Remote Sensing of Environment 315 (2024) 114431

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Machine learning forecast of surface solar irradiance from meteo satellite data

Alessandro Sebastianelli [a,*], Federico Serva [b,a], Andrea Ceschini [d], Quentin Paletta [a,c], Massimo Panella [d], Bertrand Le Saux [a]

[a] Φ-lab European Space Agency (ESA), Via Galileo Galilei 1, Frascati, 00044, Italy
[b] National Research Council (CNR), Via del Fosso del Cavaliere 100, Rome, 00133, Italy
[c] Climate Office European Space Agency (ESA), Fermi Avenue, Harwell Campus, Didcot, OX11 0FD, United Kingdom
[d] Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, Via Eudossiana 18, Rome, 00184, Italy

ARTICLE INFO

Edited by Marie Weiss

Keywords:
Surface irradiance
Forecasting
Deep learning
Meteosat
Solar energy

ABSTRACT

In order to facilitate the shift towards sustainable practices and to support the transition to renewable energy, there is a requirement for faster and more accurate predictions of solar irradiance. Surface solar energy predictions are essential for the establishment of solar farms and the enhancement of energy grid management. This paper presents a novel approach to forecast surface solar irradiance up to 24 h in advance, utilizing various machine and deep learning architectures. Our proposed Machine Learning (ML) models include both point-based (1D) and grid-based (3D) solutions, offering a comprehensive exploration of different methodologies. Our forecasts leverage two days of input data to predict the next day of solar exposure at country scale. To assess the models' performance, extensive testing is conducted across three distinct geographical areas of interest: Austria (where models were trained and validated), Switzerland and Italy (where we tested our models under a transfer learning regime), and sensitivity to the season is also discussed. The study incorporates comparisons with established benchmarks, including state-of-the-art numerical weather predictions, as well as fundamental predictors such as climatology and persistence. Our findings reveal that the ML-based methods clearly outperform traditional forecasting techniques, demonstrating high accuracy and reliability in predicting surface solar irradiance. This research not only contributes to the advancement of solar energy forecasting but also highlights the effectiveness of machine learning and deep learning models in being competitive to conventional methods for short-term solar irradiance predictions.

1. Introduction

One of the major challenges of the ongoing transition towards low-carbon energy supply is the integration of weather dependent renewable energy sources into power systems (MacKay, 2009; Dubus et al., 2018). In addition to improving energy storage capabilities and demand flexibility, the integration of intermittent energy sources into the energy mix would largely benefit from accurately anticipating the future renewable power supply at diverse spatio-temporal scales (Martinot, 2016; Heptonstall and Gross, 2021) to adapt the response of energy systems to their volatile nature (Samu et al., 2021).

Solar energy, for instance, is particularly affected by meteorological phenomenons such as clouds and aerosols causing local and global short-term fluctuations in energy generation (Kumar et al., 2020). Therefore, the development of new techniques able to model and predict the intermittent nature of the solar resource at ground level is of the utmost importance to potentially contribute to global energy security, social equity, and environmental sustainability (Johnson et al., 2020).

To this end, standard physics-based approaches have been traditionally employed to perform solar irradiance forecasts (Perez et al., 2010, 2013). These algorithms have been largely evaluated in diverse scenarios and conditions, consistently outperforming naive persistence models. These methods are based on Numerical Weather Prediction (NWP) techniques, that is physical models for modeling atmospheric conditions, including cloud formation and dissipation, and cover forecast horizons ranging from a few hours to 15 days ahead (Perez et al., 2013; Aguiar et al., 2016; Haupt et al., 2018). However, such physics-based forecasting technologies still face some difficulties currently limiting their weather forecast skills. In practice, NWP models struggle to assimilate cloud properties, model the complex dynamics of the

* Corresponding author.
E-mail address: alessandro.sebastianelli@esa.int (A. Sebastianelli).

https://doi.org/10.1016/j.rse.2024.114431
Received 19 January 2024; Received in revised form 2 September 2024; Accepted 13 September 2024
Available online 25 September 2024

system and are bound to the selected time scale, the geographic zone of the study and the model's input parameters.

Artificial Intelligence (AI) is now emerging as a promising alternative or addition to standard physics-based forecasting models. In particular, approaches based on Machine Learning (ML) or Deep Learning (DL) allow to build on the extensive amount of *in situ* and remote sensing atmospheric observations to improve weather forecasting. More specifically, these algorithms have the potential to advance solar energy meteorology (Paletta et al., 2023b) by leveraging historical solar irradiance measurements and cloud cover observations to learn complex relationships and patterns that govern solar energy generation (Succetti et al., 2020, 2021). As explained in Gailhofer et al. (2021), ML technologies can play a major role in the transition to a low-carbon economy. In the context of the energy transition, for instance, incorporating ML models into the forecasting process could provide more accurate and reliable solar irradiance predictions, thus improving the integration of increasing amount of solar energy in the electrical grid.

The adoption of ML for solar irradiance forecasting has gained traction in the renewable energy industry due to its capability to handle the non-linearities and uncertainties associated with solar radiation, while facilitating the fusion of various data sources such as *in situ* atmospheric sensors, sky cameras, and weather satellites (Paletta et al., 2023a).

ML techniques, such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), Categorical Boosting (CatBoost) and Extreme Gradient Boosting (XGBoost) have already shown promising results in various forecasting horizons, ranging from short-term (e.g., 1 to 30-min ahead) to long-term predictions (e.g., day ahead) (Voyant et al., 2017; Yagli et al., 2019; Zhou et al., 2021; Zhao et al., 2022). For instance (Ayet and Tandeo, 2018) successfully employed KNN to deliver a probabilistic forecast of the Global Horizontal Irradiance (GHI) up to 6 h in advance over a localized solar energy source using a 5-year data.

Ramadhan et al. (2021) and Cornejo-Bueno et al. (2019) compared physics-based models with some ML algorithms on hourly-sampled solar radiation data, showing that such models generally perform better when input parameters are appropriately selected. Solano et al. (2022) compared the performances of several ML algorithms for solar radiation forecasting using endogenous and exogenous input focusing on 1, 2 and 3 h lead times. Over a 6-year time window, they found out that an ensemble feature selection approach outperforms plain SVM, CatBoost and XGBoost algorithms for all prediction time horizons. Nespoli et al. (2022) presented three ML models based on NNs and RF to detect the clouds which will potentially obstruct the direct sunrays and predict solar irradiance by means of satellite images and weather data. With respect to the state of the art, their approach offers a lighter yet effective way to forecast short-term peaks and drops in solar resource availability at a particular location. However, the proposed algorithms has not been extended beyond the 15 min ahead forecast horizon, which limits its full applicability to solar energy management. Moreover, the MLP and RF models used in the study do not consider the temporal dependencies among data.

Recently, DL has emerged as the golden standard in solar radiation forecasting. The competitiveness of DL over numerical and ML models has been already demonstrated by several studies (Succetti et al., 2021, 2020; Espeholt et al., 2022; Rosato et al., 2021). Many Deep Neural Network (DNN) architectures have been proposed over time, including Recurrent Neural Networks (RNNs), Long Short-Term Memories (LSTMs) and Gated Recurrent Units (GRUs) for learning temporal patterns, Convolutional Neural Networks (CNNs) to detect spatial features, and a combination of both for a spatio-temporal analysis of solar radiation variations.

For example, Aguiar et al. (2016) focused on hourly GHI forecasting from 1 to 6 h ahead using a NN trained on both ground-based and exogenous data, i.e. satellite GHI data, solar radiation and Total Cloud Cover forecasts provided by the European Centre for Medium Range Weather Forecasts (ECMWF). They found out that the combination of exogenous satellite and ECMWF data with ground data may improve the accuracy of intra-day forecasts from 3 h onwards, whereas from 1 to 3 h ahead satellite data are sufficient to achieve satisfactory performance. Narvaez et al. (2021) proposed a site-adaptation scheme to fuse multivariate on-ground and satellite-based data, on which an Encoder–Decoder LSTM sequence-to-sequence model is used to forecast solar radiation. Results showed that the LSTM model outperformed other ML models, as well as traditional approaches. However, a potential limitation of this approach arises from the occasional unavailability of in-situ data and the occurrence of missing data. In Jiang et al. (2019), a DNN composed of a CNN and a Multi-Layer Perceptron (MLP) was applied to extract spatial features from satellite imagery and estimate hourly global solar radiation. While the task of estimating hourly solar radiation is not inherently complex, the DNN architecture was effective in characterizing varying cloud morphology. However, the model's potential could be further enhanced by incorporating temporal information through RNN models, which are able to capture changes in solar radiation patterns over time. A similar remark could be raised for Pérez et al. (2021), who proposed an intra-day forecasting DNN model based on CNN layers capable of providing up to 6 h ahead forecasts without requiring real-time data measurements during training. The proposed model used a series of time-dependant irradiance estimates near the target location as the main input; these estimates were derived from satellite images and were combined with other secondary inputs in the DNN model. Although the DNN model outperformed NWP alternatives, its use of CNN layers alone constrained its ability to effectively learn complex temporal patterns in the data. Consequently, extending the forecasting horizon beyond 6 h may not be feasible without introducing RNNs into the model's architecture. Finally, Michael et al. (2022) proposed a hybrid short-term solar irradiance prediction model based on a modified multi-step CNN-stacked LSTM architecture. The CNN was charged of extracting relevant features from incoming irradiance data, whereas a stacked LSTM network was used to predict the final output. Such a model proved to be superior compared to other works proposed in literature for 1 h ahead forecast, but no evaluation was performed for longer forecasting horizons.

In this paper, our main contribution is a comprehensive study of the most promising ML forecasting techniques for surface solar irradiance from Earth Observation (EO) data. As shown in Fig. 1, we propose and assess various models to predict solar irradiance data coming from satellite measurements. We compare their performances with physics-based forecasts and show that ML is competitive compared to standard NWP products.

The rationale behind the use of satellite data only in this work is motivated by the fact that unlike ground-based measurements, which are accurate but only available for few locations, EO data allow to map surface solar irradiance at the global scale with relatively high spatial and temporal resolution (Huang et al., 2019). In particular, geostationary satellites are able to acquire atmospheric observations all around the globe with a sub-hourly temporal resolution. The resulting data have advanced our understanding of the Earth's solar energy distribution, cloud cover dynamics, and weather patterns. By providing insights into the diurnal and seasonal variations of solar irradiance across different regions, this information is crucial for various applications such as solar energy planning. Furthermore, satellite data allows for accessible and continuous monitoring of solar energy production in isolated areas where ground-based instruments may be challenging to deploy and maintain (Yagli et al., 2020).

Fig. 2 illustrates the applicability of various solar forecasting methods based on diverse types of input data and for various spatio-temporal scales (Kumar et al., 2020; Nielsen et al., 2021). In this study, we aim to predict the surface solar irradiance up to 24 h ahead at sub-hourly frequency, a time window for which satellite-based data-driven models and NWP models exhibit comparable forecasting skill, whereas

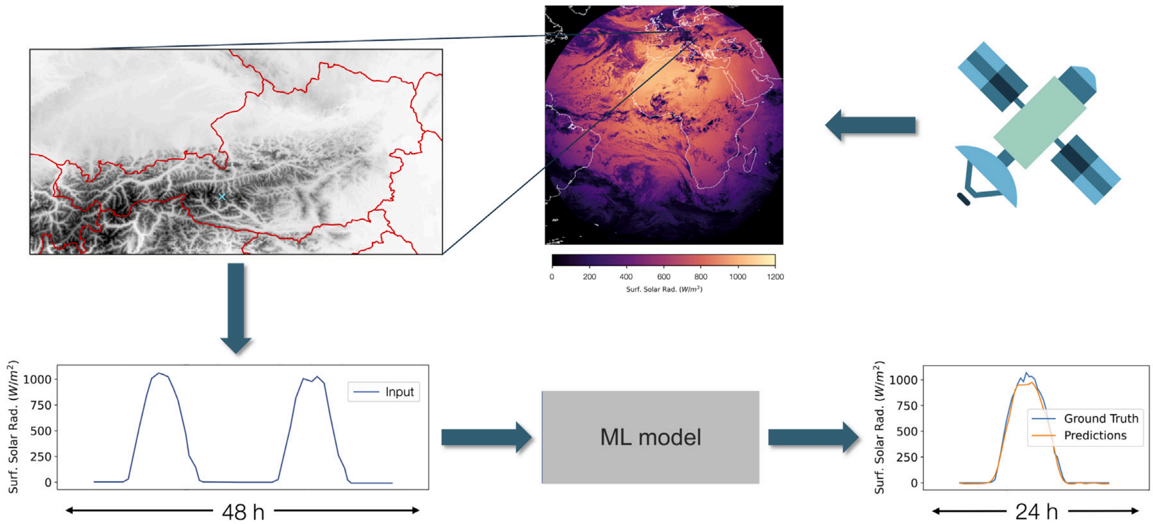

Fig. 1. Flowchart of our forecasting procedure: surface solar irradiances obtained from Meteosat data are extracted over the region of interest (elevation with gray shadings, top plots). For grid-point predictions (bottom plots), a 2-day time series (bottom left, 48 h) is used as input for ML models, providing a day ahead (bottom right, 24 h) prediction.

statistical models processing *in situ* measurements are most effective for shorter-term lead times (Perez et al., 2010, 2013; Aguiar et al., 2016; Haupt et al., 2018; Pérez et al., 2021; Nielsen et al., 2021), as illustrated in Fig. 2. A day-ahead prediction is important for enabling decision-making in energy management, especially for energy storage purposes (Qing and Niu, 2018).

We propose a new benchmark to evaluate and foster the development of ML approaches for the task of solar irradiance forecasting using satellite-derived data. To the best of our knowledge, no works appeared in the state of the art dealing with such a long lead time and without ancillary data. Moreover, we show how some of the ML models are able to compete and even to reach better performances with respect to physics-based models.

Our final objective is to assist in the effective management of the electricity grid. Additionally, as last contribution, we provide openly our pipeline encompassing diverse areas of interest, which can be used to evaluate future models. The availability of consistent global solar irradiance prediction may facilitate the development and assessment of solar energy projects on a large scale, contributing to the transition to cleaner and sustainable energy sources.

As a case study we consider the Austrian region, that is representative for the European energy system (Simoes et al., 2017), for training and validation of ML models, and we also present forecast results on Switzerland (which has similar geographical characteristics) and Italy (located further south and with different geography) to investigate potential geographic issues limiting transferability.

The conversion of the energy mix to increasingly include renewable energy sources is one of the staples of the Green Transition (Söderholm, 2020) and Austria was used as a prototype for the ESA Green Transition Information Factory (GTIF) initiative.[1] This tool allows users to interactively discover the underlying opportunities and complexities of transitioning to carbon neutrality by 2050 using the power of Earth Observation, cloud-computing and cutting edge analytics. The cloud-based integrated GTIF environment enables: (1) Decision-makers to assess and monitor the effectiveness of policies, and evaluate political objectives and outcomes using GTIF-provided data, indicators and interactive exploration tools. (2) Industry to develop novel solutions to foster the Green Economy, supported by space technologies, and connect to relevant national and international stakeholders within the wider GTIF ecosystem. (3) Citizens to engage and understand the needs for their actions through interactive exploration tools and captivating scientific narratives across key Green Transition domains.

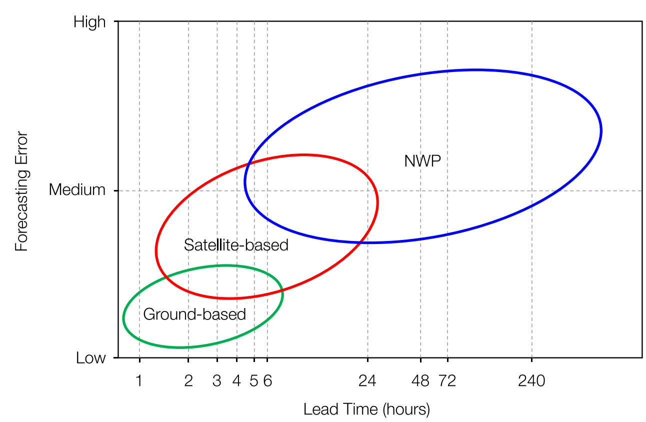

Fig. 2. Schematic illustration of the forecasting error achieved with models relying on ground, satellite and NWP-based data, as a function of time. Relative skill based on the works by Voyant et al. (2017) and Kumar et al. (2020), among others, but represented qualitatively since different metrics are used in the literature.

2. Material and methods

This section presents the data that have been used for training, validating and testing the proposed solutions, as well as all the procedure we applied to preprocess the raw data and to create the final dataset.

2.1. Observational and weather model data

The main dataset we use in this work is a satellite-based climate data record (CDR) provided by the Climate Monitoring (CM) Satellite Application Facility (SAF) of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). The Surface Solar Radiation dataset — Heliosat (SARAH-2, v2.1) provides surface incoming solar radiation (SIS, units W m^{-2}) derived from the geostationary Meteosat satellites for the period 1983–2017 (Müller et al., 2015; Pfeifroth et al., 2018, 2019). SIS is defined as the radiation flux (irradiance) reaching a horizontal plane at the Earth surface in the 0.2–4 μm wavelength region. The dataset covers the region between ±65° latitude and Spanning ±65° longitude (including Europe, Africa, the Atlantic Ocean), and data are provided in the form of a regular 0.05° × 0.05° latitude–longitude grid at a thirty-minute resolution. The coverage is reported in Fig. 3, showing SIS maps collected at different times of the day in June

[1] https://gtif.esa.int/ (last access, December 2023).

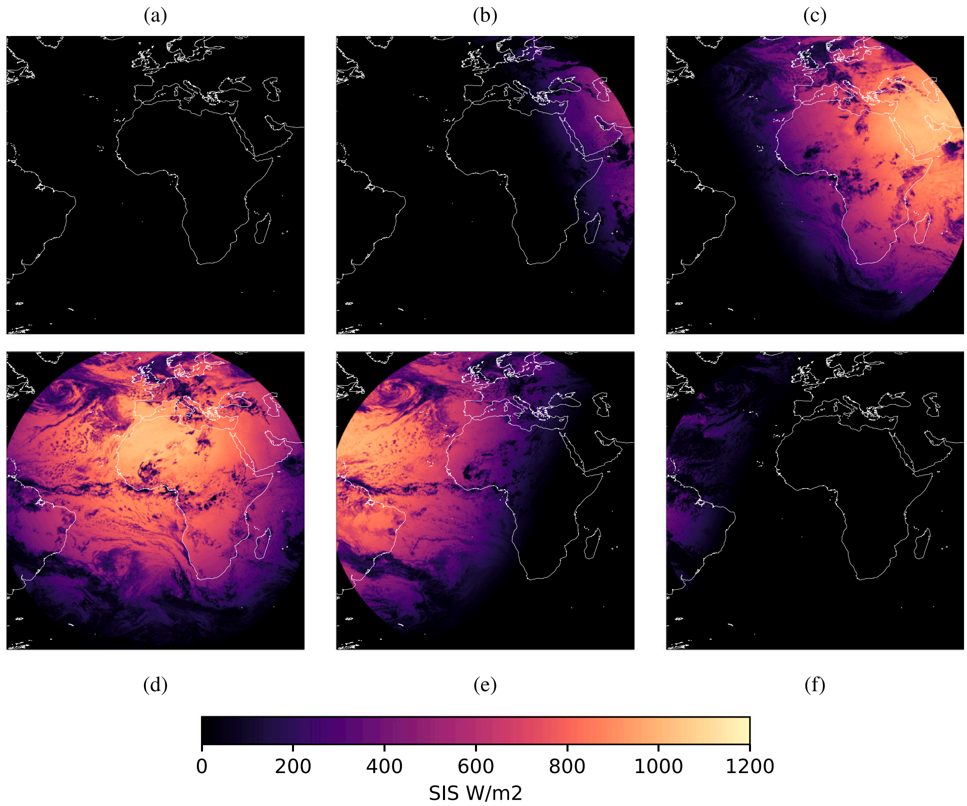

Fig. 3. The spatial coverage covers the Meteosat disk: (a) 2017-06-06 00:00 UTC, (b) 2017-06-06 04:00 UTC, (c) 2017-06-06 08:00 UTC, (d) 2017-06-06 12:00 UTC, (e) 2017-06-06 16:00 UTC, (f) 2017-06-06 20:00 UTC.

2017. To obtain a continuous CDR, passive observations of the visible channels of the first (Meteosat visible-infrared imaged, MVIRI) and second (Spinning enhanced visible-infrared imager, SEVIRI) generations of EUMETSAT satellites have been processed with a two-step approach. The modified Heliosat algorithm (MAGICSOL) is used to derive an effective cloud albedo (CAL, Cano et al., 1986), taking into account the viewing geometry and sensor calibration characteristics, including the matching between the MVIRI and SEVIRI channels, which are not identical (Müller et al., 2015).

The presence of clouds affects the solar radiation reaching the surface, and the derived CAL variable is then used to derive SIS values and other radiative parameters. The method includes empirical adjustments, also based on ground measurements. Atmospheric effects are modeled with look-up-tables calculated by radiative transfer model simulations, accounting for the presence of aerosols, water vapor, and ozone and varying surface albedo by means of climatologies. The product is not free from uncertainties, as for example errors are larger over bright surfaces, such as desert regions or snowy areas, since clear-sky and cloudy-sky reflection are similar. It is also worth noting that orography is not taken into account for clear-sky calculations, while this is important for radiative transfer modeling (Erickson, A., PhD Thesis, personal communication). Another source of error stems from the presence of aerosols, since ground measurements are sparse and aerosol loading influences solar radiation, especially in cloud-free conditions. The use of climatological aerosol data in the retrieval model affects the representation of interannual and longer term changes in the product.

Long-term observations of surface radiation and other meteorological parameters are carried out in many stations across the globe, but the quality of the measurements is uneven. The Baseline Surface Radiation Network (BSRN) (Driemel et al., 2018) currently comprise about 50 stations providing high-quality measurements with a high temporal frequency, with a common formatting and subject to quality control procedures. While multiple radiative parameters are measured, here we focus on the global shortwave radiation (SWD), measured by pyranometers, to compare with satellite-derived values.

Numerical weather prediction (NWP) data are commonly used for forecasting the renewable energy production from hours to days ahead. To provide a fair comparison with state-of-the-art weather prediction models, we considered forecasts produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) with the Integrated Forecasting System (IFS) model, which has been operational since the 1980s' and continuously updated over time. The quality of the forecasts has steadily improved over time, owing to model development and improved data assimilation methods, and its accuracy over the Alpine region is known to be relatively high (Haiden and Trentmann, 2016). As discussed by Yang et al. (2022), ECMWF irradiance forecasts are at par or perform better than other state-of-the-art systems. We retrieved the surface solar radiation (variable *ssrd*) from high resolution (HRES) forecasts initialized at 00 UTC for the period 2012–2017. Over this period, the vertical model resolution has not changed and we downloaded data at the spatial resolution used in 2012 (over a regular 0.15° × 0.15° latitude–longitude grid), including hourly ECMWF forecasts (units J m^{-2}). A temporal linear interpolation was used to obtain a twice-hourly frequency, and nearest-neighbor upsampling by a factor 3 (from ~15 to ~5 km) was used to match the Meteosat data characteristics (Leirvik and Yuan, 2021). Surface irradiance (units W m^{-2}) is then computed starting from the interpolated quantities over 30 min intervals. For simplicity, no post-processing or bias-correction methods are applied.

2.2. Basic comparison of irradiance datasets

In this work we mainly focus on satellite irradiance data, but it is informative to compare them with quality-checked BSRN measurements to assess the accuracy of remote sensing observations with

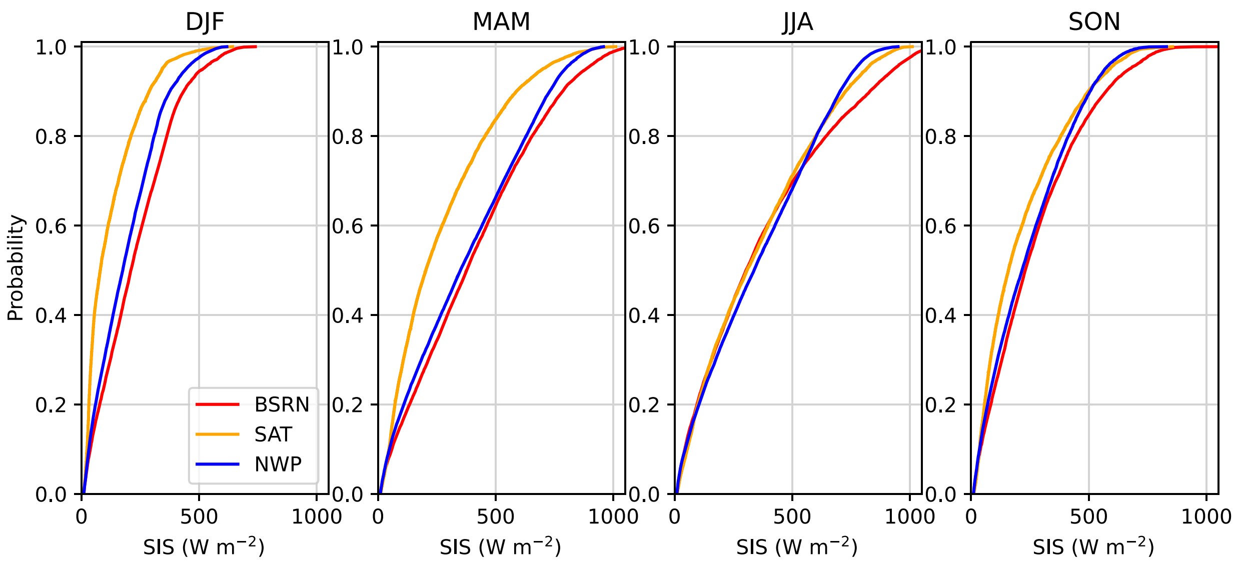

Fig. 4. Cumulative distribution functions (CDFs) for surface solar irradiance from station, satellite (SAT) and numerical weather prediction (NWP) datasets at the Sonnblick site. Based on data for the period 2012–2017, divided by meteorological season.

high-quality ground-level measurements. For this, we consider the long-term record from the Sonnblick station, located in the Swiss Alps at high elevation (3109 m). Additional information on the station instrumentation is available from Olefs (2013). Since seasonality is a major source of cloudiness variability, we analyze the data after stratification by meteorological season (December to February, DJF — winter; March to May, MAM — spring; June to August, JJA — summer; September to November, SON — autumn). To compare the various datasets, irradiance values resampled to 30-min periods – characteristic of the satellite product – and time series are extracted for the nearest grid-point to the Sonnblick geographical coordinates.

From Fig. 4 we note that inter-dataset differences are more marked in the winter and spring seasons, with the satellite-derived dataset underestimating irradiance in all conditions. We note that weather observations at Sonnblick are provided[2] to the Global Telecommunication System, hence one can expect them to match forecasts well, at least closer to forecast initialization (Haiden et al., 2018). Maximum irradiances for both satellite and NWP datasets are lower than station measurements, likely due to effects of spatial resolution and the characteristics of the Sonnblick site. Another effect is likely due to the occurrence of clouds, which can have very fine spatial and temporal scales, and the best agreement among the products is obtained in summer for values below ~500 W m^{-2}.

Summary statistics for the distribution quantiles given in Table 1 confirm the findings obtained from the CDFs. The value of point measurements from the BSRN dataset generally exceed those from other datasets, which is partly expected given the exposition of the mountain site. The order of magnitudes of the three products are generally consistent, with lower values in DJF and larger in JJA, and satellite-based irradiances differ the most from observation in the MAM season. This can be explained by the fact that in the Swiss Alps snow accumulation can be significant in late winter and spring, and the high surface albedo can negatively affect the performances of the retrieval algorithm. The year-round difference between satellite and NWP irradiances reaches ~ 150 W m^{-2} for the 95th quantile of the distribution (Q_{95}). We note that the high elevation and surface conditions of Sonnblick make it a challenging area for satellite and models alike, and as expected the agreement between satellite and NWP-based irradiance is better year-round in surrounding lowlands, as illustrated in Supplement A (Fig. A.1 and Tab. A.1).

[2] https://www.sonnblick.net/en/science/networks/wmo-gts/ (last accessed, December 2023).

Table 1
Seasonal statistics (quantiles Q from 5 to 95) for BSRN, satellite (SAT) and numerical weather prediction (NWP) datasets at the Sonnblick site. Based on data for the period 2012–2017, divided by meteorological season.

(a) DJF						(b) MAM					
Data	Q_5	Q_{25}	Q_{50}	Q_{75}	Q_{95}	Data	Q_5	Q_{25}	Q_{50}	Q_{75}	Q_{95}
BSRN	20	100	210	336	515	BSRN	28	174	376	600	883
SAT	17	36	80	183	343	SAT	29	88	203	403	699
NWP	20	76	175	286	444	NWP	26	145	350	580	793
(c) JJA						(d) SON					
Data	Q_5	Q_{25}	Q_{50}	Q_{75}	Q_{95}	Data	Q_5	Q_{25}	Q_{50}	Q_{75}	Q_{95}
BSRN	23	123	303	569	932	BSRN	22	106	230	401	672
SAT	29	127	308	542	816	SAT	25	69	159	330	588
NWP	22	136	337	560	767	NWP	21	90	216	374	567

2.3. Methods

In order to comprehensively evaluate the effectiveness of ML models in solar irradiance forecasting, we have selected various shallow and deep architectures which are usually employed for time series prediction tasks. They are hereafter briefly described and accompanied with references for the reader interested in deepening the theoretical and mathematical aspects of each model.

LSTM: it is a RNN variant renowned for its superior ability to capture sequential dependencies and long-range temporal patterns (Hochreiter and Schmidhuber, 1997). By using a system of memory cells, gates, and buffer mechanisms, LSTM excels in learning hidden patterns among time series data while mitigating vanishing gradient issues, making it a powerful candidate for modeling complex solar irradiance fluctuations.

GRU: it is another variant of the RNN architecture, with a simplified gating mechanisms compared to LSTM (Cho et al., 2014). Its streamlined design allows for efficient learning and prediction of temporal sequences, making it a valuable algorithm in solar irradiance prediction, where swift adaptation to changing conditions is of paramount importance.

bLSTM: The bidirectional LSTM (bLSTM) enriches the predictive power of traditional LSTMs by processing input sequences in both forward and backward directions simultaneously (Graves and Schmidhuber, 2005). This dual perspective enhances the model's grasp on temporal dynamics, allowing it to discern intricate relationships within solar irradiance data and yield more accurate forecasts.

ConvLSTM: The Convolutional LSTM (ConvLSTM) combines the capabilities of CNNs and LSTMs, enabling it to capture both spatial and temporal features in solar irradiance datasets (Obiora and Ali, 2021). This architecture has proved to be particularly beneficial for multi-site solar radiation forecasting due to its inherent ability to analyze complex interactions across multiple locations while correlating them with temporal variations.

RF: The RF algorithm employs an ensemble of decision trees, collectively harnessing diverse input features to make informed predictions (Breiman, 2001). Its capacity to handle complex interactions and non-linearities within solar irradiance data makes it a versatile contender for accurate forecasting, especially when fast training and inference times are required.

CB: The Categorical Boosting or CatBoost (CB) algorithm integrates decision trees and gradient boosting theory to create a more robust ensemble model (Prokhorenkova et al., 2018). Decision trees are employed as base learners to capture intricate data patterns, whereas the gradient boosting technique is used to refine these base models iteratively. This way, the CB algorithm is able to model complex interactions within data by continually adjusting and optimizing its predictions, striking a balance between model complexity and generalization. Additionally, CB is highly effective for analyzing extended solar irradiance time series due to its efficient CPU utilization.

To run comparisons with *naive* and physics-based approaches, we have chosen to consider NWP, Climatology (per pixel average irradiance level over 35 years) and Persistence (average of the previous two days) models. The comparison between NWP and ML models was performed over 47 timesteps starting from 30 min since forecast initialization, e.g., for a forecast initialized at 00 UTC, from 00.30 UTC to 23.30 UTC of the same day.

2.4. Experimental settings for ML-based models

2.4.1. Data preparation

Data reshape: the models of interest, that have been introduced in Section 2.3, are designed to take as input a SIS time series over two consecutive days in order to forecast the subsequent day of surface incoming solar radiation (SIS) values. In practice, considering a dataset $D \in \mathrm{R}^{T\times W\times H\times C}$, where $T = 35 * 364 * (24 * 2)$ represents the number of available satellite acquisitions (35 years, 364 days, 48 acquisitions per day, one every 30 min), W and H are respectively the width and height in pixels of the Meteosat disk, and C the number of channels ($C = 2$ in this study, i.e. SIS and day of year), results in creating a new dataset $D_2 = \{X, y\} \in \{\mathrm{R}^{T^{*}\times 96\times W\times H\times C}, \mathrm{R}^{T^{*}\times 48\times W\times H\times C}\}$, where T^{*} represent the number of sequences of 3 days (96 + 48 acquisitions) available in T, X and y are the inputs and the ground truths respectively.

Data normalization: before proceeding with the training of the ML-based models, we normalized our data using a *max-scaler* with the maximum values inferred using the whole dataset and rounded by excess.

Data Loader: in order to deal with the large amount of spatio-temporal data, we developed a data-loader that dynamically selects spatial (i.e., a certain location) and temporal (e.g., a certain date) indices at run-time. It is worth highlighting that models such as the RF, GRU, CB, and LSTMs are structured to work with data at pixel level, thus the number of samples in the dataset is much higher compared to image-based solutions such as the ConvLSTM. On the other hand, data are spatially correlated, hence to increase the diversity of the training set we randomly selected pixels for different locations across the domain. This further motivates the need for an efficient data loader, able to deal with 1D and 3D data structures alike. For pixel-based models, all pixels are used in validation/testing to reconstruct complete spatial maps.

Dataset split: we trained the models only on data from Austria from 1983 to 2011, while we validated the models using data from Austria (9°–7.5° E, 45.5°–49.2° N) from 2012–2017. The choice depends mostly on the fact that NWP models are constantly updated, and to be more fair in their comparison, we have chosen the most recent years for validation. Finally, we tested the models on two states not included in the training dataset, Italy (6.5°–19° E, 36.5°–47° N) and Switzerland (5.6°–0.8° E, 45.5°–49.2° N), again for the period between 2012 and 2017.

Table 2
Hyperparameter settings for models training.

	GRU	LSTM	bLSTM	ConvLSTM	RF	CB
Learning rate	0.002	0.002	0.002	0.0002	–	0.02
Layers/Depth	4	4	4	4	–	8
Hidden size (per layer)	10	10	10	48	–	–
Batch size	1024	1024	1024	96[a]	–	–
Loss function	MAE	MAE	MAE	MAE	MAE	RMSE[b]
Training epochs	30	30	30	50	20	20
Network size	~3k	~4k	~9.5k	~67k	300	100

[a] The ConvLSTM model working with relatively more complex data samples, the batch size was decreased in order to fit memory requirements.

[b] CatBoost currently only supports the RMSE loss for multi-regression tasks.

2.4.2. Models' hyperparameters

In order to achieve a fast and good convergence, the neural models were trained using the Adam optimization algorithm (Kingma and Ba, 2014). The training hyperparameters were fine-tuned through an extensive grid search procedure. A summary list of the models' hyperparameters is illustrated in Table 2.

All the experiments were performed using Python 3.9 on a machine equipped with an Intel 11th Gen Intel(R) Core(TM) i7-11800H 8-Core CPU at 2.30 GHz and with 32 GB of RAM, using for training and inference a NVIDIA GeForce RTX 3080 GPU with 16 GB of dedicated vRAM. The neural network models were implemented in PyTorch, RF is implemented with sci-kit learn.

3. Results

In this section, we show the overall robustness and skills of selected ML and DL models as viable alternatives for standard physics-based methods. A spatial analysis is conducted to highlight the difficulties in predicting solar irradiance with respect to the orography of the territory, known to be an important feature (Erickson, A., PhD Thesis, personal communication). Additionally, we illustrate the performance of the models across different seasons, allowing us to determine when each model exhibits its optimal capabilities. This seasonal analysis provides valuable insights into the dynamics of ML model behavior.

3.1. Overall forecast results

The results are organized and presented via sets composed of three consecutive months, forming triplets. This grouping approach allows for a comprehensive analysis of how the models performed in relation to different seasons throughout the studied period. Analyzing the results in this manner provides a more nuanced understanding of the models' capabilities and limitations across various seasons. It allows us to discern patterns, fluctuations, and potential correlations between the models' performance and the changing seasons. This information is invaluable for assessing the adaptability and reliability of the models in real-world scenarios that may exhibit distinct seasonal characteristics.

We compare the overall forecast results of all the models, i.e. the results of a 24 h-ahead prediction averaged over time sequences spanning from 2012 to 2017. The bLSTM model outperforms other models in terms of normalized mean absolute error (nMAE). Note that normalization is done separately by season, since annual variations are very large, therefore results are given as percentages for easier comparison. This indicates that the bLSTM is the most accurate model for forecasting

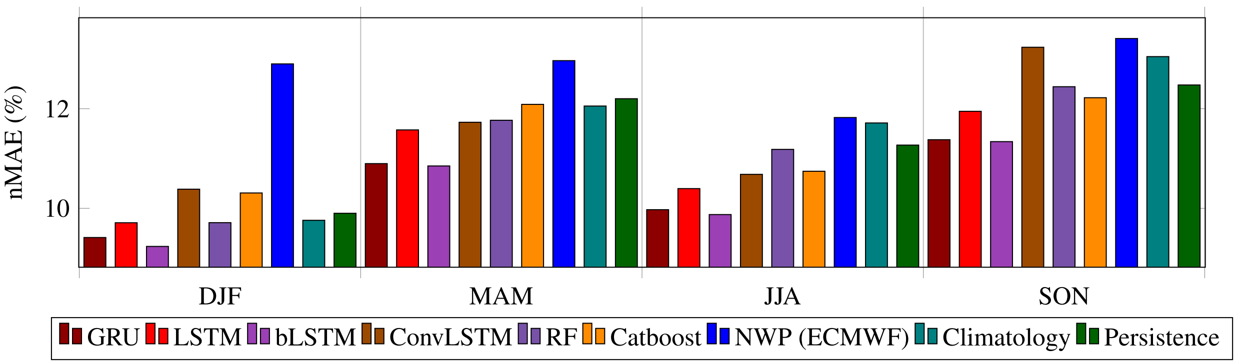

Fig. 5. Error (nMAE) values calculated for the validation period (2012–2017) for Austria divided by seasons. For all season, the bLSTM is outclassing the other models including the standard Numerical Weather Prediction (NWP).

solar irradiance values, at the cost of increased resources for training compared to other pixel-based models we tested (not shown).

Overall, the results and the study suggest that ML models, particularly deep networks, have the potential to significantly improve the accuracy and reliability of solar irradiance forecasting. These models can help to address the challenges posed by the highly intermittent nature of ground-level solar radiation, making them a valuable tool for the enhanced integration of renewable sources in the energy mix.

We compared our results against numerical weather prediction models, that are widely used for solar applications (Moreno-Garcia et al., 2019) including hours-ahead solar power forecasting. However, the reliability of their predictions can be impacted by complex local conditions (such as complex terrain) and timing accuracy of events such as cloud occurrence. Combining weather model data with machine learning techniques is currently attracting much interest in the field of renewable energy meteorology (Markovics and Mayer, 2022). We also compared our results against *naive* forecasting models, e.g., the average day obtained from the long-term record (climatology) – result obtained by averaging the full time-series (climatology) – and by averaging the input sequence (persistence).

Results are not only valid for the main region of interest, Austria, as shown in Fig. 5, but also for other countries like Switzerland, Fig. C.6, that exhibit similar meteorological conditions, and Italy, located further south and hence characterized by different climates Fig. C.15. Indeed, for the latter one, all the models had a more complex and diverse data space to work with, resulting in slightly poorer results for the transfer learning task (Nie et al., 2024). It should be noted, however, that the ConvLSTM is performing better over the wider Italian domain comparing to smaller regions, indicating that the domain size is key for spatial convolutional network as shown by Nielsen et al. (2021).

3.2. Spatial analysis by region

A widely used indicator for solar energy management is the energy integral (irradiation) available at a certain location (MacKay, 2009). This quantity is dependent on the time of the year, as in cloud-free conditions more energy is available in summer than in winter. Unlike time-resolved predictions, we cannot diagnose short-term fluctuations from time integrals, but they are nonetheless important for evaluating models for day-ahead predictions. The following section presents graphical results of the selected models across the four seasons focusing on Austria, while, for spacing and readability reasons, results for Switzerland and Italy are reported in the Supplement. These results are divided in two series of plots for each season, the first one depicts integrals calculated over the whole prediction period (24 h), the second series shows error maps, i.e, the difference between each model prediction and the ground truth.

Finally, the last set of plots display yearly averaged results (Figs. 12, 13 and 14). The spatial analysis provides a detailed understanding of the distribution of errors across different regions of Austria. This facilitate the identification of areas where the models perform well and areas where they need improvement. The analysis would also help determine the factors that contribute to errors in SIS forecasting. For example, solar irradiance levels over areas with complex terrain or persistent cloud coverage may be more challenging to predict.

By complementing the error heat-map with other spatial data such land use/land cover maps (accessible through GTIF), the analysis could provide insights into the environmental factors that affect SIS. This information could be used to improve the accuracy of the models by incorporating additional environmental data. The analysis could also help identify the temporal patterns in errors, such as seasonal variations or changes over time. This can be used to improve the models by incorporating temporal data or adjusting the model parameters to better capture fine-scale variations. The results for winter and fall times can be found in Supplement B, while the results for summer and spring times are presented below.

Spring time. Figs. 6 and 7 show results, as forecast irradiation over 24 h, for each model at a specific date during winter time for 2013. We can observe from Fig. 6 that results are close to the ground truth especially if looking at models like GRU and bLSTM, and even if maps of differences Fig. 7 are showing peaks of under- and over-estimation. Focusing on Fig. 6, we can see that these two models are the two better at catching spatial patterns, giving a more reliable prediction and overpassing other models especially NWP and persistence. Moreover we can observe that models like NWP and Persistence are struggling at catching spatial patterns, resulting into smooth and/or under-estimated results. Finally, Fig. 8 shows results at pixel level for a point located to the Sonnblick station. These results shows that, even if all the models on average are good at forecasting SIS, in general the huge variability is particularly difficult to catch. This aspect will be better detailed in the discussion section.

Summer time. Summer time is the best case scenario, as during this season the presence of clouds is limited in the area of interest and simple baselines, such as persistence, would work relatively well. We can see this from Fig. 9, were for most of the models the results are almost perfectly matching the ground truth, even if also in this case the bLSTM is better at catching even the smallest patterns. Moreover Fig. 9 shows also that other models like RF, CB and Climatology are heavily under-estimating the SIS variable, while NWP is producing a flat result. This results are also in this case confirmed and better highlighted by Fig. 10. Pixel level results in Fig. 11 are particularly interesting in this case, since it is possible to see that all the models are working well, mostly because of the season and fair weather on that specific date. Moreover by comparing results at pixel level to results at image level it is possible to note a strong discrepancy between some models and the ground truth, especially in the north-eastern area. For this reason looking at individual location may be not entirely informative, and spatial analysis is more useful to assess and compare models' results.

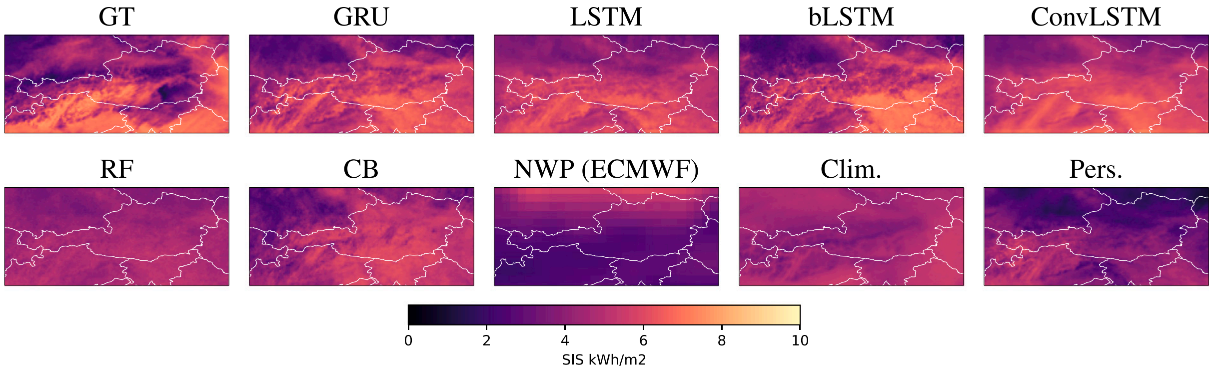

Fig. 6. Daily surface irradiation over Austria at spring time for 2013-05-03 (validation dataset). The satellite ground truth is indicated as GT, for the other acronyms refer to text.

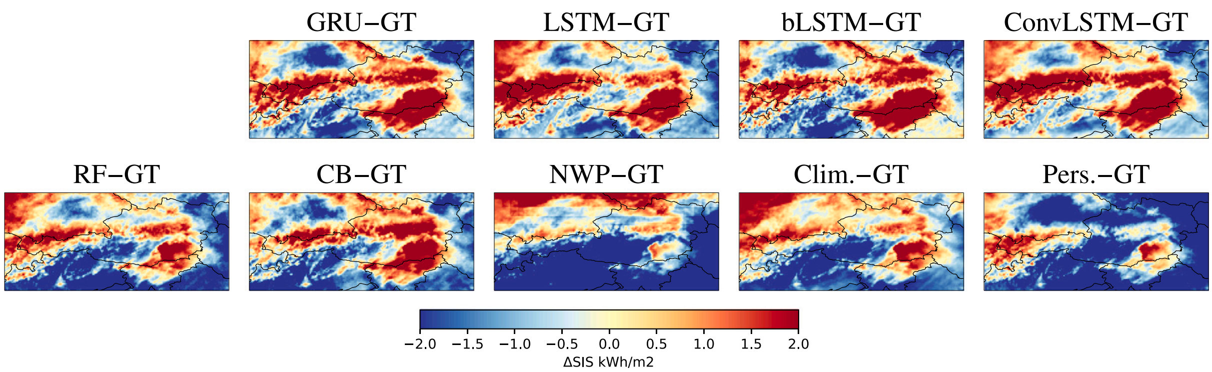

Fig. 7. Difference maps obtained by comparing the ground truth (GT) and the models' prediction at spring time for 2013-05-03 (validation dataset).

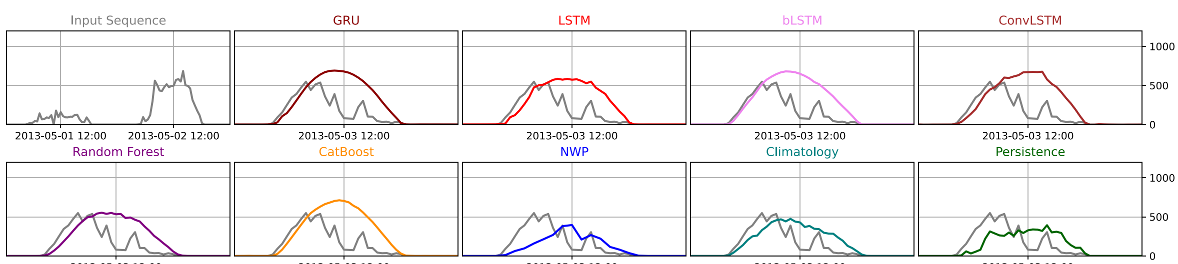

Fig. 8. Pixel level results, on Sonnblick, for 2013-05-03 (validation dataset). Observed data is shown with the gray line.

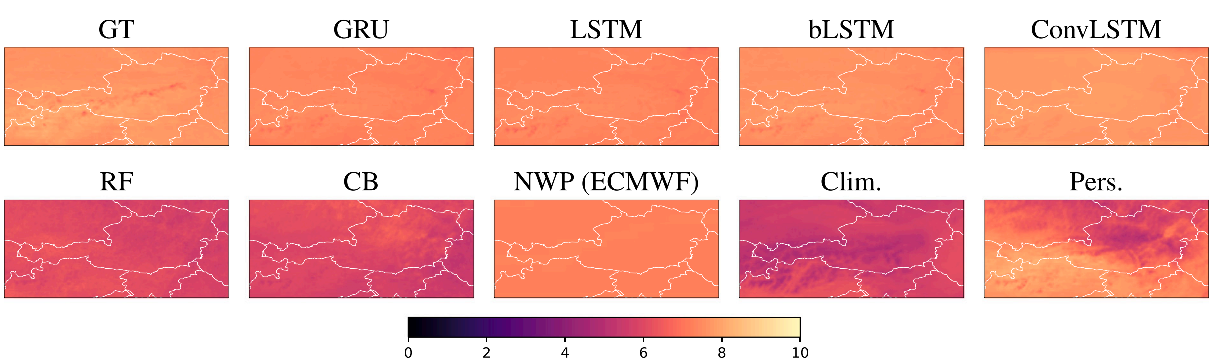

Fig. 9. Daily surface irradiation over Austria at summer time for 2013-08-01 (validation dataset).

3.3. Yearly averaged results

To gain an overall sense on the model capabilities, we discuss in this sections annual averaged model results. For time-averaged results, we selected our best model and compared it with NWP and ground truth (GT), as shown in the top row of Figs. 12, 13 and 14. As evident, on average, the bLSTM closely approximates the ground truth, as observed in the difference map (bottom row), preserving the majority of spatial features. This is in contrast to the NWP model, which tends to smooth

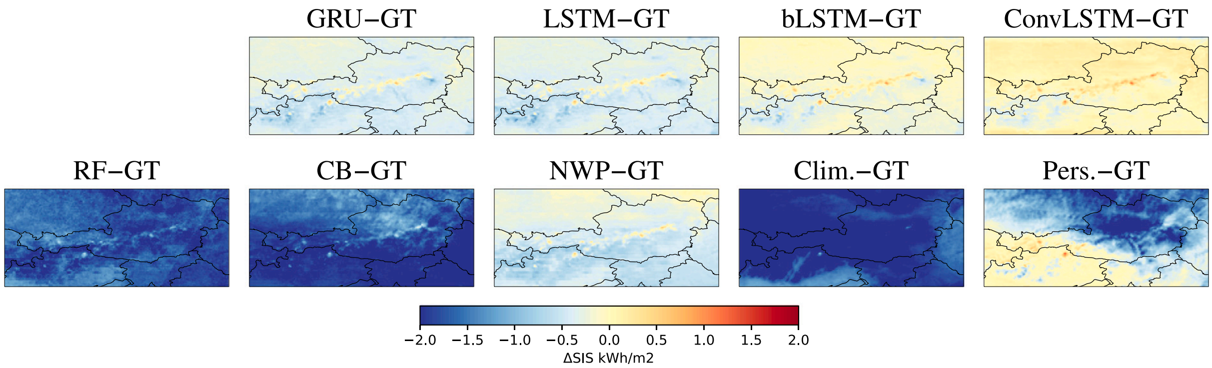

Fig. 10. Difference maps obtained by comparing the ground truth and the models' prediction at summer time for 2013-08-01 (validation dataset).

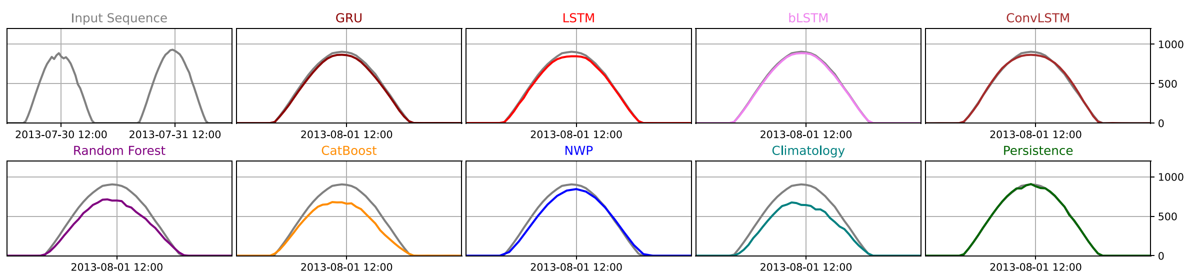

Fig. 11. Pixel level results, on Sonnblick, for 2013-08-01 (validation dataset).

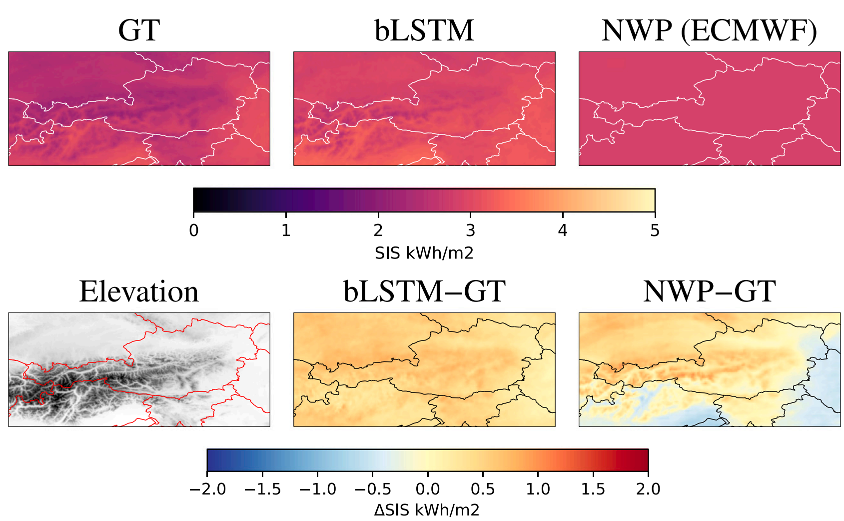

Fig. 12. Yearly averaged irradiation for 2013 (top), DEM and difference maps of the former (bottom), for Austria (validation dataset).

the prediction, and appears to be influenced by surface elevation (here obtained from a digital elevation model, DEM).

For the case of Austria, Fig. 12, the bLSTM has a negative bias (~ -0.5 kWh/m^2) while preserving most of the spatial features shown in GT, differently from the NWP model, showing a certain elevation-dependent bias.

For the case of Switzerland (where we applied transfer learning, i.e. performed inference without updating models' weights) in Fig. 13 it is possible to note that bLSTM is producing more accurate results, this can be noted also in the difference map, showing differences close to zero. As for the Austrian region, the NWP prediction is giving spatially homogeneous fields with a larger bias.

Finally for the case of Italy, where we also applied transfer learning, we can note that there is a spatial gradient across longitude both in bLSTM and NWP, Fig. 14. For the bLSTM this can be related to the process of transfer learning a model that has been trained on a slightly different distribution for the data, even though this model is returning also in this case better results than NWP. For this wider area the latitudinal variation of the surface irradiance is likely also important.

4. Discussion

In this study, several ML models are benchmarked in the context of day-ahead solar irradiance forecasting using EO data. The two preceding days derived from Meteosat geostationary satellite data and day of year are used as inputs. Our findings suggest that ML-based models can effectively complement existing solar irradiance forecasting tools for this time horizon (Kumar et al., 2020). Model performances could likely be boosted by accounting for local characteristics, such as elevation and terrain slope (Erickson, A., PhD Thesis, personal communication). As reported before (Montero-Martin et al., 2020), elevation is however not the main source of error in satellite-derived irradiance.

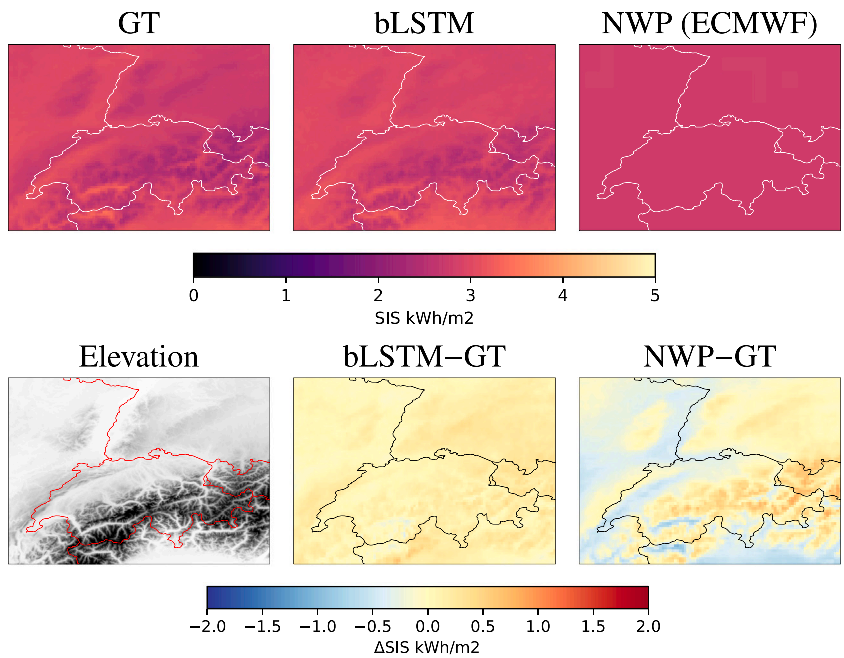

Fig. 13. Yearly averaged irradiation for 2014 (top), DEM and difference maps of the former (bottom), for Switzerland (test dataset).

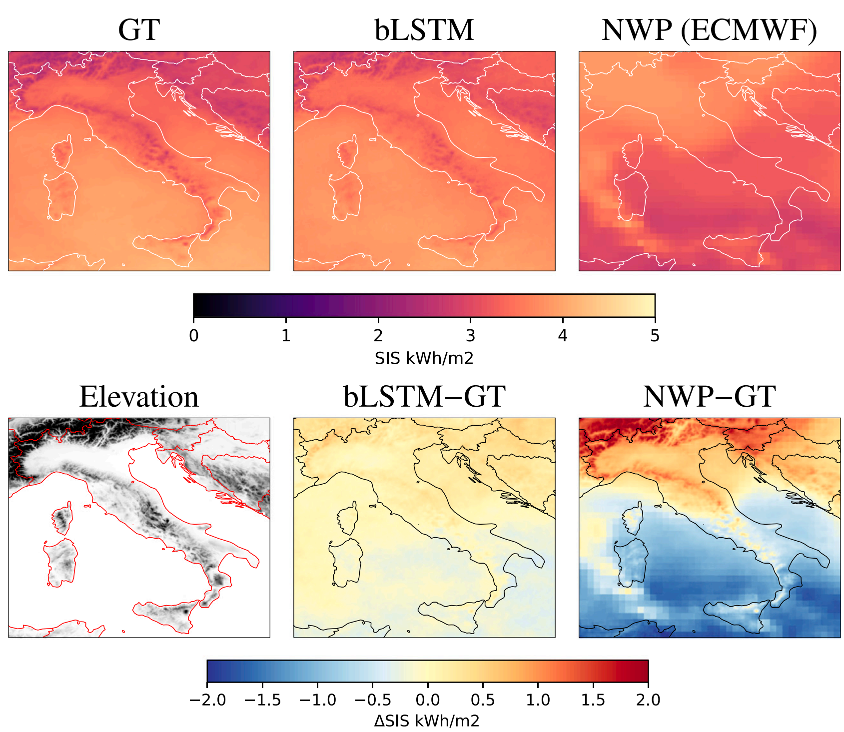

Fig. 14. Yearly averaged irradiation for 2016 (top), DEM and difference maps of the former (bottom) for Italy (test dataset).

The models were benchmarked against NWP, Climatology, and Persistence models, and were evaluated based on the nMAE. Results show that the tested deep-learning architectures (GRU, bLSTM, and LSTM models) outperform the other models, with the bLSTM model achieving the lowest error. This means that for pixel-based predictions there is an advantage in using well-tuned deep recurrent architectures. The spatial analysis highlighted how pixel-based predictions tend to produce higher variability, which is likely due to differences in the input data, while as expected (Nielsen et al., 2021) results for models with spatial convolutional cells tend to be more homogeneous. On the other hand, the seasonality analysis confirmed that there are significant dependencies on the time of the year, as weather conditions are more persistent in summer than in colder seasons. Due to the characteristics of the satellite retrieval algorithm and NWP model development, it is not possible to explore here the effects of long-term changes in surface irradiance (Wild et al., 2021). To this aim, additional data sources, in particular atmospheric aerosols and surface properties (such as land cover), should be taken into account.

Future research should focus on further improving the accuracy, robustness and efficiency of these models (Cowls et al., 2023), as well as exploring their potential applications in other EO-based frameworks. For example, it would be interesting to explore the use of multivariate prediction models (e.g., Gao et al., 2022) that incorporate high frequency *in situ* measurements, together with satellite data, also bypassing complex retrieval algorithms (Serva et al., 2022). In particular, using a more sophisticated loss function (Basri et al., 2020; Shifaz et al., 2023) would likely help obtaining a more realistic representation of high-frequency components that can be evaluated with domain-specific metrics (Vallance et al., 2017). In this way irradiance forecasts could be more performing also on sub-hourly basis, which we found to be difficult for our sequence-to-sequence models. The ability to timely predict rarer events, such as fast solar power variations, is indeed a crucial aspect of solar forecasting for which ML models trained on standard loss functions tend to underperform (Paletta et al., 2021). Additionally, investigating the integration of advanced ensemble techniques combining numerical predictions with the capabilities of various ML-based models could enhance the predictive power of these algorithms, while providing diverse scenarios for power systems operators to build on. Finally, future works could extend the spatial predictions to tie in with a larger context. For instance, they could explore the use of satellite data to provide insights on solar radiation across wider geographical areas, to be processed with a cloud-aware prediction models (Sebastianelli et al., 2022).

5. Conclusions

The seamless and dependable distribution of variable renewable energies into the power grid is critical to facilitate the transition towards a low carbon economy. Solar power, for instance, is strongly impacted by the cloud cover dynamics over plant sites. To improve the reliability of this energy source, power systems heavily rely on solar forecasting tools to best adapt the response of the electric grid to upcoming fluctuations, from minutes to days and beyond. However, current approaches based on physics models still have difficulties in accurately predicting solar radiation at ground level due to its volatile behavior.

We demonstrate the potential of ML-based techniques for reliable day-ahead prediction of surface solar irradiance, with deep learning models performing best across different seasons and geographical settings. Trained models can also effectively be deployed to perform prediction over unseen regions. Among possible improvements, the use of multivariate models, physically grounded loss functions, and combinations of different data sources (e.g. weather prediction models and ground-based observations), should be explored in future research. This could help improve the accuracy of solar irradiance forecasting from minutes to days ahead by taking into account regional characteristics and global weather patterns, providing actionable insights with low latency.

We believe that this study provides valuable insights into the use of ML for surface irradiance forecasting, and that further research on this topic is required to facilitate the integration of solar energy in the grid to implement large-scale decarbonization strategies.

CRediT authorship contribution statement

Alessandro Sebastianelli: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Federico Serva:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrea Ceschini:** Writing – original draft, Software, Methodology, Conceptualization. **Quentin Paletta:** Writing – original draft. **Massimo Panella:** Writing – original draft, Supervision, Project administration, Conceptualization. **Bertrand Le Saux:** Writing – original draft, Supervision, Project administration, Methodology, Conceptualization.

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

All the data we used are openly available for research purpose to registered users.

Acknowledgments

We acknowledge constructive comments from three reviewers. We are grateful to the providers of the data used in this work: EUMETSAT CM-SAF (Pfeifroth et al., 2019) for surface irradiance data, ECMWF for the IFS forecast outputs, and the BSRN network for *in-situ* radiative measurements (Olefs, 2013).

Code and data availability

All the data we used are openly available for research purpose to registered users. Information to access code and data resources can be found at https://irradianceai.github.io/ (last accessed Aug. 2024).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.rse.2024.114431.

References

Aguiar, L.M., Pereira, B., Lauret, P., Díaz, F., David, M., 2016. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. Renew. Energy 97, 599–610.

Ayet, A., Tandeo, P., 2018. Nowcasting solar irradiance using an analog method and geostationary satellite images. Sol. Energy 164, 301–315.

Basri, R., Galun, M., Geifman, A., Jacobs, D., Kasten, Y., Kritchman, S., 2020. Frequency bias in neural networks for input of non-uniform density. In: Proceedings of the 37th International Conference on Machine Learning. ICML '20, Vol. 119, JMLR.org, pp. 685–694.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Cano, D., Monget, J.-M., Albuisson, M., Guillard, H., Regas, N., Wald, L., 1986. A method for the determination of the global solar radiation from meteorological satellite data. Solar Energy 37 (1), 31–39.

Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y., 2014. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.

Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S., 2019. Machine learning regressors for solar radiation estimation from satellite data. Sol. Energy 183, 768–775.

Cowls, J., Tsamados, A., Taddeo, M., Floridi, L., 2023. The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. Ai Society 1–25.

Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., Denn, F.M., Duprat, T., Fukuda, M., Grobe, H., Haeffelin, M., Hodges, G., Hyett, N., Ijima, O., Kallis, A., Knap, W., Kustov, V., Long, C.N., Longenecker, D., Lupi, A., Maturilli, M., Mimouni, M., Ntsangwane, L., Ogihara, H., Olano, X., Olefs, M., Omori, M., Passamani, L., Pereira, E.B., Schmithüsen, H., Schumacher, S., Sieger, R., Tamlyn, J., Vogt, R., Vuilleumier, L., Xia, X., Ohmura, A., König-Langlo, G., 2018. Baseline surface radiation network (BSRN): structure and data description (1992–2017). Earth Syst. Sci. Data 10 (3), 1491–1501. http://dx.doi.org/10.5194/essd-10-1491-2018.

Dubus, L., Muralidharan, S., Troccoli, A., 2018. What does the energy industry require from meteorology? Weather Clim. Serv. Energy Ind. 41–63.

Espeholt, L., Agrawal, S., Sønderby, C., et al., 2022. Deep learning for twelve hour precipitation forecasts. Nature Commun. 5145 (13), http://dx.doi.org/10.1038/s41467-022-32483-x.

Gailhofer, P., Herold, A., Schemmel, J.P., Scherf, C.-S., Urrutia, C., Koehler, A.R., Braungardt, S., 2021. The role of Artificial Intelligence in the European Green Deal. Technical Report, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, URL https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662906/IPOL_STU(2021)662906_EN.pdf.

Gao, Y., Miyata, S., Akashi, Y., 2022. Interpretable deep learning models for hourly solar radiation prediction based on graph neural network and attention. Appl. Energy 321, 119288. http://dx.doi.org/10.1016/j.apenergy.2022.119288.
Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM networks. In: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.. Vol. 4, IEEE, pp. 2047–2052.
Haiden, T., Dahoui, M., Ingleby, B., de Rosnay, P., Prates, C., Kuscu, E., Hewson, T., Isaksen, L., Richardson, D., Zuo, H., Jones, L., 2018. Use of in situ surface observations at ECMWF. ECMWF Tech. Memo. http://dx.doi.org/10.21957/DJ9LPY4WA, URL https://www.ecmwf.int/node/18748.
Haiden, T., Trentmann, J., 2016. Verification of cloudiness and radiation forecasts in the greater Alpine region. Meteorol. Z. 25 (1), 3–15. http://dx.doi.org/10.1127/metz/2015/0630.
Haupt, S.E., Kosović, B., Jensen, T., Lazo, J.K., Lee, J.A., Jiménez, P.A., Cowie, J., Wiener, G., McCandless, T.C., Rogers, M., et al., 2018. Building the Sun4Cast system: Improvements in solar power forecasting. Bull. Am. Meteorol. Soc. 99 (1), 121–136.
Heptonstall, P., Gross, R., 2021. A systematic review of the costs and impacts of integrating variable renewables into power grids. Nat. Energy (6), 72–83. http://dx.doi.org/10.1038/s41560-020-00695-4.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Huang, G., Li, Z., Li, X., Liang, S., Yang, K., Wang, D., Zhang, Y., 2019. Estimating surface solar irradiance from satellites: Past, present, and future perspectives. Remote Sens. Environ. 233, 111371. http://dx.doi.org/10.1016/j.rse.2019.111371, URL https://www.sciencedirect.com/science/article/pii/S0034425719303906.
Jiang, H., Lu, N., Qin, J., Tang, W., Yao, L., 2019. A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data. Renew. Sustain. Energy Rev. 114, 109327.
Johnson, O.W., Han, J.Y.-C., Knight, A.-L., Mortensen, S., Aung, M.T., Boyland, M., Resurrección, B.P., 2020. Intersectionality and energy transitions: A review of gender, social equity and low-carbon energy. Energy Res. Soc. Sci. 70, 101774.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Kumar, D.S., Yagli, G.M., Kashyap, M., Srinivasan, D., 2020. Solar irradiance resource and forecasting: a comprehensive review. IET Renew. Power Gener. 14 (10), 1641–1656.
Leirvik, T., Yuan, M., 2021. A machine learning technique for spatial interpolation of solar radiation observations. Earth Space Sci. 8 (4), e2020EA001527.
MacKay, D.J.C., 2009. Sustainable Energy–Without the Hot Air. UIT Cambridge Ltd..
Markovics, D., Mayer, M.J., 2022. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew. Sustain. Energy Rev. 161, 112364. http://dx.doi.org/10.1016/j.rser.2022.112364, URL https://www.sciencedirect.com/science/article/pii/S136403212200274X.
Martinot, E., 2016. Grid integration of renewable energy: Flexibility, innovation, and experience. Annu. Rev. Environ. Resour. 41 (1), 223–251. http://dx.doi.org/10.1146/annurev-environ-110615-085725.
Michael, N.E., Mishra, M., Hasan, S., Al-Durra, A., 2022. Short-term solar power predicting model based on multi-step CNN stacked LSTM technique. Energies 15 (6), 2150.
Montero-Martin, J., Anton, M., Vaquero-Martinez, J., Sanchez-Lorenzo, A., 2020. Comparison of long-term solar radiation trends from CM SAF satellite products with ground-based data at the Iberian peninsula for the period 1985–2015. Atmos. Res. 236, 104839.
Moreno-Garcia, I.M., López-Luque, R., Varo-Martínez, M., Fernández-Ahumada, L.M., Ramírez-Faz, J.C., de la Torre, F.C., 2019. An Approach for the Solar Energy Assessment using Weather Medium-Range Forecasting. In: 2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe. EEEIC / I&CPS Europe, pp. 1–5. http://dx.doi.org/10.1109/EEEIC.2019.8783583.
Müller, R., Pfeifroth, U., Träger-Chatterjee, C., Trentmann, J., Cremer, R., 2015. Digging the METEOSAT Treasure—3 Decades of Solar Surface Radiation. Remote Sens. 7 (6), 8067–8101. http://dx.doi.org/10.3390/rs70608067.
Narvaez, G., Giraldo, L.F., Bressan, M., Pantoja, A., 2021. Machine learning for site-adaptation and solar radiation forecasting. Renew. Energy 167, 333–342.
Nespoli, A., Niccolai, A., Ogliari, E., Perego, G., Collino, E., Ronzio, D., 2022. Machine learning techniques for solar irradiation nowcasting: Cloud type classification forecast through satellite data and imagery. Appl. Energy 305, 117834.
Nie, Y., Paletta, Q., Scott, A., Pomares, L.M., Arbod, G., Sgouridis, S., Lasenby, J., Brandt, A., 2024. Sky image-based solar forecasting using deep learning with heterogeneous multi-location data: Dataset fusion *versus* transfer learning. Appl. Energy 369, 123467. http://dx.doi.org/10.1016/j.apenergy.2024.123467.
Nielsen, A.H., Iosifidis, A., Karstoft, H., 2021. IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. Sol. Energy 228, 659–669.
Obiora, C.N., Ali, A., 2021. Effective implementation of convolutional long short-term memory (convlstm) network in forecasting solar irradiance. In: IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society. IEEE, pp. 1–5.
Olefs, M., 2013. Horizon at station Sonnblick. http://dx.doi.org/10.1594/PANGAEA.808997.
Paletta, Q., Arbod, G., Lasenby, J., 2021. Benchmarking of deep learning irradiance forecasting models from sky images – An in-depth analysis. Sol. Energy 224, 855–867. http://dx.doi.org/10.1016/j.solener.2021.05.056.
Paletta, Q., Arbod, G., Lasenby, J., 2023a. Omnivision forecasting: Combining satellite and sky images for improved deterministic and probabilistic intra-hour solar energy predictions. Appl. Energy 336, 120818. http://dx.doi.org/10.1016/j.apenergy.2023.120818.
Paletta, Q., Terrén-Serrano, G., Nie, Y., Li, B., Bieker, J., Zhang, W., Dubus, L., Dev, S., Feng, C., 2023b. Advances in solar forecasting: Computer vision with deep learning. Adv. Appl. Energy 11, 100150. http://dx.doi.org/10.1016/j.adapen.2023.100150.
Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., K., Renné, D., Hoff, T.E., 2010. Validation of short and medium term operational solar radiation forecasts in the US. Sol. Energy 84 (12), 2161–2172.
Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Van Knowe, G., Hemker Jr., K., Heinemann, D., Remund, J., Müller, S.C., Traunmüller, W., et al., 2013. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. Sol. Energy 94, 305–326.
Pérez, E., Pérez, J., Segarra-Tamarit, J., Beltran, H., 2021. A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant. Sol. Energy 218, 652–660.
Pfeifroth, U., Kothe, S., Trentmann, J., Hollmann, R., Fuchs, P., Kaiser, J., Werscheck, M., 2019. Surface radiation data set - heliosat (SARAH) - edition 2.1. https://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=SARAH_V002_01.
Pfeifroth, U., Sanchez-Lorenzo, A., Manara, V., Trentmann, J., Hollmann, R., 2018. Trends and variability of surface solar radiation in Europe based on surface- and satellite-based data records. J. Geophys. Res.: Atmos. 123 (3), 1735–1754. http://dx.doi.org/10.1002/2017JD027418, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017JD027418.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst. 31.
Qing, X., Niu, Y., 2018. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy 148, 461–468. http://dx.doi.org/10.1016/j.energy.2018.01.177, URL https://www.sciencedirect.com/science/article/pii/S0360544218302056.
Ramadhan, R.A., Heatubun, Y.R., Tan, S.F., Lee, H.-J., 2021. Comparison of physical and machine learning models for estimating solar irradiance and photovoltaic power. Renew. Energy 178, 1006–1019.
Rosato, A., Araneo, R., Andreotti, A., Succetti, F., Panella, M., 2021. 2-d convolutional deep neural network for the multivariate prediction of photovoltaic time series. Energies 14 (9), 2392.
Samu, R., Calais, M., Shafiullah, G., Moghbel, M., Shoeb, M.A., Nouri, B., Blum, N., 2021. Applications for solar irradiance nowcasting in the control of microgrids: A review. Renew. Sustain. Energy Rev. 147, 111187. http://dx.doi.org/10.1016/j.rser.2021.111187, URL https://www.sciencedirect.com/science/article/pii/S1364032121004755.
Sebastianelli, A., Puglisi, E., Del Rosso, M.P., Mifdal, J., Nowakowski, A., Mathieu, P.P., Pirri, F., Ullo, S.L., 2022. PLFM: Pixel-level merging of intermediate feature maps by disentangling and fusing spatial and temporal data for cloud removal. IEEE Trans. Geosci. Remote Sens. 60, 1–16.
Serva, F., Ansalone, L., Mathieu, P.-P., 2022. Retrieval of marine parameters from hyperspectral satellite data and machine learning methods. In: International Conference on Applied Intelligence and Informatics. Springer, pp. 351–363.
Shifaz, A., Pelletier, C., Petitjean, F., Webb, G.I., 2023. Elastic similarity and distance measures for multivariate time series. Knowl. Inf. Syst. 65 (6), 2665–2698. http://dx.doi.org/10.1007/s10115-023-01835-4.
Simoes, S., Zeyringer, M., Mayr, D., Huld, T., Nijs, W., Schmidt, J., 2017. Impact of different levels of geographical disaggregation of wind and PV electricity generation in large energy system models: A case study for Austria. Renew. Energy 105, 183–198. http://dx.doi.org/10.1016/j.renene.2016.12.020.
Söderholm, P., 2020. The green economy transition: the challenges of technological change for sustainability. Sustain. Earth 3 (1), 6. http://dx.doi.org/10.1186/s42055-020-00029-y.
Solano, E.S., Dehghanian, P., Affonso, C.M., 2022. Solar Radiation Forecasting Using Machine Learning and Ensemble Feature Selection. Energies 15 (19), 7049.
Succetti, F., Di Luzio, F., Ceschini, A., Rosato, A., Araneo, R., Panella, M., 2021. Multivariate prediction of energy time series by autoencoded lstm networks. In: 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe. EEEIC/I&CPS Europe, IEEE, pp. 1–5.
Succetti, F., Rosato, A., Araneo, R., Panella, M., 2020. Deep neural networks for multivariate prediction of photovoltaic power time series. IEEE Access 8, 211490–211505.
Vallance, L., Charbonnier, B., Paul, N., Dubost, S., Blanc, P., 2017. Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric. Sol. Energy 150, 408–422. http://dx.doi.org/10.1016/j.solener.2017.04.064.
Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: A review. Renew. Energy 105, 569–582.

Wild, M., Wacker, S., Yang, S., Sanchez-Lorenzo, A., 2021. Evidence for clear-sky dimming and brightening in central Europe. Geophys. Res. Lett. 48 (6), e2020GL092216.

Yagli, G.M., Yang, D., Gandhi, O., Srinivasan, D., 2020. Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance? Appl. Energy 259, 114122.

Yagli, G.M., Yang, D., Srinivasan, D., 2019. Automatic hourly solar forecasting using machine learning models. Renew. Sustain. Energy Rev. 105, 487–498.

Yang, D., Wang, W., Bright, J.M., Voyant, C., Notton, G., Zhang, G., Lyu, C., 2022. Verifying operational intra-day solar forecasts from ECMWF and NOAA. Sol. Energy 236, 743–755. http://dx.doi.org/10.1016/j.solener.2022.03.004, URL https://www.sciencedirect.com/science/article/pii/S0038092X22001645.

Zhao, S., Wu, L., Xiang, Y., Dong, J., Li, Z., Liu, X., Tang, Z., Wang, H., Wang, X., An, J., et al., 2022. Coupling meteorological stations data and satellite data for prediction of global solar radiation with machine learning models. Renew. Energy 198, 1049–1064.

Zhou, Y., Liu, Y., Wang, D., Liu, X., Wang, Y., 2021. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. Energy Convers. Manage. 235, 113960.