SAFEDIVERTOR: FAITHFUL DIVERTOR HEAT FLUX RECONSTRUCTION FROM MACROSCOPIC PLASMA STATE SIGNALS VIA TIME-FREQUENCY PRIOR EXPLOITATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Divertor heat-flux analysis is essential for understanding plasma-wall interactions and protecting plasma-facing components in magnetic-confinement fusion devices, while conventional infrared-based inversion is usually performed after discharge and requires heat-conduction modeling with device-specific material properties, divertor geometry, and boundary conditions. Rather than accelerating this conventional infrared-based inversion paradigm, we introduce a new online-oriented signal-based reconstruction paradigm that directly reconstructs time-resolved radial heat-flux profiles from multi-source macroscopic plasma-state signals available during discharge. To enable systematic study of this task, we construct DivMPS2HF, a multi-source discharge dataset that provides the data foundation and benchmark for signal-based divertor heat-flux reconstruction. We further propose SafeDivertor, a task-driven framework designed to address the key challenges of signal-based heat-flux reconstruction. Specifically, we design physical prior-aware initialization to provide radial-distribution guidance for unavailable target channels, input perturbation to improve robustness against heterogeneous signal reliability, spectral-aware reconstruction optimization to preserve transient heat-flux dynamics, and progressive training to progressively optimize these complementary objectives. Experiments on DivMPS2HF demonstrate that SafeDivertor achieves the best overall performance among the evaluated time-series baselines across all five metrics, establishing a new performance benchmark for signal-based divertor heat-flux reconstruction.

1 INTRODUCTION

Divertor heat-flux analysis is essential for safe and high-performance fusion operation, as concentrated and transient heat loads can damage plasma-facing components (Yang et al., 2020). This challenge is particularly critical for future burning-plasma devices such as ITER, where divertor heat loads are subject to strict material limits (Pitts et al., 2019). Accurate time-resolved estimation is therefore important for heat-load assessment, component protection, and future online control (Aymerich et al., 2023).

As shown in Figure 1(b), conventional divertor heat-flux analysis mainly follows an infrared-based inversion paradigm, where the target-surface temperature is first measured and the heat-flux profile is then reconstructed using computationally expensive numerical solvers, typically for offline post-shot analysis (Herrmann et al., 1995; Lee et al., 2017; Shi et al., 2017; Yang et al., 2020). Recent learning-based methods, such as Physics-Informed Neural Networks (PINNs) for accelerating heat-equation solving and convolutional neural networks (CNNs) for heat-flux reconstruction, improve computational efficiency but still only replace or accelerate individual components within the conventional thermal-analysis paradigm (Aymerich et al., 2023; Looby et al., 2019; Wang et al., 2025). *However, these methods remain dependent on thermal measurements, heat-conduction modeling, or simulation-derived data, limiting their direct application to online-oriented reconstruction.* In contrast, macroscopic plasma-state signals are routinely recorded during plasma discharges and are usually available for online monitoring and control (Kates-Harbeck et al., 2019; Guo et al., 2021;

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.

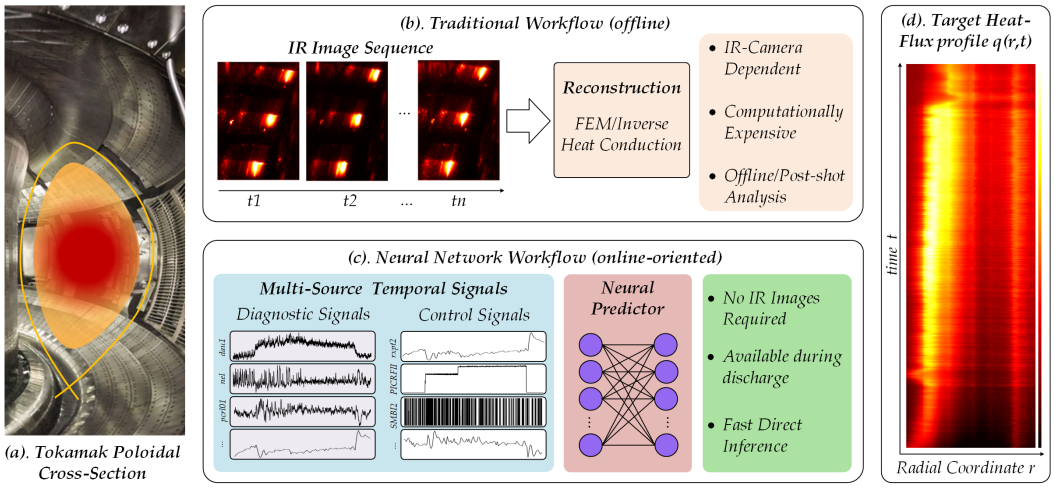


Figure 1: Illustration of the conventional IR-based heat-flux inversion workflow and the proposed signal-based neural reconstruction workflow. (a) Tokamak poloidal cross-section. (b) Offline workflow based on IR image sequences and heat-conduction inversion. (c) Proposed online-oriented workflow that reconstructs heat-flux profiles from multi-source plasma-state signals. (d) Target divertor heat-flux profile $q(r, t)$.

Zheng et al., 2023; Matos et al., 2020). This naturally raises an important question: *Can time-resolved radial divertor heat-flux profiles be reconstructed directly from macroscopic plasma-state signals, without relying on conventional infrared-based inversion?*

As illustrated in Figure 1(c), we introduce an online-oriented signal-based heat-flux reconstruction task that directly maps macroscopic plasma-state signals to time-resolved radial divertor heat-flux profiles. To support this task, we construct DivMPS2HF, a multi-source discharge dataset that aligns heterogeneous plasma-state signals with radial heat-flux profiles. We further propose SafeDivertor, which jointly models observed signals and target-channel placeholders. Because the target-channel placeholders provide no heat-flux information, physical prior-aware initialization (PPI) supplies them with a channel-level statistical heat-flux prior, providing radial-distribution guidance before reconstruction. Since heterogeneous plasma-state signals may differ in reliability and cause the model to over-rely on a small subset of channels, input perturbation (IP) regularizes the conditioning process by perturbing randomly selected observed channels during training. Moreover, because point-wise reconstruction loss alone tends to smooth rapid heat-flux variations, spectral-aware reconstruction optimization (SRO) exploits the time-frequency prior embodied in the ground-truth heat-flux profiles through a multi-scale short-time Fourier transform (STFT) loss, thereby preserving transient dynamics and spectral consistency across multiple temporal scales. As these objectives emphasize different aspects of reconstruction, progressive training (PT) introduces them in stages to stabilize optimization. Experiments on DivMPS2HF show that SafeDivertor achieves the best overall performance among the evaluated time-series baselines across all five metrics. The main contributions of this work are summarized as follows:

- To the best of our knowledge, we are the first to formulate an online-oriented signal-based reconstruction paradigm that shifts divertor heat-flux analysis from offline infrared-based inversion toward online-oriented reconstruction from macroscopic plasma-state signals available during discharge.
- We propose SafeDivertor, a task-driven framework that integrates physical prior-aware initialization, input perturbation, spectral-aware reconstruction optimization, and progressive training for faithful heat-flux reconstruction.
- We construct DivMPS2HF, a multi-source discharge dataset and evaluation benchmark for this task, and establish a strong reference baseline that achieves the best overall performance among the evaluated time-series methods across all five metrics.

2 Related Work

Accurate divertor heat-flux analysis is essential for understanding plasma-wall interactions and protecting plasma-facing components. Conventional methods mainly reconstruct heat flux from infrared surface-temperature measurements by solving heat-conduction equations with device-specific material properties, geometry, and boundary conditions (Lee et al., 2017; Herrmann et al., 1995; Shi et al., 2017). Learning-based approaches have further accelerated heat-equation solving with PINNs or reconstructed parameterized heat-flux profiles using CNNs (Aymerich et al., 2023; Looby et al., 2019). However, they still generally depend on thermal measurements or heat-conduction modeling and are primarily designed for offline analysis. Meanwhile, recent time-series methods have developed diverse architectures for general forecasting, forecasting with exogenous variables, and imputation (Liu et al., 2024; 2025; Chen et al., 2025; Hu et al., 2025; Lee & Lee, 2026; Shu & Lampos, 2026; Wang et al., 2024; Li et al., 2026; Park et al., 2026). Nevertheless, these settings differ from our online-oriented signal-based reconstruction task, where all divertor heat-flux channels are unavailable throughout the current window and must be directly reconstructed as time-resolved radial profiles from physically distinct multi-source macroscopic plasma-state signals, without infrared measurements. *More details are provided in Appendix A.*

3 Motivation

Unlike conventional time-series imputation, which reconstructs partially missing observations within the same variable space, our task requires reconstructing entirely unavailable heat-flux channels from physically distinct macroscopic plasma-state signals. Therefore, we formulate it as a structured channel-level reconstruction problem, as follows.

Let $\mathcal{D} = \left\{\left(\mathbf{X}^{(i)}, \mathbf{Y}^{(i)}\right)\right\}_{i=1}^{N}$ denote a discharge dataset, where N denotes the number of samples after window sampling. For the i-th sample, $\mathbf{X}^{(i)}$ represents the macroscopic plasma-state signals within a time window of length T, and $\mathbf{Y}^{(i)}$ denotes the corresponding time-resolved heat-flux profile on the divertor target plate. The input sequence is defined as $\mathbf{X}^{(i)} = [x_1^{(i)}, x_2^{(i)}, ..., x_C^{(i)}] \in \mathbb{R}^{C\times T}$, where $C = 67$ is the number of signal channels; and the corresponding target heat-flux profile is defined as $\mathbf{Y}^{(i)} = [y_1^{(i)}, y_2^{(i)}, ..., y_O^{(i)}] \in \mathbb{R}^{O\times T}$, where $O = 116$ is the number of heat-flux channels.

Since the actual heat-flux sequence within the current time window is unobservable, we further represent this task as a structured channel-level reconstruction problem based on target-channel initialization. Specifically, an initial representation of the target heat-flux channels, denoted as $\mathbf{Y}_{\text{init}}^{(i)} \in \mathbb{R}^{O\times T}$, is concatenated with the observed signals $\mathbf{X}^{(i)}$ along the channel dimension to form a unified input sequence:

$$\mathbf{Z}^{(i)} = [\mathbf{X}^{(i)}, \mathbf{Y}_{\text{init}}^{(i)}] \in \mathbb{R}^{F\times T}, \quad F = C + O. \tag{1}$$

It is important to note that $\mathbf{Y}_{\text{init}}^{(i)}$ does not contain the actual heat-flux information of the current sample; it is only used as an initial placeholder for the target heat-flux channels. The model reconstructs the actual heat-flux profile based on this unified input sequence:

$$\hat{\mathbf{Y}}^{(i)} = f_{\boldsymbol{\theta}}(\mathbf{Z}^{(i)}), \quad f_{\boldsymbol{\theta}} : \mathbb{R}^{F\times T} \to \mathbb{R}^{O\times T}. \tag{2}$$

The model parameters are optimized by minimizing the reconstruction error between the reconstructed heat-flux sequence and the actual heat-flux sequence:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\hat{\mathbf{Y}}^{(i)}, \mathbf{Y}^{(i)}\right). \tag{3}$$

During training and evaluation, the loss and all metrics are computed only on the O target heat-flux channels.

Although the task formulation is straightforward, it presents several task-specific challenges. First, the target-channel placeholders contain no heat-flux information, leaving the model without explicit guidance about the radial distribution to be reconstructed. Second, the heterogeneous plasma-state

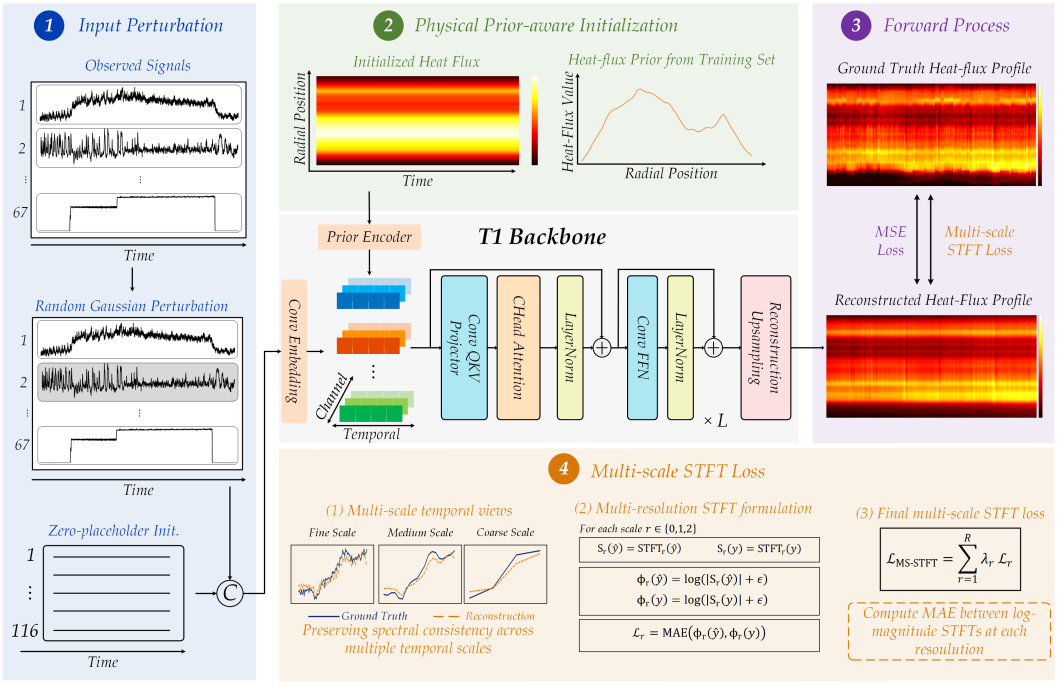


Figure 2: Overview of the proposed SafeDivertor framework for divertor heat-flux reconstruction from macroscopic plasma-state signals. SafeDivertor integrates physical prior-aware initialization, input perturbation, T1 backbone reconstruction, and multi-scale STFT-based spectral supervision.

signals differ in physical meaning and measurement reliability, which may cause the model to over-rely on a small subset of channels. Third, point-wise reconstruction losses tend to smooth rapid heat-flux variations and fail to explicitly exploit the time-frequency prior contained in the target profiles. Finally, these objectives emphasize different aspects of reconstruction and are difficult to optimize simultaneously. These challenges motivate SafeDivertor, which introduces PPI, IP, SRO, and PT to address the lack of target guidance, over-reliance on specific input channels, insufficient spectral preservation, and optimization instability, respectively, as detailed in the following section.

4 Methodology

4.1 Dataset Curation

To support the proposed online-oriented reconstruction task, we construct DivMPS2HF, a multi-source discharge dataset containing 77 shots, divided into 64, 3, and 10 shots for training, validation, and testing, respectively. The split is performed at the shot level to prevent cross-shot information leakage. All signals are temporally aligned and resampled to a unified 1 kHz grid. Each sample contains 67 input channels, including plasma-state diagnostic signals, equilibrium and configuration parameters, and signals related to external actuators and device control. These channels are selected from a broader pool of candidate signals to retain signals with reliable data quality, strong physical relevance to divertor heat-load evolution, and useful predictive information. The reconstruction target is a time-resolved radial heat-flux profile consisting of 116 channels, obtained from post-shot heat-flux analysis as supervised labels. Notably, infrared measurements and heat-conduction inversion results are used only to construct the reference labels and are not included in the model input. During inference, SafeDivertor directly reconstructs the heat-flux profile from macroscopic plasma-state signals through a single neural forward pass. *Further details on dataset are provided in Appendix B.1.*

4.2 SAFEDIVERTOR FRAMEWORK

Physical Prior-aware Initialization. Since the target heat-flux profile is unavailable during inference, the model must reconstruct the spatial heat-flux distribution solely from multi-source plasma signals. However, zero-valued representations of the unavailable target channels provide no information about the intrinsic radial heat-flux distribution across different divertor locations. Since the radial heat-flux distribution exhibits persistent spatial characteristics across discharges, we estimate a channel-wise prior from historical observations. To address this challenge, we formulate the multi-source signals and target heat-flux channels as a unified structured time series. Based on this formulation, we design Physical Prior-aware Initialization (PPI) to inject radial heat-flux distribution knowledge into unavailable target channels before reconstruction.

Specifically, for the o-th target heat-flux channel, its statistical prior is computed as the average heat-flux value over all training shots and time steps:

$$\mu_o = \frac{1}{\sum_{s=1}^{S_{\text{train}}} L_s} \sum_{s=1}^{S_{\text{train}}} \sum_{t=1}^{L_s} \mathbf{Y}_{o,t}^{(s)}, \quad o = 1, \ldots, O, \tag{4}$$

where S_{train} denotes the number of training discharge shots, L_s denotes the length of the s-th training shot, and $Y_{o,t}^{(s)}$ is the heat-flux value of the o-th target channel at time step t in the s-th training shot. For each sample, the radial heat-flux prior is defined as follows:

$$\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_O]^\top, \quad \mathbf{P} = \boldsymbol{\mu} \mathbf{1}_{1\times T} \in \mathbb{R}^{O\times T}, \tag{5}$$

where $\mathbf{1}_{1\times T} \in \mathbb{R}^{1\times T}$ denotes an all-ones row vector and T is the length of the input time window.

Based on this prior, we introduce a lightweight prior encoding branch to extract structural information. Specifically, the original input with zero-valued target placeholders is first encoded into a feature representation $\mathbf{H}_0$. According to the channel organization of the unified input, $\mathbf{H}_0$ is divided into the observed-signal features and target-placeholder features:

$$\mathbf{H}_0^{(i)} = \left[\mathbf{H}_{0,\text{obs}}^{(i)}, \mathbf{H}_{0,\text{tar}}^{(i)}\right]. \tag{6}$$

Meanwhile, the statistical heat-flux prior $\mathbf{P}$ is processed by a convolutional prior encoder $g_{\boldsymbol{\theta}}(\cdot)$ to obtain the prior feature. The prior feature is then added only to the target-placeholder features, yielding the prior-enhanced representation:

$$\mathbf{H}^{(i)} = \left[\mathbf{H}_{0,\text{obs}}^{(i)}, \mathbf{H}_{0,\text{tar}}^{(i)} + g_{\boldsymbol{\theta}}(\mathbf{P})\right]. \tag{7}$$

Here, $\mathbf{H}^{(i)}$ denotes the prior-enhanced feature representation. This design preserves the observed-signal features and injects the statistical radial heat-flux information only into the target-channel representations through an independent prior encoding branch.

Input Perturbation. During practical discharge data acquisition, multi-source diagnostic signals are inevitably affected by hardware noise, acquisition errors, and temporal fluctuations. Moreover, different diagnostic channels may exhibit varying reliability across discharges and time windows, which can lead to shortcut learning where the model relies excessively on a small subset of highly informative signals rather than capturing robust cross-diagnostic dependencies. To address this challenge, we design Input Perturbation (IP) as a channel-level robustness regularization strategy that improves model tolerance to unreliable heterogeneous diagnostics during training. As illustrated in Part 1 of Figure 2, Gaussian perturbation is applied only to selected observable signals, while the remaining observable signals and the heat-flux prior are kept unchanged. Specifically, for each training sample, we randomly select 20% of the 67 observable signals and add Gaussian noise with a standard deviation of 0.03 to the complete time series of these channels. The augmented input is defined as follows:

$$\tilde{\mathbf{X}}_{c,t} = \begin{cases} \mathbf{X}_{c,t} + \epsilon_{c,t}, & c \in \mathcal{C}_n, \\ \mathbf{X}_{c,t}, & c \notin \mathcal{C}_n, \end{cases} \quad \epsilon_{c,t} \sim \mathcal{N}(0, \sigma^2), \tag{8}$$

where $\tilde{\mathbf{X}}_{\mathbf{c},\mathbf{t}}$ and $\mathbf{X}_{c,t}$ denote the augmented and original signals of channel c at time step t, $\mathcal{C}_n$ is the randomly selected subset containing 20% of the channels, and $\epsilon_{c,t}$ denotes the added Gaussian noise. σ controls the noise intensity and is set to 0.03 in our experiments.

Basic Reconstruction Loss. Given the reconstructed heat-flux profile $\hat{\mathbf{Y}}$ and the true heat-flux profile $\mathbf{Y}$, we first adopt the mean squared error as the basic reconstruction loss:

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{BOT} \sum_{b=1}^{B} \sum_{o=1}^{O} \sum_{t=1}^{T} \left(\hat{\mathbf{Y}}_{b,o,t} - \mathbf{Y}_{b,o,t} \right)^2, \tag{9}$$

where B is the batch size, $O = 116$ denotes the number of target heat-flux channels, and T is the length of the time window. The MSE loss provides point-wise supervision for numerical fidelity in the time domain.

Spectral-aware Reconstruction Optimization. Although the MSE loss provides direct point-wise supervision in the time domain, it does not explicitly constrain the frequency characteristics of the reconstructed heat-flux sequences. In divertor heat-load evolution, rapid variations and transient events may appear as high-frequency components in the heat-flux profile. Models optimized solely by MSE may produce overly smoothed reconstructions that fail to preserve these spectral details.

Therefore, we introduce Spectral-aware Reconstruction Optimization (SRO) to explicitly preserve the temporal-frequency characteristics of heat-flux evolution. Compared to single-scale spectral constraints, multi-scale designs are better suited for heat-flux reconstruction because different time scales emphasize different dynamic modes. Shorter temporal scales are more sensitive to rapid local variations and transient heat-load dynamics, while longer temporal scales help capture broader heat-flux evolution trends. Therefore, applying spectral constraints at multiple temporal scales encourages the model to preserve both local high-frequency dynamics and global time-frequency consistency. Let $\phi(\cdot)$ denote the log-magnitude STFT operator along the temporal dimension:

$$\phi(\mathbf{Y}) = \log\left(|\mathrm{STFT}(\mathbf{Y})| + \epsilon\right), \tag{10}$$

where $\mathrm{STFT}(\cdot)$ denotes the STFT operation along the temporal dimension, and ϵ is a small constant for numerical stability. The multi-scale STFT loss is formulated as follows:

$$\mathcal{L}_{\mathrm{MS-STFT}} = \sum_{r=1}^{R} \lambda_r \mathrm{MAE}\left(\phi(\mathbf{Y}_r), \phi(\hat{\mathbf{Y}}_r)\right), \tag{11}$$

where $R = 3$, $\mathbf{Y}_r$ and $\hat{\mathbf{Y}}_r$ denote the ground-truth and reconstructed heat-flux sequences at the r-th temporal scale, respectively, and λ_r is the weight for the r-th temporal scale. By matching the log-magnitude spectra of heat-flux sequences at multiple temporal scales, this loss helps the model preserve both rapid high-frequency variations and broader temporal spectral patterns.

Overall Training Objective. The final training objective combines the basic reconstruction loss and the multi-scale STFT-based spectral loss. The MSE loss provides direct supervision for numerical accuracy in the time domain, while the multi-scale STFT loss constrains the spectral characteristics of heat-flux sequences across different temporal scales. The overall loss is defined as follows:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{MSE}} + \mathcal{L}_{\mathrm{MS-STFT}}. \tag{12}$$

By jointly optimizing these objectives, the model improves the point-wise numerical accuracy and multi-scale spectral consistency of the reconstructed heat-flux profile.

Progressive Training. Directly optimizing input robustness and spectral consistency from the beginning may introduce optimization conflicts, since the model needs to simultaneously learn basic reconstruction, robustness to perturbations, and temporal-frequency consistency. To resolve this optimization conflict, we adopt a three-stage progressive training strategy. In the first stage, the model is trained only with the basic MSE reconstruction loss, allowing it to learn a stable mapping from multi-source signals to heat-flux profiles:

$$\boldsymbol{\theta}_1^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}_{\mathrm{MSE}}(\mathbf{Y}, \hat{\mathbf{Y}}), \tag{13}$$

where $\hat{\mathbf{Y}} = f_\theta([\mathbf{X}, \mathbf{Y}_{\mathrm{init}}])$ denotes the reconstructed heat-flux profile.

In the second stage, channel-level noise augmentation is enabled while the model continues to optimize the reconstruction objective, which improves its robustness to input-channel perturbations:

$$\boldsymbol{\theta}_2^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}_{\mathrm{MSE}} \left(\mathbf{Y}, f_{\boldsymbol{\theta}}([\tilde{\mathbf{X}}, \mathbf{Y}_{\mathrm{init}}]) \right), \tag{14}$$

where $\tilde{\mathbf{X}}$ represents the perturbed observed signals. In the third stage, the input perturbation is disabled, and the multi-scale STFT loss is introduced to refine the reconstructed heat-flux profiles under clean input conditions:

$$\boldsymbol{\theta}_3^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}_{\mathrm{total}}, \tag{15}$$

where $\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{MSE}} + \mathcal{L}_{\mathrm{MS-STFT}}$ denotes the overall training objective. This stage focuses on improving spectral consistency and preserving rapid heat-flux variations. The model parameters are progressively inherited across stages, enabling a stable transition from numerical reconstruction to robust and spectral-aware heat-flux reconstruction.

5 EXPERIMENTS

Implementation Details. We construct DivMPS2HF, a multi-source time-series dataset for reconstructing divertor heat-flux profiles from macroscopic plasma-state signals. All models are evaluated using five metrics: MSE, MAE, SSIM, LSD, and LSD-HF, which measure point-wise numerical accuracy, structural similarity, and spectral consistency of the reconstructed heat-flux profiles.

During training and evaluation, fixed-length windows are sampled from the corresponding discharge shots. The input window length is set to $T = 500$ time steps, corresponding to a 0.5-second discharge segment under the unified 1 kHz sampling rate. The experiments are conducted on eight NVIDIA RTX 4090 GPUs. For SafeDivertor, the Gaussian perturbation ratio is set to 20%, and the noise standard deviation is set to 0.03. The channel-level heat-flux prior is computed only from the training set and kept fixed during validation and testing. For efficiency evaluation, all models are tested with batch size 1 and input length 500, and the reported inference latency only measures the model forward process. *More details about the dataset construction, evaluation metrics and baselines are provided in Appendix B.1 and Appendix B.2.*

Main Results. As shown in Table 1, SafeDivertor achieves the best performance across all evaluation metrics compared with representative time-series baselines. Among the baselines, T1 (Park et al., 2026) shows the strongest overall performance in numerical accuracy and spectral consistency, achieving the best baseline MSE, MAE, LSD, and LSD-HF, while TimeFilter (Hu et al., 2025) obtains the highest baseline SSIM. In contrast, several forecasting and exogenous-variable forecasting models show limited performance, indicating that directly applying existing time-series forecasting paradigms is insufficient for reconstructing target heat-flux channels from heterogeneous macroscopic plasma-state signals.

Compared with the best baseline result for each metric, SafeDivertor reduces MSE from 0.262 to 0.200, MAE from 0.333 to 0.291, LSD from 3.521 to 2.475, and LSD-HF from 3.829 to 2.729, while improving SSIM from 0.854 to 0.870. These results demonstrate that the proposed task-driven designs help SafeDivertor better preserve numerical accuracy, spatiotemporal structure, and spectral characteristics of divertor heat-flux profiles.

Table 1: Overall comparison with representative time-series baselines on the DivMPS2HF dataset. Further details of the evaluated baselines are provided in Appendix B.3.

	T1	Sonnet	Timer-XL	TimeFilter	CrossLinear	SimpleTM	TimePerceiver	TimeXer	iTransformer	SafeDivertor
MSE↓	<u>0.262</u>	0.382	0.307	0.279	0.378	0.540	0.437	0.538	0.469	**0.200**
MAE↓	<u>0.333</u>	0.414	0.367	0.354	0.427	0.499	0.426	0.488	0.446	**0.291**
SSIM↑	0.848	0.804	0.824	<u>0.854</u>	0.706	0.760	0.814	0.769	0.792	**0.870**
LSD↓	<u>3.521</u>	7.294	4.718	3.868	5.101	8.033	4.546	4.876	7.893	**2.475**
LSD-HF↓	<u>3.829</u>	7.418	5.420	4.573	4.976	8.153	5.359	5.228	8.136	**2.729**

Ablation Study. Table 2 presents the ablation results of the proposed task-driven components, including physical prior-aware initialization (PPI), input perturbation (IP), spectral-aware reconstruction optimization (SRO), and progressive training (PT). The baseline model without these components achieves an MSE of 0.262, MAE of 0.333, SSIM of 0.848, LSD of 3.521, and LSD-HF of 3.829.

When each component is introduced individually, different performance trends can be observed. PPI reduces MSE from 0.262 to 0.211, MAE from 0.333 to 0.293, and improves SSIM from 0.848 to 0.870, indicating that the physical heat-flux prior provides useful statistical radial-distribution information for unavailable target channels. IP also improves the time-domain and structural metrics, reducing MSE to 0.230 and improving SSIM to 0.860, which shows that perturbing observable signal channels helps regularize the model's dependence on specific signals. However, both PPI and IP lead to higher LSD and LSD-HF, because they do not explicitly supervise the spectral characteristics of reconstructed heat-flux sequences. In contrast, SRO contributes most significantly to spectral consistency, reducing LSD from 3.521 to 2.306 and LSD-HF from 3.829 to 2.511. This confirms the necessity of multi-scale STFT-based spectral supervision for preserving spectral characteristics and high-frequency transient variations.

When PPI, IP, and SRO are jointly used, the model achieves a more balanced performance across time-domain, structural, and spectral metrics. This demonstrates that these components play complementary roles: PPI provides statistical heat-flux prior information, IP regularizes heterogeneous observable signals, and SRO enhances spectral consistency. After further introducing PT, the full model achieves the best MSE and MAE of 0.200 and 0.291, respectively, while maintaining competitive spectral performance. These results show that the proposed task-driven components improve complementary aspects of reconstruction quality.

Table 2: Ablation study of the main components in SafeDivertor on the DivMPS2HF dataset.

	PPI	IP	SRO	PT	MSE↓	MAE↓	SSIM↑	LSD↓	LSD-HF↓
#1	✗	✗	✗	✗	0.262	0.333	0.848	3.521	3.829
#2	✓	✗	✗	✗	<u>0.211</u>	<u>0.293</u>	**0.870**	3.617	4.141
#3	✗	✓	✗	✗	0.230	0.313	0.860	3.725	4.205
#4	✗	✗	✓	✗	0.219	0.299	0.863	**2.306**	**2.511**
#5	✓	✓	✓	✗	0.211	0.296	<u>0.869</u>	<u>2.326</u>	<u>2.568</u>
#6	✓	✓	✓	✓	**0.200**	**0.291**	**0.870**	2.475	2.729

Effect of Multi-scale Spectral Loss. As shown in Table 3, we analyze the role of spectral loss in the spectral-aware reconstruction optimization module. The *w/o* SRO setting represents a model trained using only the basic MSE reconstruction loss. Introducing single-scale STFT loss significantly improves the model across all metrics. This indicates that the spectral optimization objective helps the model better preserve the spectral characteristics and rapid changes in the heat-flux sequences.

Compared to the single-scale setting, the multi-scale STFT loss function further improves most metrics. These results demonstrate that applying spectral constraints across multiple timescales is more effective than using a single scale, as different scales can capture both broader heat flux evolution patterns and local high-frequency variations. Overall, the multi-scale STFT loss function provides a better balance between numerical accuracy and spectral consistency.

Input Perturbation Strategy Analysis. We further evaluate different input perturbation strategies by applying channel masking and Gaussian perturbation to observable signal channels under different perturbation ratios $r \in \{0.1, 0.2, 0.4\}$. As shown in Table 4, channel masking provides limited and unstable improvements. Although it slightly improves MSE and SSIM at $r = 0.1$, larger masking ratios degrade the numerical metrics, suggesting that directly removing observable channels may discard useful plasma-state information.

In contrast, Gaussian perturbation consistently improves numerical accuracy and structural similarity across different ratios, indicating that perturbing observable channels is more effective than completely masking them. Although both strategies degrade spectral performance, this is expected because they do not explicitly constrain frequency-domain characteristics.

Overall, Gaussian perturbation is a more effective input perturbation strategy for improving numerical accuracy and structural similarity.

Efficiency Analysis. Table 5 compares the computational complexity and inference efficiency of different models. SafeDivertor requires 21.255 GFLOPs and achieves an inference latency of 11.313 ms per sample with a batch size of 1 and an input length of 500. Since each input win-

Table 3: Ablation study of the multi-scale STFT loss.

Metrics	*w/o* SRO	Single Scale	Multi Scale
MSE↓	0.262	0.221	**0.219**
MAE↓	0.333	0.301	**0.299**
SSIM↑	0.848	**0.867**	0.863
LSD↓	3.521	2.367	**2.306**
LSD-HF↓	3.829	2.549	**2.511**

Table 4: Effect of different input perturbation strategies.

	w/o IP	r=0.1		r=0.2		r=0.4	
Metrics	-	**CM**	**GP**	**CM**	**GP**	**CM**	**GP**
MSE↓	0.262	0.251	0.229	0.288	0.230	0.276	**0.203**
MAE↓	0.333	0.329	0.316	0.345	0.313	0.339	**0.297**
SSIM↑	0.848	0.858	0.861	0.854	0.860	0.853	**0.874**
LSD↓	**3.521**	3.653	3.816	3.696	3.725	3.599	3.784
LSD-HF↓	**3.829**	4.093	4.315	4.096	4.205	4.013	4.163

Table 5: Comparison of model parameters (M), computational cost (GFLOPs), and per-sample inference latency (ms).

	Sonnet	T1	Timer-XL	TimeFilter	CrossLinear	SimpleTM	TimePerceiver	TimeXer	iTransformer	SafeDivertor
Params.	42.239	12.783	33.825	**11.011**	33.130	34.753	16.032	51.088	34.635	12.783
GFLOPs	123.188	21.240	61.583	35.118	5.122	21.629	14.824	44.684	**2.390**	21.255
Inference	10.947	11.023	14.721	261.343	4.623	6.126	2.530	4.161	**1.732**	11.313

dow corresponds to a 0.5-second discharge segment at 1 kHz, this latency indicates its potential for online-oriented window-level heat-flux reconstruction. Compared with the T1 backbone, SafeDivertor retains the same parameter scale, with GFLOPs increasing only from 21.240 to 21.255 and comparable latency, while substantially improving reconstruction performance. These results demonstrate a favorable performance-efficiency trade-off with negligible additional inference cost.

Visualization. As shown in Figure 3, we compare the ground truth, SafeDivertor, and T1 on four samples. The original T1 backbone recovers the overall spatiotemporal heat-flux structure and main radial high-heat-flux bands, but its results are relatively over-smoothed and may underestimate local high-intensity regions, failing to preserve temporal fluctuations and fine-grained heat-flux variations. In contrast, SafeDivertor better preserves the main heat-flux structures while recovering clearer high-intensity bands and richer temporal details. These visual results are consistent with the quantitative improvements in Table 1, further demonstrating the advantage of the proposed task-driven designs. *More visualization results are provided in Appendix C.1.*

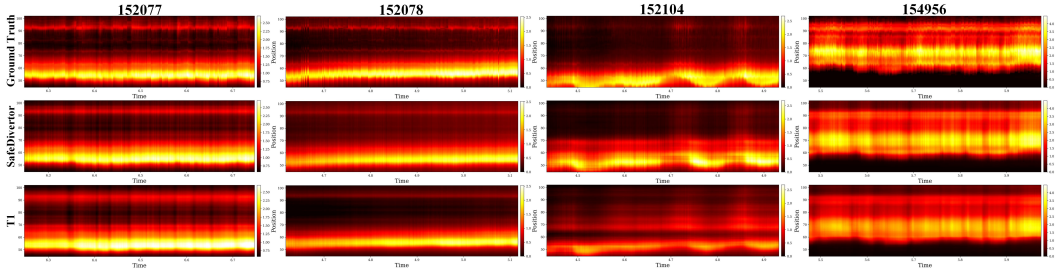


Figure 3: Qualitative visualization of four randomly selected samples from different discharge shots, showing the ground truth, SafeDivertor, and T1 from top to bottom.

6 Conclusion

In this paper, we introduce a new online-oriented paradigm for directly reconstructing divertor heat-flux profiles from macroscopic plasma-state signals, without using infrared surface-temperature measurements as model inputs. We also construct DivMPS2HF, a multi-source discharge dataset for this task, and propose SafeDivertor, a unified reconstruction framework equipped with physical prior-aware initialization, input perturbation, spectral-aware reconstruction optimization, and progressive training. Experiments on DivMPS2HF show that SafeDivertor achieves the best overall performance among the evaluated time-series baselines across all five metrics. Efficiency analysis further shows that SafeDivertor maintains inference efficiency comparable to T1, providing a favorable latency basis for future online deployment. Future work will expand the dataset, improve cross-scenario generalization, and further enhance reconstruction efficiency.

AI USE STATEMENT

We acknowledge the use of LLM solely during the writing process for language refinement, grammar checking, and improving the clarity and conciseness of academic expressions. We have reviewed all modifications assisted by AI and take full responsibility for the final content of this paper.

ETHICS STATEMENT

This study focuses on heat flux reconstruction based on nuclear fusion plasma measurement data and poses no foreseeable ethical risks. The datasets used consist entirely of experimentally measured data recorded by machines; they contain no human subjects or personally identifiable information.

REPRODUCIBILITY STATEMENT

The construction procedures of the datasets used in this study are described in detail in Appendix B.1. The source code will be made publicly available once the paper is accepted.

REFERENCES

E Aymerich, F Pisano, B Cannas, G Sias, A Fanni, Y Gao, D Böckenhoff, M Jakubowski, et al. Physics informed neural networks towards the real-time calculation of heat fluxes at w7-x. *Nuclear Materials and Energy*, 34:101401, 2023.

Hui Chen, Viet Luong, Lopamudra Mukherjee, and Vikas Singh. SimpleTM: A simple baseline for multivariate time series forecasting. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=oANkBaVci5.

BH Guo, B Shen, DL Chen, C Rea, RS Granetz, Y Huang, L Zeng, H Zhang, JP Qian, YW Sun, et al. Disruption prediction using a full convolutional neural network on east. *Plasma Physics and Controlled Fusion*, 63(2):025008, 2021.

Albrecht Herrmann, W Junker, K Gunther, S Bosch, M Kaufmann, J Neuhauser, G Pautasso, Th Richter, and R Schneider. Energy flux to the asdex-upgrade diverter plates determined by thermography and calorimetry. *Plasma Physics and Controlled Fusion*, 37(1):17–29, 1995.

Yifan Hu, Guibin Zhang, Peiyuan Liu, Disen Lan, Naiqi Li, Dawei Cheng, Tao Dai, Shu-Tao Xia, and Shirui Pan. Timefilter: Patch-specific spatial-temporal graph filtration for time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=490VcNtjh7.

Julian Kates-Harbeck, Alexey Svyatkovskiy, and William Tang. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature*, 568(7753):526–531, 2019.

HH Lee, RA Pitts, CS Kang, ST Oh, JG Bak, SH Hong, HM Wi, YS Kim, HS Kim, DC Seo, et al. Thermographic studies of outer target heat fluxes on kstar. *Nuclear Materials and Energy*, 12: 541–547, 2017.

Jaebin Lee and Hankook Lee. Timeperceiver: An encoder-decoder framework for generalized time-series forecasting. *Advances in Neural Information Processing Systems*, 38:136136–136165, 2026.

Zhengyu Li, Xiangfei Qiu, Yuhan Zhu, Xingjian Wu, Jilin Hu, Chenjuan Guo, and Bin Yang. GCGNet: Graph-consistent generative network for time series forecasting with exogenous variables. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=EO5jwQ5NCw.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 11116–11140, 2024.

Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Timer-xl: Long-context transformers for unified time series forecasting. In *International Conference on Learning Representations*, volume 2025, pp. 83982–84006, 2025.

Tom Looby, Matthew Reinke, David Donovan, Travis Gray, Mike Messineo, and Andrei Khodak. Convolutional neural networks for heat flux model validation on nstx-u. *IEEE Transactions on Plasma Science*, 48(1):3–13, 2019.

F Matos, V Menkovski, F Felici, A Pau, F Jenko, TCV Team, and EUROfusion MST1 Team. Classification of tokamak plasma confinement states with convolutional recurrent neural networks. *Nuclear Fusion*, 60(3):036022, 2020.

Dongik Park, Hyunwoo Ryu, Suahn Bae, Keondo Park, and Hyung-Sin Kim. T1: One-to-one channel-head binding for multivariate time-series imputation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=IAnIlFsPEW.

Richard A Pitts, Xavier Bonnin, Frédéric Escourbiac, Heinke Frerichs, JP Gunn, Takeshi Hirai, AS Kukushkin, Elizaveta Kaveeva, MA Miller, David Moulton, et al. Physics basis for the first iter tungsten divertor. *Nuclear Materials and Energy*, 20:100696, 2019.

Bo Shi, Zhen-Dong Yang, Bin Zhang, Cheng Yang, Kai-Fu Gan, Mei-Wen Chen, Jin-Hong Yang, Hui Zhang, Jun-Li Qi, Xian-Zu Gong, et al. Heat flux on east divertor plate in h-mode with lhcd/lhcd+ nbi. *Chinese Physics Letters*, 34(9):095201, 2017.

Yuxuan Shu and Vasileios Lampos. Sonnet: Spectral operator neural network for multivariable time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 25419–25427, 2026.

Xiao Wang, Zikang Yan, Hao Si, Zhendong Yang, Qingquan Yang, Dengdi Sun, Wanli Lyu, and Jin Tang. Revisiting heat flux analysis of tungsten monoblock divertor on east using physics-informed neural network. *arXiv preprint arXiv:2508.03776*, 2025.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long. Timexer: Empowering transformers for time series forecasting with exogenous variables. *Advances in Neural Information Processing Systems*, 37:469–498, 2024.

Zhendong Yang, Pei He, HongFeng Yan, Zhang Bin, Shu Shuangbao, Fuqiong Wang, Meiwen Chen, Guozhang Jia, and Xianzu Gong. The development of a three-dimensional finite element method code for the heat flux analysis of tungsten monoblock divertor on east. *Fusion Engineering and Design*, 152:111448, 2020.

Wei Zheng, Fengming Xue, Zhongyong Chen, Dalong Chen, Bihao Guo, Chengshuo Shen, Xinkun Ai, Nengchao Wang, Ming Zhang, Yonghua Ding, et al. Disruption prediction for future tokamaks using parameter-based transfer learning. *Communications Physics*, 6(1):181, 2023.

Pengfei Zhou, Yunlong Liu, Junli Liang, Qi Song, and Xiangyang Li. Crosslinear: Plug-and-play cross-correlation embedding for time series forecasting with exogenous variables. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 4120–4131, 2025.

A Related Work

A.1 Divertor Heat-Flux Analysis

Accurate divertor heat-flux analysis is essential for understanding plasma-wall interactions and protecting plasma-facing components in magnetic-confinement fusion devices. Existing studies mainly rely on infrared thermal imaging technology. First, the surface temperature of the divertor target plate is measured by an IR camera. Then, the heat flux is inferred by solving the heat conduction equations related to material properties and boundary conditions (Lee et al., 2017; Herrmann et al.,

1995). For example, KSTAR employs the NANTHELOT heat-flux reconstruction code to solve the heat diffusion equation from measured tile surface temperature, while EAST calculates divertor heat-flux distributions from IR measurements using DFLUX or FEM-based solvers (Lee et al., 2017; Shi et al., 2017). These pipelines are often computationally expensive and are commonly used for offline post-shot analysis or require additional acceleration for real-time deployment. To improve efficiency, recent studies have introduced learning-based methods into heat-flux analysis. For example, approaches based on physics-informed neural networks (PINNs) have been explored to accelerate heat-equation solving toward real-time heat-flux estimation, while convolutional neural network (CNN)-based methods have been used to reconstruct parameterized heat-flux profiles from simulation-generated thermocouple signals (Aymerich et al., 2023; Looby et al., 2019). However, most existing heat-flux analysis methods either rely on thermal imaging measurements or focus on solving heat-conduction equations, while direct reconstruction of time-resolved divertor heat-flux profiles from multi-source discharge signals remains less explored.

A.2 Multivariate Time-Series Modeling

Multivariate time series modeling has been extensively studied to capture temporal dependencies and cross-variable interactions across multiple channels. In recent years, general time series models have achieved significant performance from different modeling perspectives. For example, iTransformer treats each variate as a token by inverting the dimension to enhance cross-variable dependency modeling (Liu et al., 2024). Timer-XL (Liu et al., 2025) represents time series as a long context token sequence and uses Transformer for multivariate sequence modeling. SimpleTM (Chen et al., 2025) combines signal processing ideas with a lightweight attention structure to provide a simple yet effective multivariate prediction baseline. TimeFilter (Hu et al., 2025) captures local dynamic patterns by constructing a spatiotemporal graph structure and performing patch-level filtering. TimePerceiver (Lee & Lee, 2026) uses an encoder-decoder structure and learnable queries to adapt to prediction targets at different time locations. Sonnet (Shu & Lampos, 2026) captures temporal dynamics in the spectral domain and captures cross-variable relationships by utilizing spectral consistency between variables. Some methods also focus on time series prediction under exogenous variable conditions. For example, TimeXer (Wang et al., 2024) fuses endogenous and exogenous variable information through patch-wise self-attention and variate-wise cross-attention, while GCGNet (Li et al., 2026) uses graph structure consistency constraints to model the correlation between exogenous and target variables. Meanwhile, time-series imputation provides useful insights into recovering unavailable variables from observed contexts. Representative works such as T1 (Park et al., 2026) combine a CNN-Transformer hybrid architecture with a channel-head binding mechanism to achieve robust imputation. Although these methods provide useful paradigms for multivariate time-series modeling, divertor heat-flux reconstruction differs from standard forecasting, exogenous-variable prediction, and partial imputation, because all target heat-flux channels are unavailable throughout the current input window and must be reconstructed from physically different macroscopic plasma-state signals.

B Experiment

B.1 Dataset

We construct DivMPS2HF, a multi-source time-series dataset for reconstructing divertor heat-flux profiles from macroscopic plasma-state signals. The dataset contains data from 77 discharge shots, among which 64 shots are used for training, 3 shots for validation, and 10 shots for testing. To avoid cross-shot information leakage, the dataset is split at the discharge-shot level, ensuring that samples from the same discharge shot do not appear in different subsets.

Before constructing the dataset, all selected signals are aligned to a unified temporal grid with a sampling rate of 1 kHz. The selected signals have heterogeneous sampling rates: most signals are recorded around 1 kHz, while some diagnostic signals are sampled at either lower or much higher frequencies. For example, $\mathcal{D}_\alpha$-related signals can be sampled at up to 250 kHz. To obtain synchronized multi-source inputs, we resample all signals to the common 1 kHz temporal grid before model training and evaluation. Signals with higher sampling rates are downsampled, while signals with lower sampling rates are mapped to the aligned time grid using temporal resampling.

After temporal alignment, samples are generated on the fly during data loading by applying a sliding window of $T = 500$ time steps to each discharge shot, corresponding to 0.5 s at 1 kHz. A stride of one time step is used to densely capture the temporal evolution of the heat-flux profiles. This sampling procedure yields approximately $150{,}504$, $11{,}838$, and $28{,}982$ windows for training, validation, and testing, respectively.

The model input consists of 67 channels of macroscopic plasma-state signals, which include not only traditional diagnostic measurements but also plasma state diagnostic signals, equilibrium and configuration parameters and signals related to external actuators and device control. These signals characterize the global and boundary plasma states related to the evolution of divertor heat load during the discharge process.

The reconstruction target is the time-resolved heat-flux profile on the divertor target. For each time step, the heat-flux profile is represented as a 116-dimensional vector, corresponding to 116 sampling channels in the radial direction of the target. The reference heat-flux profiles are obtained from post-shot heat-flux analysis and are used as supervised labels for model training and evaluation. It should be emphasized that infrared images and heat-conduction inversion results are not used as model inputs. During inference, SafeDivertor only takes macroscopic plasma-state signals as input and reconstructs the target heat-flux profile through a direct neural forward pass. Detailed information about the input signals, including their diagnostic sources and physical meanings, is provided in Table 6. The DivMPS2HF dataset will be made available upon reasonable request for academic research purposes.

B.2 METRICS

We evaluate the reconstructed divertor heat-flux profiles using five metrics: **MSE**, **MAE**, **SSIM**, **LSD**, and **LSD-HF**. Mean Squared Error (MSE) and Mean Absolute Error (MAE) are used as standard point-wise error metrics to evaluate the point-wise numerical accuracy of reconstructed heat flux in the time domain. Given N test samples, the reconstructed heat-flux profiles $\hat{\mathbf{Y}} \in \mathbb{R}^{N\times O\times T}$ and the ground-truth heat-flux profiles $\mathbf{Y} \in \mathbb{R}^{N\times O\times T}$, **MSE** and **MAE** are defined as follows:

$$\mathrm{MSE} = \frac{1}{NOT}\sum_{n=1}^{N}\sum_{o=1}^{O}\sum_{t=1}^{T}\left(\mathbf{Y}_{n,o,t} - \hat{\mathbf{Y}}_{n,o,t}\right)^2, \tag{16}$$

$$\mathrm{MAE} = \frac{1}{NOT}\sum_{n=1}^{N}\sum_{o=1}^{O}\sum_{t=1}^{T}\left|\mathbf{Y}_{n,o,t} - \hat{\mathbf{Y}}_{n,o,t}\right|. \tag{17}$$

Here, $O = 116$ denotes the number of target heat-flux channels, and T denotes the length of the time window.

SSIM is used to evaluate the structural similarity of reconstructed spatiotemporal heat flux maps, taking into account both temporal evolution and radial profile distribution. For each sample, the heat-flux profile is treated as a two-dimensional map over radial channels and time steps. We first compute SSIM on each sample and then average the scores over all test samples:

$$\mathrm{SSIM} = \frac{1}{N}\sum_{n=1}^{N}\mathrm{SSIM}\left(\hat{\mathbf{Y}}_n, \mathbf{Y}_n\right), \tag{18}$$

where N denotes the number of test samples. For the n-th sample, the sample-level SSIM is computed by averaging the local SSIM values over all evaluated 7×7 blocks:

$$\mathrm{SSIM}\left(\hat{\mathbf{Y}}_n, \mathbf{Y}_n\right) = \frac{1}{K}\sum_{k=1}^{K}\mathrm{SSIM}_{n,k}, \tag{19}$$

where K is the number of evaluated local blocks in each heat-flux map. The local SSIM of the k-th block is defined as

$$\mathrm{SSIM}_{n,k} = \frac{(2\mu_{\mathbf{Y}_{n,k}}\mu_{\hat{\mathbf{Y}}_{n,k}} + \epsilon_1)(2\mathrm{Cov}(\mathbf{Y}_{n,k}, \hat{\mathbf{Y}}_{n,k}) + \epsilon_2)}{(\mu^2_{\mathbf{Y}_{n,k}} + \mu^2_{\hat{\mathbf{Y}}_{n,k}} + \epsilon_1)(\sigma^2_{\mathbf{Y}_{n,k}} + \sigma^2_{\hat{\mathbf{Y}}_{n,k}} + \epsilon_2)}. \tag{20}$$

Here, $\mathbf{Y}_{n,k}$ and $\hat{\mathbf{Y}}_{n,k}$ denote the k-th local 7×7 block of the ground-truth and reconstructed heat-flux maps for the n-th test sample, respectively. μ, σ^2, and $\mathrm{Cov}(\cdot,\cdot)$ denote the local mean, variance, and covariance, and ϵ_1, ϵ_2 are small constants for numerical stability.

Let $\mathbf{S}$ and $\hat{\mathbf{S}}$ denote the power spectra of the ground-truth and reconstructed heat-flux profiles after STFT:

$$\mathbf{S} = |\mathrm{STFT}(\mathbf{Y})|^2 + \epsilon, \quad \hat{\mathbf{S}} = |\mathrm{STFT}(\hat{\mathbf{Y}})|^2 + \epsilon, \tag{21}$$

where ϵ is a small constant for numerical stability.

For each evaluated spectrum, we first compute the log-spectral distance:

$$d_{n,o,m} = \sqrt{\frac{1}{F}\sum_{f=1}^{F}\left(\log \hat{\mathbf{S}}_{n,o,f,m} - \log \mathbf{S}_{n,o,f,m}\right)^2}, \tag{22}$$

where F denotes the number of frequency bins. The final **LSD** score is obtained by averaging over all test samples, target channels, and STFT frames:

$$\mathrm{LSD} = \frac{1}{NOM}\sum_{n=1}^{N}\sum_{o=1}^{O}\sum_{m=1}^{M} d_{n,o,m}, \tag{23}$$

where N is the number of test samples, O is the number of target heat-flux channels, and M is the number of STFT frames.

Similarly, **LSD-HF** is computed by restricting the frequency bins to the high-frequency set $\mathcal{F}_{\mathrm{HF}}$:

$$d^{\mathrm{HF}}_{n,o,m} = \sqrt{\frac{1}{|\mathcal{F}_{\mathrm{HF}}|}\sum_{f\in\mathcal{F}_{\mathrm{HF}}}\left(\log \hat{\mathbf{S}}_{n,o,f,m} - \log \mathbf{S}_{n,o,f,m}\right)^2}, \tag{24}$$

$$\text{LSD-HF} = \frac{1}{NOM}\sum_{n=1}^{N}\sum_{o=1}^{O}\sum_{m=1}^{M} d^{\mathrm{HF}}_{n,o,m}. \tag{25}$$

All metrics are computed only on the 116 target heat-flux channels.

B.3 Baselines

We select representative state-of-the-art time series models from recent years as benchmarks for comparison. Specifically, we include general multivariate time series forecasting models, including Sonnet (Shu & Lampos, 2026), Timer-XL (Liu et al., 2025), TimeFilter (Hu et al., 2025), SimpleTM (Chen et al., 2025), TimePerceiver (Lee & Lee, 2026), and iTransformer (Liu et al., 2024). Furthermore, since our task naturally involves using multi-source signals as external conditions to reconstruct the target heat flux sequence, we also includ exogenous variable prediction models, including CrossLinear (Zhou et al., 2025) and TimeXer (Wang et al., 2024). Additionally, we use T1 (Park et al., 2026) as the time series imputation benchmark, where multi-source signal channels are treated as observed variables, and heat flux channels are treated as structured missing variables.

C Results

C.1 Visualization

Figure 4 presents qualitative comparisons on four samples from different test shots. Most baselines can identify the approximate radial locations of the dominant heat-flux bands, but their reconstruction characteristics vary considerably. Several methods produce over-smoothed profiles with limited temporal variation, while others exhibit noticeable intensity bias, discontinuous patterns, or horizontal artifacts. Although the original T1 backbone preserves the overall spatiotemporal structure relatively well, it still tends to underestimate local high-intensity regions and suppress fine-grained temporal fluctuations. In contrast, SafeDivertor more faithfully recovers the positions and intensities of the main heat-flux bands while retaining clearer transient variations across different discharge conditions. These qualitative results are consistent with the quantitative improvements reported in the main experiments, further demonstrating the effectiveness of the proposed task-driven designs.

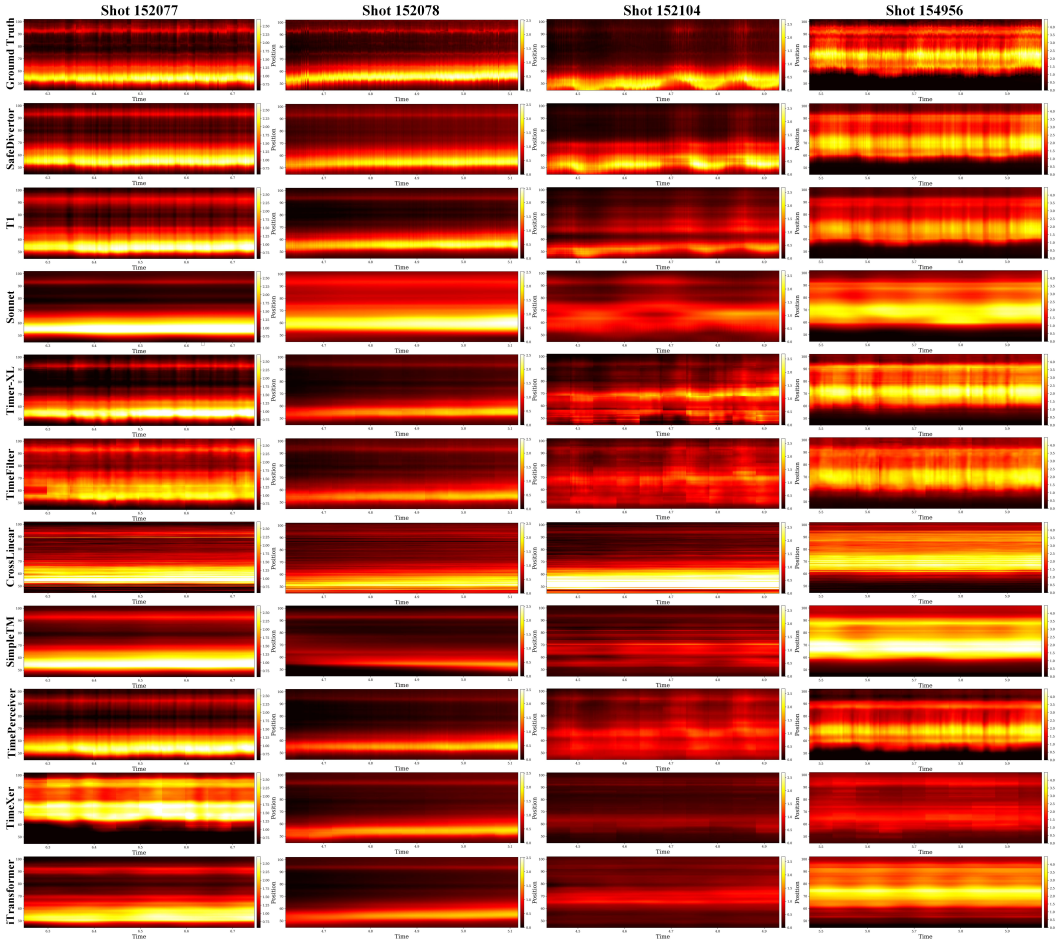


Figure 4: Qualitative comparison of four samples, one from each of four test shots. Each column corresponds to one sample, while the rows show the ground truth and reconstruction results of SafeDivertor and the evaluated baselines.

Table 6: Description of input signals used in DivMPS2HF. Signals marked with * represent multiple measurements from different channels of the same diagnostic system.

Signal names	Description
a	Plasma minor radius, describing the radial size of the plasma configuration.
B_t	Toroidal magnetic field strength at the plasma center.
β_N, β_P	Normalized and poloidal plasma beta parameters, describing plasma pressure and confinement performance.
C_{III}	C III impurity emission signal, providing information on impurity radiation in the plasma edge region.
MP*	Magnetic probe signals from different poloidal locations, providing information on magnetic field fluctuations and plasma dynamics.
XUV*	Extreme ultraviolet (XUV) signals from different channels, providing information on plasma radiation and impurity behavior
$\mathcal{D}_\alpha$*	$\mathcal{D}_\alpha$ emission signals from different channels, characterizing edge plasma radiation and recycling processes.
$n_{e,point}$	Line-averaged electron density signal measured by the POlarimeter-INTerferometer (POINT) diagnostic, providing information on plasma density evolution.
$n_{e,target}$	Density reference signal for feedback control in the Plasma Control System (PCS), specifying the target plasma density
DTC	Distance between the strike point and the corner of the divertor target plate, providing information on the strike-point position
drsep	the radial distance dRsep between the primary and secondary X-point surfaces on the outer mid-plane
ENG	Plasma stored energy measured by the diamagnetic diagnostic
NGP	Neutral gas pressure signals measured in the upper and lower divertors, providing information on divertor neutral pressure conditions.
GAP_{in}	Distance between the inboard last closed flux surface (LCFS) and the first wall, indicating the inboard plasma-wall gap
ECE*	Electron cyclotron emission signals from different channels, providing information related to electron temperature and plasma radiation.
HX*	Hard X-ray diagnostic signals from different channels.
κ	Plasma elongation parameter describing plasma cross-section shape.
HMP	High-frequency magnetic probe signal for measuring magnetic fluctuations and characterizing MHD and Alfvénic activities
l_i	Plasma internal inductance parameter.
Liii*	Lithium impurity emission signals from different channels.
Div_{is}	Divertor Langmuir probe ion saturation current measurement.
lmezx	Plasma vertical position error signal used for plasma position control.
nel	Line-averaged electron density measurement.
pcic*	Fast control coil current signals, providing information on the coil currents used for fast feedback control of plasma position and vertical stability
I_p	Plasma current measurement.
V_{loop}	Loop voltage measurement.
P_{ECRH}*	Electron cyclotron resonance heating related power signals, describing ECRH input conditions.
P_{ICRF}*	Ion cyclotron resonance heating related power signals, describing ICRF heating conditions.
P_{LHCD}*	Lower hybrid current drive absorbed power measurement.
P_{total}	Total absorbed auxiliary heating power.
q_{95}	Edge safety factor evaluated at 95% of the normalized poloidal magnetic flux, closely related to plasma MHD stability.
r_x	Radial X-point position measurements.
SMBI	Supersonic molecular beam injection related signal.
δ_{bot}, δ_{top}	Lower and upper plasma triangularity, respectively, characterizing the plasma shape and magnetic configuration
vb	Bremsstrahlung radiation measurement signal.
W_{mhd}	Plasma stored energy derived from MHD equilibrium reconstruction, providing an estimate of the plasma energy content.
z_{cur}	Vertical position of the plasma current centroid.
z_x	Vertical X-point position measurements.