TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Lanzilao, L., Pinto, F., & Meyer, A. (2026). Characterization and forecasting of national-scale solar power ramp events. *Energy and AI*, *26*, Article 100918. https://doi.org/10.1016/j.egyai.2026.100918

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Energy and AI 26 (2026) 100918

Contents lists available at ScienceDirect

Energy and AI

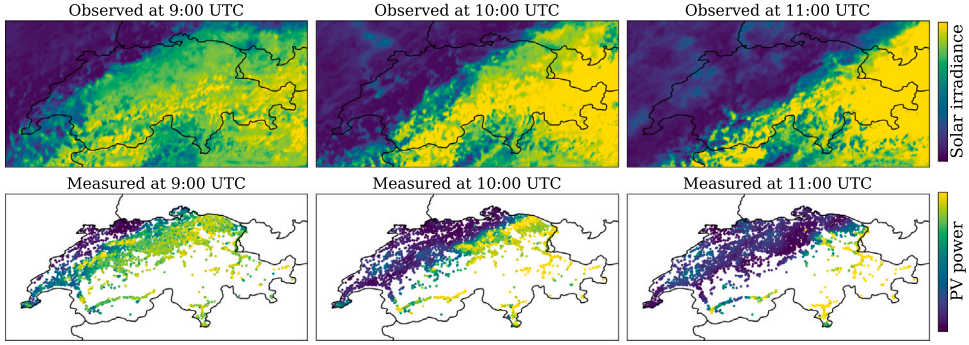
journal homepage: www.elsevier.com/locate/egyai

Characterization and forecasting of national-scale solar power ramp events

Luca Lanzilao [a,*], Francesco Pinto [a,b], Angela Meyer [a,b]

[a] School of Engineering and Computer Science, Bern University of Applied Sciences, Quellgasse 21, Biel, 2501, Bern, Switzerland
[b] Department of Geosciences and Remote Sensing, TU Delft, Stevinweg 1, Delft, 2628 CN, South-Holland, The Netherlands

GRAPHICAL ABSTRACT



HIGHLIGHTS

- Solar ramp events peak during the spring and summer months.
- Solar ramp events are driven by morning cloud clearing and afternoon cloud formation.
- SHADECast outperforms optical-flow and physics-based models during most ramp events.
- Forecast RMSE increases by up to 50% during solar ramp events at a 2-hour lead time.
- IFS outperforms nowcast models for extreme ramps, where training data are scarce.

ARTICLE INFO

Keywords:
Solar energy nowcasting
Solar ramp events
Spatiotemporal forecasting
Numerical weather prediction models
Photovoltaic energy

ABSTRACT

The rapid growth of solar energy is reshaping power system operations and increasing the complexity of grid management. As photovoltaic (PV) capacity expands, short-term fluctuations in PV generation introduce substantial operational uncertainty. At the same time, solar power ramp events intensify risks of grid instability and unplanned outages due to sudden large power fluctuations. Accurate identification, forecasting and mitigation of solar ramp events are therefore critical to maintaining grid stability. In this study, we characterize national-scale solar PV ramp events based on two years of PV power production from 6434 PV stations at 15-minute resolution. We develop quantitative metrics to define solar ramp events and systematically characterize their occurrence, frequency, and magnitude at a national scale. Furthermore, we examine the meteorological drivers of ramp events, highlighting the role of mesoscale cloud systems. In particular, we observe that ramp-up events are typically associated with cloud dissipation during the morning, while ramp-down events commonly occur when cloud cover increases in the afternoon. Additionally, we adopt a recently developed spatiotemporal forecasting framework to evaluate both deterministic and probabilistic PV power forecasts derived from deep learning and physics-based models, including SolarSTEPS, SHADECast, IrradianceNet, and Integrated

* Corresponding author.
E-mail address: luca.lanzilao@gmail.com (L. Lanzilao).

https://doi.org/10.1016/j.egyai.2026.100918
Received 27 March 2026; Received in revised form 14 September 2026; Accepted 27 September 2026
Available online 10 October 2026
2666-5468/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Forecasting System Ensemble (IFS-ENS). The results show that the generative nowcast model, SHADECast, is the most reliable model for most ramps, achieving a Continuous Ranked Probability Score 10.8% lower than that of SolarSTEPS at a two-hour lead time, whereas IFS-ENS outperforms the satellite-based models for the most extreme ramp events. State-of-the-art nowcasting models struggle to capture ramp dynamics, with forecast root mean squared error (RMSE) increasing by up to 50% compared to normal operating conditions. Overall, these results emphasize the need for improved high-resolution spatiotemporal modeling to enhance ramp prediction skill and support the reliable integration of large-scale solar generation into power systems.

1. Introduction

Rapid fluctuations in surface solar irradiance (SSI) at local to regional scales can lead to pronounced variability in photovoltaic (PV) power output. This variability, largely governed by the dynamic evolution and motion of cloud systems, represents a major challenge for power system operation and the continued expansion of solar energy [1,2]. The problem is especially pronounced for large utility-scale PV plants and regions with a high spatial density of distributed installations, where irradiance variations can translate into significant aggregate power ramps [3]. As solar penetration continues to grow, such ramp events are expected to occur more often and with greater magnitude [4]. These rapid transitions increase the complexity of maintaining grid stability, scheduling reserves, and operating in electricity markets. Consequently, reliable detection and short-term prediction of solar ramp events are crucial for improving operational flexibility, minimizing balancing requirements, and enabling a more efficient coordination between generation assets, grid infrastructure, and electricity demand [5,6].

Prior work on solar ramp events has primarily focused on identifying their atmospheric drivers and developing operational countermeasures [7]. These approaches, including model predictive control [8] and robust optimization strategies [9], are essential for mitigating rapid power fluctuations caused by ramp events that may threaten grid stability [10–12]. Observational studies employing all-sky cameras [13–15], Doppler radar measurements, and satellite imagery have investigated cloud processes responsible for abrupt irradiance fluctuations [16]. Anticipating such rapid variability through intra-hour forecasting has therefore become a central research objective [6,17,18]. Numerous approaches have been proposed, ranging from numerical weather prediction (NWP) systems to satellite-based retrievals, ground-based sensing networks, and purely data-driven or hybrid modeling frameworks [19]. The effectiveness of these methods strongly depends on the spatial and temporal scales of interest, with recent advances in deep learning-based spatiotemporal architectures yielding notable improvements in short-term irradiance prediction, particularly when applied to satellite-derived fields [20,21].

Despite these advances, many existing studies rely on geographically limited case studies or relatively small fleets of PV installations. For instance, [22] analyzed data from eight meteorological stations in Spain to introduce a novel ramp-rate metric designed to evaluate ramp events between successive forecast lead times. Similarly, [7] proposed an automated ramp detection framework based on a single-parameter swinging door algorithm and introduced a dedicated assessment metric using observations from two solar stations in Hawaii. More recently, optical flow techniques and machine learning approaches have been increasingly adopted for forecasting ramp-up and ramp-down events. [23] employed an all-sky imager in combination with in situ sensor measurements and external meteorological data. They extracted image-derived features such as cloud pixel fraction, brightness, edges, and corner points, and applied optical flow methods to estimate cloud motion and velocity. In a related study, [13] developed a multi-step forecasting framework for PV ramp-rate control, utilizing a ResNet-18 model for time-series irradiance prediction and incorporating stacked sky images to better capture cloud dynamics and improve short-term forecast accuracy. Likewise, [24] implemented a low-cost sky-imaging camera network coupled with neural networks to predict very short-term solar irradiance ramp events at two locations in California, USA.

Despite the rapidly increasing penetration of PV systems and their growing importance for distribution and transmission network operations, solar ramp events have received comparatively less attention than wind ramp events, particularly at large spatial scales [6,25]. Prior solar-ramp studies have focused predominantly on individual sites or small PV station networks and do not provide a comprehensive comparison of different forecast approaches, including NWP, optical flow models, convolutional and generative spatiotemporal next-frame prediction models [26]. To address this gap, we extend ramp characterization and forecast evaluation to nationally aggregated PV generation based on two years of operational PV power measurements at 15-minute resolution from 6434 PV systems distributed across Switzerland, representing, to the best of our knowledge, the first national-scale study of solar ramp events based on PV generation data. We introduce a systematic methodology to detect ramp events, quantify their frequency and occurrence. Moreover, we use satellite-derived SSI fields to characterize the meteorological conditions that promote the development of ramp events. Furthermore, we adopt the spatiotemporal PV forecasting framework developed by [21] to benchmark multiple forecasting models. The comparison is explicitly conducted in the presence of ramp and non-ramp events, enabling us to assess not only overall forecast skill but also reliability, sharpness, and degradation of performance during highly variable weather conditions. The evaluated satellite-based models include the probabilistic optical-flow model SolarSTEPS [27], the probabilistic deep generative diffusion model SHADECast [20], and the deterministic deep learning model IrradianceNet [28]. Additionally, we also include the high-resolution integrated forecast system (IFS-ENS) model developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), widely used for SSI forecasting in operational practice [29–31].

The main novel contributions of this study are as follows:

- we characterize national-scale solar PV ramp events across more than 6400 PV systems,
- we assess how spatiotemporal satellite-based and IFS forecast models perform during ramp and non-ramp conditions,
- we demonstrate that the generative nowcast model outperforms the other satellite-based models and the NWP for most ramps, whereas the NWP outperforms the satellite-based models for the most extreme ramps, and
- we visualize the cloud radiative footprint of mesoscale cloud systems on countrywide PV production.

The remainder of this article is structured as follows. Section 2 describes the satellite-derived SSI dataset, the PV power dataset, and the forecasting models. Section 3 introduces the ramp detection methodology and analyzes their frequency, spatial characteristics, and meteorological drivers. Section 4 details the intraday forecasting framework for both SSI and PV power. Section 5 evaluates the forecast skill of the benchmarked models, discussing their performance under both ramp and non-ramp conditions. Finally, Section 6 summarizes the main findings and outlines future research directions. In this study, we use the terms *solar ramp event* and *PV ramp event* interchangeably to refer to rapid changes in the nationally aggregated PV power production. Likewise, the terms *solar power* and *PV power* are used interchangeably to refer to solar photovoltaic power.

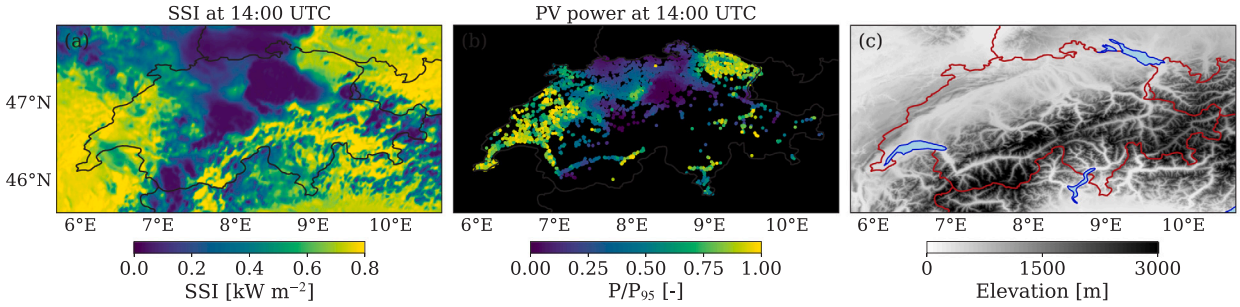


Fig. 1. (a) Satellite-based SSI field and (b) PV power output of the 6434 PV stations, observed on 26 June 2020 at 14:00 UTC. Note that the PV power output is normalized using the station-specific 95th percentile of the power time series. (c) Elevation map showing values in meters above sea level. The black and red lines denote national borders.

2. Datasets and models

Our analysis relies on satellite-derived SSI observations and operational PV power measurements. The SSI data are obtained from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard the Meteosat Second Generation (MSG) satellite, positioned in geostationary orbit at 0° longitude [32]. SEVIRI provides full-disk Earth scans with a nadir spatial sampling distance of approximately 3 km. The irradiance fields are retrieved from the EUMETSAT Climate Monitoring Satellite Application Facility (CM SAF), specifically from the High-Resolution European Surface Solar Radiation Data Record (HANNA) [33]. The dataset is projected onto a regular grid with a spatial resolution of 0.01° in both longitude and latitude, covering the domain [15°W, 35°E] × [30°N, 70°N]. The SSI fields are available for 2019–2020 at a temporal resolution of 15 min [33]. Although the temporal coverage of the HANNA product is limited, we chose HANNA over the Surface Radiation Dataset Heliosat (SARAH-3) because the latter underestimates SSI values in snow-covered regions, such as the Alpine arc [27,34]. Fig. 1(a) displays the HANNA SSI field observed on 26 June 2020 at 14:00 UTC over the domain of interest. The high spatial resolution allows for a sharp representation of cloud structures, identifiable by their radiative effect on the surface as areas with lower SSI values.

The PV data consist of 15-minute production time series from 6434 operational PV systems distributed across Switzerland. The proprietary production data were obtained under a dedicated data usage agreement. A detailed description of the quality control and filtering procedures applied to the raw measurements is provided in [21]. Assuming an average of 12 daylight hours per day and considering the two-year analysis period (i.e. 2019–2020), the dataset comprises approximately 225 million PV production data points. To enable consistent comparison across installations with different nominal capacities, PV power at each station is normalized with P_{95}, defined as the station-specific 95th percentile of its power time series. This normalization ensures comparability across systems and reduces sensitivity to rare extreme outliers. Although normalization by nominal capacity would be a natural alternative, this information is not available. The majority of stations (approximately 76%) exhibit P_{95} values between 10 and 200 kW, while only 4% exceed 300 kW, indicating that the dataset is dominated by small- to medium-scale installations. Fig. 1(b) illustrates the locations and corresponding power output of the 6434 PV stations included in this study, recorded on 26 June 2020 at 14:00 UTC while Fig. 1(c) shows the elevation map over the domain of interest. Most PV installations are located in the Swiss Plateau or in valleys within the Alpine region. The cloud patterns visible in Fig. 1(a) are mirrored in the PV power measurements. In fact, regions with reduced PV power output correspond closely to areas with low SSI, highlighting the strong relationship between solar irradiance and power generation. This demonstrates that geostationary satellite observations can effectively capture the spatial footprint of cloud cover and its immediate impact on country-wide solar energy production, providing both high spatial accuracy and low-latency information.

The forecasting models evaluated in this study include the probabilistic optical-flow model SolarSTEPS [27], the probabilistic deep generative diffusion model SHADECast [20], the deterministic deep learning model IrradianceNet [28], and the physics-based IFS-ENS model [35]. SolarSTEPS is a probabilistic nowcasting framework for satellite-derived clear-sky index (CSI) fields introduced by [27]. The model first estimates cloud motion vectors (CMVs) using the Lucas–Kanade optical flow method. CSI fields are then decomposed into spatial scales via Fast Fourier Transform, and each scale is forecast using a separate autoregressive (AR) model with spatially correlated noise to generate ensemble members. The resulting components are subsequently recombined and advected to produce the final probabilistic output. In this work, we adopt the same architecture and pre-trained weights as described in [27]. We additionally evaluate a simplified variant, SolarSTEPS-pa, representing a pure advection (PA) approach. Unlike the full SolarSTEPS framework, this version does not model temporal variability in CSI. Instead, it perturbs the estimated CMVs and advects the CSI fields accordingly. Consequently, SolarSTEPS-pa isolates the contribution of cloud advection while omitting scale decomposition and AR-based temporal modeling. IrradianceNet is a deterministic spatiotemporal deep-learning model based on Convolutional Long Short-Term Memory (ConvLSTM) layers, developed by [28]. It forecasts satellite-derived CSI fields using only past CSI inputs. Originally trained on SARAH-2.1 and subsequently retrained on HelioMont data [36,37] with an extended autoregressive forecast horizon, it serves as a strong deterministic baseline. However, it does not provide probabilistic forecasts or uncertainty quantification with reasonable computational effort. We adopt the same architecture and pre-trained weights as described in [20]. SHADECast is a deep generative diffusion model for probabilistic intraday solar forecasting [20]. The framework first compresses CSI sequences using a variational autoencoder, predicts their deterministic evolution through a latent nowcaster based on Adaptive Fourier Neural Operator blocks and a temporal transformer, and subsequently applies a conditional latent diffusion model to generate ensemble forecasts. The latent outputs are finally decoded into physical CSI fields. We use the architecture and pre-trained weights described in [20].

To benchmark these satellite-based methods against a physics-based approach, we include probabilistic SSI forecasts from the physics-based IFS-ENS model [35]. Ten randomly selected ensemble members at approximately 9 km spatial resolution are used to ensure consistency with the ensemble size of the satellite-based probabilistic models. Additionally, we apply a machine learning-based bias correction using a U-Net architecture, referring to the corrected forecasts as IFS-ENS-corrected. Further details on the correction methodology and training procedure are provided in [21]. Overall, seven distinct methods are considered, covering both deterministic and probabilistic forecasting approaches. It is important to emphasize that we do not develop

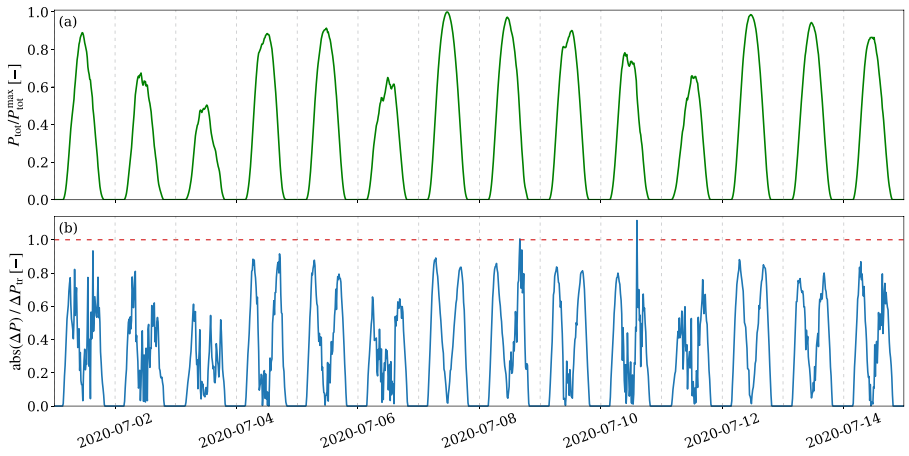


Fig. 2. (a) Time series of aggregated PV power production from 6434 PV stations, normalized by the maximum value observed over the displayed period, and (b) the corresponding magnitude of power variations, normalized by the ramp threshold ΔP_{tr}. Results are shown for the first two weeks of July 2020. The horizontal dashed red line indicates the threshold used to identify ramp events, while the vertical dashed gray lines mark midnight of each day.

specialized ramp-event forecasting models. Instead, our objective is to assess how state-of-the-art generic spatiotemporal SSI forecasting models perform in capturing solar ramp events.

3. Solar ramp events

Ramp events do not have a universally accepted definition, as their characterization depends on the specific application and operational context. Ramp identification relies on several key attributes, including the duration Δt, magnitude of the power variation ΔP, rate of change $\Delta P/\Delta t$, and the direction of the transition [30]. Ramp direction distinguishes between increases and decreases in power output. Positive ramps (ramp-up events) correspond to periods of rapidly increasing generation, whereas negative ramps (ramp-down events) denote sharp decreases in power production. The severity of a ramp event is typically determined by the interplay between its magnitude and duration. Large power variations occurring over short time intervals are generally considered more critical, as they imply higher rates of change and pose greater operational challenges for grid management.

In this study, the ramp duration is fixed to the temporal resolution of both the SSI and PV datasets, that is, 15 min. This choice is consistent with the time granularity adopted in electricity market operations [38]. The primary challenge lies in defining an appropriate threshold for ΔP to distinguish ramp events from normal variability. To address this, we adopt a data-driven approach. First, satellite observations are used to compute the spatially averaged CSI over the study domain between sunrise and sunset for each day within the two-year study period. From this distribution, we identify a subset of 43 days characterized by predominantly clear-sky conditions, defined as those with a mean CSI between sunrise and sunset exceeding the 90th percentile. Next, we aggregate the PV power production of all 6434 PV systems into a single country-level generation time series and compute the maximum 15-minute power variation observed during these clear-sky days. This value represents the upper bound of power fluctuations in clear-sky days, i.e., in the absence of cloud-induced variability and solely driven by deterministic diurnal evolution. The resulting threshold, $\Delta P_{tr} = 37.7$ MW, is adopted to differentiate between non-ramp and ramp events. This corresponds to a rate of change of 2.51 MW min^{-1}. Power variations exceeding ΔP_{tr} over a 15-minute interval are therefore classified as ramp events, as they exceed the variability expected under clear-sky conditions and are attributed to transient atmospheric phenomena such as cloud dynamics. We note that this threshold corresponds to approximately 7.1% of the aggregated P_{95} values of all PV systems considered. Importantly, this threshold is dataset-specific and depends on the spatial aggregation level, and therefore must be determined for the particular set of monitored PV systems. Section 5.2 assesses the sensitivity of the analysis to this value and to the temporal aggregation window. The threshold itself proves stable across alternative constructions, whereas the number of detected events and, above approximately 40 MW, the relative ranking of the forecasting methods depend on the threshold value.

The aggregated PV power generation of the 6434 PV stations considered in this study is shown in Fig. 2(a) for the first two weeks of July 2020. The total PV power output exhibits a pronounced diurnal cycle, with day-to-day differences in peak production driven mainly by varying cloud conditions. Periods of reduced peak power indicate the presence of extended cloud cover, while smoother and higher daytime maxima correspond to predominantly clear-sky conditions. The associated power variations over the same time period are shown in Fig. 2(b). High values of $|\Delta P|/\Delta P_{tr}$ occur primarily during morning and afternoon in clear-sky days, whereas rapid fluctuations in PV power output on cloudy or highly variable days are driven by transient cloud passages. Conversely, low values indicate stable production regimes under slowly varying irradiance conditions. Power variations exceeding the defined threshold (red dashed line) identify ramp events, highlighting time intervals with extreme short-term variability.

The distribution of the 15-minute power variations computed from the aggregated PV generation over the two-year time span is shown in Fig. 3(a). The distribution is strongly skewed toward small amplitudes, indicating that most power variations remain well below the ΔP_{tr} threshold. The frequency decreases nearly monotonically with increasing ΔP. In total, we identify 74 events with $\Delta P > \Delta P_{tr}$. We refer to these as ramp events, which may be considered extreme because they impose relevant stress on the power grid. These events reside in the upper tail of the distribution, highlighting the comparatively low probability of large-amplitude ramps. Due to their low frequency, such events are inherently challenging for machine learning models to represent accurately [39]. Fig. 3(b) shows the diurnal distribution of the 74 ramp events. Ramp-up events occur predominantly during the morning (approximately 07:00–10:00 UTC), whereas ramp-down events are concentrated in the afternoon (approximately 13:00–16:00 UTC). This asymmetry reflects systematic differences in the physical drivers of strong positive and negative PV power changes. As will be discussed later, ramp-up events are typically associated with cloud dissipation during the morning, while ramp-down events commonly occur when cloud cover increases in the afternoon. In fact, these processes reinforce the underlying diurnal cycle of clear-sky SSI, amplifying the increase in solar irradiance in the morning and the decrease in the afternoon.

The seasonal distribution of aggregated PV power variations as a function of the time of day using hexagonal binning is illustrated in

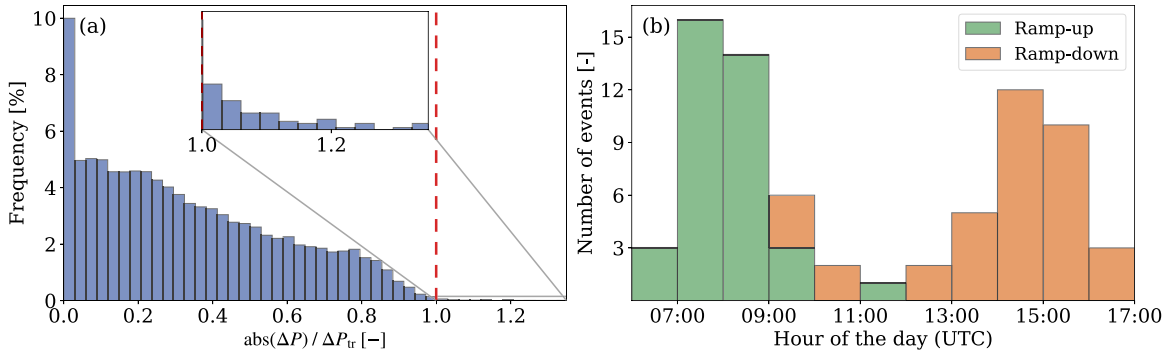


Fig. 3. (a) Empirical probability density of magnitude of the power variation over a time interval of 15 min normalized by the threshold ΔP_{tr}, measured over the year 2019–2020 for the aggregate PV power generation of 6434 PV stations. (b) Number of ramp events aggregated by hour of the day, with distinction between ramp-up and ramp-down events.

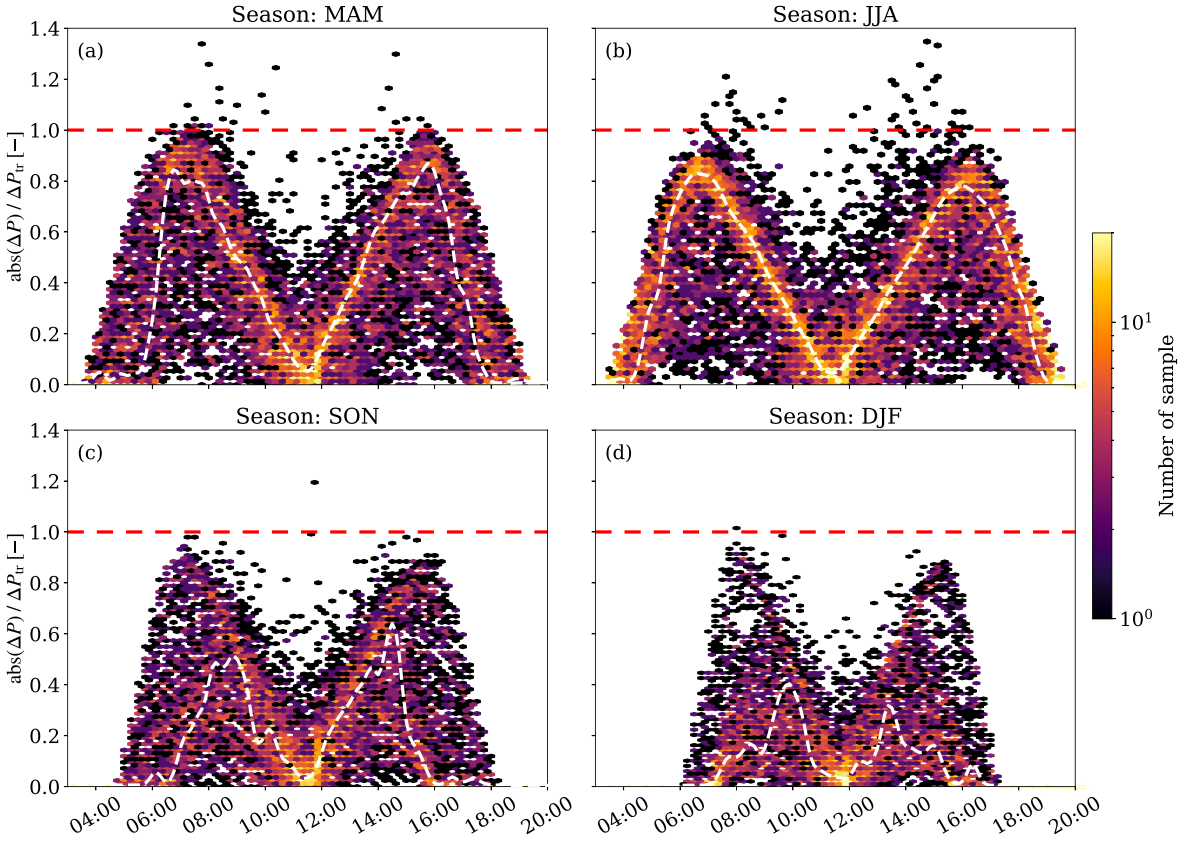


Fig. 4. Hexagonal bin distribution of ramp magnitude normalized with the threshold ΔP_{tr} as a function of hour of the day for (a) MAM, (b) JJA, (c) SON, and (d) DJF. Color shading indicates the number of samples per hexagonal bin on a logarithmic scale. The dashed white line traces the locus of maximum sample density at each hour, highlighting the most probable ramp magnitude throughout the diurnal cycle. The dashed red line marks the threshold used to identify ramp events. The bin widths are 15 min along the x-axis and 0.028 in normalized ramp units along the y-axis.

Fig. 4. The color scale indicates the number of samples within each hexagonal bin, thereby visualizing the density of occurrences. Across all seasons, most samples lie below the threshold ΔP_{tr}, confirming that ramp events are rare and confined to the upper tail of the distribution. A pronounced diurnal structure is evident in every season. Large power variations are clustered around the morning and afternoon hours, whereas variations near solar noon are typically smaller, reflecting comparatively stable generation conditions. The dashed white line indicates the regions of highest occurrence density. The latter mostly follow the derivative of the diurnal evolution of SSI in clear-sky days, where the largest gradients occur during the morning rise and afternoon decline of the sun in the sky. Seasonal differences are clearly visible. In spring (MAM) and summer (JJA), the distribution spans a broader temporal range, consistent with longer daylight duration and higher irradiance levels. In autumn (SON), the temporal window narrows and the density shifts toward moderately lower magnitudes. During winter (DJF), both the occurrence density and the amplitude of power variations are substantially reduced, and events are confined to a short midday interval, reflecting limited daylight hours and lower solar elevation. Ramp events occur predominantly in the MAM and JJA months, and are most frequent during the morning and afternoon hours, consistent with the statistics shown in Fig. 3(b).

Finally, we present an example of cloud dynamics leading to a ramp event. Specifically, Fig. 5(a–c) shows the satellite-derived SSI fields at 09:00, 10:00, and 11:00 UTC, while Fig. 5(d–f) presents the corresponding PV power production at the same times. At 09:00 UTC, high SSI

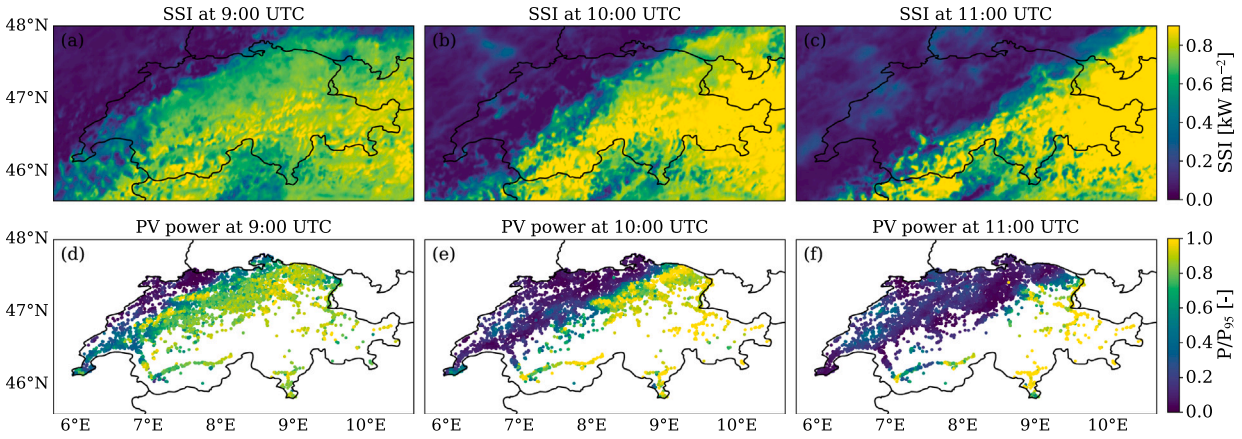


Fig. 5. (a–c) Satellite-based SSI fields and (d–f) PV power output of the 6434 PV stations, observed on 23 May 2020 at three time stamps: 09:00 UTC, 10:00 UTC and 11:00 UTC. The black lines denote national borders. Note that the PV power output is normalized using the station-specific 95th percentile of the power time series.

values prevail over large parts of the region of interest, consistent with elevated PV power production across most stations. Between 09:00 and 11:00 UTC, a coherent cloud front is advected to the southeast, covering the majority of the Swiss Plateau and causing a significant reduction in SSI. The distributed PV power production reveals a corresponding spatially coherent decline in PV power generation in the areas affected by the cloud band. The temporal sequence highlights the strong coupling between satellite-observed irradiance variability and the response of distributed PV systems. This event illustrates how the passage of a cloud band over a region with a high density of PV installations can induce a rapid decline in PV power production. In fact, between 09:00 and 11:00 UTC, the national PV power output decreased by approximately 60.7%.

At individual PV plants, grid operators may impose ramp-rate limits to restrict rapid changes in power injection. Ramp-rate limits are typically expressed in MW/min or % of rated capacity/min and can be enforced by plant or inverter controllers or curtailment. Such plant-level ramp-rate control can smooth fluctuations in the production of individual PV plants and thereby reduce their contribution to ramps in nationally aggregated PV production. Although such limits do not directly apply to nationally aggregated PV generation as investigated in our study, they illustrate the operational challenge posed by rapid power changes.

In this study, we detect ramp events directly from the nationally aggregated power generated by all 6434 PV stations. Therefore, if individual plants actively limit their power ramps, both the individual and nationally aggregated PV generation already contain the effects of that ramp-rate control. The ramps reported here are therefore not solely meteorologically determined but reflect both SSI variability and any operational plant-level controls. Nevertheless, national-scale ramps remain substantial despite ramp-rate control when spatially coherent, large-scale weather changes simultaneously affect the output of many geographically distributed PV systems, as illustrated by our study. At national scale, large PV ramps translate into system-level ramping and balancing requirements, which must be met through flexible generation, storage, demand response, or cross-border exchanges. Moreover, as the absolute ramp magnitudes in our study are derived from a fleet of approximately 6400 PV systems, the ramp-event definition and the reported ramp magnitudes may need to be adjusted when applied to PV fleets with substantially larger or smaller installed capacities.

4. Methodology

The satellite-based nowcasting models, namely IrradianceNet, SolarSTEPS, and SHADECast, are adopted to forecast satellite-derived CSI fields. All models were originally trained and calibrated on seven years of HelioMont CSI data at a spatial resolution of 0.02° – see Section 2 for more information. For evaluation, we use two years of HANNA SSI data at 0.01° resolution. To ensure consistency with the training setup, the HANNA SSI fields are downsampled to 0.02° using 2×2 pixel averaging. SSI is subsequently converted to CSI using clear-sky irradiance computed with the Ineichen model as implemented in pvlib [40,41]. Each forecast is generated from four consecutive CSI fields (corresponding to one hour of observations) and predicts the subsequent eight time steps, corresponding to a two-hour forecast horizon with lead times of 15 min. The predicted CSI fields are finally converted back to SSI, resulting in the corresponding SSI forecasts.

SHADECast and IrradianceNet operate over the spatial domain [4.41°E, 12.09°E] × [42.99°N, 50.67°N], corresponding to images of size 384×384 pixels. Due to architectural constraints, IrradianceNet is applied using a 128×128 patch-based strategy, with border interpolation employed to reconstruct the full spatial field [20,28]. SolarSTEPS requires a slightly larger input domain to mitigate edge effects associated with optical-flow-based advection. The predicted irradiance fields are subsequently converted into PV power using station-specific machine learning models, enabling comparison with production data from 6434 PV stations across the study domain. We adopt the extreme gradient boosting (XGBoost) algorithm [42] to model the conversion between solar irradiance and PV power output. The SSI serves as the primary meteorological predictor. To account for geometric and topographic effects, such as terrain-induced shading, we additionally include the solar zenith angle (SZA) and solar azimuth angle (AZI). These quantities are computed using HORAYZON, a ray-tracing algorithm that determines the local horizon and sky-view factor [43]. This is particularly important in regions with complex topography, where terrain features can induce substantial spatial variability in SSI. To capture temporal periodicity, we further incorporate the day of the year (DoY) and the hour of the day (HoD). In total, each PV system is characterized by seven predictors, i.e. SSI, SZA, AZI, and the four time-based features (i.e., HoD and DoY with cyclical encoding), and one target variable, namely the measured PV power output. To account for differences in nominal capacity, orientation, and elevation among PV systems, a separate XGBoost model is trained for each station. We refer to [21] for further details on the training strategy and model performance.

Forecast evaluation follows a user-centric setup. At each full hour, between one hour after sunrise and three hours before sunset, the models generate forecasts with a two-hour horizon and time resolution of 15 min. Although the satellite-based models are capable of producing forecasts at 15-minute intervals, the evaluation is restricted to hourly initialization to ensure comparability with IFS-ENS. This procedure is repeated over the entire two-year evaluation period, yielding a total of

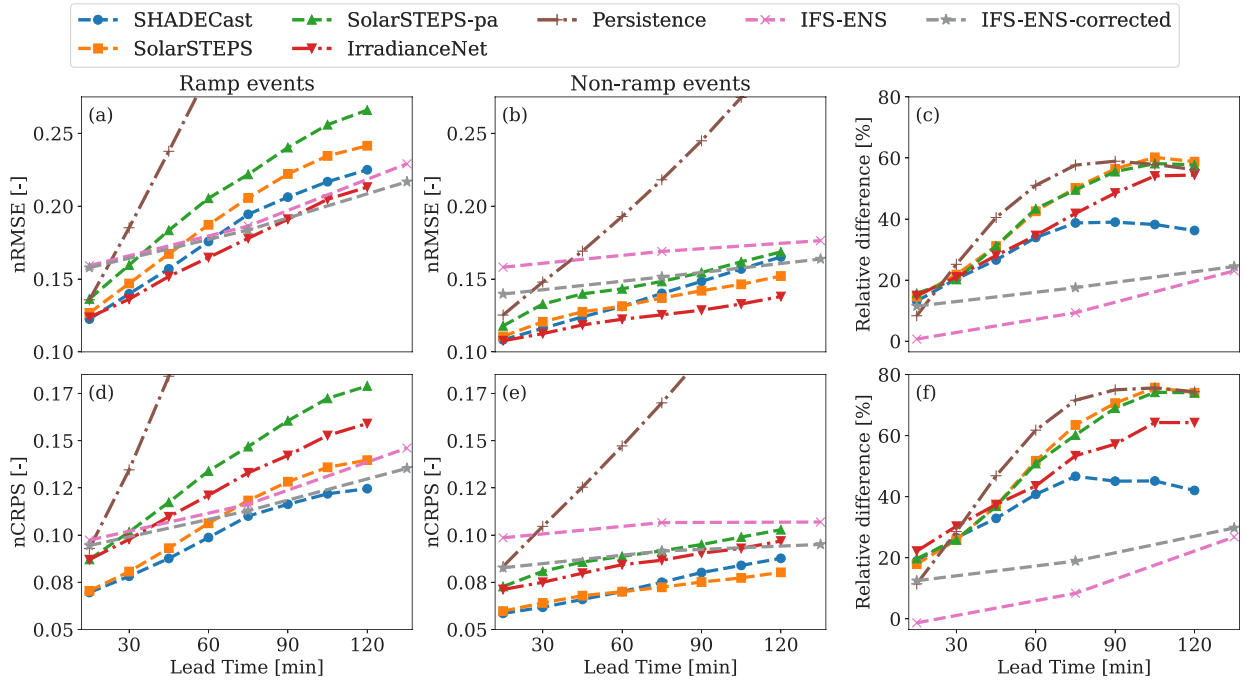


Fig. 6. (a, b) nRMSE and (d, e) nCRPS averaged across all PV stations and over the subsets of forecasts corresponding to (a, d) ramp events and (b, e) non-ramp events. Panels (c, f) show the relative difference in nRMSE and nCRPS, respectively, between forecasts during ramp and non-ramp conditions. The relative difference is defined as $(m_A - m_B)/m_B$, where m denotes a generic metric and A and B refer to ramp and non-ramp subsets, respectively. For deterministic models, nCRPS is replaced by nMAE. Dashed–dotted lines indicate deterministic models, while dashed lines represent probabilistic models.

6158 forecast inferences per model. We note that the methodology used for PV power forecasting corresponds to that implemented by [21], to which we refer for a detailed description. All simulations were performed on the Swiss high-performance computing system Alps at the Swiss National Supercomputing Centre.

5. Results

The aggregate performance of the forecasting models across all ramp events is presented in Section 5.1, along with a comparison against non-ramp conditions using both deterministic and probabilistic metrics. Subsequently, three case studies are selected to illustrate model performance under typical weather regimes associated with ramp events. Section 5.3 examines a ramp-down event driven by cloud advection, Section 5.4 presents a ramp-up event resulting from rapid morning cloud dissipation, and Section 5.5 highlights how fog and low-level clouds dissipation over the Swiss Plateau can lead to strong increases in PV power generation. A more detailed quantitative analysis of these case studies is provided in Appendix A.

5.1. Statistical evaluation across all ramp events

To assess the performance of the models in predicting upcoming ramps, we identify all forecasts that contain at least one ramp event within the 30–75 min lead-time window. This window is chosen to focus on short-term predictability while avoiding very short lead times, where persistence effects dominate, and longer lead times, where forecast skill typically decreases. As 29% of ramp events occur at consecutive time steps, this results in a total of 51 forecasts classified as ramp-event cases, of which 45 have a nowcast issued at that time. Notably, these represent only 0.83% of all forecasts over the 2019–2020 period, highlighting the rarity and extreme nature of such events. To evaluate model performance under contrasting conditions, we construct a corresponding set of forecasts without ramp events. For each ramp-event case, we select forecasts issued at the same UTC time on the preceding and following days, provided that no ramp occurs within the same lead-time window. This approach ensures a consistent comparison framework, as the selected forecasts are temporally close and therefore associated with similar solar elevation angles and typical SSI levels. Consequently, differences in forecast performance can be more directly attributed to the presence or absence of ramp dynamics rather than to seasonal or diurnal variability. Applying this procedure yields a total of 77 forecasts classified as non-ramp cases. Forecast performance is evaluated using both deterministic and probabilistic metrics. The normalized root mean square error (RMSE), abbreviated as nRMSE, quantifies the magnitude of point forecast errors, representing the average deviation between predicted and observed power. The normalized continuous ranked probability score (CRPS), abbreviated as nCRPS, assesses the full predictive distribution, capturing both reliability and sharpness of the ensemble forecasts. Detailed definitions of these metrics are provided in Appendix B.

Fig. 6 compares PV power forecast errors, averaged over the ramp and non-ramp subsets, as a function of lead time. For both deterministic (nRMSE) and probabilistic (nCRPS) metrics, errors increase with lead time across all models. However, the degradation in forecast skill is substantially more pronounced during ramp events, highlighting the difficulty of predicting PV power under rapidly evolving cloud conditions. The optical-flow-based approach that relies solely on advection (i.e. SolarSTEPS-pa) exhibits the strongest error growth in ramp situations, reflecting that the model does not account for cloud formation and dissipation processes. Among the satellite-based models, SHADECast emerges as the most reliable, achieving an nCRPS that is 10.8% lower than that of SolarSTEPS at a 120-minute lead time. Under non-ramp conditions, however, SolarSTEPS slightly outperforms SHADECast at lead times beyond 60 min, consistent with the findings of [21]. As expected, the bias-corrected IFS forecasts consistently outperform the raw forecasts, highlighting the importance of statistical post-processing. Fig. 6(c, f) further quantifies the relative increase in forecast error between ramp and non-ramp cases in terms of nRMSE and nCRPS, respectively. For most satellite-based models, ramp events lead to nRMSE increases

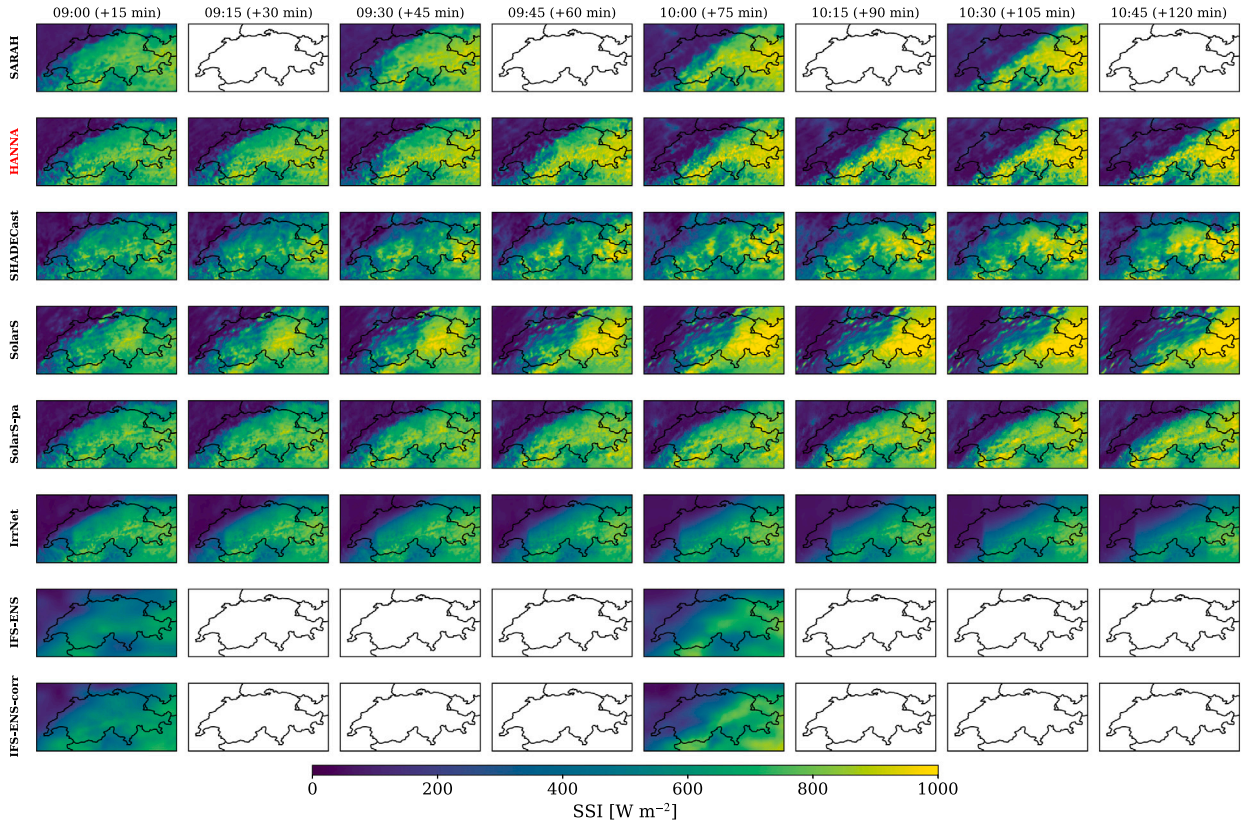


Fig. 7. Satellite-based SSI observations and model forecasts over the area of interest at lead times ranging from 15 to 120 min. The forecasts are issued at 09:00 UTC on 23 May 2020. For the probabilistic models, the ensemble member chosen is the one with the lowest RMSE. Note that the IFS-ENS SSI fields represent a one-hour temporal average, whereas all other fields are instantaneous. The black lines denote national borders. The row highlighted with the red label is used as the ground truth. Missing panels reflect differences in the temporal resolution of the satellite observations and forecast models. For illustration purposes, the SolarSTEPS, SolarSTEPS-pa and IrradianceNet models are abbreviated to SolarS, SolarS-pa and IrrNet, respectively.

of 20%–30% at short lead times, rising to approximately 60% at longer horizons. A similar pattern is observed for the nCRPS, with relative differences reaching up to 70% beyond 90-minute lead times. Although IFS-ENS also shows performance degradation during ramp events, the relative increase is less pronounced compared to the satellite-based models. Overall, the results demonstrate that ramp conditions not only reduce point forecast accuracy but also deteriorate the probabilistic skill of ensemble predictions, making ramp events a dominant source of forecast uncertainty in PV power forecasting.

Further event-based verification metrics are provided in Appendix B. To quantify the sampling uncertainty of the scores, we performed a block bootstrap over days ($B = 50\,000$ resamples) to the 74 ramp events detected for a ramp threshold of 37.7 MW; we use the day as the resampling unit because forecasts issued on the same day might have high correlation. The 95% confidence intervals of the individual scores overlap considerably and the model ranking is unchanged: SHADECast scores the lowest nCRPS at every lead time and ranks first in at least 90% of resamples from 30 min onward, maintaining its advantage over the other models at every lead time.

5.2. IFS outperforms nowcast models for extreme ramps

The threshold ΔP_{tr} separates ramp from non-ramp events. In the following, we assess different threshold values ΔP_{tr} to examine how the forecast models' performances depend on the choice of threshold and, thus, on the magnitude of the identified ramps. By construction, the threshold $\Delta P_{tr} = 37.7$ MW is the largest 15-minute change in the PV power observed on any of the clear-sky days of the study period, so that power variations exceeding it cannot be produced by solar geometry alone. The threshold value ΔP_{tr} lies in the tail of a steeply decaying distribution (Fig. 3a), therefore the count of solar ramp events is highly sensitive to it. To test the sensitivity, we vary the threshold ΔP_{tr} by +20%, +15%, +10%, +5%, −5%, −10%, −15%, and −20%, resulting in the detection of between 11 at $\Delta P_{tr} = 45.25$ MW (+20%) and 2078 ramp events at $\Delta P_{tr} = 30.16$ MW (−20%), as shown in Fig. 14. We find that the model performance ranking remains unchanged except for the largest ramp values. If we increase the threshold by 20%, we notice that IFS-ENS and IFS-ENS-corrected outperform the machine learning and optical flow models at lead times beyond approximately 50 min (Figs. 14, 15). This corresponds to only 11 ramp events during the study period of 2019–2020, of which only 7 can be considered independent, with at least multiple hours separating any pairs of these events. We find that the relative advantage of the machine-learning-based solar nowcasting models over the IFS observed under average conditions [21] shrinks and even reverses for the most extreme solar ramp events for those lead times. The superior performance of IFS-ENS over the machine-learning-based models for the most extreme ramp events is consistent with findings from other applications showing that machine-learning weather models can lose skill relative to physics-based NWP for some extreme events, including extreme heat, cold, and wind events [44,45]. The superior performance for the most extreme ramps may be related to the fact that extreme ramps are rare in the training set, so the machine learning models have only few examples to learn from and may struggle to extrapolate beyond the training distribution, and loss functions optimizing for average error tend to also smooth extremes, whereas NWP models are not constrained by the frequency of rare events in the training data. Moreover, the coarser resolution of IFS-ENS acts as a spatial average, producing a smoother SSI field that may be penalized less by squared-error metrics

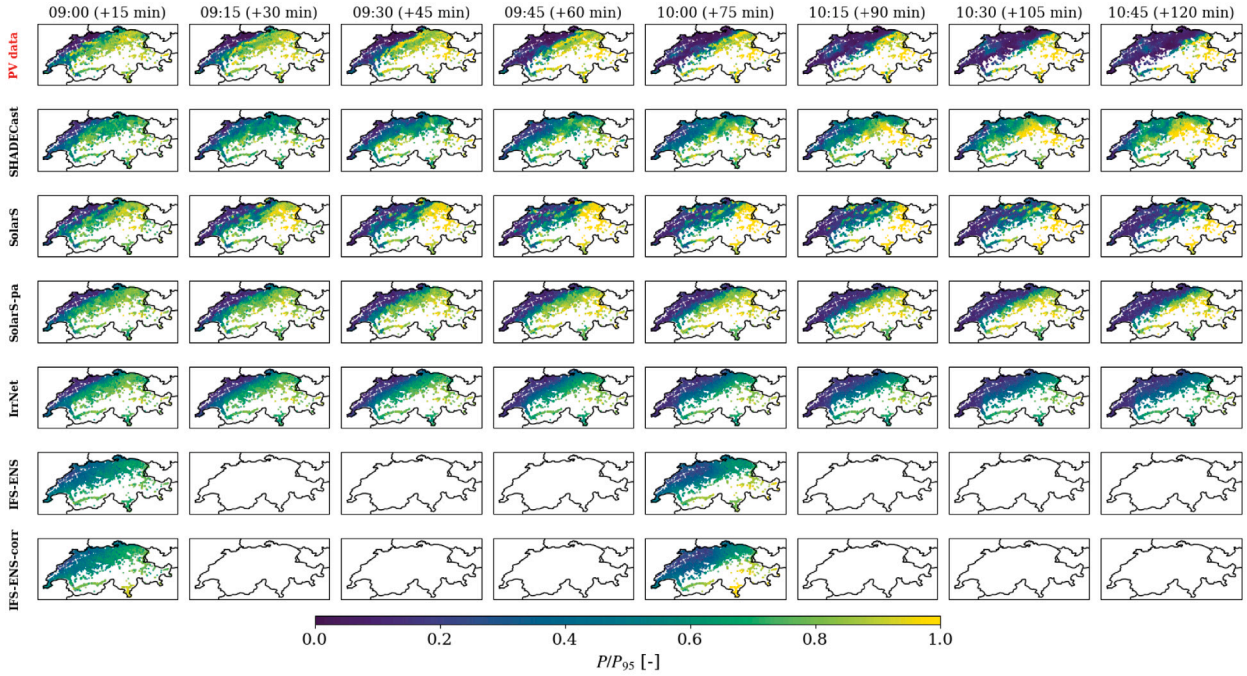


Fig. 8. PV power measurements and model forecasts at lead times ranging from 15 to 120 min. The forecasts are issued at 09:00 UTC on 23 May 2020. For probabilistic models, the same ensemble member used in Fig. 7 is shown. The row highlighted with the red label is used as the ground truth. Missing panels reflect differences in the temporal resolution of the satellite observations and forecast models. Note that the PV power output is normalized using the station-specific 95th percentile of the power time series.

than a sharp field that places a cloud edge in the wrong location We show the 7 independent forecasts with the most extreme ramps, corresponding to a threshold value of (ΔP_{tr} = 45.2 MW), in Figs. 17 to 23 of Appendix.

5.3. Case study I: ramp-down event driven by cloud advection

The SSI satellite observations and model forecasts for the ramp-down event described in Section 3 are presented in Fig. 7 for lead times ranging from 15 to 120 min. Forecasts are issued at 09:00 UTC on 23 May 2020, using satellite-observed SSI fields from 08:00 to 08:45 UTC as input. The SSI observations are shown for two retrieval methods, that is SARAH-3 and HANNA. Both datasets show a coherent cloud band that is advected across the study domain in a south-easterly direction. Due to its higher spatial and temporal resolution, HANNA more clearly resolves cloud edges, small-scale cloud structures, and their temporal evolution. This ramp-down event is primarily governed by cloud advection. Consequently, the optical-flow-based models, SolarSTEPS and SolarSTEPS-pa, succeed in forecasting the cloud band motion with relatively high accuracy. While SHADECast captures the SSI increase in eastern Switzerland, it fails to accurately predict the cloud-band advection in this ramp event, likely due to the limited representation of such events in the training data. At the same time, the absence of spatial blurring, attributable to its diffusion-based framework, is evident. The IrradianceNet model captures the overall evolution of the cloud band better but fails to reproduce the increase in SSI over the eastern regions. Moreover, the patch-based strategy introduces spatial discontinuities particularly evident at lead times exceeding 60 min, with sharp transitions along patch boundaries where interpolation is applied. In addition, its forecasts are smoother and highly diffuse. Finally, IFS-ENS correctly predicts regions of low SSI. However, its comparatively coarse spatial and temporal resolution yields smooth and more homogeneous SSI fields with less distinct cloud boundaries.

Fig. 8 presents the corresponding PV power forecasts obtained from the station-specific machine-learning-based irradiance-to-power conversion model. The top row shows the observed PV power generation, which exhibits pronounced spatial variability closely aligned with the cloud distribution. As the cloud band advects over the Swiss Plateau, national PV power output decreases by approximately 60.7% between 09:00 and 10:45 UTC. Overall, the forecast models reproduce the main SSI patterns, with low-irradiance areas corresponding to reduced PV power output and vice versa. Given that this ramp event is dominated by cloud advection, the displayed ensemble member of the SolarSTEPS-pa forecast is in close agreement with the observed PV power distribution, while SHADECast struggles to accurately capture the power gradients in this case.

5.4. Case study II: ramp-up event driven by cloud dissipation

The second illustrative case highlights a ramp-up event that occurred on 3 October 2020 at 11:45 UTC. Fig. 9 presents the SSI satellite observations and model forecasts for the ramp-up event described in Section 3, also for lead times between 15 and 120 min. Forecasts were initialized at 11:00 UTC on 3 October 2020, using SSI fields from 10:00 to 10:45 UTC. Both SARAH and HANNA retrievals indicate rapid cloud clearing over the Swiss Plateau. This large-scale cloud dissipation leads to a significant increase in SSI and, consequently, a strong ramp-up in PV power production. In contrast to the previously discussed advection-dominated case, cloud dissipation plays a major role in this event, posing a challenge for the optical-flow-based models. In particular, SolarSTEPS-pa is unable to represent the rapid reduction in cloud cover and therefore substantially underestimates the irradiance increase. While showing tendencies to increasing SSI in the forecasts, the data-driven models also struggle to reproduce the rapid clearing, likely reflecting the limited representation of such events in the training data. The IFS-ENS shows a comparable behavior, with a smoothed transition toward higher SSI values. As a result, all models underestimate both the magnitude and spatial extent of the ramp-up.

Fig. 10 displays the corresponding PV power forecasts. The observed power production increases by approximately 65.7% between 10:00

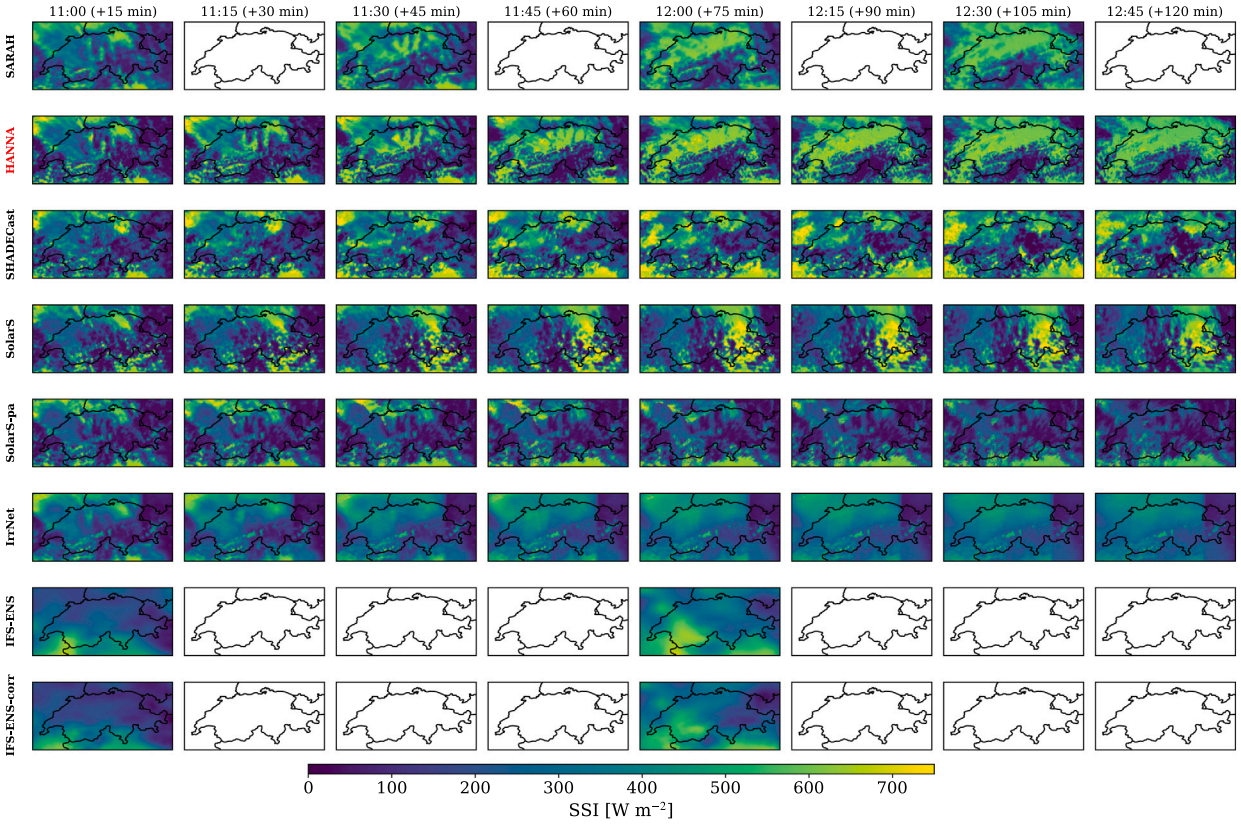


Fig. 9. Satellite-based SSI observations and model forecasts over the area of interest at lead times ranging from 15 to 120 min. The forecasts are issued at 11:00 UTC on 3 October 2020. For the probabilistic models, the ensemble member chosen is the one with the lowest RMSE. Note that the IFS-ENS SSI fields represent a one-hour temporal average, whereas all other fields are instantaneous. The black lines denote national borders. The row highlighted with the red label is used as the ground truth. Missing panels reflect differences in the temporal resolution of the satellite observations and forecast models.

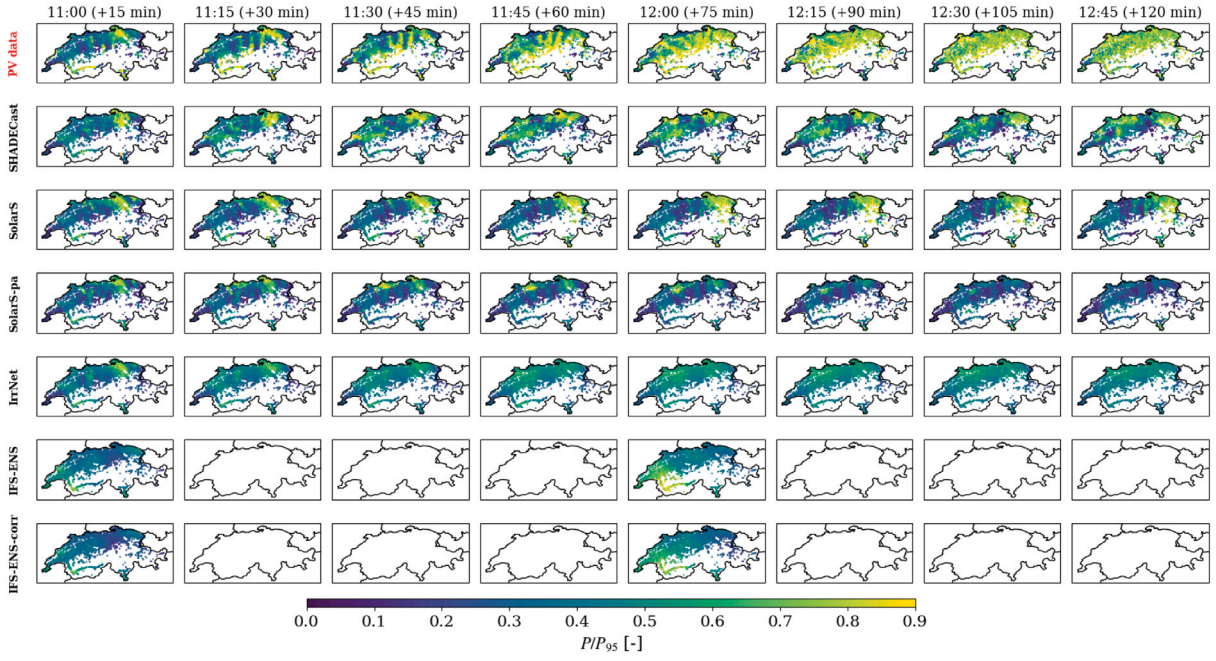


Fig. 10. PV power measurements and model forecasts at lead times ranging from 15 to 120 min. The forecasts are issued at 11:00 UTC on 3 October 2020, a day with high-variability weather. For probabilistic models, the same ensemble member used in Fig. 9 is shown. The black lines denote national borders. The row highlighted with the red label is used as the ground truth. Missing panels reflect differences in the temporal resolution of the satellite observations and forecast models. Note that the PV power output is normalized using the station-specific 95th percentile of the power time series.

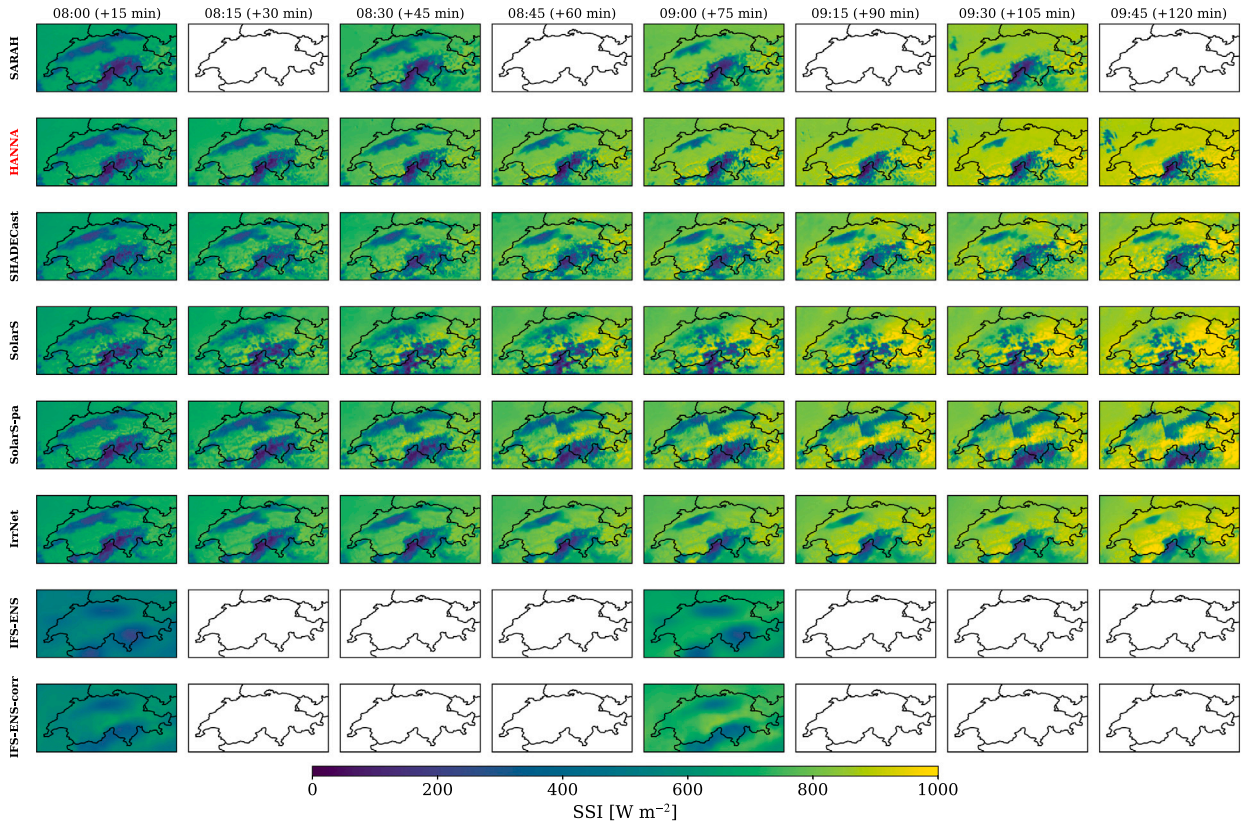


Fig. 11. Satellite-based SSI observations and model forecasts over the area of interest at lead times ranging from 15 to 120 min. The forecasts are issued at 08:00 UTC on 7 June 2019. For the probabilistic models, the ensemble member chosen is the one with the lowest RMSE. Note that the IFS-ENS SSI fields represent a one-hour temporal average, whereas all other fields are instantaneous. The black lines denote national borders. The row highlighted with the red label is used as the ground truth. Missing panels reflect differences in the temporal resolution of the satellite observations and forecast models.

and 11:45, in line with the clearing of clouds. While the models broadly reflect the relationship between irradiance and power, they fail to capture the steep spatial gradients and rapid temporal increase associated with this event. Overall, these results highlight that rapid changes in cloud cover remain difficult to predict, regardless of whether the forecasting approach is data-driven or physics-based, remaining a major source of forecast error in both SSI and PV power predictions.

5.5. Case study III: Ramp-up event in presence of low-level clouds

The Swiss Plateau is frequently affected by fog and low-level clouds due to orographic confinement and the abundant moisture supply from lakes and rivers. During winter, such stratiform cloud decks often persist throughout the day under stable atmospheric conditions. In late spring and early summer, low-level clouds typically dissipate during the morning as solar heating intensifies quickly. Because a large fraction of PV installations are located in the Swiss Plateau, this transition from fog and overcast to clear-sky conditions, combined with the increasing solar elevation, frequently triggers large variations in PV power output. One exemplary case occurred on 7 June 2019 at 08:45 UTC. The corresponding SSI satellite observations and model forecasts for lead times from 15 to 120 min are shown in Fig. 11. At 08:00 UTC, the satellite observations indicate a widespread layer of clouds covering the majority of the Swiss Plateau. As the sun rises, enhanced surface heating warms the boundary layer, increases turbulent mixing, and progressively dissipates the cloud deck. By 10:00 UTC, most clouds have dissipated and the region transitions to near clear-sky conditions. Data-driven models, such as SHADECast and IrradianceNet, reproduce the cloud pattern and its dissipation remarkably well in this case. This may reflect the relatively high frequency of such clearing events, which can enable data-driven models to learn the associated spatial and temporal evolution from the training data. In contrast, the purely advection-based optical flow model SolarSTEPS-pa performs poorly, as it can only advect existing cloud structures and cannot represent cloud dynamics. When cloud-evolution mechanisms are included, as in SolarSTEPS, the model better captures the observed clearing process. The physics-based IFS-ENS forecast correctly predicts the presence of clouds over the Swiss Plateau. However, its comparatively coarse spatial and temporal resolution limits its ability to resolve cloud edges and the detailed evolution of the dissipation process.

Fig. 12 presents the corresponding PV power measurements and forecasts. The measurements, shown in the top row, exhibit spatial and temporal patterns closely aligned with the cloud distribution. In fact, regions affected by clouds correspond to strongly reduced PV power generation, whereas cloud-free areas show enhanced power output. By 10:00 UTC, most PV stations generate power above their respective P_{95} values. Between 08:00 and 09:45 UTC, the national PV power output increased by approximately 72.0%. Overall, the forecast models reproduce the main SSI patterns, with low-irradiance areas translating into reduced PV power output and vice versa. Consistent with the SSI results, SHADECast and IrradianceNet provide the most accurate power forecasts in this case.

We expect that our results overall generalize qualitatively to regions with similar climates. While the characteristics of ramp events, their frequency, magnitude, and seasonal occurrence may be specific to climates and PV fleets similar to the one of the study domain,

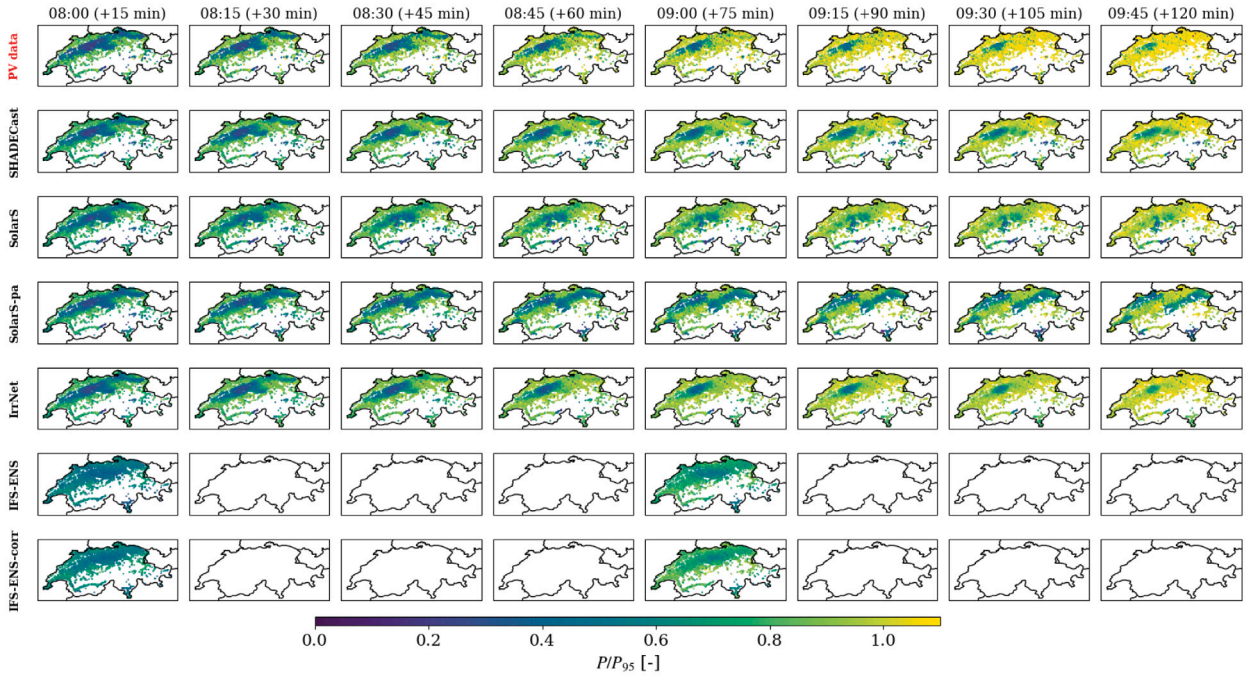


Fig. 12. PV power measurements and model forecasts at lead times ranging from 15 to 120 min. The forecasts are issued at 08:00 UTC on 7 June 2019. For probabilistic models, the same ensemble member used in Fig. 11 is shown. The black lines denote national borders. The row highlighted with the red label is used as the ground truth. Missing panels reflect differences in the temporal resolution of the satellite observations and forecast models. Note that the PV power output is normalized using the station-specific 95th percentile of the power time series.

rapid changes in spatially coherent cloud cover can produce substantial ramps in geographically distributed PV fleets also in other regions, while their rapid evolution poses challenges for both satellite-based nowcasting and NWP. Applying the proposed framework to other regions may require adapting the ramp-event definition to differences in installed PV capacity.

6. Conclusions

This study presents a comprehensive assessment of solar ramp events and their implications for short-term PV power forecasting. We analyze two years of 15-minute resolution PV power measurements from 6434 PV stations and introduce a quantitative, data-driven methodology to identify solar ramp events, defined as large variations in nationally aggregated PV generation over short time intervals. Within this framework, we systematically characterize the frequency, magnitude, seasonal distribution, and diurnal occurrence of the ramp events. Furthermore, we employ a recently developed spatiotemporal forecasting framework to evaluate both deterministic and probabilistic PV power forecasts under ramp and non-ramp conditions. The benchmarked models, that is SolarSTEPS, SHADECast, IrradianceNet, and IFS-ENS, rely on optical-flow techniques, deep-learning frameworks and physics-based approaches,

Our analysis aggregates 6434 PV systems distributed across Switzerland into a single signal. This aggregation is appropriate for characterizing variability relevant at the national level but may not capture the full distribution of ramp events in PV fleets with substantially larger or smaller installed capacities or in different climate zones, where SSI may, on average, change more rapidly or slowly. For example, regions dominated by convective clouds can experience rapid SSI changes, whereas persistently clear or stratiform-cloud regimes and seasons generally exhibit slower SSI variability. The ramp threshold ΔP_{tr} derived in this study is specific to the spatial extent and composition of the Swiss PV fleet and should therefore not be directly transferred to significantly different aggregation areas and total installed capacities. Consequently, the reported event statistics and forecasting performance should be interpreted as representative of national-scale variability rather than as a direct assessment of forecasting requirements at local grid levels.

First, our results indicate that solar ramp events occur mainly in spring and summer, with peak frequency between 07:00–09:00 and 14:00–16:00 UTC. In particular, ramp-up events typically result from the combined effect of increasing solar elevation and rapid cloud dissipation during the morning hours, whereas ramp-down events are primarily associated with decreasing solar elevation and enhanced cloud formation in the afternoon. These findings highlight the strong coupling between mesoscale cloud dynamics and variability in nationally aggregated PV power production.

Our study indicates that the optical-flow-based model relying purely on advection (e.g. SolarSTEPS-pa) experiences the largest increase in forecast errors during ramp events, due to its lack of representation of cloud formation and dissipation. Among the satellite-based approaches, SHADECast provides the most reliable forecasts under these conditions. In contrast, under non-ramp conditions, SolarSTEPS performs slightly better than SHADECast for lead times exceeding 60 min. Furthermore, the bias-corrected IFS-ENS forecasts consistently outperform their raw counterparts, underlining the important role of statistical post-processing.

Quantitatively, satellite-based models exhibit a clear degradation in performance during ramp events compared to non-ramp cases. Specifically, the nRMSE increases by approximately 20%–30% at short lead times, reaching up to 60% at longer forecast horizons. A similar pattern is observed for nCRPS, with relative increases approaching 70% beyond lead times of 90 min. These findings indicate that ramp events significantly reduce not only deterministic forecast accuracy but also the reliability and sharpness of ensemble predictions. While IFS-ENS also experiences reduced accuracy during ramp events, the magnitude

of deterioration is substantially smaller than that of the satellite-based methods.

For the most extreme ramps, the IFS-ENS and IFS-ENS-corrected outperform the machine learning and optical flow models at lead times beyond approximately 50 min (Figs. 14, 15). This corresponds to only 7 independent ramp events during the study period of 2019–2020. The relative advantage of the machine-learning-based solar nowcasting models over the IFS observed under average conditions shrinks and even reverses for the most extreme solar ramp events at those lead times. We hypothesize that this is because extreme ramps are rare in the training set, so the machine learning models have only few examples to learn from.

The data-driven models benchmarked in our study have been trained on satellite-derived SSI fields in which ramp events are very sparse. Accordingly, the models were not optimized for forecasting rapid SSI changes like the ones examined here. Hence, the models performance in ramp conditions can likely be improved by focusing the model training on ramp events, for instance, by adapting the loss function to place greater emphasis on extreme events. Moreover, future research can involve increasing the fraction of ramp events in the training set, for example, by upsampling and image augmentation of SSI field sequences related to ramp events. Overall, accurately capturing rapid cloud formation and dissipation remains a key challenge for nowcast models. Improving high-resolution spatiotemporal modeling capabilities and increasing the robustness of data-driven approaches to extreme events are essential steps toward enhancing ramp prediction skill. Such advancements are relevant for maintaining grid stability and supporting the reliable integration of growing shares of solar generation into power systems. Future research should also explore alternative spatial aggregation strategies, for example, at the level of transmission grid zones, to better align meteorological variability with operational grid constraints. Moreover, the integration of multi-source observations (e.g., satellite, radar, and ground-based measurements) and the explicit modeling of extreme-event statistics represent promising directions to further improve the prediction of solar ramp events.

Our study captures approximately 5% of the PV installations in the study region during the study period 2019–2020, with ramp magnitudes observed reaching up to 45 MW within 15 min. Extrapolated to the entire fleet at the time, ramp magnitudes would be approximately twentyfold larger, assuming a similar spatial distribution and proportional scaling of installed capacity. Moreover, as in many other countries, installed PV capacity is growing rapidly in Switzerland and is projected to increase fivefold by 2050. A simple proportional extrapolation of the observed ramps to the entire PV fleet and its projected growth would imply ramps on the GW scale within 15 min by 2050. Accurate ramp forecasts can support and complement the deployment of flexibility measures such as ramp-rate control and storage. Characterizing the frequency, timing, and predictability of these events at present penetration levels is therefore also directly relevant to future reserve dimensioning.

CRediT authorship contribution statement

Luca Lanzilao: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Francesco Pinto:** Data curation, Formal analysis, Investigation, Software, Visualization, Writing – review & editing. **Angela Meyer:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, ChatGPT was used to help with writing style. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Declaration of competing interest

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

Acknowledgments

We thank four anonymous reviewers for their constructive comments which helped to improve the quality of this paper. We acknowledge funding from the Swiss National Science Foundation (grant 200654). We thank Mathieu Schaer and Christian Steger of MeteoSwiss for the support to account for topographic shading using the HORAYZON library. We also thank Anke Tetzlaff and Uwe Pfeifroth of MeteoSwiss and the German Weather Service, respectively, for compiling and providing the HANNA dataset.

Appendix A. Quantitative forecast analysis for the selected case studies

This section provides a quantitative comparison of the forecasts for the three case studies introduced in Section 5, focusing on both SSI and PV power predictions. Fig. 13(a–c) presents SSI forecasts, interpolated to PV station locations and averaged across ensemble members and stations, against satellite observations, while Fig. 13(d–f) shows the corresponding aggregated PV power generation. As expected, SSI and PV power exhibit highly correlated temporal behavior, leading to similar trends as a function of the lead time. Case study I is dominated by cloud advection, resulting in a pronounced decrease in SSI and PV production. IrradianceNet provides the most accurate representation of both magnitude and timing. Optical-flow-based methods also capture this decreasing trend reasonably well. In contrast, IFS-ENS significantly underestimates the sharp transition, while SHADECast fails to capture the ramp event, leading to a PV power prediction at two-hour lead time that is more than twice as high as the observed generation. Case study II is characterized by rapid cloud dissipation, resulting in a steep increase in SSI and PV power output. Optical-flow methods struggle in this regime, as they are inherently limited in representing cloud formation and dissipation processes, resulting in persistently underestimated forecasts. Although IFS-ENS, SHADECast, and IrradianceNet capture the overall upward trend, the SSI and PV power variations are notably smoother than the ones observed. For example, at a two-hour lead time, IrradianceNet underestimates PV production by 34.3%. Finally, case study III involves the dissipation of low-level clouds over the Swiss Plateau, a common meteorological situation in Switzerland. In this scenario, data-driven models perform particularly well, accurately capturing the ramp-up in both SSI and PV power. SHADECast and IrradianceNet show only minor underestimation at a two-hour lead time, with errors of approximately 4.9%. While IFS-ENS also performs reasonably well, optical-flow methods, despite capturing the increasing trend, tend to underestimate the increase of both SSI and PV power generation.

Appendix B. Metrics

This section introduces the verification metrics used to evaluate the PV power forecasts in Section 5.1. We consider forecasts for a generic PV installation, which consists of $L = 8$ lead times and $E = 10$ ensemble members. We denote with $\hat{y}_{s,n,e,l}$ the predicted PV power at station s, time step n, ensemble member e and lead time l. Here, $s \in \{1, \dots, S\}$ indexes the PV installations, $n \in \{1, \dots, N\}$ indexes the time steps, $e \in \{1, \dots, E\}$ indexes the ensemble member and $l \in \{1, \dots, L\}$ indexes the lead time. The corresponding measured PV power at station s and lead time l, associated with forecast n, is denoted by $y_{s,n,l}$.

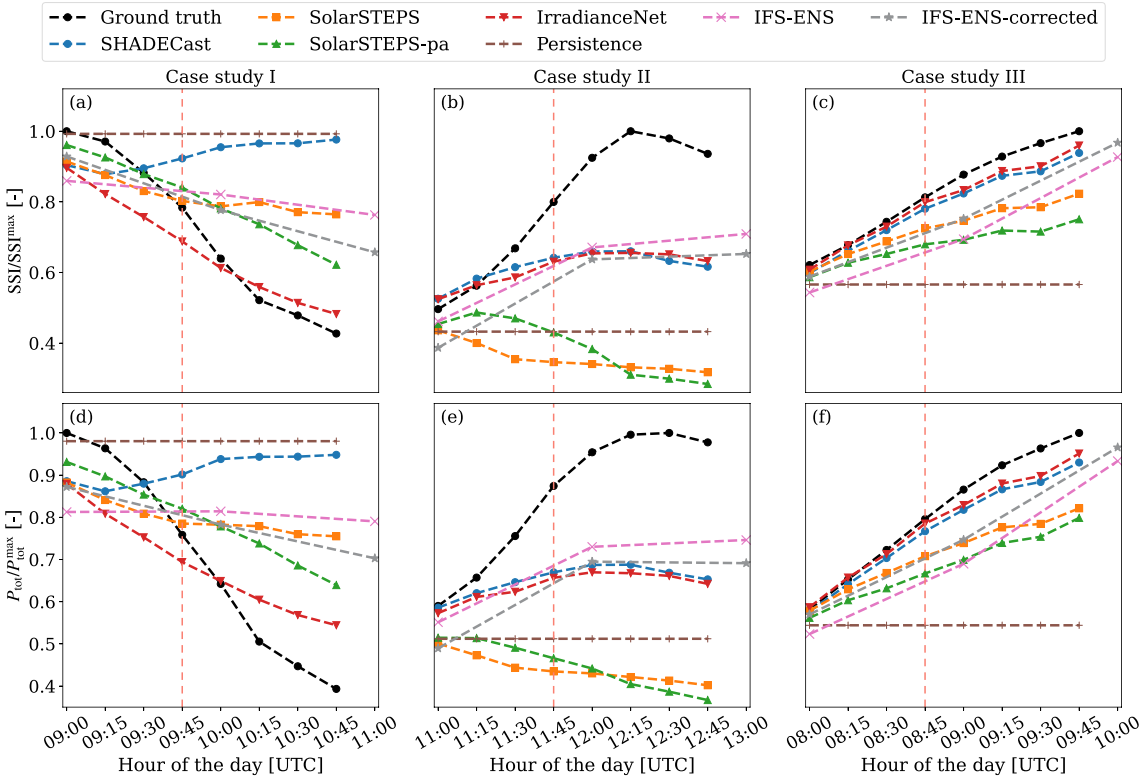


Fig. 13. Comparison of (a–c) SSI and (d–f) aggregated PV power forecasts computed for the three case studies discussed in Section 5. The SSI is spatially interpolated to each station location, averaged across all stations, and normalized by its maximum value within the corresponding 2-hour period. The PV power generated is summed over all stations and normalized by the maximum measured aggregated PV power within the same time window. The vertical dashed red line marks the occurrence time of the ramp event.

Table 1
Event-based verification on hourly ramps, which is the timescale on which IFS-ENS can be scored at all. The event is the change between two consecutive hourly means of the power.

Model	H	M	FA	FB	FAR	CS
SHADECast	52	93	46	0.67	0.47	0.27
SolarSTEPS	50	95	17	0.46	0.25	0.31
SolarSTEPS-pa	50	95	17	0.46	0.25	0.31
IrradianceNet	39	106	12	0.35	0.23	0.25
IFS-ENS	45	100	41	0.59	0.48	0.24
IFS-ENS-corrected	56	89	59	0.79	0.51	0.27
Persistence	–	145	–	–	–	–

Deterministic performance metrics are computed using the ensemble mean forecast, defined as

$$\bar{\hat{y}}_{s,n,l} = \frac{1}{E}\sum_{e=1}^{E} \hat{y}_{s,n,e,l}.$$

For deterministic models, $E = 1$, and thus $\bar{\hat{y}} = \hat{y}$.

The normalized root mean square error (nRMSE) averaged over all lead times and all forecasts for a station s, is defined as

$$\mathrm{nRMSE}_s = \frac{1}{f_s}\sqrt{\frac{1}{NL}\sum_{n=1}^{N}\sum_{l=1}^{L}\left(\bar{\hat{y}}_{s,n,l} - y_{s,n,l}\right)^2},$$

where $f_s = P_{95,s}$. This normalization ensures comparability across stations, as installations with larger $P_{95,s}$ would otherwise exhibit systematically higher errors (see Tables 1 and 2).

In addition to deterministic metrics, we evaluate the normalized continuous ranked probability score (nCRPS), which assesses the reliability and sharpness of the full predictive distribution. The nCRPS compares the predictive cumulative distribution function $F_{s,n,l}$, derived from the ensemble forecast, with the observation $y_{s,n,l}$, and is defined as

$$\mathrm{nCRPS}_s = \frac{1}{NLf_s}\sum_{n=1}^{N}\sum_{l=1}^{L}\int_{-\infty}^{\infty}\left(F_{s,n,l}(z) - \mathbb{1}(y_{s,n,l} \le z)\right)^2 dz,$$

where $\mathbb{1}(\cdot)$ denotes the indicator function. In practice, $F_{s,n,l}$ is approximated from the empirical distribution of the ensemble members $\{\hat{y}_{s,n,e,l}\}_{e=1}^{E}$.

Both deterministic and probabilistic metrics can be further averaged across all stations, yielding the global nRMSE and nCRPS illustrated in Fig. 6.

To assess the ability to anticipate a ramp, we apply the detection rule of Section 3 to the forecast aggregated series and compare it with the rule applied to the measurements. At each lead time this yields a contingency table with hits H, false alarms FA, and misses M from which we compute the false alarm ratio $\mathrm{FAR} = FA/(H+FA)$ [46], the frequency bias $\mathrm{FB} = (H+FA)/(H+M)$ [47], and the critical success index $\mathrm{CS} = H/(H+FA+M)$ [48].

The ramp duration was fixed to 15 min in Section 3 to match the resolution of both datasets and the imbalance settlement period harmonized across continental Europe under Regulation (EU) 2017/2195 [49]. Other jurisdictions settle on different time windows (e.g. 30 or 60 min), so we repeat the analysis with a time window $w \in \{15, 30, 45, 60, 75, 90, 120\}$ minutes. At $w = 15$ min the generalized detector reproduces the event set of Section 3 exactly, and all seven models reproduce the published scores to within numerical precision; the results are shown in Fig. 16.

Data availability

The authors do not have permission to share data.

Table 2
Station-averaged nCRPS and nRMSE during ramp events at 60-minute forecast lead time as a function of the ramp threshold. IFS-ENS runs on hourly steps. Compared models include SHADECast (SC), SolarSTEPS (SS), SolarSTEPS-pa (SS-pa), IrradianceNet (IN), IFS-ENS (IFS), and IFS-ENS-corr (IFS-c). The lowest nCRPS and nRMSE per row are marked in bold font. The ranking is computed on the unrounded scores and is identical at every threshold for both metrics.

Threshold	SC		SS		SS-pa		IN		IFS		IFS-c	
[MW]	nCRPS	nRMSE	nCRPS	nRMSE	nCRPS	nRMSE	nCRPS	nRMSE	nCRPS	nRMSE	nCRPS	nRMSE
30.16	**0.06**	0.12	0.06	0.13	0.08	0.14	0.07	**0.11**	0.09	0.14	0.08	0.13
33.94	**0.07**	0.14	0.08	0.15	0.10	0.16	0.09	**0.13**	0.10	0.16	0.09	0.15
35.84	**0.08**	0.15	0.09	0.16	0.11	0.18	0.10	**0.15**	0.10	0.16	0.10	0.16
37.71	**0.10**	0.18	0.11	0.19	0.13	0.21	0.12	**0.16**	0.12	0.19	0.11	0.18
40.01	**0.11**	0.20	0.12	0.21	0.15	0.23	0.14	0.19	0.12	0.19	0.11	**0.18**
41.48	0.13	0.21	0.14	0.23	0.17	0.25	0.15	0.20	0.13	0.20	**0.12**	**0.19**
45.25	0.12	0.20	0.13	0.22	0.16	0.22	0.14	0.18	0.12	0.18	**0.12**	**0.18**

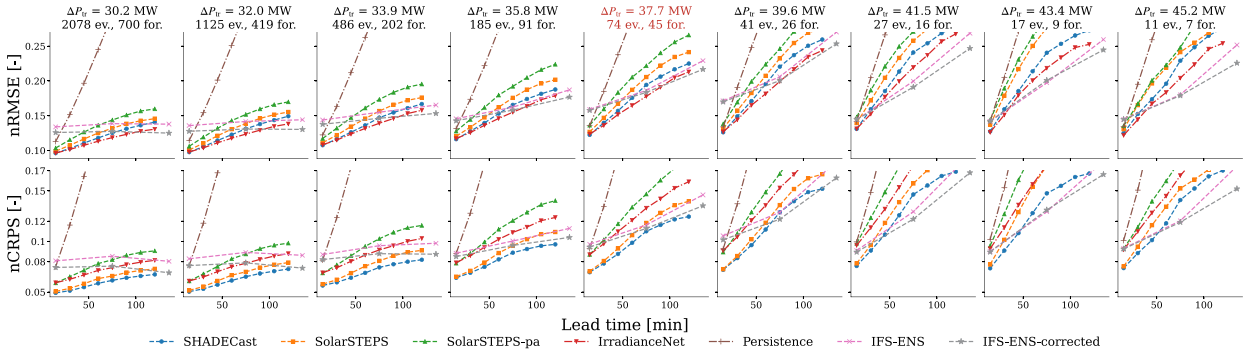


Fig. 14. Comparison of nRMSE and nCRPS for different values of ΔP_{tr}. Both the metrics tend to degrade when the threshold increases: for thresholds above 37.7MW IFS-ENS outperforms the models after approximately 50 min lead time. ev. stands for the number of events, for. stands for the number of forecasts.

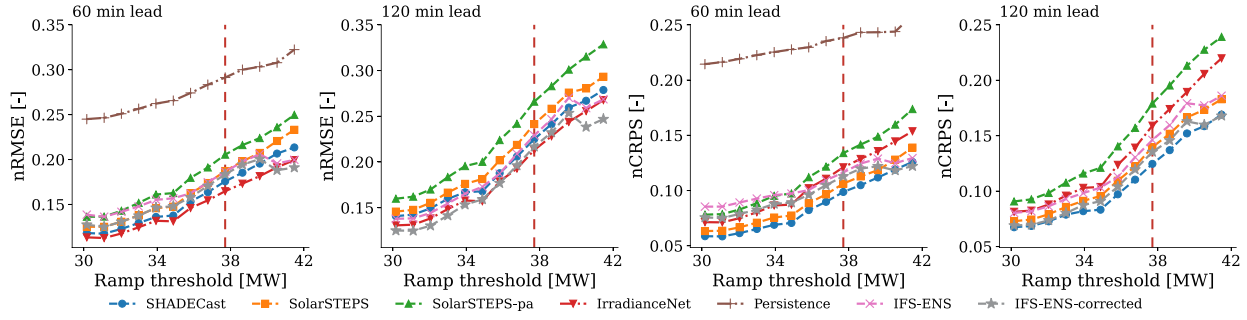


Fig. 15. Station-averaged nRMSE and nCRPS during ramp events as a function of the ramp threshold, at 60 and 120 min lead times. All models degrade as the threshold rises. The red dashed line marks ΔP_{tr}.

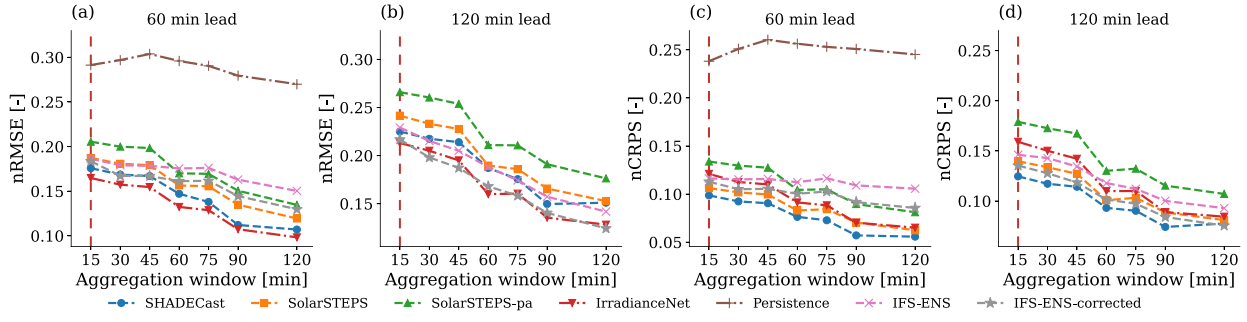


Fig. 16. Station-averaged nRMSE (a, b) and nCRPS (c, d) during ramp events as a function of the temporal aggregation window (w), at 60-minute (a, c) and 120-minute (b, d) lead times. The threshold is recalibrated at each window. Scores decrease with w for all models, while the relative ranking is almost unchanged. The red dashed line marks the 15-minute window adopted in this study. IFS-ENS is verified at its nearest hourly step.

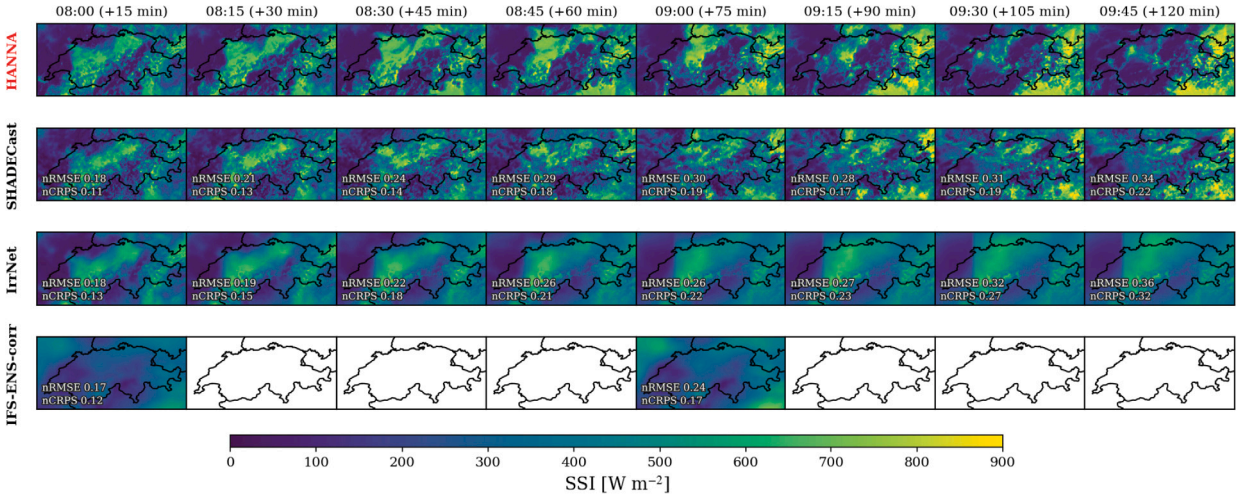


Fig. 17. 10.06.2019, ramp event at 09:15 (−58 MW).

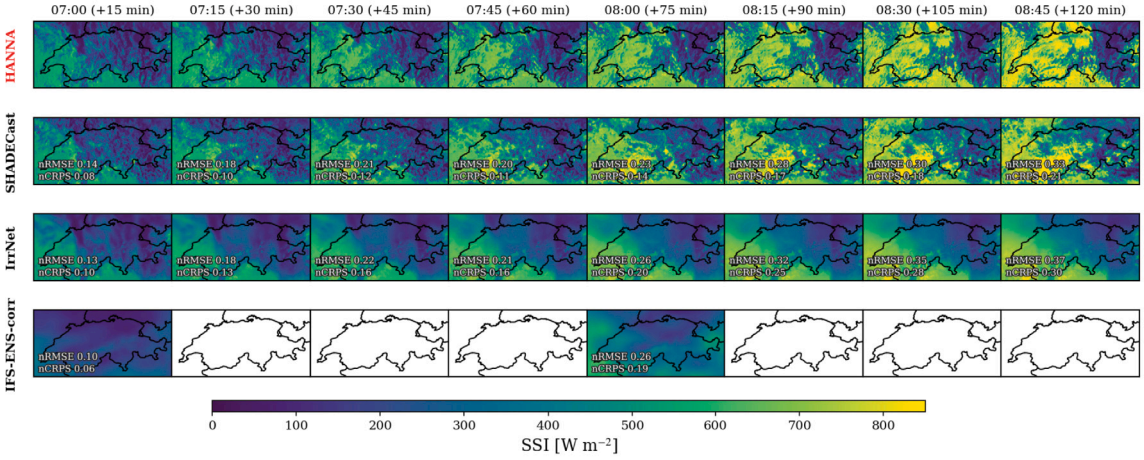


Fig. 18. 12.06.2019, ramp event at 07:30 (+46 MW).

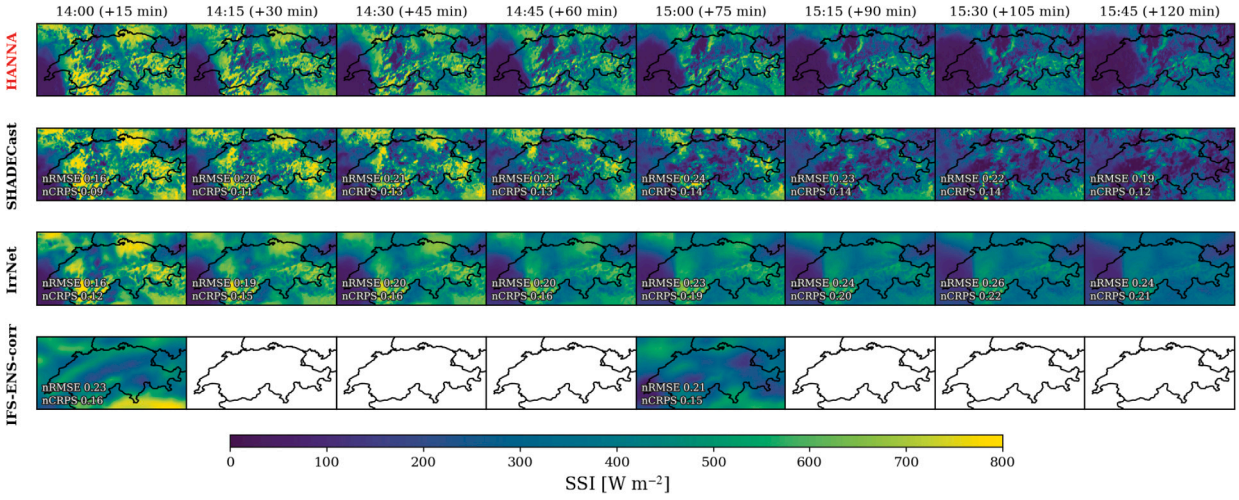


Fig. 19. 15.06.2019, ramp event at 15:15 (−50 MW).

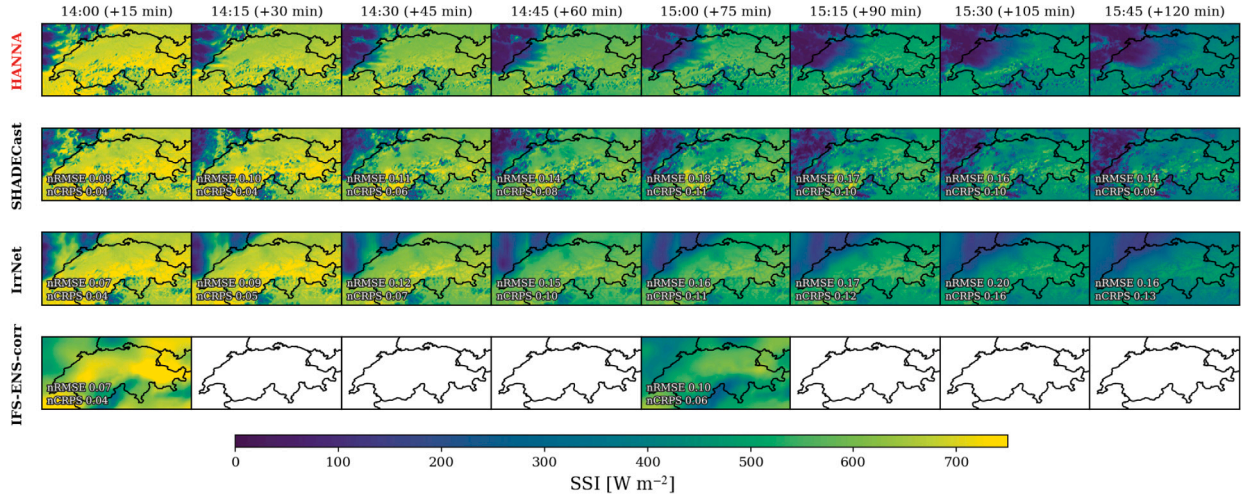


Fig. 20. 09.08.2019, ramp events at 14:30 (−47 MW), 14:45 (−51 MW), 15:00 (−45 MW), 15:15 (−45 MW).

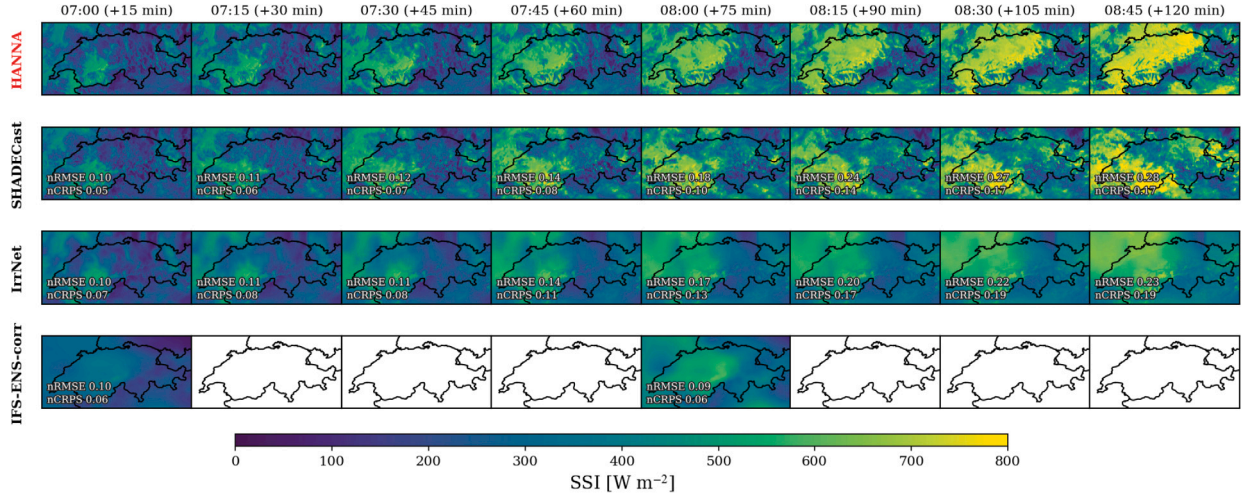


Fig. 21. 10.05.2020, ramp events at 07:45 (50 MW), 08:00 (48 MW).

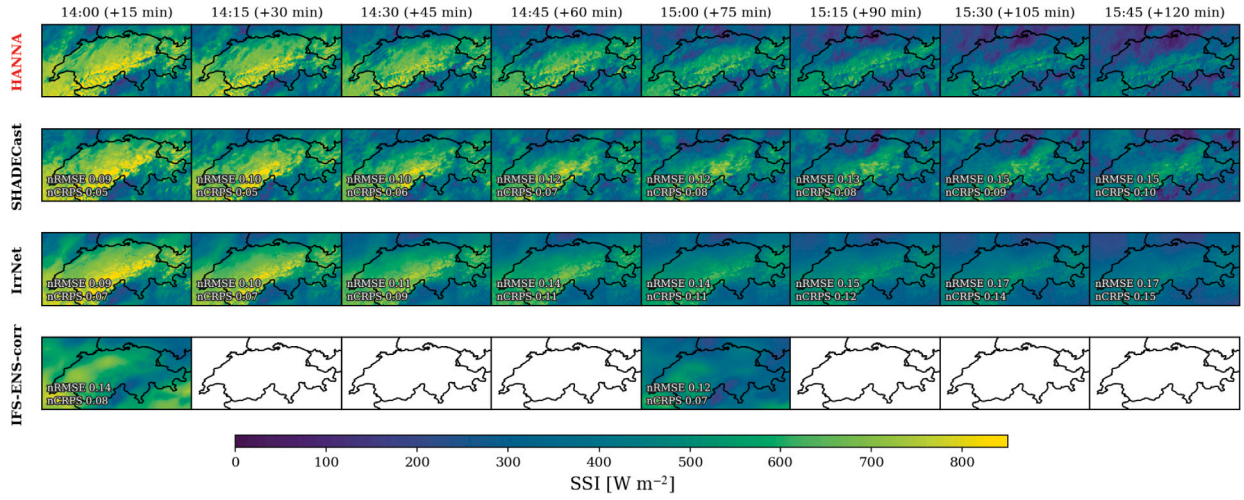


Fig. 22. 22.05.2020, ramp event at 14:45 (−49 MW).

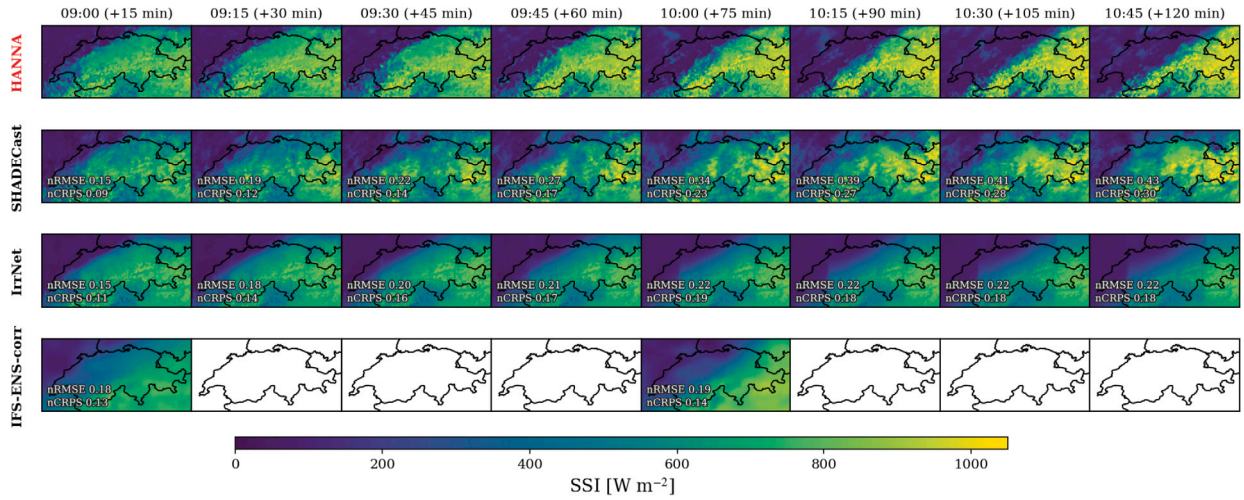


Fig. 23. 23.05.2020, ramp event at 10:15 (−47 MW).

References

[1] Morales Juan M, Conejo Antonio J, Madsen Henrik, Pinson Pierre, Zugno Marco. Integrating renewables in electricity markets: operational problems. Springer Science & Business Media; 2013.

[2] International Energy Agency. Renewables 2024. 2024, Licence: CC BY 4.0. https://www.iea.org/reports/renewables-2024.

[3] Tuohy Aidan, Zack John, Haupt Sue Ellen, Sharp Justin, Ahlstrom Mark, Dise Skip, Grimit Eric, Mohrlen Corinna, Lange Matthias, Casado Mayte Garcia, Black Jon, Marquis Melinda, Collier Craig. Solar forecasting: Methods, challenges, and performance. IEEE Power Energy Mag 2015;13(6):50–9.

[4] Gonçalves Ana CR, Costoya Xurxo, Nieto Raquel, Liberato Margarida LR. Extreme weather events on energy systems: a comprehensive review on impacts, mitigation, and adaptation measures. Sustain Energy Res 2024;11.

[5] Abuella Mohamed, Chowdhury Badrul. Forecasting of solar power ramp events: A post-processing approach. Renew Energy 2019;133:1380–92.

[6] Zhang Jie, Zhu Xinchun, Xie Yigong, Chen Guo, Liu Shuangquan. Detection and prediction of wind and solar photovoltaic power ramp events based on data-driven methods: A critical review. Energies 2025;18(13).

[7] Florita Anthony, Hodge Bri-Mathias, Orwig Kirsten. Identifying wind and solar ramping events. In: 2013 IEEE green technologies conference (greenTech). 2013, p. 147–52.

[8] Samu Remember, Calais Martina, Shafiullah GM, Moghbel Moayed, Shoeb Md Asaduzzaman, Nouri Bijan, Blum Niklas. Applications for solar irradiance nowcasting in the control of microgrids: A review. Renew Sustain Energy Rev 2021;147:111187.

[9] Pan Yushu, Ju Liwei, Yang Shenbo, Guo Xinyu, Tan Zhongfu. A multi-objective robust optimal dispatch and cost allocation model for microgrids-shared hybrid energy storage system considering flexible ramping capacity. Appl Energy 2024;369:123565.

[10] Alam MJE, Muttaqi KM, Sutanto D. A novel approach for ramp-rate control of solar PV using energy storage to mitigate output fluctuations caused by cloud passing. IEEE Trans Energy Convers 2014;29(2):507–18.

[11] Shivashankar S, Mekhilef Saad, Mokhlis Hazlie, Karimi M. Mitigating methods of power fluctuation of photovoltaic (PV) sources – a review. Renew Sustain Energy Rev 2016;59:1170–84.

[12] Sukumar Shivashankar, Marsadek Marayati, Agileswari KR, Mokhlis Hazlie. Ramp-rate control smoothing methods to control output power fluctuations from solar photovoltaic (PV) sources—A review. J Energy Storage 2018;20:218–29.

[13] Wen Haoran, Du Yang, Chen Xiaoyang, Lim Enggee, Wen Huiqing, Jiang Lin, Xiang Wei. Deep learning based multistep solar forecasting for PV ramp-rate control using sky images. IEEE Trans Ind Inform. 2021;17(2):1397–406.

[14] Paletta Quentin, Arbod Guillaume, Lasenby Joan. Omnivision forecasting: Combining satellite and sky images for improved deterministic and probabilistic intra-hour solar energy predictions. Appl Energy 2023;336:120818.

[15] Ruan Guoping, Chen Xiaoyang, Lim Eng Gee, Fang Lurui, Su Qi, Jiang Lin, Du Yang. On the use of sky images for intra-hour solar forecasting benchmarking: Comparison of indirect and direct approaches. Sol Energy 2024;276:112649.

[16] Si Zhiyuan, Yang Ming, Yu Yixiao, Ding Tingting. Photovoltaic power forecast based on satellite images considering effects of solar position. Appl Energy 2021;302:117514.

[17] Abuella Mohamed, Chowdhury Badrul. Qualifying combined solar power forecasts in ramp events' perspective. In: 2018 IEEE power and energy society general meeting. PESGM, 2018, p. 1–5.

[18] Zhu Wenli, Zhang Li, Yang Ming, Wang Bo. Solar power ramp event forewarning with limited historical observations. IEEE Trans Ind Appl 2019;55(6):5621–30.

[19] Yang Dazhi, Wang Wenting, Gueymard Christian A, Hong Tao, Kleissl Jan, Huang Jing, Perez Marc J, Perez Richard, Bright Jamie M, Xia Xiang'ao, van der Meer Dennis, Peters Ian Marius. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. Renew Sustain Energy Rev 2022;161:112348.

[20] Carpentieri A, Folini D, Leinonen J, Meyer A. Extending intraday solar forecast horizons with deep generative models. Appl Energy 2025;377:124186.

[21] Lanzilao Luca, Meyer Angela. Intraday spatiotemporal PV power prediction at national scale using satellite-based solar forecast models. Energy AI 2026;25.

[22] Nouri Bijan, Fabel Yann, Blum Niklas, Schnaus Dominik, Zarzalejo Luis F, Kazantzidis Andreas, Wilbert Stefan. Ramp rate metric suitable for solar forecasting. Sol RRL 2024;8(24):2400468.

[23] Hendrikx NY, Barhmi K, Visser LR, de Bruin TA, Pó M, Salah AA, van Sark WGJHM. All sky imaging-based short-term solar irradiance forecasting with long short-term memory networks. Sol Energy 2024;272:112463.

[24] Chu Yinghao, Pedro Hugo TC, Li Mengying, Coimbra Carlos FM. Real-time forecasting of solar irradiance ramps with smart image processing. Sol Energy 2015;114:91–104.

[25] Gallego-Castillo Cristobal, Cuerva-Tejero Alvaro, Lopez-Garcia Oscar. A review on the recent history of wind power ramp forecasting. Renew Sustain Energy Rev 2015;52:1148–57.

[26] Meyer Angela. Satellite solar radiation forecasting by next-frame prediction – advances and future opportunities. Energy AI 2026;25.

[27] Carpentieri A, Folini D, Nerini D, Pulkkinen S, Wild M, Meyer A. Intraday probabilistic forecasts of surface solar radiation with cloud scale-dependent autoregressive advection. Appl Energy 2023;351:121775.

[28] Nielsen Andreas H, Iosifidis Alexandros, Karstoft Henrik. IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. Sol Energy 2021;228:659–69.

[29] Yang Dazhi, Wang Wenting, Bright Jamie M, Voyant Cyril, Notton Gilles, Zhang Gang, Lyu Chao. Verifying operational intra-day solar forecasts from ECMWF and NOAA. Sol Energy 2022;236:743–55.

[30] Wang Wenting, Yang Dazhi, Hong Tao, Kleissl Jan. An archived dataset from the ECMWF ensemble prediction system for probabilistic solar power forecasting. Sol Energy 2022;248:64–75.

[31] Sebastianelli Alessandro, Serva Federico, Ceschini Andrea, Paletta Quentin, Panella Massimo, Le Saux Bertrand. Machine learning forecast of surface solar irradiance from meteo satellite data. Remote Sens Environ 2024;315:114431.

[32] Schmetz Johannes, Pili Paolo, Tjemkes Stephen, Just Dieter, Kerkmann Jochen, Rota Sergio, Ratier Alain. AN introduction TO METEOSAT SECOND generation (MSG). Bull. Meteorol Soc. 2002;83(7):977–92.

[33] EUMETSAT Satellite Application Facility on Climate Monitoring. Meteosat High Resolution Solar Radiation (HANNA) Demonstrational Data Record 2019–2020. 2025, https://www.cmsaf.eu/demo_hanna. (Accessed: 24 July 2025).

[34] Pfeifroth U, Drücke J, Kothe S, Trentmann J, Schröder M, Hollmann R. SARAH-3 – satellite-based climate data records of surface solar radiation. Earth Syst Sci Data 2024;16(11):5243–65.

[35] Roberts CD, Senan R, Molteni F, Boussetta S, Mayer M, Keeley SPE. Climate model configurations of the ECMWF integrated forecasting system (ECMWF-IFS cycle 43r1) for HighResMIP. Geosci Model Dev 2018;11(9):3681–712.

[36] Stöckli Reto. The HelioMont Surface Solar Radiation Processing (2022 Version). MeteoSwiss; 2013, p. 126, Scientific Report 93.

[37] Castelli M, Stöckli R, Zardi D, Tetzlaff A, Wagner JE, Belluardo G, Zebisch M, Petitta M. The HelioMont method for assessing solar irradiance over complex terrain: Validation and improvements. Remote Sens Environ 2014;152:603–13.
[38] Märkle-Huß Joscha, Feuerriegel Stefan, Neumann Dirk. Contract durations in the electricity market: Causal impact of 15min trading on the EPEX spot market. Energy Econ 2018;69:367–78.
[39] Olivetti L, Messori G. Do data-driven models beat numerical models in forecasting weather extremes? A comparison of IFS HRES, pangu-weather and GraphCast. EGUsphere 2024;2024:1–35.
[40] Ineichen Pierre. A broadband simplified version of the solis clear sky model. Sol Energy 2008;82(8):758–62.
[41] Holmgren William F, Hansen Clifford W, Mikofski Mark A. Pvlib python: a python package for modeling solar energy systems. J Open Source Softw 2018;3(29):884.
[42] Chen Tianqi, Guestrin Carlos. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. KDD '16, New York, NY, USA: Association for Computing Machinery; 2016, p. 785–94.
[43] Steger CR, Steger B, Schär C. HORAYZON v1.2: an efficient and flexible ray-tracing algorithm to compute horizon and sky view factor. Geosci Model Dev 2022;15(17):6817–40.
[44] Zhang Zhongwei, Fischer Erich, Zscheischler Jakob, Engelke Sebastian. Physics-based models outperform AI weather forecasts of record-breaking extremes. Sci Adv 2026;12.
[45] Charlton-Perez Andrew, Dacre Helen, Driscoll Simon, Gray Suzanne, Harvey Ben, Harvey Natalie, Hunt Kieran, Lee Robert, Swaminathan Ranjini, Vandaele Remy, Volonté Ambrogio. Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of storm ciarán. Npj Clim Atmospheric Sci 2024;93.
[46] EUMETRAIN. False Alarm Ratio (FAR), Post Agreement (PAG) and False Alarm Rate (F). 2026, https://resources.eumetrain.org/data/4/451/english/msg/ver_categ_forec/uos2/uos2_ko3.htm. (Accessed: 19 August 2026).
[47] EUMETRAIN. Frequency Bias (B or FBI). 2026, https://resources.eumetrain.org/data/4/451/english/msg/ver_categ_forec/uos2/uos2_ko1.htm. (Accessed: 19 August 2026).
[48] EUMETRAIN. Critical Success Index (CSI) or Threat Score (TS), and Equitable Threat Score (ETS). 2026, https://resources.eumetrain.org/data/4/451/english/msg/ver_categ_forec/uos2/uos2_ko4.htm. (Accessed: 19 August 2026).
[49] Official Journal of the European Union. COMMISSION REGULATION (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing. 2017, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R2195. (Accessed: 21 August 2026).