

Ramp Rate Metric Suitable for Solar Forecasting

Bijan Nouri,* Yann Fabel, Niklas Blum, Dominik Schnaus, Luis F. Zarzalejo, Andreas Kazantzidis, and Stefan Wilbert

Solar irradiance forecasting plays a crucial role in integrating large quantities of intermittent solar power. Forecasting systems are commonly evaluated using metrics like root-mean-square error (RMSE) and skill scores. However, these metrics aggregated over larger data sets do not adequately assess the prediction of ramp events, which are critical for many applications. This article introduces a novel, simple, and adaptable ramp rate metric that analyzes ramp events between successive lead times within forecasts. A case study on ramp rate mitigation in PV systems benchmarks suitable ramp thresholds for various solar irradiance components. The capabilities and limitations of deterministic and probabilistic forecasts from two all-sky imager-based models are evaluated for ramp prediction. A state-of-the-art data-driven vision transformer End2End model excels in RMSE and skill scores but performs poorly in ramp prediction. Conversely, a novel generative forecasting model combined with a convolutional neural network-based irradiance model shows superior ramp prediction, achieving an F1 score of ≥ 0.7 for critical ramp events. This study underscores the importance of suitable ramp rate metrics and highlights the potential of generative models for enhancing ramp forecasting.

the power system.^[1] This issue is particularly critical for large centralized solar plants or dense fleets of solar installations. Active smoothing and buffering through battery energy storage systems could mitigate the negative effects of ramps, but at significantly increased operating and capital costs.^[2] By using intra-hour forecasts, changing solar irradiance conditions can be anticipated, allowing optimized power plant operation and grid integration.^[3] Consequently, this has the potential to reduce storage system capacity requirements and costs.^[4] Numerical weather prediction (NWP), satellite observations, local sensing, and data-driven or hybrid methods are all valuable tools for generating forecasts.^[5] The choice of method depends on the specific application and the relevant temporal and spatial scales. Particularly for large centralized solar power plants or dense fleets of solar power plants, ramps per minute could be of importance.^[1]

1. Introduction

Sudden changes in solar irradiance on a local scale can have a significant impact on solar power generation. This intermittent nature of the solar resource, which is mainly due to cloud passing's, poses a challenge to the integration of solar energy into

Local sensing methods, such as all-sky imagers that capture sky conditions with high spatial and temporal resolution, could serve as suitable forecasting systems.^[6] Recent years have seen a surge in new developments with increasingly powerful forecasting models, particularly regarding direct data-driven methods using deep learning.^[5,7]

The primary metrics used to evaluate and optimize forecasting systems are root-mean-square error (RMSE) and forecast skill. The RMSE serves as a measure of the accuracy achieved by the developed model, while the forecast skill serves as a critical benchmark that evaluates the model's RMSE against that of a persistence baseline model.^[8] However, the pursuit of RMSE optimization often results in a smoothing effect on forecasts, especially in direct data-driven approaches. While these optimization strategies excel at minimizing RMSE and demonstrate high forecast skill, they may prove less adept at predicting short-term ramps.^[9] This makes some forecasts well suited for applications that require accurate energy quantity prediction for a given time interval but virtually useless for control applications where the actual ramp detection is desired (e.g., ramp rate mitigation of PV power plants).^[4] Highly resolved probabilistic approaches could be potential solution.^[10,11] Yet, when ensemble members or initial deterministic forecasts in a quantile approach are optimized for RMSE, similar smoothing effects are partially observed. These effects appear at individual probability levels in probabilistic forecasts. While the likelihood of ramps can be inferred from the sharpness of the probabilistic forecast

B. Nouri, Y. Fabel, N. Blum, S. Wilbert
Institute for Solar Research
German Aerospace Center (DLR)
Calle Doctor Carracido N°44, 04005 Almería, Spain
E-mail: bijan.nouri@dlr.de

D. Schnaus
Chair of Computer Vision & Artificial Intelligence
Technical University Munich (TUM)
85748 Garching, Germany

L. F. Zarzalejo
Energy Department–Renewable Energy Division
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT)
Av. Complutense 40, 28040 Madrid, Spain

A. Kazantzidis
Physics Department, Laboratory of Atmospheric Physics
University of Patras
Patras 26500, Greece

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/solr.202400468>.

DOI: 10.1002/solr.202400468

distribution, exact forecasts of ramp events are not provided. It is therefore important that ramp-rate specific metrics are also considered when evaluating the adequacy of probabilistic forecasts.

Ramps are characterized by both the difference in irradiance from the beginning to the end of a time interval and the difference between the minimum and maximum irradiance within that timeframe.^[12,13] This highlights the importance of considering both duration and magnitude in forecasting ramps. Several potential metrics have been described in the literature.^[14] One common approach is to assess the difference between two measured irradiance values (γ) while ramp prediction involves comparing the predicted ($\hat{\gamma}$) and measured irradiance.^[15] This is defined as:

$$\gamma(t + FH) - \gamma(t) > \varepsilon \rightarrow \text{observed ramp event} \quad (1)$$

$$\hat{\gamma}(t + FH) - \gamma(t) > \varepsilon \rightarrow \text{predicted ramp event} \quad (2)$$

FH is the forecast horizon and ε is the ramp threshold. The threshold is set to 10% of the corresponding clear sky irradiance and is limited to a discrete forecast horizon of 10 min into the future. This makes the approach rather rigid and potentially unsuitable for various applications. Additionally, using a threshold based on a fixed proportion of clear-sky irradiance is highly sensitive during morning and evening hours, potentially triggering events solely due to changes in the sun's position.

An adapted version of the method considers the derivative of normalized irradiance values ($n\gamma$), making it more suitable for varying FH and sun positions:^[14]

$$\frac{n\gamma(t + FH) - n\gamma(t)}{FH} > \varepsilon \rightarrow \text{observed ramp event} \quad (3)$$

$$\frac{n\hat{\gamma}(t + FH) - n\gamma(t)}{FH} > \varepsilon \rightarrow \text{predicted ramp event} \quad (4)$$

Normalization is performed using the maximum clear sky irradiance at the top of the atmosphere for each day. The threshold ε is defined individually for each FH as the 99th percentile of ramp rates under clear sky conditions. Although the intention is to establish a universally applicable threshold, using the 99th percentile of ramp rates under clear sky conditions makes ε extremely sensitive and thus unsuitable. Establishing an appropriate threshold for defining ramps remains a common challenge with these metrics. As in the examples presented, thresholds often vary depending on the dataset used, which complicates reproducibility and comparability. The swinging-door algorithm addresses this by dynamically adapting to identify ramp events in data streams, utilizing a variable ε_{SDA} (width of a door) to pinpoint ramp boundaries.^[16] However, the algorithm's sensitivity to ε_{SDA} is crucial, a small value may lead to false detections from noise, while a large value might miss smaller ramps. Hence, selecting an appropriate ε_{SDA} is vital to strike a balance between sensitivity and robustness. Due to its complexity, the swinging-door algorithm is less intuitive and more difficult to interpret. Further metrics have been developed specifically for hourly resolution NWP forecasts.^[12,17] But the usefulness of some of these metrics is questionable since persistence forecasts outperform NWP models in predicting ramps.^[17] By definition, persistence models cannot predict ramps caused

by cloud passages, so a meaningful ramp metric should not indicate a good ramp prediction skill for persistence approaches. This issue is particularly important for highly resolved intra-hourly forecasts, where changes in irradiance due to solar position between subsequent forecast lead times are minimal.

No metric has yet been found that is adequate for many use cases and accepted by most stakeholders.^[5,14] This is partly due to the complexity of the so far proposed metrics and partly due to the difficulty of establishing a definition of practical irradiance ramp events.

This study aims to develop a simple, adaptable ramp rate metric to effectively assess ramp events in solar irradiance forecasting. We introduce a novel ramp rate metric and conduct a case study on ramp rate mitigation in PV systems. Our study focuses on systems operating under grid codes with specific ramp limits to benchmark suitable ramp thresholds. The capabilities and limitations of ramp prediction of deterministic and probabilistic forecasts from all-sky imager-based models are evaluated for a state-of-the-art data-driven and a novel generative model.

In Section 2, we introduce and discuss the error metrics. The experimental setup is presented in Section 3, including the PV model and data sets used. Section 4 evaluates different ramp rate thresholds for the specific use case of ramp rate mitigation in PV plants. In Section 5, both a state-of-the-art data-driven forecasting model and a novel generative forecasting model are presented, along with their validation results using the previously defined ramp error metric and thresholds. Finally, Section 6 finishes with concluding remarks.

2. Error Metrics

Section 2.1 first presents typical established error metrics. Section 2.2 then introduces the novel error metric presented in this article specifically for ramp rates.

2.1. Overview on Typical Error Metrics

A detailed description of typical error metrics for deterministic forecasts is given in ref. [18], a corresponding overview of metrics for probabilistic forecasts can be found in ref. [19].

Using multiple metrics allows for a more comprehensive evaluation of forecast accuracy. Different metrics highlight different characteristics of the errors. Typical error metrics for deterministic forecasts are the Bias, the mean absolute error (MAE) and the RMSE:

$$\text{Bias} = 1/n \sum_{i=1}^n \hat{\gamma}_i - \gamma_i \quad (5)$$

$$\text{MAE} = 1/n \sum_{i=1}^n |\hat{\gamma}_i - \gamma_i| \quad (6)$$

$$\text{RMSE} = \sqrt{1/n \sum_{i=1}^n (\hat{\gamma}_i - \gamma_i)^2} \quad (7)$$

Bias quantifies systematic deviations, MAE reflects average error magnitude, and RMSE emphasizes the presence of large errors.

The quality of probabilistic forecasts can be characterized by the attributes of reliability, resolution, and sharpness.^[20] Reliability assesses whether a forecast is statistically consistent, that is, high reliability implies that the predicted probabilities match the observed outcomes. Resolution evaluates the forecasts ability to distinguish between different prevailing conditions. Sharpness describes the concentration of the forecast distribution derived solely from the forecasts themselves. A forecast can be sharp but useless if it is unreliable. Skilled probabilistic forecasts should be both reliable and have high resolution.

Typical error metrics for probabilistic forecasts are the interval score (IS), quantile score (QS), and the continuous ranked probability score (CRPS):

$$IS_{PI} = \frac{1}{N} \sum_{i=1}^N (U_{i,PI} - L_{i,PI}) + \frac{2}{\alpha} (L_{i,PI} - y_i) 1_{y_i < L_{i,PI}} + \frac{2}{\alpha} (y_i - U_i) 1_{y_i > U_{i,PI}} \quad (8)$$

$$QS_{\tau} = \frac{1}{N} \sum_{i=1}^N \psi_{\tau}(y_i - \hat{y}_{i,\tau}) \quad \text{with } \psi_{\tau}(u) \begin{cases} \tau u & \text{if } u \geq 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases} \quad (9)$$

$$CRPS = \frac{1}{N} \sum_{i=1}^N \int_0^1 (F_{\hat{y}_i}(x) - F_{y_i}(x))^2 dx \quad (10)$$

where L is the lower bound, U is the upper bound, PI is the prediction interval, $F_{\hat{y}_i}$ describes the cumulative distribution functions (CDF) of the forecasts, F_{y_i} describes the CDF of the corresponding observation as Heaviside step function shifted to the observation y_i .

α is the statistical significance level and τ is the probability level. IS focuses on the accuracy of prediction intervals, rewarding narrow, accurate intervals and penalizing intervals that miss the actual value and targets mainly sharpness and reliability. QS evaluates specific quantile forecasts, providing insights into how well the forecast captures different points in the distribution and targets mainly reliability and resolution. $CRPS$ assesses the overall accuracy of the forecast distribution, integrating the entire probabilistic forecast into a single measure and targets mainly reliability and resolution.

For both deterministic and probabilistic forecasts, it is highly recommended to utilize skill scores, which compare the model under investigation with a reference model.^[21] The skill score is defined as:

$$ss = 1 - \frac{s_{\text{model}}}{s_{\text{benchmark}}} \quad (11)$$

where s represents a suitable metric, such as RMSE for deterministic forecasts and CRPS for probabilistic forecasts.^[18,19] Skill scores are essential for evaluating and comparing the performance of forecasting models, providing a standardized measure of how well a forecast performs relative to a benchmark. For deterministic forecasts, the reference model is often persistence,^[22] while for probabilistic forecasts, it is typically climatology model.^[23] By quantifying the improvement over simpler well-known models, skill scores help identify the effectiveness and reliability of advanced forecasting methods.

Bias, MAE, RMSE, IS, QS, CRPS, and skill scores, while useful to evaluate general forecast accuracy and reliability, are not suitable for evaluating ramp predictions due to their inability to specifically capture the timing and magnitude of sudden changes. These metrics focus on overall error and distributional accuracy, but ramps involve rapid and significant variations that require specialized metrics. Consequently, models evaluated and optimized solely by these metrics may appear performant yet fail to accurately predict critical ramp events. Specialized ramp detection metrics are necessary to address these shortcomings. As noted earlier, no standard and widely accepted metric exists in the literature to evaluate ramp forecast performance.^[5,14] In the following, a simple ramp metric is proposed. It can be easily applied to different use cases by using flexible thresholds.

2.2. Proposed Ramp Rate Metric

We propose a metric to evaluate the accuracy of solar irradiance forecasts based on the detection of ramp events. A predicted ramp event is a significant change in irradiance between two subsequent lead times (LT) within a forecast. The simplicity of this ramp definition lies in its straightforward comparison of the rate of change in irradiance to a threshold ϵ . This makes it easy to implement and understand. Thresholds ϵ can and should be adapted to the specific use cases reflecting varying degrees of significance of irradiance changes. The approach evaluates ramp events based on multistep predictions, where each prediction is made from the same origin point in time t_0 . To account for minor deviations in the exact timing of ramp events, a tolerance window is introduced. This window enhances the robustness of the approach in practical scenarios where exact timing alignment between observed and predicted ramps is challenging. Within this window, if an observed ramp and a predicted ramp occur at any time, they are considered a match. The equation to detect an observed ramp event is:

$$ORE_{t,LT} = \max_{j \in \{-\Delta t + res, -\Delta t + 2res, \dots, \Delta t\}} \left\{ \frac{\gamma(t + LT + j) - \gamma(t + LT + j - res)}{res} \right\} > \epsilon \quad (12)$$

In this equation, $ORE_{t,LT}$ represents the observed ramp event at time t and lead time LT . The term $\gamma(t)$ denotes the observed irradiance at time t . The tolerance window is Δt , the ramp event detection threshold is ϵ , and res is the corresponding temporal resolution.

Similarly, the equation to detect a predicted ramp event is:

$$PRE_{t,LT} = \max_{j \in \{-\Delta t + res, -\Delta t + 2res, \dots, \Delta t\}} \left\{ \frac{\hat{\gamma}(t + LT + j) - \hat{\gamma}(t + LT + j - res)}{res} \right\} > \epsilon \quad (13)$$

where $PRE_{t,LT}$ represents the predicted ramp event at time t and lead time LT . The term $\hat{\gamma}(t)$ denotes the predicted irradiance at time t . **Figure 1** illustrates a case of a correctly detected ramp event within the tolerance window.

True positive (TP), false negative (FN), false positive (FP), and true negative (TN) matches are defined according to Equation (14) through (17).

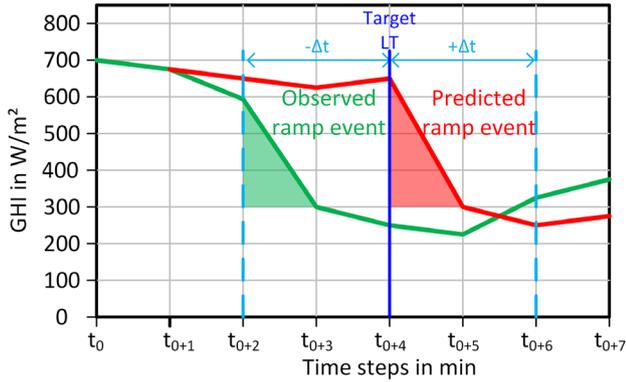


Figure 1. Observed and predicted ramp events for $LT = 4$ min, $\Delta t = \pm 2$ min and an ramp rate threshold of $\epsilon = 100$ W/(m²·min). This example illustrates a matching ramp event within the tolerance window.

$$TP_{LT} = \sum_{i=1}^n ORE_{i,LT} \wedge PRE_{i,LT} \quad (14)$$

$$FN_{LT} = \sum_{i=1}^n ORE_{i,LT} \wedge \neg PRE_{i,LT} \quad (15)$$

$$FP_{LT} = \sum_{i=1}^n \neg ORE_{i,LT} \wedge PRE_{i,LT} \quad (16)$$

$$TN_{LT} = \sum_{i=1}^n \neg ORE_{i,LT} \wedge \neg PRE_{i,LT} \quad (17)$$

Accuracy, Precision, Recall and F1 Score are calculated according to Equation (18)–(21).

$$Accuracy_{LT} = \frac{TP_{LT} + TN_{LT}}{TP_{LT} + TN_{LT} + FP_{LT} + FN_{LT}} \quad (18)$$

$$Precision_{LT} = \frac{TP_{LT}}{TP_{LT} + FP_{LT}} \quad (19)$$

$$Recall_{LT} = \frac{TP_{LT}}{TP_{LT} + FN_{LT}} \quad (20)$$

$$F1_{LT} = \frac{2 \cdot Precision_{LT} \cdot Recall_{LT}}{Precision_{LT} + Recall_{LT}} \quad (21)$$

3. Data Sets

At CIEMAT's Plataforma Solar de Almería, eight meteorological stations are distributed over an area of ≈ 1 km². Each station is equipped with ISO 9060 class A spectrally flat pyranometers and pyrhemometers, as well as automated solar trackers with sun sensors for measuring direct normal irradiance (DNI), global horizontal irradiance (GHI), and diffuse horizontal irradiance (DHI). Additionally, there are two ASIs on the site, positioned at the southwest and northeast ends. The ASI systems use off-the-shelf Mobotix Q25 cameras, which capture fully hemispherical images of the sky at a resolution of 2112×2048 pixels

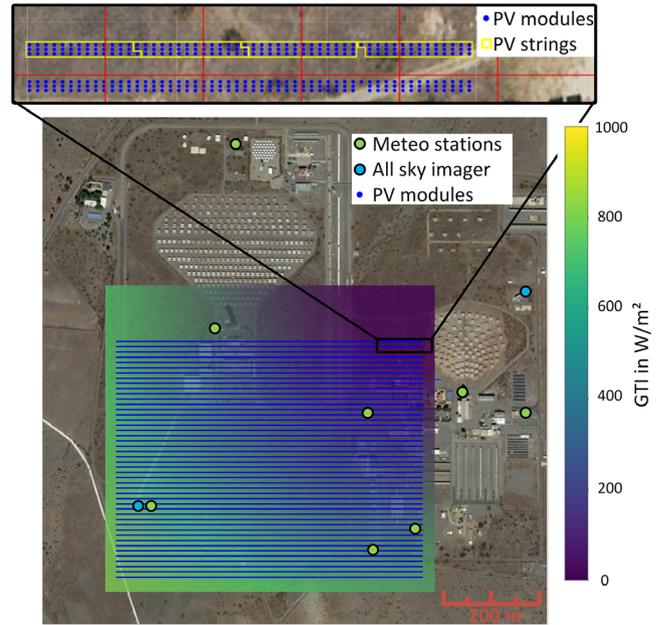


Figure 2. Positions of the measurement stations and of the simulated PV system at CIEMAT's PSA (latitude: 37.0927° N, longitude: -2.3607° E and altitude: 546 m). The plant consist of 78 400 modules (azimuth: 180° and tilt: 30°) with a maximum power of 21,128,800 W. Each yellow outline represents one string of PV modules.

at an exposure time of 160 μ s. Data is recorded with a temporal resolution of 1 min. **Figure 2** shows the location of each measurement station.

This study utilizes two datasets from these devices. The first dataset covers 359 days from September 1, 2020, to August 31, 2021, and is used to evaluate suitable ramp rate thresholds (Section 4). The second dataset spans 28 days between September and November 2019 and is employed to validate two ASI-based forecasting models using the proposed ramp rate metric (Section 5). This second dataset was introduced in a previous paper as a particularly interesting and complex benchmark dataset,^[24] which has since been used in several other benchmarks.^[9,14]

4. Ramp Rate Thresholds

The primary challenge for a simple ramp rate metric, as introduced in Section 2.2, is to determine an appropriate and flexible ramp threshold ϵ . This study is an attempt to address this challenge from an application perspective of ramp rate mitigation in PV plants. There are already a large number of grid codes in place around the world that have introduced ramp rate limits for PV power plants. A commonly used ramp rate limit is $\pm 10\%$ min of the plant's rated power.^[4] Such a PV power-based metric can be practical when the PV system specifications, such as total PV field area, tilt, tracking, etc. are known. However, these parameters vary across different applications, complicating the evaluation and comparison of forecast data. In contrast, GHI or GTI (Global Tilted Irradiance) are directly related to the physical input driving PV power production. Thus, they provide a

more fundamental basis for evaluating potential fluctuations in power output. Evaluating ramp events in terms of irradiance removes dependencies on specific PV system parameters, making it easier to compare forecasts and performance across different installations. For the evaluation of a suitable threshold, a PV model is implemented as described in Section 4.1. The actual evaluation is done in Section 4.2.

4.1. PV Model

A hypothetical PV power plant serves as a testbed to evaluate the applicability of different irradiance ramp thresholds, while complying with grid code specifications of $\pm 10\%$ min of the plant's rated power. The plant is modeled using pvlb, with the PV model originally developed and validated using production data from a real 22 MW power plant in Northwest Germany.^[4] This model was then adapted for further research on a hypothetical PV power plant located in southern Spain at CIEMAT's Plataforma Solar de Almería.^[25]

The final PV system consists of individual 78 400 modules. The modules parameters are selected from the California Energy Commission module database,^[26] with additional parameters obtained from the Sandia National Laboratories module database.^[27] These modules are connected into strings. Two strings are connected to each inverter. Each row of the array consists of four strings and two inverters. The simulation uses a simplified PVWatt inverter model.^[28] The general setup, as well as a detailed representation of the different strings, is shown in Figure 2. A full description of the model, including a complete list of parameters used, is available in ref. [25].

An estimation of GTI corresponding to the module orientation is required for the PV model. This GTI should match the spatial resolution of the solar field, with a corresponding irradiance value for each PV module. The procedure described in ref. [29] is used to determine the GTI according to Equation (22):

$$GTI(\delta, \phi) = DNI \cdot \cos(\theta(\delta, \phi)) + D_t + D_g \quad (22)$$

where θ is the incidence angle for the evaluated plane, δ is the tilt angle, ϕ is the azimuth angle, D_t is the diffuse sky irradiance in the evaluated plane, and D_g is the ground-reflected irradiance in the evaluated plane. D_t is calculated according to Equation (23):

$$D_t = DHI \cdot R_d \quad (23)$$

where, R_d is the transposition factor, determined using the Ma-Iqbal approach.^[30] Assuming isotropy of reflected irradiance, D_g is calculated from GHI and ground albedo (ρ) using Equation (24).^[30]

$$D_g = \frac{1 - \cos(\delta)}{2} \cdot \rho \cdot GHI \quad (24)$$

The GTI is calculated for each of the 8 meteorological stations. For every time step, a GTI map is created using the grid corresponding to the PV model (see Figure 2). A nearest interpolation with a Gaussian filter is applied to smooth transitions. This results in a GTI value for each grid point corresponding to the nearest station. On the basis of these maps, the cumulative

power for each inverter is calculated for each timestamp over the 359 days of the data set described in Section 3.

4.2. Evaluation Ramp Rate Thresholds

A total of 483 614 time steps of PV power output are modeled. In around 1.8% of the cases critical ramps with $\geq 10\%$ min of the rated power 21,128.8 kW were observed (see Figure 3). These critical ramps serve as the basis for determining suitable solar irradiance ramp thresholds for GHI, GHI Clear Sky Index (kGHI), and GTI (specific to the PV module plane).

Figure 4 shows the distribution of absolute irradiance ramps for GHI, kGHI, and GTI over all sun elevation angles and discretized over distinct bins of sun elevation angles for the 8754 critical PV ramp events. As expected, the mean and median ramps for the GHI increase with sun elevation. For kGHI and GTI, a more homogeneous distribution over distinct sun elevation angles is observed. However, an effect of the sun elevation angle is still visible.

In this study, the thresholds ϵ are chosen to optimize the F1 score. We also investigated the difference between a constant and a sun elevation dependent ϵ . All final thresholds are listed in Table 1 and 2.

TP, FN, and FP ramp events are described by the Equation (25)–(27):

$$TP = \sum_{i=1}^n OPVRE_i \wedge EIRE_i \quad (25)$$

$$FN = \sum_{i=1}^n OPVRE_i \wedge \neg EIRE_i \quad (26)$$

$$FP = \sum_{i=1}^n \neg OPVRE_i \wedge EIRE_i \quad (27)$$

where $OPVRE$ is the observed PV ramp event with a PV power output gradient $\geq 10\%$ min of the rated power. $EIRE$ is the expected irradiance ramp event with an irradiance gradient $\geq \epsilon$. TN ramp events do not need to be addressed as only the 8754 critical events already identified are of interest.

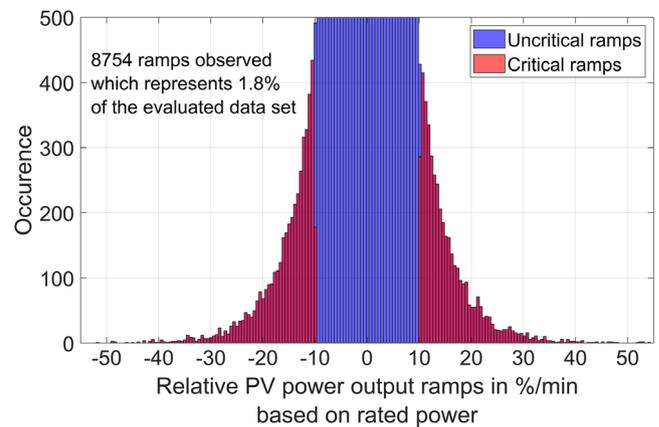


Figure 3. Observed PV power output ramps during the 359-day evaluation period.

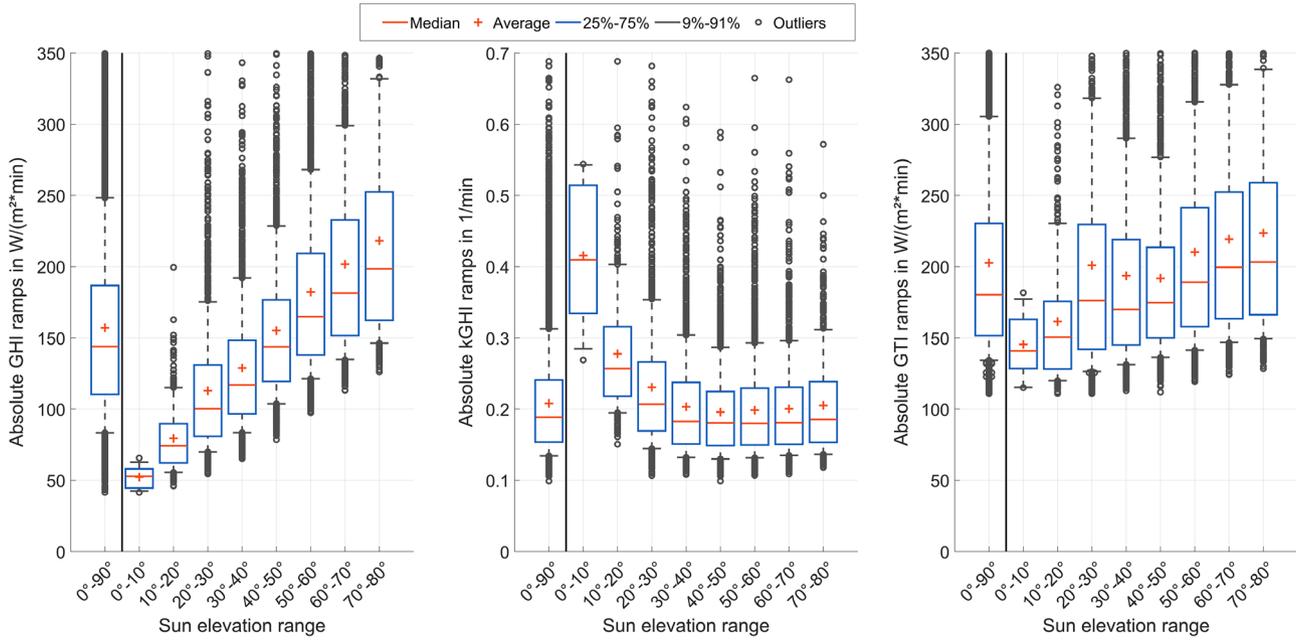


Figure 4. Observed absolute irradiance ramps corresponding to the previously identified 8754 critical PV power ramps.

Table 1. Optimized constant thresholds ϵ .

ϵ_{GHI}	ϵ_{kGHI}	ϵ_{GTI}
110 W m ⁻²	0.14	130 W m ⁻²

Table 3 summarizes the results. For GHI and kGHI-based thresholds, it is evident that incorporating a sun elevation-dependent ϵ is beneficial, as it results in an increase in the F1 score of $\approx 11\%$. Conversely, for GTI-based thresholds, the sun elevation has a relatively minor effect, with an improvement of $\approx 2.3\%$ in the F1 score. This is expected due to the relatively homogeneous distribution of the percentiles visible in Figure 4. Overall, the GTI-based thresholds are the most performant. This is not surprising, as the GTI is already aligned with the plane of the PV modules.

In the presented use case, GTI-based thresholds that account for the sun's elevation have proven effective, achieving an F1 score of 0.948. Practically, a GTI-based approach is highly feasible. Depending on the latitude, the forecasting system could consider the most reasonable orientations. For specific applications, the existing orientations of a power plant can be utilized, as this method only requires the orientation and does not involve complex PV modeling. For a more general approach, a GHI-based

threshold is recommended. It is advisable to select a threshold that depends on the sun's elevation. This recommendation applies to both GHI and kGHI.

5. Ramp Rate Validation Forecasting Approaches

Section 5.1 introduces the forecasting approaches used. Section 5.2 presents the validation results considering the novel ramp rate metric.

5.1. Used Forecasting Approaches

In the following, a state-of-the-art data-driven model optimized on RMSE (End2End model) as well as a novel generative forecasting model are presented.

5.1.1. End2End Model

The examined End2End model is a deep learning model as presented in ref. [9]. It is trained in an end-to-end manner hence, no physical processes like cloud dynamics or transmittance are explicitly modeled but the model learns the correlation of input and output data. Apart from sky images, the model is

Table 2. Optimized sun elevation dependent thresholds ϵ .

	0°-10°	10°-20°	20°-30°	30°-40°	40°-50°	50°-60°	60°-70°	70°-80°
ϵ_{GHI}	42 W m ⁻²	55 W m ⁻²	69 W m ⁻²	83 W m ⁻²	103 W m ⁻²	121 W m ⁻²	134 W m ⁻²	146 W m ⁻²
ϵ_{kGHI}	0.284	0.195	0.144	0.132	0.129	0.131	0.134	0.136
ϵ_{GTI}	111 W m ⁻²	115 W m ⁻²	121 W m ⁻²	127 W m ⁻²	132 W m ⁻²	137 W m ⁻²	142 W m ⁻²	145 W m ⁻²

Table 3. TP, FP, FN, and F1Score of detected critical PV power ramp events based on different threshold values ϵ for GHI, kGHI, and GTI. Both constant and sun elevation-dependent ϵ values are investigated. In addition, the absolute avg. PV power gradient is shown for both FP and FN ramp events. Ideally, both values should be close to the 10% min limit for critical ramp events. The color coding is individual for each column and ranges from dark green (best outcome) to dark red (worst outcome).

	TP critical ramps	FP critical ramps	FN critical ramps	F1 score	Absolute avg. PV power gradient for FP cases [% min ⁻¹]	Absolute avg. PV power gradient for FN cases [% min ⁻¹]
GHI	6597	1499	2157	0.783	8.86	12.26
kGHI	7574	3664	1180	0.758	6.07	10.87
GTI	8219	779	535	0.926	9.37	10.51
GHI: sun elevation based	8018	1730	736	0.867	7.74	10.69
kGHI: sun elevation based	8000	2145	754	0.847	7.44	10.68
GTI: sun elevation based	8316	482	438	0.948	9.51	10.39

provided with additional timeseries data, like past irradiance measurement. The output is a multi-step irradiance forecast for the location of the camera. Specifically, the model as part of this study predicts a scalar GHI and DNI value for each lead time from 1 to 20 min. To consider spatio-temporal dependencies of irradiance and cloud dynamics, the model integrates a transformer-based architecture. More precisely, it consists of two branches, each containing a separate transformer architecture, to extract features from images and timeseries data.

In the first branch, feature representations from time series data are learned using the model from ref. [31]. As input, 30-minute sequences of GHI, DNI, and DHI measurements and calculated sun elevation/azimuth values are used, allowing to learn temporal patterns. The length of the resulting representation vectors is a hyperparameter set to 512.

For image feature extraction, the timeSformer architecture is used.^[32] By applying a combined spatio-temporal attention mechanism, cloud behavior as observed by the ASI should be learned. Here, a shorter time window of 5 min is considered. While spatial attention is achieved through the underlying vision transformer architecture,^[33] temporal attention is applied over the same image patch of the whole image sequence, following the “divided space-time attention” method from ref. [32]. Like the time series representation, the length of the image sequence representation is set to 512.

To obtain the final multistep forecast both representation vectors are concatenated and fed into a multi-layer-perceptron (MLP). The multistep forecast for lead times 1 to 20 min is thus calculated at once for each timestep. A schematic of the entire approach is shown in Figure 5. As being trained end-to-end, both model branches are optimized in parallel without further pretraining.

As described in ref. [9], a nonparametric probabilistic quantile forecasting method based on ref. [11] can be applied to the End2End model to obtain probabilistic forecasts. Overall, the End2End model exhibits state-of-the-art performance, achieving average skill scores of 23.4% for deterministic forecasts and 55.1% for probabilistic forecasts up to a 20-minute horizon.^[9]

5.1.2. Generative Forecasting Model

The generative forecasting model (GFM) was previously presented in ref. [34]. It consists of a probabilistic generative model

and a deterministic irradiance model. First, the generative model predicts multiple continuations of previous frames. In the second step, the irradiance network is applied to the generated videos to predict the irradiance at each time step.^[34]

The generative model is implemented with a probabilistic denoising diffusion model,^[35] which is trained to incrementally remove noise from a noised image. During inference, the model can be recursively applied to randomly initialized images to generate photorealistic samples.^[35] In particular, a 2D U-net with spatial attention layers is used.^[36,37] Since this model typically expects images instead of videos, the past frames and the noise are concatenated in the channel dimension, the number of channels for the first convolutional layer of the U-net is adjusted, and all concatenated samples are predicted in the last layer. It uses a 64 × 64 pixel resolution and predicts eight different samples of the next five frames given the last six frames on a one frame per minute basis.

The irradiance forecast is then derived by processing each predicted synthetic sky image separately. It can be regarded as a solar estimation approach based on sky image data only. In this setup we use a CNN with ResNet34 backbone,^[38] to extract information from a single sky images and estimate the corresponding GHI and DNI. Since both processing steps (image generation and irradiance estimation) are independent, the irradiance model

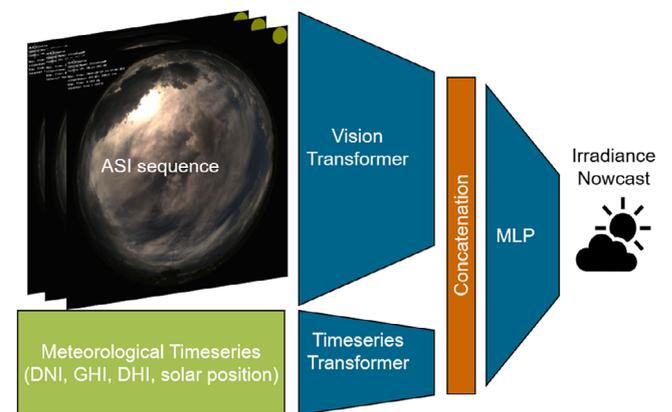


Figure 5. Schematic graph of generating forecasts with the End2End model (reproduced from ref. [9]).

is trained on real image data, separately from the diffusion model. However, Gaussian noise is added to the images and the images are resized a 64×64 pixel resolution to account for the characteristics of the synthetic images. After the training on real image data, the irradiance model is used in inference mode on the synthetic images to obtain the nowcasts. A schematic of the entire approach is shown in **Figure 6**. By processing multiple samples (N), an irradiance distribution can be derived for each lead time to create a probabilistic nowcast. In this study, eight samples are generated per lead time. A Gaussian distribution is used to generate the desired quantiles of a probabilistic nowcasts from the eight samples.

Compared to the End2End model, the GFM is not as prone to shortcut learning as discriminative models because the model has to learn the entire data generation process.^[39] Moreover, by using a probabilistic generative model, one can model the aleatoric uncertainty given by the ambiguity that multiple cloud changes are possible for a given state.

5.2. Validation Results

The validation is applied to GHI and GTI forecasts. Both forecasting systems provide DNI and GHI forecasts, allowing for the calculation of GTI forecasts according to Equation (22). Sun elevation-dependent ramp thresholds for GHI and GTI from Section 4.2 are utilized. In addition to deterministic forecasts, the quantiles 2.3, 15.85, 50, 84.15, and 97.7% of the respective probabilistic forecasts are considered.

The GFM is currently limited to a horizon of 5 min ahead, so the validation for both methods only considers multistep forecasts up to this timeframe. The tolerance window for the Ramp Rate metric is chosen to encompass all lead times up to 5 min ahead. GFM generates 8 samples per forecast and lead time. To validate the deterministic forecasts, each of the 8 samples is considered separately, and the validation results are averaged. The performance across samples is very consistent, with a

standard deviation in the lower tenths of a percent range for the F1 score.

For validation, the 28-day benchmark data set introduced in Section 3 is used. This validation dataset consists of the sky images from the southwest camera and the corresponding reference station (see Figure 2). The results of this validation are presented in **Table 4** and **5**. The GFM model generally outperforms the End2End model in terms of TP and FN, especially for deterministic forecasts, indicating superior ramp event detection. This trend extends to probabilistic forecasts, except at the extreme quantiles (2.3% and 97.7%) where the End2End model shows better TP and FN performance. However, this improvement is offset by significantly worse performance in FP and TN at these quantiles, leading to more false detections. For other quantiles, the End2End model excels in minimizing FP and maximizing TN. Overall, GFM maintains a good balance between FP and TN, showcasing robustness across variable conditions.

In terms of accuracy, the End2End model performs better overall, benefiting from the fact that only $\approx 13\%$ of the time stamps within the validation dataset include ramps. Therefore, the End2End model benefits from maximizing TN. However, accuracy significantly drops at the extreme quantiles (2.3% and 97.7%) due to increased false detections. High accuracy alone is insufficient, especially where TN is so dominant, as TN largely represents simple clear-sky conditions. Balancing precision and recall is crucial. Precision indicates the proportion of predicted positive cases that are actually positive, which is particularly important in scenarios where the cost of FP is high. Recall measures the proportion of actual positive cases that are correctly identified, which is crucial in situations where the cost of FN is high. The F1 score consistently favors GFM, indicating it effectively manages the trade-off between precision and recall. GFM not only detects true events but also minimizes false detections, making it more reliable for ramp predictions across different conditions.

Comparing GHI and GTI-based ramp predictions, GTI exhibits notably higher detection rates, as reflected by the F1

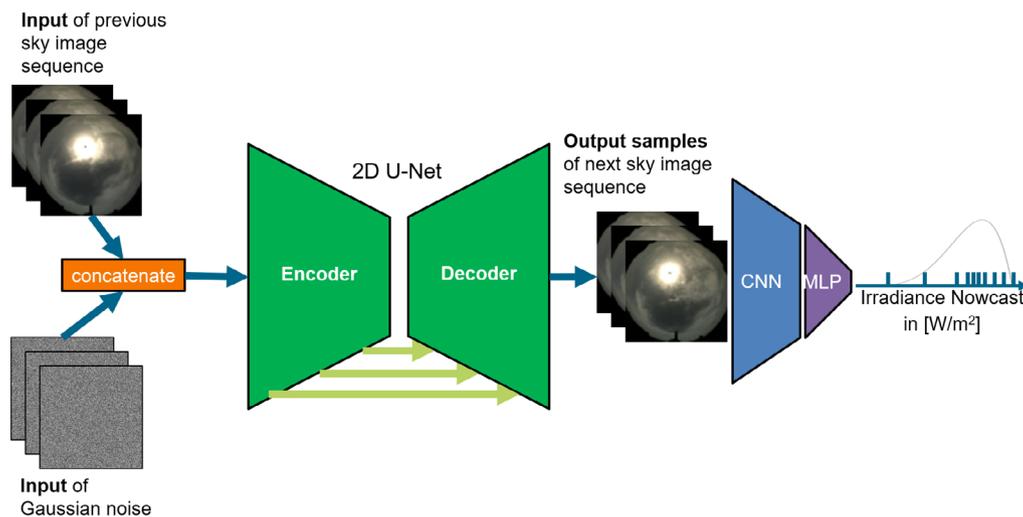


Figure 6. Schematic graph of generating forecasts with the generative model as well as CNN irradiance model. Per lead time, N samples of synthetic sky images and corresponding irradiance are generated.

Table 4. Observed TP, FN, FP, and TN ramp events for both forecasting methods. The results are discretized by deterministic and probabilistic methods as well as GHI and GTI. The color coding goes from dark green (best outcome) to dark red (worst outcome).

	TP				FN			
	End2End		GFM		End2End		GFM	
	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$
Deterministic	7.65%	33.74%	55.66%	76.11%	92.35%	66.26%	44.34%	23.89%
q2.30	81.79%	76.24%	62.01%	77.89%	18.21%	23.76%	37.99%	22.11%
q15.85	18.07%	51.08%	52.51%	72.04%	81.93%	48.92%	47.49%	27.96%
q50.00	7.65%	32.11%	43.67%	70.40%	92.35%	67.89%	56.33%	29.60%
q84.15	11.21%	51.99%	43.80%	71.31%	88.79%	48.01%	56.20%	28.69%
q97.70	84.83%	79.27%	51.45%	77.80%	15.17%	20.73%	48.55%	22.20%
	FP				TN			
	End2End		GFM		End2End		GFM	
	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$
Deterministic	0.12%	0.75%	6.56%	9.62%	99.88%	99.25%	93.44%	90.38%
q2.30	17.54%	15.43%	12.50%	9.78%	82.46%	84.57%	87.50%	90.22%
q15.85	0.57%	5.33%	7.75%	8.35%	99.43%	94.67%	92.25%	91.65%
q50.00	0.11%	0.54%	4.38%	8.15%	99.89%	99.46%	95.62%	91.85%
q84.15	0.35%	4.66%	3.43%	10.06%	99.65%	95.34%	96.57%	89.94%
q97.70	18.63%	13.72%	4.91%	13.67%	81.37%	86.28%	95.09%	86.33%

Table 5. Observed accuracy and F1 score for both forecasting methods. The results are discretized by deterministic and probabilistic methods as well as GHI and GTI. The color coding goes from dark green (best outcome) to dark red (worst outcome).

	Accuracy				F1			
	End2End		GFM		End2End		GFM	
	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$	$\epsilon(\text{GHI})$	$\epsilon(\text{GTI})$
Deterministic	93.88%	86.79%	90.98%	87.68%	0.14	0.49	0.45	0.70
q2.30	82.42%	82.98%	85.84%	87.89%	0.38	0.63	0.36	0.71
q15.85	94.13%	86.38%	89.66%	87.93%	0.29	0.59	0.40	0.69
q50.00	93.88%	86.65%	92.23%	87.78%	0.14	0.48	0.42	0.69
q84.15	93.89%	87.10%	93.13%	86.41%	0.19	0.61	0.45	0.67
q97.70	81.60%	84.95%	92.24%	84.71%	0.38	0.67	0.46	0.66

score. This is largely due to GTI's stronger sensitivity to the direct component of solar irradiance, resulting in more pronounced ramp amplitudes during cloud passing's.

This section has shown that forecasts with high performance for energy quantities can perform differently for ramp rates. Our novel metric has detected this difference well.

6. Conclusion

This study introduced a new ramp rate metric, designed for simplicity and adaptability by defining ramp events through straightforward thresholds and a tolerance window. This methodology is flexible across various forecasting scenarios and robust in detecting ramp events, accommodating minor timing deviations. By

adjusting the resolution parameter, the metric can be applied to datasets with varying time intervals between measurements.

A key challenge with a ramp rate metric is determining the appropriate ramp threshold, which should be tailored to the application at hand, as different applications have distinct requirements. Some applications may disregard ramps if only accurate energy quantities are necessary, while others prioritize accurate ramp forecasts that exceed a certain amplitude over-achieving high skill scores.

This study exemplifies the specific use case of ramp rate mitigation in PV systems, addressing a ramp rate limitation of $\pm 10\%$ min of the system's rated power, as outlined in many grid codes.^[4] This application necessitates high-resolution forecasts, ideally capable of directly predicting PV production. However, high-resolution PV production prediction requires

complex PV modeling. Evaluating ramp events in terms of irradiance simplifies usage and comparisons across different installations by removing dependencies on specific PV system parameters.

High-resolution modeled PV production data over 359 days were used to identify ideal GHI, kGHI, and GTI ramp rate thresholds for detecting critical ramps in PV production. Fixed GHI and kGHI ramp thresholds achieved F1 scores between 0.758 and 0.783, with kGHI performing better in precision (less critical FN events) and GHI performing better in recall (less critical FP events). Sun elevation-dependent GHI and kGHI ramp thresholds improved F1 scores from 0.847 to 0.867, balancing precision and recall. GTI-based thresholds performed best, with F1 scores of 0.926 and 0.948 for fixed and sun elevation-dependent thresholds, respectively. This result is unsurprising, given that the GTI is already aligned with the plane of the PV modules.

Although the presented thresholds were optimized for a specific application, they are generally applicable for evaluating the ramp prediction performance of different high-resolution forecasting systems using the proposed new metric at any arbitrary site.

To demonstrate the new metric two distinct ASI-based forecasting methods were evaluated for ramp detection suitability. The End2End model is a data-driven approach that combines a time series and vision transformers to derive irradiance. Specifically, it utilizes time series data on current irradiance and sun position, in addition to an ASI image series. The GFM employs a denoising diffusion model to generate multiple continuations of past frames and a CNN to predict irradiance from the generated synthetic sky images. Both models are capable of providing deterministic and probabilistic DNI and GHI forecasts. The DNI and GHI forecasts can be used to calculate GTI forecast. The End2End model represents the state-of-the-art, exhibiting high skill scores for both deterministic and probabilistic forecasts with horizons up to 20 min ahead.^[9] The GFM is a rather novel approach which considers cloud shape changes.^[34]

Both models undergo validation using GHI and GTI forecasts, incorporating sun elevation-dependent ramp thresholds and the proposed ramp rate metric, which considers TP, FN, FP, TN ramp events, along with overall accuracy and the F1 score. In terms of deterministic forecasts, the End2End model exhibits several characteristics. Firstly, it tends to miss most of the critical ramps, leading to low TP values and correspondingly high FN values. On the other hand, the model demonstrates very high TN values and correspondingly low FP values. Notably, since only $\approx 13\%$ of timestamps contain critical ramps, TN becomes particularly relevant for assessing accuracy. This leads to high overall accuracy but low F1 scores. This behavior is expected, as the End2End model is optimized to minimize RMSE, resulting in smoothing effects. While this makes the model well-suited for predicting accurate energy quantities, it is less effective for detecting actual ramp events. Probabilistic forecasts of the End2End model show improved F1 scores at lower and higher quantiles. The model can generally detect ramps, but the smoothing effect causes many ramps to fall below the critical thresholds. However, the lower and higher quantiles of the probabilistic forecasts amplify the ramps, making them partially detectable.

The GFM model is less affected by smoothing effects. Additionally, the GFM model accounts for the complex effects of cloud shape changes, making it particularly suitable for ramp detection. Validation results confirm this, with the GFM model demonstrating superior performance in TP and FN metrics and achieving higher F1 scores. This indicates its reliability in accurately detecting true ramp events and minimizing missed detections.

When comparing GHI and GTI-based ramp predictions, GTI shows notably higher ramp detection rates, as indicated by the F1 score. This is primarily due to GTI's greater sensitivity to the direct component of solar irradiance, resulting in more pronounced ramp amplitudes during cloud passages.

This study underscores the necessity of including specific ramp rate metrics in forecast model evaluations. Models assessed solely on RMSE and skill scores may conceal their inability to detect ramps effectively. Moreover, the potential of generative models like GFM is highlighted, though current limitations indicate a need for further improvements in image resolution and irradiance modeling. Future research will entail a comprehensive validation of an enhanced GFM model. The upgraded model will have the capability to generate multistep forecasts with a horizon of ≥ 30 min.

Acknowledgements

This research was funded by the German Federal Ministry for Economic Affairs and Climate Action through the Solrev2 project (grant agreement no. 03EE1195B) and by the German Federal Ministry for the Environment and Consumer Protection through the AuSeSol project (grant agreement no. 67K121007A), based on a decision by the German Bundestag.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

all sky imager, error metrics, forecasting, PV-production, ramp rate, solar irradiance

Received: June 28, 2024
Revised: September 17, 2024
Published online:

- [1] R. Perez, M. David, T. E. Hoff, M. Jamaly, S. Kivalov, J. Kleissl, M. Perez, *Found. Trend. Renewable Energy* **2016**, 1, 1.
- [2] A. Makibar, L. Narvarte, E. Lorenzo, *Sol. Energy* **2021**, 230, 435.
- [3] S. R. West, D. Rowe, S. Sayeef, A. Berry, *Sol. Energy* **2014**, 110, 188.
- [4] J. Schaible, B. Nouri, L. Höpken, T. Kotzab, M. Loevenich, N. Blum, S. Wilbert, *EPJ Photovolt.* **2024**, 15, 15.
- [5] Y. Chu, M. Li, C. F. Coimbra, D. Feng, H. Wang, *Iscience* **2021**, 24, 103136.

- [6] P. Kuhn, B. Nouri, S. Wilbert, C. Prah, N. Kozonek, T. Schmidt, R.R. Pitz-Paal, *Prog. Photovolt.: Res. Appl.* **2018**, 26, 608.
- [7] F. Lin, Y. Zhang, J. Wang, *Int. J. Forecast.* **2023**, 39, 244.
- [8] R. Marquez, C. F. Coimbra, *J. Sol. Energy Eng.* **2013**, 135, 011016.
- [9] Y. Fabel, B. Nouri, S. Wilbert, N. Blum, D. Schnaus, R. Triebel, L. F. Zarzalejo, E. Ugedo, J. Kowalski, R. Pitz-Paal, *Sol. RRL* **2024**, <https://doi.org/10.1002/solr.202300808>.
- [10] Q. Paletta, G. Arbod, J. Lasenby, *Appl. Energy* **2023**, 336, 120818.
- [11] B. Nouri, S. Wilbert, N. Blum, Y. Fabel, E. Lorenz, A. Hammer, R. Pitz-Paal, *Sol. Energy* **2023**, 253, 285.
- [12] L. Bianco, I. V. Djalalova, J. M. Wilczak, J. Cline, S. Calvert, E. Konopleva-Akish, J. Freedman, *Weather Forecast.* **2016**, 31, 1137.
- [13] C. Kamath, in *IEEE PES T&D 2010*, IEEE, Piscataway, NJ **2010**, pp. 1–6, <https://doi.org/10.1109/TDC.2010.5484508>.
- [14] S. A. Logothetis, V. Salamalikis, B. Nouri, J. Remund, L. F. Zarzalejo, Y. Xie, S. Wilbert, E. Ntavelis, J. Nou, N. Hendriks, L. Visser, M. Sengupta, M. Pó, R. Chauvin, S. Grieu, N. Blum, A. Kazantzidis, *Energies* **2022**, 15, 6191.
- [15] Y. Chu, H. T. Pedro, M. Li, C. F. Coimbra, *Sol. Energy* **2015**, 114, 91.
- [16] A. Florita, B. M. Hodge, K. Orwig, in *2013 IEEE Green Technologies Conf. (GreenTech)*, IEEE, Piscataway, NJ **2013**, pp. 147–152, <https://doi.org/10.1109/GreenTech.2013.30>
- [17] L. Vallance, B. Charbonnier, N. Paul, S. Dubost, P. Blanc, *Sol. Energy* **2017**, 150, 408.
- [18] D. Yang, S. Alessandrini, J. Antonanzas, F. Antonanzas-Torres, V. Badescu, H. G. Beyer, J. Zhang, *Sol. Energy* **2020**, 210, 20.
- [19] P. Lauret, M. David, P. Pinson, *Sol. Energy* **2019**, 194, 254.
- [20] M. Sengupta, A. Habte, S. Wilbert, C. Gueymard, J. Remund, Best practices handbook for the collection and use of solar resource data for solar energy applications (No. NREL/TP-5D00-77635). National Renewable Energy Lab.(NREL), Golden, CO (United States), **2021**.
- [21] T. Gneiting, A. E. Raftery, *J. Am. Stat. Assoc.* **2007**, 102, 359.
- [22] A. Kumler, Y. Xie, Y. Zhang, *Sol. Energy* **2019**, 177, 494.
- [23] J. L. G. La Salle, M. David, P. Lauret, *Sol. Energy* **2021**, 223, 398.
- [24] S. A. Logothetis, V. Salamalikis, S. Wilbert, J. Remund, L. F. Zarzalejo, Y. Xie, A. Kazantzidis, *Renewable Energy* **2022**, 199, 246.
- [25] L. Höpken, Reducing the Impact of Irradiance Ramps on PV Power Production—A Techno-Economic Analysis of Nowcasting (Master thesis, RWTH Aachen), **2023**, <https://elib.dlr.de/197700/>.
- [26] California Energy Commission. PV Module List - Full Data. <https://www.energy.ca.gov/media/2367> (accessed: 12 May 2023).
- [27] J. A. Kratochvil, W. E. Boyson, D. L. King, Photovoltaic array performance model (No. SAND2004-3535). Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA (United States) **2004**, <https://doi.org/10.2172/919131>.
- [28] A. P. Dobos, PVWatts version 5 manual (No. NREL/TP-6A20-62641). National Renewable Energy Lab.(NREL), Golden, CO (United States) **2014**, <https://doi.org/10.2172/1158421>.
- [29] N. B. Blum, S. Wilbert, B. Nouri, J. Lezaca, D. Hucklebrink, A. Kazantzidis, R. Pitz-Paal, *Sol. Energy* **2022**, 232, 232.
- [30] C. Demain, M. Journée, C. Bertrand, *Renewable Energy* **2013**, 50, 710.
- [31] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, C. Eickhoff, in *Proc. of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* **2021**, pp. 2114–2124, <https://doi.org/10.1145/3447548.3467401>.
- [32] G. Bertasius, H. Wang, L. Torresani, in *Int. Conf. on Machine Learning* **2021**, Vol. 139, p. 139, <https://doi.org/10.48550/arXiv.2102.05095>.
- [33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, N. Houslsby, An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, **2020**, <https://doi.org/10.48550/arXiv.2010.11929>.
- [34] Y. Fabel, D. Schnaus, B. Nouri, S. Wilbert, N. Blum, L. F. Zarzalejo, R. Pitz-Paal, in *IEA Workshop on Minutescale Forecasting for the Weather-Driven Energy System*, Roskilde, Denmark **2024**, <https://elib.dlr.de/204099/>.
- [35] J. Ho, A. Jain, P. Abbeel, *Adv. Neural Inf. Process. Syst.* **2020**, 33, 6840.
- [36] O. Ronneberger, P. Fischer, T. Brox, in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference*, Munich, Germany, 5–9 October 2015, proceedings, part III 18, Springer International Publishing, New York **2015**, pp. 234–241.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, I. Polosukhin, Attention is all you need. *Advances in neural information processing systems*, 30, **2017**.
- [38] K. He, X. Zhang, S. Ren, J. Sun, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE, Piscataway, NJ **2016**, pp. 770–778.
- [39] R. Geirhos, J. H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, F. A. Wichmann, *Nat. Mach. Intell.* **2020**, 2, 665.