (2016-11). http://dx.doi.org/10.1594/PANGAEA.871813.

Colle, S., 2017d. Basic measurements of radiation at station Florianopolis (2016-12). http://dx.doi.org/10.1594/PANGAEA.871814.

Colle, S., 2017e. Basic measurements of radiation at station Florianopolis (2017-01). http://dx.doi.org/10.1594/PANGAEA.871815.

Colle, S., 2017f. Basic measurements of radiation at station Florianopolis (2017-02). http://dx.doi.org/10.1594/PANGAEA.880553.

Colle, S., 2017g. Basic measurements of radiation at station Florianopolis (2017-03). http://dx.doi.org/10.1594/PANGAEA.882517.

Colle, S., 2017h. Basic measurements of radiation at station Florianopolis (2017-04). http://dx.doi.org/10.1594/PANGAEA.880554.

Colle, S., 2017i. Basic measurements of radiation at station Florianopolis (2017-05). http://dx.doi.org/10.1594/PANGAEA.880555.

Colle, S., 2017j. Basic measurements of radiation at station Florianopolis (2017-06). http://dx.doi.org/10.1594/PANGAEA.882518.

Colle, S., 2017k. Basic measurements of radiation at station Florianopolis (2017-07). http://dx.doi.org/10.1594/PANGAEA.880556.

Colle, S., 2017l. Basic measurements of radiation at station Florianopolis (2017-08). http://dx.doi.org/10.1594/PANGAEA.882519.

Colle, S., 2018a. Basic measurements of radiation at station Florianopolis (2017-09). http://dx.doi.org/10.1594/PANGAEA.892110.

Colle, S., 2018b. Basic measurements of radiation at station Florianopolis (2017-10). http://dx.doi.org/10.1594/PANGAEA.892112.

Colle, S., 2018c. Basic measurements of radiation at station Florianopolis (2017-11). http://dx.doi.org/10.1594/PANGAEA.892113.

Colle, S., 2018d. Basic measurements of radiation at station Florianopolis (2017-12). http://dx.doi.org/10.1594/PANGAEA.892114.

Combal, B., Noel, J., 2009. Projection of meteosat images into world geodetic system wgs-84 matching spot/vegetation grid. http://dx.doi.org/10.2788/2452.

Copernicus, 2022. Cams solar radiation time-series. copernicus atmosphere monitoring service (cams) atmosphere data store (ads). https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-solar-radiation-timeseries. (Accessed 15 September 2022).

Cuevas-Agulló, E., 2016a. Basic measurements of radiation at station Izana (2016-01). http://dx.doi.org/10.1594/PANGAEA.858260.

Cuevas-Agulló, E., 2016b. Basic measurements of radiation at station Izana (2016-02). http://dx.doi.org/10.1594/PANGAEA.858752.

Cuevas-Agulló, E., 2016c. Basic measurements of radiation at station Izana (2016-03). http://dx.doi.org/10.1594/PANGAEA.861205.

Cuevas-Agulló, E., 2016d. Basic measurements of radiation at station Izana (2016-04). http://dx.doi.org/10.1594/PANGAEA.860345.

Cuevas-Agulló, E., 2016e. Basic measurements of radiation at station Izana (2016-06). http://dx.doi.org/10.1594/PANGAEA.863179.

Cuevas-Agulló, E., 2016f. Basic measurements of radiation at station Izana (2016-07). http://dx.doi.org/10.1594/PANGAEA.863592.

Cuevas-Agulló, E., 2016g. Basic measurements of radiation at station Izana (2016-08). http://dx.doi.org/10.1594/PANGAEA.864737.

Cuevas-Agulló, E., 2016h. Basic measurements of radiation at station Izana (2016-09). http://dx.doi.org/10.1594/PANGAEA.865241.

Cuevas-Agulló, E., 2016i. Basic measurements of radiation at station Izana (2016-10). http://dx.doi.org/10.1594/PANGAEA.868072.

Cuevas-Agulló, E., 2016j. Basic measurements of radiation at station Izana (2016-11). http://dx.doi.org/10.1594/PANGAEA.869178.

Cuevas-Agulló, E., 2017a. Basic and other measurements of radiation at station Izana (2016-12). http://dx.doi.org/10.1594/PANGAEA.871032.

Cuevas-Agulló, E., 2017b. Basic and other measurements of radiation at station Izana (2017-01). http://dx.doi.org/10.1594/PANGAEA.871856.

Cuevas-Agulló, E., 2017c. Basic and other measurements of radiation at station Izana (2017-02). http://dx.doi.org/10.1594/PANGAEA.873293.

Cuevas-Agulló, E., 2017d. Basic and other measurements of radiation at station Izana (2017-03). http://dx.doi.org/10.1594/PANGAEA.874307.

Cuevas-Agulló, E., 2017e. Basic and other measurements of radiation at station Izana (2017-04). http://dx.doi.org/10.1594/PANGAEA.875157.

Cuevas-Agulló, E., 2017f. Basic and other measurements of radiation at station Izana (2017-05). http://dx.doi.org/10.1594/PANGAEA.875905.

Cuevas-Agulló, E., 2017g. Basic and other measurements of radiation at station Izana (2017-06). http://dx.doi.org/10.1594/PANGAEA.877415.

Cuevas-Agulló, E., 2017h. Basic and other measurements of radiation at station Izana (2017-07). http://dx.doi.org/10.1594/PANGAEA.879319.

Cuevas-Agulló, E., 2017i. Basic and other measurements of radiation at station Izana (2017-08). http://dx.doi.org/10.1594/PANGAEA.880342.

Cuevas-Agulló, E., 2017j. Basic and other measurements of radiation at station Izana (2017-09). http://dx.doi.org/10.1594/PANGAEA.881601.

Cuevas-Agulló, E., 2017k. Basic and other measurements of radiation at station Izana (2017-10). http://dx.doi.org/10.1594/PANGAEA.882526.

Cuevas-Agulló, E., 2017l. Basic and other measurements of radiation at station Izana (2017-11). http://dx.doi.org/10.1594/PANGAEA.883959.

Cuevas-Agulló, E., 2018a. Basic and other measurements of radiation at station Izana (2017-12). http://dx.doi.org/10.1594/PANGAEA.884897.

Cuevas-Agulló, E., 2018b. Basic measurements of radiation at station Izana (2016-05). http://dx.doi.org/10.1594/PANGAEA.884733.

Dorvlo, A.S., Jervase, J.A., Al-Lawati, A., 2002. Solar radiation estimation using artificial neural networks. Appl. Energy 71 (4), 307–319. http://dx.doi.org/10.1016/S0306-2619(02)00016-8, URL https://www.sciencedirect.com/science/article/pii/S0306261902000168.

Emery, W., Camps, A., 2017a. Chapter 10 - land applications. In: Emery, W., Camps, A. (Eds.), Introduction to Satellite Remote Sensing. Elsevier, pp. 701–766. http://dx.doi.org/10.1016/B978-0-12-809254-5.00010-5, URL https://www.sciencedirect.com/science/article/pii/B9780128092545000105.

Emery, W., Camps, A., 2017b. Chapter 8 - atmosphere applications. In: Emery, W., Camps, A. (Eds.), Introduction to Satellite Remote Sensing. Elsevier, pp. 597–636. http://dx.doi.org/10.1016/B978-0-12-809254-5.00008-7, URL https://www.sciencedirect.com/science/article/pii/B9780128092545000087.

Emery, W., Camps, A., 2017c. Chapter 9 - ocean applications. In: Emery, W., Camps, A. (Eds.), Introduction to Satellite Remote Sensing. Elsevier, pp. 637–699. http://dx.doi.org/10.1016/B978-0-12-809254-5.00009-9, URL https://www.sciencedirect.com/science/article/pii/B9780128092545000099.

EUMETSAT, 2022. Eumdac - python client for eumetsat data access services. https://eumetsatspace.atlassian.net/wiki/spaces/EUMDAC/overview. (Accessed 15 September 2022).

EUMETSAT, 2022a. 0 degree service. https://www.eumetsat.int/0-degree-service. (Accessed 15 September 2022).

EUMETSAT, 2022b. Eumetsat data store. https://www.eumetsat.int/eumetsat-data-store. (Accessed 15 September 2022).

EUMETSAT, 2022c. Eumetsat data tailor. https://www.eumetsat.int/data-tailor. (Accessed 15 September 2022).

EUMETSAT, 2022d. EUMETSAT meteosat second generation (msg). https://www.eumetsat.int/meteosat-second-generation. (Accessed 15 September 2022).

EUMETSAT, 2022e. Seviri. https://www.eumetsat.int/seviri. (Accessed 15 September 2022).

Fallahi, S., Amanollahi, J., Tzanis, C.G., Ramli, M.F., 2018. Estimating solar radiation using noaa/avhrr and ground measurement data. Atmos. Res. 199, 93–102. http://dx.doi.org/10.1016/j.atmosres.2017.09.006, URL https://www.sciencedirect.com/science/article/pii/S0169809517308062.

Gallo, R., Castangia, M., Macii, A., Macii, E., Patti, E., Aliberti, A., 2022. Solar radiation forecasting with deep learning techniques integrating geostationary satellite images. Eng. Appl. Artif. Intell. 116, 105493. http://dx.doi.org/10.1016/j.engappai.2022.105493, URL https://www.sciencedirect.com/science/article/pii/S0952197622004833.

Goncu, O., Koroglu, T., Ozdil, N.F., 2021. Estimation of hourly global solar radiation using artificial neural network in Adana Province, Turkey. J. Therm. Eng. 7 (5), http://dx.doi.org/10.18186/thermal.1051313.

Gschwind, B., Wald, L., Blanc, P., Lefèvre, M., Schroedter-Homscheidt, M., Arola, A., 2019. Improving the mcclear model estimating the downwelling solar radiation at ground level in cloud-free conditions - mcclear-v3. Meteorol. Z. 28 (2), 147–163. http://dx.doi.org/10.1127/metz/2019/0946.

Guijo-Rubio, D., Durán-Rosal, A., Gutiérrez, P., Gómez-Orellana, A., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S., Hervás-Martínez, C., 2020. Evolutionary artificial neural networks for accurate solar radiation prediction. Energy 210, 118374. http://dx.doi.org/10.1016/j.energy.2020.118374, URL https://www.sciencedirect.com/science/article/pii/S036054422031481X.

Haeffelin, M., 2022. Basic measurements of radiation at station Palaiseau (2003-06 et seq). http://dx.doi.org/10.1594/PANGAEA.946383.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585 (7825), 357–362. http://dx.doi.org/10.1038/s41586-020-2649-2.

Hodgetts, G., Fishwick, P., Tamlyn, J., 2022. Basic measurements of radiation at station Camborne (2001-01 et seq). http://dx.doi.org/10.1594/PANGAEA.945262.

Holmgren, W., Calama-Consulting, Hansen, C., Anderson, K., Mikofski, M., Lorenzo, A., Krien, U., bmu, Stark, C., Jensen, A.R., DaCoEx, Driesse, A., de León Peque, M.S., kt, mayudong, Heliolytics, Miller, E., Anoma, M.A., Guo, V., Boeman, L., Stein, J., Vining, W., jforbess, Lunel, T., Morgan, A., Ranalli, J., Carlosbogo, Leroy, C., A.M.R, 2022. pvlib/pvlib-python: v0.9.1. http://dx.doi.org/10.5281/zenodo.6395177.

Holmgren, W.F., Hansen, C.W., Mikofski, M.A., 2018. pvlib Python: a Python package for modeling solar energy systems. J. Open Source Softw. 3 (29), 884. http://dx.doi.org/10.21105/joss.00884.

IEA, 2022. Electricity Market Report - 2022. IEA, Paris, https://www.iea.org/reports/electricity-market-report-january-2022. (Accessed 15 September 2022).

Jiang, H., Lu, N., Huang, G., Yao, L., Qin, J., Liu, H., 2020. Spatial scale effects on retrieval accuracy of surface solar radiation using satellite data. Appl. Energy 270, 115178. http://dx.doi.org/10.1016/j.apenergy.2020.115178, URL https://www.sciencedirect.com/science/article/pii/S0306261920306905.

Kallis, A., 2022. Basic and other measurements of radiation at station Toravere (1999-01 et seq). http://dx.doi.org/10.1594/PANGAEA.946387.

Kato, S., Ackerman, T.P., Mather, J.H., Clothiaux, E.E., 1999. The k-distribution method and correlated-k approximation for a shortwave radiative transfer model. J. Quant. Spectrosc. Radiat. Transfer 62 (1), 109–121. http://dx.doi.org/10.1016/S0022-4073(98)00075-2, URL https://www.sciencedirect.com/science/article/pii/S0022407398000752.

Khan, M., Huque, S., Mohammad, A., 2014. A neural network model for estimating global solar radiation on horizontal surface. pp. 1–4. http://dx.doi.org/10.1109/EICT.2014.6777857.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. arXiv:1412.6980.

Klüser, L., Killius, N., Gesell, G., 2015. Apollo_ng - a probabilistic interpretation of the apollo legacy for avhrr heritage channels. Atmos. Meas. Tech. 8 (10), 4155–4170. http://dx.doi.org/10.5194/amt-8-4155-2015, URL https://amt.copernicus.org/articles/8/4155/2015/.

Knap, W., 2022. Basic and other measurements of radiation at station Cabauw (2005-02 et seq). http://dx.doi.org/10.1594/PANGAEA.940531.

Koo, Y., Oh, M., Kim, S.-M., Park, H.-D., 2020. Estimation and mapping of solar irradiance for korea by using coms mi satellite images and an artificial neural network model. Energies 13 (2), http://dx.doi.org/10.3390/en13020301, URL https://www.mdpi.com/1996-1073/13/2/301.

Kurniawan, A., Harumwidiah, A., 2021. An evaluation of the artificial neural network based on the estimation of daily average global solar radiation in the city of Surabaya. Indones. J. Electr. Eng. Comput. Sci. 22, 1245. http://dx.doi.org/10.11591/ijeecs.v22.i3.pp1245-1250.

Kurniawan, A., Shintaku, E., 2021. Two-step artificial neural network to estimate the solar radiation at java island. Int. J. Electr. Comput. Eng. (IJECE) 11, 3559. http://dx.doi.org/10.11591/ijece.v11i4.pp3559-3566.

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J.W., Morcrette, J.-J., 2013. Mcclear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos. Meas. Tech. 6 (9), 2403–2418. http://dx.doi.org/10.5194/amt-6-2403-2013, URL https://amt.copernicus.org/articles/6/2403/2013/.

McArthur, L., 2005. Baseline surface radiation network (bsrn). operations manual. wmo/td-no. 1274, wcrp/wmo. https://bsrn.awi.de/fileadmin/user_upload/bsrn.awi.de/Publications/McArthur.pdf. (Last accessed 30 September 2022).

Mimouni, M., 2016a. Basic measurements of radiation at station Tamanrasset (2016-01). http://dx.doi.org/10.1594/PANGAEA.858083.

Mimouni, M., 2016b. Basic measurements of radiation at station Tamanrasset (2016-02). http://dx.doi.org/10.1594/PANGAEA.858706.

Mimouni, M., 2016c. Basic measurements of radiation at station Tamanrasset (2016-03). http://dx.doi.org/10.1594/PANGAEA.859374.

Mimouni, M., 2016d. Basic measurements of radiation at station Tamanrasset (2016-04). http://dx.doi.org/10.1594/PANGAEA.860678.

Mimouni, M., 2016e. Basic measurements of radiation at station Tamanrasset (2016-05). http://dx.doi.org/10.1594/PANGAEA.861270.

Mimouni, M., 2016f. Basic measurements of radiation at station Tamanrasset (2016-06). http://dx.doi.org/10.1594/PANGAEA.865125.

Mimouni, M., 2016g. Basic measurements of radiation at station Tamanrasset (2016-07). http://dx.doi.org/10.1594/PANGAEA.867459.

Mimouni, M., 2016h. Basic measurements of radiation at station Tamanrasset (2016-08). http://dx.doi.org/10.1594/PANGAEA.868940.

Mimouni, M., 2016i. Basic measurements of radiation at station Tamanrasset (2016-09). http://dx.doi.org/10.1594/PANGAEA.868941.

Mimouni, M., 2016j. Basic measurements of radiation at station Tamanrasset (2016-10). http://dx.doi.org/10.1594/PANGAEA.867781.

Mimouni, M., 2016k. Basic measurements of radiation at station Tamanrasset (2016-11). http://dx.doi.org/10.1594/PANGAEA.869110.

Mimouni, M., 2017a. Basic measurements of radiation at station Tamanrasset (2016-12). http://dx.doi.org/10.1594/PANGAEA.870586.

Mimouni, M., 2017b. Basic measurements of radiation at station Tamanrasset (2017-01). http://dx.doi.org/10.1594/PANGAEA.874637.

Mimouni, M., 2017c. Basic measurements of radiation at station Tamanrasset (2017-02). http://dx.doi.org/10.1594/PANGAEA.874641.

Mimouni, M., 2017d. Basic measurements of radiation at station Tamanrasset (2017-03). http://dx.doi.org/10.1594/PANGAEA.874902.

Mimouni, M., 2017e. Basic measurements of radiation at station Tamanrasset (2017-04). http://dx.doi.org/10.1594/PANGAEA.876323.

Mimouni, M., 2017f. Basic measurements of radiation at station Tamanrasset (2017-05). http://dx.doi.org/10.1594/PANGAEA.876325.

Mimouni, M., 2017g. Basic measurements of radiation at station Tamanrasset (2017-06). http://dx.doi.org/10.1594/PANGAEA.878224.

Mimouni, M., 2017h. Basic measurements of radiation at station Tamanrasset (2017-07). http://dx.doi.org/10.1594/PANGAEA.879786.

Mimouni, M., 2017i. Basic measurements of radiation at station Tamanrasset (2017-08). http://dx.doi.org/10.1594/PANGAEA.881068.

Mimouni, M., 2017j. Basic measurements of radiation at station Tamanrasset (2017-09). http://dx.doi.org/10.1594/PANGAEA.881883.

Mimouni, M., 2017k. Basic measurements of radiation at station Tamanrasset (2017-10). http://dx.doi.org/10.1594/PANGAEA.883056.

Müller, R., Pfeifroth, U., 2022. Remote sensing of solar surface radiation – a reflection of concepts, applications and input data based on experience with the effective cloud albedo. Atmos. Meas. Tech. 15 (5), 1537–1561. http://dx.doi.org/10.5194/amt-15-1537-2022, URL https://amt.copernicus.org/articles/15/1537/2022/.

Ntsangwane, L., 2017a. Basic measurements of radiation at station De Aar (2016-01). http://dx.doi.org/10.1594/PANGAEA.879358.

Ntsangwane, L., 2017b. Basic measurements of radiation at station De Aar (2016-02). http://dx.doi.org/10.1594/PANGAEA.879359.

Ntsangwane, L., 2017c. Basic measurements of radiation at station De Aar (2016-03). http://dx.doi.org/10.1594/PANGAEA.879360.

Ntsangwane, L., 2017d. Basic measurements of radiation at station De Aar (2016-04). http://dx.doi.org/10.1594/PANGAEA.879361.

Ntsangwane, L., 2017e. Basic measurements of radiation at station De Aar (2016-05). http://dx.doi.org/10.1594/PANGAEA.879362.

Ntsangwane, L., 2017f. Basic measurements of radiation at station De Aar (2016-06). http://dx.doi.org/10.1594/PANGAEA.879363.

Ntsangwane, L., 2017g. Basic measurements of radiation at station De Aar (2016-07). http://dx.doi.org/10.1594/PANGAEA.879364.

Ntsangwane, L., 2017h. Basic measurements of radiation at station De Aar (2016-08). http://dx.doi.org/10.1594/PANGAEA.879365.

Ntsangwane, L., 2017i. Basic measurements of radiation at station De Aar (2016-09). http://dx.doi.org/10.1594/PANGAEA.879473.

Ntsangwane, L., 2017j. Basic measurements of radiation at station De Aar (2016-10). http://dx.doi.org/10.1594/PANGAEA.879366.

Ntsangwane, L., 2017k. Basic measurements of radiation at station De Aar (2016-11). http://dx.doi.org/10.1594/PANGAEA.879367.

Ntsangwane, L., 2017l. Basic measurements of radiation at station De Aar (2016-12). http://dx.doi.org/10.1594/PANGAEA.879368.

Ntsangwane, L., 2017m. Basic measurements of radiation at station De Aar (2017-01). http://dx.doi.org/10.1594/PANGAEA.879474.

Ntsangwane, L., 2017n. Basic measurements of radiation at station De Aar (2017-02).

Ntsangwane, L., 2017o. Basic measurements of radiation at station De Aar (2017-03). http://dx.doi.org/10.1594/PANGAEA.879476.

Ntsangwane, L., 2017p. Basic measurements of radiation at station De Aar (2017-04). http://dx.doi.org/10.1594/PANGAEA.879477.

Ntsangwane, L., 2017q. Basic measurements of radiation at station De Aar (2017-05). http://dx.doi.org/10.1594/PANGAEA.880334.

Ntsangwane, L., 2017r. Basic measurements of radiation at station De Aar (2017-06). http://dx.doi.org/10.1594/PANGAEA.879363.

Ntsangwane, L., 2017s. Basic measurements of radiation at station De Aar (2017-07). http://dx.doi.org/10.1594/PANGAEA.879479.

Ntsangwane, L., 2017t. Basic measurements of radiation at station De Aar (2017-08). http://dx.doi.org/10.1594/PANGAEA.880335.

Ntsangwane, L., 2017u. Basic measurements of radiation at station De Aar (2017-09). http://dx.doi.org/10.1594/PANGAEA.881309.
Ntsangwane, L., 2017v. Basic measurements of radiation at station De Aar (2017-10). http://dx.doi.org/10.1594/PANGAEA.882383.
Ntsangwane, L., 2017w. Basic measurements of radiation at station De Aar (2017-11). http://dx.doi.org/10.1594/PANGAEA.883723.
Ntsangwane, L., 2018. Basic measurements of radiation at station De Aar (2017-12). http://dx.doi.org/10.1594/PANGAEA.884839.
Olano, X., 2021. Basic measurements of radiation at station Cener (2009-07 et seq). http://dx.doi.org/10.1594/PANGAEA.931893.
Olefs, M., 2022. Basic measurements of radiation at station Sonnblick (2013-01 et seq). http://dx.doi.org/10.1594/PANGAEA.946389.
Pereira, E.B., 2018a. Basic measurements of radiation at station Petrolina (2016-01). http://dx.doi.org/10.1594/PANGAEA.889520.
Pereira, E.B., 2018aa. Basic measurements of radiation at station São Martinho da Serra (2016-02). http://dx.doi.org/10.1594/PANGAEA.889302.
Pereira, E.B., 2018ab. Basic measurements of radiation at station São Martinho da Serra (2017-02). http://dx.doi.org/10.1594/PANGAEA.889428.
Pereira, E.B., 2018ac. Basic measurements of radiation at station São Martinho da Serra (2016-03). http://dx.doi.org/10.1594/PANGAEA.889361.
Pereira, E.B., 2018ad. Basic measurements of radiation at station São Martinho da Serra (2017-03). http://dx.doi.org/10.1594/PANGAEA.889519.
Pereira, E.B., 2018ae. Basic measurements of radiation at station São Martinho da Serra (2016-04). http://dx.doi.org/10.1594/PANGAEA.889303.
Pereira, E.B., 2018af. Basic measurements of radiation at station São Martinho da Serra (2017-04). http://dx.doi.org/10.1594/PANGAEA.889429.
Pereira, E.B., 2018ag. Basic measurements of radiation at station São Martinho da Serra (2016-05). http://dx.doi.org/10.1594/PANGAEA.889304.
Pereira, E.B., 2018ah. Basic measurements of radiation at station São Martinho da Serra (2017-05). http://dx.doi.org/10.1594/PANGAEA.889430.
Pereira, E.B., 2018ai. Basic measurements of radiation at station São Martinho da Serra (2016-06). http://dx.doi.org/10.1594/PANGAEA.889306.
Pereira, E.B., 2018aj. Basic measurements of radiation at station São Martinho da Serra (2017-06). http://dx.doi.org/10.1594/PANGAEA.889431.
Pereira, E.B., 2018ak. Basic measurements of radiation at station São Martinho da Serra (2016-07). http://dx.doi.org/10.1594/PANGAEA.889307.
Pereira, E.B., 2018al. Basic measurements of radiation at station São Martinho da Serra (2016-08). http://dx.doi.org/10.1594/PANGAEA.889308.
Pereira, E.B., 2018am. Basic measurements of radiation at station São Martinho da Serra (2016-09). http://dx.doi.org/10.1594/PANGAEA.889311.
Pereira, E.B., 2018an. Basic measurements of radiation at station São Martinho da Serra (2016-10). http://dx.doi.org/10.1594/PANGAEA.889309.
Pereira, E.B., 2018ao. Basic measurements of radiation at station São Martinho da Serra (2016-11). http://dx.doi.org/10.1594/PANGAEA.889426.
Pereira, E.B., 2018ap. Basic measurements of radiation at station São Martinho da Serra (2016-12). http://dx.doi.org/10.1594/PANGAEA.889310.
Pereira, E.B., 2018b. Basic measurements of radiation at station Petrolina (2017-01). http://dx.doi.org/10.1594/PANGAEA.890317.
Pereira, E.B., 2018c. Basic measurements of radiation at station Petrolina (2016-02). http://dx.doi.org/10.1594/PANGAEA.889521.
Pereira, E.B., 2018d. Basic measurements of radiation at station Petrolina (2017-02). http://dx.doi.org/10.1594/PANGAEA.890318.
Pereira, E.B., 2018e. Basic measurements of radiation at station Petrolina (2016-03). http://dx.doi.org/10.1594/PANGAEA.889522.
Pereira, E.B., 2018f. Basic measurements of radiation at station Petrolina (2017-03). http://dx.doi.org/10.1594/PANGAEA.890319.
Pereira, E.B., 2018g. Basic measurements of radiation at station Petrolina (2016-04). http://dx.doi.org/10.1594/PANGAEA.889524.
Pereira, E.B., 2018h. Basic measurements of radiation at station Petrolina (2017-04). http://dx.doi.org/10.1594/PANGAEA.890320.
Pereira, E.B., 2018i. Basic measurements of radiation at station Petrolina (2016-05). http://dx.doi.org/10.1594/PANGAEA.889525.
Pereira, E.B., 2018j. Basic measurements of radiation at station Petrolina (2017-05). http://dx.doi.org/10.1594/PANGAEA.890321.
Pereira, E.B., 2018k. Basic measurements of radiation at station Petrolina (2016-06). http://dx.doi.org/10.1594/PANGAEA.889526.
Pereira, E.B., 2018l. Basic measurements of radiation at station Petrolina (2017-06). http://dx.doi.org/10.1594/PANGAEA.890322.
Pereira, E.B., 2018m. Basic measurements of radiation at station Petrolina (2016-07). http://dx.doi.org/10.1594/PANGAEA.889527.
Pereira, E.B., 2018n. Basic measurements of radiation at station Petrolina (2017-07). http://dx.doi.org/10.1594/PANGAEA.890323.
Pereira, E.B., 2018o. Basic measurements of radiation at station Petrolina (2016-08). http://dx.doi.org/10.1594/PANGAEA.889528.
Pereira, E.B., 2018p. Basic measurements of radiation at station Petrolina (2017-08). http://dx.doi.org/10.1594/PANGAEA.890324.
Pereira, E.B., 2018q. Basic measurements of radiation at station Petrolina (2016-09). http://dx.doi.org/10.1594/PANGAEA.889529.
Pereira, E.B., 2018r. Basic measurements of radiation at station Petrolina (2017-09). http://dx.doi.org/10.1594/PANGAEA.890325.
Pereira, E.B., 2018s. Basic measurements of radiation at station Petrolina (2016-10). http://dx.doi.org/10.1594/PANGAEA.889530.
Pereira, E.B., 2018t. Basic measurements of radiation at station Petrolina (2017-10). http://dx.doi.org/10.1594/PANGAEA.890326.
Pereira, E.B., 2018u. Basic measurements of radiation at station Petrolina (2016-11). http://dx.doi.org/10.1594/PANGAEA.889531.
Pereira, E.B., 2018v. Basic measurements of radiation at station Petrolina (2017-11). http://dx.doi.org/10.1594/PANGAEA.890327.
Pereira, E.B., 2018w. Basic measurements of radiation at station Petrolina (2016-12). http://dx.doi.org/10.1594/PANGAEA.889532.
Pereira, E.B., 2018x. Basic measurements of radiation at station Petrolina (2017-12). http://dx.doi.org/10.1594/PANGAEA.890328.
Pereira, E.B., 2018y. Basic measurements of radiation at station São Martinho da Serra (2016-01). http://dx.doi.org/10.1594/PANGAEA.889301.
Pereira, E.B., 2018z. Basic measurements of radiation at station São Martinho da Serra (2017-01). http://dx.doi.org/10.1594/PANGAEA.889427.
Qu, Z., Oumbe, A., Blanc, P., Espinar, B., Gesell, G., Gschwind, B., Klüser, L., Lefèvre, M., Saboret, L., Schroedter-Homscheidt, M., Wald, L., 2017. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The heliosat-4 method. Meteorol. Z. 26 (1), 33–57. http://dx.doi.org/10.1127/metz/2016/0781.
Rigollier, C., Lefèvre, M., Wald, L., 2004. The method heliosat-2 for deriving shortwave solar radiation from satellite images. Sol. Energy 77 (2), 159–169. http://dx.doi.org/10.1016/j.solener.2004.04.017, URL https://www.sciencedirect.com/science/article/pii/S0038092X04001082.
Sahan, M., Yakut, E., 2016. Estimation of monthly global solar radiation in the eastern mediterranean region in Turkey by using artificial neural networks. EPJ Web Conf. 128, 06001. http://dx.doi.org/10.1051/epjconf/201612806001.
Satpy, 2022. Satpy's documentation. https://satpy.readthedocs.io/en/stable/index.html. (Accessed 15 September 2022).
Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., Ratier, A., 2002. An introduction to meteosat second generation (msg). Bull. Am. Meteorol. Soc. 83, 977–992. http://dx.doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2, URL https://journals.ametsoc.org/view/journals/bams/83/7/1520-0477_2002_083_0977_aitmsg_2_3_co_2.xml.
Schmithüsen, H., Sieger, R., König-Langlo, G., 2012. BSRN toolbox V2.0 – a tool to create quality checked output files from BSRN datasets and station-to-archive files. http://dx.doi.org/10.1594/PANGAEA.774827.
Schroedter-Homscheidt, M., Azam, F., Betcke, J., Hanrieder, N., Lefèvre, M., Saboret, L., Saint-Drenan, Y.-M., 2022. Surface solar irradiation retrieval from msg/seviri based on apollo next generation and heliosat-4 methods. Meteorol. Z. http://dx.doi.org/10.1127/metz/2022/1132.
Shaikh, M.R., Waghmare, S., Labade, S., Tekale, A., 2017. A review paper on electricity generation from solar energy. Int. J. Res. Appl. Sci. Eng. Technol. 887, http://dx.doi.org/10.22214/ijraset.2017.9272.
SoDa, 2022. Solar radiation data website. https://www.soda-pro.com/home. (Accessed 15 September 2022).
Tang, W., Qin, J., Yang, K., Liu, S., Lu, N., Niu, X., 2016. Retrieving high-resolution surface solar radiation with cloud parameters derived by combining modis and mtsat data. Atmos. Chem. Phys. 16 (4), 2543–2557. http://dx.doi.org/10.5194/acp-16-2543-2016, URL https://acp.copernicus.org/articles/16/2543/2016/.
Tang, W., Yang, K., Sun, Z., Qin, J., Niu, X., 2017. Global performance of a fast parameterization scheme for estimating surface solar radiation from modis data. IEEE Trans. Geosci. Remote Sens. 55 (6), 3558–3571. http://dx.doi.org/10.1109/TGRS.2017.2676164.
Trueblood, C., Coley, S., Key, T., Rogers, L., Ellis, A., Hansen, C., Philpot, E., 2013. Pv measures up for fleet duty : Data from a tennessee plant are used to illustrate metrics that characterize plant performance. IEEE Power Energy Mag. 11 (2), 33–44. http://dx.doi.org/10.1109/MPE.2012.2234405.
Vuilleumier, L., Heimo, A., 2022. Basic and other measurements of radiation at station Payerne (1992-10 et seq). http://dx.doi.org/10.1594/PANGAEA.946384.
Wacker, S., Behrens, K., 2022. Basic measurements of radiation at station Lindenberg (1994-10 et seq). http://dx.doi.org/10.1594/PANGAEA.946382.
Woldegiyorgis, T.A., Admasu, A., Benti, N.E., Asfaw, A.A., 2022. A comparative evaluation of artificial neural network and sunshine based models in prediction of daily global solar radiation of lalibela, ethiopia. Cogent Eng. 9 (1), 1996871. http://dx.doi.org/10.1080/23311916.2021.1996871.
WRMC-BSRN, 2022. Wrmc-bsrn. word radiation monitoring center-baseline surface radiation network. https://bsrn.awi.de/. (Accessed 15 September 2022).
Yadav, A.K., Chandel, S., 2014. Solar radiation prediction using artificial neural network techniques: A review. Renew. Sustain. Energy Rev. 33, 772–781. http://dx.doi.org/10.1016/j.rser.2013.08.055, URL https://www.sciencedirect.com/science/article/pii/S1364032113005959.
Yang, H., Wang, L., Huang, C., Luo, X., 2021. 3d-cnn-based sky image feature extraction for short-term global horizontal irradiance forecasting. Water 13 (13), http://dx.doi.org/10.3390/w13131773, URL https://www.mdpi.com/2073-4441/13/13/1773.
Zhang, J., Zhao, L., Deng, S., Xu, W., Zhang, Y., 2017. A critical review of the models used to estimate solar radiation. Renew. Sustain. Energy Rev. 70, 314–329. http://dx.doi.org/10.1016/j.rser.2016.11.124, URL https://www.sciencedirect.com/science/article/pii/S1364032116308115.