This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2026.3714841

Weakly Informative Priors Flow Matching for Probabilistic Regional Solar Irradiance Forecasting

Tao Jing, Shanlin Chen, Jiannong Fang, Fernando Porté-Agel, Ming Chun Lam and Mengying Li

***Abstract*—Probabilistic forecasting of regional surface solar irradiance provides critical technical support for the reliable operation and grid integration of large-scale and distributed photovoltaic power systems. However, stochastic fluctuations driven by cloud dynamics substantially complicate the characterization of solar irradiance uncertainty at the regional scale, thereby posing significant challenges to probabilistic regional forecasting. To address this, we propose WIP-Flow, a novel Weakly Informative Priors Flow Matching framework that leverages satellite-derived historical irradiance fields and cloud optical thickness to capture complex spatio-temporal dynamics. Central to WIP-Flow is an unsupervised autoregressive integrated correction algorithm, which combines optical flow with multi-scale decomposition to generate a physically informed initial distribution based on cloud advection. Furthermore, a Flow Matching model with a stochastic interpolation path is introduced to correct boundary distortions and characterize stochastic cloud formation, diffusion, and dissipation processes. Experimental validation on a satellite-derived regional solar irradiance dataset demonstrates that WIP-Flow generates physically coherent probabilistic forecasts and achieves competitive or superior performance across marginal and multivariate evaluation metrics. Notably, WIP-Flow achieves a root mean square error of 104.3 W/m^2 and a continuous ranked probability score of 47.81 W/m^2 in 4-hour-ahead forecasts, indicating its potential value for uncertainty-aware solar power grid integration.**

***Index Terms*—Regional Solar Irradiance Forecasting, Probabilistic Forecasting, Weakly Informative Priors, Flow Matching**

I. Introduction

As the global energy transition accelerates and renewable energy capacity continues to expand, solar power generation is becoming increasingly critical in modern power systems [1], [2]. Surface Solar Irradiance (SSI), the primary driver of photovoltaic (PV) power output, exhibits pronounced temporal and spatial variability, which is largely governed by cloud dynamics. This variability introduces significant challenges to the safe and reliable operation of power systems, efficient scheduling, and electricity market operation [3], [4]. For grid integration of large-scale PV, traditional deterministic forecasting methods often fail to effectively account for SSI uncertainty and the risks associated with extreme weather conditions [5]. Probabilistic forecasting has emerged as a promising alternative, offering insights into the distribution, range, and potential scenarios of future SSI. This approach provides crucial decision support for power generation planning, grid optimization, load management, energy storage strategies, and risk-aware dispatch [6].

Probabilistic forecasting of SSI at the regional scale remains an underexplored yet crucial area of research. While numerical weather prediction (NWP) remains the dominant approach, its ability to quantify uncertainties comes at the cost of high computational demands associated with ensemble forecasts, which limits its scalability and real-time applicability [7]. To address these limitations, researchers have explored alternative probabilistic forecasting approaches that extend beyond NWP models. One such approach involves probabilistic graphical models that integrate ground-based irradiance measurements from multiple monitoring stations with advanced spatial interpolation techniques, such as spatio-temporal Kriging, to construct regional probability distribution maps of SSI [8]. Another innovative direction employs graph neural networks, leveraging graph-based convolutional structures to explicitly model spatial dependencies among distributed sites while capturing complex nonlinear relationships between irradiance and meteorological factors [9]. However, the effectiveness of these methods is fundamentally constrained by the spatial density of ground-based monitoring stations. In areas where station distribution is sparse, the reliability of the constructed regional mapping deteriorates, significantly limiting the applicability of these approaches to regions without well-developed monitoring infrastructures.

The increasing adoption of validated satellite-derived irradiance data has facilitated the widespread application of computer vision techniques in probabilistic SSI forecasting. From the perspective of model construction paradigms, these approaches can be broadly categorized into two technical pathways: extending deterministic prediction models to enable the generation of probabilistic outputs; and directly utilizing generative models to learn and simulate the uncertainty in future irradiance distributions in an end-to-end manner. Regarding the former, Nielsen et al. [10] modified a deterministic ConvLSTM to output per-pixel SSI probabilities over predefined bins, which greatly inflates output dimensionality.

This work is substantially supported by the Innovation and Technology Commission (Project No. ITS/188/23) of the Hong Kong Special Administrative Region, China. *Corresponding authors: Mengying Li*

Corresponding author: Mengying Li

T. Jing is with both the Department of Mechanical Engineering, The Hong Kong Polytechnic University (PolyU), Kowloon, Hong Kong SAR, and Wind Engineering and Renewable Energy Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland. Email: taojoty.jing@connect.polyu.hk.

M. Li is with Department of Mechanical Engineering, The Hong Kong Polytechnic University (PolyU), Kowloon, Hong Kong SAR. Email: mengying.li@polyu.edu.hk.

S. Chen, J. Fang, and F. Porté-Agel are with Wind Engineering and Renewable Energy Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland. Email: shanlin.chen@epfl.ch, jiannong.fang@epfl.ch, fernando.porte-agel@epfl.ch.

M. C. Lam is with Radiation Monitoring and Assessment Branch, Hong Kong Observatory, Hong Kong SAR. Email: mclam@hko.gov.hk.

Carpentieri et al. [11] performed deterministic SSI forecasting by combining semi-Lagrangian extrapolation, cascade decomposition, and autoregressive models. They then introduced locally spatially correlated noise to model uncertainties and generate probabilistic predictions. For the latter, Wen et al. [12] proposed a regional-scale probabilistic prediction method for SSI by combining satellite-derived irradiance data with a multiscale Generative Adversarial Networks (GAN) framework, effectively addressing the shortcomings of traditional single-point prediction methods. Similarly, Liu et al. [13] developed a spatio-temporal deep learning model leveraging variational inference to capture the complex spatio-temporal variability of SSI. While these approaches demonstrate significant potential, they are hindered by several inherent challenges, including unstable training processes, inadequate temporal modeling capabilities, and subpar performance in uncertainty quantification.

Recently, generative models such as the denoising diffusion probabilistic model (DDPM) and the Flow Matching model have offered novel perspectives on quantifying uncertainty [14], [15]. However, these models, originally designed for image generation tasks, still face fundamental challenges when directly applied to SSI forecasting, which exhibits strong periodic characteristics. Carpentier et al. [16] proposed guiding the diffusion process by incorporating deterministic forecast conditions. However, this method inherently relies on generating future trajectories from random Gaussian noise, which not only limits its ability to capture complex cloud-driven dynamics in the atmosphere but may also reduce the accuracy of clear-sky forecasts by introducing unnecessary random perturbations into the predictions. Jing et al. [17] proposed DiffSolar, which decouples the deterministic prediction of regional SSI from the stochastic residuals induced by cloud dynamics. DiffSolar employs a smart persistent forecasting model (SPFM) to compute the deterministic component, allowing the generative model to focus on predicting probabilistic residuals, which significantly enhance prediction accuracy and interpretability. Nevertheless, the complexity and lengthy training cycles of DDPM still constrain its further application and development in real-time SSI forecasting.

Compared with DDPM, Flow Matching learns a velocity field that transports a source distribution to the target distribution. By leveraging ordinary differential equation (ODE) solvers, it enables fast and accurate generation tasks. Unlike DDPM's noise reduction process, which requires a large number of iterative steps to progressively optimize samples, Flow Matching directly models transformation dynamics, significantly reducing computational overhead while maintaining high-quality results [15]. However, applying Flow Matching techniques to regional solar irradiance forecasting faces two major challenges. Firstly, simple initial distributions (such as Gaussian distributions) often struggle to accurately capture the spatiotemporal dynamics of cloud evolution through Flow Matching. This limitation stems from such distributions' inability to encode the complex spatiotemporal dependencies inherent in cloud formation and motion. To address this issue, designing more informative initial distributions that incorporate prior knowledge of cloud behavior can significantly enhance the model's ability to learn cloud dynamics. Secondly, a meticulously chosen initial distribution might inadvertently hinder Flow Matching from effectively learning the mapping between distributions. An overly complex or restrictive initial distribution may introduce bias or limit the flexibility of the velocity field, making it difficult for the model to accurately simulate the evolution of cloud dynamics.

The primary motivation of this study is to explore how to adapt and optimize the Flow Matching methodology for probabilistic regional SSI forecasting, offering an innovative pathway to enhance the reliability of solar forecasting and its integration into power systems. In this work, we propose a weakly informative priors Flow Matching framework for probabilistic regional SSI forecasting, denoted as WIP-Flow. By leveraging the temporal dynamics of historical SSI fields, WIP-Flow not only aligns with the physical essence of cloud evolution and its impact on SSI but also enhances the reliability of the initial distribution in Flow Matching. The proposed method provides a physically guided generative forecasting strategy for improving regional SSI probabilistic prediction while extending Flow Matching to weather-related spatio-temporal forecasting. The main contributions are as follows:

- We develop a weak-prior-guided Flow Matching framework for probabilistic regional SSI forecasting. By leveraging historical SSI fields and cloud optical thickness (COT), the proposed framework models cloud-driven spatio-temporal uncertainty while avoiding direct generation from an uninformative Gaussian prior.
- We propose an unsupervised autoregressive integrated correction (AIC) algorithm to construct a physically informed source distribution for Flow Matching. The algorithm combines Lucas–Kanade optical flow, semi-Lagrangian advection, multi-scale cascade decomposition, and autoregressive integrated correction to capture the dominant cloud-advection structure in historical SSI fields.
- To rectify AIC-induced boundary distortions and represent stochastic cloud formation, diffusion, and dissipation, we propose a stochastic interpolation method based on Stochastic Differential Equations (SDEs) to model the probabilistic mapping path between the initial distribution and the target distribution. Furthermore, we provide a distributional-transport interpretation to show that the AIC-informed source distribution reduces the effective source–target discrepancy, allowing Flow Matching to focus on stochastic residual cloud evolution rather than learning the entire future irradiance field from an uninformative prior.

The remainder of the paper is organized as follows. Section II formalizes the regional SSI prediction task. Section III presents the proposed weakly informative priors Flow Matching framework. In Section IV, we describe the experimental setup. Experimental results and discussion are presented in Section V. Finally, Section VI provides the conclusion.

II. Task Definition

A prevalent strategy in SSI forecasting involves selecting the Clear-Sky Index (CSI) as the primary predictive variable.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2026.3714841

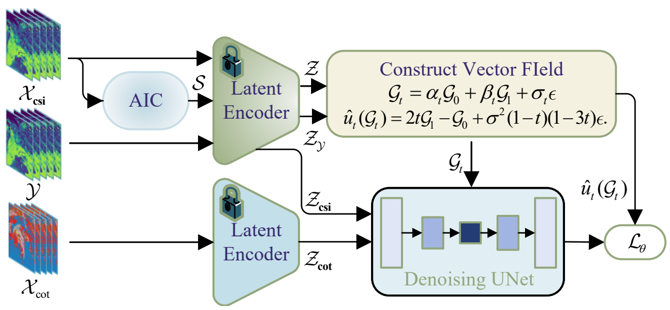

Fig. 1: The framework of WIP-Flow.

The fundamental motivation is to decouple the stochastic fluctuations of irradiance from the strong, deterministic influence of seasonal and diurnal solar cycles. This approach effectively isolates the transient transmittance variations introduced by atmospheric conditions. Consequently, the CSI serves as a dimensionless metric of cloud-induced modulation, directly capturing variations in optical thickness, cloud cover, phase state, and geometric distribution [18], [19]. The CSI is defined as:

$$\mathrm{CSI}^i = \frac{\mathrm{SSI}^i}{\mathrm{SSI}^i_{\mathrm{cs}}}, \tag{1}$$

where SSI^i_{cs} refers to the SSI value under clear sky conditions, which can be derived from a physical or empirical clear sky model. McClear [20] is used to calculate clear sky irradiance in this work. Regional-scale CSI forecasting is formulated as a spatio-temporal forecasting problem. The observed CSI and COT are expressed as two spatio-temporal sequences $\mathcal{X}_{\mathrm{csi}} = [\mathbf{x}^i_{\mathrm{csi}}]^0_{i=-T_h} \in \mathbb{R}^{1\times T_h\times H\times W}$ and $\mathcal{X}_{\mathrm{cot}} = [\mathbf{x}^i_{\mathrm{cot}}]^0_{i=-T_h} \in \mathbb{R}^{1\times T_h\times H\times W}$, where T_h represents input time steps, and H and W denote the spatial resolution. The goal of probabilistic forecasting is to model the conditional probability distribution $p(\mathcal{Y}|\mathcal{X}_{\mathrm{csi}}, \mathcal{X}_{\mathrm{cot}})$ of the T_f-step-ahead target CSI sequence $\mathcal{Y} = [\mathbf{y}^i]^{T_f}_{i=1} \in \mathbb{R}^{1\times T_f\times H\times W}$, given observations $\mathcal{X}_{\mathrm{csi}}$ and $\mathcal{X}_{\mathrm{cot}}$. A superscript i is employed to indicate the i-th temporal frame.

III. Methodology

In this work, short-term (0–4 hours ahead) SSI forecasting is considered, which is greatly influenced by daytime cloud dynamics. Fundamentally, cloud dynamics can be partitioned into four dominant processes: cloud advection, formation, diffusion, and dissipation. Accordingly, WIP-Flow is designed as a physically guided probabilistic framework: AIC first captures the dominant deterministic advection component, while Flow Matching corrects the remaining stochastic cloud-evolution residuals. To characterize the advection component, AIC is introduced to initialize per-pixel optical-flow vectors and perform semi-Lagrangian extrapolation, yielding a predicted advection field. In addition to advective extrapolation, an autoregressive integrated correction is used to capture local temporal persistence and cross-scale evolution.

However, AIC tends to introduce inherent boundary distortion issues during the advection extrapolation process due to invalid back-traced departure points near the domain boundary [21]. More importantly, as a deterministic prior, AIC cannot fully represent cloud formation, diffusion, dissipation, or forecast uncertainty. To address the limitations of AIC, a flow-matching framework is introduced to enhance the modeling of cloud formation, diffusion, and dissipation. Specifically, stochastic Gaussian boundary completion is applied only to the distorted boundary regions in the AIC output before latent encoding. This operation provides stochastic degrees of freedom for uncertain boundary inflow, while preserving the physically informed AIC advection structure in the valid region. The boundary-completed AIC field then serves as the weakly informative source distribution for the Flow Matching phase. A VAE further compresses the high-dimensional CSI and COT fields into a tractable latent space. Then, we formulate an SDE-based interpolation path that transports the initial distribution to the target distribution. The SDE path not only mitigates boundary artifacts but also captures cloud formation, diffusion, and dissipation processes, which are inadequately represented by AIC. Thus, the performance gain arises from the coupling of a weakly informative AIC prior with stochastic Flow Matching correction, rather than from either component alone. The overall framework of WIP-Flow is presented in Fig. 1.

A. Autoregressive Integrated Correction

Starting from a sequence of historical CSI frames $\mathcal{X}_{\mathrm{csi}}$, we estimate inter-frame motion using the Lucas–Kanade method [21]. Specifically, dense optical-flow vectors are computed for each pair of consecutive frames $\{\mathbf{x}^{i-1}_{\mathrm{csi}}, \mathbf{x}^i_{\mathrm{csi}}\}$ within the history window. To obtain a stable representation of the dominant transport, these flow fields are temporally averaged to produce a mean optical-flow field (CMV):

$$\mathbf{v}_i = \mathrm{LK}\{\mathbf{x}^{i-1}_{\mathrm{csi}}, \mathbf{x}^i_{\mathrm{csi}}\}, \tag{2}$$

$$\hat{\mathbf{v}} = \frac{1}{T_h}\sum_{i=1}^{T_h} \mathbf{v}_i, \tag{3}$$

where LK denotes the Lucas–Kanade algorithm described in [21], and $\hat{\mathbf{v}}$ denotes the mean CMV. Using the CMV, we then extrapolate the last historical CSI frame with the backward interpolation-once semi-Lagrangian scheme described in [22]. For each forecast time step, pixel locations are traced backward along the mean-flow trajectories to their departure points in the previous frame, and CSI values are recovered via bilinear interpolation. For a given pixel location $\mathbf{m}^0$ in $\mathbf{x}^0_{\mathrm{csi}}$, the departure point is computed as:

$$\mathbf{m}^{-1} = \mathbf{m}^0 - \hat{\mathbf{v}}\Delta t \tag{4}$$

The CSI value is then obtained by bilinear interpolation at the departure point:

$$\mathcal{S}_{csi} = \begin{bmatrix} \mathbf{x}^0_{\mathrm{csi}} \\ \mathbf{x}^1_{\mathrm{csi}} \\ \vdots \\ \mathbf{x}^{T_f}_{\mathrm{csi}} \end{bmatrix} = \begin{bmatrix} \mathcal{I}\big(\mathbf{x}^{-1}_{\mathrm{csi}},\, \mathbf{m}^0 - \hat{\mathbf{v}}\,\Delta t\big) \\ \mathcal{I}\big(\mathbf{x}^{0}_{\mathrm{csi}},\, \mathbf{m}^1 - \hat{\mathbf{v}}\,\Delta t\big) \\ \vdots \\ \mathcal{I}\big(\mathbf{x}^{T_f-1}_{\mathrm{csi}},\, \mathbf{m}^{T_f} - \hat{\mathbf{v}}\,\Delta t\big) \end{bmatrix} \tag{5}$$

where $\mathcal{S}_{\text{csi}} \in \mathbb{R}^{1\times T_f \times H \times W}$ represents the results of the semi-Lagrangian extrapolation, $\mathcal{I}$ is the spatial bilinear interpolation operator, and Δt denotes the temporal resolution. This procedure yields a forward-propagated CSI sequence while suppressing frame-to-frame noise and preserving coherent motion patterns.

In previous studies on precipitation forecasting, cascade decomposition has commonly been employed to model the evolutionary characteristics of advection processes across different spatial scales [23], [24]. Considering the scale interactions driven by both cloud dynamics and micro-physical processes, the spatiotemporal evolution of SSI also exhibits significant scale heterogeneity and non-uniformity. Therefore, we employ a method based on the Discrete Fourier Transform (DFT) to perform cascade decomposition on the CSI advection field obtained through semi-Lagrangian extrapolation. The DFT and a set of bandpass filters are applied to the CSI advection field, and each scale is obtained as:

$$\mathcal{S}_j = \mathcal{F}^{-1}(w_j\, \mathcal{F}(\mathcal{S}_{\text{csi}})), \tag{6}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the DFT operator and the inverse DFT operator, respectively; w_j is the weight of the Fourier wavenumbers $\mathbf{k} = (k_x, k_y)$ for different scale levels j:

$$w_{|\mathbf{k}|,j} = \mathbf{h}_{|\mathbf{k}|,j}, \mathbf{g}_{|\mathbf{k}|,j}, \tag{7}$$

where $\sum_{j=1}^{n} \mathbf{h}_{|\mathbf{k}|,j} = 1$ are the normalization constants. The $\mathbf{g}_{|\mathbf{k}|,j}$ are defined as follows:

$$\mathbf{g}_{|\mathbf{k}|,j} = \begin{cases} e^{\left(-\frac{(\log_a |\mathbf{k}| - b_j)^2}{2c_j^2}\right)}, & 0 < |\mathbf{k}| \leq \frac{H}{4}, \\ 0, & |\mathbf{k}| = 0 \\ \begin{cases} 0, & j = 1, 2, \ldots, n-1 \\ 1, & j = n \end{cases}, & |\mathbf{k}| > \left\lfloor \frac{H}{4} \right\rfloor \end{cases} \tag{8}$$

where

$$\begin{aligned} a &= \left(\frac{H}{8}\right)^{\frac{1}{n-2}} \\ b_j &= \begin{cases} 0, & j = 1 \\ \log_a 2 + j - 2, & j = 2, 3, \ldots, n \end{cases} \\ c_j &= \begin{cases} 0.15a, & j = 0 \\ 0.1a, & j = 1, 2, \ldots, n \end{cases} \end{aligned} \tag{9}$$

where the number of scale levels $n = 6$. All hyperparameter settings are referenced from [24].

For the n cascade levels, we employ an autoregressive integrated algorithm to predict the growth and decay trends of cloud motion advection. The autoregressive iterative formula applied at each scale in the Lagrange coordinates can be formulated as:

$$\mathbf{A}_{p,j,x,y}(1 - \mathbf{B})^d \mathcal{S}_j^t(x, y) = \phi_0, \tag{10}$$

where $\mathbf{A}_{p,j,x,y} = 1 - \phi_{1,j,x,y}\mathbf{B} - \cdots - \phi_{1,j,x,y}\mathbf{B}^p$, $\mathbf{B}^p(\mathcal{S}_j^t(x, y)) = \mathcal{S}_j^{t-p}(x, y)$, x, and y are the spatial coordinates, d denotes the order of differencing, and p is the order of autoregression. Based on computational complexity, we set $p = 2$, $d = 1$, and $\phi_0 = 0$. During each iteration, the model parameters $\phi_{1,j,x,y} \cdots \phi_{p,j,x,y}$ in Eq. (10) are estimated using the Yule–Walker equations [25] and by computing the empirical lag-l autocorrelation functions:

$$\gamma_{i,x,y}(l) = \frac{\text{cov}_{x,y}\left(\Delta\mathcal{S}_j^{t_0}, \Delta\mathcal{S}_j^{t_0-l}\right)}{\sqrt{\text{std}_{x,y}\left(\Delta\mathcal{S}_j^{t_0}\right)\text{std}_{x,y}\left(\Delta\mathcal{S}_j^{t_0-l}\right)}}, \tag{11}$$

$$\text{cov}_{x,y}\left(\Delta\mathcal{S}^{t_0}, \Delta\mathcal{S}^{t_0-l}\right) = \frac{1}{s_{x,y}} \sum_{\Omega} w_{x,y} \Delta\mathcal{S}_{\Omega}^{t_0} \Delta\mathcal{S}_{\Omega}^{t_0-l}, \tag{12}$$

$$\text{std}_{x,y}\left(\Delta\mathcal{S}^{t_0}\right) = \sqrt{\frac{1}{s_{x,y}} \sum_{\Omega} w_{x,y} (\Delta\mathcal{S}_{\Omega}^{t_0})^2} \tag{13}$$

where

$$s_{x,y} = \sum_{\Omega} w_{x,y}, \tag{14}$$

$$\Delta\mathcal{S}_j^t = \mathcal{S}_j^t - \mathcal{S}_j^{t-1}, \tag{15}$$

Ω denotes the domain, and $w_{x,y}$ denotes the bivariate Gaussian function. The autoregressive integrated algorithm requires $p + 2$ frames of input CSI advection fields. After parameter estimation is completed, the resulting parameters are projected onto the nearest point within the defined stationary region to ensure that they satisfy the stationarity requirement:

$$\begin{aligned} &|\phi_{1,j,x,y}| < 2, \\ &\phi_{2,j,x,y} < 1 - \phi_{1,j,x,y}, \\ &|\phi_{2,j,x,y}| < 2. \end{aligned} \tag{16}$$

After the iterations are completed, each cascade level is reorganized to generate the pre-completion corrected CSI field:

$$\bar{\mathcal{S}} = \sum_{j=1}^{n} \mathcal{S}_j. \tag{17}$$

Boundary distortions may arise after semi-Lagrangian extrapolation when back-traced departure points fall outside the valid spatial domain. To construct a valid source state for the subsequent Flow Matching process, stochastic Gaussian boundary completion is applied only to these ill-defined boundary pixels. Let $\mathcal{M}$ and $\mathcal{B}$ denote the valid AIC-predicted region and the boundary-distorted region, respectively, with $\mathcal{M} \cup \mathcal{B} = \Omega$ and $\mathcal{M} \cap \mathcal{B} = \emptyset$. Denoting the pre-completion AIC field by $\bar{\mathcal{S}}$, the final AIC output $\hat{\mathcal{S}}$ is defined as

$$\hat{\mathcal{S}}(\mathbf{p}) = \mathbf{1}_{\mathcal{M}}(\mathbf{p})\bar{\mathcal{S}}(\mathbf{p}) + \mathbf{1}_{\mathcal{B}}(\mathbf{p})\epsilon_b(\mathbf{p}), \qquad \epsilon_b(\mathbf{p}) \sim \mathcal{N}(0, \sigma_b^2), \tag{18}$$

where $\mathbf{p}$ denotes a spatial grid point, ϵ_b is used only for boundary-distorted pixels, and $\mathbf{1}_{\mathcal{M}}$ and $\mathbf{1}_{\mathcal{B}}$ are the corresponding indicator functions. This construction preserves the valid AIC-predicted region and introduces stochastic degrees of freedom only for uncertain boundary inflow. The Gaussian boundary component is not regarded as the final boundary prediction; spatial coherence is recovered through the learned conditional Flow Matching transport from the AIC-informed source distribution to the target distribution. The details of AIC are shown in Algorithm 1.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2026.3714841

Algorithm 1: Autoregressive Integrated Correction

Input:
Dataset: $\{\mathcal{X}_{\text{csi}}\}$;
Hyperparameters: p and n;
Boundary noise: $\epsilon_b \sim \mathcal{N}(0, \sigma_b^2)$.
Output:
Pre-completion AIC field $\bar{\mathcal{S}}$ and boundary-completed AIC field $\hat{\mathcal{S}}$.

Calculate optical flow based on Eqs. (2) and (3);
Semi-Lagrangian extrapolation based on Eq. (5);
Cascade decomposition with n levels based on Eqs. (6)–(8);
for $j = 1, \ldots, n$ **do**
    for $i = 1, \ldots, p$ **do**
        Iterate the autoregressive integrated algorithm based on Eq. (10);
        Estimate the model parameters based on Eqs. (11)–(13);
Recompose the cascade levels to obtain the pre-completion AIC field $\bar{\mathcal{S}}$ based on Eq. (17);
Apply stochastic Gaussian boundary completion only to the boundary-distorted region to obtain $\hat{\mathcal{S}}$;
return $\hat{\mathcal{S}}$

B. Weakly Informative Priors Flow Matching

Although the AIC algorithm effectively models cloud advection, it remains limited in representing uncertainty associated with cloud formation and dissipation, which in turn limits the accuracy of regional SSI predictions. To mitigate this issue, an SDE-based interpolation path is introduced to transport the AIC-informed source distribution toward the target distribution of $\mathcal{Y}$. Here, interpolation refers to a latent probability path in Flow Matching, rather than spatial interpolation or image super-resolution, since the source and target fields share the same satellite-derived grid. To accelerate training and inference, a pre-trained variational autoencoder (VAE) [26] encodes all inputs into a latent space:

$$\{\mathcal{Z}_{\text{csi}}, \mathcal{Z}_{\text{cot}}, \hat{\mathcal{Z}}, \mathcal{Z}_{\mathcal{Y}}\} = \mathcal{E}\{\mathcal{X}_{\text{csi}}, \mathcal{X}_{\text{cot}}, \hat{\mathcal{S}}, \mathcal{Y}\}. \tag{19}$$

Flow Matching learns a time-dependent vector field that deterministically transports a source distribution p_0 to a target distribution p_1 along a probability path p_t [15], [27]. At inference time, generating a new sample reduces to drawing from p_0 and integrating the ODE defined by the velocity field. In the proposed framework, the mapping from the biased prediction $\hat{\mathcal{Z}}$ to the ground truth $\mathcal{Z}_{\mathcal{Y}}$ is modeled as

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi_t(g) = u_t(\psi_t(g)), \tag{20}$$

where $t \in [0, 1]$ denotes the time step, ψ_t defines the flow, and the velocity field u_t generates the probability path p_t subject to the boundary conditions

$$\psi_0(g) = \mathcal{G}_0 = \hat{\mathcal{Z}} \sim p_0, \quad \psi_1(g) = \mathcal{G}_1 = \mathcal{Z}_{\mathcal{Y}} \sim p_1. \tag{21}$$

The use of the distribution of $\hat{\mathcal{Z}}$ as the source distribution can be interpreted from a distributional-transport perspective. Let p_0^G denote a standard Gaussian prior, p_0^{AIC} denote the distribution of the encoded AIC estimate $\hat{\mathcal{Z}}$, and p_1 denote the distribution of the encoded target $\mathcal{Z}_{\mathcal{Y}}$. The 2-Wasserstein distance between two distributions is defined as

$$W_2^2(p, q) = \inf_{\pi \in \Pi(p,q)} \mathbb{E}_{(X,Y)\sim\pi}\|X - Y\|^2, \tag{22}$$

where $\Pi(p, q)$ denotes the set of all couplings with marginals p and q. Since the paired samples $(\hat{\mathcal{Z}}, \mathcal{Z}_{\mathcal{Y}})$ naturally define a feasible coupling between p_0^{AIC} and p_1, we have

$$W_2^2(p_0^{\text{AIC}}, p_1) \leq \mathbb{E}\|\hat{\mathcal{Z}} - \mathcal{Z}_{\mathcal{Y}}\|^2. \tag{23}$$

This inequality shows that the source–target discrepancy induced by the AIC prior is upper-bounded by its mean squared residual to the target. In contrast, when a Gaussian prior is used, the velocity field needs to learn both the large-scale cloud-advection structure and the stochastic residual correction from a less informative source. Therefore, the AIC prior provides a physically meaningful initialization that can reduce the effective transport distance when its residual discrepancy to the target is smaller than that of an uninformative prior.

While various constructions of probabilistic paths p_t are available [28], [29], standard linear paths are ill-suited to solar irradiance prediction, where $\hat{\mathcal{Z}}$ and $\mathcal{Z}_{\mathcal{Y}}$ are highly correlated yet differ in high-frequency details induced by cloud dynamics. An overly simple (e.g., linear) interpolation path makes it difficult to disentangle the necessary correction signal from noise. To this end, a stochastic interpolant path with a non-linear schedule is adopted to amplify the gradient signal near the target:

$$\mathcal{G}_t = \alpha_t\mathcal{G}_0 + \beta_t\mathcal{G}_1 + \sigma_t\epsilon, \tag{24}$$

with $\mathcal{G}_0 = \hat{\mathcal{Z}}$, $\mathcal{G}_1 = \mathcal{Z}_{\mathcal{Y}}$, and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$. Equation (24) defines a bridge between the corrected prediction and the target, allowing the velocity field u_t together with the score field implied by σ_t to capture the complex non-linear residuals caused by cloud dynamics. The interpolation coefficients are chosen as

$$\alpha_t = 1 - t, \quad \beta_t = t^2, \quad \sigma_t = \sigma^2 t(1-t)^2, \tag{25}$$

so that the convex trajectory induced by $\beta_t = t^2$ steers the velocity field more aggressively toward the target structure as $t \to 1$. The corresponding conditional vector field, which serves as the regression target for the neural network, is then

$$\hat{u}_t(\mathcal{G}_t|\mathcal{G}_0, \mathcal{G}_1) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G}_t = -\mathcal{G}_0 + 2t\mathcal{G}_1 + \sigma^2(1-t)(1-3t)\epsilon. \tag{26}$$

The model is trained by minimizing the Flow Matching loss

$$\mathcal{L}_\theta = \mathbb{E}_{t,p_0,p_1,\epsilon}\left[\left\|u_t^\theta(\mathcal{G}_t, t, \hat{\mathcal{Z}}, \mathcal{Z}_{\text{csi}}, \mathcal{Z}_{\text{cot}}) - \hat{u}_t(\mathcal{G}_t|\mathcal{G}_0, \mathcal{G}_1)\right\|^2\right]. \tag{27}$$

During training, a denoising UNet [17] parameterizes u_t^θ, with the full procedure summarized in Algorithm 2. At inference, ensemble members are generated by independently sampling stochastic Gaussian boundary completion and integrating the learned velocity field with a forward Euler solver, as detailed in Algorithm 3.

IV. Experimental Setup

A. Datasets PreProcessing

This study employs two key satellite-derived datasets: SSI and COT. The SSI dataset is obtained from SARAH (Surface Solar Radiation Dataset – Heliosat), which is based on observations from the Meteosat Second Generation (MSG)

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2026.3714841

Algorithm 2: WIP-Flow Training

Input:
Training dataset: $\{\mathcal{X}_{\text{csi}}, \mathcal{X}_{\text{cot}}, \mathcal{Y}\}$;
VAE encoder: $\mathcal{E}$;
Noise distribution: $\epsilon \sim \mathcal{N}(0, \mathbf{I})$;
Noise scale: σ_s.
Output:
Learned vector field $u_t^{\theta^*}$.
Perform AIC to obtain $\hat{\mathcal{S}}$ based on Algorithm 1;
Encode $\{\mathcal{Z}_{\text{csi}}, \mathcal{Z}_{\text{cot}}, \hat{\mathcal{Z}}, \mathcal{Z}_{\mathcal{Y}}\} = \mathcal{E}(\mathcal{X}_{\text{csi}}, \mathcal{X}_{\text{cot}}, \hat{\mathcal{S}}, \mathcal{Y})$;
repeat
  Initialize $\mathcal{G}_0 = \hat{\mathcal{Z}}$ and $\mathcal{G}_1 = \mathcal{Z}_{\mathcal{Y}}$ based on Eq. (21);
  Sample $t \sim \mathcal{U}(0, 1)$ and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$;
  Construct $\mathcal{G}_t$ based on Eq. (24);
  Compute $\hat{u}_t(\mathcal{G}_t|\mathcal{G}_0, \mathcal{G}_1)$ based on Eq. (26);
  Update θ by minimizing
  $\mathbb{E}_{t,\mathcal{G}_0,\mathcal{G}_1,\epsilon}\left[\left\|u_t^{\theta}(\mathcal{G}_t, t, \hat{\mathcal{Z}}, \mathcal{Z}_{\text{csi}}, \mathcal{Z}_{\text{cot}}) - \hat{u}_t(\mathcal{G}_t|\mathcal{G}_0, \mathcal{G}_1)\right\|^2\right]$;
until *convergence*
return $u_t^{\theta^*}$

Algorithm 3: WIP-Flow Inference with Forward Euler

Input:
Testing dataset: $\{\mathcal{X}_{\text{csi}}, \mathcal{X}_{\text{cot}}\}$;
VAE encoder $\mathcal{E}$ and decoder $\mathcal{D}$;
Learned vector field $u_t^{\theta^*}$;
Number of ensemble members N;
Number of Euler steps T.
Output:
Forecast ensemble $\{\hat{\mathcal{Y}}^{(r)}\}_{r=1}^{N}$.
Encode $\{\mathcal{Z}_{\text{csi}}, \mathcal{Z}_{\text{cot}}\} = \mathcal{E}(\mathcal{X}_{\text{csi}}, \mathcal{X}_{\text{cot}})$;
for $r = 1, \ldots, N$ **do**
  Perform AIC with an independent boundary-noise realization to obtain $\hat{\mathcal{S}}^{(r)}$ based on Algorithm 1;
  Encode $\hat{\mathcal{Z}}^{(r)} = \mathcal{E}(\hat{\mathcal{S}}^{(r)})$;
  Initialize $\mathcal{G}^{(r)} \leftarrow \hat{\mathcal{Z}}^{(r)}$;
  for $k = 0, \ldots, T-1$ **do**
    Set $t_k = k/T$;
    $\mathcal{G}^{(r)} \leftarrow \mathcal{G}^{(r)} + \frac{1}{T} u_{t_k}^{\theta^*}(\mathcal{G}^{(r)}, t_k, \hat{\mathcal{Z}}^{(r)}, \mathcal{Z}_{\text{csi}}, \mathcal{Z}_{\text{cot}})$;
  Decode $\hat{\mathcal{Y}}^{(r)} = \mathcal{D}(\mathcal{G}^{(r)})$;
return $\{\hat{\mathcal{Y}}^{(r)}\}_{r=1}^{N}$

satellites [30]. The COT dataset is derived from the third-generation CLoud property dAtAset (CLAAS-3), a climate data record generated using data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument on the same MSG satellite [31]. The COT values are computed using the Cloud Physical Properties (CPP) algorithm, which utilizes the 3.8 μm channel to effectively capture cloud microphysical and radiative characteristics. Both datasets cover the European region, spanning from 5.1°W to 20.3°E longitude and 36.9°N to 62.3°N latitude. They are provided at a spatial resolution of $0.2° \times 0.2°$ and a temporal resolution of 30 minutes. Missing values in the time series were imputed to ensure temporal continuity using observations from the same calendar date of the previous year. Nighttime periods were excluded to focus solely on daylight hours critical for solar irradiance forecasting, based on a valid data index derived from the SARAH cloud albedo product. To enhance model stability, all input features are normalized to the range $[-1, 1]$ via min-max scaling. Following established practices [30], [31], we applied this transformation using fixed bounds: $[0.05, 1.2]$ for CSI and $[0.0, 150]$ for COT. Ground-based solar irradiance

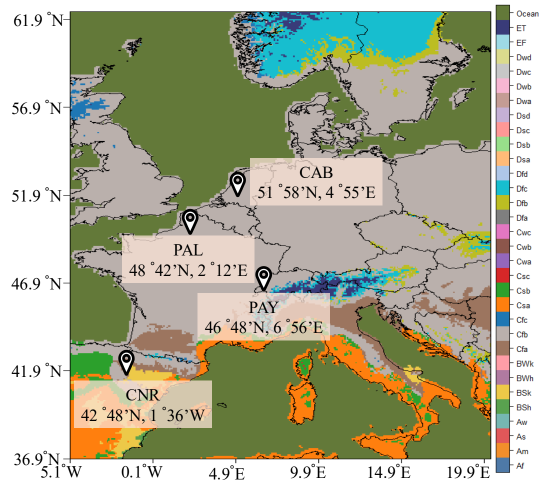

Fig. 2: Geographical distribution of the selected area and the four selected BSRN stations in the updated Köppen–Geiger climate classification system [32].

measurements were used as an independent validation dataset to identify potential inaccuracies in the models. Validation data were sourced from four European Baseline Surface Radiation Network (BSRN) stations [33] as shown in Fig. 2, equipped with high-precision pyranometers. A slight positive bias in SARAH irradiance at these stations has been reported [34]; therefore, the BSRN measurements are used as an independent reference for assessing potential satellite-related deviations. The selected stations span diverse climatic regions, covering the cloudy and variable northwest and the stable, mountainous south, ensuring comprehensive validation.

B. Training Procedure

All training experiments are performed on one NVIDIA A6000 graphics card. In terms of dataset segmentation, we use data from 2017 to 2020 for training, data from 2021 for validation, and data from 2022 for testing. The training process of WIP-Flow is divided into two stages: firstly, training the VAE, and then training Flow Matching. The VAE architecture, as implemented in the latent diffusion model (LDM) [35], is adopted for feature compression to improve training efficiency. Specifically, the VAE encodes an input frame $\mathcal{X} \in \mathbb{R}^{T_h \times 1 \times H \times W}$ into a latent representation $\mathcal{Z} \in \mathbb{R}^{T_h \times 32 \times H/8 \times W/8}$. For the training configuration of Flow Matching, we enabled the exponential moving average mechanism to enhance model stability, set the initial learning rate to 5×10^{-4}, and used a cosine annealing restart strategy for learning rate scheduling, which includes 500 warm-up steps and 2 restart cycles. The AdamW optimizer is utilized for model optimization, setting $\beta_1 = 0.95$, $\beta_2 = 0.999$, and weight decay to 1×10^{-5}. To improve training efficiency, we use the mixed-precision training strategy with 50 Euler steps in the inference stage of the WIP-Flow model. The number of ensemble members generated for all probabilistic forecasts is $N = 20$.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2026.3714841

C. Benchmarks

To comprehensively evaluate the effectiveness of WIP-Flow, we conducted comparative experiments against six probabilistic models: SPFM$_{en}$, DiffSolar [17], SHADECast [16], LDM [35], GAN [12], and Rectified Flow [27], where LDM and Rectified Flow use a standard Gaussian as their initial distribution, and the same UNet backbone is used to parameterize the corresponding generative dynamics where applicable. Additionally, we compared WIP-Flow with two deterministic forecasting models, Earthformer [36] and ConvLSTM [10], which do not have a probabilistic decoding layer, as well as two deterministic NWP models: GFS and HRES. The SPFM ensemble SPFM$_{en}$ is constructed as a baseline based on the historical series. Specifically, the CSI at the same moment of the previous *N* days is extracted from the historical series and combined with the theoretical clear-sky irradiance at the target moment to form the final prediction. To maintain experimental fairness, we configured all benchmark models according to their original hyperparameter settings.

D. Evaluation Metrics

To quantify the forecasting performance, we construct a comprehensive assessment framework that combines probabilistic and deterministic metrics to evaluate the reliability, sharpness, regional dependence structure, and deterministic accuracy of the prediction ensemble [37]. The Continuous Ranked Probability Score (CRPS) is used as the primary marginal probabilistic metric. It measures the distance between the cumulative distribution function $\hat{F}^i$ of the ensemble forecast $\hat{\mathcal{Y}}^i$ and the Heaviside function H centered at the observation $\mathcal{Y}^i$:

$$\mathrm{CRPS} = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\left(\hat{F}^i(x) - H(x-\mathcal{Y}^i)\right)^2 \mathrm{d}x, \tag{28}$$

where n denotes the number of valid grid–time evaluation points. A lower CRPS indicates better marginal probabilistic performance. To provide a standardized comparison against the reference ensemble, the Continuous Ranked Probability Skill Score (CRPSS) is further used:

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}}{\mathrm{CRPS}_{en}}, \tag{29}$$

where CRPS_{en} denotes the CRPS calculated from the reference SPFM_{en} ensemble.

Since CRPS evaluates the marginal predictive distribution at each grid–time location, it does not fully characterize the joint dependence structure of regional probabilistic forecasts. Therefore, two multivariate scoring rules, namely the Energy Score (ES) and the Variogram Score (VS), are further adopted to assess the spatial and temporal coherence of the generated ensemble scenarios [37], [38]. For each forecast case and lead time, each ensemble member is flattened over the regional grid to form a multivariate vector. The ES and VS values reported in this study are computed separately at each forecast lead time and then averaged over all valid daytime test cases. The Energy Score extends CRPS to multivariate forecasts and is defined as

$$\mathrm{ES}(F,\mathbf{y}) = \mathbb{E}_F\left\|\mathbf{X}-\mathbf{y}\right\|_2 - \frac{1}{2}\mathbb{E}_F\left\|\mathbf{X}-\mathbf{X}'\right\|_2, \tag{30}$$

where $\mathbf{X}$ and $\mathbf{X}'$ are independent ensemble members drawn from the forecast distribution F, and $\mathbf{y}$ denotes the observed regional irradiance vector. ES evaluates the overall distance between the forecast ensemble distribution and the observed regional field.The Variogram Score of order p is used to evaluate pairwise dependence structures:

$$\mathrm{VS}_p(F,\mathbf{y}) = \sum_{i,j} w_{ij}\left(|y_i - y_j|^p - \mathbb{E}_F|X_i - X_j|^p\right)^2, \tag{31}$$

where i and j denote two spatial locations in the flattened regional field. The weight is defined as $w_{ij} = \mathbb{I}(d_{ij} \le d_0)$, where d_{ij} is the normalized distance between locations i and j, and d_0 is a prescribed neighborhood threshold. Thus, only local neighboring pairs are included in the score, which emphasizes cloud-driven local dependence structures. Following [38], $p = 0.5$ is adopted for robustness to extreme irradiance values. Both ES and $\mathrm{VS}_{0.5}$ are negatively oriented.

In addition to probabilistic metrics, the root mean square error (RMSE), mean absolute error (MAE), and forecast skill score ($\mathrm{S_f}$) are used to evaluate the deterministic prediction accuracy of the ensemble mean:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\mathcal{Y}^j - \bar{\mathcal{Y}}^j\right)^2}, \tag{32}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}\left|\mathcal{Y}^j - \bar{\mathcal{Y}}^j\right|, \tag{33}$$

where $\bar{\mathcal{Y}}^j$ denotes the ensemble mean prediction at the j-th valid evaluation point. The forecast skill score is calculated relative to the reference SPFM_{en} forecast:

$$\mathrm{S_f} = 1 - \frac{\mathrm{RMSE_{Forecast}}}{\mathrm{RMSE}_{\mathrm{SPFM}_{en}}}. \tag{34}$$

A larger $\mathrm{S_f}$ indicates stronger deterministic forecasting skill relative to the reference model.

V. Results and Discussion

A. Forecast Assessment

Table I reports the probabilistic performance across four forecast horizons under both marginal scoring rules (CRPS and CRPSS) and multivariate scoring rules (ES and $\mathrm{VS}_{0.5}$). All metrics are averaged over the complete satellite-derived spatial domain and all valid daytime forecast cases. CRPS and CRPSS evaluate marginal probabilistic accuracy at each grid–time location, whereas ES and $\mathrm{VS}_{0.5}$ assess the joint spatio–temporal distribution of the generated ensemble scenarios. WIP-Flow achieves the lowest CRPS at the 0.5-, 1-, and 2-hour horizons, corresponding to CRPSS values of 0.338, 0.342, and 0.329, respectively. At the 4-hour horizon, DiffSolar marginally outperforms WIP-Flow in CRPS, whereas WIP-Flow achieves the best $\mathrm{VS}_{0.5}$, indicating stronger pairwise spatio–temporal coherence. The multivariate scores further

TABLE I: Probabilistic predictive performance under marginal and multivariate scoring rules.

Method	0.5-hour-ahead				1-hour-ahead				2-hour-ahead				4-hour-ahead			
	CRPS	CRPSS	ES	$VS_{0.5}$	CRPS	CRPSS	ES	$VS_{0.5}$	CRPS	CRPSS	ES	$VS_{0.5}$	CRPS	CRPSS	ES	$VS_{0.5}$
$SPFM_{en}$	53.32	–	108.2	3.182	71.91	–	146.5	4.487	84.87	–	172.6	5.276	74.39	–	152.4	4.713
GAN [12]	38.70	0.274	81.47	2.508	57.52	0.200	119.8	3.706	67.93	0.199	142.3	4.385	57.85	0.222	121.6	3.842
LDM [35]	42.66	0.199	89.74	2.738	53.33	0.258	112.6	3.463	64.51	0.240	135.4	4.176	54.07	0.273	114.8	3.624
SHADECast [16]	40.18	0.246	84.58	2.586	52.30	0.273	109.8	3.358	60.45	0.288	127.5	3.917	50.30	0.324	107.2	3.392
DiffSolar [17]	38.47	0.279	80.78	2.453	49.17	0.316	103.5	3.158	58.13	0.315	122.3	3.742	**47.11**	**0.367**	**100.4**	3.165
Rectified Flow [27]	42.35	0.205	88.93	2.713	58.37	0.188	121.7	3.738	59.25	0.301	124.7	3.864	56.11	0.245	117.8	3.658
WIP-Flow	**36.28**	**0.338**	**76.24**	**2.314**	**47.29**	**0.342**	**99.68**	**3.041**	**56.91**	**0.329**	**117.4**	**3.547**	47.81	0.357	101.8	**3.073**

demonstrate the advantage of WIP-Flow in preserving physical cloud-driven patterns. WIP-Flow achieves the lowest $VS_{0.5}$ across all four horizons, with values of 2.314, 3.041, 3.547, and 3.073. Compared with DiffSolar, the reductions in $VS_{0.5}$ are 5.7%, 3.7%, 5.2%, and 2.9% at the 0.5-, 1-, 2-, and 4-hour horizons, respectively. Compared with SHADECast, the corresponding reductions are 10.5%, 9.4%, 9.4%, and 9.4%. On ES, WIP-Flow ranks first at the 0.5-, 1-, and 2-hour horizons and remains close to the best-performing model at 4 hours. These results indicate that WIP-Flow improves marginal probabilistic accuracy at short forecast horizons and consistently enhances the regional dependence structure of the forecast ensemble, as reflected by the best $VS_{0.5}$ across all horizons.

Table II summarizes the deterministic forecasting performance of various models. For probabilistic approaches, the evaluation metrics are computed based on the ensemble mean of 20 generated samples. WIP-Flow achieves the lowest RMSE and the highest S_f across all four prediction horizons, while maintaining competitive MAE values. Specifically, for the 0.5-hour prediction horizon, WIP-Flow achieved the lowest MAE of 47.91 W/m^2 and RMSE of 83.15 W/m^2, while simultaneously attaining the highest S_f of 0.238. Compared with existing generative models and deep learning benchmarks, WIP-Flow effectively enhances deterministic forecast accuracy through the coupling of AIC-informed initialization and flow-matching correction. Notably, even over longer time spans, WIP-Flow maintains a highly competitive RMSE (104.3 W/m^2) and the highest S_f score (0.278), demonstrating its robustness in capturing complex irradiance dynamics. Figure 3 illustrates a visual comparison of the WIP-Flow model against ground truth, AIC, and Rectified Flow across four forecast horizons (0.5 to 4.0 hours), highlighting the high consistency of WIP-Flow's predicted SSI with the ground truth in terms of cloud motion. The WIP-Flow samples remain spatially coherent near the advected boundaries, indicating that the stochastic Gaussian boundary completion is effectively corrected by the subsequent Flow Matching transport rather than being propagated as noise-like artifacts.

Furthermore, the AIC exhibits predictive capabilities that are superior to those of traditional NWP models. Despite its relatively simple structure, the AIC consistently yields MAE and RMSE values that are significantly lower than those of GFS and HRES across all prediction horizons. For instance, in the 0.5-hour forecast, AIC's MAE (54.07 W/m^2) is substantially lower than that of GFS (94.28 W/m^2). This indicates that

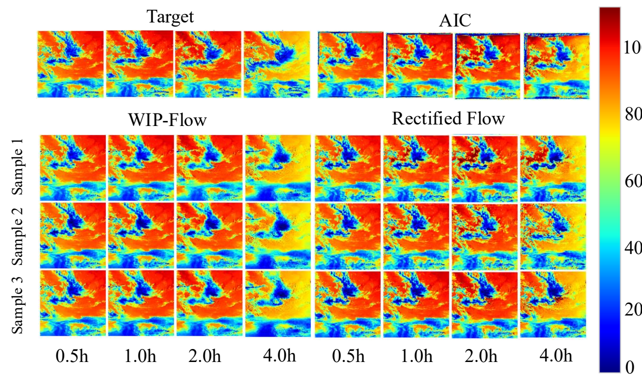

Fig. 3: Visual comparison of target SSI fields, AIC outputs, WIP-Flow samples, and Rectified Flow samples at 11:00 on 2022-06-02. Three generated members are shown for each probabilistic model across four forecast horizons.

advection extrapolation based on optical flow can effectively capture cloud motion trends in nowcasting, providing a solid physical foundation for more complex models like WIP-Flow. However, as the prediction horizon extends, the advantages of WIP-Flow become even more pronounced, as it successfully combines the short-term precision of advection-based transport with the long-term generative modeling of complex cloud dynamics.

B. Ablation Experiments

To systematically evaluate the contribution of each key component in WIP-Flow, we conduct ablation experiments from four perspectives: the AIC module, the weakly informative prior construction, the stochastic interpolation path, and the COT input. For generative ablations, all variants use the same VAE latent representation, denoising UNet backbone, training strategy, and testing split unless otherwise specified. These controlled comparisons are designed to clarify whether the final performance gain arises from a single dominant component or from the coupling between the AIC-informed prior and the stochastic Flow Matching correction.

1) Effect of the AIC Module: To evaluate the contribution of the AIC module, we conduct an internal ablation with three deterministic variants: (i) semi-Lagrangian extrapolation only, which advects the latest CSI field using the estimated cloud motion vectors; (ii) semi-Lagrangian extrapolation with cascade decomposition, which introduces multi-scale separation

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2026.3714841

TABLE II: Deterministic Predictive Performance Comparison of Different Models

Method	Prediction Horizons											
	0.5-hour-ahead			1-hour-ahead			2-hour-ahead			4-hour-ahead		
	MAE	RMSE	S_f	MAE	RMSE	S_f	MAE	RMSE	S_f	MAE	RMSE	S_f
$SPFM_{en}$	63.77	109.2	-	81.30	136.1	-	93.12	155.7	-	85.21	144.4	-
GAN [12]	55.32	95.46	0.126	71.69	120.6	0.114	81.75	135.7	0.128	68.53	122.4	0.152
LDM [35]	63.38	106.1	0.028	79.58	130.1	0.044	87.21	141.0	0.094	71.81	123.9	0.141
SHADECast [16]	55.53	95.71	0.124	72.19	121.4	0.109	82.18	136.3	0.124	68.89	122.6	0.150
DiffSolar [17]	52.31	84.02	0.231	62.13	104.2	0.234	69.77	116.5	0.252	**60.28**	105.9	0.267
Rectified Flow [27]	53.78	95.78	0.123	70.80	123.3	0.094	84.37	143.1	0.081	73.84	134.3	0.070
ConvLSTM [10]	63.23	94.37	0.136	71.76	109.4	0.195	76.75	124.9	0.198	66.02	117.5	0.186
Earthformer [36]	55.74	86.60	0.211	64.85	108.4	0.243	**63.09**	114.7	0.263	64.62	107.8	0.253
GFS	94.28	123.5	-0.131	97.53	136.8	-0.005	99.59	146.9	0.056	107.5	158.9	-0.100
HRES	93.50	121.9	-0.116	95.41	127.6	0.062	98.08	136.9	0.121	103.4	145.8	-0.010
AIC	54.07	97.39	0.108	70.52	123.8	0.091	83.41	143.2	0.081	72.39	132.6	0.082
WIP-Flow	**47.91**	**83.15**	**0.238**	**61.66**	**102.4**	**0.248**	67.57	**110.4**	**0.290**	61.74	**104.3**	**0.278**

TABLE III: Ablation study of the AIC module. RMSE is reported at different lead times.

Variant	0.5 h	1 h	2 h	4 h
Semi-Lagrangian only	104.8	132.6	153.9	146.8
Semi-Lagrangian + cascade	101.2	128.7	148.6	139.4
Full AIC	**97.39**	**123.8**	**143.2**	**132.6**

but removes autoregressive correction; and (iii) the full AIC module, which applies autoregressive integrated correction at each cascade level.

Table III shows that each component in AIC improves the deterministic prior. Cascade decomposition reduces the 4-hour-ahead RMSE from 146.8 to 139.4 W/m^2, while the full AIC further reduces it to 132.6 W/m^2, confirming that autoregressive integrated correction helps compensate scale-dependent cloud growth and decay after advection. Nevertheless, AIC alone remains much weaker than the full WIP-Flow model, which achieves a 4-hour-ahead RMSE of 104.3 W/m^2. This indicates that AIC provides a useful but incomplete deterministic prior, and that the subsequent Flow Matching correction is essential for modeling stochastic cloud evolution and probabilistic uncertainty.

2) Effect of Weakly Informative Prior Construction: To evaluate the effectiveness of using AIC predictions as the initial distribution for WIP-Flow training, we designed three comparative experiments. In these experiments, the initial distributions were set as standard Gaussian noise, SPFM predictions, and AIC predictions, respectively. For each case, the proposed probabilistic interpolation path was applied to map the vector field between the initial and target distributions. The 4-hour-ahead deterministic prediction results across 20 samples are illustrated in Fig. 4.

Among the tested source distributions, the AIC-based initialization yields the lowest RMSE values, primarily concentrated between 100 and 110 W/m^2. This indicates that AIC provides a more informative source distribution than Gaussian noise or SPFM predictions for the subsequent generative transport. This result confirms that AIC provides a more informative initial distribution for Flow Matching by reducing the source–target discrepancy. However, it does not imply that AIC alone is responsible for the final forecasting performance. As reported in Table II, the standalone AIC baseline obtains a 4-hour-ahead RMSE of 132.6 W/m^2, whereas the full WIP-Flow model reduces the error to 104.3 W/m^2. The large gap between AIC and WIP-Flow demonstrates that the deterministic AIC prior is insufficient for accurate regional SSI forecasting, especially because it cannot model cloud formation, dissipation, deformation, boundary uncertainty, or the ensemble spread required for probabilistic prediction. Therefore, AIC should be interpreted as a weakly informative physical prior rather than as a complete forecasting model. Its main functionality is to initialize the Flow Matching process from a physically plausible cloud-advection state, allowing the velocity network to focus on learning the stochastic residual transformation from the AIC-biased estimate to the true future distribution.

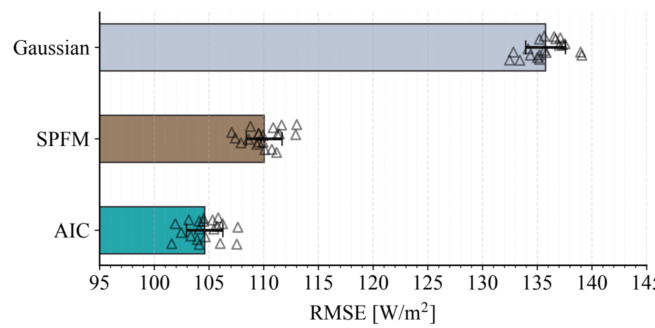

Fig. 4: RMSE comparison of 4-hour-ahead predictions using different initial distributions (Gaussian, SPFM, and AIC) of 20 samples. Individual sample is plotted as triangles along the x-axis.

3) Effect of Stochastic Interpolation Path: To validate the effectiveness of the proposed SDE-based stochastic interpolation path, we conducted comparative experiments against two established baselines: the deterministic straight-line interpolation path proposed by Liu et al. [27] (Rectified Flow) and the Optimal Transport Schrödinger Bridge (OT-SBFM) proposed by Tong et al. [39]. As illustrated in Fig. 5, the proposed WIP-Flow consistently achieves the lowest RMSE across all forecast horizons. While the prediction error naturally increases for all models as the forecast horizon extends, WIP-Flow demonstrates superior stability. It maintains an RMSE

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2026.3714841

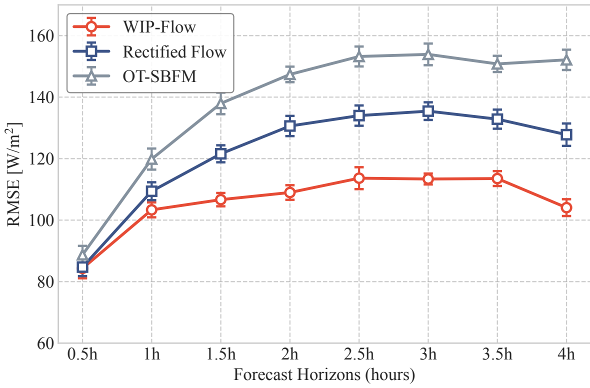

Fig. 5: Comparison of forecasting performance of different interpolation path across different prediction horizons.

significantly below 120 W/m^2 even at the 4-hour horizon. In sharp contrast, both baseline methods exhibit considerably higher error values, with OT-SBFM showing the largest errors in this forecasting setting, reaching an RMSE peak exceeding 150 W/m^2. Rectified Flow relies on a deterministic straight-line interpolation to connect the noise distribution to the data distribution. However, the evolution of solar irradiance is a highly non-linear dynamic process influenced by complex atmospheric factors. The rigid assumption of a straight trajectory oversimplifies these complex temporal dynamics, failing to capture the intrinsic non-linear manifold of the meteorological data, which leads to lower prediction accuracy. Although OT-SBFM theoretically minimizes transport costs based on Optimal Transport principles, this mathematical optimality does not necessarily align with the physical reality of weather evolution. This rigid coupling constraint limits the exploration of the state space, resulting in the highest RMSE among the tested methods.

This comparison further confirms that the performance gain of WIP-Flow is not merely inherited from the AIC prior. If the deterministic prior were the dominant or sole source of improvement, different transport paths using the same AIC-informed source distribution would lead to similar forecasting errors. However, the results in Fig. 5 show substantial differences among the interpolation strategies. The proposed stochastic interpolation path consistently achieves the lowest RMSE, indicating that the Flow Matching formulation and the choice of probability path provide an additional correction capability beyond deterministic AIC initialization. In particular, the stochastic path allows the model to explore non-deterministic residual cloud evolution around the AIC estimate, thereby improving the representation of cloud formation, diffusion, dissipation, and boundary uncertainty.

4) Effect of COT Input: To examine whether the performance improvement of WIP-Flow is mainly attributed to the additional COT input, we conduct an input-channel ablation experiment by removing COT from the conditioning variables. The resulting variant, denoted as WIP-Flow w/o COT, uses only historical CSI fields together with the AIC-informed source distribution, while keeping the VAE architecture, denoising UNet backbone, Flow Matching path, training strategy, and inference configuration unchanged.

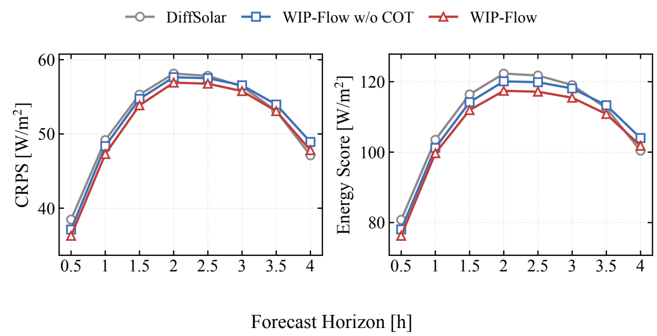

Fig. 6: Ablation study on the contribution of COT input.

As shown in Fig. 6, removing COT causes a moderate increase in both CRPS and ES, confirming that COT provides useful cloud optical information for probabilistic SSI forecasting. Nevertheless, WIP-Flow w/o COT remains close to the full WIP-Flow model and remains competitive with DiffSolar across most forecast horizons. This indicates that the performance gain of WIP-Flow is not solely attributable to the additional COT input; rather, the AIC-informed source distribution and Flow Matching correction remain the primary methodological contributors, while COT serves as a complementary cloud-related input.

C. Sensitivity to AIC Prior

Since the optical-flow-based AIC prior may become less reliable when brightness constancy or spatial smoothness is violated, we further assess the sensitivity of WIP-Flow to AIC-prior quality. Specifically, we perform a post-hoc stratified analysis using the standalone AIC 4-hour-ahead regional RMSE. This grouping is used only for diagnostic evaluation and is not involved in training or inference. The testing samples are ranked by the AIC error and divided into three equally sized groups using the empirical 33.3% and 66.7% quantiles. The low-error group represents cases with relatively reliable AIC priors, whereas the high-error group corresponds to strongly biased priors, which may result from rapid cloud formation, dissipation, deformation, boundary effects, or other violations of optical-flow assumptions.

As shown in Fig. 7, WIP-Flow consistently outperforms standalone AIC across all reliability groups. The RMSE is reduced from 82.5 to 70.8 W/m^2 in the low-error group, from 138.7 to 110.9 W/m^2 in the medium-error group, and from 176.6 to 131.2 W/m^2 in the high-error group, corresponding to relative improvements of 14.2%, 20.0%, and 25.7%, respectively. These results directly address the sensitivity of WIP-Flow to degraded AIC priors and indicate that the proposed model learns corrective stochastic transport beyond direct propagation of the AIC trajectory.

D. Computational Cost

The computational cost of WIP-Flow and representative generative baselines is summarized in Table IV. All measurements are conducted on a workstation equipped with an Intel

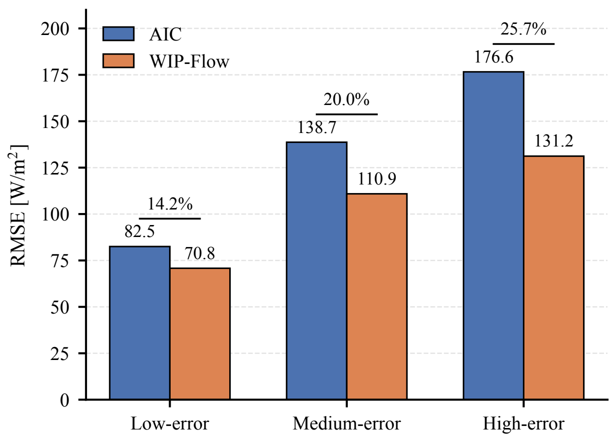

Fig. 7: AIC-error-stratified 4-hour-ahead RMSE comparison between standalone AIC and WIP-Flow. Percentages indicate the relative RMSE reduction achieved by WIP-Flow over AIC.

TABLE IV: Computational cost on the NVIDIA A6000 GPU under FP16. Inference latency is reported per probabilistic forecast of 20 ensemble members.

Method	Train (h)	Steps	Inference (s)
LDM [35]	105.9	100	136
SHADECast [16]	136.8	100	178
DiffSolar [17]	116.0	100	129
Rectified Flow [27]	74.90	50	28
WIP-Flow	**71.20**	50	**26**

Xeon Platinum 8352V CPU and NVIDIA A6000 48GB GPU. Inference latency is measured on a single NVIDIA A6000 GPU under FP16 precision with batch size 1, and corresponds to generating a 20-member ensemble over the full study region and all prediction horizons from 0.5 h to 4 h. As shown in Table IV, WIP-Flow requires 71.2 h for training, including 48.6 h for VAE training and 22.6 h for Flow Matching training, whereas DiffSolar requires 116.0 h in total, corresponding to a 38.6% reduction in training time. At inference time, WIP-Flow requires 26 s to generate a 20-member ensemble, compared with 129 s for DiffSolar, achieving a 79.8% latency reduction or a 4.96× speedup. This efficiency gain mainly comes from the flow-matching formulation, which generates samples through ODE integration with fewer neural-network evaluations than diffusion-based iterative denoising. Although WIP-Flow includes AIC preprocessing, AIC is parameter-free and introduces limited overhead. Compared with Rectified Flow, WIP-Flow has a similar computational cost but achieves better forecasting performance, indicating that the AIC-informed prior improves accuracy without increasing model complexity.

E. Independent Station-Based Validation against BSRN Measurements

Figure 8 presents a comparative analysis of the conditional distributions generated by DiffSolar and WIP-Flow against ground-based measurements from four BSRN stations. While both models exhibit strong linear consistency with the measured data, WIP-Flow achieves lower RMSE and MAE values at all four stations. Visually, the joint density of WIP-Flow is more tightly concentrated around the identity line than that of DiffSolar, particularly in the medium- and high-irradiance ranges. This indicates that WIP-Flow provides more accurate station-level irradiance estimates with fewer large deviations.

Specifically, the RMSE values decrease from 123.3 to 121.4 W/m^2 at CAB, from 135.7 to 132.7 W/m^2 at CNR, from 126.1 to 122.3 W/m^2 at PAL, and from 120.4 to 118.2 W/m^2 at PAY. The corresponding MAE values decrease from 84.4 to 80.7 W/m^2, from 88.6 to 85.4 W/m^2, from 84.4 to 79.9 W/m^2, and from 80.8 to 76.3 W/m^2, respectively. The largest RMSE reduction is observed at the PAL station, where WIP-Flow reduces the RMSE by approximately 3.0% relative to DiffSolar. Furthermore, the marginal distributions predicted by WIP-Flow show better agreement with the measured distributions, whereas DiffSolar exhibits a slightly broader spread and more pronounced deviations in several low-irradiance cases. These results confirm that WIP-Flow improves independent station-based SSI prediction accuracy over the diffusion-based baseline.

As shown in Fig. 9, the WIP-Flow model demonstrates superior capability in capturing the complex temporal dynamics of solar irradiance compared to the baseline DiffSolar model. While both models generally follow the diurnal trend, WIP-Flow exhibits significantly higher sensitivity to rapid, high-frequency fluctuations caused by cloud cover. In contrast to DiffSolar, which tends to produce overly smoothed predictions that lag behind sudden changes, WIP-Flow aligns closely with the ground truth even during periods of extreme volatility, such as the sharp oscillations observed between 10:00 and 14:00 at the PAL station.

F. Downstream PV Power Forecasting Validation

To assess whether the SSI forecasting improvements can be transferred to downstream PV power prediction, an additional plant-level validation is conducted using the Stutensee PVOutput system. The plant is located at $49.12°$N, $8.51°$E, within the spatial coverage of the satellite-derived SSI dataset used in this study. It provides measured AC active power and essential system metadata for irradiance-to-power conversion, including a reported DC capacity of 9.92 kW, an inverter AC capacity of 9.00 kW, an array azimuth of $225°$, and a tilt angle of $35°$. For both WIP-Flow and DiffSolar, the SSI forecasts at the nearest satellite grid cell are converted into normalized PV AC power using the same PVWatts pipeline implemented in pvlib [40] and identical plant metadata. Deterministic errors are then calculated using the ensemble-mean PV power forecasts. PV power and the corresponding forecasting errors are reported in per-unit (p.u.) values, defined as:

$$P_{\text{p.u.}} = \frac{P}{P_{\text{rated}}}, \tag{35}$$

where P denotes the PV AC power and P_{rated} is the reported DC capacity of the PV system, i.e., 9.92 kW in this experiment.

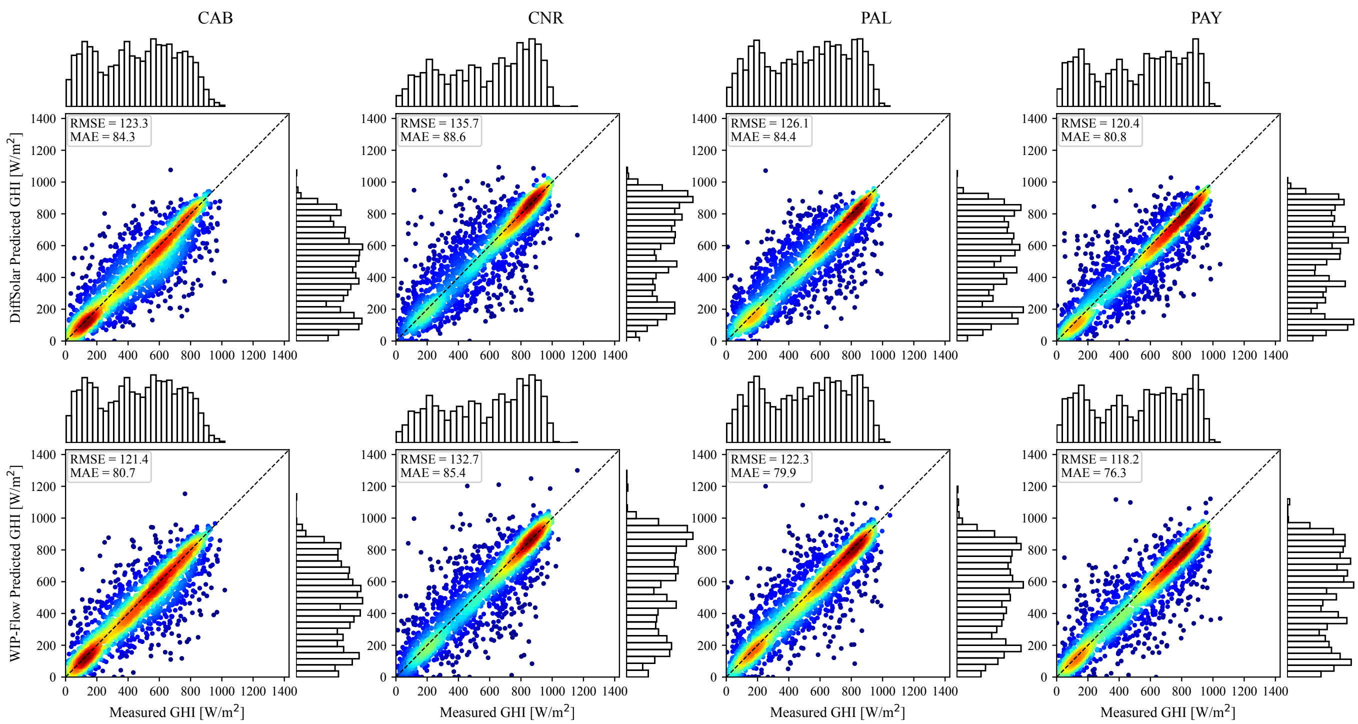

Fig. 8: Joint and marginal distributions of measured and 4-hour-ahead predicted SSI using DiffSolar and WIP-Flow when evaluated at CAB, CNR, PAL, and PAY stations. The heat maps show the two-dimensional kernel density estimates.

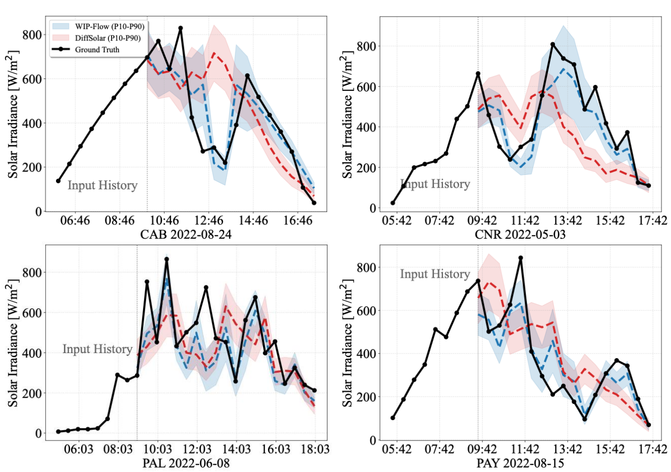

Fig. 9: Visual comparison of forecasting performance between DiffSolar and WIP-Flow against the ground truth. The first 4 hours are continuously generated predictions, and the data for each subsequent time point are taken from the last time step of each rolling prediction sample. The plots illustrate the P10-P90 prediction intervals across different dates and sites.

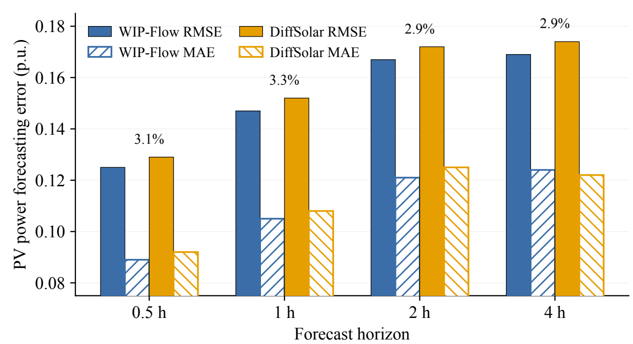

Fig. 10: Downstream PV power forecasting performance on the Stutensee PVOutput system. Percentages denote the relative RMSE reduction of WIP-Flow compared with DiffSolar at each forecast horizon.

Figure 10 presents the downstream PV power forecasting performance of WIP-Flow and DiffSolar across four forecast horizons, where deterministic errors are computed using the ensemble-mean PV power forecasts. WIP-Flow consistently achieves lower RMSE than DiffSolar at all evaluated horizons. The RMSE is reduced from 0.129–0.174 p.u. to 0.125–0.169 p.u., corresponding to relative reductions of 3.1%, 3.3%, 2.9%, and 2.9% from 0.5 to 4 hours. For MAE, WIP-Flow achieves lower errors at the 0.5-, 1-, and 2-hour horizons, while its 4-hour-ahead MAE is slightly higher than that of DiffSolar. Overall, the results indicate that WIP-Flow provides consistent RMSE improvements and comparable MAE performance in downstream PV power forecasting, demonstrating that its SSI forecasting gains can be transferred to PV power prediction under a unified irradiance-to-power conversion pipeline.

VI. Conclusion

In this study, we addressed the critical challenge of characterizing irradiance uncertainty in regional SSI forecasting, a task complicated by stochastic cloud dynamics. We introduced WIP-Flow, a novel Flow Matching-based framework that introduces Flow Matching into probabilistic regional SSI forecasting. By integrating an unsupervised autoregressive correction algorithm with a stochastic interpolation method, WIP-Flow effectively synthesizes physical insights, specifically cloud advection and multi-scale decomposition, with advanced generative modeling. Our experimental results on satellite-derived datasets validate that the proposed framework not only captures complex spatiotemporal dependencies but also rectifies common boundary distortions associated with cloud formation and dissipation. Achieving a RMSE of 104.3 W/m^2 and a CRPS of 47.81 W/m^2 for 4-hour-ahead predictions, WIP-Flow demonstrates competitive and physically coherent probabilistic forecasting performance. Furthermore, the added plant-level PVOutput validation demonstrates that the SSI forecasting gains of WIP-Flow can be transferred to downstream PV power forecasting under a consistent irradiance-to-power conversion pipeline. These findings establish WIP-Flow as a robust tool for generating physically coherent probabilistic predictions, offering significant potential to enhance the stability and management of large-scale photovoltaic grid integration.

REFERENCES

[1] D. Yang and J. Kleissl, *Solar irradiance and photovoltaic power forecasting*. CRC Press, 2024.

[2] K. Wang, T. Wang, S. Shan, W. Dou, J. Zhang, H. Wei, and K. Zhang, "Multimodal ensemble photovoltaic power forecasting incorporating diffusion-based satellite image prediction," *IEEE Transactions on Sustainable Energy*, pp. 1–11, 2025.

[3] T. Jing, S. Chen, D. Navarro-Alarcon, Y. Chu, and M. Li, "Solarfusionnet: Enhanced solar irradiance forecasting via automated multimodal feature selection and cross-modal fusion," *IEEE Transactions on Sustainable Energy*, vol. 16, no. 2, pp. 761–773, 2025.

[4] B. Schubnel, J. Simeunović, C. Tissier, P.-J. Alet, and R. E. Carrillo, "Solarcrossformer: Improving day-ahead solar irradiance forecasting by integrating satellite imagery and ground sensors," *IEEE Transactions on Sustainable Energy*, pp. 1–14, 2025.

[5] M. Kakimoto, Y. Endoh, H. Shin, R. Ikeda, and H. Kusaka, "Probabilistic solar irradiance forecasting by conditioning joint probability method and its application to electric power trading," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 983–993, 2018.

[6] K. Doubleday, S. Jascourt, W. Kleiber, and B.-M. Hodge, "Probabilistic solar power forecasting using bayesian model averaging," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 325–337, 2020.

[7] P. A. Jimenez, J. P. Hacker, J. Dudhia, S. E. Haupt, J. A. Ruiz-Arias, C. A. Gueymard, G. Thompson, T. Eidhammer, and A. Deng, "Wrf-solar: Description and clear-sky assessment of an augmented nwp model for solar power prediction," *Bulletin of the American Meteorological Society*, vol. 97, no. 7, pp. 1249–1264, 2016.

[8] H. Kim and D. Lee, "Probabilistic solar power forecasting based on bivariate conditional solar irradiation distributions," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 2031–2041, 2021.

[9] M. Khodayar, S. Mohammadi, M. E. Khodayar, J. Wang, and G. Liu, "Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 571–583, 2019.

[10] A. H. Nielsen, A. Iosifidis, and H. Karstoft, "Irradiancenet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting," *Solar Energy*, vol. 228, pp. 659–669, 2021.

[11] A. Carpentieri, D. Folini, D. Nerini, S. Pulkkinen, M. Wild, and A. Meyer, "Intraday probabilistic forecasts of surface solar radiation with cloud scale-dependent autoregressive advection," *Applied Energy*, vol. 351, p. 121775, 2023.

[12] H. Wen, Y. Du, X. Chen, E. G. Lim, H. Wen, and K. Yan, "A regional solar forecasting approach using generative adversarial networks with solar irradiance maps," *Renewable Energy*, vol. 216, p. 119043, 2023.

[13] Y. Liu, H. Qin, Z. Zhang, S. Pei, C. Wang, X. Yu, Z. Jiang, and J. Zhou, "Ensemble spatiotemporal forecasting of solar irradiation using variational bayesian convolutional gate recurrent unit network," *Applied Energy*, vol. 253, p. 113596, 2019.

[14] Y. Hatanaka, Y. Glaser, G. Galgon, G. Torri, and P. Sadowski, "Diffusion models for high-resolution solar forecasts," *arXiv preprint arXiv:2302.00170*, 2023.

[15] Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow matching for generative modeling," *arXiv preprint arXiv:2210.02747*, 2022.

[16] A. Carpentieri, D. Folini, J. Leinonen, and A. Meyer, "Extending intraday solar forecast horizons with deep generative models," *Applied Energy*, vol. 377, p. 124186, 2025.

[17] T. Jing, S. Chen, T. Du, and M. Li, "Diffsolar: Prior knowledge-guided residual diffusion model for enhanced regional-scale probabilistic solar irradiance forecasting," *10.36227/techrxiv.176115866.65619398/v2*, 2025.

[18] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, and C. F. Coimbra, "History and trends in solar irradiance and pv power forecasting: A preliminary assessment and review using text mining," *Solar Energy*, vol. 168, pp. 60–101, 2018.

[19] S. Chen, T. Jing, M. Li, H. H. Lee, and S. Bu, "Improved satellite-based estimation of direct normal solar irradiance using clearness index with deep learning methods," *Solar Energy*, vol. 302, p. 113991, 2025.

[20] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola *et al.*, "McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmospheric Measurement Techniques*, vol. 6, no. 9, pp. 2403–2418, 2013.

[21] J.-Y. Bouguet *et al.*, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel corporation*, vol. 5, no. 1-10, p. 4, 2001.

[22] U. Germann and I. Zawadzki, "Scale-dependence of the predictability of precipitation from continental radar images. part i: Description of the methodology," *Monthly Weather Review*, vol. 130, no. 12, pp. 2859–2873, 2002.

[23] S. Pulkkinen, V. Chandrasekar, and A.-M. Harri, "Fully spectral method for radar-based precipitation nowcasting," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 5, pp. 1369–1382, 2019.

[24] S. Pulkkinen, V. Chandrasekar, A. von Lerber, and A.-M. Harri, "Nowcasting of convective rainfall using volumetric radar observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7845–7859, 2020.

[25] L. H. Koopmans, *The spectral analysis of time series*. Elsevier, 1995.

[26] D. P. Kingma, M. Welling *et al.*, "An introduction to variational autoencoders," *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.

[27] X. Liu, C. Gong, and Q. Liu, "Flow straight and fast: Learning to generate and transfer data with rectified flow," *arXiv preprint arXiv:2209.03003*, 2022.

[28] Y. Lipman, M. Havasi, P. Holderrieth, N. Shaul, M. Le, B. Karrer, R. T. Chen, D. Lopez-Paz, H. Ben-Hamu, and I. Gat, "Flow matching guide and code," *arXiv preprint arXiv:2412.06264*, 2024.

[29] Y. Chen, M. Goldstein, M. Hua, M. S. Albergo, N. M. Boffi, and E. Vanden-Eijnden, "Probabilistic forecasting with stochastic interpolants and follmer processes," *arXiv preprint arXiv:2403.13724*, 2024.

[30] U. Pfeifroth, J. Drücke, S. Kothe, J. Trentmann, M. Schröder, and R. Hollmann, "Sarah-3–satellite-based climate data records of surface solar radiation," *Earth System Science Data*, vol. 16, no. 11, pp. 5243–5265, 2024.

[31] N. Benas, I. Solodovnik, M. Stengel, I. Hüser, K.-G. Karlsson, N. Håkansson, E. Johansson, S. Eliasson, M. Schröder, R. Hollmann *et al.*, "Claas-3: the third edition of the cm saf cloud data record based on seviri observations," *Earth System Science Data*, vol. 15, no. 11, pp. 5153–5170, 2023.

[32] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, "Present and future Köppen-Geiger climate classification maps at 1-km resolution," *Scientific data*, vol. 5, no. 1, pp. 1–12, 2018.

[33] A. Driemel, J. Augustine, K. Behrens, S. Colle, C. Cox, E. Cuevas-Agulló, F. M. Denn, T. Duprat, M. Fukuda, H. Grobe *et al.*, "Baseline surface radiation network (bsrn): structure and data description (1992–2017)," *Earth System Science Data*, vol. 10, no. 3, pp. 1491–1501, 2018.

[34] D. Yang and J. M. Bright, "Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and

overall metrics for hourly data over 27 years," *Solar Energy*, vol. 210, pp. 3–19, 2020.

[35] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[36] Z. Gao, X. Shi, H. Wang, Y. Zhu, Y. B. Wang, M. Li, and D.-Y. Yeung, "Earthformer: Exploring space-time transformers for earth system forecasting," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25 390–25 403, 2022.

[37] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 69, no. 2, pp. 243–268, 2007.

[38] M. Scheuerer and T. M. Hamill, "Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities," *Monthly Weather Review*, vol. 143, no. 4, pp. 1321–1334, 2015.

[39] A. Tong, K. Fatras, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, G. Wolf, and Y. Bengio, "Improving and generalizing flow-based generative models with minibatch optimal transport," *arXiv preprint arXiv:2302.00482*, 2023.

[40] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: a python package for modeling solar energy systems," *Journal of Open Source Software*, vol. 3, no. 29, p. 884, 2018.