Predicting Photovoltaic Power Production using High-Uncertainty Weather Forecasts

Tomas Polasek[a,*] (Researcher), Martin Čadík[a] (Researcher)

[a] *Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic*

ARTICLE INFO

Keywords:
Solar Power Forecasting
Photovoltaic Dataset
Prediction Uncertainty
Machine Learning Model

ABSTRACT

A growing interest in renewable power increases its impact on the energy grid, posing significant challenges to reliability, stability, and planning. Weather-based prediction methods help relieve these issues. However, their real-world accuracy is limited by weather forecast errors. To help resolve this limitation, we introduce the SolarPredictor model. Publicly available weather forecasts are used to predict solar power production by a target photovoltaic power plant. To achieve high prediction accuracy, the model is trained on genuine weather forecasts, including errors and mispredictions. Further, we introduce the SolarDB dataset, comprising one year of power production data for 16 power plants. The dataset includes hourly weather forecasts with seven days of history, allowing our model to anticipate errors in the meteorological features. The prediction accuracy is evaluated on a wide range of weather forecast ages, accurately reflecting real-world performance. The SolarPredictor model is compared against 17 techniques, reaching an average RRMSE of 6.15 for 1-day, 8.54 for 3-day, and 8.89 for 7-day predictions on the SolarDB dataset. Finally, we analyze the effects of weather forecast uncertainty on prediction accuracy, showing there is at least a 23 % performance gap compared to using zero-error weather. Data and additional resources are available at cphoto.fit.vutbr.cz/solar.

1. Introduction

Renewable energy utilization grows every year. Its increasing importance is notable by the average yearly growth rate of almost 0.68 % within the European Union [1] and targets of reaching 32 % coverage by 2030. However, many renewable energy sources are inherently dependent on meteorological factors, leading to challenging problems with energy grid balancing and stability [2]. Precise predictions are highly desirable from the point of smart power grids [3], competitive pricing [4], and energy utilization [5]. Many countries also require production forecasts [6], with deviations leading to penalty charges [7]. Predictive systems are used to alleviate these issues, thus guaranteeing a degree of certainty and allowing fallback to other power sources when needed [8, 9]. Several such systems specifically target photovoltaic power plants, predicting their power output directly from the input weather data [10]. However, while the prediction accuracy of weather forecasting techniques is constantly improving, even short-term weather forecasts are often weighed down by significant errors [11]. As a result, using error-free weather in predictive models leads to unexpected performance degradation in real-world scenarios when realistic weather forecasts are used.

Aims and Objectives: There are three primary goals we focus on in this paper:

- Predicting power production of solar power plants using localized meteorological data. We target specific locations to model their local properties, focusing on real-world performance by using weather forecast data. We solve this task with a novel SolarPredictor architecture (Sec. 2.2), taking into consideration the error-prone nature of weather forecasts.
- Creating a suitable dataset for the training of forecast-based prediction models. We propose the SolarDB dataset (Sec. 2.1) covering one year of detailed data for 16 photovoltaic power plants. We make this dataset publicly available to other researchers, facilitating further research in this area.
- Analysing of what affects the prediction accuracy and ways of improving it further. We use the SolarDB dataset to evaluate a wide range of scenarios – including weather, season, location, and forecast age – in order to determine which factors are important.

We make the dataset freely available for research purposes. For additional data, resources, and source codes, see the supplementary materials and the project website at cphoto.fit.vutbr.cz/solar.

*Corresponding author
ipolasek@fit.vutbr.cz (T. Polasek); cadik@fit.vutbr.cz (M. Čadík)
cphoto.fit.vutbr.cz (T. Polasek); cphoto.fit.vutbr.cz (M. Čadík)
ORCID(s): 0000-0002-1441-0241 (T. Polasek); 0000-0001-7058-9912 (M. Čadík)

Nomenclature

Abbreviations

CD	Clear Day
CNN	Convolutional Neural Network
DNI	Direct Normal Irradiance
DNN	Deep Neural Network
GHI	Global Horizontal Irradiance
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
MLP	Multilayer Perceptron
NN	Neural Network
NWP	Numerical Weather Prediction
OC	Overcast Day
RF/RFor	Random Forest
RNN	Recurrent Neural Network
SeH	Multi-output Model with History
Seq	Multi-output Model
Sin	Single-output Model
SPred	Solar Predictor
SVM	Support Vector Machine
TCN	Temporal Convolutional Network
Trf	Transformer

Variables

Δ	Acceptance threshold
$f, \hat{f}$	Measurement frequency
P, P'	Production over interval
p_t	Production at time index t
t	Time of day
w_t	Weather features at time index t
x_t	Input features at time index t
$y_t/\hat{y}_t$	Ground-truth/prediction at time index t

Parameters

MSE	Mean Squared Error
PError	Power error
RMSE	Root Mean Square Error
RRMSE	Relative RMSE
R^2	Determination
WError	Weather error
cor_p	Pearson correlation coefficient
cor_s	Spearman correlation coefficient
$V_{x\text{th}}$	xth percentile of V
$Var_x[f]$	Variance of f with respect to x

1.1. Related Work

Review of Solar Prediction Models

The prediction task takes a set of input values – including endogenous (internal) and exogenous (external) parameters – and calculates a power production forecast for the target time frame. The accuracy is largely determined by the precision of the input parameters and the prediction time horizon [12] – i.e., short ($t \leq 1$ day), medium ($t \leq 14$ days), and long-term ($t > 14$ days). The prediction methods include physical, statistical, and machine learning models [13], their uses ranging from real-time scheduling [14] to long-term planning [15].

Physical models use the properties of the photovoltaic system without relying on historical data. The inputs include numerical weather predictions (NWP), monitoring data, and properties of the plant [16, 17], and output global horizontal irradiance (GHI) and direct normal irradiance (DNI), proportional to the plant's production [18]. Although physically correct, these methods rely on precise meteorological data. They generalize poorly when provided with imprecise weather data [19], resulting in unreliable predictions when using real-world weather forecasts. These approaches include ASHRAE [20] and Hottel equations [21].

Statistical methods consider historical data and attempt to automatically derive a pattern – mapping the input variables to the output power production [22]. Common approaches used with solar power include Markov Chains [23], fuzzy logic [24] or, auto-regressive models [25, 26] such as the NARX [12] or NARMAX [27]. Although generally less complex than the physical models, the historical data allows them to better model the specifics of a power plant [28]. However, to do so, they require a large amount of data for each target plant [25], resulting in difficulties when rapid growth is needed.

Machine learning models use a trained mapping function. They derive the mapping through a process of training on a dataset of input and output samples [29]. Each prediction model is suited to a different use case. The models include Decision Trees [30], Random Forests (RFor) [31], and their ensembles [32], which are well suited to represent the variability of different power plant sites. Similarly, Support Vector Machines (SVM) [33] automatically determine important input parameters and allow a degree of interpretability. Deep Networks (DNN) [34] are used sparingly since they provide a looser constraint on the time-series prediction data. Recurrent Networks (RNN) [35] model the temporal data, enhancing multi-step prediction stability [36]. Both Long Short-Term Memory (LSTM) [37] and Gated Recurrent Unit (GRU) [38] are used to further stabilize the training process. Conversely, Convolutional Networks (CNN) [39] are used to model the spatial aspect of the power production data [40] and perform localized associations [41].

Hybrid architectures combine both temporal and spatial constraints of the models [42]. The ConvLSTM [43] uses convolutional layers within the LSTM unit to provide translation invariance and a wider receptive field. The WaveNet [44] and Temporal Convolutional Network (TCN) [45] employ causal dilated convolutions to better model the temporal nature of the predictions. The Transformer [46] utilizes an attention mechanism, selectively choosing which time steps to use in the prediction. Ghimire *et al.* [39] and Zang *et al.* [47] use a CNN-LSTM hybrid for pattern recognition and time-series analysis to lower the data requirements. This is further enhanced when using attention to model the short-term and long-term temporal modes [48]. Each architecture has its respective merits and shortcomings. The RNNs are challenging to train due to vanishing gradients [49] and numerical instability [50]. CNNs can reach wide receptive fields but require deep architectures and large amounts of training data [41].

This study introduces a hybrid predictor motivated by the spatio-temporal nature of the power production signal. We name the model SolarPredictor. The model is designed to predict power production from low-reliability weather forecasts by adapting the UNet [51] to use residual connections [52]. The model performs a spatial analysis with 1D causal dilated convolutions [53], modeling the short and long-term relations. Then, the signal is passed through an aggregator to reconstruct the temporal nature of the prediction [44]. Details of this architecture are found in Sec. 2.2.

Review of Renewable Energy Datasets

A robust dataset is essential for the training of machine learning models, which require a large amount of data. Based on their coverage, solar power datasets can be divided into [54]: large-scale, planet-covering datasets with lower spatial resolution; regional, providing added detail and granularity, although at reduced scale; and site-specific, a set of locations, usually including power production values along with meteorological features.

The large-scale datasets include the Retrospective Analysis datasets – MERRA [55] and MERRA-2 [56] – focusing on climate data spanning the Earth from 1980. They provide hourly meteorological data with a spatial resolution of 50 km. The Derived Renewable Energy (EMHIRES) datasets comprise solar [57] and wind [58] power generation data. Both datasets cover Central Europe, providing 30 years of hourly data without weather features. Finally, the Solar Radiation Database (NSRDB) [59] is a solar-specific dataset with an hourly resolution for specific locations or a 4 km grid. The NSRDB contains solar irradiance information without direct power production data.

The OpenSolar [60] initiative provides tools for solar-specific regional datasets. The Solar Power Data (SPDIS) [61, 62] dataset consists of one year of 5-minute data covering 5,020 locations, including simulated power production and weather data based on the NSRDB [59]. The Solar Radiation Research Laboratory (SRRL) [63] dataset includes data from a single site dating from 1981, comprising 75 instruments at a 1–10-minute resolution. However, the site is located at 1829 m above sea level. The Microgen Database [64] dataset includes data from over 7,000 installations with only power generation data available. Lastly, the Dataport [65] dataset consists of production data of over 318 installations at a 1-minute resolution with no weather features.

The *Comprehensive* dataset [66] consists of 3 years of data for a single installation, providing production data at a 1-minute resolution along with seven weather variables. Data is sourced from on-site measurements and the Mesoscale Forecast System. However, the dataset does not include any weather forecast data. Finally, the DKA Solar [67] dataset includes 62 installations in central-northern Australia, consisting of production and weather features at 1-minute to 5-minute intervals, although no weather forecasts are provided.

Under optimal conditions, decades of historical data are needed for each power plant site to reach reliable predictions [54]. However, a dataset of this magnitude is currently not available and collecting years of data for each new installation is untenable. Moreover, to the best of our knowledge, all sufficiently sized datasets provide only zero-error weather, which is not representative of the real-world weather forecast used during inference.

Considering the outlined limitations, we present the SolarDB dataset (Sec. 2.1). It comprises one year of data for 16 solar power plants, providing power production at 5-minute intervals. Weather features include 12 meteorological variables available at an hourly resolution. Crucially, the dataset includes weather forecast features from commodity sources, providing seven days of historical forecasts with a 1-hour spacing between the data points.

ID	Freq.	Interval	Days	Power	Invs.
#1	5 min	Oct 2019	365	122.6 kWp	12
#2	5 min	May 2019	361	999.8 kWp	15
#3	5 min	May 2019	365	36.00 kWp	4
#4	5 min	May 2019	365	11.05 kWp	1
#5	5 min	May 2019	365	56.39 kWp	6
#6	5 min	May 2019	365	162.0 kWp	4
#7	5 min	Nov 2019	365	252.5 kWp	7
#8	5 min	May 2019	365	30.03 kWp	4
#9	5 min	Nov 2019	365	75.90 kWp	2
#10	15 min	May 2019	365	909.3 kWp	89
#11	10 min	Mar 2019	324	672.6 kWp	39
#12	5 min	Nov 2019	363	655.3 kWp	21
#13	5 min	Nov 2019	354	686.4 kWp	10
#14	10 min	Nov 2019	352	646.7 kWp	38
#15	5 min	Dec 2019	365	834.9 kWp	3
#16	5 min	Nov 2019	360	99.00 kWp	3

Table 1
(1-column) Power Plant Sites: A list of photovoltaic installations included in the SolarDB dataset. Their reporting frequencies, nominal power, and inverter counts are provided. Each site is represented by 1-year of data starting with the Interval. The number of days without missing values is specified; other days contain gaps not exceeding four samples.

2. Method

To resolve the conflicting requirements of precise power predictions on one side and utilization of error-prone weather forecasts on the other, we propose the SolarPredictor system. It uses a hybrid neural network model, combining multi-scale spatiotemporal analysis with residual connections. The input consists of meteorological features for the target time frame and near historical production data. In contrast to other contemporary methods, the prediction is based on realistic weather forecasts and limited historical data for each site. This results in a more practical prediction task, taking into account the inherent uncertainty present in weather forecasts [68] and reflecting its real-world performance. To train the SolarPredictor system, we introduce the SolarDB dataset covering production and weather data. Please see the supplementary materials and resources available at cphoto.fit.vutbr.cz/solar for further details and data.

2.1. The SolarDB Dataset

The dataset consists of 1 year of data from 16 photovoltaic installations. It includes power production, on-site weather forecasts, and additional meta information. The data spans 1.46 M power records, 16.3 M inverter power records, 139 K weather records, and 22.8 M forecast records. Through careful curation and pre-processing, it is further expanded into the overall 40 M records and made publicly available through the provided Python SolarDB API.

2.1.1. Data Acquisition

A total of 16 installations spanning 258 power inverters are included in the dataset, as detailed in Tab. 1. The data was collected over the interval of two years, after which the continuous intervals of the highest quality were selected on a per-site basis. The SolarDB dataset includes four types of data: power, weather, exogenous, and meta-data. For a full description, please see App. A.

Power data is provided separately for the complete power plant and each of its inverters. Both pure production (Power_{DC}) and power after inversion (Power_{AC}) are provided in 5-minute intervals. The values are collected from various reporting systems, resulting in a non-uniform frequency (Tab. 1, Freq.). They are unified to the 5-minute frequency using linear interpolation for the in-between values as detailed in Sec. 2.1.2. We also provide status data for each installation and inverter, aggregating the status and error codes from reporting systems into a simplified codebook detailed in the supplementary materials.

Weather data consists of measured weather and weather forecasts with 12 features each, provided at a 1-hour resolution. Weather records are annotated with their age in hours. Measured weather represents the actual weather situation, while the forecasts may contain errors based on their age. Seven days of past forecasts are provided with a 1-hour step, resulting in 1 measured weather and $7 \times 24 = 168$ forecasts for each hourly time-point. Meteorological forecasts are collected from the Dark Sky API [69].

2.1.2. Data Pre-processing

The data pre-processing consists of filtration, frequency unification, interpolation, and extrapolation. First, erroneous measurements and inconsistent data are removed, converting both power and weather features into standard units and ranges. For details, see App. A and the extended quantities table in the supplementary materials.

Next, the frequency of the data points is unified to the common 5-minute interval. This step simplifies downstream tasks and generalized predictions. The missing values are calculated through linear interpolation and marked with a special Interpolated flag. However, only values between two valid records are interpolated to ensure high quality.

Finally, missing Power and Weather data is completed through a sequence of interpolation and extrapolation steps. This ensures data availability for all time points. In contrast to the frequency unification, the weather features are kept at an hourly resolution to reduce the size of the dataset. Linear interpolation is applied to intervals of $1 - 2$ missing records. Any remaining missing time steps are then extrapolated by using the next closest day. Records created in this way are marked as Interpolated or Extrapolated, respectively.

2.1.3. Data Augmentation

The dataset is augmented with complementary Exogenous features, including temporal, solar, and clarity components. Temporal features focus on the periodic properties of the production. Importantly, each Weather record is provided with its Age implying its veracity. The solar features comprise simulated properties of the Sun – altitude, azimuth, and irradiance [18] – calculated with the PySolar library [70] using the precise position of the power plant.

The apparent clarity is expressed by the daily clarity feature, splitting the data into clear (CD) and overcast (OC) days. The division is made on a per-day basis and is consistent with the concept of clear-day models [18]. The clarity is defined through the clear-day likelihood L_{CD} as

$$P'(t) = \max_t \left[\, P(t) \,\right] \, \left(1 - (\, 2t - 1\,)^2 \,\right), \tag{1}$$

$$L_{OD}(D) = \mathrm{Var}\left[\frac{1}{\sqrt{\mathrm{Var}\left[\, P(t) \,\right]}} \, \left(\, P'(t) - P(t) \,\right) \right], \tag{2}$$

$$L_{CD}(D) = \max\left[0, \, \min\left[1, \, 1 - L_{OD}(D) \right] \right], \tag{3}$$

where $t \in [0, 1]$ is the time of day, P' is the estimated clear-day production, L_{OD} and L_{CD} are the likelihoods of the day being overcast and clear, respectively, and D is the target day. A day is marked as clear when $L_{CD} >= 0.5$. This results in a roughly equal split over the 16 power plants – 50.3 % of the days are marked as clear and 49.7 % as overcast. For qualitative comparison, see the random selection of clear and overcast production curves in App. A.

2.2. The SolarPredictor System

As visualized in Fig. 1, its inputs consist of historical power production and weather features for each target time step. To better reflect the weather uncertainty, forecasts aged $\langle 1 \ldots 168 \rangle$ are used, denoting measured weather with age 0. Predictions are performed on a per-day basis, with outputs representing power production for each 5-minute step.

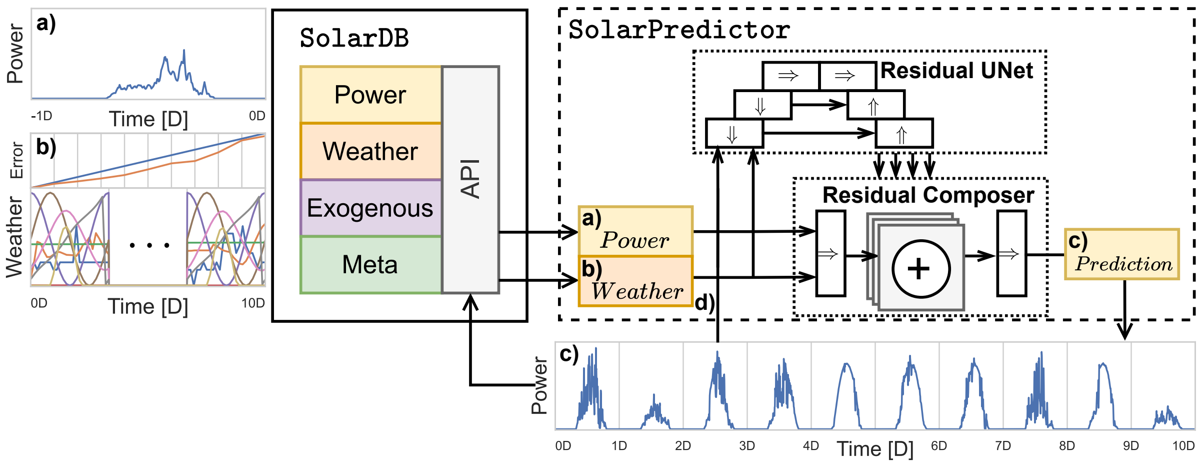

Figure 1: (Color, 2-column) System Overview: The SolarPredictor system is trained on the SolarDB dataset. The inputs consist of historical power **(a)** and weather features **(b)**. Power predictions are estimated in daily blocks **(c)**. The initial Power data **(a)** is historical, while consequent predictions recurrently use the previous outputs **(d)**. Notably, the predictions use weather forecasts with increasing age **(b, blue)**, which leads to progressively increasing weather error **(b, orange)**.

2.2.1. Predictor Architecture

The SolarPredictor model architecture is visualized in Fig. 2. It uses a fully convolutional architecture consisting of two blocks: *Residual UNet* and *Residual Composer*. The upper *Residual U-Net* performs a multi-scale analysis of the time series and produces the residual signals. While the lower *Residual Composer* constructs the prediction estimate by gradually adding up the residual signals. For parametrization of the architecture, see the supplementary materials.

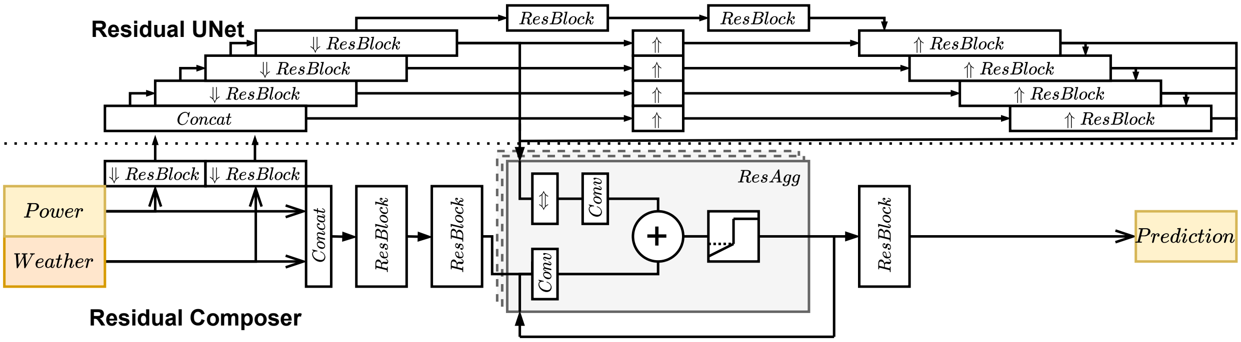

Figure 2: (Color, 2-column) SolarPredictor Model: The model combines a *Residual UNet* for time-series analysis (**top**) with a *Residual Composer* (**bottom**). The inputs consist of weather features and historical power data, producing the power prediction estimate. ⇓, ⇑ represent the down/up-sampling operations, while ⇕ is the conditional re-sampling.

Residual UNet is inspired by the UNet [51] and ResNet [52] architectures (Fig. 2, top). It contains a cascade of $ResBlock$ units with 1d causal convolutions [53] divided into four stages. First, both Power and Weather data are pre-processed and concatenated. Next, the contractive stage on the left performs a down-sampling (⇓ $ResBlock$) to extract spatio-temporal information, which passes through the bottleneck stage. Finally, the expansive stage performs the multi-scale analysis (⇑ $ResBlock$), combining previous outputs with skip connections from the contractive stage.
Residual Composer is inspired by the residual modules in WaveNET [44], aggregating the individual residual signals into the estimated power prediction (Fig. 2, bottom). A pre-processing phase takes both the Power and Weather features and reshapes them into the required output dimensions. Next, a series of residual aggregation modules ($ResAgg$) add the residual signals from the upper part of the model to the ongoing signal. The adaptive up-down sampling module (⇕) is used to unify the dimensions between the residual tiers. Finally, the estimated power production is produced by stacking the aggregation modules in a sequence, gradually building the output signal.

2.2.2. Model Training

The proposed SolarPredictor architecture is further enhanced with data augmentations to improve predictor accuracy and training stability. A total of six techniques are used, and their efficacy is further explored in Sec. 3.
Filtering & Scaling: Outlier analysis and data scaling is performed to prevent anomalies and stabilize the model training. The outliers are detected using absolute z-scores calculated over the base Power and Weather features. Samples with a z-score above an empirical threshold of $\gamma = 20$ are considered anomalous. These are consequently removed (Power) or set to zero (Weather). Each Power and Weather feature is then separately processed with a Positive Robust scaler [71], defined as

$$\hat{v} = \frac{v - \mu_V}{V_{75^{\text{th}}} - V_{25^{\text{th}}}} - \frac{\min[V] - \mu_V}{V_{75^{\text{th}}} - V_{25^{\text{th}}}}, \tag{4}$$

where $\hat{v}$ is the scaled output, $v \in V$ is the input, $V = \{v_1, v_2 \dots\}$ is the set of all inputs, μ_V is the mean value, and $V_{25^{\text{th}}}, V_{75^{\text{th}}}$ are the 25th and 75th percentiles of V.
Feature augmentation: Highly correlated features (Sec. 3.1) are removed to reduce the model size and accelerate the training procedure. Inverted variants of the Cloud Cover and Visibility features are calculated and scaled using the Positive Robust scaler (Eq. 4). Finally, cyclic time features are generated, facilitating phase detection. This approach is similar to the positional encoding in Transformer models [46]. We generate two features containing the sine and cosine transforms of selected temporal features. For details, please see App. B.

Weather Sampling: Training the model with only zero-error weather leads the model to implicit trust in the input weather, resulting in a loss of accuracy in real-world scenarios. The cause of these errors is connected to the correlation between the forecast age and prediction error (Sec. 3.7). However, directly using the weather forecasts from the beginning of the training leads to error accumulation during the recurrent predictions. Therefore, we train the model on both the measured and forecast weather features, utilizing multiple forecast ages.

Three types of sampling schemes (App. B) are used to construct the training data. The *Measured* weather samples use the zero-error weather features for each target time point. The *Forecast* samples utilize weather forecasts delayed by a constant offset d. Since downstream applications receive fresh forecasts at a set frequency, the training uses multiple delays $d \in \{ 1D, \cdots, 7D \}$. These delays allow the model to learn the relationship between the age of the weather and its veracity. Lastly, the *Realistic* samples simulate the real-world scenario by using the latest available weather forecast at the time of prediction. Thus, the age of the features increases linearly, as does its error.

Sample Weighing: The input data contains imbalanced classes with under-represented weather conditions. A sample weighing scheme is used to account for their frequency. It consists of two parts: clear-day weights wg_{cd} and temporal weights wg_t. The clear-day weights wg_{cd} balance the data based on day clarity, based on Eq. 2. The value of wg_{cd} is minimized for clear days while being at its maximum for overcast days. Conversely, the temporal weight wg_t is designed to favor samples based on their relative frequency. Finally, the total sample weight $wg = wg_{cd} \cdot wg_t$ is calculated for each sample in the training set and used as a loss multiplier. For detailed calculation, please see App. B.

Importance Pruning: Many of the samples are assigned with small weights wg, resulting in a low contribution to the overall training. However, completely removing them results in a degradation in prediction performance, as evaluated later in Sec. 3.5. Thus, we use stochastic importance pruning instead of simply removing the samples.

Three acceptance probabilities govern this procedure: δ_d, δ_h, δ_s, representing day, hour, and smooth (5-minute) samples, respectively. Depending on each sample's prediction window, it is assigned with a threshold $\Delta := \delta_d \mid \delta_h \mid \delta_s$. Then, a uniform random variable $A \sim \mathcal{U}(0, 1)$ is used to determine whether a sample is accepted ($A \leq \Delta$) or rejected. By choosing $\delta_d = 1.0$, $\delta_h = 0.5$, and $\delta_s = 0.01$, we see a 95 % reduction in data without noticeable degradation in performance. The stochastic nature of the pruning also provides regularization and improves training (Sec. 3.5).

Training Procedure: The SolarPredictor model is trained on the SolarDB dataset. The ADAM [72] optimizer is used with the AMSGrad [73] modification, utilizing the reduce-on-plateau technique. Through empirical observation, the initial learning rate is set to $lr = 0.005$ while setting $\beta_1 = 0.9$ and $\beta_2 = 0.999$ – the values recommended by the authors [73]. The training proceeds with a mini-batch size of 256, shuffling samples at the end of each epoch. Further, Smooth Loss $\mathcal{L}$ is used as the training objective, combining the properties of Mean Squared Error $\mathcal{L}_{\mathrm{M}}$, Huber Loss $\mathcal{L}_{\mathrm{H}}$, and Differential Smooth Loss $\mathcal{L}_{\Delta}$. For details, see App. B. Lastly, Gaussian noise ($\mu = 0$ and $\sigma = 0.0005$) is injected into the power history data, improving the training stability and generalization of the model.

3. Results and Discussion

The results of the proposed system are presented in this section. First, a deeper look at the SolarDB dataset is provided, focusing on its statistical properties. Second, the accuracy of the SolarPredictor model is evaluated. To better convey its properties, we compare it against other predictors from the point of accuracy, performance, and size. Next, the proposed model improvements are gauged in an ablation study and further studied in a cross-validation experiment. Finally, the relevance of the presented results is discussed, along with limitations and potential uses.

3.1. Dataset Overview

An overview of the SolarDB dataset can be found in Fig. 3. The 16 power plants show the expected sinusoidal tendency consistent with the Earth's temperate zone. Notably, the similar precipitation and temperatures throughout the year are caused by the physical proximity of the power plants, facilitating experiments with their distance.

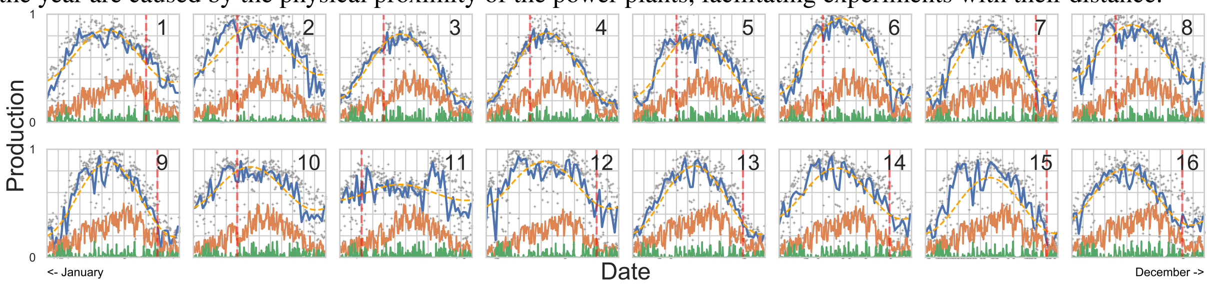

Figure 3: (Color, 2-column) SolarDB Overview: Each graph displays a year of data for one of the 16 power plants from January (**left**) to December (**right**). The weekly power production is marked in **blue**; **grey** dots represent daily values, along with temperature (**orange**) and precipitation (**green**). The data is aligned, starting date marked by the **red** vertical line.

3.2. Feature Analysis

There is a wide range of meteorological features contained in the SolarDB dataset (App. A). However, their importance is uncertain. To determine which features are essential in the prediction task, we calculate the Pearson correlation [74] between the weather features and the power production. The resulting matrix is visualized in Fig. 4. It exhibits a strong correlation between power, temperature, and sun-related features. Further, the TimeX feature corresponds with the sinusoidal tendency of the production, which usually peaks around noon. As expected, during clear days (orange bars), the correlation is higher since the pattern is not disrupted by the weather conditions. Comparatively, the overcast days (blue bars) have a higher relative correlation with the weather features and are less dependent on the time. Finally, shifted power productions (PwrHist) are highly correlated for both clear and overcast days, indicating that historical production data may lead to improved accuracy. This effect is further analyzed in Sec. 3.4.

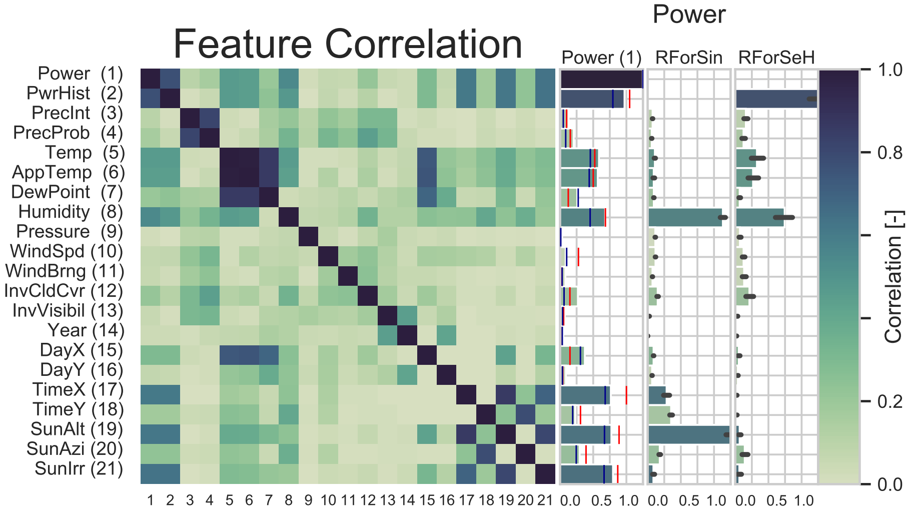

Figure 4: (Color, 1/1.5-column) Feature Analysis: The correlation matrix (**left**) shows Pearson correlation [74] for weather features. Power importance (**right**) details how vital these features are to the prediction. The **Power (1)** column contains correlations of the Power to all other features – filled bars show average results for all days, while the vertical bars represent the Clear (**orange**) and Overcast (**blue**) days. The following columns contain feature analysis using single (**RForSin**) and multi-value (**RForSeH**) Random Forest prediction models (Sec. 3.4) using normalized Gini Importance [75] as the metric.

To corroborate our findings, deeper analysis is performed by utilizing a single-value (RForSin) and multi-value (RForSeH) Random Forest models, each configured as detailed in Sec. 3.4. The RForSin receives weather features for a single time point, while the RForSeH is provided with weather features for the entire day along with historical power production. To increase the significance of the results, 30 cross-validation runs are performed and averaged. The importance analysis is based on the normalized Gini Importance [75]; details can be found in App. C.

The results for both RForSin and RForSeH models (Fig. 4) show the mean (I_f, solid bar) and the variance (σ_f^2, tail) values. Both temperature and humidity are strongly utilized, yet precipitation and cloud cover features are less critical. Notably, the single-value predictor (RForSin) mostly depends on the humidity, time, and sun features – possibly for the day-cycle phase detection. Conversely, the multi-value predictor (RForSeH) replaces the temporal features with power history, suggesting that recurrent models may utilize past values for phase detection.

3.3. Evaluation Procedure

The accuracy of the proposed model is evaluated using the task of power prediction. Unless otherwise specified, the inputs consist of the weather forecasts. Thus, the age of the weather features increases monotonically (Fig. 1, b), resulting in progressively increasing forecast error, as it would in real-use scenarios. Further, in multi-day predictions, the predicted power is used recurrently – supplying the outputs of the previous prediction step as an input to the next step. This approach better approximates the real-world use case and shows any potential error accumulation issues.

Several quantitative metrics are used to gauge the accuracy of the models. For weather and power data, the errors are calculated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Relative RMSE (RRMSE), and Coefficient of Determination (R^2). Hourly aggregates of weather (WError) and power (PError) errors are also used to gauge the overall performance. For implementation details, please see App. C.

A cross-validation protocol is used in all experiments. The cross-validation procedure ensures that the results are representative of various seasons and weather conditions. Each individual experiment consists of 12 separate runs. In each run, the one year of data is split into training and test sets of 11 and 1 month, respectively. The test set is rotated so that each month is selected exactly once within the 12 runs. The prediction results for the 12 test sets are then concatenated, making up one full year of predictions, which are then used in the evaluation.

Prediction models are implemented using the Python programming language, utilizing the Tensorflow [76] framework, SciPy [77], and SciKit-Learn [71] packages. Training and inference measurements are performed on a work node equipped with an Intel Xeon 12-core @ 2.2GHz CPU, 96GiB DDR4 @ 2666GHz memory, and an NVidia GeForce RTX 2080 Ti GPU. Specific model and training parameters can be found in the supplementary materials.

3.4. Model Comparison

A performance baseline consisting of 24 prediction models are tested on the SolarDB dataset. The results are presented in Tab. 2. Where applicable, alternative input and output vectors are also tested, resulting in single-value (Sin), multi-value (Seq), and multi-value with history (SeH) variants. To ensure a fair comparison, all of the models utilize the same set of augmentations and weather sources.

The evaluation is performed on power plant #8, chosen as the overall average within the SolarDB dataset. Each model is first trained using the cross-validation protocol. Then, predictions for 1 to 10 days ahead are calculated for each of the 12 testing months. The months are concatenated into a single continuous array, making up a full year of predictions. Next, the predicted data is split into Clear and Overcast subsets (Sec. 2.1.3). This allows separate evaluation in each category (Clear, Overcast) as well as the overall performance (All). Finally, the performance metrics are calculated for each of the 1 to 10-day sets and then averaged into a single aggregate value.

					Clear Days		Overcast Days		All Days		
	Model	Power	Weather	Out	RRMSE	PError	RRMSE	PError	RMSE	RRMSE	PError
Classical	SVMSin	N/A	1:5	1:5	19.09	70.361	42.17	83.056	7031	29.66	78.882
	TreeSin	N/A	1:5	1:5	9.71	29.860	31.42	59.199	4446	21.99	49.464
	TreeSeq	N/A	288:5	288:5	9.39	26.909	24.19	55.075	4028	18.05	45.717
	TreeSeH	288:5	288:5	288:5	9.37	**21.992**	24.31	51.305	3975	17.99	41.479
	RForSin	N/A	1:5	1:5	8.56	22.955	33.65	62.592	3973	21.80	49.384
	RForSeq	N/A	288:5	288:5	7.47	26.175	25.92	53.353	3288	17.46	44.248
	RForSeH	288:5	288:5	288:5	**7.42**	24.373	**19.30**	**50.026**	**3050**	**13.82**	**41.311**
DNN	DNNSin	N/A	1:5	1:5	17.09	42.106	30.29	60.910	5607	23.69	54.403
	DNNSeq	N/A	288:5	288:5	12.62	25.500	28.10	70.295	4758	21.18	55.558
	DNNSeH	288:5	288:5	288:5	**11.36**	**24.291**	**19.19**	**52.701**	**3894**	**15.52**	**43.169**
Recurrent	LSTMVan	288:5	288:5	288:5	17.02	40.179	31.31	61.396	5658	24.39	54.240
	LSTMSta	288:5	288:5	288:5	17.08	40.796	31.03	61.867	5657	24.20	54.716
	LSTMCNN	288:5	288:5	288:5	16.97	36.410	24.78	57.747	5416	20.86	50.532
	LSTMCon	288:5	288:5	288:5	17.38	36.264	34.56	63.265	5920	26.35	54.187
	LSTMBid	288:5	288:5	288:5	10.03	19.818	31.69	66.416	4340	22.07	51.051
	SPredSeRc	36:40	4:10	8:5	16.34	47.334	26.68	**56.843**	5399	21.89	53.636
	SPredPaRc	36:40	36:5	36:5	**10.24**	**16.051**	**20.86**	60.477	**3710**	**16.00**	**45.508**
Convolutional	SPredCNN	288:5	72:20	288:5	9.18	38.394	15.85	46.770	3311	13.11	43.820
	SPredCNNO	288:5	288:5	288:5	8.06	30.601	20.25	58.196	3507	15.41	49.032
	SPredUNet	288:5	288:5	288:5	7.85	28.160	20.64	51.936	3376	15.47	44.066
	SPredTCN	288:5	288:5	288:5	7.46	28.643	19.03	51.032	3173	14.38	43.542
	SPredWave	288:5	288:5	288:5	7.89	25.623	24.85	55.232	3439	17.23	45.317
	SPredTrf	288:5	288:5	288:5	9.50	27.625	18.29	53.575	3573	14.69	44.997
	SPred	288:5	288:5	288:5	**5.68**	**16.886**	**16.16**	**44.351**	**2533**	**11.78**	**35.097**
Enh	RForEnh	288:5	288:5	288:5	6.90	19.838	17.34	44.982	2775	12.46	36.468
	SPredEnh	288:5	288:5	288:5	**4.50**	**15.308**	**11.29**	**32.480**	**1897**	**8.45**	**26.304**

Table 2
(2-column) Prediction Models: The experiments are split by model type into Classical, DNN, RNN, CNN, and Enhanced categories. The sizing of the input (Power, Weather) and output (Out) vectors are specified as *count* : *stride*, where the *count* is the number of values provided and *stride* the time step in minutes – e.g., 288:5 covers 1440 minutes, i.e., one day. The performance is evaluated separately for Clear and Overcast days, with All Days covering both categories.

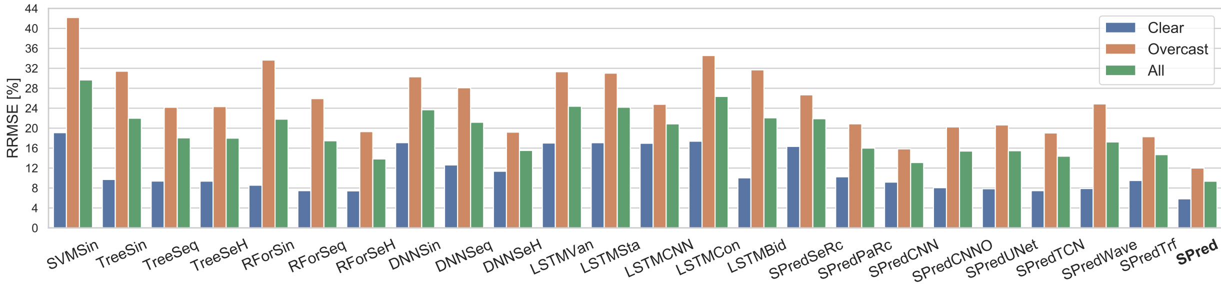

Figure 5: (Color, 2-column) Model Comparison: Evaluation of model prediction accuracy. The models are evaluated on their performance during **Clear**, **Overcast**, and **All** days. Notably, the clear days are easier to predict, leading to a lower error, while the overcast days are more challenging due to higher forecast uncertainty.

The result of the experiments is visualized in Fig. 5. The best-performing model among the Classical approaches is the Random Forest. Notably, expanding the model's range of inputs from a single value (Sin) to a day (Seq) and adding historical data (SeH) leads to an overall improvement of 36.61 %. This result indicates that a broader context is critical to prediction accuracy. A similar tendency is also seen in the DNN models, further confirming this theory.

Surprisingly, most of the LSTM [49] Recurrent models perform poorly. There is notable error accumulation in the longer prediction time-frames. Experiments with GRU [78] units lead to similar results. Although they achieve comparable Clear Day performance, they fail to predict the Overcast Days. Hybrid models combining LSTMs phase detection with DNN signal processing are an attempt to balance the results – serial (SPredSeRc) and parallel (SPredPaRc) configurations. However, both variants are still worse compared to the Random Forest models.

Purely convolutional architecture (SPredCNNO) and its combination with dense layers (SPredCNN) achieve RRMSE $\approx$ 15.41 and 13.11, respectively, with a significant accuracy trade-off between clear and overcast days. Further, we experiment on four backbone architectures based on the UNet (SPredUNet) [51], Temporal Convolutional Network (SPredTCN) [53], WaveNet (SPredWave) [44], and the Transformer (SPredTrf) [46]. Although these models show a minor improvement (2.11 % - 7.44 %) to the Clear Day accuracy of SPredCNNO, they are generally (9.69 % - 31.42 %) worse when compared to the SPredCNN.

The SolarPredictor model (SPred) utilizes the architecture proposed in Sec. 2.2.1. It achieves an average 14.76 % improvement in All Day accuracy over the second-best Random Forest model (RForSeH), gaining 23.45 % on clear days and 16.27 % on overcast days. The performance improvement is even more notable for the enhanced version of the model (SPredEnh), which is further analyzed in Sec. 3.5.

Finally, we also compare the models from the point of time and space complexity, visualized in Fig. 6. The results show that the Random Forest model requires 340× more memory and is 2.9× slower than the SolarPredictor model. Please see the paper supplement for detailed additional comparison data.

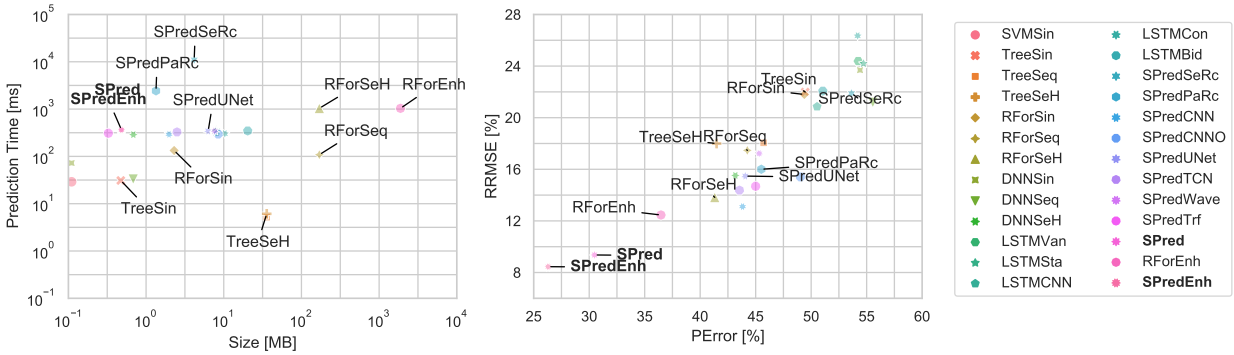

Figure 6: (Color, 2-column) Model Properties: (**left**) Comparison of time and space complexity of the prediction models. The size includes all model parameters necessary for the prediction. The prediction time is the total time taken to predict seven days of data. (**right**) Comparison of model prediction accuracy displaying the power error (PError) and RRMSE.

3.5. Ablation Study

To better demonstrate the effect of the augmentations, we present an extensive ablation study covering the enhancements proposed in Sec. 2.2.2. The experiments point wto their general applicability, improving the prediction accuracy for both the SolarPredictor (SPred) and the second-best Random Forest model (RForSeH). For additional experiments and complete data, please see the supplementary materials.

									Clear Days		Overcast Days		All Days		
	Model	Scl	Sel	Aug	Smpl	Meas	Fore	Real	RRMSE	PError	RRMSE	PError	RMSE	RRMSE	PError
	SPredBaseline	✗	✗	✗	✗	✓	✗	✗	9.13	20.265	32.33	74.819	4005	21.30	56.980
Feature	SPredScaler	✓	✗	✗	✗	✓	✗	✗	6.63	20.397	17.13	45.412	2755	12.57	37.036
	SPredOutliers	✓	O	✗	✗	✓	✗	✗	6.75	18.906	17.20	45.919	2756	12.75	36.977
	SPredWeights	✓	O+W	✗	✗	✓	✗	✗	6.59	18.945	17.22	47.076	2749	12.71	37.662
	SPredCyclicAge	✓	O+W	C	✗	✓	✗	✗	**5.62**	18.099	16.98	**43.323**	**2496**	**11.79**	**34.857**
	SPredAugmented	✓	O+W	C+F	✗	✓	✗	✗	**5.68**	**16.886**	16.16	**44.351**	**2533**	**11.78**	**35.097**
Forecast	SPredMeasured	✓	O+W	C+F	Day	✓	✗	✗	16.63	55.805	26.92	58.248	5503	22.56	57.451
	SPredAllData	✓	O+W	C+F	All	✓	✗	✗	5.54	17.911	15.63	45.557	2377	11.24	36.322
	SPredForecast	✓	O+W	C+F	All	✓	0D	✗	6.26	21.286	13.76	41.379	2418	10.45	34.523
	SPredMultiple	✓	O+W	C+F	All	✓	7D	✗	**4.65**	**12.878**	15.63	42.795	2274	10.98	32.785
	SPredRealistic	✓	O+W	C+F	All	✓	7D	7D	**4.50**	15.308	**11.29**	**32.480**	**1897**	**8.45**	**26.304**

Table 3
(2-column) Ablation Study: Experimental results dealing with the efficacy of the proposed augmentations. The baseline model (**top**) is first (**middle**) enhanced with scaling (Scl), sample selection (Sel: **O**utliers, **W**eights), and augmentation (Aug: **C**yclic, **F**eature selection). This improves Clear Day performance. Next, training data augmentation is introduced (**bottom**) with sampling (Smpl: single per Day, All), measured weather (Meas), forecasts of up to a given age (Fore), and realistic samples (Real). Using the realistic sampling scheme improves both Clear and Overcast days.

The results of the ablation study are presented in Tab. 3. The starting model (SPredBaseline) is gradually modified with feature and forecast augmentations. Note that the baseline already predicts Clear Day accurately with RRMSE $\approx$ 9.13, while the accuracy for overcast days is only 32.33. The overall prediction performance is improved by over 44 % by gradually adding scaling, outlier removal, sample weights, cyclic, and solar features. To further improve the performance, we consider the utilization of weather forecasts during training. Notably, the augmented model limited to using only a single training vector per day (SPredMeasured) leads to degraded prediction performance. Adding intraday sampling (SPredAllData) reduces the training data requirements to only 5.23 % of those used by the augmented model while improving the performance by 4.58 %. Up until this point, the model was trained only on the measured weather without seeing any inaccurate weather forecasts. However, simply training the SolarPredictor on the forecasts results in lower reliance on weather features, decreasing the Clear day accuracy.

				Clear Days		Overcast Days		All Days		
Model	Meas	Fore	Real	RRMSE	PError	RRMSE	PError	RMSE	RRMSE	PError
SPredAugmented	✓	✗	✗	5.68	16.886	16.16	44.351	2533	11.78	35.097
SPredW1D	✓	1D	✗	5.41	18.022	14.84	41.780	2321	10.77	33.776
SPredW4D	✓	4D	✗	5.42	18.216	15.13	40.635	2307	10.92	33.112
SPredW5D	✓	5D	✗	5.25	17.278	15.65	42.922	2324	11.14	34.306
SPredRealistic	✓	7D	✗	**4.65**	**12.878**	15.63	42.795	2274	10.98	32.785
SPredW7R	✓	7D	7D	**4.50**	15.308	**11.29**	**32.480**	**1897**	**8.45**	**26.304**

Table 4
(2-column) Weather Ablation: Experiments with training on weather forecast data, using the SPredAugmented (Tab. 3). Starting with measured weather (Meas), additional forecast ages (Fore) are added in the **middle** section. Diminishing effects are notable at around five days of data. The deterioration is corrected by using the realistic sampling (Sec. 2.2.2).

The study presented in Tab. 4 shows the effect of training the model on different forecast ages and quantities. They confirm that multiple forecast ages improve the overall accuracy. However, diminishing returns are seen at five days of data. This problem is solved by using the realistic sampling (Sec. 2.2.2) – training the model on linearly aging forecast data. The final model reaches improvements of 3.23 % and 27.77 % for Clear and Overcast days, respectively. This disparity indicates that the sampling allows the model to better gauge the forecast uncertainty. The importance of training on forecast data is also visible when comparing it against the original SPredAugmented model. Its overall results are improved by 28.27 %. Finally, the experiment concerning vector sizing, presented in the supplementary materials, shows that providing the model with historical data is highly beneficial, leading to a 15.86 % improvement.

3.6. Cross-Validation Results

The SolarPredictor model, along with the proposed augmentation, was shown to improve the prediction accuracy. However, the generalization of this approach is uncertain. Thus, we perform a cross-validation experiment spanning all 16 power plants from the SolarDB dataset to confirm the general validity of these conclusions. The two best-performing models were chosen – RForEnh and SPredEnh (Tab. 2) – both with the same set of proposed augmentations, improving their accuracy from the base models by 9.84 % and 28.27 %, respectively.

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	All Power Plants	
Model	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	RRM	PError
RForEnh	13.78	12.39	10.31	7.90	10.90	14.62	18.08	12.46	17.24	18.42	17.01	17.96	9.51	20.03	18.19	14.07	14.55	37.864
SPredEnh	**7.44**	**8.29**	**6.10**	**4.80**	**6.70**	**8.33**	**8.15**	**8.45**	**9.18**	**8.77**	**8.47**	**9.50**	**6.38**	**8.45**	**9.24**	**9.30**	**7.97**	**25.350**

Table 5
(2-column) Cross-Validation Experiments: Evaluation of the two best-performing models – RForEnh and SPredEnh – for each of the 16 power plants. Each column represents a single power plant with the prediction RRMSE (RRM). The Final three columns show the overall results, calculated as an average over all power plants.

The results of this experiment are presented in Tab. 5. The SolarPredictor model (SPredEnh) provides consistently higher accuracy than the second-best model (RForEnh), ranging from 32.18 % to 57.81 %. On average, the SolarPredictor model is 43.73 % more accurate. A detailed overview of the performance is visualized in Fig. 7. As expected, the Clear days are overall easier to predict (**D**). Interestingly, the relative difficulty of Clear and Overcast days is different for each power plant, which is possibly a combination of site-specific environmental and meteorological factors. Finally, a calendar view of the predictions is provided in Fig. 8. The results for all 16 power plants are overlaid, averaging their daily metrics. Overall, the calendar for RForEnh (**A**) is darker in hue, corresponding to its lower prediction accuracy.

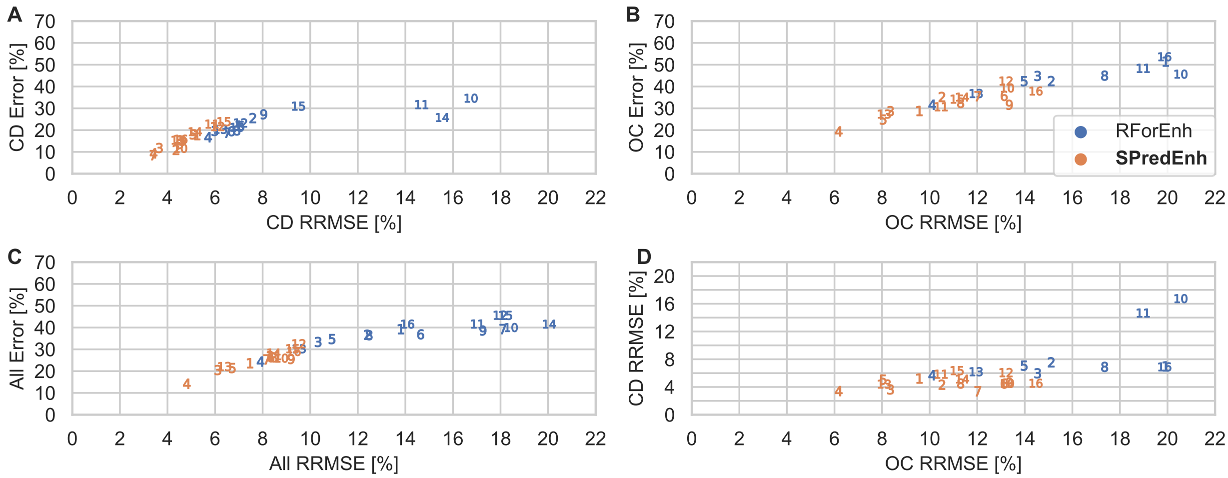

Figure 7: (Color, 2-column) Power Plant Results: The prediction accuracy of the two best-performing models – RForEnh (**blue**) and SPredEnh (**orange**) – spanning all 16 power plants. The graphs contain results for Clear (**A**), Overcast (**B**), and all (**C**) days. Finally, a correlation between the accuracy of clear (CD) and overcast (OC) days is shown in **D**. Notably, the results show that some power plants are easier to predict (e.g., #1 or #4) when compared to others (e.g., #12 or #14).

3.7. Prediction Difficulty

This section considers the cross-validation data to better understand which factors determine the difficulty of prediction. Essentially, the goal is to analyze the results detailed in Sec. 3.4 and 3.6 from the point of conditions leading to inferior prediction performance. This experiment allows deeper understanding of various modes of failure the prediction model exhibits and sets an expectation for the feasibility of future prediction accuracy.

Weather Accuracy Aspects contributing to the prediction difficulty are evaluated in Fig. 9. First is the length of the prediction horizon **(A)**. As expected, the further into the future a model predicts, the lower its average accuracy. The cause of this drop-off is probably caused by the increasing uncertainty in weather forecasts (bottom inset) since all types of models are affected. This hypothesis is further supported by Fig. 9 **(B)**, correlating the increased weather error with the lower prediction accuracy.

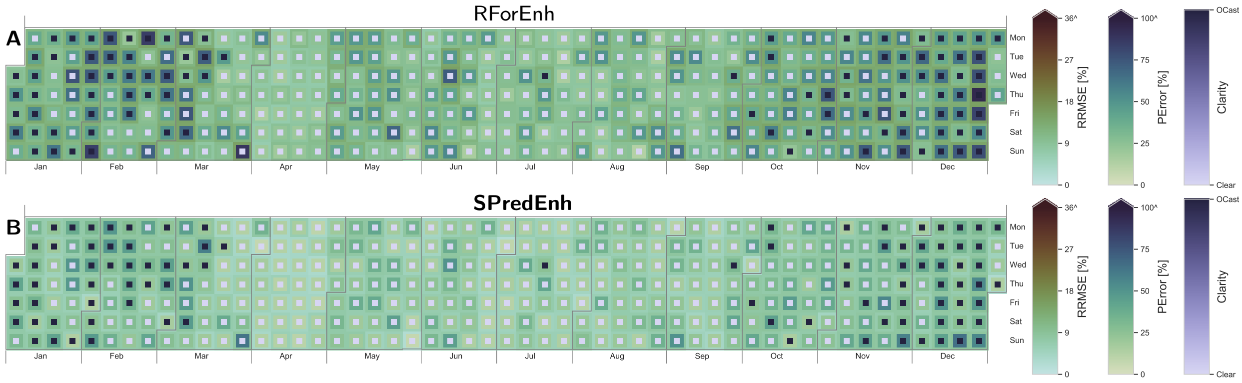

Figure 8: (Color, 2-column) Cross-Validation Calendar: Evaluation of predictions spanning the 16 available power plants for the RForEnh (**A**) and SPredEnh (**B**) models. Individual cells represent aggregate daily metrics, which are calculated as a mean over all power plants. Months are bracketed by vertical lines, while rows represent days of the week. Each cell conveys three types of information, from outside to inside: RRMSE, prediction error, and clarity – light/dark for clear/overcast days. Notably, the overcast days are more challenging to predict, corresponding with lower accuracy.

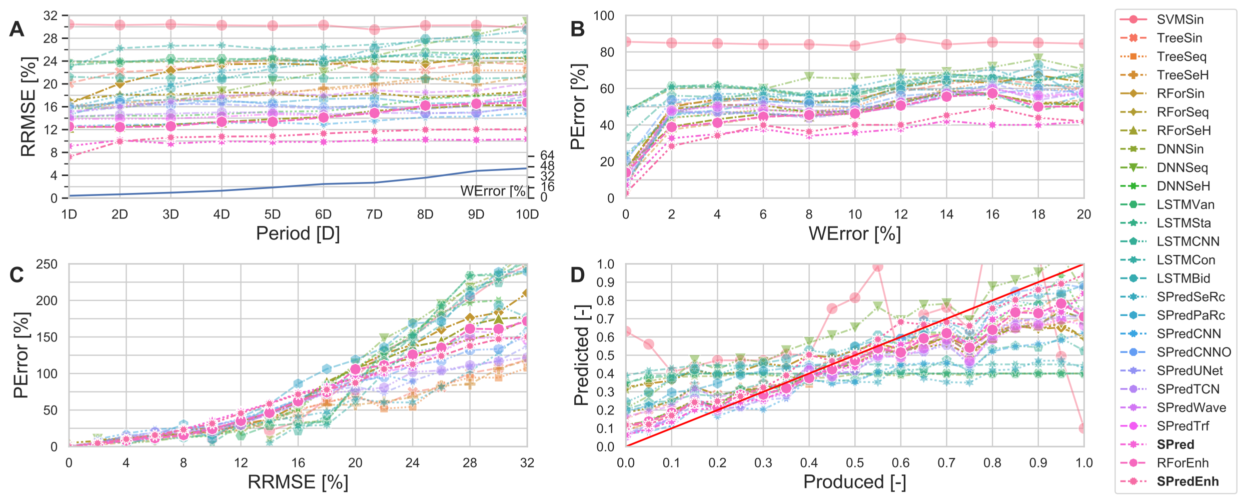

Figure 9: (Color, 2-column) Prediction Difficulty: The effect of various factors on the prediction accuracy for models from Tab. 2. **(A)** examines the relationship between prediction length and accuracy – the bottom inset displays the growing weather forecast error. **(B)** shows the detrimental effect of weather error on the prediction error, while **(C)** relates the daily prediction error and the RRMSE. Finally, **(D)** investigates the correlation between ground truth and predicted values – the diagonal line represents a perfect prediction accuracy. For the complete data, please see the supplementary materials.

Under optimal conditions, the curves in Fig. 9 **(A)** and **(B)** would be low and as flat as possible. However, this is an ill-posed problem, implying that the model can completely compensate for the forecast error. Nevertheless, the prediction model can be expected to utilize up-to-date forecasts more efficiently. This can be quantified by the relative accuracy decrease over the $1-10$ day predictions. Using the data available in the supplementary materials, this results in an average RRMSE increase of 32.35 % (RForSeH), 34.65 % (RForEnh), 14.84 % (SPred), and 51.06 % (SPredEnh). The increased percentage indicates that the model enhancements lead to better utilization of forecast uncertainty.

Fig. 9 **(C)** analyzes the modes of model accuracy by plotting the relationship between RRMSE and PError. Thus, the models can be split into three categories. The smoothing models predict correctly on average, matching the lower frequency parts of the signal. Conversely, the pinpoint models try to match the high-frequency changes. Finally, balanced models cover both, leading to models with curves close to the diagonal – e.g., RForEnh and SPredEnh.

Fig. 9 **(D)** presents a correlation between ground truth and predicted power production. The results show that extreme values are more challenging to predict. This is caused by the rarity of anomalously low or high production values in the training data. The probability of their occurrence is low. Thus, models avoid them in deference to more probable results. The problem is improved by sample weighing, as seen by the SPredEnh being closer to the diagonal.

Seasonality and Climate: Seasonality presents a noticeable effect on the difficulty of prediction, as seen by the results displayed in Fig. 8. The days categorized as Overcast are concentrated during the winter months, directly corresponding to increased prediction error and weather forecast error (see supplementary materials). Some power plants are also inherently more difficult to predict, as seen in Fig. 7. The causes include varied climates and location-specific properties, such as the position of the plant and its orientation.

			1D		3D		7D		Clear Days		Overcast Days		All Days		
Model	Meas	Fore	RRMSE	PError	RRMSE	PError	RRMSE	PError	RRMSE	PError	RRMSE	PError	RMSE	RRMSE	PError
RForEnh	✗	✓	10.85	31.166	10.88	31.886	12.86	35.711	6.90	19.838	17.34	44.982	2775	12.46	36.468
SPredEnh	✗	✓	**6.15**	**19.057**	**8.54**	**26.514**	**8.89**	**27.228**	**4.50**	**15.308**	**11.29**	**32.480**	**1897**	**8.45**	**26.304**
RForEnhFW	✓	✗	9.47	24.432	9.47	24.435	9.47	24.448	4.32	8.635	14.22	33.832	2127	9.59	24.659
SPredEnhFW	✓	✗	**6.39**	**21.699**	**6.43**	**21.428**	**6.41**	**21.363**	**2.97**	**9.427**	**8.88**	**28.485**	**1387**	**6.43**	**21.526**

Table 6
(2-column) Forecast Difficulty: A comparison of training/testing on forecasts (RForEnh, SPredEnh) and measured weather (RForEnhFW, SPredEnhFW). The results show that using zero-error weather is 23 % easier compared to weather forecasts.

Zero-Error Weather: Tab. 6 analyzes the disparity in prediction accuracy between weather forecasts and measured weather. The same models are trained and evaluated on weather forecasts (RForEnh, SPredEnh) and measured weather (RForEnhFW, SPredEnhFW). Notably, the comparison shows a 23.03 % and 23.90 % difference in accuracy for RForEnh and SPredEnh, respectively. This confirms that predictions made using weather forecasts are considerably more difficult. In conclusion, we note that with improved weather forecasts, the prediction accuracy can potentially improve by a further 23 %, thus emphasizing the need for high-quality weather forecasts.

3.8. Limitations and Discussion

The SolarPredictor system has several limitations. First is the apparent dependence of the models on weather forecast data. Although the models expect unreliable forecast data – due to the training regime – their performance is limited by the veracity of the weather forecast. For the SolarDB dataset, we quantify the gap between the performance possible with completely accurate weather at around 23 %. Further, the model is also dependant on training data. While the requirements are lower compared to other models, the SolarPredictor still needs at least a month of historical data for reliable predictions. For a wider deployment, a transfer learning scheme could be used to expedite the process. A base model can be trained on the SolarDB corpus and then fine-tuned on several days of data for the target power plant. The resulting model can then be used for production and re-trained as more data becomes available.

To test our approach in practice, we integrated the SolarPredictor into a commercial solar management system for private residences. The resulting software allowed the user to monitor their photovoltaic panels and use our predictions to gauge how much power they can expect in the horizon of $1 - 10$ days. To allow quick deployment, we utilized the transfer learning approach outlined above. A base model was trained on the SolarDB dataset. Then, adding a new power plant consisted of the following steps. First, a ramp-up phase included re-training the model once a day. The pre-trained base model was fine-tuned on all up-to-date available data. Thus, as the power plant collected more data, the training corpus grew. Finally, the second phase began when a month of historical production data was available – the re-training frequency was reduced to a monthly basis. The resulting system was able to quickly adapt to new power plants while also allowing the expansion of the pre-training dataset.

However, there are other uses for SolarPredictor, specifically in the power grid management sector. Transmission system operators (TSO) have various means of controlling the power grid to ensure stable and reliable operation [79]. One includes the control reserves – backup energy sources that are employed when power consumption fluctuates or due to a power plant outage. Another way is to dis/connect interruptible loads – large consumers which allow outages. Finally, redispatch measures allow the operators to adjust the feed from connected power plants to reduce power congestion and restore balance. However, a growing percentage of renewable energy sources – such as photovoltaic or wind power plants – introduces inherent instability into the grid. These sources are heavily dependent on weather conditions and, as such, are difficult to predict. With the SolarPredictor system, we provide a step towards increasing power grid stability. If employed, the system allows the TSOs to expect excess or lack of energy in advance. Therefore, interruptible loads can be notified, and an early plan for control reserve activation can be put in place.

The system is also useful concerning the requirements of the growing electromobility industry. Longer journeys utilizing electric vehicles (EV) require careful route planning to account for the shorter drive distance. Similar concerns will be even more prevalent once EVs are used for logistics – i.e., automated trucks and deliveries. The recharge stations require a stable energy supply, and long-term storage is currently not available. However, the SolarPredictor system can be used to predict power availability along the route, thus allowing automated routing based on energy availability.

4. Conclusions

In this study, we proposed a deep-learning system called SolarPredictor, facilitating the prediction of solar power plant production. The proposed SolarPredictor model is novel in two respects. First, its architecture combines elements of UNet for spatio-temporal analysis with residual aggregation modules composing the prediction signal. Second, the training and augmentation regime utilizes weather forecasts to allow the model to better estimate real-world forecast errors and automatically adapt to specific power plant sites. Further, we presented the SolarDB dataset, covering a year of data for 16 power plants along with seven days of hourly weather forecast data. We make this dataset freely available to the research community, facilitating further research. The SolarDB dataset is used to evaluate the prediction performance of various models. The novel SolarPredictor architecture results in a 14.76 % improvement against the second-best model. The proposed training regime is shown to effectively allow the model to gauge uncertainty in weather forecasts, resulting in an additional 28.27 % improvement over the base model trained only on current weather. Notably, the presented results are consistent with real-world performance since weather forecast data is used during the evaluation. Finally, we present an empirical analysis of the factors determining the prediction difficulty, showing the importance of considering power plant location, seasonality, and quality of weather forecasts.

Similarly to other machine learning approaches, the proposed SolarPredictor model requires a considerable amount of training data for each power plant, slowing down its deployment. Our preliminary experiments indicate that transfer learning is highly viable. By training a baseline model on the SolarDB dataset and fine-tuning it to a target power plant, we are able to provide predictions with only several days of data instead of months. Our future work will focus on amending this issue, quantifying how well the models generalize between power plants.

Declaration of Competing Interest

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

Acknowledgement

The SolarDB dataset was created in collaboration with Forena s.r.o., who provided access to the power plant data.

A. Dataset Details

For an overview of quantities contained within the SolarDB dataset, see Tab. 7. An random selection of Clear and Overcast production curves can be seen in Fig. 10.

Type	Name	Description	Type	Name	Description
Power	Power_{AC}	Final output after conversion	Meta	Freq	Frequency of power observation
	$\text{Energy}_{\text{Int}}$	Energy produced over interval		Capacity	Installed capacity and number of inverters
	Power_{DC}	Pure output of the panels		Location	Anonymized latitude and longtitude
Weather	Summary	String summary of weather	Exogenous	Interpolated	Flag for original (0), interpolated (1)
	PrecipInt	Precipitation intensity		Extrapolated	Flag for original (0), extrapolated (1)
	PrecipProb	Precipitation probability		Day	Observation Day of the year
	Temp	Measured temperature		Year	Year of observation
	ApparentTemp	Perceived temperature		Time	Observation second of the day
	DewPoint	Dew point temperature		Age	Age of forecast, 0 for measured
	Humidity	Humidity from dry (0) to humid (1)		SunAltitude	Sun altitude from ground plane
	Pressure	Pressure at ground level		SunAzimuth	Sun azimuth, north at 0, clockwise
	WindSpeed	Average wind speed		SunIrradiance	Estimated clear sky irradiance
	WindBearing	Wind bearing, north at 0°, clockwise		Status	Status code for power record
	CloudCover	Cloud cover, clear (0) to overcast (1)		Error	Error code for power record
	Visibility	Visibility up to 16 km		Clear	Daily clarity, clear (0) or overcast (1)

Table 7
(2-column) Dataset Quantities: Categorization of data included in the SolarDB dataset. Power production and exogenous variables are provided in 5-minute intervals for plants and inverters separately. Weather variables include on-site weather and weather forecasts, both at a 1-hour resolution. Meta-data is provided for each site and inverter.

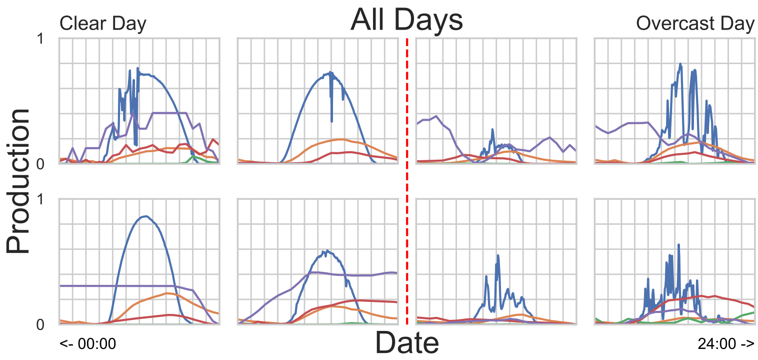

Figure 10: (Color, 1-column) Day Clarity: Examples of clear (**left**) and overcast (**right**) productions from the SolarDB dataset. Each graph contains the power production (blue), temperature (orange), precipitation (green), wind speed (red), and humidity (purple). Generally, prediction for clear days is less complicated due to reduced weather-related effects.

B. Predictor Details

Cyclic Time features consist of two sub-features containing the sine and cosine components, calculated as

$$v_x = 0.5 \cos\left(2\pi \frac{v}{v_{max}}\right) + 0.5, \tag{5}$$

$$v_y = 0.5 \sin\left(2\pi \frac{v}{v_{max}}\right) + 0.5, \tag{6}$$

where v is the input value and v_{max} is the maximum value. The resulting values, along with the Age and Sun features, facilitate the model's phase detection.

Sample weights include the clear-day weights wg_{cd}, which balance the data based on day clarity (Sec. 2.1.3), setting $wg_{cd} = \max\left[0.5, \min[1.0, 2.0 - L_{OC}]\right] \in \langle 0.5, 1.0\rangle$, where L_{OC} is the clear-day likelihood defined in Eq. 2. The temporal weights wg_t favor samples based on their frequency

$$wg_t(dt) = \begin{cases} \lambda_d, & \text{if } dt \text{ is the beginning of the day} \\ \lambda_h \cdot \frac{1}{24}, & \text{if } dt \text{ is the beginning of the hour} \\ \lambda_s \cdot \frac{1}{288}, & \text{otherwise,} \end{cases} \tag{7}$$

where dt is the sample's timestamp, $\lambda_d = 1$, $\lambda_h = 12$, $\lambda_s = 32$ are the importance multipliers, and $1/24$, $1/288$ are the relative frequencies for hourly and 5-minute samples, respectively.

Weather Sampling is used during the training to gradually teach the model the veracity of the input weather features. The three types of sampling schemes can be found in Tab. 8.

	Power	t	$t+1$	$\dots$	$t+o$
Weather	Measured	$\left.{}^{t}_{t}\right\|0$	$\left.{}^{t+1}_{t+1}\right\|0$	$\dots$	$\left.{}^{t+o}_{t+o}\right\|0$
	Forecast	$\left.{}^{t-d}_{t}\right\|d$	$\left.{}^{t+1-d}_{t+1}\right\|d$	$\dots$	$\left.{}^{t+o-d}_{t+o}\right\|d$
	Realistic	$\left.{}^{t}_{t}\right\|0$	$\left.{}^{t}_{t+1}\right\|1$	$\dots$	$\left.{}^{t}_{t+o}\right\|o$

Table 8
(1-column) Weather Sampling: The sampling scheme used for weather feature selection. Target power values at a given time $t + x$ are paired with weather features $\left.{}^{t+s}_{t+g}\right| a$ from source time $t + s$ for a target time $t + g$, imposing age $a = g - s$. Measured weather provides zero-error features, Forecast uses a fixed age, and Realistic increases the age linearly with time.

Loss Function used during the training is a combination of Mean Squared Error $\mathcal{L}_{\mathrm{M}}$, Huber Loss $\mathcal{L}_{\mathrm{H}}$, and Differential Smooth Loss $\mathcal{L}_{\Delta}$. The Smooth Loss $\mathcal{L}$ is defined as

$$\mathcal{L}_{\mathrm{M}}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2, \tag{8}$$

$$\mathcal{L}_{\mathrm{H}}(y, \hat{y}) = \begin{cases} \frac{1}{2} \left(y_i - \hat{y}_i \right)^2, & \text{if } \left| y_i - \hat{y}_i \right| \leq \delta \\ \frac{1}{2} \delta^2 + \delta \left(\left| y_i - \hat{y}_i \right| - \delta \right), & \text{otherwise,} \end{cases} \tag{9}$$

$$\mathcal{L}_{\Delta}(y, \hat{y}) = \sum_{i=1}^{n-1} \left| \hat{y}_{i+1} - \hat{y}_i \right|, \tag{10}$$

$$\mathcal{L}(y, \hat{y}) = wg(y) \cdot \left(\lambda_{\mathrm{M}}\, \mathcal{L}_{\mathrm{M}}(y, \hat{y}) + \lambda_{\mathrm{H}}\, \mathcal{L}_{\mathrm{H}}(y, \hat{y}) + \lambda_{\Delta}\, \mathcal{L}_{\Delta}(y, \hat{y}) \right), \tag{11}$$

with wg being the sample weight, experimentally choosing $\delta = 1$, $\lambda_{\mathrm{M}} = 0.5$, $\lambda_{\mathrm{H}} = 0.7$, and $\lambda_{\Delta} = 0.5$.

Predictor Architecture, visualized in Fig. 2, was implemented using the TensorFlow framework [76]. The AMSGrad [73] variant of the ADAM [72] optimizer is used in training. The Initial $lr = 0.005$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ is modified with the reduce-on-plateau technique ($f = 0.1$, $p = 5$), utilizng mini-batch size of 256. The model is trained for 60 epochs with the Smooth Loss objective. The inputs include 288 values of weather features and 288 values of historical power values. The model produces 288 values at each prediction step. The Conv1D layers comprise causal dilated 1D convolutions [53] using the LeakyReLU ($\alpha = 0.3$) activation function. Glorot normal and uniform [80] initialization is used for the kernel and bias, respectively. Additionally, kernel regularization is used, setting $l_2 = 0.005$. The architectural details for the diagram visualized in Fig. 2 are as follows:

- $\Downarrow$ ResBlock [F]: In → Conv1D[F, 3] → BatchNormalization → Conv1D[F, 3] → BatchNormalization → AveragePooling1D[$s = 2$]] → Add[In → Conv1D[F, 1, $s = 2$] → BatchNormalization] → Out

- ResBlock [F]: In → Conv1D[F, 3] → BatchNormalization → Conv1D[F, 3] → BatchNormalization → Add[In → Conv1D[F, 1] → BatchNormalization] → DropoutOut

- $\Uparrow$ ResBlock [F]: In → UpConv1D[$s = 2$] → Conv1D[F, 3] → BatchNormalization → Concatenate[Skip] → Conv1D[F, 3] → BatchNormalizationConv1D[F, 3] → BatchNormalization → Add[In → Conv1DTranspose[F, 1, $s = 2$] → BatchNormalization] → Out

- ResAgg: In → UpDownConv1D[288] → Conv1D[1, 1] → Add[In] → Out

- RUNet: Concatenate[Power History →$\Downarrow$ ResBlock[4], Weather Features →$\Downarrow$ ResBlock[4]] →$\Downarrow$ ResBlock[8] →$\Downarrow$ ResBlock[16] →$\Downarrow$ ResBlock[32] → ResBlock[64] → ResBlock[64] →$\Uparrow$ ResBlock[32] →$\Uparrow$ ResBlock[16] →$\Uparrow$ ResBlock[8] →$\Uparrow$ ResBlock[4]

- RComposer [UN]: Concatenate[Power History →$\Downarrow$ ResBlock[4], Weather Features →$\Downarrow$ ResBlock[4]] → ResBlock[16] → ResBlock[8] → ResAgg[Concatenate[$\Downarrow$ ResBlock[4]$_{\mathrm{UN}}$, $\Downarrow$ ResBlock[4]$_{\mathrm{UN}}$]] → ResAgg[$\Downarrow$ ResBlock[8]$_{\mathrm{UN}}$] → ResAgg[$\Downarrow$ ResBlock[16]$_{\mathrm{UN}}$] → ResAgg[$\Downarrow$ ResBlock[32]$_{\mathrm{UN}}$] → ResAgg[$\Uparrow$ ResBlock[32]$_{\mathrm{UN}}$] → ResAgg[$\Uparrow$ ResBlock[16]$_{\mathrm{UN}}$] → ResAgg[$\Uparrow$ ResBlock[8]$_{\mathrm{UN}}$] → ResAgg[$\Uparrow$ ResBlock[4]$_{\mathrm{UN}}$] → ResBlock[4] → Out

- SolarPredictor: Input → RComposer[RUNet] → Output

C. Evaluation

Importance Analysis uses the normalized Gini Importance [75] as the metric, defined as

$$i_f = 1/n_s^{root} \sum \left(n_s^p n_i^p - n_s^l n_i^l - n_s^r n_i^r \right) \tag{12}$$

for each feature f, where n_s^x is the number of samples reaching node x and n_i^x is the Gini Impurity for node x. Furthermore, *root* represents the root node, and p, l, r are the parent, left child, and right child nodes. Finally, the i_f values are calculated for each ensemble and aggregated through their mean I_f and variance σ_f^2 values. However, as noted by other authors [81, 82], Gini importance may lead to biased results. Thus, we confirm our results by computing the Permutation Importance [81], leading to the same relative ordering.

Quantitative Metrics used during evaluation include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Relative RMSE (RRMSE), and Coefficient of Determination (R^2), defined as

$$\mathrm{MSE} = \frac{1}{n}\sum_i \left(y_i - \hat{y}_i\right)^2, \tag{13}$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}, \tag{14}$$

$$\mathrm{RRMSE} = \sqrt{\frac{\mathrm{MSE}}{\frac{1}{n}\sum_i y_i^2}}, \tag{15}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_i \left(y_i - \hat{y}_i\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2}, \tag{16}$$

where y_i, $\hat{y}_i$ are the ground-truth and predicted values, respectively. Hourly aggregates of weather (WError) and power (PError) are defined as $\left(\sum_i \hat{y}_i \hat{f}\right) / \left(\sum_i y_i f\right)$, where f and $\hat{f}$ represent the measurement frequency. Unless otherwise specified, the measurement frequency for both values is $f = \hat{f} = 5 / 60$ – i.e., 5-minute intervals. Data correlations are examined using Pearson (cor_p) [74] and Spearman (cor_s) [83] correlation coefficients.

References

[1] Merce Almuni, Ils Moorkens, Risto Juhana Saarikivi, and Mihai Tomescu. ETC/CME Report 7/2020: Renewable energy in Europe 2020 - Recent growth and knock-on effects. *ETC/CME, December 2020, 113 pp.*, 1(Eionet report ETC/CME 7/2020), December 2020. URL https://www.eionet.europa.eu/etcs/etc-cme/products/etc-cme-reports/etc-cme-report-7-2020-renewable-energy-in-europe-2020-recent-growth-and-knock-on-effects.

[2] Gideon A.H. Laugs, René M.J. Benders, and Henri C. Moll. Balancing responsibilities: Effects of growth of variable renewable energy, storage, and undue grid interaction. *Energy Policy*, 139:111203, 2020. ISSN 0301-4215. doi: https://doi.org/10.1016/j.enpol.2019.111203. URL https://www.sciencedirect.com/science/article/pii/S0301421519307876.

[3] Lukas Nonnenmacher, Amanpreet Kaur, and Carlos F.M. Coimbra. Day-ahead resource forecasting for concentrated solar power integration. *Renewable Energy*, 86:866–876, 2016. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2015.08.068. URL https://www.sciencedirect.com/science/article/pii/S0960148115302688.

[4] Jennifer Luoma, Patrick Mathiesen, and Jan Kleissl. Forecast value considering energy pricing in california. *Applied Energy*, 125:230–237, 2014. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2014.03.061. URL https://www.sciencedirect.com/science/article/pii/S0306261914003018.

[5] Wei Wei, Feng Liu, and Shengwei Mei. Energy pricing and dispatch for smart grid retailers under demand response and market price uncertainty. *IEEE Transactions on Smart Grid*, 6(3):1364–1374, 2015. doi: 10.1109/TSG.2014.2376522.

[6] Tanveer Ahmad, Hongcai Zhang, and Biao Yan. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. *Sustainable Cities and Society*, 55:102052, 2020. ISSN 2210-6707. doi: https://doi.org/10.1016/j.scs.2020.102052. URL https://www.sciencedirect.com/science/article/pii/S2210670720300391.

[7] Birk Kraas, Marion Schroedter-Homscheidt, and Reinhard Madlener. Economic merits of a state-of-the-art concentrating solar power forecasting system for participation in the spanish electricity market. *Solar Energy*, 93:244–255, 2013. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2013.04.012. URL https://www.sciencedirect.com/science/article/pii/S0038092X13001527.

[8] R. Hanna, J. Kleissl, A. Nottrott, and M. Ferry. Energy dispatch schedule optimization for demand charge reduction using a photovoltaic-battery storage system with solar forecasting. *Solar Energy*, 103:269–287, 2014. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2014.02.020. URL https://www.sciencedirect.com/science/article/pii/S0038092X1400098X.

[9] Maurizio Delfanti, Davide Falabretti, and Marco Merlo. Energy storage for pv power plant dispatching. *Renewable Energy*, 80:61–72, 2015. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2015.01.047. URL https://www.sciencedirect.com/science/article/pii/S0960148115000658.

[10] R. Ahmed, V. Sreeram, Y. Mishra, and M.D. Arif. A review and evaluation of the state-of-the-art in pv solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews*, 124:109792, 2020. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2020.109792. URL https://www.sciencedirect.com/science/article/pii/S1364032120300885.

[11] Bruce Rose and Eric Floehr. Analysis of high-temperature forecast accuracy of consumer weather forecasts from 2005–2016. *Forecast Watch Report.(Dublin, OH, USA)*, 2017.

[12] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley, 2015. ISBN 9781118674925. URL https://books.google.cz/books?id=rNt5CgAAQBAJ.

[13] Huaizhi Wang, Yangyang Liu, Bin Zhou, Canbing Li, Guangzhong Cao, Nikolai Voropai, and Evgeny Barakhtenko. Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Conversion and Management*, 214:112909, 2020. ISSN 0196-8904. doi: https://doi.org/10.1016/j.enconman.2020.112909. URL https://www.sciencedirect.com/science/article/pii/S0196890420304477.

[14] Sobrina Sobri, Sam Koohi-Kamali, and Nasrudin Abd. Rahim. Solar photovoltaic generation forecasting methods: A review. *Energy Conversion and Management*, 156:459–497, 2018. ISSN 0196-8904. doi: https://doi.org/10.1016/j.enconman.2017.11.019. URL https://www.sciencedirect.com/science/article/pii/S0196890417310622.

[15] Marzieh Mokarram, Mohammad Jafar Mokarram, Mohsen Gitizadeh, Taher Niknam, and Jamshid Aghaei. A novel optimal placing of solar farms utilizing multi-criteria decision-making (mcda) and feature selection. *Journal of Cleaner Production*, 261:121098, 2020. ISSN 0959-6526. doi: https://doi.org/10.1016/j.jclepro.2020.121098. URL https://www.sciencedirect.com/science/article/pii/S0959652620311458.

[16] Steven D. Miller, Matthew A. Rogers, John M. Haynes, Manajit Sengupta, and Andrew K. Heidinger. Short-term solar irradiance forecasting via satellite/model coupling. *Solar Energy*, 168:102–117, 2018. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2017.11.049. URL https://www.sciencedirect.com/science/article/pii/S0038092X17310435. Advances in Solar Resource Assessment and Forecasting.

[17] Yan Hao and Chengshi Tian. A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. *Applied Energy*, 238:368–383, 2019. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2019.01.063. URL https://www.sciencedirect.com/science/article/pii/S0306261919300637.

[18] Francis M. Lopes, Hugo G. Silva, Rui Salgado, Afonso Cavaco, Paulo Canhoto, and Manuel Collares-Pereira. Short-term forecasts of ghi and dni for solar energy systems operation: assessment of the ecmwf integrated forecasting system in southern portugal. *Solar Energy*, 170: 14–30, 2018. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2018.05.039. URL https://www.sciencedirect.com/science/article/pii/S0038092X18304729.

[19] Akinobu Murata, Hideaki Ohtake, and Takashi Oozeki. Modeling of uncertainty of solar irradiance forecasts on numerical weather predictions with the estimation of multiple confidence intervals. *Renewable Energy*, 117:193–201, 2018. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2017.10.043. URL https://www.sciencedirect.com/science/article/pii/S0960148117309813.

[20] G.V. Parishwad, R.K. Bhardwaj, and V.K. Nema. Estimation of hourly solar radiation for india. *Renewable Energy*, 12(3):303–313, 1997. ISSN 0960-1481. doi: https://doi.org/10.1016/S0960-1481(97)00039-6. URL https://www.sciencedirect.com/science/article/pii/S0960148197000396.

[21] L.W. Florschuetz. Extension of the hottel-whillier model to the analysis of combined photovoltaic/thermal flat plate collectors. *Solar Energy*, 22(4):361–366, 1979. ISSN 0038-092X. doi: https://doi.org/10.1016/0038-092X(79)90190-7. URL https://www.sciencedirect.com/science/article/pii/0038092X79901907.

[22] Hadrien Verbois, Robert Huva, Andrivo Rusydi, and Wilfred Walsh. Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning. *Solar Energy*, 162:265–277, 2018. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2018.01.007. URL https://www.sciencedirect.com/science/article/pii/S0038092X18300197.

[23] Joakim Munkhammar, Dennis van der Meer, and Joakim Widén. Probabilistic forecasting of high-resolution clear-sky index time-series using a markov-chain mixture distribution model. *Solar Energy*, 184:688–695, 2019. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2019.04.014. URL https://www.sciencedirect.com/science/article/pii/S0038092X19303469.

[24] Laith M. Halabi, Saad Mekhilef, and Monowar Hossain. Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation. *Applied Energy*, 213:247–261, 2018. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2018.01.035. URL https://www.sciencedirect.com/science/article/pii/S0306261918300357.

[25] Jin Dong, Mohammed M. Olama, Teja Kuruganti, Alexander M. Melin, Seddik M. Djouadi, Yichen Zhang, and Yaosuo Xue. Novel stochastic methods to predict short-term solar radiation and photovoltaic power. *Renewable Energy*, 145:333–346, 2020. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2019.05.073. URL https://www.sciencedirect.com/science/article/pii/S0960148119307463.

[26] Tanveer Ahmad, Dongdong Zhang, and Chao Huang. Methodological framework for short-and medium-term energy, solar and wind power forecasting with stochastic-based machine learning approach to monetary and energy policy applications. *Energy*, 231:120911, 2021. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2021.120911. URL https://www.sciencedirect.com/science/article/pii/S0360544221011592.

[27] Mohamed Louzazni, Heba Mosalam, Ahmed Khouya, and Khalid Amechnoue. A non-linear auto-regressive exogenous method to forecast the photovoltaic power output. *Sustainable Energy Technologies and Assessments*, 38:100670, 2020. ISSN 2213-1388. doi: https://doi.org/10.1016/j.seta.2020.100670. URL https://www.sciencedirect.com/science/article/pii/S221313881930952X.

[28] David P. Larson, Lukas Nonnenmacher, and Carlos F.M. Coimbra. Day-ahead forecasting of solar power output from photovoltaic plants in the american southwest. *Renewable Energy*, 91:11–20, 2016. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2016.01.039. URL https://www.sciencedirect.com/science/article/pii/S0960148116300398.

[29] Umit Agbulut, Ali Etem Gurel, and Yunus Bicen. Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable and Sustainable Energy Reviews*, 135:110114, 2021. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2020.110114. URL https://www.sciencedirect.com/science/article/pii/S1364032120304056.

[30] Ellysia Jumin, Faridah Bte Basaruddin, Yuzainee Bte. Md Yusoff, Sarmad Dashti Latif, and Ali Najah Ahmed. Solar radiation prediction using boosted decision tree regression model: A case study in malaysia. *Environmental Science and Pollution Research*, 28(21):26571–26583, Jun 2021. ISSN 1614-7499. doi: 10.1007/s11356-021-12435-6. URL https://doi.org/10.1007/s11356-021-12435-6.

[31] L. Benali, G. Notton, A. Fouilloy, C. Voyant, and R. Dizene. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renewable Energy*, 132:871–884, 2019. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2018.08.044. URL https://www.sciencedirect.com/science/article/pii/S0960148118309947.

[32] Alberto Torres-Barran, Álvaro Alonso, and Jose R. Dorronsoro. Regression tree ensembles for wind energy and solar radiation prediction. *Neurocomputing*, 326-327:151–160, 2019. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2017.05.104. URL https://www.sciencedirect.com/science/article/pii/S0925231217315229.

[33] Alireza Zendehboudi, M.A. Baseer, and R. Saidur. Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of Cleaner Production*, 199:272–285, 2018. ISSN 0959-6526. doi: https://doi.org/10.1016/j.jclepro.2018.07.164. URL https://www.sciencedirect.com/science/article/pii/S095965261832153X.

[34] Abdul Rahim Pazikadin, Damhuji Rifai, Kharudin Ali, Muhammad Zeesan Malik, Ahmed N. Abdalla, and Moneer A. Faraj. Solar irradiance measurement instrumentation and power solar generation forecasting based on artificial neural networks (ann): A review of five years research trend. *Science of The Total Environment*, 715:136848, 2020. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2020.136848. URL https://www.sciencedirect.com/science/article/pii/S0048969720303582.

[35] Fei Wang, Zhiming Xuan, Zhao Zhen, Kangping Li, Tieqiang Wang, and Min Shi. A day-ahead pv power forecasting method based on lstm-rnn model and time correlation modification under partial daily pattern prediction framework. *Energy Conversion and Management*, 212:112766, 2020. ISSN 0196-8904. doi: https://doi.org/10.1016/j.enconman.2020.112766. URL https://www.sciencedirect.com/science/article/pii/S0196890420303046.

[36] Ahmed Tealab. Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*, 3(2):334–340, 2018.

[37] Jinhua Zhang, Jie Yan, David Infield, Yongqian Liu, and Fue sang Lien. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and gaussian mixture model. *Applied Energy*, 241:229–244, 2019. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2019.03.044. URL https://www.sciencedirect.com/science/article/pii/S0306261919304532.

[38] Hui Liu, Xiwei Mi, Yanfei Li, Zhu Duan, and Yinan Xu. Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional gated recurrent unit network and support vector regression. *Renewable Energy*, 143:842–854, 2019. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2019.05.039. URL https://www.sciencedirect.com/science/article/pii/S0960148119306998.

[39] Sujan Ghimire, Ravinesh C. Deo, Nawin Raj, and Jianchun Mi. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy*, 253:113541, 2019. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2019.113541. URL https://www.sciencedirect.com/science/article/pii/S0306261919312152.

[40] Na Dong, Jian-Fang Chang, Ai-Guo Wu, and Zhong-Ke Gao. A novel convolutional neural network framework based solar irradiance prediction method. *International Journal of Electrical Power & Energy Systems*, 114:105411, 2020. ISSN 0142-0615. doi: https://doi.org/10.1016/j.ijepes.2019.105411. URL https://www.sciencedirect.com/science/article/pii/S0142061518332915.

[41] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012. URL http://arxiv.org/abs/1207.0580.

[42] Zhongyang Han, Jun Zhao, Henry Leung, King Fai Ma, and Wei Wang. A review of deep learning models for time series prediction. *IEEE Sensors Journal*, 2019.

[43] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *CoRR*, abs/1506.04214, 2015. URL http://arxiv.org/abs/1506.04214.

[44] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew W. Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *CoRR*, abs/1609.03499, 2016. URL http://arxiv.org/abs/1609.03499.

[45] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *CoRR*, abs/1803.01271, 2018. URL http://arxiv.org/abs/1803.01271.

[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[47] Haixiang Zang, Ling Liu, Li Sun, Lilin Cheng, Zhinong Wei, and Guoqiang Sun. Short-term global horizontal irradiance forecasting based on a hybrid cnn-lstm model with spatiotemporal correlations. *Renewable Energy*, 160:26–41, 2020. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2020.05.150. URL https://www.sciencedirect.com/science/article/pii/S0960148120308557.

[48] Jiaqi Qu, Zheng Qian, and Yan Pei. Day-ahead hourly photovoltaic power forecasting using attention-based cnn-lstm neural network embedded with multiple relevant and target variables prediction pattern. *Energy*, 232:120996, 2021. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2021.120996. URL https://www.sciencedirect.com/science/article/pii/S0360544221012445.

[49] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[50] Arun Venkatraman, Martial Hebert, and J. Andrew Bagnell. Improving multi-step prediction of learned time series models. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, page 3024–3030. AAAI Press, 2015. ISBN 0262511290.

[51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[52] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[53] Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 156–165, 2017.

[54] Frank Vignola, Cathy Grover, Nick Lemon, and Andrew McMahan. Building a bankable solar radiation dataset. *Solar Energy*, 86(8):2218–2229, 2012. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2012.05.013. URL https://www.sciencedirect.com/science/article/pii/S0038092X1200182X. Progress in Solar Energy 3.

[55] Michele M. Rienecker et al. Merra: Nasa's modern-era retrospective analysis for research and applications. *Journal of Climate*, 24 (14):3624 – 3648, 2011. doi: 10.1175/JCLI-D-11-00015.1. URL https://journals.ametsoc.org/view/journals/clim/24/14/jcli-d-11-00015.1.xml.

[56] Ronald Gelaro et al. The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate*, 30 (14):5419 – 5454, 2017. doi: 10.1175/JCLI-D-16-0758.1. URL https://journals.ametsoc.org/view/journals/clim/30/14/jcli-d-16-0758.1.xml.

[57] Gonzalez Aparicio Iratxe, Huld Thomas, Careri Francesco, Monforti-Ferrario Fabio, and Zucker Andreas. *EMHIRES dataset.Part II, Solar power generation* . Publications Office, 2017. doi: 10.2760/044693. URL https://data.europa.eu/doi/10.2760/044693.

[58] Gonzalez Aparicio Iratxe, Zucker Andreas, Careri Francesco, Monforti-Ferrario Fabio, Huld Thomas, and B. Jake. *EMHIRES dataset. Part I, Wind power generation.* Publications Office, 2016. doi: 10.2790/831549. URL https://data.europa.eu/doi/10.2790/831549.

[59] Manajit Sengupta, Yu Xie, Anthony Lopez, Aron Habte, Galen Maclaurin, and James Shelby. The national solar radiation data base (nsrdb). *Renewable and Sustainable Energy Reviews*, 89:51–60, 2018. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2018.03.003. URL https://www.sciencedirect.com/science/article/pii/S136403211830087X.

[60] Cong Feng, Dazhi Yang, Bri-Mathias Hodge, and Jie Zhang. Opensolar: Promoting the openness and accessibility of diverse public solar datasets. *Solar Energy*, 188:1369–1379, 2019. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2019.07.016. URL https://www.sciencedirect.com/science/article/pii/S0038092X19306693.

[61] D Lew, R Piwko, G Jordan, N Miller, K Clark, L Freeman, and M Milligan. Western wind and solar integration study. Technical report, National Renewable Energy Lab. (NREL), Golden, CO (United States), 1 2011. URL https://www.osti.gov/biblio/1046882.

[62] Aaron Bloom, Aaron Townsend, David Palchak, Joshua Novacheck, Jack King, Clayton Barrows, Eduardo Ibanez, Matthew O'Connell, Gary Jordan, Billy Roberts, Caroline Draxl, and Kenny Gruchalla. Eastern renewable generation integration study. Technical report, National Renewable Energy Lab. (NREL), Golden, CO (United States), 8 2016. URL https://www.osti.gov/biblio/1318192.

[63] T. Stoffel and A. Andreas. Nrel solar radiation research laboratory (srrl): Baseline measurement system (bms); golden, colorado (data). Technical report, National Renewable Energy Lab. (NREL), Golden, CO (United States), 7 1981. URL https://www.osti.gov/biblio/1052221.

[64] Sheffield Solar. Microgen database, 2016. URL http://www.microgen-database.org.uk. Last visited 2021-04-22.

[65] Pecan Street. Pecan street dataport database, 2021. URL https://dataport.pecanstreet.org/. Last visited 2021-04-22.

[66] Hugo T. C. Pedro, David P. Larson, and Carlos F. M. Coimbra. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. *Journal of Renewable and Sustainable Energy*, 11(3):036102, 2019. doi: 10.1063/1.5094494. URL https://doi.org/10.1063/1.5094494.

[67] Knowledge Australia Centre. Dka solar centre dataset, 2021. URL http://dkasolarcentre.com.au/historical-data/download.

[68] Federico Cassola and Massimiliano Burlando. Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output. *Applied Energy*, 99:154–166, 2012. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2012.03.054. URL https://www.sciencedirect.com/science/article/pii/S0306261912002747.

[69] Dark Sky. Darksky weather api, 2021. URL https://darksky.net/dev. Last visited 2021-04-22.

[70] Brandon Stafford. Pysolar documentation, 2015. URL https://pysolar.readthedocs.io/.

[71] Fabian Pedregosa et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[72] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

[73] Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of adam and beyond. *CoRR*, abs/1904.09237, 2019. URL http://arxiv.org/abs/1904.09237.

[74] Karl Pearson and Olaus Magnus Friedrich Erdmann Henrici. Vii. mathematical contributions to the theory of evolution.—iii. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 187:253–318, 1896. doi: 10.1098/rsta.1896.0007. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1896.0007.

[75] A. Hapfelmeier and K. Ulm. A new variable selection approach using random forests. *Computational Statistics & Data Analysis*, 60:50–69, 2013. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2012.09.020. URL https://www.sciencedirect.com/science/article/pii/S0167947312003490.

[76] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

[77] Pauli Virtanen et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[78] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*, abs/1406.1078, 2014. URL http://arxiv.org/abs/1406.1078.

[79] Andrea Biancardi, Matteo Di Castelnuovo, and Iain Staffell. A framework to evaluate how european transmission system operators approach innovation. *Energy Policy*, 158:112555, 2021. ISSN 0301-4215. doi: https://doi.org/10.1016/j.enpol.2021.112555. URL https://www.sciencedirect.com/science/article/pii/S0301421521004250.

[80] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In Yee Whye Teh and Mike Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR. URL http://proceedings.mlr.press/v9/glorot10a.html.

[81] Andre Altmann, Laura Tolosi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 04 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq134. URL https://doi.org/10.1093/bioinformatics/btq134.

[82] A. L. Boulesteix, A. Bender, J. Lorenzo Bermejo, and C. Strobl. Random forest gini importance favours snps with large minor allele frequency: impact, sources and recommendations. *Brief Bioinform*, 13(3):292–304, May 2012. [DOI:10.1093/bib/bbr053] [PubMed:21908865].

[83] C. Spearman. *The Proof and Measurement of Association Between Two Things.* Studies in individual differences: The search for intelligence. Appleton-Century-Crofts, East Norwalk, CT, US, 1961. doi: 10.1037/11491-005. URL https://doi.org/10.1037/11491-005.