SOLARSPECNET: PHASE-GUIDED STRUCTURE-INTENSITY MODELING FOR SHORT-TERM SOLAR RADIATION FIELD FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Short-term solar-radiation forecasting is essential for photovoltaic integration, power scheduling, and environmental analysis. However, most current methods either focus on station-level time-series prediction or treat gridded radiation fields as generic image sequences. Consequently, they fail to explicitly model two tightly coupled aspects of future radiation-field evolution. 1) Where will regions of high and low radiation emerge? 2) How will the radiation intensity at each location change over time? To address this limitation, we propose **SolarSpecNet**, which explicitly disentangles these two aspects for specialized modeling while preserving their interaction through spectral-spatial fusion. Specifically, a Structural Evolution Branch exploits Fourier phase information to forecast the spatial evolution of radiation patterns, while a Radiative Energy Branch captures radiation-intensity dynamics from both spatial-domain field values and frequency-domain representations. Then, a Spectral-Spatial Fusion Module integrates these complementary predictions through spectral recombination and spatial refinement to produce the final radiation-field forecast. Experiments on gridded radiation fields show that SolarSpecNet consistently improves multi-step forecasting over representative statistical, recurrent, convolutional, spectral, and foundation time-series baselines. Additional visualizations show that its forecasts evolve smoothly, retain clear day–night boundaries, and preserve the main radiation patterns over time. Codes are attached.

1 INTRODUCTION

Solar radiation is a key driver of photovoltaic generation, grid balancing, and environmental decision making. Accurate short-term forecasts are therefore important for renewable-energy integration and risk-aware scheduling. Recent research has moved beyond purely statistical point forecasting toward interpretable deep models, physics-informed predictors, sky-image systems, multimodal fusion, and AI-weather-assisted forecasting Yang et al. (2018); Li et al. (2025); Peng et al. (2026); Ansong et al. (2025); Abad-Alcaraz et al. (2025); Yan et al. (2025); Jia et al. (2026). Despite this progress, most methods still predict a station-level series, a local visual target, or photovoltaic power at a small set of sites. They do not directly address continuous radiation-field forecasting, where a model must track the movement of illumination patterns while estimating irradiance at every grid point.

ERA5 surface radiation provides a temporally consistent gridded view of this evolution and supports short-term field forecasting over broad regions Hersbach et al. (2020); Ruan et al. (2026). Unlike a one-dimensional series, a radiation field has explicit spatial organization, including day–night boundaries, latitudinal gradients, and region-dependent attenuation. Unlike a generic image sequence, each pixel has direct physical meaning because it records radiative magnitude. A suitable model should therefore preserve the motion and deformation of radiation structures while accurately predicting their location-dependent intensities.

Deep spatiotemporal models provide an important foundation. ConvLSTM, PredRNN, PhyDNet, TAU, and SimVP learn field dynamics through recurrent, disentangled, attention-based, or convolutional mechanisms Shi et al. (2015); Wang et al. (2017); Le Guen & Thome (2020); Tan et al. (2023); Gao et al. (2022a). Earthformer, FourCastNet, GraphCast, Pangu-Weather, and newer geometric or

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.

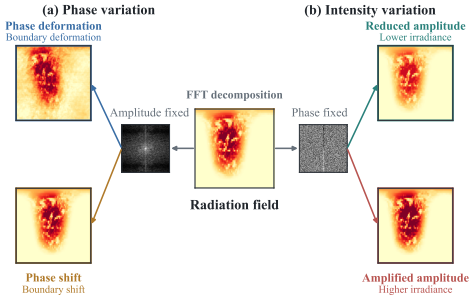


Figure 1: Fourier phase mainly determines radiation-field layout, whereas amplitude reflects radiation strength.

Fourier variants further demonstrate the potential of data-driven gridded weather forecasting Gao et al. (2022b); Pathak et al. (2022); Lam et al. (2023); Bi et al. (2023); Bonev et al. (2025); Lin et al. (2026). Frequency-domain operator learning offers a complementary perspective. The Fourier Neural Operator (FNO) learns mappings between functions through Fourier-space parameterization; the Adaptive Fourier Neural Operator (AFNO) performs efficient spectral token mixing; the Wavelet Neural Operator (WNO) introduces multi-resolution representations; and the Information-Coupled Neural Operator (ICNO) combines cross-frequency and spectral-spatial interactions Li et al. (2021); Guibas et al. (2022); Tripura & Chakraborty (2023); Chen et al. (2026); Azizzadenesheli et al. (2024). However, these general frameworks do not explicitly separate two coupled properties of a solar radiation field: spatial structure, which determines where radiation patterns appear and move, and radiative magnitude, which determines the values required by downstream energy applications.

This observation motivates **SolarSpecNet**. Following the classical relationship between Fourier phase and spatial structure, and between amplitude and spectral energy Oppenheim & Lim (1981), we design a spectral-spatial architecture that gives these two aspects distinct but interacting roles. As illustrated in Figure 1, phase-related information preserves the main radiation layout, whereas amplitude-related information reflects intensity distribution. SolarSpecNet therefore uses a Structural Evolution Branch for phase-guided spatial organization, a Radiative Energy Branch for intensity estimation, and a Spectral-Spatial Fusion Module that reconstructs and refines the final field.

Our contributions are threefold. First, we formulate short-term solar radiation forecasting as a continuous field prediction problem requiring both spatial consistency and magnitude accuracy. Second, we introduce a spectral-spatial architecture that explicitly couples phase-guided structural evolution with amplitude-related radiative-energy modeling. Third, experiments against representative statistical, recurrent, convolutional, spectral, and foundation time-series baselines show that SolarSpecNet produces accurate and coherent multi-step radiation forecasts.

2 Related Work

2.1 Solar Radiation Forecasting

Solar radiation has been forecast with statistical time-series methods, clear-sky correction, machine-learning regression, and deep temporal models Box et al. (2015); Hochreiter & Schmidhuber (1997); Yang et al. (2018); Voyant et al. (2025). Recent studies improve nonlinear forecasting through hybrid convolutional–recurrent designs, interpretable multi-scale decomposition, and physics-informed learning with satellite-derived cloud information İrem Fatma Şener & İhsan Tuğal (2025); Li et al. (2025); Peng et al. (2026). Other work incorporates sky images, multimodal measurements, satellite products, reanalysis fields, AI-weather outputs, or language-model prompts to supply richer atmospheric and spatial context Ansong et al. (2025); Abad-Alcaraz et al. (2025); Yan et al. (2025); Hersbach et al. (2020); Ruan et al. (2026); Jia et al. (2026). These advances mainly target point irradiance, local nowcasting, or photovoltaic power. Our work instead forecasts a gridded radiation field and jointly models its spatial organization and radiative intensity.

2.2 SPATIOTEMPORAL FORECASTING

Surface solar radiation is a structured geophysical field and is therefore related to spatiotemporal prediction and weather forecasting. ConvLSTM, PredRNN, PhyDNet, TAU, and SimVP model field evolution using convolutional recurrence, spatiotemporal memory, disentangled dynamics, temporal attention, or simplified convolutional mixing Shi et al. (2015); Wang et al. (2017); Le Guen & Thome (2020); Tan et al. (2023); Gao et al. (2022a). Earthformer, FourCastNet, GraphCast, and Pangu-Weather scale data-driven prediction to global grids, while FourCastNet 3 and the Fourier Advection Neural Operator (FANO) introduce geometric and advection-aware spectral designs Gao et al. (2022b); Pathak et al. (2022); Lam et al. (2023); Bi et al. (2023); Bonev et al. (2025); Lin et al. (2026). Recent work also studies predictive representation learning, spatiotemporal Fourier operators, and foundation forecasting models Nie et al. (2023); Nag et al. (2026); Ansari et al. (2025). These methods motivate our setting, but SolarSpecNet is tailored to radiation fields whose evolution couples illumination geometry, regional attenuation, and radiation magnitude.

2.3 SPECTRAL AND FREQUENCY-DOMAIN MODELING

Spectral representations are widely used for global correlation modeling in scientific machine learning. FNO parameterizes operators in Fourier space, AFNO adapts spectral mixing to tokenized representations, and WNO uses wavelets to capture multiple resolutions Li et al. (2021); Guibas et al. (2022); Tripura & Chakraborty (2023); Azizzadenesheli et al. (2024). Recent neural-operator variants improve operator learning through factorized Fourier representations, cross-frequency coupling, and transformer-based continuous operator architectures Tran et al. (2023); Chen et al. (2026); Wang et al. (2025). Practical analyses and local–global hybrids further clarify how spectral parameterization can be combined with spatial feature extraction Duruisseaux et al. (2026); Liu et al. (2025). SolarSpecNet differs from these generic spectral mixers by assigning explicit forecasting roles to phase-guided structure, amplitude-related energy, and their spectral-spatial recombination.

3 METHODOLOGY

3.1 PROBLEM FORMULATION

We study short-term forecasting of surface solar radiation fields. Let $\mathbf{S}_t \in \mathbb{R}^{C\times H\times W}$ denote the radiation field at valid time t, where $C = 1$ denotes a single scalar radiation channel in this work. Given T_{in} historical fields $\mathbf{X}_t = [\mathbf{S}_{t-T_{\text{in}}+1}, \ldots, \mathbf{S}_t] \in \mathbb{R}^{T_{\text{in}}\times C\times H\times W}$, the goal is to predict T_{out} future fields $\mathbf{Y}_t = [\mathbf{S}_{t+1}, \ldots, \mathbf{S}_{t+T_{\text{out}}}] \in \mathbb{R}^{T_{\text{out}}\times C\times H\times W}$. The forecasting model is written simply as

$$\widehat{\mathbf{Y}}_t = f_\theta(\mathbf{X}_t). \tag{1}$$

Unlike generic image-sequence prediction, each grid value in a solar radiation field is physically meaningful, and the field structure is strongly affected by solar geometry, diurnal illumination, atmospheric transmission, and regional radiative heterogeneity. SolarSpecNet is therefore designed to answer two questions at the same time: how the spatial layout of the radiation field evolves, and how irradiance changes.

3.2 CONSTRUCTION OF CONTINUOUS RADIATION FIELD

We construct continuous radiation fields based on ERA5 reanalysis records. The raw records are indexed by forecast initialization time, lead step, latitude, and longitude. Since that the same physical valid time may be reached from different initialization-step combinations, we reorganize the records by valid time, sort all fields chronologically, and retain a single field for duplicated valid times. Each field is then downsampled or resized to a fixed $H \times W$ grid using area interpolation, yielding a temporally aligned radiation sequence with consistent spatial resolution.

Before windowing, the valid-time-ordered sequence is divided chronologically into non-overlapping training, validation, and test intervals. Sliding windows with stride $\Delta = 1$ are generated independently within each interval, and boundary-crossing windows are discarded, so no hourly field is shared across subsets. For each starting index n, the first T_{in} frames form $\mathbf{X}_n$ and the follow-

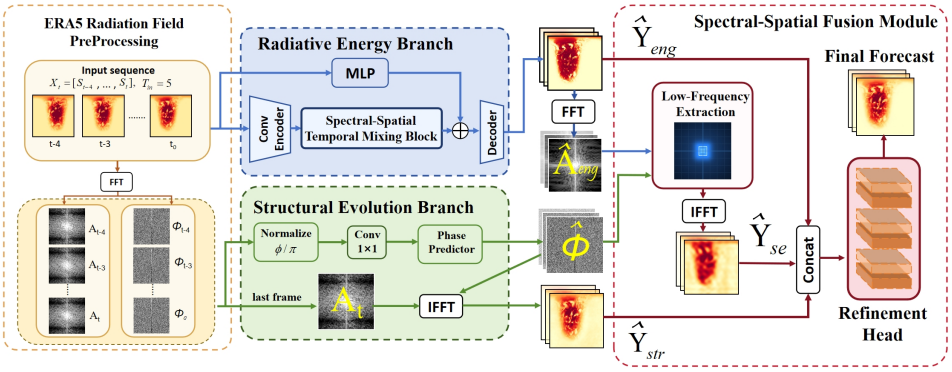


Figure 2: The overall SolarSpecNet flowchart for multi-step radiation-field forecasting, consisting mainly of phase-guided structural evolution, radiative-energy prediction, and spectral-spatial fusion.

ing T_{out} frames form $\mathbf{Y}_n$. Z-score statistics are computed only from the training interval, with $\overline{\mathbf{S}}_t = (\widetilde{\mathbf{S}}_t - \mu_{\text{train}})/(\sigma_{\text{train}} + \epsilon)$. Samples are stored in HDF5 as $(T_{\text{in}} + T_{\text{out}}, 1, H, W)$ tensors.

3.3 Spectral Decomposition

A radiation field contains large-scale components related to diurnal illumination and solar-geometry gradients, as well as local components caused by atmospheric transmission and regional heterogeneity. For a normalized field $\overline{\mathbf{S}}_t \in \mathbb{R}^{C\times H\times W}$, each grid value $\overline{\mathbf{S}}_t(h,w)$ represents the radiation magnitude at a spatial location, while its Fourier coefficient at frequency (u,v) is written as

$$\widehat{\mathbf{S}}_t(u,v) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} \overline{\mathbf{S}}_t(h,w) e^{-2\pi \mathrm{i}\left(\frac{uh}{H}+\frac{vw}{W}\right)}. \tag{2}$$

We then decompose the spectrum into amplitude and phase:

$$\begin{aligned} \mathbf{A}_t &= |\widehat{\mathbf{S}}_t|, \qquad \mathbf{\Phi}_t = \mathrm{Arg}(\widehat{\mathbf{S}}_t), \\ \overline{\mathbf{S}}_t &= \mathrm{Re}\left[\mathcal{F}^{-1}\left(\mathbf{A}_t \odot e^{\mathrm{i}\mathbf{\Phi}_t}\right)\right]. \end{aligned} \tag{3}$$

In this representation, amplitude describes the strength distribution of spectral components and is related to radiative magnitude, whereas phase determines the spatial arrangement of these components and is important for preserving illumination boundaries, latitudinal gradients, and regional layouts. To make the spectral position explicit, we also construct two fixed normalized frequency-coordinate maps $\mathbf{U}$ and $\mathbf{V}$ following the FFT frequency indexing. They indicate where each spectral component is located in the horizontal and vertical frequency axes, and are used as positional cues for phase evolution. We further use a low-frequency region $\Omega_k = \{(u,v) \mid |u| \le k, |v| \le k\}$ for auxiliary spectral supervision, because stable large-scale organization is dominated by low-frequency components.

3.4 SolarSpecNet Architecture

As shown in Figure 2, SolarSpecNet takes a normalized historical radiation sequence $\mathbf{X}_t \in \mathbb{R}^{T_{\text{in}}\times C\times H\times W}$ as input and produces a multi-step forecast $\widehat{\mathbf{Y}}_t \in \mathbb{R}^{T_{\text{out}}\times C\times H\times W}$. The forward process follows the same logic as the architecture diagram. For each historical frame, the model first obtains its Fourier amplitude and phase. The phase sequence is sent to the Structural Evolution Branch to predict future field organization, while the original radiation sequence is sent to the Radiative Energy Branch to estimate future intensity. These two branches do not produce the final forecast independently. Instead, they provide complementary intermediate views: a structure-oriented spatial prediction, an energy-oriented spatial prediction, and a recombined low-frequency spectral prediction. The Spectral-Spatial Fusion Module then integrates these views and outputs the final radiation fields.

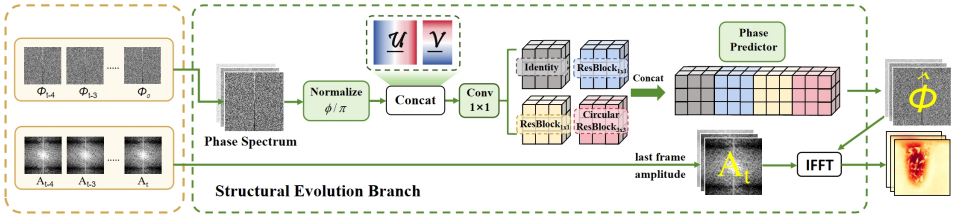


Figure 3: Detailed architecture of the SEB. Historical phase maps are normalized and augmented with fixed spectral-coordinate maps, processed by multiscale residual transformations and the phase predictor, and recombined with the latest observed amplitude through an inverse Fourier transform to obtain the structure-oriented forecast.

Structural Evolution Branch (SEB). As detailed in Figure 3, SEB is responsible for the geometric part of the forecast. The concerned input is the Fourier phase sequence extracted from the historical radiation fields. Since phase values alone do not tell the model which spatial frequency each value belongs to, we augment the normalized phase with two fixed spectral-coordinate maps. This gives the branch both the phase state and the frequency position of each spectral component:

$$\begin{aligned} \mathbf{Q}_\phi &= \mathrm{Concat}\left[\frac{\mathbf{\Phi}_{t-T_{\mathrm{in}}+1:t}}{\pi}, \mathbf{U}, \mathbf{V}\right], \\ \widehat{\mathbf{\Phi}}_{t+1:t+T_{\mathrm{out}}} &= B_\phi\left(\mathbf{Q}_\phi\right). \end{aligned} \tag{4}$$

Here B_ϕ denotes the follow-up SEB, operating on the temporal sequence of phase maps and projecting the historical structural state to future lead steps. In other words, B_ϕ learns how the phase pattern evolves over time rather than how bright each grid point should be. This makes the branch suitable for modeling the displacement of large-scale illumination boundaries, the movement of latitude-related gradients, and the coherent evolution of regional radiation layouts.

The output of SEB is a future phase sequence $\widehat{\mathbf{\Phi}}_{t+1:t+T_{\mathrm{out}}}$. To make this structural prediction usable by the later fusion module, we then convert it back to the spatial domain. Specifically, the predicted phase is paired with the latest observed amplitude $\mathbf{A}_t$, which provides a stable magnitude anchor from the most recent radiation field. The inverse Fourier transform then gives a phase-guided auxiliary forecast:

$$\widehat{\mathbf{Y}}_{\mathrm{str}} = \mathrm{Re}\left[\mathcal{F}^{-1}\left(\mathbf{A}_t \odot e^{\mathrm{i}\widehat{\mathbf{\Phi}}_{t+1:t+T_{\mathrm{out}}}}\right)\right]. \tag{5}$$

This auxiliary forecast is not used as the final answer. The expected role is to provide an explicit spatial estimate of where future radiation structures are likely to appear, while leaving the accurate estimation of radiation magnitude to the energy branch and the fusion module.

Radiative Energy Branch (REB). As detailed in Figure 4, the function of REB is to estimate the magnitude component, with input being the original normalized radiation sequence rather than only the amplitude spectrum. This choice is important since the grid values directly encode surface radiation intensity, while the spectrum only provides a transformed view of its spatial distribution. The branch therefore starts from the physical field representation and learns an intensity-oriented evolution path.

REB contains three conceptual operations. First, a residual spatial encoder maps each historical radiation field into hidden features, preserving local intensity patterns and regional attenuation information. Second, a temporal projection maps the encoded history from T_{in} input to T_{out} future steps, so that the branch produces a feature sequence rather than a single-step estimate. Third, the projected features are refined by combining broad spatial-frequency mixing with local convolutional correction. The frequency-side operation helps capture large-scale energy redistribution, such as the expansion or weakening of illuminated regions, while the grid-side refinement keeps local radiation values and regional variations consistent.

Given these operations, a decoder maps the features back to the radiation-field space and produces an energy-oriented auxiliary forecast $\widehat{\mathbf{Y}}_{\mathrm{eng}}$. This output has the same spatial and temporal shape

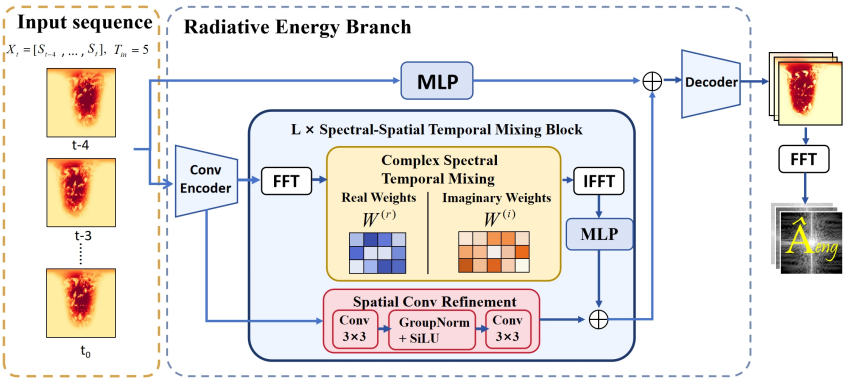


Figure 4: Detailed architecture of the REB. The historical radiation sequence is encoded and processed by repeated spectral-spatial temporal mixing blocks, in which complex spectral temporal mixing captures global evolution and spatial convolutions refine local details. The decoder produces the energy-oriented forecast, from which the predicted amplitude spectrum is extracted.

as the target sequence and can be interpreted as the estimate of future radiation intensity. We then compute its Fourier amplitude as the predictions:

$$\widehat{\mathbf{A}}_{\text{eng}} = \left| \mathcal{F}\left(\widehat{\mathbf{Y}}_{\text{eng}}\right) \right|. \tag{6}$$

Thus, REB provides two pieces of information: a spatial-domain energy forecast that describes future radiation values, and a spectral-domain amplitude estimate that describes how radiation energy is distributed across spatial frequencies.

Spectral-Spatial Fusion Module (SSF). SSF is used to connect the two branches, which receives the structure-oriented auxiliary forecast $\widehat{\mathbf{Y}}_{\text{str}}$, the energy-oriented auxiliary forecast $\widehat{\mathbf{Y}}_{\text{eng}}$, and a low-frequency spectral reconstruction. Note the spectral reconstruction is formed by pairing the predicted amplitude from REB with the predicted phase from SEB. Technically, we apply a low-frequency mask $\mathbf{M}_k$ during this operation:

$$\widehat{\mathbf{Y}}_{\text{se}} = \text{Re}\left[\mathcal{F}^{-1}\left(\mathbf{M}_k \odot \widehat{\mathbf{A}}_{\text{eng}} \odot e^{\text{i}\widehat{\mathbf{\Phi}}_{t+1:t+T_{\text{out}}}} \right) \right]. \tag{7}$$

This reconstruction explicitly links "where the radiation pattern moves" with "how strong the radiation field is" at the dominant spatial scales. The mask suppresses unstable high-frequency components in this coupling step, so the model first aligns the large-scale structure and energy distribution before relying on spatial refinement for local details.

The final fusion is performed in the spatial domain. SSF concatenates the structural forecast, energy component, and low-frequency reconstructed field along the channel dimension. These three inputs play different roles: $\widehat{\mathbf{Y}}_{\text{str}}$ emphasizes layout and boundary placement, $\widehat{\mathbf{Y}}_{\text{eng}}$ predicts radiation magnitude, and $\widehat{\mathbf{Y}}_{\text{se}}$ provides an explicitly recombined spectral-spatial prior. A lightweight convolutional mixer then compares and refines these complementary views, correcting local inconsistencies and producing the final prediction:

$$\widehat{\mathbf{Y}}_t = R\Big(\text{Concat}\left[\widehat{\mathbf{Y}}_{\text{str}}, \widehat{\mathbf{Y}}_{\text{eng}}, \widehat{\mathbf{Y}}_{\text{se}}\right]\Big). \tag{8}$$

To sum up, the outcome is not produced by simply a single black-box image predictor. Instead, it is obtained through an explicit route: phase-guided structural evolution, intensity-oriented energy estimation, low-frequency spectral recombination, and spatial refinement.

Training Objective. SolarSpecNet is trained end-to-end with paired historical-and-future radiation sequences. During each iteration, the input sequence is passed through two branches and the fusion module to obtain the final prediction $\widehat{\mathbf{Y}}_t$, while the structural and energy branches also produce their auxiliary outputs. The main supervision is applied to the final forecast in the physical grid

space. This term directly optimizes the quantity required for solar-radiation forecasting: the future radiation value at each grid location and horizon.

Auxiliary supervision is used to make the intermediate outputs match their intended roles. The energy branch is supervised in the spatial domain by comparing $\widehat{\mathbf{Y}}_{\mathrm{eng}}$ with the future radiation fields, and in the frequency domain by comparing the amplitude spectrum of the energy forecast with the target spectrum. The structural branch is supervised with low-frequency phase consistency. We use a phase consistency term rather than a direct regression penalty because phase is periodic; the loss is therefore computed in a way that encourages angular agreement in the stable low-frequency region.

The total training objective is

$$\begin{aligned} \mathcal{L} =& \mathcal{L}_{\mathrm{pred}} + \lambda_{\mathrm{eng}}\mathcal{L}_{\mathrm{eng}} + \lambda_{\mathrm{amp}}\mathcal{L}_{\mathrm{amp}} + \lambda_{\phi}\mathcal{L}_{\phi}, \\ \mathcal{L}_{\mathrm{pred}} =& \|\widehat{\mathbf{Y}}_t - \mathbf{Y}_t\|_2^2, \quad \mathcal{L}_{\mathrm{eng}} = \|\widehat{\mathbf{Y}}_{\mathrm{eng}} - \mathbf{Y}_t\|_2^2. \end{aligned} \tag{9}$$

Here $\mathcal{L}_{\mathrm{amp}}$ measures amplitude-spectrum consistency between the energy forecast and the target, and $\mathcal{L}_{\phi}$ measures low-frequency phase consistency between the structural prediction and the target. They are defined as

$$\begin{aligned} \mathcal{L}_{\mathrm{amp}} &= \left\| \widehat{\mathbf{A}}_{\mathrm{eng}} - |\mathcal{F}(\mathbf{Y}_t)| \right\|_2^2, \\ \mathcal{L}_{\phi} &= \frac{1}{|\Omega_k|} \sum_{(u,v)\in\Omega_k} \left[1 - \cos\left(\widehat{\mathbf{\Phi}}(u,v) - \mathbf{\Phi}^{Y}(u,v) \right) \right]. \end{aligned} \tag{10}$$

where $\mathbf{\Phi}^{Y} = \mathrm{Arg}(\mathcal{F}(\mathbf{Y}_t))$ is the target phase spectrum. The amplitude-spectrum weight is gradually reduced during early training, so it helps stabilize the energy branch at the beginning but does not dominate the final grid-space forecast. At inference time, the model follows the same forward route from the historical radiation fields to the final prediction, but no target fields or auxiliary losses are required.

4 Experiments

4.1 Experimental Setup

Dataset. We use valid-time-ordered surface radiation fields at 128×128 resolution. A chronological split before sliding-window construction yields 2,138, 118, and 120 training, validation, and test samples, respectively; each sample contains 5 historical and 20 future hourly fields. Training-only statistics are used for normalization, and all metrics are reported in Z-score-normalized space. Full preprocessing and split details are provided in the appendix.

Baselines. We compare our proposal against ARIMA Box et al. (2015), LSTM Hochreiter & Schmidhuber (1997), ConvLSTM Shi et al. (2015), SimVP Gao et al. (2022a), ICNO Chen et al. (2026), and Chronos-2 Ansari et al. (2025), covering statistical, recurrent, convolutional, spectral, and foundation time-series forecasting paradigms.

Implementation details. SolarSpecNet uses hidden width 48, three energy-evolution blocks, bandwidth $k = 10$, and AdamW with batch size 4 and learning rate 5×10^{-5}. We set $(\lambda_{\mathrm{eng}}, \lambda_{\mathrm{amp}}, \lambda_{\phi}) = (0.05, 0.01, 0.01)$ and anneal the amplitude term during early training. All baselines follow the same input-output and normalization protocol; complete settings appear in the appendix.

Evaluation metrics. We report MSE, MAE, RMSE, PSNR, SSIM, and LPIPS; arrows in the tables indicate the preferred direction.

4.2 Quantitative Comparison

SolarSpecNet leads the 20-step averages in Table 1 for MSE, MAE, RMSE, and PSNR. Against Chronos-2, SimVP, and ICNO, it reduces MSE by 48.9%, 46.1%, and 68.7%; MAE by 37.1%,

36.1%, and 71.2%; and RMSE by 28.5%, 27.7%, and 44.3%, while improving PSNR by 2.83, 2.98, and 5.18 dB, respectively. These consistent gains confirm its advantage in point-wise radiation accuracy.

ICNO obtains the best SSIM (0.8964) and LPIPS (0.1870), while SolarSpecNet ranks second (0.8361 and 0.2503). Compared with Chronos-2, SolarSpecNet also improves SSIM by 8.7% and reduces LPIPS by 24.0%. Thus, its primary strength lies in point-wise radiation accuracy, complemented by solid, albeit not uniformly superior, structural fidelity.

Model	MSE↓	MAE↓	RMSE↓
ARIMA	0.2834	0.3323	0.5160
LSTM	0.0972	0.1596	0.2993
ConvLSTM	0.0818	0.1358	0.2745
SimVP	<u>0.0210</u>	<u>0.0486</u>	<u>0.1441</u>
ICNO	0.0361	0.1080	0.1870
Chronos-2	0.0221	0.0494	0.1456
SolarSpecNet	**0.0113**	**0.0311**	**0.1041**

Model	PSNR↑	SSIM↑	LPIPS↓
ARIMA	6.1802	0.3886	0.5823
LSTM	10.8781	0.5749	0.3941
ConvLSTM	11.6091	0.6104	0.4357
SimVP	16.8848	0.7626	0.4244
ICNO	14.6907	**0.8964**	**0.1870**
Chronos-2	<u>17.0383</u>	0.7695	0.3291
SolarSpecNet	**19.8660**	<u>0.8361</u>	<u>0.2503</u>

Table 1: Average performance over 20 lead steps in Z-score-normalized space. Bold and underline denote the best and second-best results.

Figure 5a further shows that SolarSpecNet maintains the lowest MAE and RMSE at every lead step, while its SSIM remains consistently second only to ICNO.

4.3 Qualitative Results

Figure 5b shows the first five forecast hours on a globe projection. High-radiation regions move smoothly across the visible hemisphere and the day–night transition remains temporally coherent, indicating that the model learns continuous field evolution rather than independent frame-wise outputs. The boundary shape changes gradually from one hour to the next, without isolated bright regions or abrupt discontinuities that would be inconsistent with short-term solar geometry.

Across the displayed horizons, the illuminated footprint translates and deforms as the boundary advances, while intensity within illuminated areas changes more smoothly. The model also avoids frame-to-frame flicker, isolated hotspots, and visible map-seam discontinuities. Neighboring lead times retain common spatial support and gradual boundary motion, supporting a coherent forecast trajectory rather than unrelated images.

Together, the visualizations show coherent global propagation, regional trends, and large-scale spectral structure; discrepancies remain near sharp transitions and high-frequency components. Additional analysis appears in the appendix.

4.4 Ablation Study

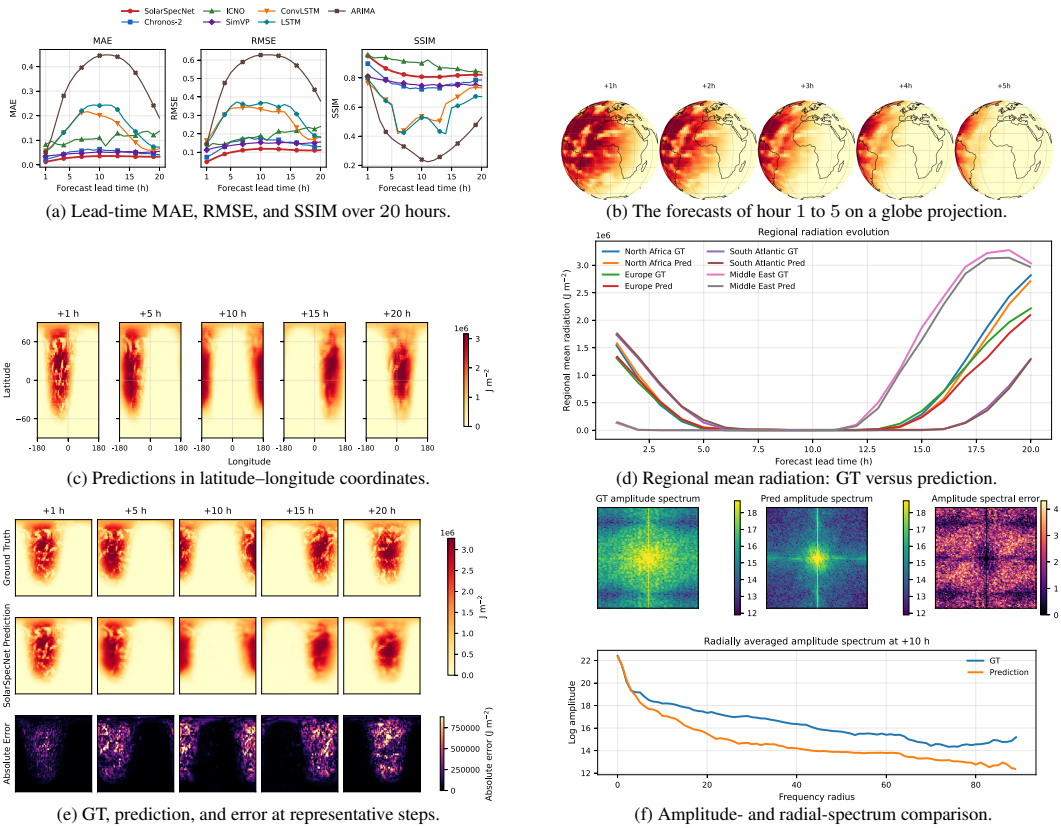
We conduct six ablations: three remove SEB, REB, or SSF, and three remove their auxiliary losses. In the fusion variant, both branches remain active, but amplitude and phase are combined by masked inverse Fourier transform without three-view residual refinement.

Variant	MSE↓	MAE↓	RMSE↓
Full Model	0.0113	0.0311	0.1041
w/o SEB	0.0122	0.0330	0.1083
w/o REB	0.0242	0.0524	0.1507
w/o SSF	0.1004	0.1842	0.3155
w/o $\mathcal{L}_{\text{amp}}$	0.0211	0.0481	0.1432
w/o $\mathcal{L}_{\text{eng}}$	0.0210	0.0484	0.1427
w/o $\mathcal{L}_{\phi}$	0.0200	0.0461	0.1394

Variant	PSNR↑	SSIM↑	LPIPS↓
Full Model	19.8660	0.8361	0.2503
w/o SEB	19.5304	0.8265	0.2775
w/o REB	16.7422	0.7578	0.3050
w/o SSF	10.0571	0.1420	0.7536
w/o $\mathcal{L}_{\text{amp}}$	17.0210	0.7708	0.3307
w/o $\mathcal{L}_{\text{eng}}$	17.0839	0.7718	0.3236
w/o $\mathcal{L}_{\phi}$	17.2643	0.7768	0.3279

Table 2: Ablation results averaged over 20 lead steps.

The full model performs best on every metric. Removing SSF and using direct masked reconstruction raises MSE from 0.0113 to 0.1004 and reduces SSIM from 0.8361 to 0.1420. Because this variant retains both predicted amplitude and phase, the degradation shows that spectral recombination alone cannot replace learned spatial refinement. Removing REB more than doubles MSE to



(a) Lead-time MAE, RMSE, and SSIM over 20 hours.

(b) The forecasts of hour 1 to 5 on a globe projection.

(c) Predictions in latitude–longitude coordinates.

(d) Regional mean radiation: GT versus prediction.

(e) GT, prediction, and error at representative steps.

(f) Amplitude- and radial-spectrum comparison.

Figure 5: Performance, qualitative, and diagnostic analyses of SolarSpecNet forecasts.

0.0242, confirming its central role in radiation-intensity prediction. Removing SEB causes a smaller MSE increase to 0.0122 but lowers SSIM to 0.8265 and raises LPIPS to 0.2775, indicating a stronger effect on spatial organization and perceptual fidelity.

Across the auxiliary-loss ablations, MSE increases to approximately 0.020–0.021, showing that amplitude, direct-field, and phase supervision all contribute. Removing phase consistency gives the lowest MSE among the three variants but still degrades SSIM and LPIPS, consistent with its structural role. Removing either amplitude-spectrum or direct energy supervision produces slightly larger regression errors. Detailed ablation definitions and additional interpretation are provided in the appendix.

4.5 Discussion

The results indicate complementary roles: REB contributes most to point-wise accuracy, SEB improves spatial coherence, and SSF reconciles intensity, phase-guided structure, and spectral reconstruction. Errors remain concentrated near moving boundaries, where small displacements affect SSIM and LPIPS more strongly than field-wise regression metrics. SolarSpecNet currently uses only historical radiation fields; incorporating cloud, solar-geometry, and atmospheric variables is a natural next step. Extended interpretation is provided in the appendix.

5 Conclusion

SolarSpecNet separates phase-guided structural evolution from radiative-intensity estimation and recombines them through learned spectral-spatial fusion. Experiments demonstrate strong point-wise accuracy and coherent spatial–spectral behavior over 20 forecast hours. Ablations confirm the complementary branches, learned fusion, and auxiliary losses. Boundary errors motivate future refinement with atmospheric inputs and multi-scale features.

AI USE STATEMENT

Generative AI tools were used to assist with language editing, LaTeX formatting, and manuscript organization. All AI-assisted content was reviewed and verified by the authors, who take full responsibility for the final manuscript, including its scientific claims, experimental results, and presentation.

REPRODUCIBILITY STATEMENT

Data preprocessing, implementation details, training configurations, and ablation definitions are provided in the appendix. Anonymous source code is included in the supplementary material.

REFERENCES

V. Abad-Alcaraz, M. Castilla, J. A. Carballo, J. Bonilla, and J. D. Álvarez. Multimodal deep learning for solar radiation forecasting. *Applied Energy*, 393:126061, 2025. doi: 10.1016/j.apenergy.2025.126061.

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C. Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. Chronos-2: From univariate to universal forecasting, 2025. URL https://arxiv.org/abs/2510.15821.

Martin Ansong, Gan Huang, Thomas N. Nyang'onda, Robinson J. Musembi, and Bryce S. Richards. Very short-term solar irradiance forecasting based on open-source low-cost sky imager and hybrid deep-learning techniques. *Solar Energy*, 294:113516, 2025. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2025.113516. URL https://www.sciencedirect.com/science/article/pii/S0038092X25002798.

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, April 2024. doi: 10.1038/s42254-024-00712-5. URL https://doi.org/10.1038/s42254-024-00712-5.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, July 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06185-3. PMID: 37407823.

Boris Bonev, Thorsten Kurth, Ankur Mahesh, Mauro Bisson, Jean Kossaifi, Karthik Kashinath, Anima Anandkumar, William D. Collins, Michael S. Pritchard, and Alexander Keller. Fourcastnet 3: A geometric approach to probabilistic machine-learning weather forecasting at scale, 2025. URL https://arxiv.org/abs/2507.12144.

George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 5th edition, 2015.

Yihan Chen, Wenbin Lu, Junnan Xu, Yuhang He, Wei Li, and Jianwei Zheng. Information-coupled neural operator for computational mechanics and parametric pdes. *Computer Methods in Applied Mechanics and Engineering*, 453:118851, 2026. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2026.118851. URL https://www.sciencedirect.com/science/article/pii/S0045782526001258.

Valentin Duruisseaux, Jean Kossaifi, and Anima Anandkumar. Fourier neural operators explained: A practical perspective, 2026. URL https://arxiv.org/abs/2512.01421.

Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z. Li. Simvp: Simpler yet better video prediction. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3160–3170, 2022a. URL https://api.semanticscholar.org/CorpusID:249605809.

Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Wang, Mu Li, and Dit-Yan Yeung. Earthformer: exploring space-time transformers for earth system forecasting. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022b. Curran Associates Inc. ISBN 9781713871088.

John Guibas, Morteza Mardani, Zongyi Li, Andrew Tao, Anima Anandkumar, and Bryan Catanzaro. Efficient token mixing for transformers via adaptive fourier neural operators. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=EXHG-A3jlM.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: https://doi.org/10.1002/qj.3803. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

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. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(45):38644–38652, March 2026. doi: 10.1609/aaai.v40i45.41207. URL https://ojs.aaai.org/index.php/AAAI/article/view/41207.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023. doi: 10.1126/science.adi2336. URL https://www.science.org/doi/abs/10.1126/science.adi2336.

Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11471–11481, 2020. doi: 10.1109/CVPR42600.2020.01149.

You Li, Weisheng Zhou, Yafei Wang, Sheng Miao, Wanxiang Yao, and Weijun Gao. Interpretable deep learning framework for hourly solar radiation forecasting based on decomposing multi-scale variations. *Applied Energy*, 377:124409, 2025. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2024.124409. URL https://www.sciencedirect.com/science/article/pii/S0306261924017926.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Kenghong Lin, Huiwei Lin, Baoquan Zhang, Yangjinxi Ge, Dong Liu, Xutao Li, Yunming Ye, and Chuyao Luo. FANO: Fourier Advection Neural Operator for Weather Prediction. *IEEE Transactions on Geoscience and Remote Sensing*, 64:TGRS.2026, January 2026. doi: 10.1109/TGRS.2026.3655726.

Chaoyu Liu, Davide Murari, Lihao Liu, Yangming Li, Chris Budd, and Carola-Bibiane Schönlieb. Enhancing fourier neural operators with local spatial features, 2025. URL https://arxiv.org/abs/2503.17797.

Pratik Nag, Andrew Zammit-Mangion, Sumeetpal Singh, and Noel Cressie. Spatio-temporal modeling and forecasting with fourier neural operators, 2026. URL https://arxiv.org/abs/2601.01813.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

A. V. Oppenheim and J. S. Lim. The importance of phase in signals. *IEEE Proceedings*, 69:529–541, May 1981. doi: 10.1109/PROC.1981.12022.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators, February 2022. URL https://doi.org/10.48550/arXiv.2202.11214.

Wanshan Peng, Tao He, Yichuan Ma, and Yueming Zheng. A physics-informed machine learning method for short-term solar radiation forecast with satellite-derived cloud optical depth. *Applied Energy*, 406:127284, 2026. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2025.127284. URL https://www.sciencedirect.com/science/article/pii/S0306261925020148.

Zhaohui Ruan, Lang Wu, Hongrong Shi, Meiqin Ni, Menghui Zhang, Weijun Zhu, Chunlin Huang, and Jiamin Chen. Spatiotemporal solar radiation forecasting driven by satellite-based and reanalysis data for distributed pv integration using fully-convolutional neural network. *Renewable Energy*, 256:124556, 2026. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2025.124556. URL https://www.sciencedirect.com/science/article/pii/S0960148125022207.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional lstm network: a machine learning approach for precipitation nowcasting. In *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, pp. 802–810, Cambridge, MA, USA, 2015. MIT Press.

Cheng Tan, Zhangyang Gao, Lirong Wu, Yongjie Xu, Jun Xia, Siyuan Li, and Stan Z. Li. Temporal attention unit: Towards efficient spatiotemporal predictive learning. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18770–18782, 2023. doi: 10.1109/CVPR52729.2023.01800.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tmIiMPl4IPa.

Tapas Tripura and Souvik Chakraborty. Wavelet neural operator for solving parametric partial differential equations in computational mechanics problems. *Computer Methods in Applied Mechanics and Engineering*, 404:115783, 2023. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2022.115783. URL https://www.sciencedirect.com/science/article/pii/S0045782522007393.

Cyril Voyant, Milan Despotovic, Gilles Notton, Yves-Marie Saint-Drenan, Mohammed Asloune, and Luis Garcia-Gutierrez. On the importance of clearsky model in short-term solar radiation forecasting. *Solar Energy*, 294:113490, 2025. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2025.113490. URL https://www.sciencedirect.com/science/article/pii/S0038092X25002531.

Sifan Wang, Jacob H Seidman, Shyam Sankaran, Hanwen Wang, George J. Pappas, and Paris Perdikaris. CVit: Continuous vision transformer for operator learning. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=cRnCcuLvyr.

Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and Philip S. Yu. Predrnn: recurrent neural networks for predictive learning using spatiotemporal lstms. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 879–888, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Zihuang Yan, Xianghui Lu, Lifeng Wu, Haina Zhang, Fa Liu, Xulei Wang, Wenhao Xu, and Wei Liu. Enhancing short-term solar radiation forecasting with hybrid vmd and graphcast-based machine learning models. *Expert Systems with Applications*, 285:128042, 2025. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2025.128042. URL https://www.sciencedirect.com/science/article/pii/S095741742501663X.

Dazhi Yang, Jan Kleissl, Christian A. Gueymard, Hugo T.C. Pedro, and Carlos F.M. Coimbra. History and trends in solar irradiance and pv power forecasting: A preliminary assessment and review using text mining. *Solar Energy*, 168:60–101, 2018. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2017.11.023. URL https://www.sciencedirect.com/science/article/pii/S0038092X17310022. Advances in Solar Resource Assessment and Forecasting.

İrem Fatma Şener and İhsan Tuğal. Optimized cnn-lstm with hybrid metaheuristic approaches for solar radiation forecasting. *Case Studies in Thermal Engineering*, 72:106356, 2025. ISSN 2214-157X. doi: https://doi.org/10.1016/j.csite.2025.106356. URL https://www.sciencedirect.com/science/article/pii/S2214157X25006161.

A Dataset Construction Details

Valid-time reconstruction. We construct the surface solar radiation sequence from ERA5 reanalysis records. Each record is indexed by forecast initialization time and forecast step, and different initialization–step pairs may refer to the same valid time. We first compute the valid time of every radiation field, flatten the initialization and step dimensions, and sort the fields chronologically. Duplicate timestamps are then removed by retaining one field for each valid time. The resulting hourly sequence is

$$\mathcal{S} = \{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_N\}, \qquad \mathbf{S}_t \in \mathbb{R}^{1\times H_0 \times W_0}, \tag{11}$$

where the temporal index follows valid time rather than forecast-cycle order.

Spatial resampling and normalization. All radiation fields are resampled to a common 128×128 latitude–longitude grid using area interpolation. This preserves the single-channel field representation and avoids the block artifacts introduced by nearest-neighbor resizing. Invalid numerical entries are replaced before normalization so that they do not propagate through the Fourier transform or the loss computation. The mean and standard deviation are computed from the training split only and are reused for validation and testing. All quantitative metrics are evaluated in the resulting Z-score-normalized space. Inverse normalization is applied only when visualizing radiation fields in physical units.

Sliding-window sample construction. We form forecasting samples with a sliding window of $T_{\text{in}} + T_{\text{out}} = 25$ hourly fields. For a window starting at index n, the first five fields are used as input and the following twenty fields as targets:

$$\begin{aligned} \mathbf{X}_n &= [\mathbf{S}_n, \ldots, \mathbf{S}_{n+4}], \\ \mathbf{Y}_n &= [\mathbf{S}_{n+5}, \ldots, \mathbf{S}_{n+24}]. \end{aligned} \tag{12}$$

The valid-time-ordered sequence is first divided into non-overlapping chronological training, validation, and test intervals. The window is then shifted by one hour independently within each interval, and windows crossing a split boundary are discarded. Therefore, no hourly field is reused across subsets. The resulting samples are stored in HDF5 format; during training, each tensor of shape $25 \times 1 \times 128 \times 128$ is separated into a five-frame input and a twenty-frame target.

B Model Implementation Details

Computing Environment. All experiments were conducted on a Linux server running Ubuntu 20.04.6 LTS, equipped with an Intel Xeon Silver 4210R CPU at 2.40 GHz, 187 GiB of system memory, and two NVIDIA GeForce RTX 3090 GPUs, each with 24 GB of GPU memory. The implementation was developed using Python 3.9.15 and PyTorch 2.4.1 with CUDA 12.1 and cuDNN 9.1.0. The NVIDIA driver version was 550.163.01.

SolarSpecNet follows the three-module architecture presented in the main paper. The hidden width is $D = 48$. Given an input tensor of shape $B \times 5 \times 1 \times 128 \times 128$, the Structural Evolution Branch (SEB) predicts twenty future phase maps, the Radiative Energy Branch (REB) predicts twenty future radiation fields, and the Spectral-Spatial Fusion Module (SSF) produces the final tensor of shape $B \times 20 \times 1 \times 128 \times 128$.

Structural Evolution Branch. A two-dimensional real Fourier transform is applied to each historical field, yielding amplitude and phase maps on the 128×65 half-spectrum. The five phase maps are normalized by π and concatenated with fixed horizontal and vertical frequency-coordinate maps. The input at each spectral location therefore contains both its phase history and its frequency position.

A 1×1 projection first predicts a coarse twenty-step phase residual. This estimate is refined by three parallel paths: two pointwise residual paths for location-wise corrections and one circularly padded 3×3 path for interactions between neighboring frequencies. Each path contains two residual blocks with width 48. Their outputs, together with the coarse prediction, are concatenated and projected back to the twenty-step phase representation. The predicted residual is added to the most recent observed phase. Pairing this phase with the latest observed amplitude and applying the inverse Fourier transform gives the structure-oriented auxiliary prediction $\widehat{\mathbf{Y}}_{\text{str}}$.

Radiative Energy Branch. REB operates on the normalized radiation fields in the spatial domain. Each input frame is lifted from one channel to 48 channels by two residual encoding blocks followed by a pointwise projection.

The encoded sequence is then processed by three energy-evolution blocks. Each block combines a grid-domain temporal projection with complex-valued temporal mixing in the spatial-frequency domain. The frequency-domain output is transformed back to the grid and added to the grid-domain prediction. Two 3×3 convolutions with normalization and nonlinear activation further refine the merged features. The first block expands the temporal length from 5 to 20, while the next two blocks retain the twenty-step length and refine the future representation.

A decoder with two residual blocks and a 1×1 projection maps the hidden features to a single radiation channel. After the bounded output activation, the branch produces the energy-oriented prediction $\widehat{\mathbf{Y}}_{\mathrm{eng}}$. Its Fourier magnitude, $\widehat{\mathbf{A}}_{\mathrm{eng}} = |\mathcal{F}(\widehat{\mathbf{Y}}_{\mathrm{eng}})|$, is used as the predicted amplitude in the fusion stage.

Spectral-Spatial Fusion Module. The fusion module first reconstructs a low-frequency amplitude–phase prediction. A mask with bandwidth $k = 10$ retains the dominant coefficients around the zero-frequency region. The retained amplitude from REB is combined with the phase from SEB:

$$\widehat{\mathbf{Y}}_{\mathrm{se}} = \mathrm{Re}\Big[\mathcal{F}^{-1}\Big(\mathbf{M}_{10} \odot \widehat{\mathbf{A}}_{\mathrm{eng}} \odot e^{\mathrm{i}\widehat{\mathbf{\Phi}}}\Big)\Big]. \tag{13}$$

This reconstruction transfers the predicted large-scale amplitude and phase into a common spatial representation.

For each forecast step, $\widehat{\mathbf{Y}}_{\mathrm{str}}$, $\widehat{\mathbf{Y}}_{\mathrm{eng}}$, and $\widehat{\mathbf{Y}}_{\mathrm{se}}$ are concatenated along the channel dimension. Two residual blocks with width 48 refine the three inputs, and a final 1×1 projection produces the single-channel forecast. The same fusion parameters are shared across all twenty lead times.

Optimization settings. Table 3 lists the training configuration. We use AdamW for at most 20 epochs with batch size 4, initial learning rate 5×10^{-5}, momentum coefficients $(0.9, 0.95)$, and weight decay 10^{-5}. The learning rate follows a cosine schedule after 500 warm-up steps. An exponential moving average with decay 0.95 is used for evaluation. The main prediction loss has weight 1. The energy auxiliary, amplitude-spectrum, and phase-consistency losses use weights 0.05, 0.01, and 0.01, respectively. The amplitude-spectrum weight is linearly annealed to zero over the first 2000 optimization steps; the other loss weights remain fixed.

Setting	Value
Input / output length	5 / 20
Spatial resolution	128×128
Hidden width	48
Structural refinement depth	3 paths, 2 residual blocks each
Energy-evolution depth	3 blocks
Fusion refinement depth	2 residual blocks
Low-frequency bandwidth	$k = 10$
Loss weights $(\lambda_{\mathrm{eng}}, \lambda_{\mathrm{amp}}, \lambda_{\phi})$	$(0.05, 0.01, 0.01)$
Optimizer	AdamW
Learning rate / weight decay	5×10^{-5} / 10^{-5}
Training epochs / batch size	20 / 4
Warm-up / amplitude annealing	500 / 2000 steps

Table 3: Implementation settings used for SolarSpecNet.

C Ablation Details

All six variants reported in the main-paper ablation table use the same data construction, input-output setting, optimizer, training budget, and checkpoint-selection rule as the full model. For module ablations, we remove the specified component and any auxiliary loss that requires its output. For loss ablations, the architecture is unchanged and only the corresponding coefficient is set to zero.

Structural Evolution Branch Ablation. This variant removes phase-history encoding, frequency-coordinate conditioning, future-phase prediction, and the structure-oriented inverse-Fourier output. The phase-consistency loss is also disabled because the model no longer predicts future phase. The remaining pathway therefore forecasts radiation without an explicit phase-based structural representation.

Radiative Energy Branch Ablation. This variant removes the spatial encoder, all three energy-evolution blocks, the decoder, and the predicted future amplitude. The energy auxiliary loss and amplitude-spectrum loss are disabled. This variant relies on the remaining structural pathway without a separately learned radiation-magnitude estimate.

Spectral-Spatial Fusion Module Ablation. Both forecasting branches and their auxiliary losses are retained. We remove the learned three-view fusion mixer and use the direct masked amplitude–phase reconstruction as the final output,

$$\widehat{\mathbf{Y}}_{\text{no-fus}} = \mathrm{Re}\Big[\mathcal{F}^{-1}\Big(\mathbf{M}_{10} \odot \Big|\mathcal{F}(\widehat{\mathbf{Y}}_{\text{eng}})\Big| \odot e^{\mathrm{i}\widehat{\mathbf{\Phi}}}\Big)\Big]. \tag{14}$$

No concatenation with $\widehat{\mathbf{Y}}_{\text{str}}$ or $\widehat{\mathbf{Y}}_{\text{eng}}$ and no residual spatial refinement are applied. This variant isolates the benefit of learned fusion beyond direct spectral recombination.

Without the Amplitude-Spectrum Loss. The full architecture is used with $\lambda_{\text{amp}} = 0$. REB remains supervised in the spatial domain, but no loss is applied directly to the Fourier amplitude of its prediction.

Without the Energy Auxiliary Loss. The full architecture is used with $\lambda_{\text{eng}} = 0$. The energy-oriented intermediate output is not directly matched to the target field, but it still receives gradients through the amplitude-spectrum loss, the spectral reconstruction, and the final prediction loss.

Without the Phase Consistency Loss. The full architecture is used with $\lambda_{\phi} = 0$. The predicted phase still contributes to the structure-oriented output and the fusion module, but it is not directly supervised against the target phase in the retained low-frequency region.

Further Interpretation of the Ablation Results

The main-paper ablation table confirms complementary roles for the architecture and objectives. Removing learned fusion causes the largest degradation (MSE 0.1004, SSIM 0.1420), showing that direct spectral recombination cannot replace learned spatial refinement. Removing REB more than doubles MSE, whereas removing SEB has a stronger effect on SSIM and LPIPS. The three auxiliary-loss ablations also degrade performance in distinct ways, supporting complementary supervision of radiation magnitude and spatial structure.

Figure 6 visualizes the distinct structural, magnitude, and fusion failure modes summarized by the quantitative ablation results.

Lead-Time Ablation Profiles

Figure 7 separates the module-ablation results by forecast horizon. The full model remains the most stable across the twenty output steps, while the three module ablations show distinct failure profiles. This lead-time view confirms that the improvements are not driven by only one favorable horizon.

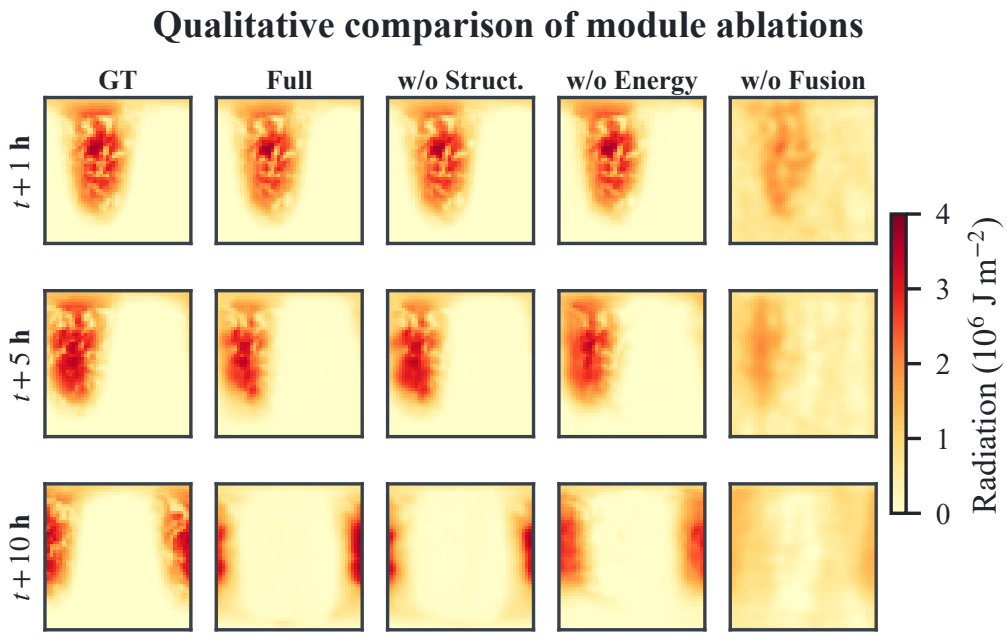


Figure 6: Full model and module ablations at lead times 1, 5, and 10 on fixed test sample 0. All panels share the same physical-unit color scale.

D Extended Discussion

Complementary branch roles. The quantitative and ablation results suggest a clear division of labor. REB determines most of the point-wise accuracy because it predicts future radiation magnitude directly from the historical fields. SEB produces a smaller but consistent gain in spatial coherence by explicitly evolving phase-guided structure. SSF is essential because it learns how to reconcile the direct intensity estimate, the phase-guided spatial estimate, and the masked spectral reconstruction; direct amplitude–phase recombination alone leaves substantial spatial artifacts.

Metric trade-off and error localization. Residuals concentrate near moving illumination boundaries rather than across illuminated interiors. Small boundary displacements can create large local structural differences even when the broad irradiance pattern and regional magnitude are correct. This explains why SolarSpecNet leads the point-wise error metrics while ICNO retains higher SSIM and lower LPIPS: the perceptual metrics are more sensitive to narrow transition misalignment than the global regression metrics.

Limitations and future directions. The current model uses only historical radiation fields and does not explicitly condition on cloud variables, solar geometry, or atmospheric predictors. Its low-frequency reconstruction also favors stable large-scale organization over the finest boundary details. Future work will incorporate exogenous atmospheric variables and multi-scale boundary refinement to improve rapidly changing local conditions and sharpen high-gradient transitions.

E Additional Experimental Analysis

Figure 5e compares ground truth (GT), prediction, and absolute error at representative horizons. SolarSpecNet preserves the major illuminated regions from short to long lead times; the remaining errors concentrate near strong gradients and moving illumination boundaries, where small spatial shifts produce larger point-wise deviations. Interior regions with slowly varying radiation remain substantially cleaner, suggesting that the model estimates broad radiative magnitude more reliably than narrow transition zones.

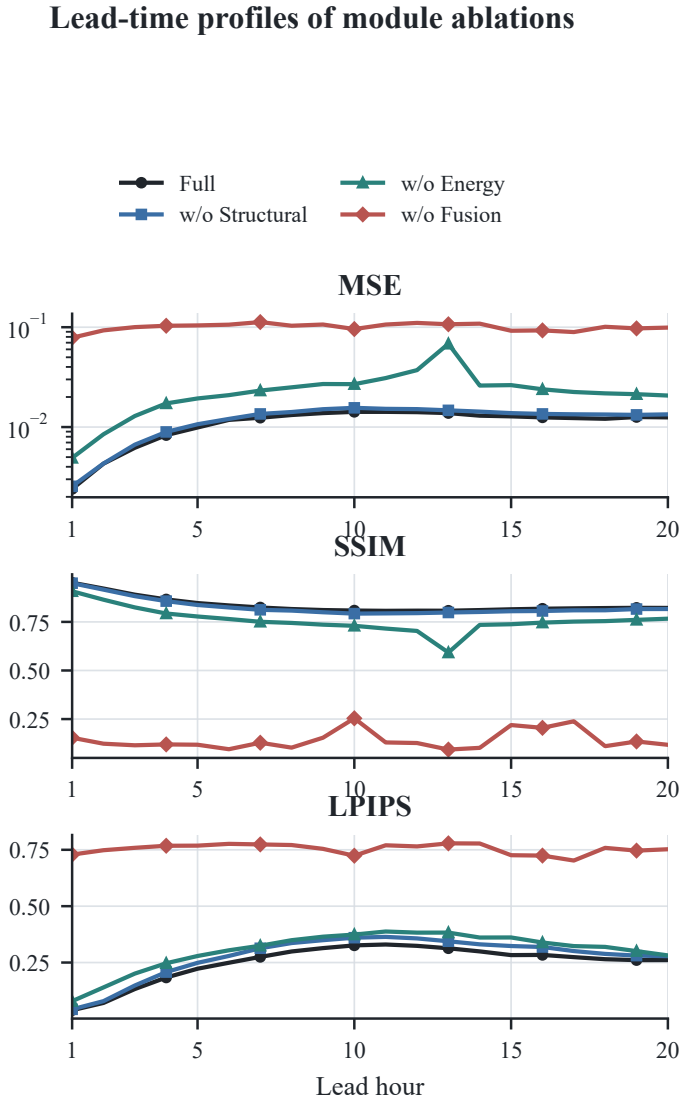


Figure 7: Per-lead MSE, SSIM, and LPIPS for the module ablations. The MSE axis is logarithmic to retain the substantially larger error of the model without learned fusion.

The latitude–longitude view in Figure 5c exposes the full domain and confirms global continuity. The forecasts retain smooth illumination transitions together with regional radiation-magnitude variations, preserving broad latitudinal organization without visible longitudinal seams.

Figure 5d shows that regional mean predictions follow the principal increasing and decreasing trends of the ground truth. SolarSpecNet therefore preserves both spatial patterns and aggregated temporal behavior relevant to regional solar-energy applications. Small deviations are most visible around rapid turning points, but the timing and direction of the dominant changes remain aligned across the forecast window.

The spectral comparison in Figure 5f shows that the prediction preserves the dominant low-frequency energy and follows the target radial spectrum. Larger residuals occur mainly at high frequencies, consistent with prioritizing stable large-scale organization before local refinement. This behavior agrees with the quantitative results: strong regression accuracy is achieved without collapsing the large-scale spectral structure that determines the global illumination pattern.

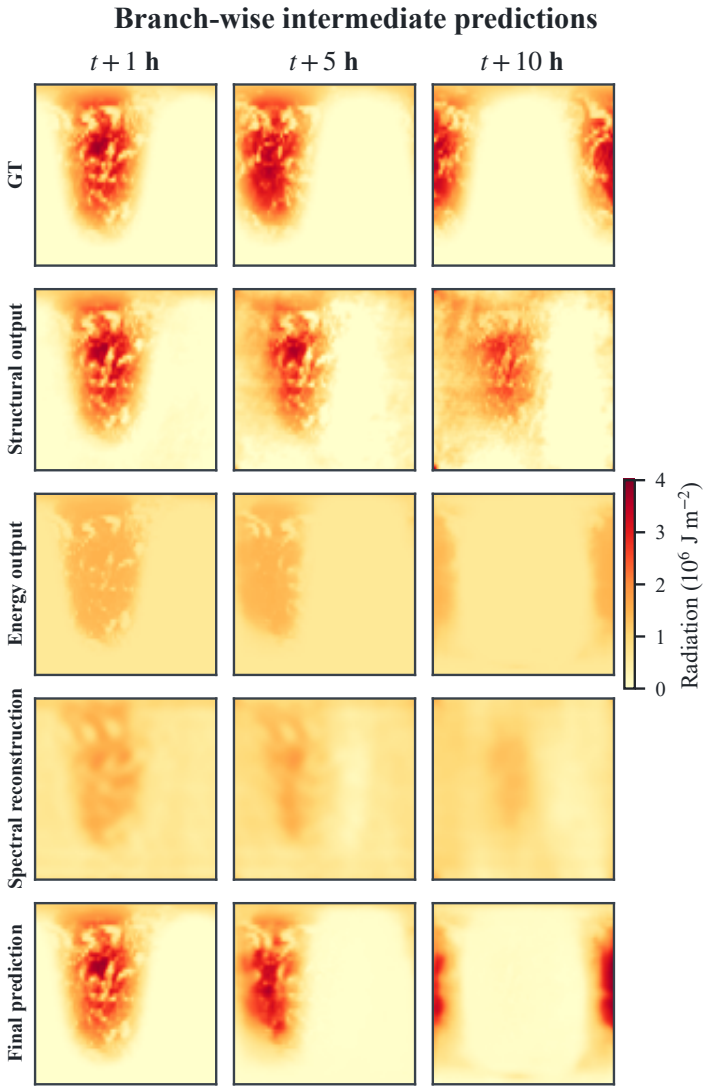


Figure 8: Ground truth, branch-wise intermediate predictions, masked spectral reconstruction, and final output at lead times 1, 5, and 10 for fixed test sample 0. Columns share a common physical-unit color scale.

Branch-wise Intermediate Predictions

Figure 8 exposes the intermediate predictions used by the fusion module. The structural output emphasizes the evolving spatial support, the energy output better follows radiation magnitude, and masked spectral recombination provides a coherent low-frequency estimate. The final prediction combines these complementary cues and removes artifacts that remain in individual intermediates.

Spectral Characteristics

The spectral comparison in the main paper shows that dominant low-frequency energy is preserved, maintaining large-scale illumination and regional magnitude. The remaining discrepancies concentrate at high frequencies associated with sharp boundaries and fine variation. This supports masked amplitude–phase reconstruction for global structure followed by spatial refinement, and motivates future boundary-aware refinement.

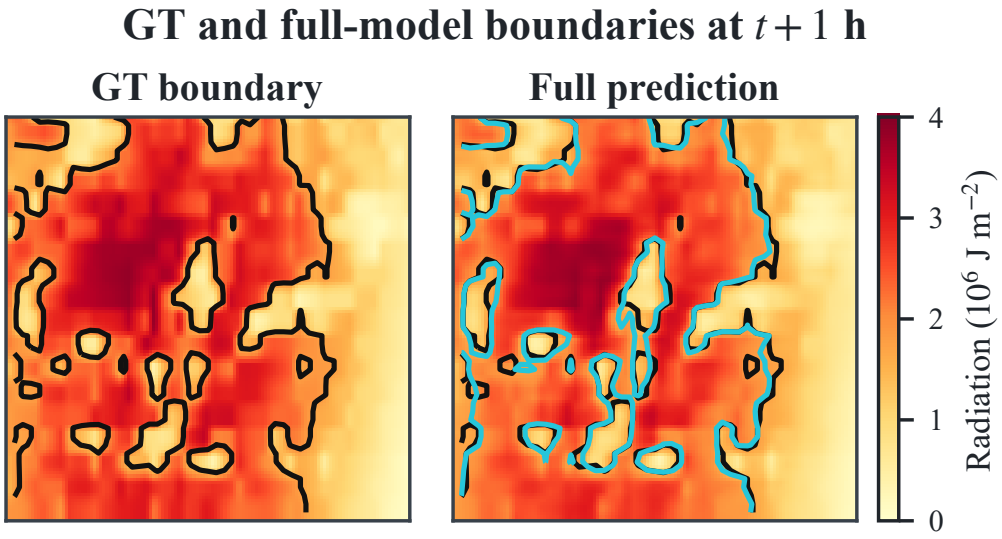


Figure 9: Ground-truth and full-model boundaries at lead time 1 on fixed test sample 0. The ground-truth contour is shown in black and the predicted contour in cyan.

Temporal Consistency and Global Propagation

The globe-projection sequence in the main paper provides a direct view of multi-step consistency. The illuminated footprint translates and deforms gradually with the forecast horizon, while radiation values inside already illuminated regions vary more smoothly. This behavior matches the intended branch roles: phase-guided features emphasize geometric evolution, whereas the energy pathway adjusts magnitude. Consecutive forecasts preserve common spatial support without isolated hotspots, visible map-seam discontinuities, or frame-to-frame flicker, indicating a connected trajectory rather than independently generated frames.

Boundary-focused Analysis

Figure 9 enlarges the highest-gradient region of the full-model forecast at lead time 1. The ground-truth and predicted contours remain close over most of the region, while their small local displacements explain why residuals concentrate around high-gradient structures.

Error Localization and Regional Behavior

The ground-truth, prediction, and error maps in the main paper show that residuals concentrate around moving illumination boundaries and steep gradients. Small boundary displacements can produce large point-wise differences even when the broad illuminated region is correctly located, whereas slowly varying interiors are reconstructed more accurately. This also explains the metric trade-off in the main comparison table: SolarSpecNet leads the regression metrics, while ICNO performs better on SSIM and LPIPS, which are more sensitive to local structural displacement.

The full-domain latitude–longitude view retains broad latitudinal organization and regional magnitude differences without a visible longitudinal seam. The regional-mean curves follow the main rises, declines, and turning directions of the target. Larger deviations occur near rapid turning points, showing that abrupt local changes remain harder than gradual evolution.