

Extending intraday solar forecast horizons with deep generative models

Carpentieri, A.; Folini, D.; Leinonen, J.; Meyer, A.

DOI

[10.1016/j.apenergy.2024.124186](https://doi.org/10.1016/j.apenergy.2024.124186)

Publication date

2024

Document Version

Final published version

Published in

Applied Energy

Citation (APA)

Carpentieri, A., Folini, D., Leinonen, J., & Meyer, A. (2024). Extending intraday solar forecast horizons with deep generative models. *Applied Energy*, 377, Article 124186. <https://doi.org/10.1016/j.apenergy.2024.124186>

Important note

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

Copyright

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.



Extending intraday solar forecast horizons with deep generative models

A. Carpentieri^{a,b,*}, D. Folini^a, J. Leinonen^{c,1}, A. Meyer^{b,d}

^a Institute for Atmospheric and Climate Science, ETH Zurich, Universitaetstrasse 16, 8092 Zurich, Switzerland

^b School of Engineering and Computer Science, Bern University of Applied Sciences, Quellgasse 21, 2501 Biel, Switzerland

^c Federal Office of Meteorology and Climatology MeteoSwiss, Via ai Monti 146, 6605 Locarno-Monti, Switzerland

^d Department of Geoscience and Remote Sensing, TU Delft, Stevinweg 1, 2628 CN Delft, Netherlands

ARTICLE INFO

Keywords:

Solar irradiance
Spatiotemporal forecasting
Renewable energy integration
Generative deep learning
Diffusion models
Ensemble forecasts

ABSTRACT

Surface solar irradiance (SSI) plays a crucial role in tackling climate change — as an abundant, non-fossil energy source, exploited primarily via photovoltaic (PV) energy production. With the growing contribution of SSI to total energy production, the stability of the latter is challenged by the intermittent character of the former, arising primarily from cloud effects. Mitigating this stability challenge requires accurate, uncertainty-aware, near real-time, regional-scale SSI forecasts with lead times of minutes to a few hours, enabling robust real-time energy grid management. State-of-the-art nowcasting methods typically meet only some of these requirements. Here we present SHADECast, a deep generative diffusion model for the probabilistic spatiotemporal nowcasting of SSI, conditioned on deterministic aspects of cloud evolution to guide the probabilistic ensemble forecast, and based on near real-time satellite data. We demonstrate that SHADECast provides improved forecast quality, reliability, and accuracy in all weather scenarios. Our model produces realistic and spatiotemporally consistent predictions extending the state-of-the-art forecast horizon by 26 min over different regions with lead times of 15–120 min. Our physics-informed generative approach leads to up to 60% performance improvement in extreme value prediction over the state-of-the-art deterministic models, showcasing the advantage of probabilistic modeling of cloudiness over the classical deterministic approach. It also surpasses the probabilistic benchmarks in predicting extreme values. Finally, SHADECast empowers grid operators and energy traders to make informed decisions, ensuring stability and facilitating the seamless integration of PV energy across multiple locations simultaneously.

1. Introduction

Harvesting solar energy resources is an essential pillar in efforts to mitigate climate change [1,2]. Photovoltaic (PV) power generation increased by 26% in 2022, accounting for two-thirds of the increase in global renewable capacity for 2023 [3]. In concert with the growing relevance of PV for total energy production, the challenge arising from the intermittent character of surface solar irradiance (SSI) increases [4]. Power production and consumption, linked via transmission and storage capacities, should be closely balanced at any moment in time [5]. The naturally arising volatility of PV production, primarily due to changing cloudiness, impacts the reliability of the electricity grid [6]. A key element in dealing with this challenge – and the topic of this paper – are regional-scale, near real-time, uncertainty-aware SSI forecasts with lead times of minutes to hours. Such forecasts enable operational management of energy production from alternate sources, such as gas turbines [3,7], facilitate the proactive scheduling of energy-intensive

industrial operations [1,8], and thus reduce operation uncertainty and stand-by costs [1,9].

The relevance of the topic spurred progress in SSI forecasting. Yet, there remains ample room and an urgent need for substantial further improvement. State-of-the-art methods span a wide range of approaches. For short lead times of up to a few hours, which are the focus of this work, data-driven methods prevail, with numerical weather prediction [10] playing only a minor role. One distinguishing feature is the input data used. Ground-based in-situ measurements of SSI have the advantage of being highly accurate, but their limited spatial representativeness [11] discourages their exclusive use for regional-scale forecasts.

Satellite-derived solar irradiance estimates offer a trade-off between accuracy and spatial coverage, which makes them highly suitable for

* Corresponding author at: Institute for Atmospheric and Climate Science, ETH Zurich, Universitaetstrasse 16, 8092 Zurich, Switzerland.

E-mail address: acarpentieri@usys.ethz.ch (A. Carpentieri).

¹ Currently at NVIDIA Corporation.

<https://doi.org/10.1016/j.apenergy.2024.124186>

Received 25 March 2024; Received in revised form 2 July 2024; Accepted 7 August 2024

Available online 5 September 2024

0306-2619/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

List of Abbreviations

<i>AFNO</i>	Adaptive Fourier neural operator
<i>AR</i>	Autoregressive
<i>ConvLSTM</i>	Convolutional long short-term memory
<i>CRPS</i>	Continuous ranked probability score
<i>CSI</i>	Clear-sky index
<i>FSS</i>	Fraction skill score
<i>MAE</i>	Mean absolute error
<i>MSE</i>	Mean squared error
<i>PICP</i>	Prediction interval coverage probability
<i>PINAW</i>	Prediction interval normalized average width
<i>PLMS</i>	Pseudo-linear multistep
<i>PV</i>	Photovoltaic
<i>RMSE</i>	Root mean squared error
<i>SEVIRI</i>	Spinning enhanced visible and infrared imager
<i>SSI</i>	Surface solar irradiance
<i>SZA</i>	Solar zenith angle
<i>VAE</i>	Variational autoencoder

short-term SSI forecasting over extended regions, enabling simultaneous SSI forecasts for multiple sites [12]. Various data-driven SSI forecast methods that rely on satellite data exist, notably statistical methods [13–18], deep learning models [19–23], and hybrid approaches that also incorporate numerical weather predictions [24,25].

The majority of these approaches, despite using satellite data as input, provide forecasts only at individual locations [12,23], which is not suitable for managing arbitrarily large grids [23]. Approaches providing regional scale forecasts are mostly deterministic [26], which again limits their practical use for lack of forecast uncertainty quantification. Also, existing deterministic models tend to generate blurry forecasts, as illustrated by recent studies comparing convolutional recurrent neural networks and optical flow methods [19]. The blurriness results from the mean squared error (MSE) minimization, which causes predictions to converge towards the mean of the distribution of all possible future SSI evolutions [27,28]. The resulting forecasts lack the spatial granularity required to accurately represent the stochastic spatiotemporal behavior of SSI. Furthermore, the tendency to converge toward the mean restricts deterministic models from accurately predicting extreme SSI values, which are crucial for estimating future PV production.

Spatiotemporal regional scale SSI forecasts with uncertainty quantification are still scarce. In [15], an Analog Ensemble method is applied to retrieve past SSI field sequences (analogs) based on four similarity metrics and project them into the future to generate an ensemble of forecasts. The analog-based approach can be effective but requires a huge amount of past data and a complete search in the dataset for each forecast. A more flexible ensemble-based approach is proposed in [18], where scale-dependent autoregressive (AR) models are applied to probabilistically forecast clear-sky index (CSI) fields in a Monte Carlo sampling approach. However, linear AR models assume stationarity in the data, making the model unable to predict distribution shifts. Differently, in [20], a deterministic convolutional long short-term memory (ConvLSTM) model is modified to directly forecast the probability of each pixel CSI value inside different ranges. This classification based procedure drastically increases the dimensionality of the output by a factor of 240, making it impractical for large-area and multi-step forecasts. Both approaches make use of CSI, a dimensionless index measuring the atmosphere effect on the incoming SSI.

Here we present the Solar High-resolution Adaptive Diffusion Ensemble forecasting model (SHADECast), producing uncertainty-aware

regional-scale SSI forecasts that model probabilistic cloud formation, evolution, and dissipation, conditioned on a data-driven deterministic cloud field forecast. Our approach is novel in that it combines insight from atmospheric physics – leading us to split the task into a deterministic part upon which a probabilistic part then acts – with inspiration from probabilistic video forecasting, where generative deep learning models have emerged as the new state of the art due to their adeptness in modeling data distributions, enabling the sampling of realistic future scenarios [29]. Notably, diffusion models [30,31] have exhibited superior performance in image and video generation tasks [29,32]. In precipitation nowcasting, they provide superior characterization of the distribution of possible outcomes compared to generative adversarial networks [28,33].

SHADECast is, to the best of our knowledge, the first uncertainty-aware, physics-inspired, satellite-based regional-scale forecast model for intraday SSI forecasts. As we are going to demonstrate, SHADECast produces skillful, sharp and reliable solar forecasts without blurring under variable weather conditions, thanks also to our innovative, physics motivated splitting of the task at hand.

We assess our model's performance by comparing it with three benchmark models: two probabilistic models, SolarSTEPS [34] and an adaptation of the precipitation nowcasting model LDCast [33], as well as a deterministic model based on a ConvLSTM architecture (IrradianceNet [20]). By addressing the challenges of dynamic weather conditions and the need for accurate extreme value predictions, our approach represents a significant advancement in the field of SSI forecasting. The following sections will detail our methodology and the specific problem our model aims to solve, demonstrating the innovative aspects of our approach.

This paper is organized as follows. Section 2 details the proposed generative short-term forecasting approach. The dataset, benchmarks, and metrics used for characterizing the forecast quality are described in Section 3. Section 4 discusses the results of our approach. Section 5 presents conclusions and opportunities for future research

2. Generative short-term forecasting

Our goal is to generate an ensemble forecast consisting of future CSI fields \hat{C} that are consistent with previous CSI fields C observed shortly before the time when the forecast is made. Based on a sequence of m observed fields $C_{t-m+1:t}$, we want to forecast n future fields $\hat{C}_{t+1:t+n}$ by means of a forecasting process f_θ starting at time t ,

$$\hat{C}_{t+1:t+n} = f_\theta(C_{t-m+1:t}, \epsilon) \quad (1)$$

with free parameters θ whose optimal values θ^* are determined by minimizing the distance between the estimated conditional probability distribution of forecasted cloudiness fields $p_\theta(\hat{C}_{t+1:t+n}|C_{t-m+1:t})$ and the actual distribution of the future fields $p(C_{t+1:t+n}|C_{t-m+1:t})$. The normally distributed random variable ϵ is sampled multiple times to draw individual ensemble members of the forecast from p_θ according to Eq. (1). SHADECast offers a concrete realization of this general concept.

The SHADECast forecast generation pipeline, depicted in Fig. 1, integrates a variational autoencoder (VAE) for data compression, a latent deterministic nowcaster based on Adaptive Fourier Neural Operator (AFNO) blocks [35,36], and a latent diffusion model represented by the denoiser. These components collaboratively forecast an ensemble of future cloudiness field sequences. The nowcaster's deterministic forecast guides the ensemble generation by the denoiser [31]. With respect to previous SSI nowcasting methods and to LDCast, an important conceptual innovation of our model lies in the decomposition of the forecasting task into a deterministic forecast (nowcaster) for large-scale dynamics and a probabilistic ensemble generation (diffusion) to model high-uncertainty regions.

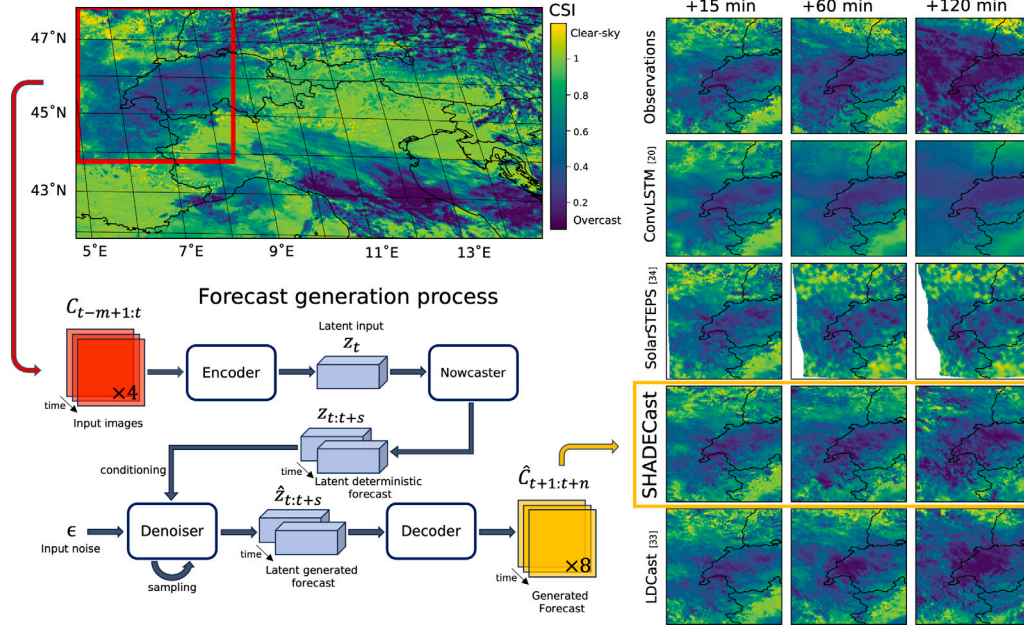


Fig. 1. SHADECast forecast generation pipeline. **Upper left panel:** Example CSI field of 24 Feb. 2016 at 11:45 UTC. The red box highlights the region forecasted in the right panel. **Lower left panel:** SHADECast forecast generation pipeline. The input CSI fields ($C_{t-m+1:t}$) are fed to the encoder, which projects the image sequence to the latent space, obtaining z_t . Then, the deterministic nowcaster forecasts the future latent representation of the CSI fields ($z_{t+1:t+s}$), where s is the lead time in the latent space, which can differ from n due to data compression. The latent forecast is, then, fed to the denoiser together with Gaussian noise ϵ . The pseudo linear multi-step (PLMS) sampler employs the denoiser to generate an ensemble member. The decoder finally decompresses the latent ensemble forecast, obtaining $\hat{C}_{t+1:t+n}$. **Right panel:** Forecasts made by SHADECast (yellow box) and benchmark models for lead times up to 120 min. For SHADECast, LDCast and SolarSTEPS the ensemble member chosen is the one with the lowest average root mean squared error (RMSE). The first row shows the satellite-derived CSI fields.

2.1. Surface solar irradiance

SSI can be expressed as the product of the clear-sky SSI, SSI_{cs} , and the clear-sky index, CSI, so $SSI = CSI \cdot SSI_{cs}$. The clear-sky SSI is an estimate of SSI in the absence of clouds. SSI_{cs} mainly depends on the solar zenith angle (SZA), its diurnal and annual cycle, and to a minor degree on aerosols and atmospheric trace gases like water vapor. The remaining most relevant factor affecting SSI are clouds, which are also the most difficult component to forecast. CSI is a dimensionless variable that quantifies the degree of cloudiness, which makes CSI a particularly suitable variable to forecast [20,34].

The temporal evolution of cloudiness fields may be seen as a composite of wind-driven cloud advection and cloud evolution – the formation, growth, and dissipation of clouds – governed by processes such as microphysics and turbulence [34]. While cloud advection and cloud evolution cannot be separated from each other in a strict physical sense, it pays off to do so in the context of forecasting, as we demonstrate below.

2.2. SHADECast

SHADECast is a conditional latent diffusion model incorporating AFNO blocks [35], known for their efficacy in modeling chaotic systems like weather [36]. With respect to current SSI forecasting models and LDCast [33], the architectural innovation of SHADECast is the incorporation of an independently-trained AFNO-based forecasting model as conditioning model (nowcaster in Fig. 1). The nowcaster focuses on forecasting large-scale components of the dynamics of cloudiness, while the diffusion model (denoiser) is responsible for forecasting the chaotic dynamics of small scales, thus generating ensembles of possible future evolutions.

The core concept of diffusion models entails forward diffusion and backward denoising processes [30,31]. The forward diffusion process iteratively introduces disruptive Gaussian noise into training data samples, whereas the backward process iteratively removes the noise from

the noisy output of the forward process, restoring the data sample to its original state. Fundamentally, the denoising process is implemented to enable the model to learn the mapping of a known simple distribution (usually an uncorrelated Gaussian) to the data distribution, enabling the generation of realistic and accurate data samples.

Our conditional latent diffusion model consists of three main components as depicted in Fig. 2:

1. A VAE, which compresses (decompresses) the data into (from) the latent space. Following the approach in [37], modeling diffusion in the latent space achieves an optimal trade-off between accuracy and efficiency.
2. A latent AFNO-based deterministic nowcaster. It takes the latent representation of the input CSI maps and forecasts consecutive maps in the latent space. The number of latent time steps is increased using a temporal transformer [38]. It can be used as an independent forecasting model.
3. A latent denoiser, which maps Gaussian noise to the future CSI maps in the latent space. Based on a U-Net architecture [39], it is conditioned on the nowcaster's output through AFNO Cross Attention blocks (Fig. 2).

The forecast generation process shown in Fig. 1 involves the encoding of m past input CSI fields $C_{t-m+1:t}$ into the latent space, resulting in the latent tensor z_t with an overall compression factor of 2. Then, the nowcaster performs a forecast in the latent space ($z_{t+1:t+s}$). Here, s represents the number of forecasted steps in the latent space, which are related to n by $s = \frac{n}{c_t}$, where c_t is the compression factor along the time dimension. Then, $z_{t+1:t+s}$ is employed to condition the denoiser that generates the forecast ensemble. The conditioning is performed by downsampling the deterministic forecast to match the dimensions of the U-Net layers of the denoiser (Fig. 2). The conditioning step is essential to guide the denoising process towards realistic future scenarios of cloudiness evolution. In summary, the goal of the denoiser is to project the input noise tensor (ϵ) to the latent representation of the future n satellite observations ($C_{t+1:t+n}$). It does so by iteratively

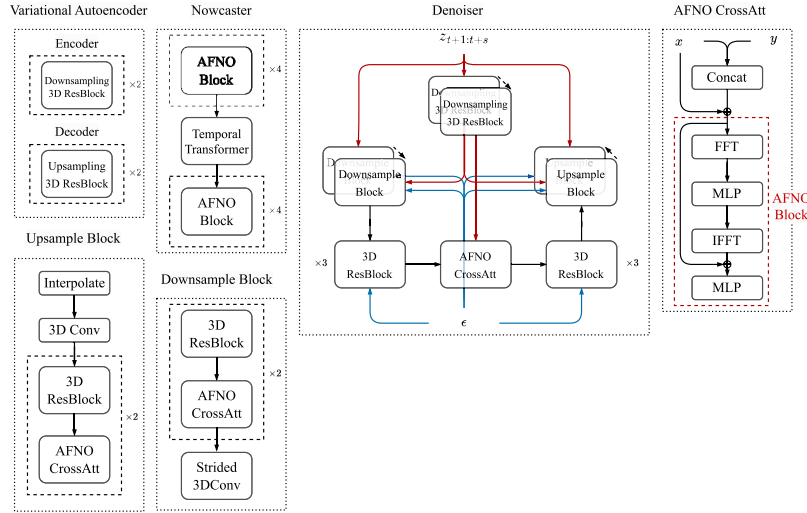


Fig. 2. SHADECast architecture. A AFNO-based U-Net architecture is employed in our denoiser, alongside principal blocks integrated into the SHADECast architecture. The symmetrical design of the denoiser includes two downsampling and upsampling blocks. The latent forecast $z_{t+1:t+s}$ undergoes 3-dimensional strided residual blocks to match spatial dimensions with U-Net components, followed by concatenation with the output of AFNO cross attention blocks. In the right panel, x represents the input from the previous layer, and y is the conditioning input. Downsampling is achieved through strided 3D convolutional layers, whereas upsampling utilizes spatial axis interpolation. The 3-dimensional residual blocks consist of two convolutional layers connected by a skip connection.

performing numerous denoising steps [31]. In our case, the CSI field sequence generation is governed by a pseudo-linear multistep sampler (PLMS) [40] to reduce the number of required denoising steps. PLMS permits to decrease the number of steps from ≈ 1000 to 25, maintaining the sample quality (refer to Supplementary Table A.1). Finally, the sampled sequence $z_{t+1:t+s}$ is decoded back by the decoder to the final forecast ensemble member $\hat{C}_{t+1:t+n}$.

Variational autoencoder

The VAE exhibits a symmetrical architecture, as in [41]. The VAE's encoder processes 4-dimensional inputs, specifically sequences of CSI fields. This encoding phase consists of two downsampling 3-dimensional residual blocks, outputting two tensors, μ and Σ , namely the mean and covariance matrices of a Gaussian distribution. They serve as the foundation for the decoder's sampling process that samples a latent vector from the latent Gaussian and reconstructs it into a sequence of CSI fields. The downsampling and upsampling blocks in the VAE mirror those in Fig. 2, with the exception of the cross-attention layer.

The CSI field sequence is represented as a four-dimensional tensor with dimensions (C_h, T, W, H) , where C_h denotes the number of image channels, T is the time dimension, and (W, H) represents the width and height of a single CSI map. In the latent space, dimensions T , W , and H are reduced by a factor of 4, while C_h is increased by a factor of 32, resulting in an overall compression factor of 2, as done in [33].

Regularization in the latent space is achieved through the Kullback-Leibler (KL) divergence between the latent data distribution and $N(0, 1)$. The reconstruction loss is quantified by the mean absolute error that measures the disparity between the VAE's input and the decoder output. The final loss is an interpolation between the two losses, with a coefficient of 0.05 for the KL loss.

The VAE comprises approximately 800,000 parameters. For detailed architecture parameters, please refer to the training configuration file available in our GitHub repository.

Nowcaster

The AFNO-based nowcaster consists of four AFNO blocks, a temporal transformer [38], and another four AFNO blocks. The AFNO blocks [35] (Fig. 2) transform the input using a 3-dimensional Fast Fourier Transform (FFT) applied to the temporal and spatial axes of the input tensor. Subsequently, a multilayer perceptron (MLP) processes the transformed data along the channel dimension. Finally, the data

undergoes inverse-FFT (IFFT), is summed with the original input, and processed by another MLP.

The temporal transformer is employed to increase the time steps through cross attention between the input and a sinusoidal time embedding tensor [33]. The time steps are increased by a factor of 2, resulting in $s = 2$ in Fig. 1.

The nowcaster operates in the latent space following the approach in [33]. Computing the AFNO in the latent space aligns with the method in [36], where the authors utilized an embedding procedure to increase the channel dimension at the expense of H and W . Through channel mixing in the Fourier space, we approximate global attention [35], as each pixel in the Fourier space contains information on the entire image.

The loss chosen is the Mean Absolute Error (MAE) and it is computed in the latent space. By computing the loss in the latent space we noticed two major improvements. First, we save one iteration (the decoding). Second, the forecasts result more detailed and less blurry even at longer lead times.

Overall, the architecture of the nowcaster comprises ~ 6 M parameters.

Denoiser

The denoiser's AFNO-based U-Net architecture, depicted in Fig. 2, is symmetrical and comprises two main components: downsampling and upsampling blocks. The denoising process begins with the latent forecast $z_{t+1:t+s}$, which is downsampled with 3-dimensional strided residual blocks. This step is crucial for achieving spatial dimension alignment with the U-Net's downsampling and upsampling blocks. The resulting output is then concatenated with the output of AFNO cross attention blocks, denoted as x and y for the input from the previous layer and the conditioning input, respectively.

For downsampling, we employ strided 3D convolutional layers, effectively reducing spatial dimensions (height and width). Conversely, upsampling is achieved through interpolation on the height and width axis of the tensor. The backbone of the architecture consists of 3-dimensional residual blocks, featuring two convolutional layers with a skip connection to enhance feature extraction. The denoiser is trained to predict the noise as done in [37] and the chosen loss function is the MSE. Moreover, as also done in [37], the exponential moving average method is employed to stabilize the training.

This detailed architecture is visually represented in Fig. 2. The denoiser is defined by approximately 320M parameters.

3. Performance evaluation

The probabilistic spatiotemporal forecasting model presented in Section 2 is trained to forecast cloudiness fields and, consequently, SSI regional fields. The dataset, training procedure, evaluation metrics and benchmark models are presented in the following.

3.1. Dataset

The CSI fields employed for this study are derived from spectral measurements of Earth taken by the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) on board the Meteosat Second Generation geostationary satellite [42]. The raw satellite images are processed by the HelioMont radiative transfer algorithm [43] to produce two-dimensional CSI fields. We refer to [18] for a comprehensive review of the dataset. The dataset spans 10 years from 2007 to 2016 at a temporal resolution of 15 min. The time period is motivated by constraints on data availability. The HelioMont CSI fields are only available for solar zenith angle (SZA) lower than 88° . The spatial resolution is approximately $0.02^\circ \times 0.02^\circ$. The region covered ranges from 8.3°E , 44.8°N to 12.8°E , 49.1°N corresponding to images of size $384 \text{ px} \times 768 \text{ px}$ in the native Geostationary projection as shown in Fig. 1. Missing pixels are filled by a linear three-dimensional (time, longitude and latitude) interpolation if they cover less than 2% of the image, otherwise the image is discarded.

Seven years of data are used for the model training (2007–2013) and one for the validation (2014), while two years are kept for the final testing (2015–2016). For training and validation, we cropped the maps into $18 \times 128 \text{ px} \times 128 \text{ px}$ patches as shown in Fig. 3. Therefore, for the training set we have 18 regions and 365×7 days of data split into overlapping 12-step sequences (4 input and 8 output maps).

To create the test set, we randomly sampled 200 days from the 2-year period (2015–2016), and then randomly sampled 4 input sequences from each of the 200 days, resulting in 800 CSI satellite image sequences for every test set region. We make use of three $256 \text{ px} \times 256 \text{ px}$ regions, namely the areas corresponding to patches (a), (b), and (c) as illustrated in Fig. 3. Using larger images for the validation (with respect to the training set) accounts for the advection effect during the forecast lead time, aiming to reduce areas completely generated by the model. At maximum speed, clouds can cross most of the 128 pixels ($\approx 250 \text{ km}$) in less than 2 h, and so the model would generate most of the forecast with no information on coming clouds. The use of smaller image patches in training was driven by memory and computational constraints. Notably, our model's architecture enables the forecast of arbitrarily large images.

In Fig. 3 we show average and standard deviation of CSI for every pixel covered by HelioMont dataset. The values are computed daily for 500 randomly sampled days from the training set and then averaged.

3.2. Data processing and training

The number of CSI fields from Meteosat SEVIRI in a day depends on the daylight hours, resulting in a higher number of available data samples during summer as HelioMont cannot derive CSI at night. To mitigate this bias in our model, we generate each training sample by randomly selecting one day from the 365×7 available days and then selecting a sequence of maps from that day. This ensures that the models are trained on a balanced dataset, exposing them to an equal number of summer and winter sequences.

The validation set follows a similar sampling approach as the training set but with a fixed structure: sequences and days are sampled once, and these validation samples remain constant throughout the validation process. This is done to obtain a consistent validation through the training process.

To facilitate model convergence and performance, the data are normalized by mapping the values to the $[-1, 1]$ range. The normalization

is straightforward as HelioMont CSI values are bounded in the $[0.05, 1.2]$ range.

The three components of SHADECast (autoencoder, nowcaster and denoiser) are trained independently. The training and validation sets are the same for the three training processes. The training is terminated if the validation loss does not decrease for at least 10 epochs (early stopping). Moreover, after 5 epochs without improvement, the learning rate is divided by a factor of 4. The initial learning rate for the VAE and the nowcaster is 10^{-3} , while for the Denoiser we set 10^{-4} . Similarly, the batch size is set to 256, 240 and 96 for the VAE, Nowcaster, and Denoiser, respectively. The batch sizes are chosen to maximally exploit the available GPU memory and stabilize the training.

Computational requirements

The training of the SHADECast diffusion model requires approximately 500 GPU hours on 24 Nvidia P100 GPUs. Generating a 2-hour forecast ensemble of 10 members at 15-min resolution for 256×256 pixels images, takes 1 min on a single Nvidia P100 GPU.

3.3. Benchmark models

To assess our model's performance, we compare it to two probabilistic ensemble-based benchmark models: SolarSTEPS [34] and LDCast [33]. In order to use the latter, we adapted and trained the original precipitation nowcasting model to forecast cloudiness. It shares the same architecture as SHADECast (see Fig. 2) for a fair comparison. In LDCast, the forecaster component is trained together with the denoiser and is, in fact, a feature extractor on the input maps. Therefore, the model is not conditioned on a deterministic forecast but, indirectly, on the input CSI fields. LDCast is chosen to illustrate that our physics-motivated choice of a separate nowcaster indeed improves the forecast accuracy and reliability of the ensemble. Overall, the training procedure and the data used are the same as for SHADECast training.

SolarSTEPS [34] is an optical-flow based approach, which was shown to outperform state-of-the-art models in the task of probabilistically forecasting satellite-derived CSI maps over Switzerland. Therefore, we consider it a valuable benchmark case in the present paper. The SolarSTEPS approach is based on the scale-dependent temporal variability of cloudiness: the small scales have a shorter lifetime with respect to bigger scale. The different scales' temporal evolution is thus modeled independently by different linear AR models. The approach permits the model to predict both the motion (optical-flow) and evolution (AR models) of cloudiness. The ensemble generation is governed by perturbing the AR models with a novel technique to generate spatially correlated CSI fields based on the short-space Fourier transform [44]. The method presented in [44] is modified to take into account the variability of the input maps in the generation of the perturbing fields. Moreover, SolarSTEPS has shown to outperform trivial benchmark models such as the persistence model. The parameterization used in our evaluation reflects the one presented in [34].

Moreover, we investigate the advantages of our probabilistic approach in comparison to the state-of-the-art deterministic model in cloudiness forecasting, IrradianceNet [20]. The model architecture remains consistent with the original paper. We retrained the model using only cloudiness fields on our training set as done for LDCast and SHADECast. In the original paper, the authors conducted a 2-step forecast. For comparability with other models, we autoregressively forecast 8 steps into the future. The model is trained on 128×128 images and tested on 256×256 similarly as done in [20]. Due to the model architecture limitations, forecasting arbitrarily large images is not possible. Consequently, a linear interpolation is applied on the borders of the individual forecasts, as detailed in [20]. It is important to note that the output interpolation introduces visible artifacts along the borders of single forecasts.

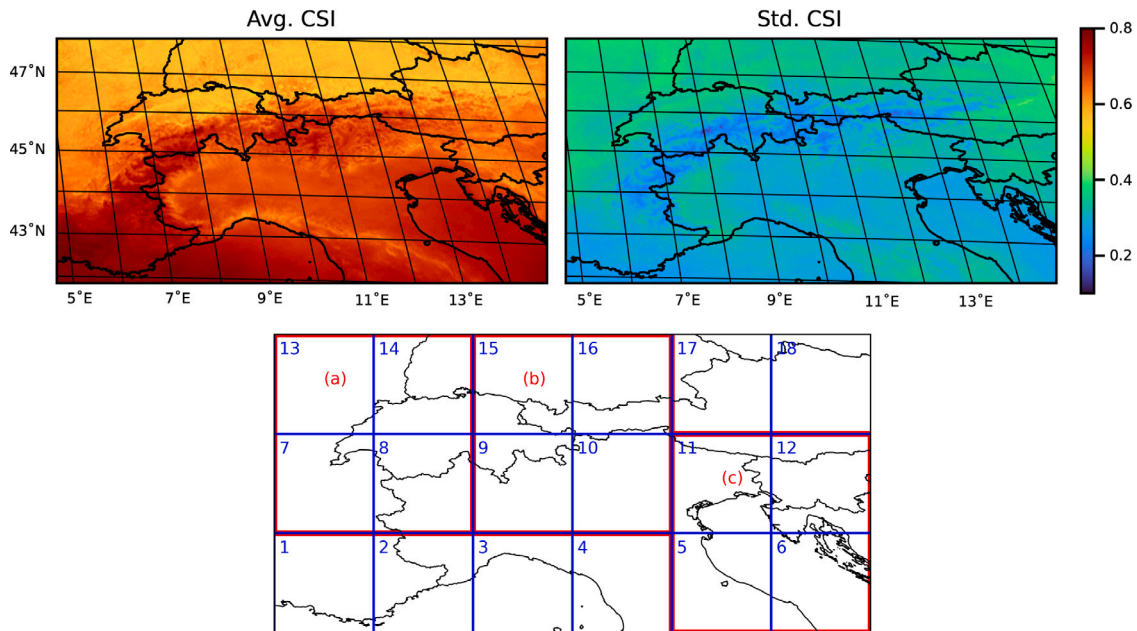


Fig. 3. Dataset overview. **Upper panel:** Average and standard deviation at pixel level of CSI values computed on 500 days sampled from the training set. The average daily CSI standard deviation is computed along the time dimension. For every sampled day, the standard deviation along the time dimension is computed for every pixel and then averaged over the 500 days. **Lower panel:** area covered by the Heliomont dataset [43]. The patches outlined in blue define the cropping applied to create the training set. For the test set we used three 256×256 patches identified by the red borders: (a), (b) and (c).

3.4. Metrics

The evaluation of the forecast ensembles is carried out by using probabilistic and deterministic metrics. For probabilistic forecasts, the main properties we evaluate are the reliability and sharpness of the forecast ensembles [45].

A reliable forecast ensemble is characterized by the observed value falling within the predicted ensemble. In an ideal scenario where the model accurately captures the uncertainty of the dynamics, the observations should be uniformly distributed within the ensemble. To assess this distribution, rank histograms [46] depict the frequency of the observed value's location among the ensemble members. In practical terms, a concave histogram signals under-confidence, indicating that the model tends to overestimate uncertainty. This results in forecasts with excessively high variance, suggesting a wider range of possibilities than observed. Conversely, a convex histogram signals overconfidence, indicating that the ensemble is too narrow and fails to adequately capture the actual uncertainty in the system dynamics. In such cases, the forecasted range is too restrictive, leading to potential underestimation of the true variability in the observed values.

Reliability is also described by the Prediction Interval Coverage Probability (PICP). PICP measures the percentage of observed values that lie in the ensemble prediction interval. We randomly sample 1000 pixels for each lead time image and check whether they fall inside the 5% and 95% percentiles of our forecast. However, PICP does not provide any information on the forecast informativeness, as an overdispersive model could lead to high PICP values. For this reason, we also measure the Prediction Interval Normalized Averaged Width (PINAW). PINAW measures the width of the prediction interval and so, it provides information on the forecast sharpness. An ideal forecast should reflect high PICP values and a low PINAW.

The CRPS is employed to evaluate the overall quality of probabilistic SSR forecasts [45,47]. CRPS accounts for both reliability and sharpness. It does so by measuring the distance between the cumulative density function of the ensemble F and the Heaviside function H centered on the observation y . The normalized CRPS for the i th pixel is then defined as:

$$\text{nCRPS}_i = \frac{1}{\text{CSI}_{\max}} \int_{-\infty}^{+\infty} (F_i(c) - H(c - y_i))^2 dc \quad (2)$$

The Heaviside function centered in y_i represents the ideal cumulative distribution for a perfect probabilistic forecast and F_i is the forecasted cumulative distribution for the i th pixel. CRPS, then, measures the distance between the Heaviside function and F_i for every point c in the F_i domain. It is computed at pixel level and averaged for every forecast step, ending up with a CRPS value for every forecasted pixel. We consider the normalized CRPS (nCRPS) by normalizing the CRPS with the maximum clear-sky index value, which is $\text{CSI}_{\max} = 1.2$.

The models performance in predicting extreme cloudiness values is measured by using the fraction skill score (FSS) [48,49]. Using a sliding window, FSS computes the accuracy of predicting the frequency of a particular event. In our case, the event corresponds to CSI values exceeding a threshold. A perfect forecast is defined by FSS equal 1, vice versa the worst forecast would have FSS = 0. We define $O_{i,j}^n$ and $F_{i,j}^n$ as two neighborhoods of size n centered in i, j from the observed and forecasted fields, respectively. $O_{i,j}^n$ and $F_{i,j}^n$ define the frequency of exceeding a set threshold inside the neighborhood, then FSS can be defined as:

$$F.S.S = 1 - \frac{\sum_i \sum_j (O_{i,j}^n - F_{i,j}^n)^2}{\sum_i \sum_j (O_{i,j}^n)^2 + (F_{i,j}^n)^2} \quad (3)$$

Finally, the normalized Root Mean Square Error (nRMSE) is used to measure the accuracy of the ensemble mean. The ensemble mean serves as a representative estimate of the central tendency of the forecasted distribution. Evaluating its accuracy provides insights into how well the ensemble captures the expected or average outcome. The nRMSE for a forecasted map is defined as:

$$\text{nRMSE} = \frac{1}{\text{CSI}_{\max}} \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (4)$$

4. Results and discussion

In the following, the forecast performance of SHADECast is discussed and compared to state-of-the-art SSI forecasting models.

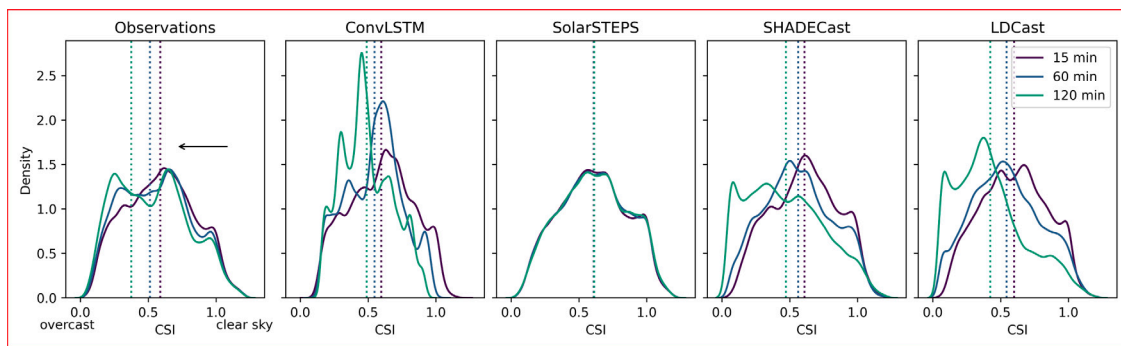


Fig. 4. Probability density distributions of the forecasted fields. The probability density distributions are computed on the forecasted CSI fields relative to the case study presented in Fig. 1 for three lead times (+15 min, +60 min, +120 min). The distributions are shown for the ground truth satellite-derived CSI fields (Observations), for SHADECast and three benchmark models. For SHADECast, LDCast and SolarSTEPS, the chosen ensemble member is the best performing one in terms of RMSE. The RMSE is computed for the different ensemble members over all pixels and timesteps, then the ensemble member with the lowest RMSE is selected for each model. The dotted vertical lines represent the distribution mean.

4.1. Clouds forming, evolving, dissipating

In the right panel of Fig. 1, we show an example of a forecast generated by SHADECast (highlighted in yellow). We present the ground truth in the first row, a deterministic forecast generated by a convolutional LSTM model based on [20] in the second row, SHADECast in the fourth row and the two benchmark models in the remaining rows. This particular case study is selected to exemplify the dynamic nature of cloud evolution throughout the forecast period. This phenomenon is visually represented by observing the shift in the mean of the CSI distribution towards lower values, as illustrated in Fig. 4.

The ConvLSTM forecast is relatively accurate within the initial 15 min, but its quality gradually diminishes afterwards due to increasing blurriness and the inability of the deterministic model to handle uncertainty. The observed lack of small-scale structures is linked to the convergence towards the mean [27] due to the pixel-level MSE minimization performed in the training. As highlighted in the introduction, our objective is modeling the distribution of potential outcomes, as the average of all outcomes (MSE minimum) does not necessarily align with the most probable outcome. SHADECast effectively simulates diverse cloudiness evolution in high-uncertainty regions, providing insights into variations that might appear indistinct in deterministic forecasts. On the other hand, the model can recognize low-uncertainty regions and keep them relatively unaltered among the ensemble members. In Fig. 1, the Alps region (bottom right area in the map) remains cloud-free throughout the 2-hour period. Similar patterns in the same region are evident in the SHADECast ensemble members but not in the benchmark probabilistic models (LDCast and SolarSTEPS). This case study demonstrates the adaptability of SHADECast in capturing ground truth uncertainty and projecting it into the forecast ensemble while retaining the less uncertain patterns. Additional forecast examples for the three test regions are also presented (see Figs. 8, 9, 10).

A distinguishing feature of SHADECast is that it allows for changes of the CSI field probability density distribution over time, as shown in Fig. 4. A scene can get more or less cloudy with time. This is a clear asset as compared to SolarSTEPS, which is limited by its underlying linear AR model to forecast stationary time series. This leads SolarSTEPS to produce fields that have approximately the same CSI distribution as the input, making it incapable of predicting scenarios where the weather situation drastically changes. This limitation is clearly visible in Fig. 1, where the cloudy region expands significantly during the forecasted period, and even more so in Fig. 4, which illustrates the distributions of CSI values for individual fields at three lead times. Also apparent is the narrowing of the distribution in the case of ConvLSTM, consistent with the overall tendency of this deterministic forecast to damp the tails of the CSI distribution in favor of mean values. This effect drastically reduces the accuracy in predicting extreme CSI values. On the other

hand, SHADECast accurately follows the observed distributional shift and outperforms the benchmark models in predicting extreme values (see Fig. 7).

4.2. Reliability and sharpness

Common measures to evaluate ensemble forecast performance include (see Methods for further details) rank histograms, PICP and PINAW, as well as the CRPS. Rank histograms shown in Fig. 5 demonstrate that SHADECast produces significantly more reliable probabilistic forecasts compared to the benchmark models. One can notice the tendency of LDCast and SolarSTEPS to generate ensembles that tend to be overconfident, underestimating the uncertainty of cloudiness evolution. LDCast overestimates, in particular, the occurrence of overcast situations (low CSI). On the other hand, SHADECast can better model the uncertainty, providing significantly more reliable ensembles. The rank histograms are computed on the test set across three different regions (see Fig. 3). In Supplementary Fig. A.2, we provide the rank histograms for the three test regions, individually. The reliability of the models does not depend on the considered location.

Model reliability can also be quantified via the PICP, shown in the second row of Fig. 5, and the PINAW, also presented in Fig. 5. The first metric calculates the average number of pixels that fall within the ensemble prediction interval, with its width (prediction sharpness) determined by the second metric. The average PICP is $\approx 70\%$ for SHADECast compared to 65% and 60% of SolarSTEPS and LDCast, respectively. The major improvement of SHADECast over LDCast can be noticed in the low-variability samples, where the model provides sharper predictions (lower PINAW) and achieves higher PICP. Instead, SolarSTEPS generally provides ensembles with lower variance, consequently achieving a lower PICP.

The CRPS serves as a compound metric, encompassing both reliability and sharpness to offer a holistic evaluation of model performance. It quantifies the distance between the ensemble and the optimal cumulative distribution for each pixel. This metric is then averaged across the entire test set (All-sky) and separately for low- and high-variability subsets, where variability is measured by the standard deviation computed on the input CSI fields. In the upper panel of Fig. 6, CRPS values are averaged across the test set for each pixel within the three test regions. Interestingly, similar spatial patterns are present in the right panel in Fig. 3, indicating a relation between standard deviation (variability) and CRPS values for the three models. In low-variability areas (Alps region in Fig. 3), the models, especially SHADECast and LDCast, exhibit a low CRPS. Conversely, the lower panel displays aggregated CRPS values averaged over all pixels, presenting the average, 25th, and 75th percentiles for each lead time.

SHADECast exhibits a 15% improvement in overall CRPS compared to SolarSTEPS and a 7% improvement over LDCast. A 120-minute

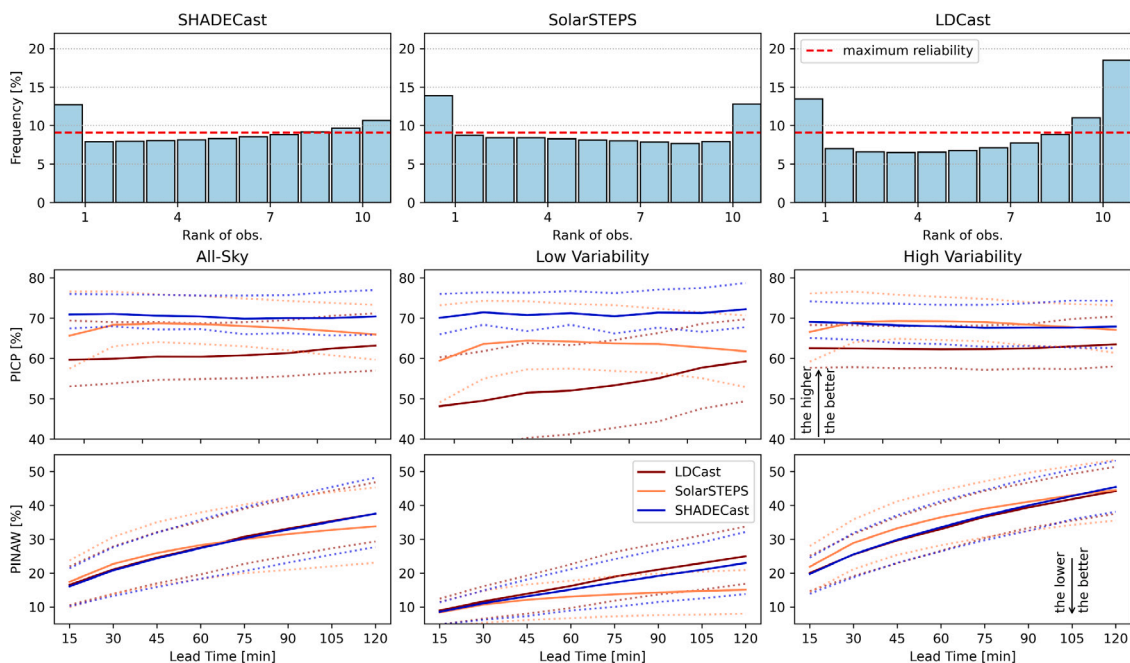


Fig. 5. Reliability evaluation. **Upper panel:** rank histogram for the 10 ensemble members (x-axis), comprising the entire test set. Our model (SHADECast) clearly provides more reliable forecasts — frequencies closer to the maximum reliability line. The high external columns on LDCast and SolarSTEPS rank histograms highlight the models overconfidence as more than 30% of CSI values fall outside the ensemble forecasts. **Lower panel:** PICP and PINAW metrics computed on the test set across different lead times. The first measures the reliability (number of ground truth pixels falling inside the prediction interval), while the second measures the sharpness of the forecast (normalized width of the prediction interval). Both metrics are measured using a confidence interval of 90%. The dotted lines represent the 25th and 75th percentile of the correspondent metric values over the entire test set.

SHADECast forecast is, then, as skillful as a 96-minute and 106-minute forecasts of SolarSTEPS and LDCast, respectively (see Fig. 6). This improvement, particularly evident in high-variability situations, suggests superior modeling of cloudiness evolution by SHADECast. The substantial enhancement over SolarSTEPS is attributed to differences in their CSI field generation mechanisms. SolarSTEPS simulates cloud evolution by random perturbations, generating CSI fields that share the spatial structure of the input satellite CSI maps but lack spatiotemporal information. In contrast, SHADECast models the spatiotemporal distribution of CSI maps, capturing information on spatial structure and temporal dynamics. The hypothesis is further supported by the smaller improvement in the low-variability subset, where cloudiness evolution is more static, resulting in similar performance between SolarSTEPS and SHADECast. This analysis underscores the importance of considering both, variability levels and the underlying dynamics of cloud evolution when assessing the efficacy of probabilistic forecasting models.

In low-variability situations, we notice a significant improvement of SHADECast over LDCast measured by an average improvement of $\sim 15\%$ in terms of CRPS. We attribute this finding to the conditioning nowcaster in SHADECast, which can better direct the forecast in low-variability situations, where a deterministic forecast contains more information with respect to a high-variability weather scenario. In these situations, the high-uncertainty regions are scarcer, so we expect the SHADECast ensemble to be closer to the nowcaster's forecast.

4.3. Forecasting extreme values

Deterministic forecasting models, such as the ConvLSTM-based model IrradianceNet [50], tend to produce blurry forecasts after a few steps. This is because such models are typically trained to minimize the mean square error of their forecast, favoring an 'average future outcome' instead of honoring, in a probabilistic sense, the inherently uncertainty evolution of weather patterns and spatial structure of the observed cloudiness field. This is due to the training process of minimizing a pixel-level loss function, such as mean squared or mean absolute

error [51]. In practice, this translates to forecasts converging towards a mean value, impeding the accurate forecast of extreme values. In our case, extreme values represent extreme overcast or complete clear-sky weather situations. We define extreme overcast and clear-sky with CSI values below 0.15 and over 0.95, respectively.

To measure the ability of the forecasting models to predict such extremes, we make use of the Fraction Skill Score (FSS) metric [48,52]. FSS evaluates the fraction of correctly predicted area for a specific threshold of interest, indicating how well a model captures the spatial distribution of an event. A higher FSS suggests better spatial agreement between predicted and observed phenomena.

In Fig. 7, FSS values are shown for SHADECast and the benchmark models for clear-sky and overcast situations for two window sizes. On average, the probabilistic models perform better than ConvLSTM due to their sharp forecasts, which do not suffer from increasing blurriness. In fact, ConvLSTM forecasts perform discretely well in low variability situations (central column) and on average in the first 15 to 30-minute. After few steps, the accuracy degrades. On the other hand, SHADECast outperforms the benchmark models on predicting clear-sky CSI values higher than 0.95 with an average 14% improvement over ConvLSTM. The improvement over the deterministic forecasts increases with the lead time, reaching 60% improvement in forecasting extreme overcast situations at 2-hours lead time.

5. Conclusions

We have introduced a novel method for probabilistically forecasting SSI satellite maps that significantly outperforms existing approaches across diverse weather situations, from low to high variability scenarios. Our model stands out as the first ensemble-based approach capable of forecasting SSI satellite maps while adapting to dynamic weather conditions without suffering from blurriness and without requiring additional information beyond the input CSI fields. We showcase our generative approach's superiority over classical deterministic models, offering realistic forecasts that outperform benchmarks in predicting extreme values.

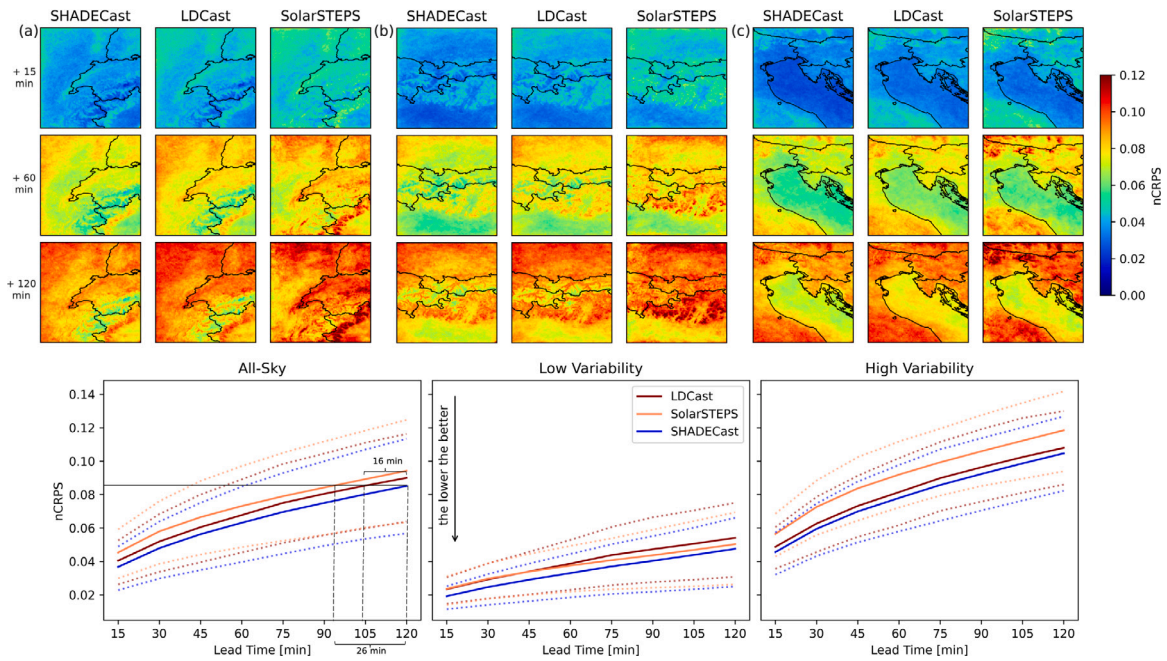


Fig. 6. Normalized Continuous Ranked Probability Score (nCRPS) plots. **Upper panel:** the nCRPS is averaged over the entire test set and reported at pixel level for the three test regions (a, b, c). The metric is shown for three lead times (+15, +60, +120 min) for SHADECast and the benchmark models. **Lower panel:** Average, 25th and 75th percentiles of nCRPS are shown for the 8 lead times and for the three models. The metric is computed for all the forecasts in the test set for every pixel and then averaged. The solid lines represent the mean value for every lead time, while the dotted lines represent the percentiles. The values shown are averaged for the entire test set (All-sky) and for two subsets, representative of low-variability and high-variability cloudiness situations.

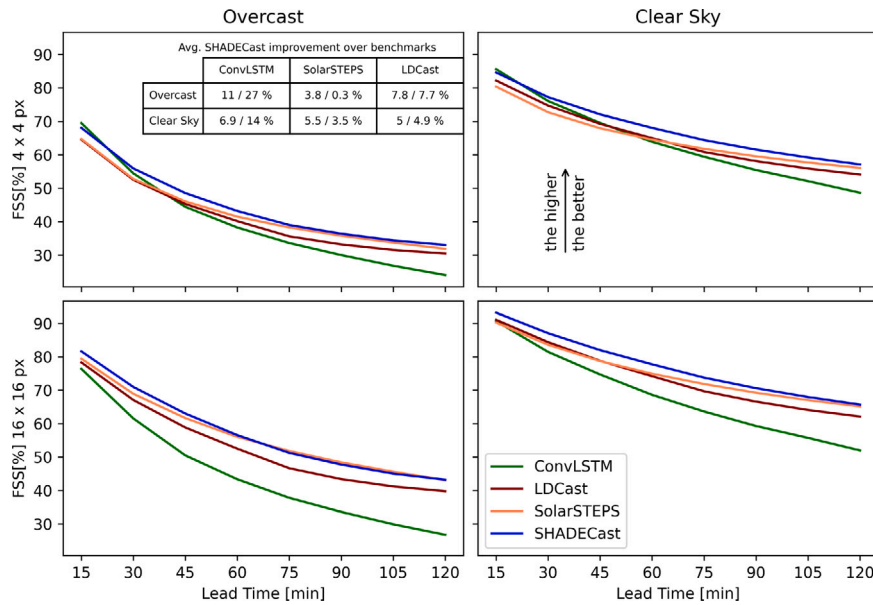


Fig. 7. Extreme values prediction accuracy. Fraction skill score (FSS) with threshold set to 0.95 (Clear Sky) and 0.15 (Overcast) relative to two window sizes: 4×4 pixels and 16×16 pixels. For SHADECast, LDCast and SolarSTEPS, the FSS is computed for every ensemble member and then averaged. The table provides the average improvement of SHADECast over the benchmark models (4×4 px and 16×16 px window sizes) for the Overcast and Clear Sky cases.

Our model exhibits superior performance, consistently outperforming benchmarks: 15% and 7% average improvement over SolarSTEPS and LDCast, respectively. This increased reliability is attributed to the incorporation of a deterministic latent nowcaster, which conditions the ensemble generation process. The modularity of our approach not only improves the performance but also permits the incorporation of alternative deterministic forecasting algorithm in our framework.

Built upon AFNO blocks and leveraging insights into cloudiness dynamics, SHADECast tackles the forecasting challenge by dividing it into a deterministic and a probabilistic components. The deterministic

nowcaster forecasts low-uncertainty large-scale dynamics, whereas the probabilistic aspect is managed by the diffusion model, responsible for simulating the stochastic evolution of cloudiness fields at smaller scales. In this way, the generated ensemble can simulate the spatial structure and dynamics of cloudiness, enabling the prediction of extreme values.

Our contribution extends beyond theoretical advances, as SHADECast provides grid and trading operators with accurate and reliable forecast ensembles. This empowers them to enhance the integration of photovoltaic energy into the grid, mitigating the volatility impact on grid resilience.

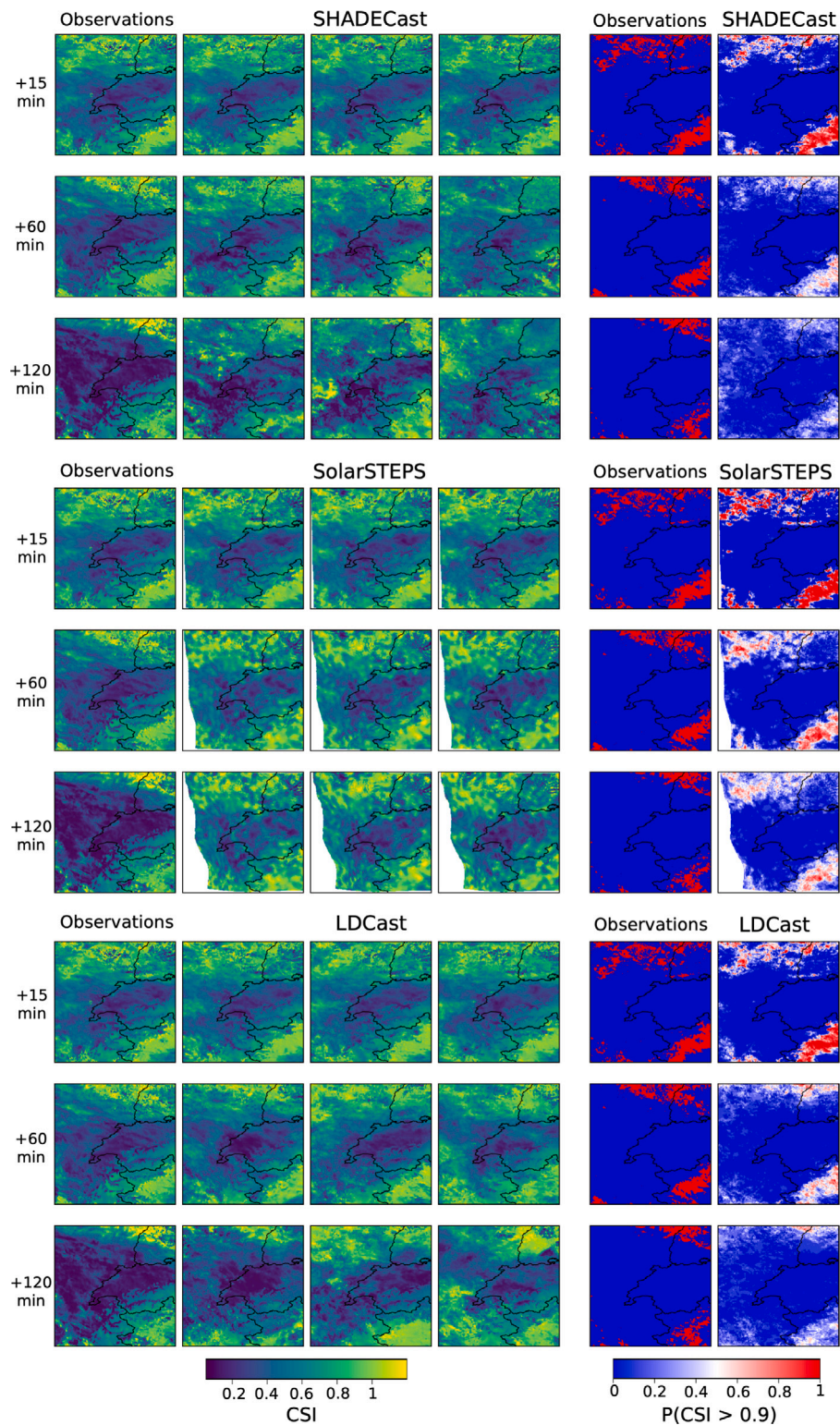


Fig. 8. Visualization of generated ensembles for test region a. The date (24 Feb. 2016) and starting time (11.45 am) are chosen to show a changing weather situation in which the cloudy surface (blue pixels) increases through the forecast. On the first column, the satellite CSI images are shown (Observations). The second, third and fourth columns show the best, average and worst ensemble members, respectively. The ensemble members are evaluated by their average RMSE over the entire forecast. The last two columns show the ground truth and forecasted probabilities of CSI exceeding 0.9 (clear-sky). The forecasted region is patch (a) (see Fig. 3).

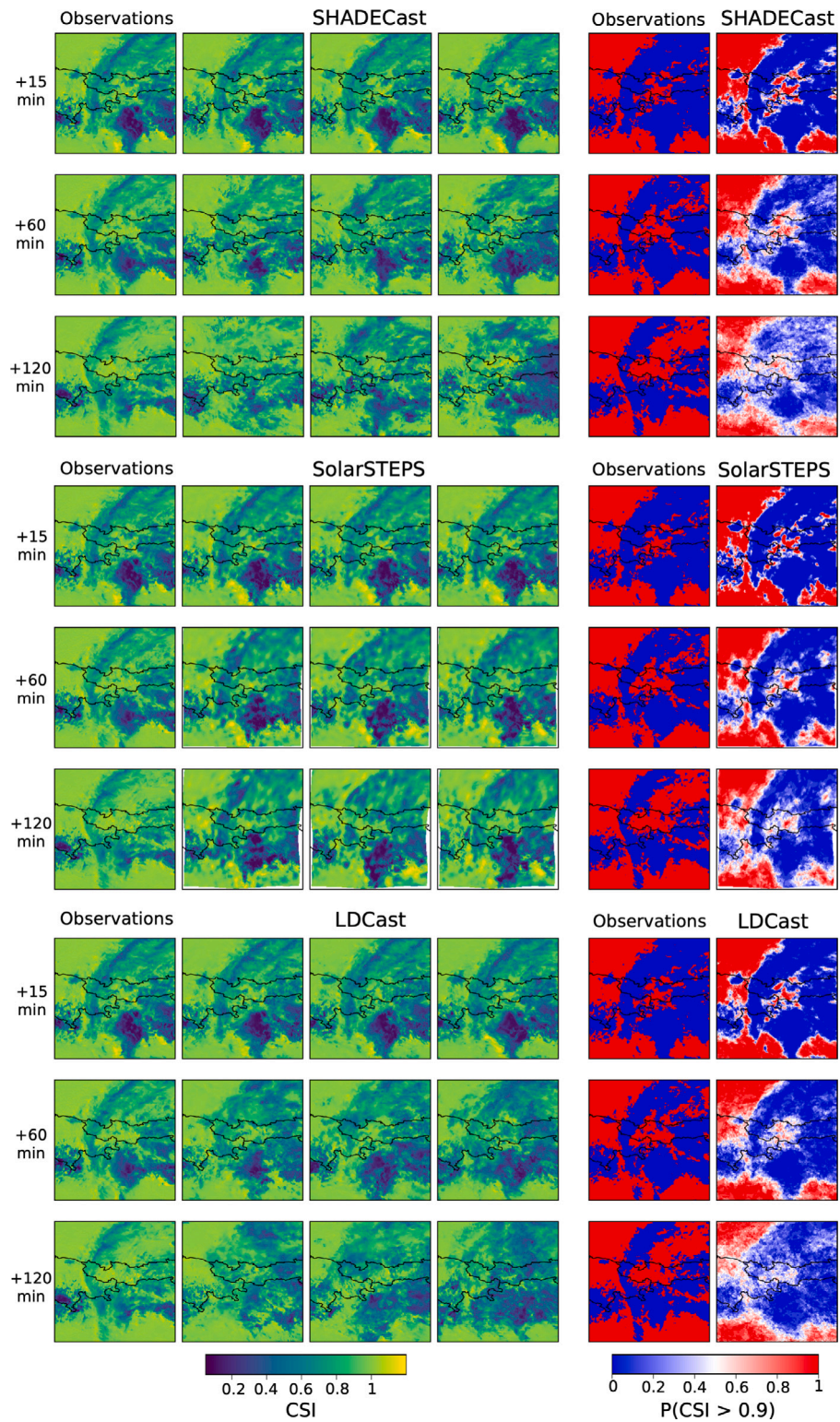


Fig. 9. Visualization of generated ensembles for test region b. The date (24 Jul. 2015) and starting time (05.30 am) are chosen to show a dissipation example of clouds on the bottom of the region. The figure is structured as Fig. 8. The forecasted region is patch (b) (see Fig. 3).

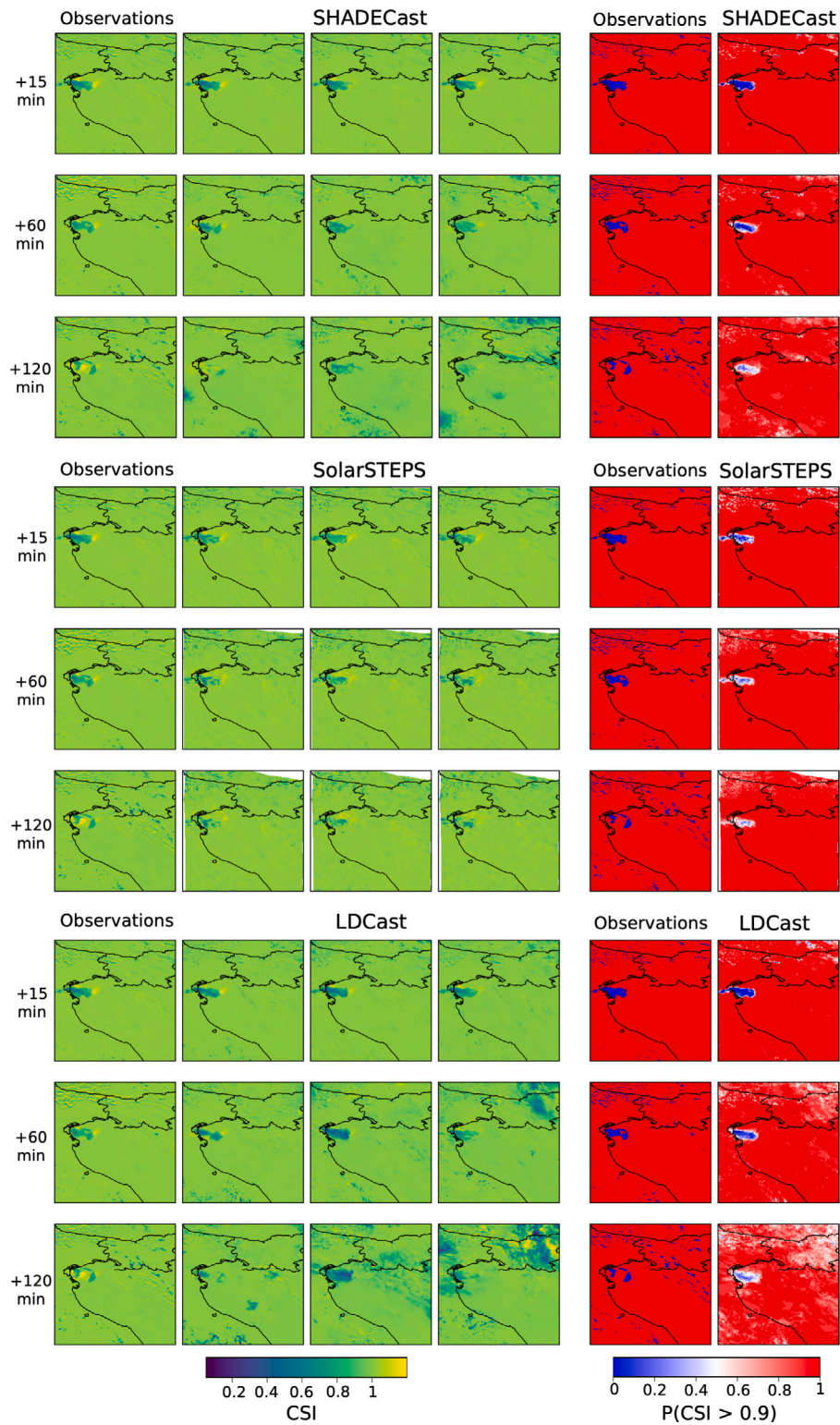


Fig. 10. Visualization of generated ensembles for test region c. The date (19 Mar. 2016) and starting time (11.15 am) are chosen to show a clear-sky and low variability weather example. The figure is structured as Fig. 8. The forecasted region is patch (c) (see Fig. 3).

One of the current limitation of the model is the lack of spatial information (longitude and latitude), whose presence could further enhance the adaptability of the model to different regions. Further enhancements of the model can be reached by adding different sources of data.

In conclusion, our model not only introduces a novel approach to SSI forecasting but also establishes a new standard in reliability and performance. By addressing the challenges of dynamic weather conditions and providing enhanced forecast ensembles, SHADECast contributes significantly to the advancement of energy meteorology and renewable energy integration.

Code availability

The code to train and test SHADECast is made available at: <https://github.com/EnergyWeatherAI/GenerativeNowcasting>.

The code for SolarSTEPS is made available at: <https://github.com/EnergyWeatherAI/SolarSTEPS>.

The original LDCast model is available at: <https://github.com/MeteoSwiss/lldcast>, whereas in the SHADECast repository there is the LDCast architecture adapted to forecast cloudiness fields.

CRedit authorship contribution statement

A. Carpentieri: Writing – original draft, Software, Methodology, Data curation, Conceptualization. **D. Folini:** Writing – review & editing, Project administration, Conceptualization. **J. Leinonen:** Software, Methodology. **A. Meyer:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Angela Meyer reports financial support was provided by Swiss National Science Foundation. Jussi Leinonen reports financial support was provided by EUMETSAT. Alberto Carpentieri reports computing resources were provided by Swiss National Supercomputing Centre. During the writing of this manuscript, J. L. changed employment from Meteoswiss to Nvidia. During the revision of the manuscript, A. C. changed employment from BFH to Nvidia. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

Angela Meyer acknowledges funding from the Swiss National Science Foundation (project ID 200654). Part of the computing was done on the Swiss National Supercomputing Centre project ID s1144. J.L. was supported by the fellowship “Seamless Artificially Intelligent Thunderstorm Nowcasts” from the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). The hosting institution of this fellowship was MeteoSwiss in Switzerland.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.apenergy.2024.124186>.

References

- [1] Yang D, Wang W, Gueymard CA, Hong T, Kleissl J, Huang J, Perez MJ, Perez R, Bright JM, Xia X, van der Meer D, Peters IM. 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, <https://www.sciencedirect.com/science/article/pii/S1364032122002593>.
- [2] Wang Y, Wang R, Tanaka K, Ciais P, Penuelas J, Balkanski Y, Sardans J, Hauglustaine D, Liu W, Xing X, Li J, Xu S, Xiong Y, Yang R, Cao J, Chen J, Wang L, Tang X, Zhang R. Accelerating the energy transition towards photovoltaic and wind in China. *Nature* 2023;619(7971):761–7, <https://doi.org/10.1038/s41586-023-06180-8>.
- [3] IEA. Renewable energy market update. International Energy Agency; 2023, <https://www.iea.org/reports/renewable-energy-market-update-june-2023>.
- [4] Nijse FJMM, Mercure J-F, Ameli N, Larosa F, Kothari S, Rickman J, Vercoulen P, Pollitt H. The momentum of the solar energy transition. *Nature Commun* 2023;14(1):6542, <https://doi.org/10.1038/s41467-023-41971-7>.
- [5] Sarmas E, Dimitropoulos N, Marinakis V, Mylona Z, Doukas H. Transfer learning strategies for solar power forecasting under data scarcity. *Sci Rep* 2022;12(1):14643, <https://doi.org/10.1038/s41598-022-18516-x>.
- [6] Smith O, Cattell O, Farcot E, O’Dea RD, Hopcraft KI. The effect of renewable energy incorporation on power grid stability and resilience. *Sci Adv* 2022;8(9):eabj6734, eprint: <https://www.science.org/doi/pdf/10.1126/sciadv.abj6734>. <https://www.science.org/doi/abs/10.1126/sciadv.abj6734>.
- [7] Brancucci Martinez-Anido C, Botor B, Florita AR, Draxl C, Lu S, Hamann HF, Hodge B-M. The value of day-ahead solar power forecasting improvement. *Sol Energy* 2016;129:192–203, <https://www.sciencedirect.com/science/article/pii/S0038092X16000736>.
- [8] Manso-Burgos Á, Ribó-Pérez D, Mateo-Barcos S, Carnero P, Gómez-Navarro T. Market value and agents benefits of enhanced short-term solar PV power generation forecasting. *Machines* 2022;10(9). <https://www.mdpi.com/2075-1702/10/9/730>.
- [9] Haupt SE, Garcia Casado M, Davidson M, Dobschinski J, Du P, Lange M, Miller T, Mohrlen C, Motley A, Pestana R, Zack J. The use of probabilistic forecasts: Applying them in theory and practice. *IEEE Power Energy Mag* 2019;17(6):46–57.
- [10] Wang P, van Westrhenen R, Meirink JF, van der Veen S, Knap W. Surface solar radiation forecasts by advecting cloud physical properties derived from meteosat second generation observations. *Sol Energy* 2019;177:47–58, <https://www.sciencedirect.com/science/article/pii/S0038092X18310521>.
- [11] Huang G, Li X, Huang C, Liu S, Ma Y, Chen H. Representativeness errors of point-scale ground-based solar radiation measurements in the validation of remote sensing products. *Remote Sens Environ* 2016;181:198–206, <https://www.sciencedirect.com/science/article/pii/S003442571630147X>.
- [12] Paletta Q, Terrén-Serrano G, Nie Y, Li B, Bieker J, Zhang W, Dubus L, Dev S, Feng C. Advances in solar forecasting: Computer vision with deep learning. *Adv Appl Energy* 2023;11:100150, <https://www.sciencedirect.com/science/article/pii/S266679242300029X>.
- [13] Hammer A, Heinemann D, Lorenz E, Lücke B. Short-term forecasting of solar radiation: a statistical approach using satellite data. *Sol Energy* 1999;67(1):139–50, <https://www.sciencedirect.com/science/article/pii/S0038092X0000384>.
- [14] Urbich I, Bendix J, Müller R. A novel approach for the short-term forecast of the effective cloud albedo. *Remote Sens* 2018;10(6). <https://www.mdpi.com/2072-4292/10/6/955>.
- [15] Ayet A, Tandeo P. Nowcasting solar irradiance using an analog method and geostationary satellite images. *Sol Energy* 2018;164:301–15, <https://www.sciencedirect.com/science/article/pii/S0038092X18301993>.
- [16] Wang P, van Westrhenen R, Meirink JF, van der Veen S, Knap W. Surface solar radiation forecasts by advecting cloud physical properties derived from meteosat second generation observations. *Sol Energy* 2019;177:47–58, <https://www.sciencedirect.com/science/article/pii/S0038092X18310521>.
- [17] Aicardi D, Musé P, Alonso-Suárez R. A comparison of satellite cloud motion vectors techniques to forecast intra-day hourly solar global horizontal irradiation. *Sol Energy* 2022;233:46–60, <https://www.sciencedirect.com/science/article/pii/S0038092X21011154>.
- [18] Carpentieri A, Folini D, Wild M, Vuilleumier L, Meyer A. Satellite-derived solar radiation for intra-hour and intra-day applications: Biases and uncertainties by season and altitude. *Sol Energy* 2023;255:274–84, <https://www.sciencedirect.com/science/article/pii/S0038092X23001810>.
- [19] Knol D, de Leeuw F, Meirink JF, Krzhizhanovskaya VV. Deep learning for solar irradiance nowcasting: A comparison of a recurrent neural network and two traditional methods. *Comput Sci – ICCS 2021 2021*;309–22.
- [20] Nielsen AH, Iosifidis A, Karstoft H. IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. *Sol Energy* 2021;228:659–69, <https://www.sciencedirect.com/science/article/pii/S0038092X21008306>.
- [21] Gallo R, Castangia M, Macii A, Macii E, Patti E, Aliberti A. Solar radiation forecasting with deep learning techniques integrating geostationary satellite images. *Eng Appl Artif Intell* 2022;116:105493, <https://www.sciencedirect.com/science/article/pii/S0952197622004833>.

- [22] Son Y, Zhang X, Yoon Y, Cho J, Choi S. LSTM-GAN based cloud movement prediction in satellite images for PV forecast. *J Ambient Intell Humaniz Comput* 14(9):12373–86, <https://doi.org/10.1007/s12652-022-04333-7>.
- [23] Wen H, Du Y, Chen X, Lim EG, Wen H, Yan K. A regional solar forecasting approach using generative adversarial networks with solar irradiance maps. *Renew Energy* 2023;216:119043, <https://www.sciencedirect.com/science/article/pii/S0960148123009576>.
- [24] Arbizu-Barrena C, Ruiz-Arias JA, Rodríguez-Benítez FJ, Pozo-Vázquez D, Tovar-Pescador J. Short-term solar radiation forecasting by advecting and diffusing MSG cloud index. *Sol Energy* 2017;155:1092–103, <https://www.sciencedirect.com/science/article/pii/S0038092X17306308>.
- [25] Hatanaka Y, Glaser Y, Galgon G, Torri G, Sadowski P. Diffusion models for high-resolution solar forecasts. 2023, [arXiv:2302.00170](https://arxiv.org/abs/2302.00170) [physics]. <http://arxiv.org/abs/2302.00170>.
- [26] Zhang X, Zhen Z, Sun Y, Zhang Y, Ren H, Ma H, Yang J, Wang F. Solar irradiance prediction interval estimation and deterministic forecasting model using ground-based sky image. In: 2022 IEEE/IAS 58th industrial and commercial power systems technical conference. I&CPS, 2022, p. 1–8.
- [27] Babaeizadeh M, Finn C, Erhan D, Campbell RH, Levine S. Stochastic variational video prediction. In: International conference on learning representations. 2018, <https://openreview.net/forum?id=rk49Mg-CW>.
- [28] Ravuri S, Lenc K, Willson M, Kangin D, Lam R, Mirowski P, Fitzsimons M, Athanassiadou M, Kashem S, Madge S, Prudden R, Mandhane A, Clark A, Brock A, Simonyan K, Hadsell R, Robinson N, Clancy E, Arribas A, Mohamed S. Skilful precipitation nowcasting using deep generative models of radar. *Nature* 2021;597(7878):672–7, <https://doi.org/10.1038/s41586-021-03854-z>.
- [29] Yang R, Srivastava P, Mandt S. Diffusion probabilistic modeling for video generation. *Entropy* 2023;25(10). <https://www.mdpi.com/1099-4300/25/10/1469>.
- [30] Jascha S-D, Eric W, Niru M, Surya G. Deep unsupervised learning using nonequilibrium thermodynamics. *Proc Mach Learn Res* 2015;37:2256–65, <https://proceedings.mlr.press/v37/sohl-dickstein15.html>.
- [31] Jonathan H, Ajay J, Pieter A. Denoising diffusion probabilistic models. In: NIPS'20. Red Hook, NY, USA: Curran Associates Inc.; 2020.
- [32] Dhariwal P, Nichol AQ. Diffusion models beat GANs on image synthesis. *Adv Neural Inf Process Syst* 2021. <https://openreview.net/forum?id=AAWuCVzaVt>.
- [33] Leinonen J, Hamann U, Nerini D, Germann U, Franch G. Latent diffusion models for generative precipitation nowcasting with accurate uncertainty quantification. 2023, [arXiv:2304.12891](https://arxiv.org/abs/2304.12891) [physics.a0-ph].
- [34] 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, <https://www.sciencedirect.com/science/article/pii/S030626192301139X>.
- [35] Guibas J, Mardani M, Li Z, Tao A, Anandkumar A, Catanzaro B. Efficient token mixing for transformers via adaptive Fourier neural operators. In: International conference on learning representations. 2022, <https://openreview.net/forum?id=EXHG-A3jIM>.
- [36] Pathak J, Subramanian S, Harrington P, Raja S, Chattopadhyay A, Mardani M, Kurth T, Hall D, Li Z, Azizzadenesheli K, Hassanzadeh P, Kashinath K, Anandkumar A. FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. 2022, [arXiv:2202.11214](https://arxiv.org/abs/2202.11214) [physics.a0-ph].
- [37] Robin R, Andreas B, Dominik L, Patrick E, Björn O. High-resolution image synthesis with latent diffusion models. 2022, p. 10684–95.
- [38] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Lu, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- [39] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention – MICCAI 2015. Cham: Springer International Publishing; 2015, p. 234–41.
- [40] Liu L, Ren Y, Lin Z, Zhao Z. Pseudo numerical methods for diffusion models on manifolds. In: International conference on learning representations. 2022, <https://openreview.net/forum?id=PIKWVd2yBkY>.
- [41] Kingma DP, Welling M. Auto-encoding variational Bayes. 2022, [arXiv:1312.6114](https://arxiv.org/abs/1312.6114) [stat.ML].
- [42] Schmetz J, Pili P, Tjemkes S, Just D, Kerkmann J, Rota S, Ratier A. An introduction to MeteoSat second generation (MSG). *Bull Am Meteorol Soc* 2002;83(7):977–92, https://journals.ametsoc.org/view/journals/bams/83/7/1520-0477_2002_083_0977_aitmsg_2_3_co_2.xml.
- [43] Castelli M, Stöckli R, Zardi D, Tetzlaff A, Wagner J, 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, <https://www.sciencedirect.com/science/article/pii/S0034425714002673>.
- [44] Nerini D, Besic N, Sideris I, Germann U, Foresti L. A non-stationary stochastic ensemble generator for radar rainfall fields based on the short-space Fourier transform. *Hydrol Earth Syst Sci* 2017;21(6):2777–97, <https://hess.copernicus.org/articles/21/2777/2017/>.
- [45] Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. *R Stat Soc* 2007;69(2):243–68.
- [46] Hamill TM. Interpretation of rank histograms for verifying ensemble forecasts. *Mon Weather Rev* 2001;129(3):550–60, https://journals.ametsoc.org/view/journals/mwre/129/3/1520-0493_2001_129_0550_iorhfv_2.0.co_2.xml.
- [47] Broecker J. Probability forecasts. John Wiley & Sons, Ltd; 2011, p. 119–39, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119960003.ch7>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119960003.ch7>.
- [48] Skok G, Roberts N. Analysis of fractions skill score properties for random precipitation fields and ECMWF forecasts. *Q J R Meteorol Soc* 2016;142(700):2599–610.
- [49] Mittermaier MP. A “meta” analysis of the fractions skill score: The limiting case and implications for aggregation. *Mon Weather Rev* 2021;149(10):3491–504, <https://journals.ametsoc.org/view/journals/mwre/149/10/MWR-D-18-0106.1.xml>.
- [50] Nielsen AH, Iosifidis A, Karstoft H. IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. *Sol Energy* 2021;228:659–69, <https://www.sciencedirect.com/science/article/pii/S0038092X21008306>.
- [51] Babaeizadeh M, Finn C, Erhan D, Campbell RH, Levine S. Stochastic variational video prediction. In: International conference on learning representations. 2018, <https://openreview.net/forum?id=rk49Mg-CW>.
- [52] Leinonen J, Hamann U, Nerini D, Germann U, Franch G. Latent diffusion models for generative precipitation nowcasting with accurate uncertainty quantification. 2023, [arXiv:2304.12891](https://arxiv.org/abs/2304.12891) [physics.a0-ph].