PVCAST: CLOUD EVOLUTION FOR JOINT PROBABILISTIC FORECASTING OF DISTRIBUTED SOLAR POWER

Anonymous authors
Paper under double-blind review

ABSTRACT

Evolving clouds are an important source of PV power variability, and uncertainty in their evolution affects forecasts across locations and time. These effects shape both marginal power distributions and their spatiotemporal dependence. Prior work largely treats uncertainty in cloud evolution and dependence among PV outputs as separate modeling problems. We introduce PVCast, a joint probabilistic framework that couples marginal PV forecasts and spatiotemporal dependence through shared cloud evolution. PVCast models this evolution with learned transport–source–sink dynamics and represents its uncertainty with an ensemble of future cloud-field sequences. Within each sequence, local cloud conditions inform marginal power forecasts, while the common evolution induces dependence across systems and forecast steps. Across three real-world PV datasets, PVCast outperforms strong time-series, multimodal, and joint probabilistic baselines in marginal and aggregate forecasting, while better capturing the spatial and temporal dependence of joint power trajectories.

1 INTRODUCTION

As distributed photovoltaic (PV) systems become more widespread, grid operation increasingly relies on forecasts of both individual system outputs and how these outputs change together across locations and future steps (Wang et al., 2024; Traverse et al., 2017). Clouds alter the sunlight reaching PV modules by attenuating direct solar radiation and scattering light, changing their power output (Traverse et al., 2017; Boussif et al., 2023). As clouds evolve, they can affect multiple PV systems, producing correlated power fluctuations (Wang et al., 2024; Yang et al., 2025). Forecasting power across these systems therefore requires accounting for uncertainty in how clouds evolve.

Satellite sequences provide observations of evolving cloud fields. Existing work combines satellite imagery with power and weather data (Ma et al., 2024; Jia et al., 2026), incorporates cloud motion into irradiance or power forecasting (Jing et al., 2024; Wan et al., 2025), or converts predicted solar-radiation fields into system-level power forecasts (Lanzilao & Meyer, 2026). These studies establish the value of cloud information but connect cloud evolution and power uncertainty in different ways. Figure 1 summarizes these approaches as *Direct Forecast* and *Staged Forecast*. The former leaves physical cloud evolution implicit, whereas the latter converts predicted cloud fields into power without explicitly modeling cloud-induced dependence across systems and forecast steps. PVCast introduces a third paradigm, *Coupled Forecast*, in which each uncertain cloud trajectory conditions both the individual power distributions and their spatiotemporal dependence. This unified coupling uses the same physical evolution to represent marginal and joint power uncertainty.

We introduce PVCast, a joint probabilistic forecasting framework for distributed PV systems. In PVCast, a transport–source–sink operator generates future cloud trajectories by representing cloud movement, formation, and dissipation. For each trajectory, a decoder shared across PV systems combines local cloud conditions with historical power, system information, and future-known numerical weather prediction (NWP) to produce conditional power distributions. The same trajectory determines the dependence used to couple these distributions, yielding coherent joint power forecasts. Across three real-world datasets, PVCast consistently improves individual and aggregate probabilistic accuracy and the representation of spatial and temporal power variations.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

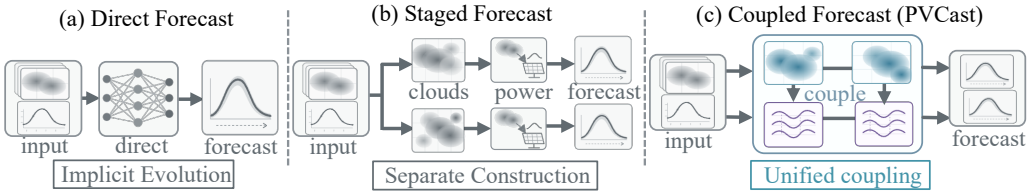

Figure 1: **Comparison of PV forecasting paradigms.** (a) Direct Forecast maps multimodal inputs directly to PV forecasts. (b) Staged Forecast first predicts cloud evolution and subsequently predicts PV power. (c) Coupled Forecast, implemented by PVCast, couples evolving cloud and power states to construct joint PV forecasts across systems and time.

Our contributions are threefold:

- We formulate distributed PV forecasting as a joint predictive problem conditioned on uncertain cloud evolution, providing a shared physical basis for system-level power distributions and their spatiotemporal dependence.
- We develop the Cloud-Driven Joint Forecasting module, combining transport–source–sink cloud evolution and cloud-conditioned spatiotemporal dependence across PV systems to construct joint power trajectories.
- We show on three real-world datasets that PVCast outperforms strong baselines in marginal and aggregate probabilistic forecasting, and in capturing spatial and temporal dependence.

2 Related Work

Multimodal solar forecasting. Weather information used in solar forecasting can be broadly grouped into ground measurements, all-sky imagery, satellite observations, and NWP forecasts (Chu et al., 2024; Nie et al., 2023). CrossViViT and SatSolarCast combine satellite observations with station irradiance histories, and Omnivision additionally incorporates sky imagery (Boussif et al., 2023; Dai et al., 2026; Paletta et al., 2022). For PV power forecasting, FusionSF and Cross-Unet use historical power and NWP with satellite imagery and local meteorological observations, respectively (Ma et al., 2024; Zhang et al., 2026). Recent methods further incorporate optical-flow cloud motion, multilayer cloud structure, or spatiotemporal text prompts (Jing et al., 2024; Wan et al., 2025; Jia et al., 2026) into multimodal irradiance and PV forecasting. These methods improve irradiance or PV forecasts by fusing heterogeneous weather information, but primarily optimize point forecasts or system-level predictive distributions.

Cloud dynamics for solar forecasting. Satellite-based methods forecast regional cloud cover or irradiance from successive observations. Classical approaches estimate cloud motion and advect the latest field (Hammer et al., 1999; Aicardi et al., 2022). Their transport assumption cannot represent local cloud formation and dissipation, motivating neural models that learn complete future fields using ConvLSTM or U-Net architectures (Nielsen et al., 2021; Marchesoni-Acland et al., 2023). Other approaches use solar geometry to extract the cloud regions relevant to individual PV systems, establishing an explicit mapping from regional cloud observations to power forecasts (Si et al., 2021). SolarSTEPS and SHADECast represent cloud-field uncertainty through autoregressive perturbations or conditional diffusion (Carpentieri et al., 2023; 2025). Recent cloud-to-power pipelines propagate these field forecasts through system-specific power models, producing spatially coherent PV scenarios (Lanzilao & Meyer, 2026). In these staged pipelines, cloud-field forecasting and individual power conversion are constructed as separate components.

Multi-site and joint probabilistic forecasting. Distributed PV models use graph structures or cross-site observations to capture dependencies among power series, primarily for deterministic forecasting (Simeunović et al., 2021; Yang et al., 2025). PV-specific probabilistic methods construct spatio-temporal trajectories by coupling marginal forecasts through covariance models, copulas, or historical dependence patterns (Golestaneh et al., 2016; van Der Meer, 2021). General-purpose joint forecasting methods learn multivariate distributions using covariance models, copulas, autoregres-

sive generation, or graph-based diffusion (Salinas et al., 2019; Zheng & Sun, 2024; Drouin et al., 2022; Ashok et al., 2024; Rasul et al., 2021; Wen et al., 2023). These approaches capture dependence from observed output series or forecast errors without explicitly tying it to model future cloud evolution. PVCast uses each uncertain cloud trajectory to condition both system-level power distributions and their spatial and temporal dependence.

3 PVCAST

3.1 FORECASTING TASK AND OVERVIEW

PVCast estimates a joint predictive distribution for N PV systems, as illustrated in Figure 2. Its inputs are historical power $\mathbf{X}_{\text{ts}} : [T_{\text{in}}, N]$, satellite imagery $\mathbf{X}_{\text{sat}} : [T_{\text{in}}, n_{\text{sat}}, H, W]$, future-known NWP covariates $\mathbf{X}_{\text{NWP}} : [T_{\text{out}}, N, n_{\text{NWP}}]$, and system information. Here T_{in} and T_{out} are the input and output window lengths, n_{sat} and n_{NWP} are the numbers of satellite channels and NWP features, and H and W are the spatial dimensions of the satellite fields. Cloud evolution in Figure 2(a) generates S possible future cloud trajectories, stacked as $\mathbf{C}_{\text{traj}} : [S, T_{\text{out}}, H, W]$. For each trajectory, system-specific cloud conditions enter the spatiotemporal dependence component in Figure 2(b) and the shared power decoder in Figure 2(d). The system context in Figure 2(c) combines input variables. Conditional power distributions from the decoder and cloud-conditioned dependence together yield joint power trajectories. The batch dimension is omitted from the task-level notation above. In the following module descriptions, B denotes the batch size.

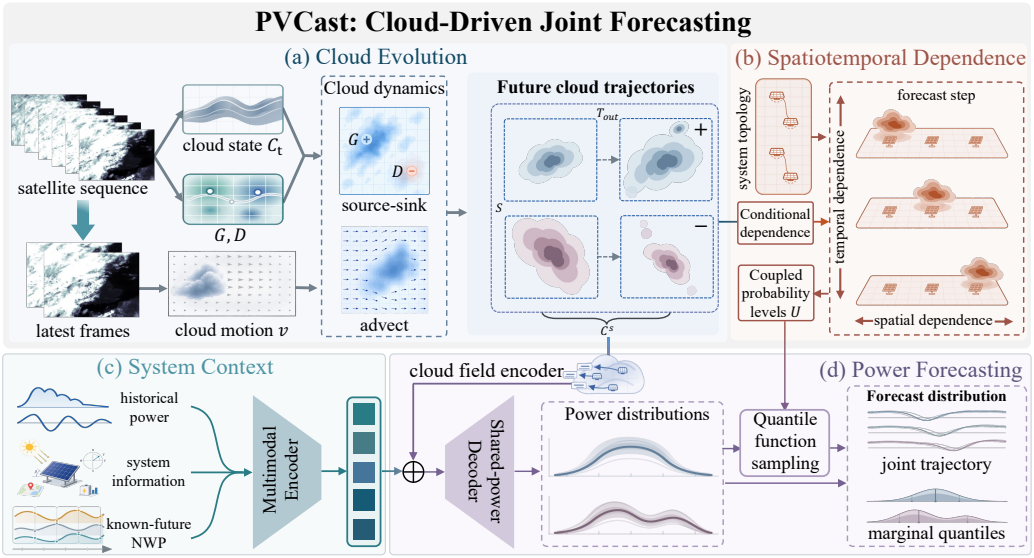

Figure 2: **Overview of PVCast.** (a) Cloud evolution infers transport and source–sink dynamics from satellite sequences to generate future cloud trajectories. (b) Multiscale cloud conditions and system topology determine coupled probability levels $\mathbf{U}$ that represent spatial and temporal dependence. (c) Historical power, system information, and future-known NWP are encoded as system context. (d) The shared power decoder maps system context and local cloud conditions to conditional quantile functions. Composing these functions with the coupled levels $\mathbf{U}$ constructs the joint power distribution.

3.2 SYSTEM CONTEXT

Figure 2(c) consists of a multimodal context encoder and the resulting system-context representations $\mathbf{H}_{\text{sys}}$. Because the inputs have different structures, the encoder assigns a separate encoding pathway to each input. Historical power and its first differences are divided into temporal patches and encoded independently: $\mathbf{Z}_{\text{ts}} = \text{MLP}_{\text{ts}}(\text{Patch}_{L,R}(\mathbf{X}_{\text{ts}}))$ and $\mathbf{Z}_{\Delta} = \text{MLP}_{\Delta}(\text{Patch}_{L,R}(\Delta\mathbf{X}_{\text{ts}}))$, where L and R are the patch length and stride. Each patch contains L consecutive input steps,

and adjacent patches start R steps apart. This produces $J_{\mathrm{p}} = \left\lfloor \frac{T_{\mathrm{in}}-L}{R} \right\rfloor + 1$ patches and representations $\mathbf{Z}_r \in \mathbb{R}^{B\times J_{\mathrm{p}}\times N\times n_{\mathrm{hid}}}$. The NWP covariates are projected into the hidden space $\mathbf{Z}_{\mathrm{NWP}} = \mathrm{Linear}(\mathbf{X}_{\mathrm{NWP}})$. Then, the NWP representations act as queries that retrieve the relevant historical information from each power branch:

$$\mathbf{H}_m = \mathrm{CrossAttn}(Q = \mathbf{Z}_{\mathrm{NWP}}, K = \mathbf{Z}_m, V = \mathbf{Z}_m)\,, \qquad m \in \{\mathrm{ts}, \Delta\mathrm{ts}\}. \tag{1}$$

System information is incorporated into the attended features, and a learned gate fuses the power and power-difference representations into $\mathbf{H}_{\mathrm{sys}} \in \mathbb{R}^{B\times T_{\mathrm{out}}\times N\times n_{\mathrm{hid}}}$. The details of the system context are in Appendix A.1.

3.3 Cloud Evolution

Consecutive cloud fields change through both the transport of existing cloud structures and their local formation or dissipation. PVCast separates these processes so that motion propagates coherent cloud structures, while source–sink dynamics represent local changes that cannot be explained by displacement alone (Tiedtke, 1993). Given the satellite history $\mathbf{X}_{\mathrm{sat}}$, Figure 2(a) estimates cloud motion and source–sink dynamics before combining them to generate future cloud fields.

Cloud-dynamics estimation. Following the Heliosat procedure, the visible satellite channel is converted to a cloud-index sequence and normalized to $[0, 1]$ to define the cloud state (Cano et al., 1986; Rigollier et al., 2004). Cloud indices outside daylight hours are set to zero. The final two causal fields are denoted by $\mathbf{C}_{-1}, \mathbf{C}_0$, with $\mathbf{C}_0$ initializing cloud forecasting. PyramidFlow estimates the cloud motion $\mathbf{v} = \mathrm{PyramidFlow}(\mathbf{C}_{-1}, \mathbf{C}_0) \in \mathbb{R}^{B\times 2\times H\times W}$, following pyramidal motion estimation with feature warping and cost-volume matching (Sun et al., 2018).

In parallel, the complete satellite history is encoded as $\mathbf{Z}_{\mathrm{sat}} = \mathrm{GRU}(\mathbf{X}_{\mathrm{sat}})$. The satellite representation, current cloud field, and multiscale Laplacian responses are combined to estimate the source and sink fields:

$$[\mathbf{G}, \mathbf{D}] = \mathrm{Softplus}\big(\mathrm{Conv}_{K_{\mathrm{conv}}\times K_{\mathrm{conv}}}\big(\mathrm{Concat}_{\mathrm{ch}}\big[\mathbf{Z}_{\mathrm{sat}}, \mathbf{C}_0, \nabla_{\mathrm{f}}^2\mathbf{C}_0, \nabla_{\mathrm{c}}^2\mathbf{C}_0\big]\big)\big)\,, \tag{2}$$

where ∇_{f}^2 and ∇_{c}^2 are fine- and coarse-scale Laplacian operators, $\mathrm{Concat}_{\mathrm{ch}}$ denotes channel-wise concatenation, and K_{conv} is the convolutional kernel width. Softplus produces the nonnegative source and sink fields $\mathbf{G}, \mathbf{D}$, representing local cloud formation and dissipation, respectively.

Transport–source–sink evolution. The predicted source and sink fields determine how the transported cloud state approaches a local equilibrium:

$$\mathbf{C}^{+} = \underbrace{\left[1 - \beta\left(1 - e^{-\mathbf{r}}\right)\right] \odot \mathcal{A}_{\mathbf{v}}(\mathbf{C})}_{\text{retained transported cloud}} + \underbrace{\beta\left(1 - e^{-\mathbf{r}}\right) \odot \mathbf{C}^{\mathrm{eq}}}_{\text{source–sink contribution}}, \tag{3}$$

where $\mathbf{r} = \mathbf{G} + \mathbf{D}$ and $\mathbf{C}^{\mathrm{eq}} = \mathbf{G} \oslash \mathbf{r}$ are the total source–sink intensity and equilibrium cloud state, respectively. The operator $\mathcal{A}_{\mathbf{v}}$ denotes semi-Lagrangian transport under the cloud-displacement field $\mathbf{v}$ (Staniforth & Côté, 1991). The learned scalar gate $\beta = \sigma(\beta_{\mathrm{logit}}) \in [0, 1]$ controls the contribution of the source–sink transition: $\beta = 0$ retains pure transport, whereas $\beta = 1$ recovers the complete update. The operators $\odot$ and $\oslash$ denote elementwise multiplication and division, with $0/0 := 0$. Applying this transition over the forecast window produces $\widehat{\mathbf{C}} \in [0, 1]^{B\times T_{\mathrm{out}}\times H\times W}$.

Cloud trajectory generation. The base cloud dynamics are expanded into S trajectories by perturbing their motion and source–sink parameters. With the base fields implicitly shared across trajectories and forecast steps, the perturbed motion is $\mathbf{v}_{\mathrm{traj}} = \mathbf{v} + \Delta\mathbf{v}$. The equilibrium cloud fields are $\mathbf{C}^{\mathrm{eq}}_{\mathrm{traj}} = \sigma\Big(\log \frac{\mathbf{C}^{\mathrm{eq}}}{1-\mathbf{C}^{\mathrm{eq}}} + \boldsymbol{\epsilon}_{\mathrm{eq}}\Big)$, where σ denotes the sigmoid function. The perturbed source–sink intensity is $\boldsymbol{\eta}_{\mathrm{traj}} = \exp(\log(\mathbf{G} + \mathbf{D}) + \boldsymbol{\epsilon}_{\eta})$, from which $\mathbf{G}_{\mathrm{traj}} = \boldsymbol{\eta}_{\mathrm{traj}} \odot \mathbf{C}^{\mathrm{eq}}_{\mathrm{traj}}$ and $\mathbf{D}_{\mathrm{traj}} = \boldsymbol{\eta}_{\mathrm{traj}} \odot (1 - \mathbf{C}^{\mathrm{eq}}_{\mathrm{traj}})$. The motion perturbation $\Delta\mathbf{V}$ is spatially smooth, while $\boldsymbol{\epsilon}_{\mathrm{eq}}$ and $\boldsymbol{\epsilon}_{\eta}$ combine learned spatial modes with temporally correlated Gaussian coefficients. Applying Equation 3 to the S perturbed parameter sets produces $\mathbf{C}_{\mathrm{traj}} \in [0, 1]^{B\times S\times T_{\mathrm{out}}\times H\times W}$.

3.4 CLOUD-CONDITIONED SPATIOTEMPORAL DEPENDENCE

Each cloud trajectory describes a future cloud field over the entire region, while dependence modeling requires cloud conditions aligned with individual PV systems. The dependence component applies topology-adaptive multiscale filters to each cloud trajectory and extracts the resulting cloud contrasts at the PV-system locations. These representations determine the spatial factor loadings and temporal correlations used to generate coupled probability levels.

Multiscale cloud dependence. Each cloud trajectory is smoothed at topology-adaptive scales $\mathcal{R}_{\mathrm{sm}} = \mathrm{Quantile}_{\boldsymbol{\tau}_{\mathrm{dist}}}(\mathcal{D}(\mathcal{X}))$, where $\mathcal{D}(\mathcal{X})$ contains pairwise system distances after mapping the system locations to the satellite grid, and $\boldsymbol{\tau}_{\mathrm{dist}}$ specifies the selected distance quantiles. The elements of $\mathcal{R}_{\mathrm{sm}}$ are discretized smoothing radii measured in satellite-grid cells. The smoothed fields are then evaluated at the PV-system locations and standardized across systems, producing the multiscale cloud deviations $\boldsymbol{\Psi}$. A constant regional component and the multiscale deviations define the unconstrained factor loadings $\widetilde{\mathbf{A}} = [\mathbf{1}, \boldsymbol{\Psi}]\,\mathrm{diag}(\mathbf{s}_{\mathrm{fac}})$, where $\mathbf{1}$ assigns the same regional loading to all systems, and $\mathbf{s}_{\mathrm{fac}}$ contains the calibrated amplitudes assigned to the regional component and each multiscale cloud deviation. The loadings are rescaled as $\mathbf{A} = \frac{\widetilde{\mathbf{A}}}{\max\left(1,\, \|\widetilde{\mathbf{A}}\|_{2,\mathrm{fac}}/a_{\max}\right)}$, where the norm is taken over the factor dimension. This ensures $\|\mathbf{A}\|_{2,\mathrm{fac}} \le a_{\max} < 1$, retaining a nonzero system-specific component in Equation 5. The spatial scales, factor amplitudes, and norm bound are specified in Appendix A.3.

Coupled probability levels. For each cloud trajectory, shared Gaussian factors $\boldsymbol{\zeta}$ represent cloud-related variation affecting multiple systems, while independent system-specific factors $\mathbf{e}$ represent the remaining local uncertainty. Both evolve over the forecast window through the unit-variance AR(1) recursion

$$\begin{aligned} \boldsymbol{\zeta}^{+} &= \boldsymbol{\rho}_{\mathrm{shared}} \odot \boldsymbol{\zeta} + \sqrt{1 - \boldsymbol{\rho}_{\mathrm{shared}}^{2}} \odot \boldsymbol{\epsilon}_{\mathrm{shared}}, \\ \mathbf{e}^{+} &= \boldsymbol{\rho}_{\mathrm{local}} \odot \mathbf{e} + \sqrt{1 - \boldsymbol{\rho}_{\mathrm{local}}^{2}} \odot \boldsymbol{\epsilon}_{\mathrm{local}}, \end{aligned} \tag{4}$$

where $(\cdot)^{+}$ denotes the next forecast step. The cloud-conditioned loadings $\mathbf{A}$ combine the shared and system-specific factors as

$$\mathbf{Z} = \sqrt{1 - \|\mathbf{A}\|_2^2} \odot \mathbf{e} + \langle \mathbf{A}, \boldsymbol{\zeta} \rangle, \qquad \mathbf{U} = \Phi(\mathbf{Z}), \tag{5}$$

where the norm and inner product are taken over the factor dimension. Since the shared and system-specific factors are independent standard normal and $\|\mathbf{A}\|_2 < 1$, each element of $\mathbf{Z}$ is marginally standard normal. Consequently, $\mathbf{U} = \Phi(\mathbf{Z})$ is marginally uniform. Shared factors and their AR(1) evolution determine spatial and temporal dependence without changing the conditional power distributions. Using R realizations for each cloud trajectory produces $\mathbf{U} \in (0,1)^{B \times S \times R \times T_{\mathrm{out}} \times N}$.

3.5 SHARED POWER DECODER AND JOINT FORECASTING

Conditional power distributions. The cloud trajectories generated above describe complete future cloud evolutions. For each trajectory, the cloud field encoder separates the motion-only evolution $\mathbf{C}_{\mathrm{adv}}$ from the local source–sink contribution $\Delta\mathbf{C}_{\mathrm{react}} = \mathbf{C}_{\mathrm{traj}} - \mathbf{C}_{\mathrm{adv}}$. The transport branch encodes changes in $\mathbf{C}_{\mathrm{adv}}$, cloud-boundary information, and cloud conditions along the transport path. The process branch encodes $\mathbf{C}_{\mathrm{traj}}$ and $\Delta\mathbf{C}_{\mathrm{react}}$. These fields are sampled bilinearly at the PV-system locations and augmented with temporal features and convolutional field observations. Learned gates control the contributions of the transport and process representations. Their gated outputs are concatenated along the feature dimension, producing the system-specific cloud representation. The shared power decoder uses the system context to modulate the cloud representation, producing a cloud-conditioned residual $\boldsymbol{\varphi}$. The residual adjusts the context-based quantile prediction as

$$\mathbf{Q}_{\mathrm{cond}} = \mathrm{sort}_{n_{\mathrm{q}}}(\mathbf{Q}_{\mathrm{ctx}}(\mathbf{H}_{\mathrm{sys}}) + \mathbf{Q}_{\mathrm{cloud}}(\boldsymbol{\varphi}))\,. \tag{6}$$

After broadcasting the system context over the cloud trajectories, $\boldsymbol{\varphi}$ has shape $[B, S, T_{\mathrm{out}}, N, n_{\mathrm{hid}}]$, and $\mathbf{Q}_{\mathrm{cond}}$ has shape $[B, S, T_{\mathrm{out}}, N, n_{\mathrm{q}}]$. The last dimension corresponds to the fixed probability levels $\boldsymbol{\tau} : [n_{\mathrm{q}}]$.

Table 1: Marginal and joint probabilistic forecasting performance on three PV datasets. Throughout all tables, ↓ and ↑ indicate that lower and higher values are better, respectively. Cell colors mark the first , second , and third best scores.

Model	UKPV			MMSP			PVOD		
	CRPS ↓	CRPS-sum ↓	ES ↓	CRPS ↓	CRPS-sum ↓	ES ↓	CRPS ↓	CRPS-sum ↓	ES ↓
Smart Persistence	0.0721	0.0317	0.1696	0.1629	0.1184	0.4152	0.0976	0.0717	0.2150
DLinear	0.0456	0.0275	0.1829	0.1096	0.1016	0.4574	0.0572	0.0336	0.2784
PatchTST	0.0430	0.0204	0.1510	0.1011	0.0805	0.3738	0.0557	0.0312	0.2069
TFT	0.0411	0.0176	0.1648	0.0672	0.0380	0.3218	0.0536	0.0318	0.2182
MTGNN	0.0428	0.0165	0.1589	0.1005	0.0896	0.3895	0.0530	0.0314	0.2074
Cross-Unet	0.0463	0.0262	0.1755	0.1017	0.0868	0.4204	0.0541	0.0349	0.2236
Correlated-Errors	0.1760	0.1656	0.3029	0.2102	0.1841	0.4508	0.1475	0.1288	0.3407
TACTiS-2	0.0573	0.0263	0.1925	0.1447	0.1148	0.5017	0.0588	0.0307	0.2071
TimeGrad	0.0576	0.0332	0.1611	0.1490	0.1127	0.4367	0.0967	0.0601	0.2302
DiffSTG	0.0817	0.0617	0.1590	0.2496	0.2377	0.4620	0.0810	0.0504	0.1929
SHADECast-PV	0.0427	0.0219	0.1747	0.0710	0.0447	0.3556	0.0556	0.0325	0.2122
FusionSF	0.0609	0.0418	0.2139	0.0666	0.0395	0.3095	0.0497	0.0301	0.2022
PVCast	0.0394	0.0119	0.1294	0.0660	0.0333	0.2994	0.0486	0.0248	0.1692

Joint power distribution. Evaluating the conditional quantile functions at the coupled probability levels produces $\widehat{\mathbf{P}}_{\text{joint}} = \mathcal{Q}_{\text{cond}}(\mathbf{U}) \in \mathbb{R}^{B\times S\times R\times T_{\text{out}}\times N}$. Each (s, r) slice forms one joint power trajectory, and their collection defines the empirical joint predictive distribution. The conditional CDF $F^{(s)}$ associated with cloud trajectory s is marginalized over cloud uncertainty as

$$F_{\text{mix}} = \frac{1}{S}\sum_{s=1}^{S} F^{(s)}, \qquad \mathbf{Q}_{\text{marg}} = F_{\text{mix}}^{-1}(\boldsymbol{\tau}). \tag{7}$$

PVCast is estimated in two stages. First, the cloud evolution and conditional power distributions are learned using $\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \left[\mathcal{L}_{\text{pin}}(\mathbf{Q}_{\text{marg}}, \mathbf{Y}) + \lambda_C \mathcal{L}_{\text{cloud}}\right]$, where the pinball loss $\mathcal{L}_{\text{pin}}$ is averaged over valid power targets and λ_C is adapted using GECO (Rezende & Viola, 2018). The cloud loss supervises one-step transport and forecast-window cloud evolution relative to holding the latest observed cloud field fixed. The dependence parameters $\boldsymbol{\phi}$ are fitted using complementary scores for spatial aggregates, temporal power changes (ramps), and complete multistep trajectories: $\boldsymbol{\phi}^* = \arg\min_{\boldsymbol{\phi}} \left[\lambda_{\text{agg}}\mathcal{L}_{\text{agg}} + \lambda_{\text{ramp}}\mathcal{L}_{\text{ramp}} + \lambda_{\text{traj}}\mathcal{L}_{\text{traj}}\right]$, where the three terms score spatial aggregates, temporal power changes, and complete trajectories constructed from $\mathcal{Q}_{\text{cond}}(\mathbf{U}_{\boldsymbol{\phi}})$.

4 Experiments

Data and Experimental Setup. We evaluate on three real-world PV datasets. **UKPV** is the PV generation dataset for Great Britain, constructed from three public sources: half-hourly generation from 841 distributed PV systems (Open Climate Fix, 2025), EUMETSAT SEVIRI imagery from the Meteosat Second Generation satellites with a native 176×260 resolution (EUMETSAT), and future-known NWP from DWD ICON-EU (Open Climate Fix, 2023). Following Fulton et al. (2024), we forecast up to 8 hours ahead. As their study does not specify the historical input window, we use the preceding 24 hours for all three datasets to maintain a consistent input duration. **MMSP** contains hourly power, satellite, and NWP data from 88 PV systems in Shandong, China (Ma et al., 2024). Following the experimental setup of Ma et al. (2024), we use a 24-hour historical input window and a 24-hour prediction window, with 64×64 satellite fields. **PVOD** provides one year of 15-minute power, system metadata, and corresponding NWP for 10 PV systems in China (Yao et al., 2021). We add satellite observations from JMA Himawari-8/9, processed to 75×150 fields and available through NOAA (Japan Meteorological Agency (JMA), 2024). Following the protocol of Cross-Unet (Zhang et al., 2026), we use 24 hours of history to predict the following 4 hours. More details of the dataset and experimental setting can be found in Appendix C.

Baseline Models. We compare PVCast against five groups of baselines. Smart Persistence provides a learning-free reference (Pedro & Coimbra, 2012). General time-series models include DLinear (Zeng et al., 2023), PatchTST (Nie et al., 2022), and TFT (Lim et al., 2021). MTGNN provides a graph-based multivariate forecasting baseline (Wu et al., 2020). General joint probabilistic models include Correlated Errors (Zheng & Sun, 2024), TACTiS-2 (Ashok et al., 2024), TimeGrad (Rasul et al., 2021), and DiffSTG (Wen et al., 2023). Solar-specific models include Cross-Unet, which combines historical PV records with future weather forecasts (Zhang et al., 2026); FusionSF, which fuses PV power, satellite imagery, and NWP (Ma et al., 2024); and SHADECast–PV, which converts probabilistic irradiance forecasts into PV power forecasts (Carpentieri et al., 2025; Lanzilao & Meyer, 2026). All neural baselines use the same NWP forecasts available at the forecast origin, except PatchTST and DiffSTG, which retain their original PV-only inputs to preserve their architectures.

Evaluation Metrics. We report three complementary metrics. The Continuous Ranked Probability Score (CRPS) evaluates each system's power distribution at each forecast step (Matheson & Winkler, 1976). CRPS-sum evaluates the predictive distribution of aggregate PV power, an input to net-load forecasts used in grid operations (Lin et al., 2025). The distribution of aggregate power is sensitive to how system outputs co-vary, making the metric sensitive to spatial dependence (Salinas et al., 2019). The Energy Score (ES) compares sampled power-change sequences with the observed sequence and accounts for differences among samples. By evaluating each sequence over the full forecast horizon, it assesses the joint distribution of successive changes and their temporal dependence (Gneiting & Raftery, 2007). Evaluating changes also reduces the influence of the diurnal cycle. Definitions are given in Appendix D.1.

4.1 Results

Overall probabilistic forecasting performance. Table 1 shows that PVCast achieves the best forecasting performance across all metrics and datasets. Relative to the second-best result in each column, PVCast reduces CRPS, CRPS-sum, and ES by an average of 2.42%, 19.29%, and 9.95%, respectively. The smaller CRPS improvement reflects its system-wise evaluation, which is less sensitive to joint forecast structure. The larger CRPS-sum and ES reductions demonstrate clearer improvements in spatially aggregated forecasts and the temporal organization of power changes. Averaged across the three datasets, PVCast reduces these metrics by 12.81%, 34.95%, and 19.70% over FusionSF, which uses the same three input modalities, and by 9.12%, 31.62%, and 20.67% over SHADECast–PV, which also models future cloud evolution.

Table 2: **Cloud-evolution ablations.** Cloud skill is measured relative to persistence. CRPS gains use persistence as the reference, while CRPS-sum and Ramp CRPS gains use the no-cloud model. Higher values are better.

Ablation	Cloud skill ↑	CRPS gain ↑	CRPS-sum gain ↑	Ramp gain ↑
Persistence	0.0000	0.00%	+1.14%	+6.37%
No cloud	–	+0.25%	0.00%	0.00%
Advection only	+0.0060	+0.61%	+11.40%	+14.00%
Source–sink only	+0.6496	+1.20%	+14.06%	+11.62%
Shuffled dynamics	+0.6599	+0.51%	+12.71%	+12.28%
Full model	+0.6603	+1.36%	+20.12%	+17.14%

Ablation of cloud evolution We conduct ablation experiments on UKPV, a large-scale and spatially distributed PV dataset. Table 2 evaluates cloud-evolution components. Ramp CRPS evaluates within-system power changes and is defined in Appendix D.1. The full model achieves the best cloud skill and power forecasting performance. Source–sink evolution accounts for most of the cloud-field improvement, while advection alone still improves CRPS-sum and Ramp CRPS by 11.40% and 14.00%. Shuffling the dynamics preserves similar cloud skill but weakens all power metrics, showing that accurate and correctly aligned cloud evolution is important for downstream forecasting. Combining both components yields gains of 20.12% in CRPS-sum and 17.14% in Ramp CRPS.

Performance across cloud regimes. Forecast origins are stratified by the recent cloud-field change $V = \frac{1}{HW} \|\mathbf{C}_0 - \mathbf{C}_{-1}\|_1$, where $\mathbf{C}_{-1}$ and $\mathbf{C}_0$ are the final two observed cloud fields. We stratify test origins into stable, moderate, and variable regimes using the lower and upper tertiles of the recent cloud-field variation V, estimated on the validation set. Figure 3 shows that

cloud-field skill remains high across all regimes, decreasing only moderately from 67.35% under stable conditions to 63.46% under variable conditions. The downstream power benefit exhibits the opposite pattern. Relative to the no-cloud model, the CRPS-sum reduction increases from 13.65% to 21.90% and 31.98%, while the aggregate ramp CRPS reduction increases from 12.97% to 18.47% and 23.04%. Cloud-field skill remains broadly stable, while the value of cloud evolution for predicting aggregate power and its changes grow as recent cloud variability increases. This distinction reflects the different references: cloud skill is measured against persistence, whereas the power improvements are measured against a model without satellite information.

Ablation of dependence Table 3 evaluates joint dependence by permuting the same predicted trajectories. Relative to independent sampling, the full model combines gains across all joint metrics. Spatial dependence provides an 18.75% reduction in CRPS-sum, while temporal dependence reduces Ramp CRPS and ES by 10.77% and 11.43%, respectively. Their negligible effects on the complementary metrics confirm that the two components capture distinct and complementary aspects of the joint forecast.

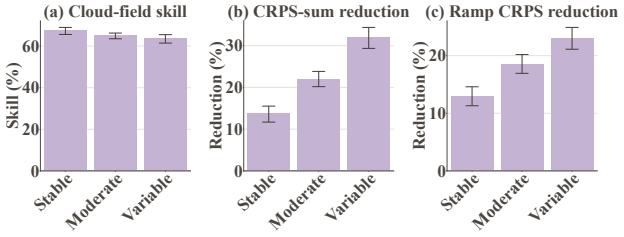

Figure 3: **Cloud-regime analysis on UKPV.** (a) Cloud-field skill over persistence. (b) CRPS-sum reduction over the no-cloud model. (c) Aggregate Ramp CRPS reduction over the no-cloud model. Error bars are 95% bootstrap percentile intervals over forecast origins. Positive values indicate an improvement over the baseline.

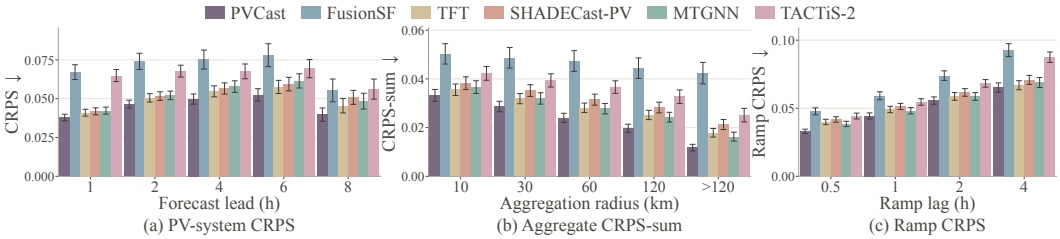

Figure 4: **Forecast performance across spatiotemporal scales.** (a) PV-system CRPS across forecast leads. (b) Aggregate CRPS-sum across aggregation radii. (c) Ramp CRPS across ramp lags. Bars report mean scores, and error bars show 95% bootstrap percentile intervals. Lower values are better.

Forecast performance across spatiotemporal scales. Figure 4 examines UKPV forecasting performance across prediction horizons, spatial aggregation scales, and temporal differences. Panel (a) reports PV-system CRPS at each forecast lead, averaged over systems and forecast origins, and therefore evaluates the marginal forecasts as the horizon increases. Panel (b) evaluates spatial aggregation. For each PV system serving as a spatial anchor, systems within radius are combined using their capacity weights, and CRPS-sum evaluates the resulting aggregate-power distribution. Scores are averaged over all PV systems used as spatial anchors. Panel (c) evaluates temporal changes by applying Ramp CRPS to power differences separated by lag, with scores averaged over systems and valid forecast-step pairs. The analysis uses 256 test origins sampled uniformly and fixed before evaluation, with origins weighted equally. We select representative competitive baseline, and PVCast achieves lower PV-system CRPS throughout the forecast window, showing that its marginal advantage is maintained as the forecast horizon increases. The separation becomes

Table 3: **Cloud-conditioned dependence ablations.** Values are relative error reductions over independent sampling. Higher values are better; negative values indicate degradation.

Configuration	CRPS-sum	Ramp CRPS	ES
Independent	0.000%	0.000%	0.000%
No spatial dependence	−0.002%	+10.766%	+11.428%
No temporal dependence	+18.749%	+0.003%	−0.001%
Full	**+18.749%**	**+10.766%**	**+11.428%**

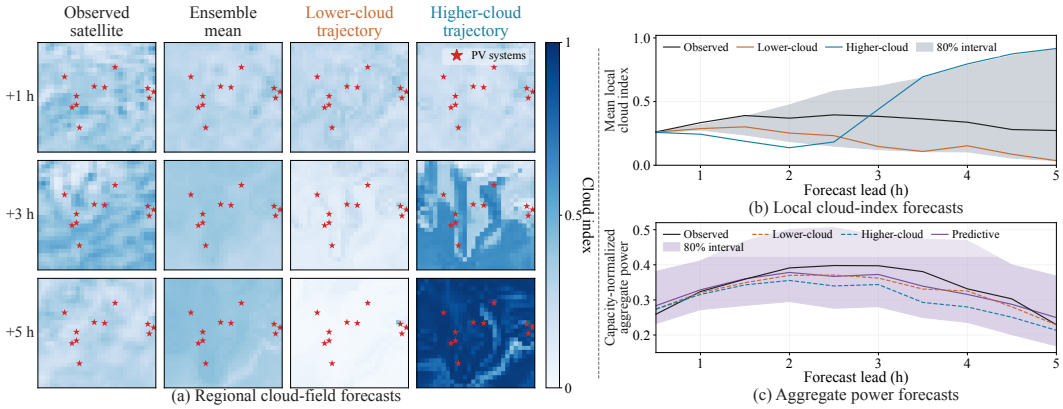

Figure 5: **Cloud-to-power uncertainty propagation over a local UKPV region.** (a) Observed cloud fields, the ensemble mean, and lower- and higher-cloud trajectories at three forecast leads. Red markers denote the selected PV systems. (b) Mean local cloud index and its 80% interval across cloud trajectories. (c) Conditional aggregate-power forecasts under the two trajectories and the final predictive distribution.

larger for spatially aggregated power in panel (b). Its CRPS-sum advantage persists as the radius expands from local neighborhoods to the complete network, indicating that the generated trajectories retain the cross-system variation needed to represent aggregate uncertainty at different spatial scales. In panel (c), PVCast maintains lower Ramp CRPS for lags from 0.5 to 4 hours, showing that the trajectories also represent power changes over both short and extended intervals.

Cloud-to-power uncertainty propagation. To examine how PVCast propagates cloud-evolution uncertainty into power forecasts, Figure 5 visualizes an evolving cloud event over 11 UKPV systems within 60 km of a spatial anchor. Local cloud indices are computed from 5×5 grid-cell neighborhoods and aggregated using installed-capacity weights. The lower- and higher-cloud examples are the trajectories with the smallest and largest mean regional cloud indices. Panel (a) compares the observed cloud fields, the ensemble mean, and lower- and higher-cloud trajectories at different forecast horizons. The trajectories remain similar during the first forecast hour but subsequently diverge over the 60-km neighborhood. The lower-cloud trajectory becomes progressively clearer, whereas the higher-cloud trajectory brings increasingly dense cloud cover. Panel (b) summarizes these spatial fields by averaging the local cloud index over the systems. This converts the field-level divergence in panel (a) into the cloud conditions directly relevant to their power forecasts. The two trajectories (lower-cloud and higher-cloud) separate after approximately two hours and span contrasting parts of the cloud-trajectory distribution, while the observed local cloud index remains within the pointwise 80% interval. Panel (c) shows how these cloud conditions affect the capacity-normalized aggregate power of the same systems. Integrating these cloud trajectories with conditional power uncertainty therefore yields an accurate aggregate-power distribution.

5 Conclusion

We presented PVCast, a joint probabilistic forecasting framework that treats uncertain cloud evolution as a shared physical driver of distributed PV generation. PVCast combines transport–source–sink cloud evolution with conditional power distributions and cloud-conditioned spatiotemporal dependence to construct coherent power trajectories. Across three real-world datasets, PVCast consistently improves marginal and aggregate probabilistic accuracy while better representing spatial aggregation and temporal power changes. These results demonstrate the effectiveness of a physics-guided forecasting paradigm, where uncertainty in a shared environmental process provides a common basis for both marginal forecasts and their dependence.

AI USE STATEMENT

In this work, we used generative AI tools solely for language polishing (grammar, spelling, and style improvement) of the manuscript. We have not used generative AI tools for generating research ideas. We have carefully reviewed all AI-assisted edits to ensure that the meaning and accuracy of the text are preserved. We take full responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The paper does not involve human subjects, private data, or potentially harmful applications. Datasets are public or used under their licenses.

REPRODUCIBILITY STATEMENT

We provide the task definition, mathematical formulation, and complete description of PVCast in Section 3, with additional architectural details in Appendix A. Cloud preprocessing, dataset alignment, chronological splits, and forecast settings are documented in Appendices B and C and summarized in Table 6. Tables 7 report the model, optimization, sampling, and baseline-specific hyperparameters. The probabilistic evaluation metrics and aggregation procedures are defined in Appendix D. All datasets are publicly available or used under their respective licenses, with their sources provided in the references.

REFERENCES

Daniel Aicardi, Pablo Musé, and Rodrigo Alonso-Suárez. A comparison of satellite cloud motion vectors techniques to forecast intra-day hourly solar global horizontal irradiation. *Solar Energy*, 233:46–60, 2022.

Arjun Ashok, Étienne Marcotte, Valentina Zantedeschi, Nicolas Chapados, and Alexandre Drouin. Tactis-2: Better, faster, simpler attentional copulas for multivariate time series. In *International conference on learning representations*, volume 2024, pp. 22897–22924, 2024.

Oussama Boussif, Ghait Boukachab, Dan Assouline, Stefano Massaroli, Tianle Yuan, Loubna Benabbou, and Yoshua Bengio. Improving day-ahead solar irradiance time series forecasting by leveraging spatio-temporal context. *Advances in Neural Information Processing Systems*, 36: 2342–2367, 2023.

Daniel Cano, Jean-Marie Monget, Michel Albuisson, Hervé Guillard, Nathalie Regas, and Lucien Wald. A method for the determination of the global solar radiation from meteorological satellite data. *Solar energy*, 37(1):31–39, 1986.

Alberto Carpentieri, Doris Folini, David Nerini, Seppo Pulkkinen, Martin Wild, and Angela Meyer. Intraday probabilistic forecasts of surface solar radiation with cloud scale-dependent autoregressive advection. *Applied Energy*, 351:121775, 2023. doi: 10.1016/j.apenergy.2023.121775.

Alberto Carpentieri, Doris Folini, Jussi Leinonen, and Angela Meyer. Extending intraday solar forecast horizons with deep generative models. *Applied Energy*, 377:124186, 2025.

Yinghao Chu, Yiling Wang, Dazhi Yang, Shanlin Chen, and Mengying Li. A review of distributed solar forecasting with remote sensing and deep learning. *Renewable and Sustainable Energy Reviews*, 198:114391, 2024.

Kuai Dai, Hui Su, Chengxing Zhai, Huiwei Lin, and Mingliang Bai. Satsolarcast: A flexible framework for multimodal solar irradiance forecasting via memory-alignment learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 38331–38339, 2026.

Alexandre Drouin, Étienne Marcotte, and Nicolas Chapados. Tactis: Transformer-attentional copulas for time series. In *International Conference on Machine Learning*, pp. 5447–5493. PMLR, 2022.

EUMETSAT. Rapid scan high rate SEVIRI level 1.5 image data—MSG. EUMETSAT User Portal. URL https://user.eumetsat.int/catalogue/EO:EUM:DAT:MSG:MSG15-RSS. Collection reference: EO:EUM:DAT:MSG:MSG15-RSS.

D. James Fulton, Jacob Bieker, Peter Dudfield, Solomon Cotton, Zakari Watts, and Jack Kelly. Forecasting regional pv power in great britain with a multi-modal late fusion network. In *Tackling Climate Change with Machine Learning Workshop at ICLR*, 2024. URL https://www.climatechange.ai/papers/iclr2024/46.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Faranak Golestaneh, Hoay Beng Gooi, and Pierre Pinson. Generation and evaluation of space–time trajectories of photovoltaic power. *Applied Energy*, 176:80–91, 2016.

Annette Hammer, Detlev Heinemann, Elke Lorenz, and Bertram Lückehe. Short-term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy*, 67(1-3):139–150, 1999.

Japan Meteorological Agency (JMA). JMA Himawari-8/9. Registry of Open Data on AWS, 2024. URL https://registry.opendata.aws/noaa-himawari/. Accessed: 2026-09-16.

Pengfei Jia, Jianghong Ma, Baoquan Zhang, Kenghong Lin, Xinyu Zhang, Chuyao Luo, Xutao Li, and Yunming Ye. Satellite-text-prompted large language model for photovoltaic power forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 38644–38652, 2026.

Tao Jing, Shanlin Chen, David Navarro-Alarcon, Yinghao Chu, and Mengying Li. Solarfusionnet: Enhanced solar irradiance forecasting via automated multi-modal feature selection and cross-modal fusion. *IEEE Transactions on Sustainable Energy*, 16(2):761–773, 2024.

Luca Lanzilao and Angela Meyer. Intraday spatiotemporal pv power prediction at national scale using satellite-based solar forecast models. *Energy and AI*, pp. 100786, 2026.

Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 37(4):1748–1764, 2021.

F. Lin, Y. Zhang, H. Zhao, W. Huo, and J. Wang. Short-term probabilistic forecasting for regional PV power based on convolutional graph neural network and parameter transferring. *IEEE Transactions on Power Systems*, 40(3):2724–2736, May 2025. doi: 10.1109/TPWRS.2024.3503288.

Ziqing Ma, Wenwei Wang, Tian Zhou, Chao Chen, Bingqing Peng, Liang Sun, and Rong Jin. FusionSF: Fuse heterogeneous modalities in a vector quantized framework for robust solar power forecasting. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5532–5543. Association for Computing Machinery, 2024. doi: 10.1145/3637528.3671509. URL https://doi.org/10.1145/3637528.3671509.

Franco Marchesoni-Acland, Andrés Herrera, Franco Mozo, Ignacio Camiruaga, Alberto Castro, and Rodrigo Alonso-Suárez. Deep learning methods for intra-day cloudiness prediction using geostationary satellite images in a solar forecasting framework. *Solar Energy*, 262:111820, 2023.

James E Matheson and Robert L Winkler. Scoring rules for continuous probability distributions. *Management science*, 22(10):1087–1096, 1976.

Sean Nassimiha, Peter Dudfield, Jack Kelly, Marc Peter Deisenroth, and So Takao. Short-term prediction and filtering of solar power using state-space gaussian processes. *arXiv preprint arXiv:2302.00388*, 2023. URL https://arxiv.org/abs/2302.00388.

Yuhao Nie, Xiatong Li, Andea Scott, Yuchi Sun, Vignesh Venugopal, and Adam Brandt. Skipp'd: A sky images and photovoltaic power generation dataset for short-term solar forecasting. *Solar Energy*, 255:171–179, 2023.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Andreas Højlund Nielsen, Alexandros Iosifidis, and Henrik Karstoft. IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. *Solar Energy*, 228:659–669, 2021. doi: 10.1016/j.solener.2021.09.073.

Open Climate Fix. DWD ICON-EU Forecast, 2023. URL `https://huggingface.co/datasets/openclimatefix/dwd-icon-eu`. Accessed: 2026-09-14.

Open Climate Fix. UK PV Dataset, 2025. URL `https://huggingface.co/datasets/openclimatefix/uk_pv`. Accessed: 2026-09-14.

Quentin Paletta, Guillaume Arbod, and Joan Lasenby. Omnivision forecasting: combining satellite observations with sky images for improved intra-hour solar energy predictions. *arXiv preprint arXiv:2206.03207*, 2022.

Hugo TC Pedro and Carlos FM Coimbra. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7):2017–2028, 2012.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International conference on machine learning*, pp. 8857–8868. PMLR, 2021.

Danilo Jimenez Rezende and Fabio Viola. Taming vaes. *arXiv preprint arXiv:1810.00597*, 2018.

Christelle Rigollier, Mireille Lefèvre, and Lucien Wald. The method heliosat-2 for deriving short-wave solar radiation from satellite images. *Solar energy*, 77(2):159–169, 2004.

David Salinas, Michael Bohlke-Schneider, Laurent Callot, Roberto Medico, and Jan Gasthaus. High-dimensional multivariate forecasting with low-rank gaussian copula processes. *Advances in neural information processing systems*, 32, 2019.

Zhiyuan Si, Ming Yang, Yixiao Yu, and Tingting Ding. Photovoltaic power forecast based on satellite images considering effects of solar position. *Applied Energy*, 302:117514, 2021.

Jelena Simeunović, Baptiste Schubnel, Pierre-Jean Alet, and Rafael E Carrillo. Spatio-temporal graph neural networks for multi-site pv power forecasting. *IEEE Transactions on Sustainable Energy*, 13(2):1210–1220, 2021.

Andrew Staniforth and Jean Côté. Semi-lagrangian integration schemes for atmospheric models—a review. *Monthly weather review*, 119(9):2206–2223, 1991.

Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8934–8943. IEEE, 2018.

Michael Tiedtke. Representation of clouds in large-scale models. *Monthly Weather Review*, 121 (11):3040–3061, 1993.

Christopher J Traverse, Richa Pandey, Miles C Barr, and Richard R Lunt. Emergence of highly transparent photovoltaics for distributed applications. *Nature Energy*, 2(11):849–860, 2017.

Dennis van Der Meer. A benchmark for multivariate probabilistic solar irradiance forecasts. *Solar Energy*, 225:286–296, 2021.

Meng Wan, Tiantian Liu, Yuxuan Bi, Jue Wang, Hui Cui, Rongqiang Cao, Jiaxiang Wang, Peng Shi, Ningming Nie, and Yangang Wang. Mcloudnet: An ultra-short-term photovoltaic power forecasting framework with multi-layer cloud coverage. In *IJCAI*, pp. 9908–9917, 2025.

Yuqing Wang, Wenjie Fu, Junfa Chen, Junlong Wang, Zhao Zhen, Fei Wang, Fei Xu, Neven Duić, Di Yang, and Yuntong Lv. Spatiotemporal federated learning based regional distributed pv ultra-short-term power forecasting method. *IEEE Transactions on Industry Applications*, 60(5):7413–7425, 2024.

Haomin Wen, Youfang Lin, Yutong Xia, Huaiyu Wan, Qingsong Wen, Roger Zimmermann, and Yuxuan Liang. Diffstg: Probabilistic spatio-temporal graph forecasting with denoising diffusion models. In *Proceedings of the 31st ACM international conference on advances in geographic information systems*, pp. 1–12, 2023.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 753–763, 2020.

Yanru Yang, Yu Liu, Yihang Zhang, Shaolong Shu, and Junsheng Zheng. Dest-gnn: A double-explored spatio-temporal graph neural network for multi-site intra-hour pv power forecasting. *Applied Energy*, 378:124744, 2025.

Tiechui Yao, Jue Wang, Haoyan Wu, Pei Zhang, Shigang Li, Yangang Wang, Xuebin Chi, and Min Shi. A photovoltaic power output dataset: Multi-source photovoltaic power output dataset with python toolkit. *Solar Energy*, 230:122–130, 2021.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11121–11128, 2023.

Yujia Zhang, Yuzhou Zhang, Zhixiang Dai, and Rita Zhang. Rethinking the use of deep learning methods for photovoltaic power forecasting. *Nature Communications*, 2026.

Vincent Zhihao Zheng and Lijun Sun. Multivariate probabilistic time series forecasting with correlated errors. *Advances in Neural Information Processing Systems*, 37:54288–54329, 2024.

A Method Details

A.1 System Context

Power-history encoding. The patch embeddings are augmented with learned positional information and refined by branch-specific mixers:

$$\mathbf{Z}_m = \mathrm{Mixer}_r\big(\mathrm{Linear}_m\big(\mathrm{Patch}(\mathbf{X}_m)\big) + \mathbf{P}_{\mathrm{pos}}\big), \qquad m \in \{\mathrm{ts}, \Delta\mathrm{ts}\}, \tag{8}$$

where $\mathbf{P}_{\mathrm{pos}}$ is a learned patch-position embedding. The two branches use independent projections and mixer parameters.

Lead-specific context. A Fourier encoding of the forecast lead is added to $\mathbf{Z}_{\mathrm{NWP}}$, giving the query $\mathbf{Q}$ used in Equation 1.

System information combines location, asset, and lead-dependent solar geometry:

$$\mathbf{e}^{\mathrm{sys}} = \mathbf{e}^{\mathrm{loc}} + \mathbf{e}^{\mathrm{asset}} + \mathbf{e}^{\mathrm{geom}}. \tag{9}$$

The location representation is obtained from Fourier-encoded satellite-grid coordinates. Asset information describes panel orientation and the availability of its attributes, while the geometry representation combines panel orientation, solar direction, and incidence angle.

Gated fusion. The level branch combines attended power history, NWP, and system information, whereas the difference branch combines attended power changes with the temporal change in system geometry:

$$\mathbf{h}^{\mathrm{level}} = \mathbf{H}_{\mathrm{ts}} + \mathbf{Q} + \mathbf{e}^{\mathrm{sys}}, \qquad \mathbf{h}^{\Delta} = \mathbf{H}_{\Delta} + \Delta\mathbf{e}^{\mathrm{sys}}. \tag{10}$$

Here Δ denotes first differences along the forecast-time dimension, with the initial step set to zero to preserve the sequence length. Static location and asset embeddings are broadcast over forecast steps, so only the lead-dependent geometry contributes to $\Delta\mathbf{e}^{\mathrm{sys}}$. Both branch representations have shape $[B, T_{\mathrm{out}}, N, n_{\mathrm{hid}}]$. A channel-wise gate $\boldsymbol{\gamma}$, with the same shape as the branch representations,

adaptively combines them. The learned projection $\mathbf{W}_{\mathrm{g}}$ maps the concatenated $2n_{\mathrm{hid}}$ features to n_{hid} gate logits:

$$\begin{aligned}\boldsymbol{\gamma} &= \sigma\big(\mathbf{W}_{\mathrm{g}}\,\mathrm{Concat}_{\mathrm{ch}}[\mathbf{h}^{\mathrm{level}}, \mathbf{h}^{\Delta}]\big)\,,\\ \mathbf{h}^{\mathrm{mix}} &= 2\boldsymbol{\gamma} \odot \mathbf{h}^{\mathrm{level}} + 2(1-\boldsymbol{\gamma}) \odot \mathbf{h}^{\Delta},\\ \mathbf{H}_{\mathrm{sys}} &= \mathbf{h}^{\mathrm{mix}} + \mathrm{GRN}\big(\mathrm{Concat}_{\mathrm{ch}}[\mathbf{h}^{\mathrm{level}}, \mathbf{h}^{\Delta}]\big)\,.\end{aligned} \tag{11}$$

The gate is initialized to assign equal weight to the two branches, making their initial combination additive. Here GRN is a gated residual network that maps the concatenated branch features to n_{hid} channels and captures interactions between power levels and changes.

A.2 Cloud Evolution

The cloud rollout uses bilinear backward sampling with border extension. Accumulated displacements sample the initial field once per forecast lead, while retention and source–sink contributions are propagated separately to limit smoothing from interpolation. Motion perturbations are held fixed across the forecast window. For each perturbed quantity, let $\mathbf{M}_a : [B, T_{\mathrm{out}}, n_a, H, W]$ contain learned spatial modes and $\mathbf{z}_a : [B, S, T_{\mathrm{out}}, n_a]$ denote their Gaussian coefficients, where $a \in \{\mathrm{eq}, \eta\}$ identifies the perturbed quantity and n_a is its number of modes. The modes are conditioned on the base forecast, its gradients, the latest cloud field, and motion. The perturbations are $\boldsymbol{\epsilon}_a = \boldsymbol{\alpha}_a \odot \langle \mathbf{M}_a, \mathbf{z}_a \rangle_{\mathrm{mode}}$, where $\boldsymbol{\alpha}_a : [B, T_{\mathrm{out}}]$ contains learned nonnegative amplitudes. The contraction sums over modes, broadcasting modes over trajectories and amplitudes over trajectories and spatial cells. Coefficients start from independent standard normal draws and use the unit-variance AR(1) construction in Equation 4 with learned correlations. Antithetic pairs negate the motion perturbations and Gaussian innovations. Logarithm and logit inputs are clipped away from singularities.

A.3 Cloud-conditioned Dependence

Let $n_{\mathrm{sm}} = |\mathcal{R}_{\mathrm{sm}}|$ denote the number of topology-adaptive smoothing scales. Average pooling at these scales, followed by bilinear sampling at the PV-system locations, produces $\mathbf{V} : [B, S, T_{\mathrm{out}}, N, n_{\mathrm{sm}}]$. The pooled fields are stacked along the final dimension. Their standardized deviations are

$$\boldsymbol{\Psi} = \frac{\mathbf{V} - \mathrm{Mean}_N(\mathbf{V})}{\max\left(\mathrm{RMS}_N(\mathbf{V} - \mathrm{Mean}_N(\mathbf{V}))\,, \epsilon\right)}. \tag{12}$$

Here Mean_N and RMS_N compute the mean and root mean square over PV systems, retaining a singleton system dimension for broadcasting. The positive constant ϵ prevents division by zero. Normalization is performed separately for each forecast origin, cloud trajectory, forecast step, and smoothing scale.

Adding the constant regional component gives $n_{\mathrm{fac}} = 1 + n_{\mathrm{sm}}$ factors and loadings $\mathbf{A} : [B, S, T_{\mathrm{out}}, N, n_{\mathrm{fac}}]$. Shared factors $\boldsymbol{\zeta}$ and local factors $\mathbf{e}$ have shapes of $[B, S, R, T_{\mathrm{out}}, n_{\mathrm{fac}}]$ and $[B, S, R, T_{\mathrm{out}}, N]$, respectively. Their initial states and innovations are independent standard normal. Shared and local temporal correlations are separately parameterized with magnitudes below one. The loadings are broadcast over conditional realizations, and the shared factors are broadcast over PV systems.

A.4 Shared Power Decoder

Let $\mathbf{E}_{\mathrm{adv}}$, $\mathbf{E}_{\mathrm{proc}}$, and $\mathbf{E}_{\mathrm{field}}$ denote the transport, process, and convolutional field representations described in the main text. Each has shape of $[B, S, T_{\mathrm{out}}, N, n_{\mathrm{hid}}]$. The transport and field gates $\mathbf{q}_{\mathrm{adv}}$ and $\mathbf{q}_{\mathrm{field}}$ have the shape of $[B, S, T_{\mathrm{out}}, N, 1]$ and scale the corresponding features:

$$\mathbf{E}^{C} = \mathrm{Concat}_{\mathrm{ch}}[\mathbf{q}_{\mathrm{adv}} \odot \mathbf{E}_{\mathrm{adv}}, \mathbf{q}_{\mathrm{field}} \odot (\mathbf{E}_{\mathrm{proc}} + \mathbf{E}_{\mathrm{field}})]\,. \tag{13}$$

Thus $n_{\mathrm{cloud}} = 2n_{\mathrm{hid}}$. The system context is broadcast over cloud trajectories and modulates these cloud features through a learned channel-wise scale. The projected cloud features are denoted by

$\mathbf{H}_C$ and their nonlinear activation is $\mathbf{H}_C^{\mathrm{act}}$. The context-conditioned residual is

$$\begin{aligned}\mathbf{H}_C &= \mathrm{Linear}_C\big([1+\tanh(\mathrm{Linear}_{\mathrm{scale}}(\mathbf{H}_{\mathrm{sys}}))]\odot\mathbf{E}^C\big)\,,\\ \mathbf{H}_C^{\mathrm{act}} &= \mathrm{GELU}(\mathbf{H}_C),\\ \boldsymbol{\varphi} &= \mathbf{H}_C+\mathrm{MLP}_C\big(\mathrm{Concat}_{\mathrm{ch}}\big[\mathbf{H}_C^{\mathrm{act}},\mathbf{H}_C^{\mathrm{act}}\odot\mathrm{Linear}_{\mathrm{ctx}}(\mathbf{H}_{\mathrm{sys}})\big]\big)\,.\end{aligned} \tag{14}$$

The scale projection outputs n_{cloud} channels. The cloud projection, context projection, and residual MLP output n_{hid} channels, so $\mathbf{H}_C$, $\mathbf{H}_C^{\mathrm{act}}$, and $\boldsymbol{\varphi}$ all have the shape of $[B,S,T_{\mathrm{out}},N,n_{\mathrm{hid}}]$. The cloud projection and residual MLP are bias-free, ensuring $\boldsymbol{\varphi}=\mathbf{0}$ when $\mathbf{E}^C=\mathbf{0}$.

During conditional sampling, quantiles are broadcast over the R realizations. Probability levels outside the quantile grid take the nearest boundary value.

B Implementation and Training Details

B.1 Cloud-index Preprocessing

PVCast uses the same daylight criterion and Heliosat transformation across datasets, while the visible channel and reflectance preprocessing depend on the satellite product. Here, e_t denotes the solar elevation at time t. The daylight mask and retrieved cloud index are

$$m_t=\mathbb{I}\{e_t>5^\circ\},\qquad n_t=m_t\ \mathrm{clip}_{[0,1.2]}\left(\frac{\rho_t-p_5}{p_{95}-p_5}\right), \tag{15}$$

where ρ_t is visible reflectance and p_5,p_{95} are its daylight envelope percentiles, estimated from at most 64 daylight frames. Satellite and cloud-index fields are set to zero outside daylight hours, and the Heliosat transformation is not evaluated at night.

UKPV derives ρ_t from SEVIRI VIS006, whose archived values are already reflectance. PVOD uses Himawari B04 and applies $\rho_t=I_t/\max(\sin e_t,\sin 5^\circ)$. MMSP uses Himawari Albedo_03, which is also provided as reflectance.

The field-level daylight mask is evaluated at one reference location per dataset: $54.0^\circ\mathrm{N}, 2.5^\circ\mathrm{W}$ for UKPV, $38.1^\circ\mathrm{N}, 115.5^\circ\mathrm{E}$ for PVOD, and $36.4^\circ\mathrm{N}, 118.0^\circ\mathrm{E}$ for MMSP. PVOD additionally applies system-specific daylight masks to power and NWP variables, whereas UKPV and MMSP use their corresponding field-level masks. The retrieved index n_t is supplied to the cloud-state module. The two causal fields used for motion estimation and cloud rollout are clipped to $[0,1]$, defining the model state $\mathbf{C}^{\mathrm{obs}}$. Non-finite values in precomputed cloud indices are replaced with zero, with no additional rescaling. The cloud index is distinct from the multichannel satellite context $\mathbf{X}_{\mathrm{sat}}$. UKPV uses VIS008, IR_108, and WV_062; PVOD uses B01, B04, and B16; and MMSP uses four albedo channels. These context channels are set to zero at night but are not transformed by Equation 15.

B.2 Implementation Settings

Table 4 collects the system-encoder, cloud, dependence, decoder, and training settings. Quantile-based models use their quantile functions for system-level CRPS and sampled trajectories for joint metrics. Sample-based models use empirical distributions for all metrics.

Baseline implementation. Trainable baselines use the shared optimizer, learning rate, batch size, epoch budget, and early-stopping patience in Table 4. The checkpoint with the best validation objective is retained. Input windows, forecast horizons, chronological splits, and evaluation masks follow the same dataset protocol as PVCast. Each baseline retains its supported input modalities. Where applicable, hidden width, attention heads, feed-forward width, and encoder depth use the system-encoder settings in that table. Table 5 lists model-specific settings and exceptions. Smart Persistence requires no gradient-based training.

Table 4: PVCast architecture and shared training settings.

Component	Hyperparameter	Value
System encoder	Hidden width	64
	Feed-forward width	128
	Attention heads	4
	Mixer layers	2
	Patch length (steps)	16
	Patch stride (steps)	8
	LayerScale initialization	10^{-2}
	Attention dropout	0
	Lead-encoding frequencies	4
	NWP features	8
Cloud evolution	Source–sink kernel	3×3
	Equilibrium modes	4
	Intensity modes	4
Power decoder	Quantile grid $\boldsymbol{\tau}$	$0.1, 0.2, \ldots, 0.9$
Cloud sampling	Training trajectories	4
	Inference trajectories	32
Dependence	Distance quantiles	$(0.10, 0.30, 0.55, 0.85)$
	Factor-norm bound	0.95
	AR range	$[0.05, 0.95]$
	Regional amplitude	0.25
Shared training	Optimizer	Adam
	Learning rate	10^{-3}
	Learning-rate schedule	Constant
	Batch size	32
	Epoch budget	8
	Early-stopping patience (epochs)	4
	Random seed	2026

C Dataset Details and Experimental Settings

C.1 Dataset Details

Table 6 summarizes the dataset scale, temporal resolution, forecasting windows, chronological splits, and retained forecast origins. Input and prediction lengths are reported in time steps. The split ratios specify chronological boundaries rather than the proportions of the final samples. Within each split, we retain daytime forecast origins for which all daylight satellite frames in the preceding 24-hour input window are available; satellite inputs outside daylight hours are set to zero. Consequently, differences in daylight duration and data availability produce origin counts that do not exactly follow the nominal split ratios. All metrics are computed over valid daylight prediction steps.

PVOD. PVOD provides 15-minute PV power and NWP observations from ten systems in China (Yao et al., 2021). We combine the satellite data JMA Himawari-8/9 data from the NOAA AWS public data constructed by mosaicking tiles `h05v03` and `h06v03` at $0.01°$ resolution. Each 75×150 regional field contains bands B01, B04, and B16, with 10,392 satellite tiles covering the PV observation period. All modalities are aligned to the 15-minute forecasting grid. Aligning the evaluation protocol from Cross-Unet (Zhang et al., 2026), we use 96 input steps, 16 prediction steps, and an 80/10/10 training, validation, and testing split.

MMSP. MMSP is the solar power dataset released with FusionSF (Ma et al., 2024), comprising three modalities: solar power generation, satellite imagery, and NWP forecasts. It contains hourly observations from 88 PV plants over an approximately $36.4°\mathrm{N}$, $118.0°\mathrm{E}$ region in Shandong, China, from 2 January 2021 to 20 June 2022. Timestamps are interpreted in the `Asia/Shanghai` time

Table 5: Baseline-specific hyperparameters. Unlisted settings follow Table 4 where applicable.

Model	Hyperparameter	Value
DLinear	Moving-average window	25
MTGNN	Graph propagation depth	2
Correlated Errors	Factor rank	$\min(8, N)$
TACTiS-2	Flow layers	2
	Copula layers	2
	Resolution	10
TimeGrad	Diffusion steps	100
	Denoiser residual blocks	8
DiffSTG	Diffusion steps	200
	Reverse steps	50
Cross-Unet	Encoder scales	3
	Segment-merging window	2
FusionSF	Hidden width	128
	Depth	4
	Dropout	0.3
SHADECast	Latent width	32
	Diffusion steps	100

Table 6: **Dataset and forecast settings.** Split ratios define chronological time boundaries, while origin counts are reported after daylight and data-availability filtering.

Dataset	Country	# sites	Interval	Input T_{in}	Pred. T_{out}	Chronological split	Retained origins (train/val/test)
UKPV	UK	841	30 min	48	16	60/20/20	7,774 / 3,879 / 2,212
MMSP	China	88	1 h	24	24	60/20/20	3,633 / 949 / 1,286
PVOD	China	10	15 min	96	16	80/10/10	8,599 / 1,033 / 822

zone. Because the released plant coordinates are spatially desensitized, solar elevation is calculated at the provincial centroid, $36.4°\text{N}$, $118.0°\text{E}$. The satellite input comprises four Himawari-8/9 AHI albedo channels at wavelengths of 0.47, 0.51, 0.64, and $0.86\,\mu\text{m}$, provided as 64×64 regional fields. Each forecast origin uses 24 hours of historical power and satellite observations together with 24 hours of future ECMWF forecasts at each plant. The NWP variables include global horizontal and direct irradiance, total and low cloud cover, 2-m temperature, surface pressure, and 100-m wind.

UKPV. We construct UKPV by aligning Open Climate Fix PV measurements (Open Climate Fix, 2025), EUMETSAT SEVIRI imagery as satellite observation (EUMETSAT), and DWD ICON-EU forecasts(Open Climate Fix, 2023) on a common 30-minute grid. Satellite observations are obtained from the public SEVIRI Rapid Scan archive prepared by Open Climate Fix. We use data from January 2020 to December 2021, matching the availability of the archived ICON-EU forecasts. Starting from 22,972 systems observed during 2020–2021, we construct the station panel following the daytime preprocessing protocol of Nassimiha et al. (2023). Raw generation is first restricted to its physically valid range and normalized by the nameplate capacity of each system. We then exclude systems that report non-zero generation overnight or contain substantial missing observations. The station coverage is evaluated between 08:00 and 16:00 UTC. Using the training fold only, we retain systems with at least 50% coverage, complete system geometry, and locations inside the SEVIRI crop. Intersecting the resulting set with the 993 identifiers in the OCF high-frequency reporting panel gives the fixed set of 841 systems used throughout our experiments. Missing values are masked without forward filling. Satellite inputs contain only observations available by the forecast origin, and NWP inputs use the eligible 00 UTC initialization at their forecast-valid times. Evaluation uses a lead-specific daylight mask based on solar elevation inspired by the protocol from Fulton et al. (2024).

The system-information encoder has the same structure across datasets, while the availability of panel attributes differs. UKPV provides tilt and azimuth, PVOD provides tilt but not azimuth, and MMSP provides neither. Unavailable attributes are set to zero and identified by their corresponding masks. For MMSP, the panel normal defaults to the zenith direction.

D Evaluation Details

The observed power is denoted by $\mathbf{Y} \in \mathbb{R}^{B\times T_{\text{out}}\times N}$, and PVCast produces $\widehat{\mathbf{P}}_{\text{joint}} \in \mathbb{R}^{B\times S\times R\times T_{\text{out}}\times N}$. For evaluation, the cloud and conditional-sampling dimensions are flattened as $j = (s-1)R + r$, giving $J = SR$ joint power trajectories $\widehat{\mathbf{P}}^{(j)}$. The final marginal quantiles are $\mathbf{Q}_{\text{marg}} \in \mathbb{R}^{B\times T_{\text{out}}\times N\times n_{\text{q}}}$ at probability levels $\boldsymbol{\tau}$. All power values are normalized by installed capacity, and all reported scores are averaged over valid daytime targets.

For an empirical predictive distribution $\{x^{(j)}\}_{j=1}^{J}$, we use

$$\text{CRPS}\left(\{x^{(j)}\}_{j=1}^{J}, y\right) = \frac{1}{J}\sum_{j=1}^{J}|x^{(j)} - y| - \frac{1}{2J^2}\sum_{j=1}^{J}\sum_{j'=1}^{J}|x^{(j)} - x^{(j')}|. \tag{16}$$

D.1 Evaluation Metrics

CRPS. The Continuous Ranked Probability Score evaluates the marginal power distribution of each PV system and forecast step (Gneiting & Raftery, 2007):

$$\text{CRPS}(F, y) = 2\int_0^1 \rho_\alpha(y - Q(\alpha))\,\mathrm{d}\alpha, \qquad \rho_\alpha(u) = u\big(\alpha - \mathbb{I}\{u < 0\}\big), \tag{17}$$

where F and Q are the predictive CDF and quantile function, respectively. For PVCast, Q is the piecewise-linear quantile function defined by $\mathbf{Q}_{\text{marg}}$, with constant values beyond its boundary probability levels. For a sample-based forecast $\{x^{(j)}\}_{j=1}^{J}$, we use the empirical form defined in Equation 16. The marginal CRPS is averaged over forecast origins, systems, and forecast steps.

CRPS-sum. CRPS-sum evaluates the predictive distribution after spatial aggregation. For the valid systems $\mathcal{I}_{b,k}$, the normalized capacity weights are defined by

$$w_{b,k,i} = \frac{K_i}{\sum_{i'\in\mathcal{I}_{b,k}} K_{i'}}, \qquad i \in \mathcal{I}_{b,k},$$

where K_i is the installed capacity of system i. The observed and sampled aggregate power are

$$Y_{b,k}^{\text{sum}} = \sum_{i\in\mathcal{I}_{b,k}} w_{b,k,i} Y_{b,k,i}, \qquad \widehat{P}_{b,k}^{(j),\text{sum}} = \sum_{i\in\mathcal{I}_{b,k}} w_{b,k,i}\widehat{P}_{b,k,i}^{(j)}. \tag{18}$$

We compute $\text{CRPS-sum}_{b,k} = \widehat{\text{CRPS}}\Big(\{\widehat{P}_{b,k}^{(j),\text{sum}}\}_{j=1}^{J}, Y_{b,k}^{\text{sum}}\Big)$ and average the score over valid forecast origins and steps. Because the same trajectory index j is retained across systems, CRPS-sum depends on both the marginal power distributions and their cross-system dependence. For the scale-wise analysis in Figure 4, the aggregation set is replaced by the radius-r neighborhood

$$\mathcal{N}_r(a) = \{i : \text{dist}(\mathbf{x}_i, \mathbf{x}_a) \le r\}$$

of each anchor system a, with capacity weights normalized within the neighborhood. Scores are averaged over all PV systems used as spatial anchors.

Ramp CRPS. Ramp CRPS evaluates the predictive distribution of within-system power changes. Let $\mathbf{Y}_0 \in \mathbb{R}^{B\times N}$ be the most recent observed power. Prepending it to the observed forecast and every sampled forecast gives the sequences $\mathbf{Y}^+ = (\mathbf{Y}_0, \mathbf{Y}_1, \ldots, \mathbf{Y}_{T_{\text{out}}})$ and $\widehat{\mathbf{P}}^{(j),+} = (\mathbf{Y}_0, \widehat{\mathbf{P}}_1^{(j)}, \ldots, \widehat{\mathbf{P}}_{T_{\text{out}}}^{(j)})$. For any sequence $\mathbf{Z}^+ = (\mathbf{Z}_0, \ldots, \mathbf{Z}_{T_{\text{out}}})$, the lag-$\ell$ difference operator is

$$\Delta_\ell \mathbf{Z}^+ = (\mathbf{Z}_\ell - \mathbf{Z}_0, \mathbf{Z}_{\ell+1} - \mathbf{Z}_1, \ldots, \mathbf{Z}_{T_{\text{out}}} - \mathbf{Z}_{T_{\text{out}}-\ell}). \tag{19}$$

Applying this operator to $\mathbf{Y}^{+}$ and $\widehat{\mathbf{P}}^{(j),+}$ gives $\Delta_\ell \mathbf{Y}$ and $\Delta_\ell \widehat{\mathbf{P}}^{(j)}$, both with the shape of $[B, T_{\text{out}} + 1 - \ell, N]$. The default one-step Ramp CRPS uses $\ell = 1$:

$$\text{RampCRPS}_{b,k,i} = \widehat{\text{CRPS}}\Big(\{\Delta_1 \widehat{P}^{(j)}_{b,k,i}\}_{j=1}^{J}, \Delta_1 Y_{b,k,i}\Big). \tag{20}$$

The score is averaged over forecast origins, systems, and valid forecast steps. Scale-wise results additionally evaluate Equation 19 at larger ℓ.

Energy Score. The Energy Score evaluates the joint temporal distribution of each system's complete one-step ramp trajectory (Gneiting & Raftery, 2007). For forecast origin b and system i, denote the observed ramp vector by $\Delta_1 \mathbf{Y}_{b,i} \in \mathbb{R}^{T_{\text{out}}}$ and its sampled counterparts by $\{\Delta_1 \widehat{\mathbf{P}}^{(j)}_{b,i}\}_{j=1}^{J}$. Then

$$\begin{aligned} \text{ES}_{b,i} = \frac{1}{J}\sum_{j=1}^{J} \Big\| \Delta_1 \widehat{\mathbf{P}}^{(j)}_{b,i} - \Delta_1 \mathbf{Y}_{b,i} \Big\|_2 \\ - \frac{1}{2J^2}\sum_{j=1}^{J}\sum_{j'=1}^{J} \Big\| \Delta_1 \widehat{\mathbf{P}}^{(j)}_{b,i} - \Delta_1 \widehat{\mathbf{P}}^{(j')}_{b,i} \Big\|_2 . \end{aligned} \tag{21}$$

The score is averaged over forecast origins and systems. Applying ES to ramps reduces the dominant contribution of the diurnal power cycle and emphasizes the temporal organization of power changes.

Cloud-field skill. Cloud-field skill evaluates the unperturbed cloud forecast $\widehat{\mathbf{C}} \in [0,1]^{B \times T_{\text{out}} \times H \times W}$ against the future verification field $\mathbf{C}_{\text{fut}}$ of the same shape, obtained by applying the preprocessing in Appendix B.1 to future observations. The persistence reference repeats the current cloud field $\mathbf{C}_0 \in [0,1]^{B \times H \times W}$ over the complete output window:

$$\widehat{\mathbf{C}}_{\text{pers}} = \text{Broadcast}_{T_{\text{out}}}(\mathbf{C}_0) \in [0,1]^{B \times T_{\text{out}} \times H \times W}.$$

Let $\mathbf{M}^C$ be the binary mask of valid daytime pixels and let $\boldsymbol{\lambda} \in \mathbb{R}_{+}^{T_{\text{out}}}$, with $\mathbf{1}^\top \boldsymbol{\lambda} = 1$, contain the forecast-step weights. Broadcasting these weights to the cloud-field shape gives $\mathbf{L}_\lambda$. For any cloud forecast $\widetilde{\mathbf{C}}$, its masked field error is

$$\mathcal{E}_C(\widetilde{\mathbf{C}}) = \frac{\Big\langle \mathbf{L}_\lambda \odot \mathbf{M}^C, \Big(\widetilde{\mathbf{C}} - \mathbf{C}_{\text{fut}}\Big)^{\odot 2} \Big\rangle}{\langle \mathbf{L}_\lambda, \mathbf{M}^C \rangle}, \tag{22}$$

where $\odot$ denotes elementwise operations and $\langle \cdot, \cdot \rangle$ sums over all tensor entries. Cloud-field skill relative to persistence is

$$\text{Skill}_C = 1 - \frac{\mathcal{E}_C(\widehat{\mathbf{C}})}{\mathcal{E}_C(\widehat{\mathbf{C}}_{\text{pers}})}. \tag{23}$$

Positive, zero, and negative values indicate performance above, equal to, and below persistence, respectively.

Aggregate Ramp CRPS. Aggregate Ramp CRPS evaluates the distribution of capacity-weighted power changes over the complete PV network. It differs from the system-level Ramp CRPS in Equation 20: system-level Ramp CRPS scores each system separately before averaging, whereas Aggregate Ramp CRPS first combines the system ramps and then evaluates their joint sample distribution.

The most recent observed power is prepended to the observed and sampled forecasts as in Equation 19. For the default one-step lag, define

$$\Delta Y_{b,k,i} = Y^{+}_{b,k,i} - Y^{+}_{b,k-1,i}, \qquad \Delta \widehat{P}^{(j)}_{b,k,i} = \widehat{P}^{(j),+}_{b,k,i} - \widehat{P}^{(j),+}_{b,k-1,i}. \tag{24}$$

Consequently, the first predicted ramp is measured relative to the latest observation at the forecast origin.

Let $M_{b,k,i}$ indicate whether system i is valid at the later endpoint k, and let K_i denote its installed capacity. The lead-specific weights are

$$w_{b,k,i} = \frac{M_{b,k,i} K_i}{\sum_{i'=1}^{N} M_{b,k,i'} K_{i'}}. \tag{25}$$

The observed and sampled aggregate ramps are

$$\Delta Y_{b,k}^{\mathrm{agg}} = \sum_{i=1}^{N} w_{b,k,i} \Delta Y_{b,k,i}, \qquad \Delta \widehat{P}_{b,k}^{(j),\mathrm{agg}} = \sum_{i=1}^{N} w_{b,k,i} \Delta \widehat{P}_{b,k,i}^{(j)}. \tag{26}$$

Validity and capacity normalization are determined only by the later endpoint k. Aggregate Ramp CRPS is then

$$\mathrm{AggregateRampCRPS}_{b,k} = \widehat{\mathrm{CRPS}}\left(\left\{ \Delta \widehat{P}_{b,k}^{(j),\mathrm{agg}} \right\}_{j=1}^{J}, \Delta Y_{b,k}^{\mathrm{agg}} \right), \tag{27}$$

and is averaged over forecast origins and steps.

E Detailed Experimental Analysis

E.1 Ablation Protocols

All ablations use the evaluation pipeline and sampling budget of the main experiments. Cloud-evolution and system-context ablations alter the corresponding forecasting paths and therefore recompute both marginal and joint predictions. Dependence ablations instead reuse the conditional quantile functions from the same model forward pass and modify only the probability-level coupling. Their marginal CRPS is consequently identical up to numerical precision.

Cloud evolution. *Persistence* repeats the latest observed cloud field over the forecast window. *No cloud* removes cloud inputs while retaining power and NWP information. *Advection only* disables the source and sink fields and transports the latest observed cloud field. *Source–sink only* sets the transport displacement to zero while retaining source–sink estimates conditioned on the inferred flow. *Shuffled cloud fields* shuffles cloud fields G, D and v. *Full* retains both transport and source–sink evolution.

Cloud-conditioned dependence. *No spatial dependence* removes the shared cross-system factors while retaining the system-specific temporal factors. *No temporal dependence* retains the shared spatial factors but sets both the shared and local AR(1) coefficients to zero. Thus, systems remain coupled within each forecast step, while different forecast steps are temporally independent. *Independent* assigns independent probability levels to every system and forecast step. The implementation draws independent Gaussian values and converts their ranks into stratified uniform levels on the fixed conditional quantile functions.

System context. *No power history* suppresses the attended power-level branch. *No power difference* suppresses the attended difference branch. *Shared history summary* replaces lead-specific cross-attention with the mean historical patch representation while retaining NWP. *No NWP* sets the NWP query to zero. *No system information* removes the system-information embedding. *Sum fusion* replaces the learned channel-wise gate with an unweighted sum of the two power branches.

E.2 System-Context Ablation

Table 7 evaluates the six components of the system-context encoder. Removing any component degrades all four forecasting metrics. The largest changes occur when learned gating is replaced by summation or when lead-specific cross-attention is replaced by a shared history mean. These changes increase CRPS-sum by approximately 240%, demonstrating the importance of adaptive feature fusion and forecast-specific summaries of power history. Historical power contributes more than its differences to marginal accuracy, although both representations improve performance. Removing system information increases CRPS-sum from 0.0119 to 0.0262, highlighting its importance when combining heterogeneous PV systems. NWP has the smallest effect on marginal CRPS, but its removal still increases CRPS by 14.2% and CRPS-sum by 50.3%. Overall, the context encoder provides complementary historical, meteorological, and system-specific information for cloud-driven joint forecasting.

Table 7: **System-context ablations on UKPV.** Lower values are better.

Context configuration	CRPS ↓	CRPS-sum ↓	Ramp CRPS ↓	ES ↓
No power history	0.0603	0.0286	0.0395	0.1569
No power differences	0.0491	0.0198	0.0339	0.1360
Cross-attention → history mean	0.0890	0.0403	0.0455	0.1781
No NWP	0.0450	0.0179	0.0330	0.1359
No system information	0.0478	0.0262	0.0351	0.1399
Learned gate → summation	0.1003	0.0405	0.0739	0.2778
Full encoder	**0.0394**	**0.0119**	**0.0316**	**0.1294**

Table 8: **Sensitivity and computational cost with respect to the number of cloud trajectories on UKPV.**

S	Forecast performance				Computational cost	
	CRPS ↓	CRPS-sum ↓	Ramp CRPS ↓	ES ↓	Time (ms/origin) ↓	Peak memory (GB) ↓
4	0.0399	0.0140	0.0398	0.1634	13.0	17.9
8	0.0397	0.0127	0.0353	0.1449	21.1	18.4
16	0.0395	0.0122	0.0330	0.1357	38.6	18.6
32	**0.0394**	**0.0119**	**0.0316**	**0.1294**	**73.5**	**19.0**
64	0.0395	0.0118	0.0314	0.1288	143.2	19.9

E.3 Sensitivity and Computational Cost

Table 8 examines the effect of the number of cloud trajectories S on forecast performance and computational cost. Forecast scores are evaluated on the UKPV test set with the calibrated dependence parameters held fixed. Computational cost is measured over 256 forecast origins on one NVIDIA L20X with batch size 32 and 841 PV systems after warm-up; the reported inference time is amortized per origin. PVCast contains 648,097 trainable parameters for every S, because increasing the number of trajectories does not change the learned model. Lower values are better, and bold denotes the configuration used in the main experiments. Doubling the default trajectory budget raises inference time by 94.8%, while reducing CRPS-sum by only 1.0% and Ramp CRPS and ES by approximately 1.8%. Peak memory increases by 4.7% because trajectories are decoded in fixed-size chunks.

E.4 Detailed Cross-dataset Experiments

Additional scale-wise results. For a more complete assessment beyond UKPV, Figures 6 and 7 extend the scale-wise analysis to MMSP and PVOD. Panel (a) evaluates marginal accuracy across forecast leads. On PVOD, PVCast achieves the lowest or tied system-level CRPS at three of the four displayed leads, showing that its joint construction preserves competitive marginal forecasts.

Panel (b) evaluates aggregate-power uncertainty across spatial scales. On MMSP, PVCast attains the joint-best CRPS-sum at the widest aggregation scale. The result is more pronounced on PVOD, where PVCast achieves the lowest CRPS-sum at every radius. Its reduction over the next-best method increases from approximately 3% at $10\,\mathrm{km}$ to 10% at the widest displayed scale. The growing separation indicates that the benefit becomes more visible as forecasts from additional systems are combined.

Panel (c) evaluates the distribution of power changes over different temporal lags. PVCast obtains the lowest or tied Ramp CRPS across all lags on MMSP and the lowest score at every lag on PVOD. On PVOD, the reduction over the next-best method is approximately 17% at 0.25 hours and 9% at 0.5 hours, showing a particularly clear advantage for short-term power changes. These results com-

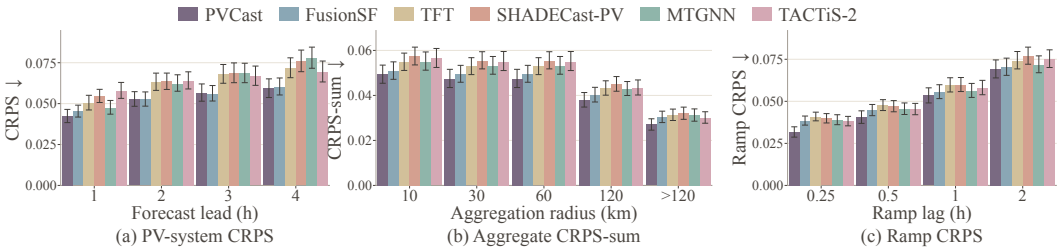

Figure 6: **Forecast performance across spatiotemporal scales on PVOD.**

plement the UKPV analysis by showing that cloud-driven coupling improves aggregate uncertainty and temporal evolution across datasets with different sampling intervals and system topologies.

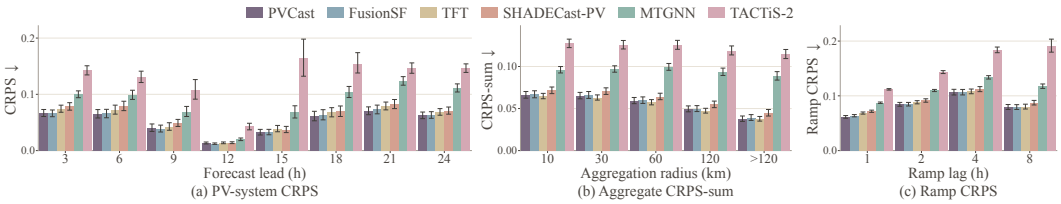

Figure 7: **Forecast performance across spatiotemporal scales on MMSP.**

Cloud evolution across regimes. To complement the regime-wise quantitative results, Figures 8, 9, and 10 visualize cloud-field and aggregate-power forecasts under stable, moderate, and variable cloud conditions, respectively. In each figure, the upper row shows the cloud observations at 0.5, 2, and 4 hours, and the middle row shows the corresponding PVCast forecasts. These comparisons examine whether the learned rollout preserves the location, orientation, and large-scale evolution of the observed cloud structures as the forecast horizon increases. The lower panel shows the observed capacity-normalized aggregate power together with the predictive median and pointwise 10th–90th percentile interval obtained after propagating the cloud trajectories through the power model.

Under the stable regime, the observed cloud field retains a broadly consistent spatial organization over the forecast window. At the 0.5-hour lead, PVCast reproduces both the location and orientation of the elongated cloud structures. At 2 and 4 hours, the forecast becomes smoother but continues to preserve the dominant diagonal organization and its gradual change in cloud index. The aggregate-power profile exhibits a modest initial increase followed by a gradual decline. PVCast follows this turning behavior with only a small difference between the predictive median and observation, while the observation remains within the predictive interval. This case illustrates that the rollout maintains coherent regional structure when recent cloud variation is limited.

The moderate case contains a clearer transition in both cloud coverage and intensity. The observed field is initially dominated by relatively high cloud-index values, which weaken and reorganize over the subsequent four hours. PVCast captures this domain-scale transition: the 0.5-hour forecast retains the initial cloud texture, while the 2- and 4-hour forecasts reproduce the broad reduction in cloud index and the redistribution of the major cloud regions. Although some fine-scale texture is smoothed at longer leads, the forecast retains the spatial gradients relevant to the systems distributed across the domain. The predictive median also follows the declining aggregate-power trajectory throughout the forecast window, with the observations covered by the pointwise interval. This agreement indicates that the evolving cloud representation remains compatible with the accompanying system context and solar conditions.

The variable case presents the strongest spatial evolution. The observed fields contain a pronounced boundary between clearer and cloudier regions, together with smaller cloud structures inside the clearer region. PVCast reproduces the location and orientation of this boundary at the 0.5-hour lead and maintains its large-scale displacement through the 2- and 4-hour forecasts. The model also captures the broad weakening of the cloud index over the forecast window, while naturally

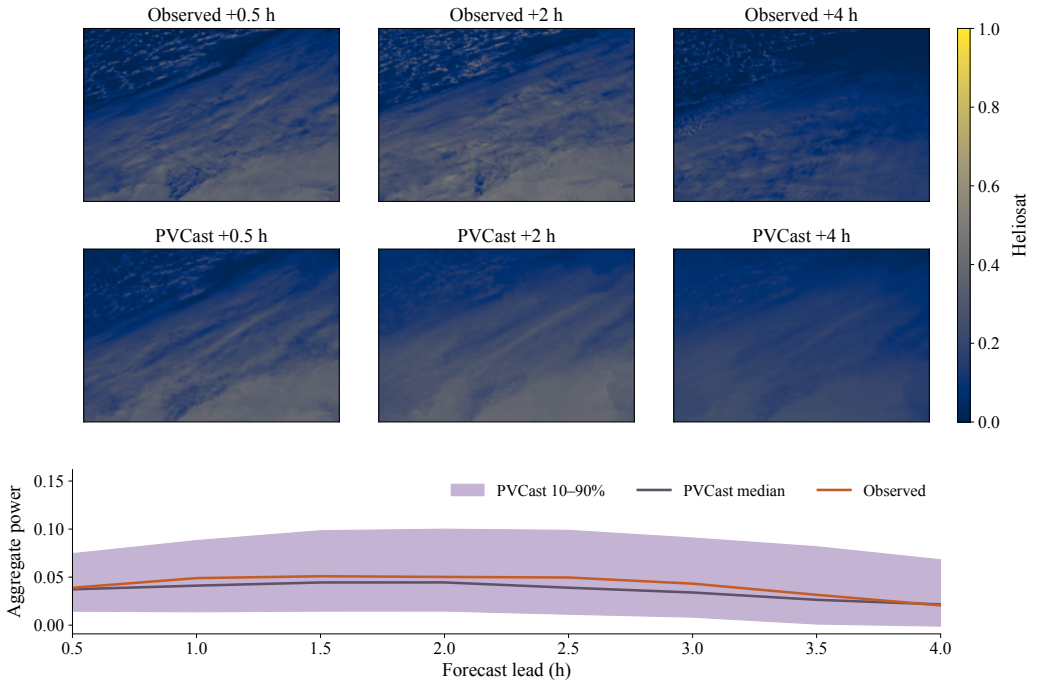

Figure 8: **Cloud-to-power forecasting under a stable cloud regime.**

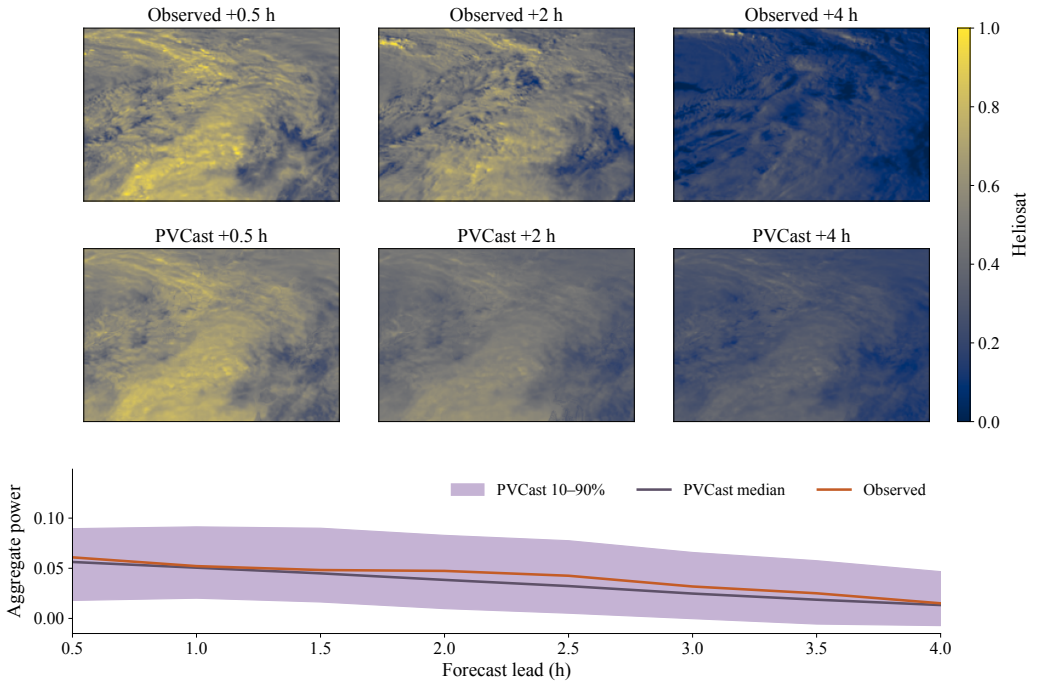

Figure 9: **Cloud-to-power forecasting under a moderate cloud regime.**

smoothing some of the smaller-scale structures whose evolution is less predictable. This case is accompanied by the largest change in aggregate power among the three examples. The predictive median closely follows the observed decline from its initial level toward zero, and the predictive interval covers the complete observed profile.

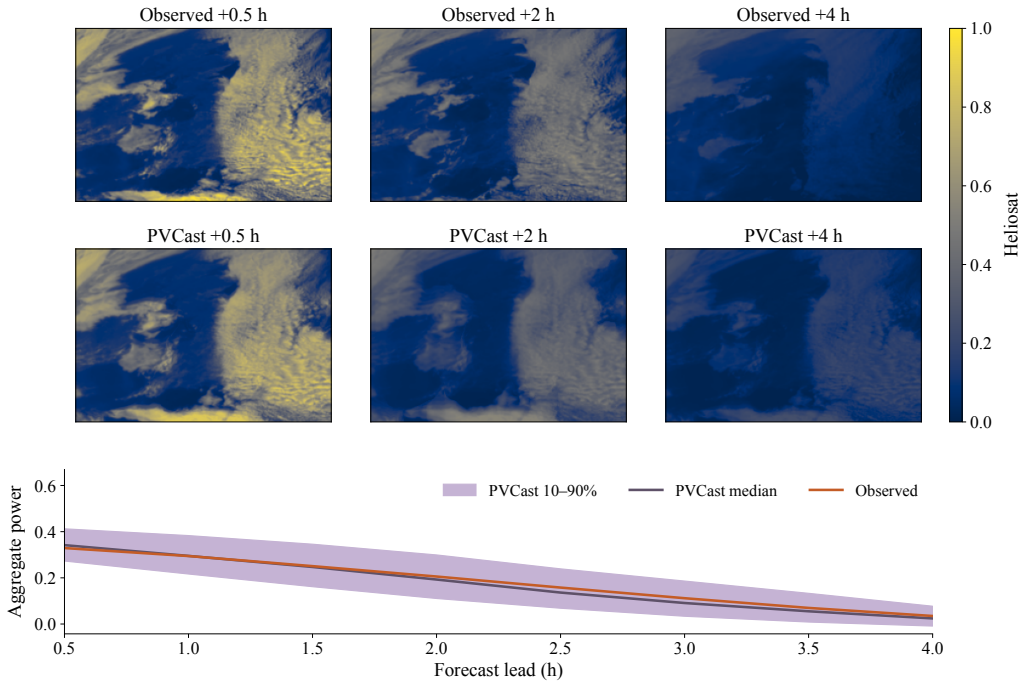

Figure 10: **Cloud-to-power forecasting under a variable cloud regime.**

Across the three regimes, the short-lead agreement is consistent with the one-step transport supervision, whereas the preservation of large-scale organization at 2 and 4 hours reflects the combined transport and source–sink rollout. The accompanying power panels show how these cloud evolutions are propagated into aggregate predictive distributions. These qualitative examples therefore complement the regime-wise score reductions reported in the main results.

Site-level ramp coherence. Aggregate scores summarize network-level behavior but do not reveal whether a joint forecast reproduces the timing and coordination of power ramps at individual systems. Nearby systems may ramp almost simultaneously under a common regional cloud pattern, or at different times as changing cloud conditions progress across the region. A joint forecast should represent both dependence patterns while retaining the distinct power trajectory of each system. Figures 11 and 12 examine these two cases using geographically nearby UKPV systems. In each figure, the map shows the system locations, while the remaining panels compare the observed power with the PVCast median and pointwise 10th–90th percentile interval.

In Figure 11, the six systems exhibit a broadly synchronous reduction in generation. Although their initial power levels and short-term fluctuations differ, the main decline occurs over a similar part of the forecast horizon. PVCast reproduces this shared transition without forcing identical site-level profiles: its medians follow the system-specific observations, while the predictive intervals accommodate local deviations around the regional event.

Figure 12 presents a different dependence pattern: the decline progresses across the region rather than occurring simultaneously. The ordering of the site-level ramps is consistent with their spatial arrangement, producing distinct transition times across the six systems. PVCast captures this staggered evolution by shifting the predicted decline between locations while preserving a coherent regional event.

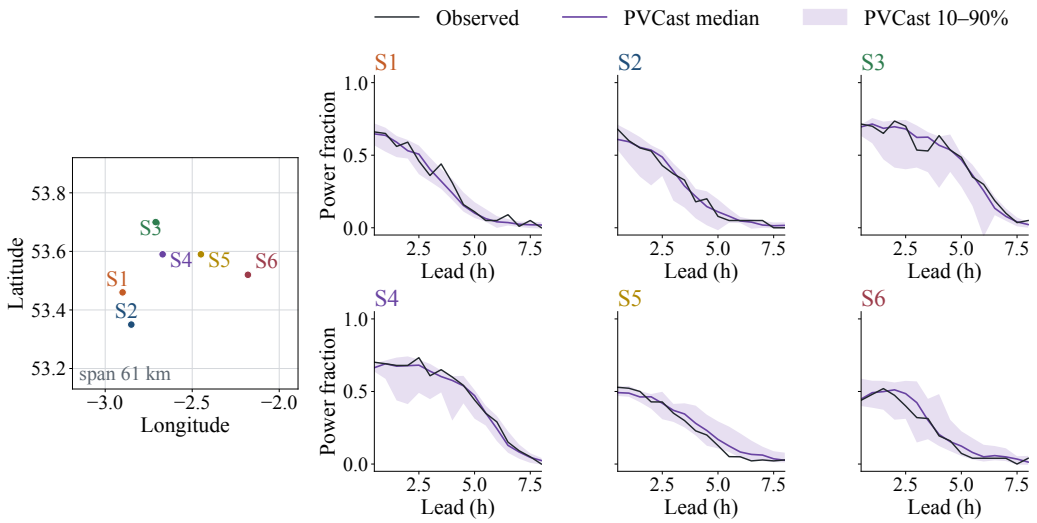

Figure 11: **Synchronous ramps across nearby PV systems.** The six systems span approximately 61 km and undergo a common decline in power.

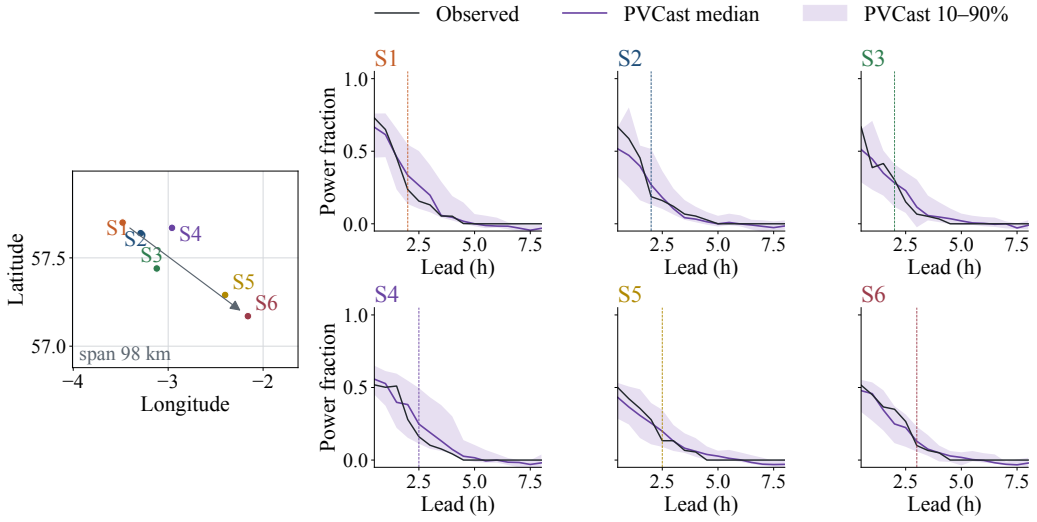

Figure 12: **Spatially lagged ramps across nearby PV systems.** The six systems span approximately 98 km and experience the power decline at different forecast leads. Dashed markers indicate the identified site-level ramp times, and the map summarizes their spatial progression.