Combining Deep Learning and Physical Models: A Benchmark Study on All-Sky Imager-Based Solar Nowcasting Systems

Yann Fabel,* Bijan Nouri, Stefan Wilbert, Niklas Blum, Dominik Schnaus, Rudolph Triebel, Luis F. Zarzalejo, Enrique Ugedo, Julia Kowalski, and Robert Pitz-Paal

Intermittent solar irradiance due to passing clouds poses challenges for integrating solar energy into existing infrastructure. By making use of intrahour nowcasts (very short-term forecasts), changing conditions of solar irradiance can be anticipated. All-sky imagers, capturing sky conditions at high spatial and temporal resolution, can be the basis of such nowcasting systems. Herein, a deep learning (DL) model for solar irradiance nowcasts based on the transformer architecture is presented. The model is trained end-to-end using sequences of sky images and irradiance measurements as input to generate point-forecasts up to 20 min ahead. Further, the effect of integrating this model into a hybrid system, consisting of a physics-based model and smart persistence, is examined.

A comparison between the DL and two hybrid models (with and without the DL model) is conducted on a benchmark dataset. Forecast accuracy for deterministic point-forecasts is analyzed under different conditions using standard error metrics like root-mean-square error and forecast skill. Furthermore, spatial and temporal aggregation effects are investigated. In addition, probabilistic nowcasts for each model are computed via a quantile approach. Overall, the DL model outperforms both hybrid models under the majority of conditions and aggregation effects.

1. Introduction

As climate change is advancing, the need for renewable energy sources becomes increasingly urgent. Solar power is one of the most promising forms of renewable energy, offering a clean and abundant source of electricity. However, the amount of solar radiation reaching the Earth's surface, i.e., solar irradiance, is highly variable. Apart from seasonal and diurnal changes, solar irradiance depends primarily on atmospheric conditions. Variations in conditions are mainly caused by aerosols and clouds. This variability poses a challenge for the efficient operation and integration of solar power plants into the grid. For example, a sudden drop of solar irradiance of a few $100\,\mathrm{W\,m^{-2}\,min^{-1}}$ due to passing clouds, directly impacts the power output of a photovoltaic (PV) plant. Especially for grids with an increased share of solar energy, such volatile conditions can lead to large mismatches in energy supply and demand. As a result, plant operation and grid integration have to become more flexible to allow further expansion of these intermittent energy sources.[1,2] A prerequisite for dealing with such problems is to anticipate future conditions. In previous studies, it has been shown that solar forecasts could significantly reduce mitigation measures and corresponding costs.[3–5]

There are generally two distinct approaches to solar forecasting. First, data-driven approaches which obtain forecasts purely

Y. Fabel, B. Nouri, S. Wilbert, N. Blum
German Aerospace Center (DLR)
Institute for Solar Research
Calle Doctor Carracido No 42, floor 1, 04005 Almería, Spain
E-mail: yann.fabel@dlr.de

D. Schnaus, R. Triebel
Chair of Computer Vision & Artificial Intelligence
Technical University Munich
85748 Garching, Germany

L. F. Zarzalejo, E. Ugedo
CIEMAT Energy Department
Renewable Energy Division
28040 Madrid, Spain

J. Kowalski
Chair of Methods for Model-based Development in Computational Engineering
RWTH Aachen University
52056 Aachen, Germany

R. Pitz-Paal
Chair of Solar Technology
RWTH Aachen University
52056 Aachen, Germany

R. Pitz-Paal
German Aerospace Center (DLR)
Institute for Solar Research
51147 Cologne, Germany

from learned patterns on historic training data. In this category, all statistical or machine learning methods fall. Second, physics-based approaches that model physical processes in the atmosphere to generate forecasts, like cloud kinematics or transmittance. It should be emphasized that also physical models often include some data-driven methods in their processing pipeline[6] and vice versa.[7] Thus, in this context, we categorize models as physics-based when they incorporate implementations of physical processes and as data-driven when they rely on data without explicit modeling of physics.

Apart from data-driven versus physics-based, solar forecasting can be categorized into three main groups based on forecast horizon and exogenous data[8]: numerical weather predictions (NWP), satellite-based, and approaches based on local observations (e.g., irradiance measurements from ground stations or sky images). NWP are usually applied for larger forecast horizons like day-ahead forecasts,[9] although in recent studies also hour-ahead forecasts have been investigated using NWP.[10] Generally, satellite-based methods are typically used for intraday forecasts.[11,12] Forecasts based on local observations such as all-sky imagers (ASI) are a common choice for intrahour forecasts.[6,13,14] Combined with a network of cameras, also an increased nowcast horizon and spatial coverage can be achieved.[15] The rationale behind this categorization stems from the stochastic behavior of cloud dynamics on a local scale. Since temporal and spatial resolutions are lower for NWP and satellite-based models than for models derived from local observations, these irradiance forecasts are typically limited in considering such dynamics.

Hence, to deal with short-term fluctuations induced by cloud passings, highly resolved intrahour forecasts are needed. Such shortest-term forecasts for the next minutes to the next hour are here referred to as nowcasts. Classic ASI-based approaches are built upon physical modeling.[6,16] Recent developments however indicate a trend toward data-driven approaches,[13,14,17] reflecting the great success of deep learning (DL) in computer vision. However, there are still limitations regarding the accuracy of such models. As stated in ref. [13], the analyzed DL models are not yet capable of anticipating ramp events, i.e., a sudden increase or drop of irradiance, but they rather behave like better persistence models (as defined in 3.1.1).

Another category of forecasting models is so-called hybrid models. Here, predictions of distinct forecasting models are combined to generate an improved forecast. Previous studies have shown that combining distinct models in such a hybrid setup can improve the forecast accuracy.[6,18]

A major challenge for all methods is the inherent uncertainty coming with the forecast. For end users like plant operators, this information is crucial as it can determine actions the operator will take. Therefore, the solar forecasting community encourages to provide probabilistic information instead of purely deterministic predictions.[19] As a result, recent studies also focus on probabilistic forecasts.[20–22] There are basically two types of probabilistic methods applied in solar forecasting: quantile and ensemble forecasts.[23] The latter is common practice in NWP, where an ensemble forecast is generated by running the model multiple times with slightly modified initial conditions. In case of satellite and sky image data, and thus for solar nowcasting, quantile forecasts are more widely used. Quantiles are used to summarize a predictive distribution and represent threshold values below which an event occurs with a certain probability. Given two quantiles with probability levels τ_1 and τ_2, a corresponding prediction interval (PI) can be derived. A PI quantifies the uncertainty by providing a range within which the true value is expected to fall with a certain level of confidence. For instance, a centered PI with coverage rate of 95%, can be estimated when quantiles $q_{\tau=0.025}$ and $q_{\tau=0.975}$ are known. Using quantile regression models, quantiles can be predicted directly.[24] Like in regular regression, the model's parameters are optimized by minimizing the so-called quantile loss, i.e., the error of predicting a certain quantile. Alternatively, quantiles from a distribution of historic prediction errors can be combined with the deterministic forecasts in a postprocessing procedure to generate PIs.[22] In both cases, parametric, i.e., assuming a distribution function, and nonparametric methods are applicable.

Furthermore, the accuracy of nowcasting systems heavily depends on the prevailing cloud conditions and thus location and time. For an expressive comparison to other methods, a common database is therefore required. A recent benchmark study[25] has compared several state-of-the-art ASI nowcasting systems using selected dates of versatile cloud conditions. Evaluating most common error metrics for solar forecasting, the study can serve as baseline for new developments.

In this work, we evaluate a new DL model based on the transformer architecture, as well as the combination of this DL model with a physics-based model, representing a new hybrid model. Both approaches show significant enhancements in nowcast accuracy compared to previous results shown in the benchmark study.[25]

2. Data

In the following sections, we describe the data basis of this work. First, we give an overview on the measured data. Next, a more detailed look is taken on the training data for the DL model. Lastly, we briefly present the validation data from the benchmark study.

2.1. Data Overview

All measurement data utilized in this work were acquired from ground-based meteorological stations at CIEMAT's (Spanish research institute: Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas) Plataforma Solar de Almería (PSA), located in southern Spain (37° 5′ 38″ N and 2° 21′ 32″ W). Being surrounded by multiple mountain ranges, very complex cloud conditions are often observed at PSA, making it a challenging site for irradiance nowcasts, as also previously analyzed in ref. [26]. In another study,[27] it was also shown that the distribution of clouds, with respect to cloud base height, is quite uniform at PSA, characterized by a slightly higher prevalence of low-layer clouds overall.

Regarding measurement devices, we used ISO 9060 class A pyranometers for global and diffuse horizontal irradiance (GHI/DHI) and class A pyrheliometers for direct normal irradiance (DNI). Sky images were recorded by off-the-shelf Q24 and Q25 Mobotix surveillance cameras. These sky images and

measurement data are utilized for different applications in this study. First, we define a dataset for training the DL model. These data come from a single station (ID 1). Second, we evaluate our nowcasting models on a 28 days benchmark dataset. For this purpose, the DL model uses images and irradiance data from the station ID 2. The physical model, as part of the hybrid setup, employs a stereoscopic approach to detect cloud height and rely on supplementary sky images from a secondary ASI (ID 10).[6] Third, validation is conducted on the 28 days using reference measurements from the stations with the IDs 1–9. Fourth, the probabilistic nowcasts are obtained by employing a quantile approach, which is individually applied to each of the investigated nowcasting models. The necessary quantile analysis utilizes a dataset of 657 days (excluding validation data), consisting of reference data from station IDs 1–9 and corresponding to deterministic nowcasts from each nowcasting model. The individual outcomes of the quantile analysis are then applied to the deterministic nowcasts of the 28 days lasting validation dataset, thereby generating probabilistic nowcasts for each model under investigation. Lastly, the climatological probabilistic baseline model relies on a 13-year-long dataset comprising irradiance measurements from station ID 1. An overview of the different datasets and their applications is given in **Table 1** and the position of each station is drawn in **Figure 1**.
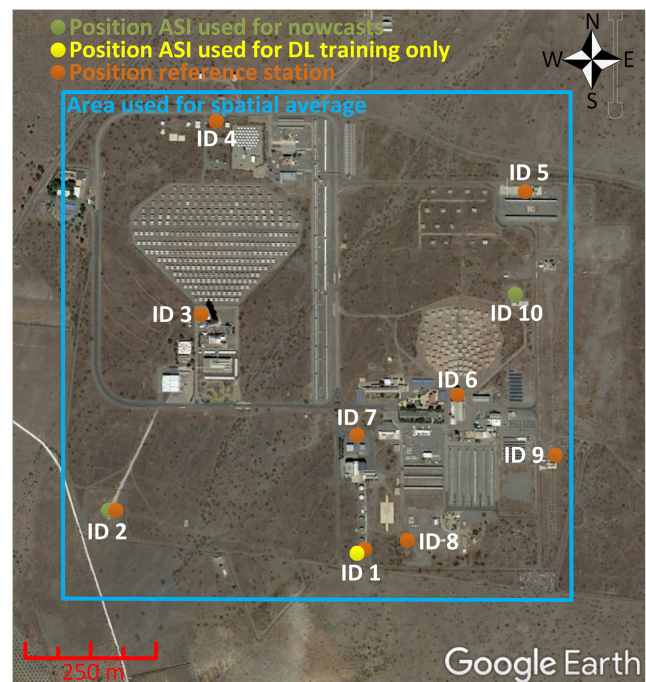
2.2. Training Data Selection

For training our DL models, we use sky images from a single ASI located at the station 1 (see Figure 1). In addition to the sky images, irradiance measurements from the same station are included. The dataset encompasses more than 3 million images distributed over the years 2015–2021. Thus, versatile conditions for all seasons and solar positions are considered. It should be highlighted that these data (i.e., sky images and irradiance measurements) were only used for model training and not for validating the DL approach against the other nowcasting systems.

We filter out training data for all cases with a sun elevation lower than 10°. Conditions with a sun elevation angle of less than 10° are more strongly affected by disturbing objects (vegetation, buildings, etc.) in the image as well as increased distortion effects of the ASI fish-eye camera lens. Moreover, we eliminate instances of low variability, such as clear sky and complete overcast conditions. Thereby, we obtain a more balanced dataset in relation to


Figure 1. Positions of reference stations and ASIs (Source: Google Earth Google 2021, accessed 25 May 2021). Spatial averaging is conducted for validation (Section 3.4).

cloud conditions and we reduce training effort as instances of low variability offer no insights into how transient clouds impact solar irradiance. This is particularly necessary because PSA predominantly experiences clear sky conditions. To identify these conditions, a DNI variability classification based on ref. [28] and adapted according to ref. [29] was applied. The classes give information on the optical properties of the prevailing clouds and therefore an implicit information on the cloud types. A variability class is determined by analyzing DNI over the past 15 min, ranging from clear sky (class 1) to overcast (class 8). The classification procedure combines 13 established variability indices from the literature, like average clear sky index. An overview of all classes is given in **Table 2**. Instances of classes of 1, 2, and 8 are then partly removed. This corresponds to a reduction of around 90% and 75% for classes 1 and 2, respectively, and almost 60% for

Table 1. Utilized datasets and their application in this work. It should be emphasized that the 28 days benchmark dataset was used exclusively for nowcast validation.

Application	Station ID (camera)	Station ID (reference)	Time period
DL training	1	1	2015–2021
Nowcast validation (deterministic/probabilistic)	2, 10	1–9	28 days (2019)
Analysis of error quantiles (probabilistic nowcasts)	2, 10	1–9	657 days (2019–2021)
Climatological baseline model (CSD-Clim)	–	1	2006–2018

Table 2. Overview of DNI variability classes according to ref. [28].

Class	Sky conditions	Clear sky index	Variability
1	Mostly clear sky	Very high	Low
2	Almost clear sky	High	Low
3	Almost clear sky	High/intermediate	Intermediate
4	Partly cloudy	Intermediate	High
5	Partly cloudy	Intermediate	Intermediate
6	Partly cloudy	Intermediate/low	High
7	Almost overcast	Low	Intermediate
8	Mostly overcast	Very low	Low
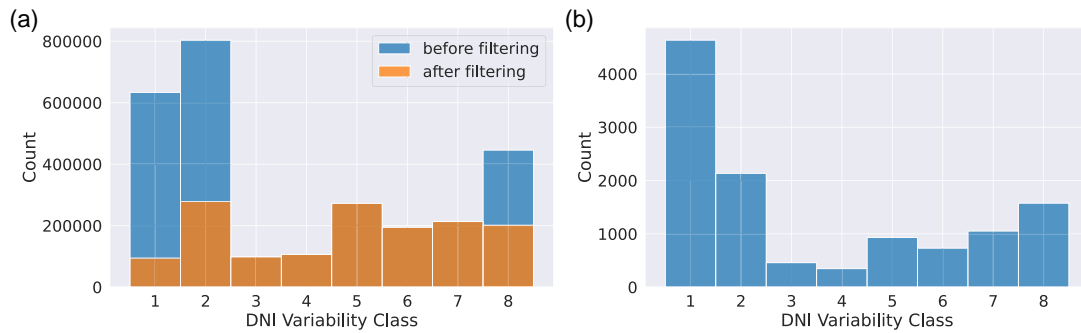
Figure 2. Distribution of DNI variability classes for: a) training dataset before and after filtering out data of low variability conditions and b) benchmark (test) dataset.[25] A larger share of clear sky conditions is included as completed days were selected.

class 8. To obtain temporally contiguous data within the dataset, only full hours that are completely dominated by one of these three classes are removed. The distribution of the data before and after filtering by DNI variability classes is shown in **Figure 2**a.

Besides reference measurements, model-based data are also incorporated for training the DL model. Solar azimuth and elevation values for the reference station are computed using the NREL SPA algorithm[30] while clear sky irradiance for GHI, DNI, and DHI is obtained based on the Linke turbidity (TL) according to ref. [31].

Regarding data splitting, we split training and validation by randomly selecting complete days with a ratio 0.8–0.2. The assumption is that each day is independent of its predecessor in terms of short-term variability.

2.3. Data Preprocessing

To exploit the temporal resolution of the sky images (resolution of 30 s) compared to the irradiance measurements (resolution of 1 min), the temporal resolution of irradiance measurements is increased to 30 s by linear interpolation. We further normalize GHI, DNI, and DHI, by the corresponding clear sky irradiance, which is derived according to ref. [31], obtaining clear sky indices k_{irr}^*:

$$k_{irr}^* = irr/irr_{clear} \qquad (1)$$

where irr refers to GHI, DNI, and DHI, respectively. Thereby, seasonal and daily variations due to varying solar positions and atmospheric turbidities are taken into account. Moreover, the input data x for the DL models, i.e., time series and image data, is standardized using data from the training set, as defined in Equation (2).

$$x_{f,t}^* = (x_{f,t} - \mu_f)/\sigma_f \qquad (2)$$

where f defines the feature (e.g., k_{GHI}^*) for time series data or the RGB channel for sky images, t specifies a timestamp, and μ and σ represent the (feature/channel-wise) mean and standard deviations over the training dataset. Images for DL training are further cropped to the fish-eye section and resized to squares of shape 128×128 whereby the round fish-eye view is retained.

2.4. Benchmark Dataset

In this work, we evaluate our models on a recent benchmark dataset presented in ref. [25] In this benchmark, five ASI-nowcasting models were compared on data of 28 manually selected days between September and November 2019. The days were selected to obtain a versatile dataset of varying cloud conditions. Despite choosing mainly variable days with cloud passages for the benchmark, clear sky conditions remain dominant as can be seen in the distribution of DNI variability classes shown in Figure 2b.

3. Methodology

In this section, we describe the applied methodology of this work. First, we give an overview of the applied nowcasting models, in particular the newly developed DL model in Section 3.1 and the combination of models into a hybrid approach (Section 3.2). We briefly summarize the calculation of probabilistic nowcasts (Section 3.3) and present the utilized baseline model (Section 3.3.1), however, for a more detailed description, the interested reader is referred to ref. [22]. Afterward, we present the validation procedure for deterministic and probabilistic nowcasting (Section 3.4).

3.1. Nowcasting Models

In the following subsections, we briefly present the individual deterministic nowcasting models that are analyzed in this work and which are used for the hybrid approach. We start with the smart persistence (SP) model, continue with the physical model and finally describe the new DL model. Each of these models produces minute-resolution predictions of GHI for every future time step (lead time (LT)) throughout the entire 20 min forecast horizon. The focus is on GHI, as for nonconcentrating PV applications, it is the most relevant quantity of solar irradiance.

3.1.1. SP Model

Assuming mostly stationary conditions, the so-called SP model is one of the simplest time series forecasting models. SP irradiance

nowcasts can be calculated according to ref. [31] The formula described in ref. [31] for GHI can be applied directly. I_0 describes the solar constant and h the height at the site. The TL parameter is calculated from most recent GHI measurements and kept constant, while the solar elevation angle α and air mass AM are adjusted according to the required LTs. Hence, the only difference to the most recent GHI measurement is due to changes in the sun position influencing α and AM.

$$
\begin{aligned}
&\text{GHI} = a_1 \cdot I_0 \cdot \sin(\alpha) \cdot \exp(-a_2 \cdot \text{AM} \cdot (f_{h1} + f_{h2} \cdot (\text{TL} - 1))) \\
&a_1 = 5.09 \times 10^{-5} \cdot h + 0.868 \\
&a_2 = 3.92 \times 10^{-5} \cdot h + 0.0387 \\
&f_{h1} = \exp(-h/8000) \\
&f_{h2} = \exp(-h/1250)
\end{aligned}
\tag{3}
$$

3.1.2. Physical Model

The second nowcasting model that is part of this study is the physical model as presented in ref. [6] The basis are two Q25 ASI which are positioned 890 m apart as well as ground-based DNI and DHI measurements at one of the ASI sites (see Figure 1). Nowcasts are generated by the following processing pipeline: 1) detect clouds from the individual images via a convolutional neural network;[32] 2) compute 3-D cloud models based on a stereoscopic approach;[33] 3) estimate cloud motion vectors from a sequence of images using cross correlation;[33] 4) predict future cloud positions by extrapolating cloud movement; 5) apply shadow projection of modeled clouds via ray tracing;[34] and 6) predict irradiance by analyzing radiative effect of clouds using cloud height and DNI measurements.[27]

The resulting outputs are irradiance maps (GHI/DNI) covering a space of about 64 km² up to 20 min ahead, with a spatial resolution of 20 m, a temporal resolution of 1 min, and an update rate of 30 s. **Figure 3** shows an overview of the processing pipeline of the physical model.

3.1.3. DL Model

The applied DL model serves as a multistep point forecast model for predicting solar irradiance (GHI). This entails generating a singular scalar value for each LT within the forecast horizon (20 min). It can be considered as a multimodal model, consisting of two branches for extracting features from irradiance measurements and images. Unlike many previous data-driven approaches in solar forecasting,[17] we do not utilize recurrent neural networks to learn temporal dependencies, instead, we apply two attention-based transformer architectures.

For time series data, consisting of irradiance measurements of GHI, DNI, and DHI, and sun elevation/azimuth values, we use the transformer architecture presented in ref. [35] based on the implementation of the Python package tsai.[36] The sequence length, i.e., the size of the input time series, is fixed to 30 min with one value per minute. In preliminary studies, this value has proven to work best for 20 min forecast horizons. As described in the original article, each time series input vector $x_{ts,t}$ for timestamp t is first linearly projected to a new vector $u_{ts,t}$. After adding a learnable positional encoding vector to $u_{ts,t}$, the encoded input is fed into a standard transformer encoder architecture as defined in ref. [37], resulting in time series representation vectors $z_{ts,t}$ of shape $[512 \times 1]$.

To extract features from ASI sequences, we use the timeSformer architecture.[38] Here, spatial and temporal dependencies in the image sequence are learned via a combined attention mechanism. The ASI input sequence length is set to 5 min, as larger values significantly prolong training duration. At the same time, longer ASI sequences do not lead to significant improvements in this setup as preliminary experiments
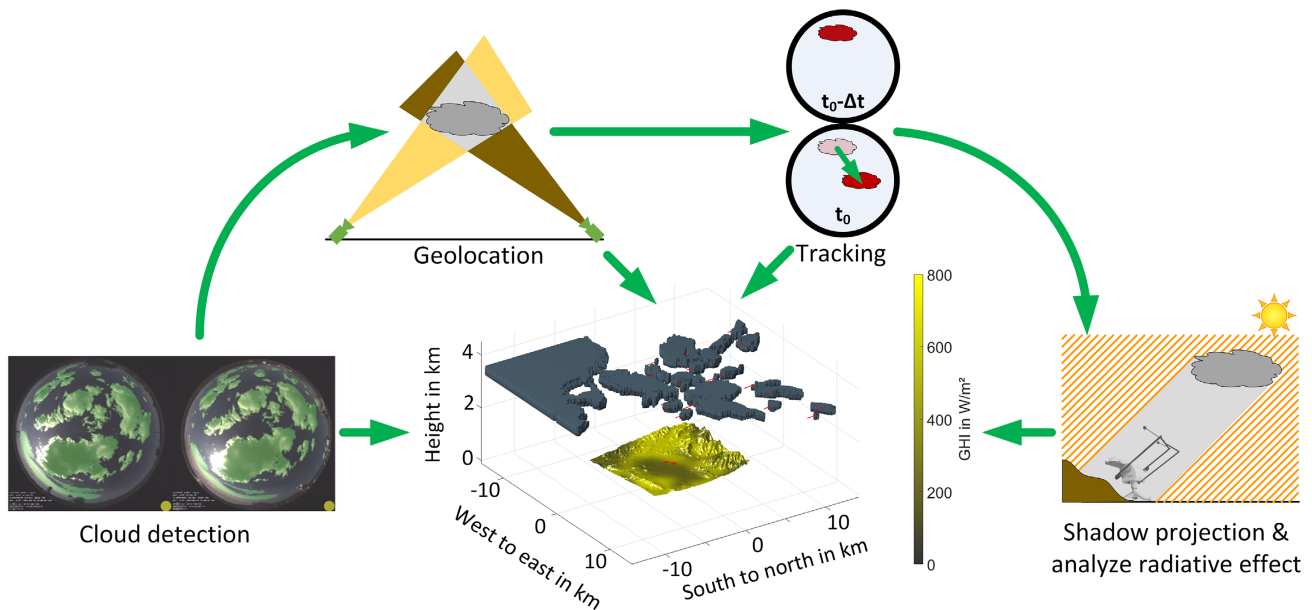
Figure 3. Illustration of processing pipeline of physical ASI model (reproduced from[6]).

indicated. While temporal attention is achieved by attending over the sequence of images, spatial attention is done via the ViT architecture,[39] dividing each image into patches of 16×16 pixels. For the combination of both attention mechanisms, we follow the "Divided Space-Time Attention" method as defined in ref. [38]. In this case, the model applies temporal and spatial attention separately. Temporal attention is limited to the respective patch over the image sequence whereas spatial attention is carried out only on the patches of the same image. The representation of each sky image sequence $z_{asi,t}$ of shape [512,1] is then concatenated to the time series representation $z_{ts,t}$ forming the final representation vector of shape [1024,1]. By feeding this vector to a two-layer fully-connected network, the multistep forecasts are obtained. A schematic graph is depicted in **Figure 4**.

As mentioned before, the DL model is trained end-to-end on predicting GHI. Thus, we optimize both parts, the time series transformer and the vision transformer together in a single training without any pretraining. Batch-wise data augmentation techniques, i.e., random rotations and flips, are applied to artificially enlarge the training data and prevent overfitting. Optimization is conducted via the AdamW optimizer[40,41] with weight decay of 0.01 and mean-square-error (MSE) as loss function. We use the one-cycle policy[42] for learning rate scheduling with $lr_{max} = 0.001$. The model is trained for 10 epochs with a batch size of 16 using the Python fastai library.[43] An overview of all training hyperparameters is given in **Table 3**.

3.2. Hybrid Model

In a previous work, a hybrid model has been presented, combining predictions from the aforementioned physical and persistence models.[6] The combination is based on real-time accuracy weighting as defined in ref. [44] The weighting is determined by a sliding validation based on computing recent root-mean-square errors (RMSE) of the respective GHI nowcasts to the corresponding reference measurements (ref) for the past 5 min. For each LT, the weighting is calculated separately depending on how well a model predicted GHI at this specific

**Table 3.** Training setup and hyperparameter selection of DL model.

Training parameter	Value
ASI input shape	$128 \times 128 \times 3 \times 5$ (height $\times$ width $\times$ channel $\times$ sequence length [min])
Time series input shape	5×30 (num feats $\times$ sequence length [min])
Time series features	k^*_{GHI}, k^*_{DNI}, k^*_{DHI}, sun elevation [°], sun azimuth [°]
ASI normalization mean (R, G, B)	(0.174, 0.170, 0.172)
ASI normalization std (R, G, B)	(0.138, 0.123, 0.118)
Time series normalization mean	(0.883, 0.686, 2.39, 37.09, 179.96)
Time series normalization std	(0.273, 0.373, 1.48, 17.25, 61.59)
Data augmentation (ASI)	Rotations, horizontal/vertical flips
Validation size	20%
Loss function	MSE
Optimizer	AdamW ($\beta_1 = 0.9$, $\beta_1 = 0.99$)
Weight decay	0.01
Maximum learning rate	0.001
Learning rate scheduler	one-cycle policy
Batch size	16
Number of epochs	10

LT in the recent past. In this work, we now add our novel DL model as third nowcasting method. The final hybrid nowcast is thus defined as:

$$RMSE_{LT,j} = \left(\frac{1}{N} \sum_{i=1}^{N=5} (GHI_{ref}(t_i) - GHI_{LT,j}(t_i))^2 \right)^{0.5} \tag{4}$$

$$GHI_{LT} = \frac{1}{\sum_{j=1}^{3} \frac{1}{RMSE_{LT,j}}} \sum_{j=1}^{3} \frac{GHI_{LT,j}}{RMSE_{E,j}} \tag{5}$$

where LT specifies a lead time between 1 and 20 min and the index j determines the nowcasting model, i.e., persistence, physical, or DL model.
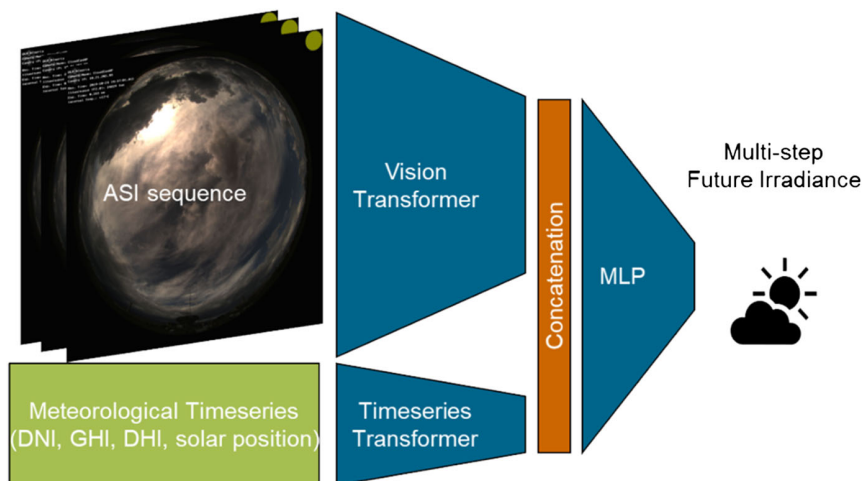
Figure 4. Schematic graph of generating forecasts with our DL model.

3.3. Probabilistic Nowcasts

The major drawback of deterministic nowcasts is the lack of information on how certain the model is with respect to its prediction. Solar energy applications however need reliable uncertainty estimations to effectively use such nowcasts. A well-suited approach to obtain information on a model's uncertainty is by calculating probability distributions over future events. From this, intervals can be derived that provide a measure of uncertainty. In ref. [22], a nonparametric probabilistic quantile nowcasting method has been presented which serves as a basis in this work. This approach and a climatological baseline model serving as validation model are presented in the following sections.

### 3.3.1. Baseline Model Clear-Sky-Dependent Climatology (CSD-Clim)

As baseline model, we chose the commonly used CSD-Clim model according to ref. [45] It is a climatological model solely based on historical data, in our case GHI measurements from 13 years period in 1 min resolution. CSD-Clim is statistically consistent and independent of forecast horizons or LTs. The main principle is the creation of clear sky index (k^*) distributions by grouping historic irradiance data into equal-sized bins of clear sky irradiance. Thereby, effects of varying sun positions and atmospheric turbidities are considered. We chose the same number of bins (30) as in the original study. A probabilistic nowcast from CSD-Clim is then obtained by taking the distribution of k^* that corresponds to the clear sky irradiance at nowcast time. Irradiance distribution in W m^{-2} and hence nowcast PIs are determined by multiplying the k^* distribution by the corresponding clear sky irradiance at nowcast time.

3.3.2. Quantile-Based Approach

The applied approach is based on the idea that the primary source of uncertainties in solar nowcasts lies in the variability of prevailing irradiance.[22] As a nonparametric method, no specific distribution function (e.g., Gaussian) is assumed; instead, the probability distribution for each nowcast is derived from historical prediction errors observed under similar conditions. To identify the prevailing conditions, we employ the identical DNI variability classification as introduced in Section 2.2. Using a dataset spanning 657 d, irradiance prediction errors and corresponding reference measurements, both normalized to k^*, are calculated, sorted, and used to compute quantiles of the error Δk^* with probability level (τ) from 1% to 99% for each LT. These quantiles are grouped by DNI variability classes to create a look-up table. Since DNI variability classes are determined from DNI values of the past 15 min, pure reference measurements cannot be used for LTs >0. Instead, a combination of DNI measurements and DNI predictions based on the LT represents the predicted DNI variability class. Given the LT-specific DNI variability classes of the nowcast, the quantiles of the expected error Δk^* are used to compute the quantiles of the nowcast:

$$k_\tau^* = k^* - \Delta k_{1-\tau}^* \qquad (6)$$

By multiplication with the respective clear sky irradiance, the irradiance quantile in W m^{-2} is obtained. Finally, the PIs for a given nowcast and LT can be deduced by the corresponding lower and upper-bound nowcast quantiles. **Figure 5** provides an overview of how probabilistic nowcasts are generated using this approach. Separate look-up tables are constructed for each of the three models under evaluation.
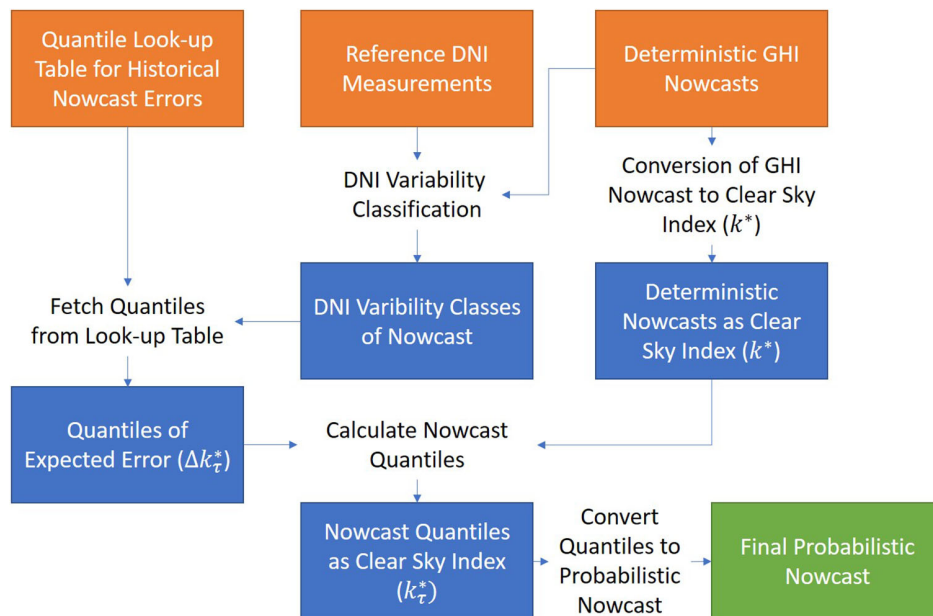
Figure 5. Process flow of quantile-based postprocessing approach to generate probabilistic nowcasts from deterministic ones as described in ref. [22] As input (orange), the current deterministic nowcast, recent reference measurements, and the look-up table of quantiles of historical errors are required. Interim results are depicted in blue and the output is the final probabilistic nowcast (green).

3.4. Validation Procedure

All our models are validated on the benchmark dataset from ref. [25], briefly described in Section 2.4. Thereby, a direct comparison to the other models from this benchmark, particularly the predecessor of the hybrid (DL) model can be made. Since all presented models provide multistep forecasts, i.e., for each time step, the model generates predictions for LTs 1–20 min into the future, we also evaluate many metrics for individual LTs. To give a more expressive insight on the accuracy of the nowcasting models depending on atmospheric conditions, we further discretize the dataset by DNI variability classes as described in Section 2.2. For the validation, DNI variability classification of each timestamp is done using reference DNI measurements only, to have the same distribution for all models.

3.4.1. Validation for Deterministic Nowcasts

For the deterministic nowcasts, we evaluate our models on standard forecasting metrics RMSE, MAE, mean bias error (MBE), and forecast skill (FS), often also referred to as skill score. The latter is of particular value in solar forecasting research as it enables comparisons to results from different locations and times, and thus should always be incorporated.[19] A more detailed definition is given in **Table 4**.

Apart from evaluating point forecasts matching the reference station, we also evaluate our models on the surrounding area. Hereby, we exploit the availability of multiple meteorological measurement stations at PSA. In total, we evaluate our nowcasting on 9 different stations as depicted in Figure 1 and calculate error metrics with respect to the individual station and to an arithmetic field average. For the hybrid models, nowcasts of the individual stations are obtained by looking at the respective

Table 4. Utilized error metrics for deterministic nowcasting with observations γ, i.e., measured reference GHI, and predictions $\hat{\gamma}$, i.e., predicted GHI, over all data points N in the validation.

Mathematical notation	Short description
$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\gamma}_i - \gamma_i)^2}$	RMSE: Measures forecast accuracy of a model by quantifying the average magnitude of the differences between predicted $\hat{\gamma}_i$ and observed (γ_i) values. Larger deviations are heavily penalized due to squaring.
$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{\gamma}_i - \gamma_i|$	MSE: Assesses the model accuracy by measuring the average magnitude of absolute differences between predicted and observed values. However, it does not penalize outliers as much as RMSE.
$MBE = \frac{1}{N}\sum_{i=1}^{N}(\hat{\gamma}_i - \gamma_i)$	MBE: Quantifies the average bias or systematic error of a forecast model. It calculates the average difference between predicted and observed values.
$FS = 1 - \frac{RMSE_{model}}{RMSE_{ref}}$	FS: Measures the forecast accuracy of a model to a baseline model. In this work, we use the SP model as described in Section 3.1.1 as baseline and RMSE as error metric. A positive skill score indicates that the forecasting model performs better than SP, while a negative score implies that SP outperforms the model.

pixel of the provided irradiance maps which matches the station's position. Persistence and DL predictions, however, are pure point forecasts, thus the same predictions are applied for all reference positions.

As mentioned before, we also calculate the arithmetic mean of the 9 reference stations to get a field average of the entire site (PSA). In case of the hybrid models, the average is calculated by the arithmetic mean on a sub-field of the irradiance map as highlighted in yellow in Figure 1. The DL and SP point predictions are compared directly with the arithmetic mean from the 9 reference stations.

Lastly, we also take a look into temporal aggregation of nowcasts up to 15 min, as such averages represent a more relevant measure when it comes to estimating short-term energy yield. Thereby, we calculate the average of the last nowcasts for each LT. For instance, when calculating the 10 min temporal average for LT 7, we take the arithmetic mean of the last 10 predictions of LT 7. This value is then compared to the mean of the reference measurements corresponding to these predictions.

3.4.2. Validation for Probabilistic Nowcasts

The evaluation of probabilistic nowcasts involves assessing three key attributes: sharpness, reliability, and resolution. Sharpness represents a measure of concentration of the predictive distribution independent of reference measurements. Reliability determines statistical consistency of nowcasts by examining whether the predicted probabilities align with the observed reference distribution. Resolution evaluates the ability to account for varying initial ambient conditions. In this work, we validate our probabilistic approach with a subset of the most relevant scores as proposed by refs. [46,47] For PI and quantile-dependent metrics, we evaluate interval score (IS) and quantile score (QS). Furthermore, we analyze continuous ranked probability score (CRPS), a single value score, aggregated over the predictive distribution. IS, QS, and CRPS are negatively oriented, meaning lower is better. As common reliability metric, we use PI coverage probability (PICP). We discuss the RMSE between PICP and PI, as PI can be treated as reference for the PICP. Thus, it is possible to evaluate an aggregated metric over the entire distribution. Lastly, we calculate a probabilistic nowcast skill based on CRPS, referred to as CRPSs. CSD-Clim serves here as baseline model. **Table 5** gives a brief summary of the utilized probabilistic error metrics.

4. Validation Results

This section presents the validation results of our nowcasting approaches as described in Section 3.1–3.3. For a comprehensive examination of our validation methodology, refer to Section 3.4. We initiate the analysis by comparing deterministic nowcasts, with a specific emphasis on the integration of the DL approach within the hybrid setup. Subsequently, we delve into the outcomes of probabilistic nowcasting. In this analysis, we refer to the deep learning model as *DL*, the hybrid model including the DL as Hybrid (DL), and the hybrid model as presented in ref. [6] as Hybrid. To put the absolute error values in relation, we included **Table 6** of the mean GHI of the 28 days validation set.

Table 5. Utilized error metrics for probabilistic nowcasting with observations y, i.e., measured reference GHI, and predictions $\hat{y}$), i.e., predicted GHI.

Mathematical notation	Attribute	Short description
$PICP_{PI} = \frac{1}{N}\sum_{i=1}^{N} 1_{y_i \in [L_{i,PI}, U_{i,PI}]}$	Reliability	PICP: Evaluates whether an observations y_i lies within the lower and upper bounds $(L_{i,PI}, U_{i,PI})$ of a specific PI.
$IS_{PI} = \frac{1}{N}\sum_{i=1}^{N}(U_{i,PI} - L_{i,PI})$ $+ \frac{2}{\alpha}(L_{i,PI} - y_i)1_{y_i < L_{i,PI}}$ $+ \frac{2}{\alpha}(y_i - U_{i,PI})1_{y_i > U_{i,PI}}$	Sharpness and reliability	IS: Measures the width of PIs, adding penalty values for observation falling outside of the interval. The amount of penalty is determined by the statistical significance level α.
$QS = \frac{1}{N}\sum_{i=1}^{N}\psi_\tau(y_i - \hat{y}_{i,\tau})$ $\psi_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases}$	Reliability and resolution	QS: Quantifies the deviation of observations to a predicted quantile τ, thus checking if a quantile is over- or underestimated.
$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_0^1 [F_{\hat{y}_i}(x) - F_{y_i}(x)]^2 dx$	Reliability and resolution	CRPS: Measures the discrepancy between the cumulative distribution function (CDF) of the nowcast $F_{\hat{y}_i}$ and the CDF of the observation F_{y_i}. It takes into account both the location and shape of the predictive distribution.
$CRPSs = 1 - \frac{CRPS_{model}}{CRPS_{ref}}$	Reliability and resolution	CRPS skill: Probabilistic nowcast skill based on CRPS and CSD-Clim as baseline model. Analogously to the FS for deterministic nowcasts, a positive CRPSs indicate a better probabilistic nowcast than obtained by the baseline model.

Table 6. Mean GHI values from reference measurements for the whole validation dataset and discretized by DNI variability classes.

	Total	DNI variability classes							
		1	2	3	4	5	6	7	8
Mean GHI [Wm^{-2}]	504	622	519	510	504	489	445	352	274

4.1. Comparison of Deterministic Models

We first evaluate the nowcasting models on RMSE, MAE, MBE, and FS for the entire 28 days benchmark dataset. In **Figure 6**, the error metrics for the different models are depicted. The solid line represents the base reference station (ID 2) and surrounding measurement stations are represented as shaded area. We see that for RMSE, all methods outperform SP starting from LT 2, as persistence is still quite accurate for LT1. However, while the difference to SP is rather small for the Hybrid model, the Hybrid (DL) model, and even more the standalone DL model achieve much lower RMSE values for LT > 3 min. This behavior is directly reflected in the FS plot, showing a great advantage of the DL model in terms of RMSE. Even for the more distant stations, a clear advantage of the DL model is visible. However, such a clear difference between the models is not visible when looking at MAE. Here, the Hybrid model and SP are almost equivalent, SP even performing slightly better for low LTs. The Hybrid (DL) model represents the best model in terms of MAE, as the DL has weaker accuracy on LT < 4 min. Still, when looking at other stations, the DL model occasionally achieves the best MAE for LT > 4 min. Regarding MBE, SP reaches the best value for the base station (ID 2), but with a clear overestimation when looking at other stations. This overestimation is even stronger for the hybrid models while the DL is the only model showing a tendency to underestimate.

Next, we evaluate the models on the arithmetic field average, as presented in **Figure 7**. Overall, the results are similar to the results at individual stations. DL has a clear advantage over the other models in terms of RMSE and thus FS. But here, DL also shows the best overall performance for MAE, however very close to the MAE of the Hybrid (DL). Only for the first 2 min, the Hybrid (DL) provides a clear benefit. This is to be expected due to the inclusion of SP. For MBE, all models show a slight increase compared to the MBE of station 2, resulting in positive values for all models and LTs and thus indicating some overestimation in general. This behavior can already be observed in the spread of the MBE in Figure 6, where the MBE of station 2 is close to the lower edge of the spread. For the remaining reference stations, mainly a positive and more pronounced bias is observed. This is not unexpected, given that the nowcasts were generated using station 2, where the predictions undergo a self-calibrating process.

An advantage of the DL model is also evident when analyzing nowcast accuracy for different irradiance conditions. In **Figure 8**, RMSE on the field average is shown discretized by DNI variability classes. As expected there is no significant difference between the models for mostly static conditions, like class 1 and 8, as solar irradiance changes mainly due to the sun position. For variable conditions, however, the DL model almost always reaches the lowest RMSE with a few exception at some isolated LTs where the Hybrid (DL) model is better.

As shown in **Figure 9**, higher temporal aggregation leads to reductions of RMSE and MAE for all models, while MBE remains almost the same. Reductions up to 50% are visible, when applying temporal aggregations of 15 min. Such behavior can be expected, as averaging over longer time spans smoothens fluctuations of the predictions. However, as can be seen in the plots for FS, SP also benefits from temporal aggregation, resulting in a decrease of FS for all models. Overall the DL model remains the most accurate approach.

In total, the data-driven DL approach has shown to be very effective in deterministic solar irradiance nowcasting. Due to MSE optimization of the DL model, a good score in RMSE and thus FS could be expected when analyzing GHI point forecasts at a single reference station. However, the DL model has turned out to generate better predictions than the hybrid models even when analyzing a limited surrounding area ($\approx$1 km^2). This is also surprising to us, as the hybrid models incorporate spatially
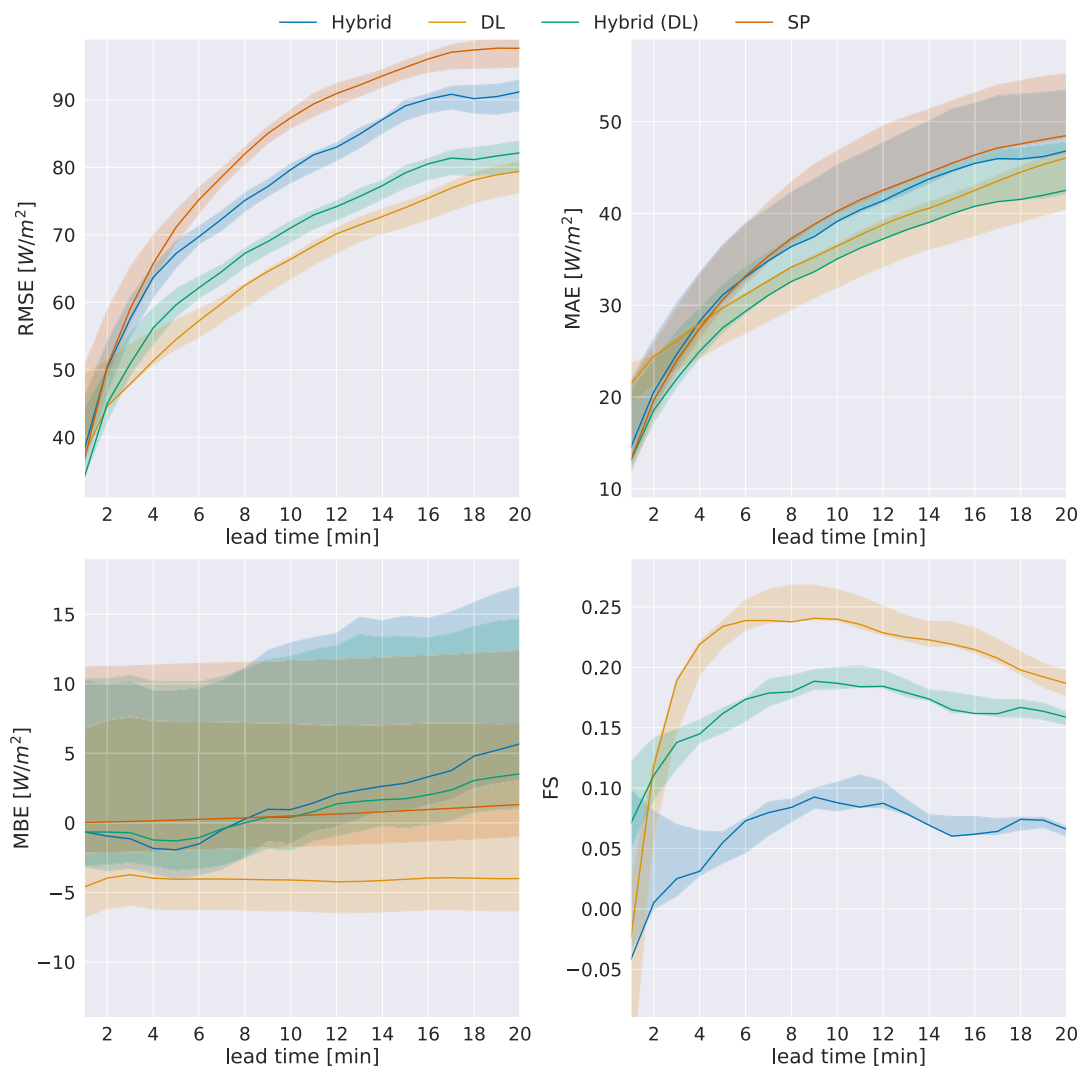
Figure 6. Comparison of deterministic nowcasting models for GHI predictions. The Hybrid (DL) model includes nowcasts from the presented DL model. Error metrics corresponding to the base reference station (ID 2) are highlighted as lines while the error metrics corresponding to the remaining eight reference stations are depicted as shaded area. FS is computed with respect to RMSE and SP.
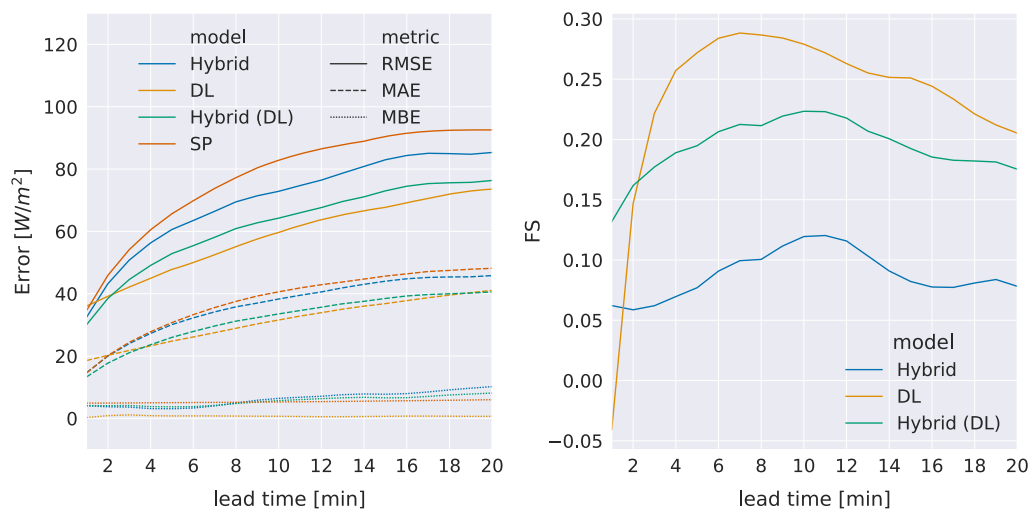
Figure 7. Deterministic nowcasts on field average. Left: RMSE, MAE, and MBE for all models. Right: FS.
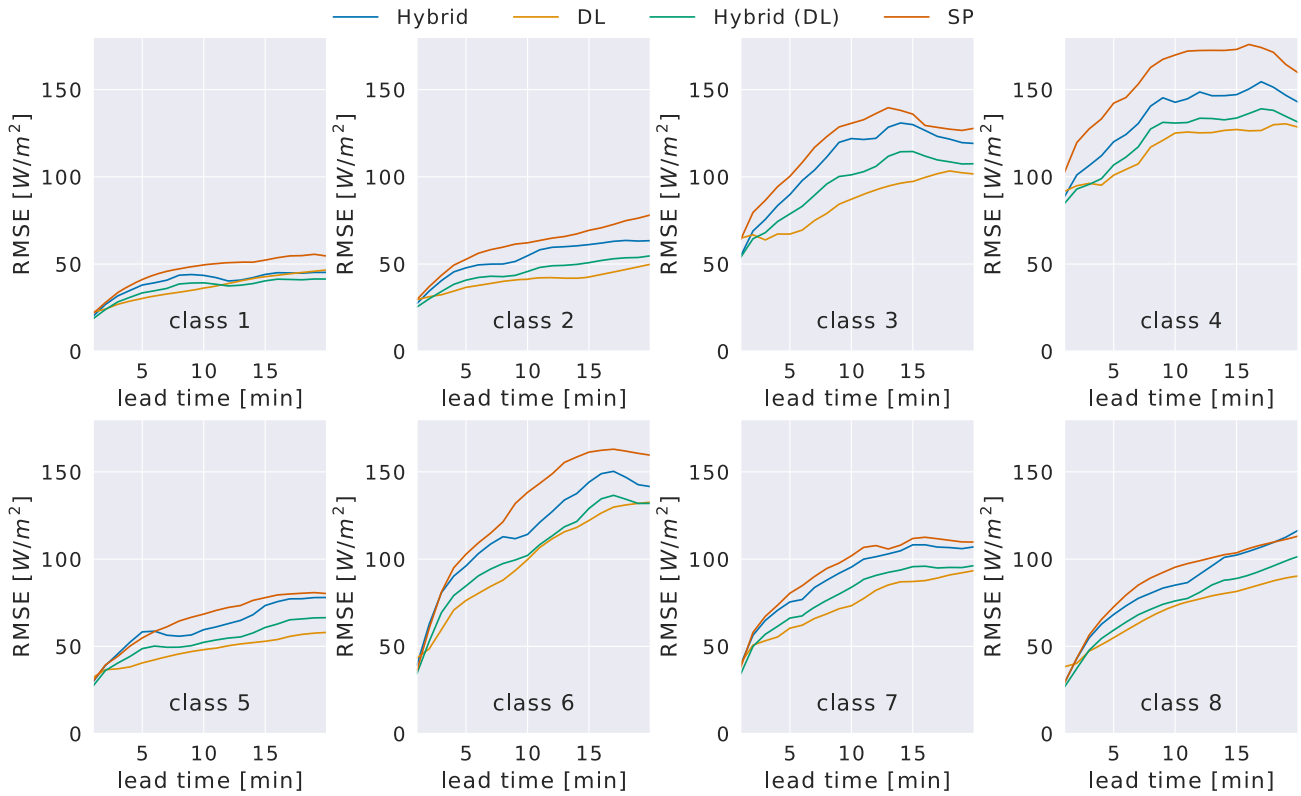
Figure 8. RMSE on field average discretized by DNI variability classes. MAE and MBE can be found in Appendix A1.
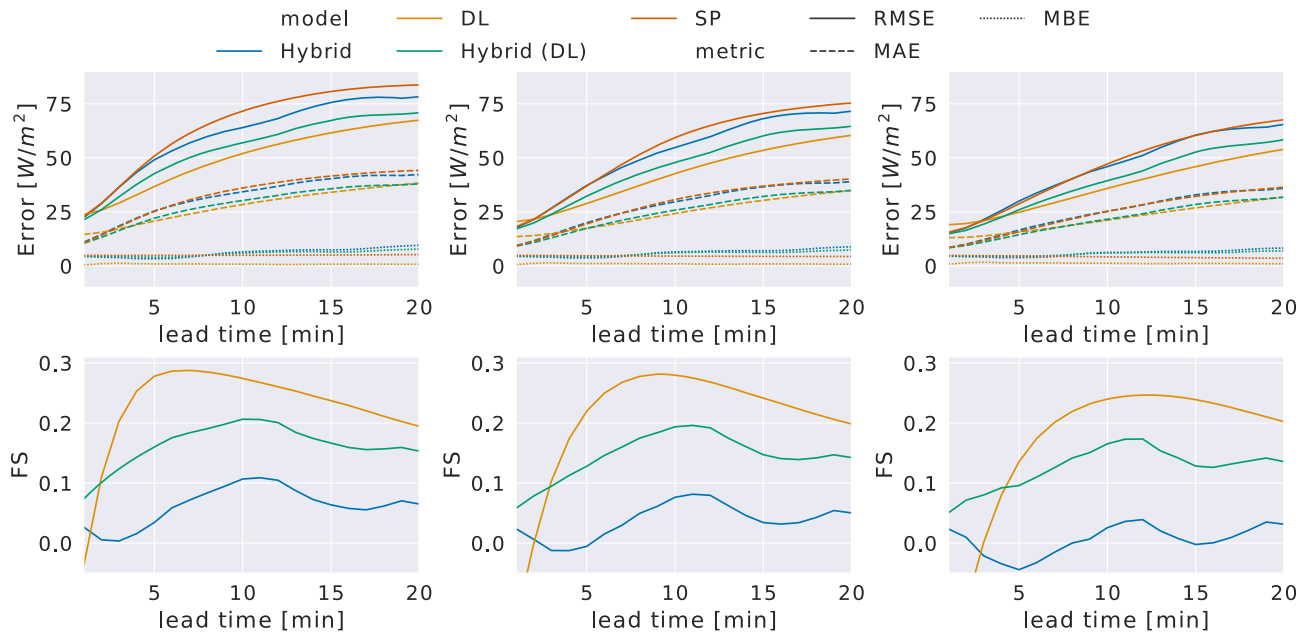
Figure 9. Error metrics of nowcasting models on field average for temporal aggregations of 5 (left), 10 (center), and 15 min (right).

resolved nowcasts from the physical model. The following factors could account for these results. First of all, throughout the last decade, DL has proven to be very effective in all types of computer vision applications compared to traditional image

processing methods. Moreover, an advantage of our new DL approach compared to the physical model is certainly the end-to-end training. As the DL model gets raw data in form of images and measurement values as input, it can learn to extract features

and complex patterns that are not evident or too complex to be integrated into a physical model. In particular for the sky images, information on cloud dynamics can be learned instead of clouds being modeled as rigid objects moving in a single direction. Lastly, there is no error propagation in such an end-to-end model compared to a physical model consisting of a long process chain.

Still, we should be careful to declare the DL model as unreservedly superior. First, it can be seen in Figure 8 that in some cases, the Hybrid (DL) model performs equally well or better. Second, the spatial resolution of the irradiance maps from the hybrid models is much larger ($>60\,km^2$) than the analyzed area ($\approx 1\,km^2$). Third, due to the higher penalization of larger errors, the DL model often predicts more conservatively. When the irradiance fluctuates significantly, the DL model tends to generate predictions around an average irradiance without capturing the actual conditions. Consequently, ramp events are less likely to be anticipated. In contrast, the hybrid or physical model can represent such highly variable conditions. At the same time, small temporal inaccuracies automatically lead to high errors. Hence, to analyze such characteristics, further error metrics are required, as also pointed out by ref. [48]

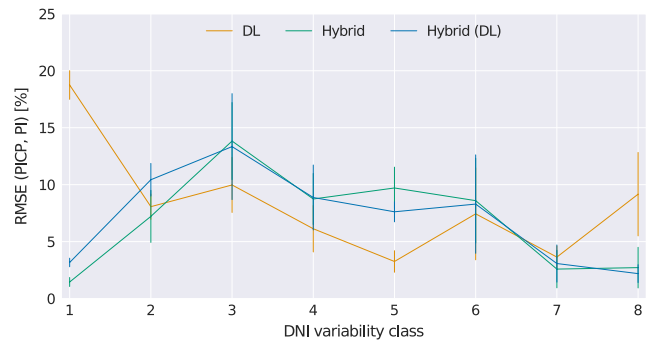## 4.2. Evaluation of the Probabilistic Model

In this section, probabilistic validation results are presented. The results are computed as before on the field average, calculated by the 9 reference stations. As indicated earlier, we evaluate CRPS, CRPSs, IS, QS, and PICP for DL, and the two hybrid models and make a comparison.

An overview of aggregated scores, averaging over all LTs, is presented in **Table 7**. In case of PICP, we calculate the RMSE of PICP and PI to specify nowcasting reliability, since PICP would match PI for a perfectly reliable nowcast distribution.

To compute PICP, quantiles 1,2,...,99% were used for each LT separately. Here, an advantage of the hybrid models is visible on average. Poor reliability from the DL model can be observed in particular for static irradiance conditions, as shown in **Figure 10**. For DNI variability classes 1 and 8, representing mostly clear sky and overcast conditions, RMSE(PICP,PI) of up to 20% occurs in
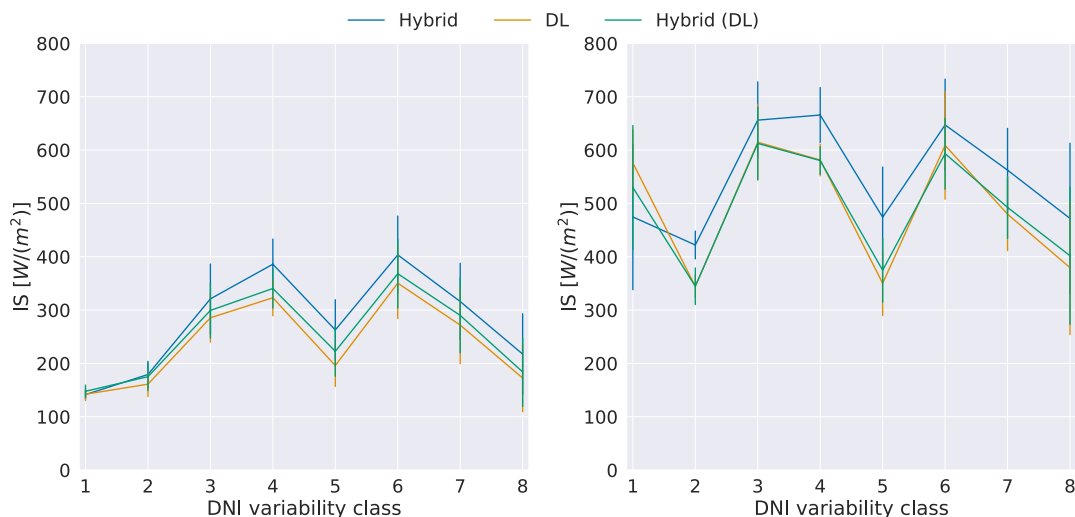

Figure 10. RMSE(PICP, PI) grouped by DNI variability classes in percentage. The line represents the average of overall LTs while the error bars show the standard deviation for each class.

the predictions of the DL model. From the deterministic predictions (Figure 8 and A1), it can be seen that the error for classes 1 and 8 is relatively low, leading to smaller PIs in general. Consequently, even slight deviations of a few $W\,m^{-2}$, e.g., due to aerosol variability or very thin clouds, can cause predictions to fall outside the PI. For the DL model, the high RMSE(PICP,PI) for classes 1 and 8 indicate that for these conditions the prevailing bias is sufficient to cause such outliers (refer also to Figure **A2**). In case of more variable conditions, however, the RMSE(PICP,PI) curve of the DL model is mostly below the ones of the hybrid models, showing that under these conditions the bias does not influence PICP as much. Due to the high share of class 1 conditions in the dataset, the DL model still gets a higher RMSE(PICP,PI) on average (Table 7).

For the IS, we look at two specific PIs, namely the ones with coverage probability corresponding to one and two standard deviations (σ, 2σ) from a normal distribution. As shown in Table 7, all models achieve scores in a similar range of around $200\,W\,m^{-2}$ for $PI_{68.3\%}$, with the DL model having the best mean over all LTs of $181.9\,W\,m^{-2}$. At $PI_{95.4\%}$ differences between the models in IS are again within $20\,W\,m^{-2}$. Though, this time the Hybrid (DL) model reaches the best mean of $487.6\,W\,m^{-2}$. Analyzing IS on DNI variability class level, as depicted in **Figure 11**, it becomes evident that DL and the Hybrid (DL) model show similar behaviors under most conditions. The largest differences are visible for classes 1, 5, 7, and 8. A higher average IS for DL at $PI_{95.4\%}$ comes mainly from the aforementioned relatively poor reliability under class 1 conditions.

Similar behavior can be observed for the QS. Independent of the probability level τ, all models reach average scores deviating only about 1 or $2\,W\,m^{-2}$ from each other. Nevertheless, the DL model has the best scores. This coincides with the results plotted in **Figure 12**. The typical shape of higher scores for the most variable conditions and lower scores for more static conditions is clearly visible for all models, whereas DL mostly outperforms the hybrid approaches.

Lastly, we look at CRPS and CRPSs. Described as one of the most important scores in probabilistic solar forecasting,[47] CRPS represents a comprehensive probabilistic metric compressed into a single numerical value. Moreover, by computing the nowcast skill of the CRPS using CSD-Clim as baseline model, a relative analysis is carried out. The last row of Table 7 shows the

Table 7. Comparison of probabilistic nowcasting between DL, Hybrid, and Hybrid (DL) based on PICP, CRPS, IS, and QS. Nowcasts are evaluated on the field average of the 9 reference stations. For each model, the mean score and its standard deviation over LTs are given. Best values highlighted in bold..

| Metric	DL	Hybrid (DL)	Hybrid
RMSE (PICP,PI) [%]	8.8 ± 0.7	4.5 ± 0.9	$\mathbf{3.0 \pm 0.7}$
IS ($PI_{68.3\%}$) [$W\,m^{-2}$]	$\mathbf{181.9 \pm 30.0}$	192.8 ± 30.8	202.9 ± 32.4
IS ($PI_{95.4\%}$) [$W\,m^{-2}$]	506.5 ± 16.6	$\mathbf{487.6 \pm 34.0}$	495.2 ± 41.0
QS ($\tau = 20\%$) [$W\,m^{-2}$]	$\mathbf{18.3 \pm 2.7}$	19.5 ± 2.9	19.7 ± 2.9
QS ($\tau = 50\%$) [$W\,m^{-2}$]	$\mathbf{20.0 \pm 3.4}$	20.3 ± 3.9	22.0 ± 4.2
QS ($\tau = 80\%$) [$W\,m^{-2}$]	$\mathbf{13.3 \pm 2.7}$	13.7 ± 2.6	15.6 ± 3.0
CRPS [$W\,m^{-2}$]	$\mathbf{23.5 \pm 4.9}$	24.5 ± 5.6	26.0 ± 6.3

Figure 11. IS grouped by DNI variability classes for PIs of coverage probabilities 68.3% and 95.4%. The line represents the average of overall LTs while the error bars show the standard deviation for each class.
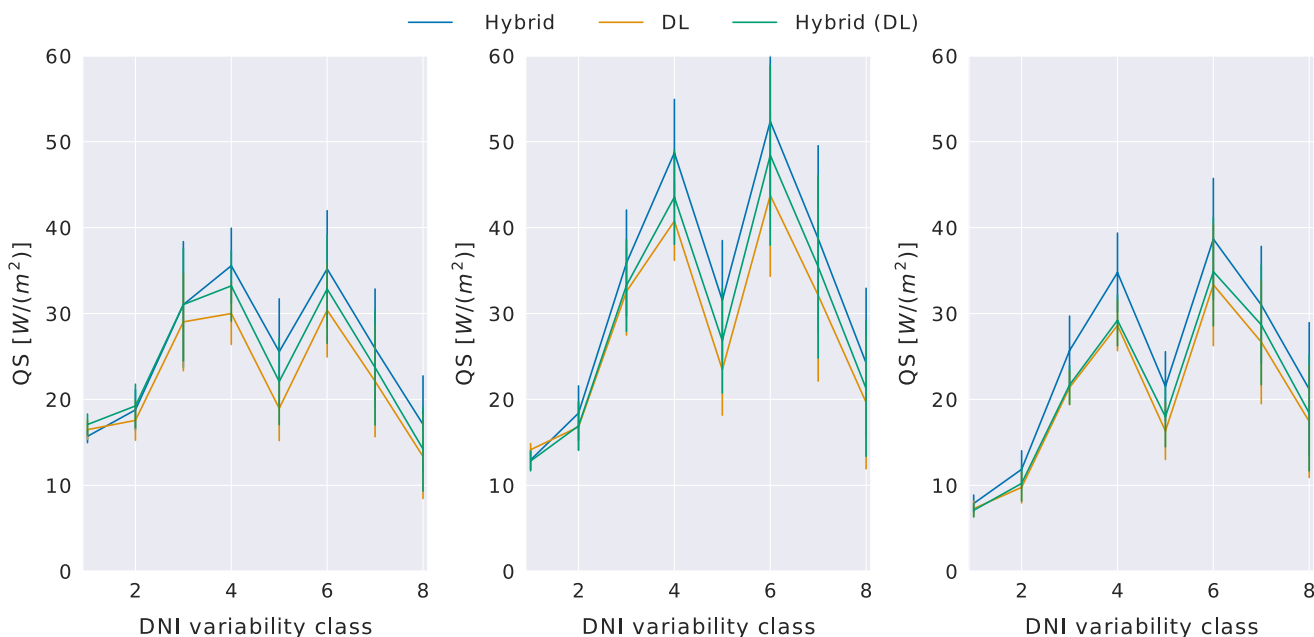
Figure 12. QS grouped by DNI variability classes for quantiles corresponding to probability levels 20%, 50%, and 80%. The line represents the average of overall LTs while the error bars show the standard deviation for each class.

average CRPS over all LTs, where the DL outperforms the hybrid models slightly by 1 to 1.5 W m^{-2}. In **Figure 13**, the CRPSs are presented grouped by LT and DNI variability classes. A noteworthy observation is the presence of a plateau among all models at LTs ranging from 9 to 13 min (left plot of Figure 13), indicating that CRPS of CSD-Clim increases more slowly around these LTs. The biggest advantages of the DL model compared to the hybrid models are visible for intermittent and high variability classes 3–6. Under these conditions, the wide distributions of climatological baseline models are comparatively strong, leading to

lower skills in general. This is contrary to the behavior of FS for deterministic predictions. Deterministic persistence approaches are by definition incapable to predict ramps, and therefore perform poorly during variable conditions. This means that in particular for conditions of highest uncertainty, the DL model can provide more reliable predictions compared to the hybrid models. This outcome is not surprising, as all models rely on the same quantile approach for estimating the expected distribution. Considering that DL scored best for deterministic nowcasting, it was expected to see an advantage also for the probabilistic
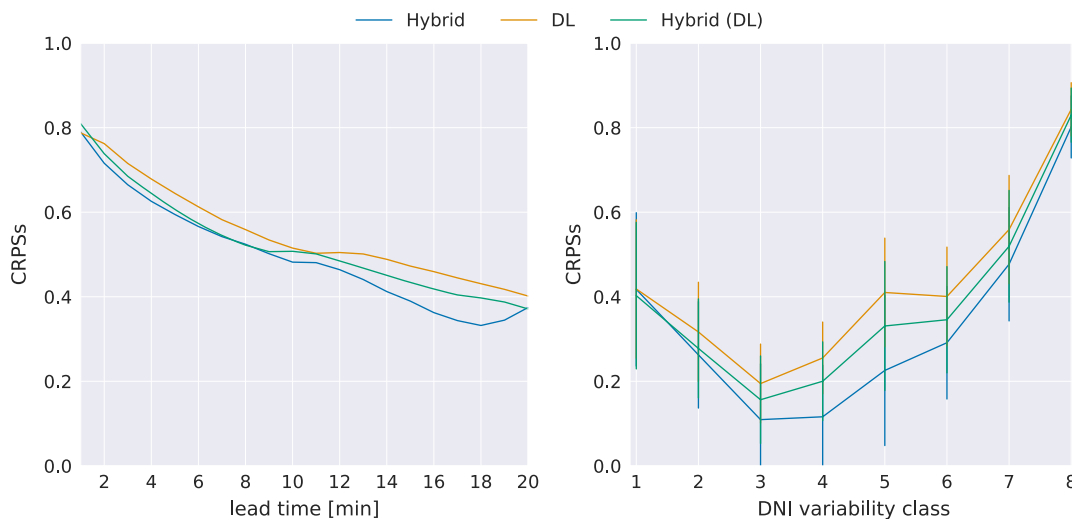
Figure 13. CRPSs over LTs (left) and DNI variability classes (right) with CSD-Clim as reference model. In the right plot, the average of overall LTs is depicted as line with the error bars as standard deviations for each class.

nowcasts. Moreover, the smallest standard deviations can be observed for the DL model (see Table 7), indicating higher stability over LTs. Nonetheless, all models still demonstrate significant uncertainties under highly variable conditions as can be seen in Figure 11 and 12, where the highest scores are dominated by classes 4 and 6, followed by classes 3 and 7.

5. Conclusion

In the last years, purely data-driven approaches for solar irradiance nowcasting have gained significant attention in the solar research community for their ease of implementation and strong performance. In this work, we showed that state-of-the-art transformer architectures are a well-suited choice for this task.

We analyzed the deterministic nowcast accuracy in terms of RMSE, MAE, MBE, and FS for our standalone DL model and compared it to a hybrid system, including and excluding the DL model. Achieving the best FS values, the strengths of the data-driven DL model became clear. Optimization based on MSE loss naturally leads to lower RMSE and higher RMSE-based FS values. Although the hybrid model benefits from the DL model, it does not reach the RMSE of the purely data-driven approach. Even for limited areas ($\approx 1\,\mathrm{km}^2$), the DL model provides better RMSE than the hybrid approach with its irradiance maps. Here, the DL even achieves the best MAE values for most LTs. For temporal aggregation, the error for all models, including SP, reduces for higher aggregation windows. It shows that the temporal resolution is decisive when analyzing models. A comparison of nowcasting systems with different resolutions is only legitimate if all systems are aligned in space and time.

In terms of probabilistic nowcasting, similar tendencies are visible with an advantage of DL compared to the hybrid models. However, the Hybrid (DL) model comes close to the DL model in terms of overall performance. All three models clearly achieve a positive skill score. Learning to minimize the error however often

comes with some artifacts. As previous studies have already shown,[13] such data-driven models often generate "smoothed" irradiance curves. These could lead to a reduced-average RMSE but may not accurately capture the present conditions of highly variable irradiance. A smoothed nowcast curve might exhibit similar mean absolute errors but with a significantly lower frequency and magnitude of pronounced errors. Precisely those outliers are penalized strongly by the RMSE. Already showing great potential for solar irradiance nowcasting, the limitations of data-driven approaches should be the focus of future research. Accurately anticipating the influence of individual clouds causing ramp events should be the main objective in upcoming investigations, requiring error metrics apart from RMSE, MAE, and FS.

Finally, the advantages of the DL approach presented in this study should not be interpreted as physics-based models being obsolete. The potential for spatial forecasts on larger areas has not been discussed here. A good example is the application of such physical models in a network[15] or in solar power plants covering large areas. Here, DL techniques could also further improve network forecasts. Future research should therefore concentrate on combining the strengths of both approaches further.

Appendix

A1 Deterministic error metrics by DNI variability classes

Figure A1 and **A2** show MAE and MBE for DL, SP, and hybrid models over LTs discretized by DNI variability classes. Qualitatively, the MAE curves look similar to the ones from RMSE. For classes 3–8, DL performs best also in MAE. In case of MBE, a significant bias is visible for the DL model for class 8. This is likely to cause the high RMSE (PICP, PI) in the probabilistic nowcasting.
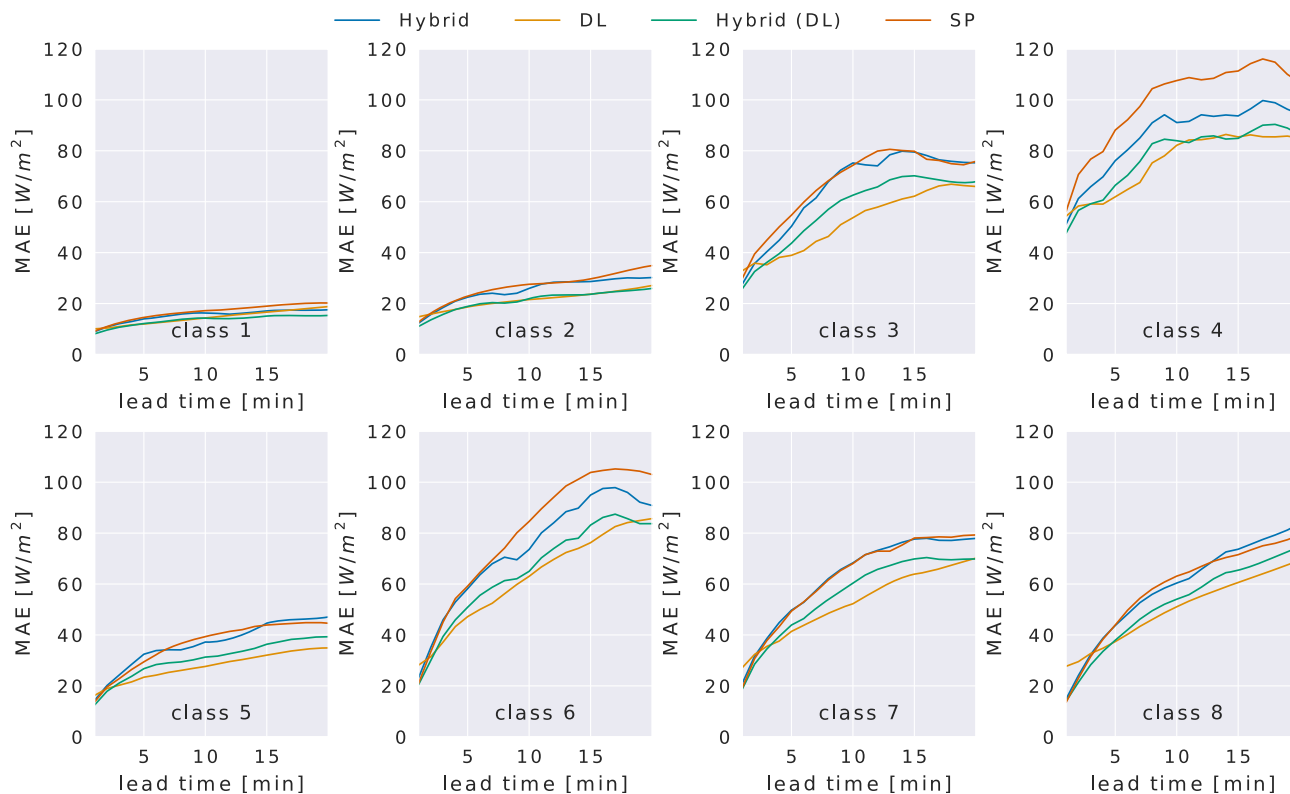
Figure A1. MAE on field average discretized by DNI variability classes. Upper row shows classes 1–4, lower row shows classes 5–8 (left to right).
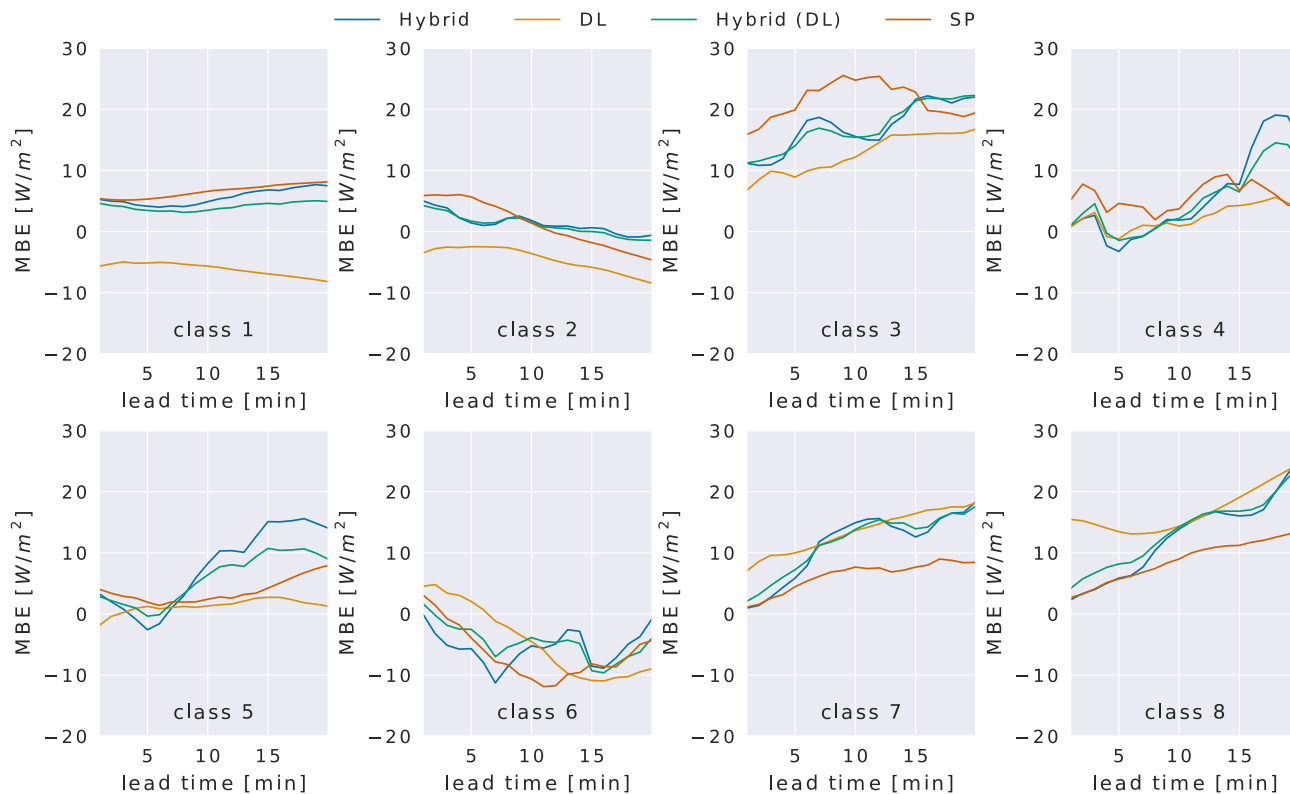
Figure A2. MBE on field average discretized by DNI variability classes. Upper row shows classes 1–4, lower row shows classes 5–8 (left to right).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Research data are not shared.

[1] R. Perez, K. R. Rábago, M. Trahan, L. Rawlings, B. Norris, T. Hoff, M. Putnam, M. Perez, *Energy Policy* **2016**, *96*, 27.

[2] M. Emmanuel, R. Rayudu, *Renewable Sustainable Energy Rev.* **2017**, *67*, 207.

[3] R. Perez, E. Lorenz, S. Pelland, M. Beauharnois, G. Van Knowe, K. Hemker, D. Heinemann, J. Remund, S. C. Müller, W. Traunmüller, G. Steinmaurer, D. Pozo, J. A. Ruiz-Arias, V. Lara-Fanego, L. Ramirez-Santigosa, M. Gaston-Romero, L. M. Pomares, *Sol. Energy* **2013**, *94*, 305.

[4] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, Z. Hu, *CSEE J. Power Energy Syst.* **2015**, *1*, 38.

[5] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. M. de Pison, F. Antonanzas-Torres, *Sol. Energy* **2016**, *136*, 78.

[6] B. Nouri, N. Blum, S. Wilbert, L. F. Zarzalejo, *Sol. RRL* **2022**, *6*, 5.

[7] J. O. Kamadinata, T. L. Ken, T. Suwa, *Renewable Energy* **2019**, *134*, 837.

[8] D. Yang, W. Wang, C. A. Gueymard, T. Hong, J. Kleissl, J. Huang, M. J. Perez, R. Perez, J. M. Bright, X. Xia, D. van der Meer, I. M. Peters, *Renewable Sustainable Energy Rev.* **2022**, *161*, 112348.

[9] D. Z. Yang, J. Kleissl, C. A. Gueymard, H. T. C. Pedro, C. F. M. Coimbra, *Sol. Energy* **2018**, *168*, 60.

[10] G. Zhang, D. Yang, G. Galanis, E. Androulakis, *Renewable Sustainable Energy Rev.* **2022**, *154*, 111768.

[11] P. Blanc, J. Remund, L. Vallance, in *Renewable Energy Forecasting, Woodhead Publishing Series in Energy* (Ed: G. Kariniotakis), Woodhead Publishing, Sawston, CA **2017**, pp. 179–198, ISBN 978-0-08-100504-0, https://www.sciencedirect.com/science/article/pii/B9780081005040000068 (accessed: December 2023)

[12] S. D. Miller, M. A. Rogers, J. M. Haynes, M. Sengupta, A. K. Heidinger, *Sol. Energy* **2018**, *168*, 102.

[13] Q. Paletta, G. Arbod, J. Lasenby, *Sol. Energy* **2021**, *224*, 855.

[14] H. R. Wen, Y. Du, X. Y. Chen, E. Lim, H. Q. Wen, L. Jiang, W. Xiang, *IEEE Trans. Ind. Inf.* **2021**, *17*, 1397.

[15] N. B. Blum, S. Wilbert, B. Nouri, J. Stührenberg, J. E. Lezaca Galeano, T. Schmidt, D. Heinemann, T. Vogt, A. Kazantzidis, R. Pitz-Paal, *Remote Sens.* **2022**, *14*, 22.

[16] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, B. Washom, *Sol. Energy* **2011**, *85*, 2881.

[17] P. Kumari, D. Toshniwal, *J. Cleaner Prod.* **2021**, *318*, 128566.

[18] Y. H. Chu, H. T. C. Pedro, C. F. M. Coimbra, *Sol. Energy* **2013**, *98*, 592.

[19] D. Z. Yang, *J. Renewable Sustainable Energy* **2019**, *11*, 2.

[20] Y. H. Chu, C. F. M. Coimbra, *Renewable Energy* **2017**, *101*, 526.

[21] D. W. van der Meer, J. Widen, J. Munkhammar, *Renewable Sustainable Energy Rev.* **2018**, *81*, 1484.

[22] B. Nouri, S. Wilbert, N. Blum, Y. Fabel, E. Lorenz, A. Hammer, T. Schmidt, L. F. Zarzalejo, R. Pitz-Paal, *Sol. Energy* **2023**, *253*, 285.

[23] M. Sengupta, A. Habte, S. Wilbert, C. Gueymard, J. Remund, NREL/TP-5D00-77635 **2021**, https://doi.org/10.2172/1778700.

[24] P. Lauret, M. David, H. T. C. Pedro, *Energies* **2017**, *10*, 10.

[25] S. A. Logothetis, V. Salamalikis, S. Wilbert, J. Remund, L. F. Zarzalejo, Y. Xie, B. Nouri, E. Ntavelis, J. Nou, N. Hendrikx, L. Visser, M. Sengupta, M. Po, R. Chauvin, S. Grieu, N. Blum, W. van Sark, A. Kazantzidis, *Renewable Energy* **2022**, *199*, 246.

[26] B. Nouri, S. Wilbert, N. Blum, P. Kuhn, T. Schmidt, Z. Yasser, T. Schmidt, L. F. Zarzalejo, F. M. Lopes, H. G. Silva, M. Schroedter-Homscheidt, A. Kazantzidis, C. Raeder, P. Blanc, R. Pitz-Paal, in *SolarPACES*, AIP Publishing, College Park, MD, October **2020**.

[27] B. Nouri, S. Wilbert, L. Segura, P. Kuhn, N. Hanrieder, A. Kazantzidis, T. Schmidt, L. Zarzalejo, P. Blanc, R. Pitz-Paal, *Sol. Energy* **2019**, *181*, 251.

[28] M. Schroedter-Homscheidt, M. Kosmale, S. Jung, J. Kleissl, *Meteorol. Z.* **2018**, *27*, 161.

[29] B. Nouri, S. Wilbert, P. Kuhn, N. Hanrieder, M. Schroedter-Homscheidt, A. Kazantzidis, L. Zarzalejo, P. Blanc, S. Kumar, N. Goswami, R. Shankar, R. Affolter, R. Pitz-Paal, *Remote Sens.* **2019**, *11*, 9.

[30] I. Reda, A. Andreas, *Sol. Energy* **2004**, *76*, 577.

[31] P. Ineichen, R. Perez, *Sol. Energy* **2002**, *73*, 151.

[32] Y. Fabel, B. Nouri, S. Wilbert, N. Blum, R. Triebel, M. Hasenbalg, P. Kuhn, L. F. Zarzalejo, R. Pitz-Paal, *Atmos. Meas. Tech.* **2022**, *15*, 797.

[33] B. Nouri, P. Kuhn, S. Wilbert, N. Hanrieder, C. Prahl, L. Zarzalejo, A. Kazantzidis, P. Blanc, R. Pitz-Paal, *Sol. Energy* **2019**, *177*, 213.

[34] B. Nouri, *Ph.D. Thesis*, RWTH Aachen University, **2020**.

[35] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, C. Eickhoff, in *Kdd '21: Proc. of the 27th ACM SIGKDD Conf. on Knowledge Discovery & Data Mining*, Association for Computing Machinery, New York, NY **2021**, pp. 2114–2124.

[36] I. Oguiza, *tsai - A State-of-the-Art Deep Learning Library for Time Series and Sequential Data*, Github **2022**, https://github.com/timeseriesAI/tsai (accessed: June 2023)

[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, Neural Information Processing Systems (NIPS), La Jolla, CA **2017**, p. 30.

[38] G. Bertasius, H. Wang, L. Torresani, in *Int. Conf. on Machine Learning*, Vol. *139*, **2021**, p. 139.

[39] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby (Preprint), arXiv:2010.11929, v2, submitted: Jun. 2020, https://ui.adsabs.harvard.edu/abs/2020arXiv201011929D.

[40] D. P. Kingma, J. Ba (Preprint), arXiv:1412.6980, v9, submitted: Jan. **2017**, https://arxiv.org/abs/1412.6980.

[41] I. Loshchilov, F. Hutter (Preprint), arXiv:1711.05101, v3, submitted: Jan. **2019**, https://arxiv.org/abs/1711.05101.

[42] L. N. Smith, N. Topin (Preprint), arXiv:1708.07120, v3, submitted: May **2018**, https://arxiv.org/abs/1708.07120.

[43] J. Howard, S. Gugger, *Information* **2020**, *11*, 2.

[44] R. Meyer, J. Torres Butron, G. Marquardt, M. Schwandt, N. Geuder, C. Hoyer-Klick, E. Lorenz, A. Hammer, H. G. Beyer, in *SolarPACES Symp.*, Vol. *2008*, Las Vegas, US March **2008**.

[45] J. L. G. L. Salle, M. David, P. Lauret, *Sol. Energy* **2021**, *223*, 398.

[46] P. Lauret, M. David, P. Pinson, *Sol. Energy* **2019**, *194*, 254.

[47] D. Yang, D. van der Meer, J. Munkhammar, *Sol. Energy* **2020**, *206*, 628.

[48] L. Vallance, B. Charbonnier, N. Paul, S. Dubost, P. Blanc, *Sol. Energy* **2017**, *150*, 408.