Energy and AI 26 (2026) 100881

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.elsevier.com/locate/egyai

Relaxed Shape Alignment and Temporal Distortion Loss for improving solar ramp events prediction

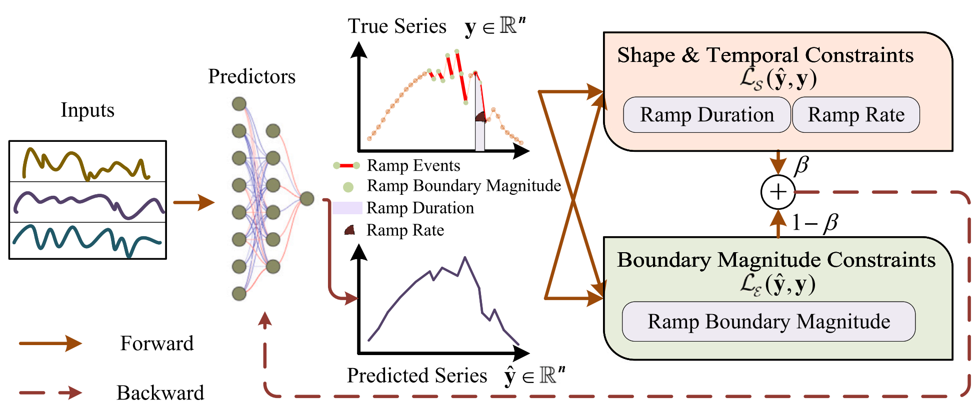
Tao Jing [a], Shanlin Chen [b,*], Yinghao Chu [c], Mengying Li [a,*]

[a] Department of Mechanical Engineering & Research Center for Data Sciences and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

[b] Wind Engineering and Renewable Energy Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland

[c] Department of Systems Engineering, City University of Hong Kong, Hong Kong Special Administrative Region

GRAPHICAL ABSTRACT



HIGHLIGHTS

- A pioneering loss function to improve solar ramp event forecasting.
- Theory-based modeling constrains solar ramp rates, durations, and boundary magnitudes.
- The proposed loss boosts ramp event predictive accuracy and outperforms existing losses.
- The proposed loss exhibits excellent robustness.

ARTICLE INFO

Keywords:
Solar ramp forecasting
Deep learning
Loss function
Temporal distortion
Time series forecasting

ABSTRACT

The inherent uncertainty of the weather system frequently causes solar ramp events, limiting the efficient integration of solar power into the grid. Although the use of multi-modal data sources with deep learning models has improved solar forecasting accuracy as measured by statistical error metrics, the common reliance on Euclidean distance loss functions often fails to accurately extract the characteristics of ramp events, which is essential for grid operators to integrate solar power. Therefore, we introduced the Relaxed Shape Alignment and Temporal Distortion Loss (RSTLoss), a novel differentiable loss function designed for deep learning-based solar

* Corresponding authors.
E-mail addresses: shanlin.chen@epfl.ch (S. Chen), mengying.li@polyu.edu.hk (M. Li).

https://doi.org/10.1016/j.egyai.2026.100881
Received 19 March 2026; Received in revised form 16 August 2026; Accepted 25 August 2026
Available online 9 September 2026

prediction models. The effectiveness of RSTLoss is thoroughly verified through extensive experiments, showing improvements of up to 7.48% in the Ramp Score and 13.17% in the Temporal Distortion Index compared to existing loss functions. These findings indicate that RSTLoss significantly enhances ramp prediction accuracy while also exhibiting excellent model compatibility.

1. Introduction

In response to the challenges posed by global climate change, the exploitation and utilization of renewable energy have become critically important [1,2]. Solar energy, renowned for its inexhaustible supply, sustainability, and zero emissions, has emerged as a principal driver of the energy revolution [3]. However, a major challenge in the widespread and efficient use of solar energy is the seamless integration of these intermittent power sources into existing grid systems. Accurate forecasting of solar irradiance variations across different time scales is essential to optimize the use of renewable energy [4,5]. However, the inherent complexity and stochastic nature of weather systems present formidable challenges for accurately predicting ground-level solar irradiance [6], especially during cloudy periods characterized by intermittent, sharp fluctuations known as "ramp events" [7]. Due to the sensitivity of solar power systems to irradiance fluctuations, severe ramp events can disconnect photovoltaic (PV) units from the grid, thus compromising overall grid stability [8,9]. Consequently, the accurate prediction of solar ramp events is essential for distribution operators to implement preventive measures and minimize restoration costs. Accurate forecasts of ramp timing, duration, rate, and magnitude can support dispatch planning, reserve scheduling, and energy-storage coordination. By reducing unexpected PV fluctuations and supply–demand imbalances, improved ramp forecasting enhances grid reliability and facilitates higher photovoltaic penetration.

In the context of specific application requirements, the definition of ramp events is subject to variation; consequently, researchers have proposed numerous approaches to establish these definitions [10]. According to the Alberta Pilot Project technical report, a ramp event is defined as a significant variation in power output that occurs within a relatively short time interval [11]. Kamath [12] defined ramp events based on time intervals, where a ramp event occurs when the amplitude difference between the start and end points exceeds a predetermined threshold. Subsequently, Kamath [13] proposed a definition based on extreme values: a ramp event is considered to have occurred if the difference between the maximum and minimum values exceeds a specific threshold during a given time interval. Despite varying definitions, ramp events are consistently characterized by three key parameters: ramp duration, ramp rate, and ramp boundary magnitude [7]. Ramp duration refers to the time span of a ramp event. Ramp rate describes the rate of change in irradiance during the event, effectively capturing the steepness of the ramp. The ramp boundary magnitude is defined as the difference in irradiance between the start and end points of the ramp event. Ramp events are classified into two categories: up-ramp and down-ramp. In solar energy systems, the primary concern is down-ramp events, which occur when cloud cover obscures the sun [10].

1.1. Current advances on detection and prediction of solar ramps

In the field of solar forecasting, the majority of research primarily aims to enhance the accuracy of predictions, which are evaluated using statistical metrics such as the Root Mean Square Error (RMSE) across the entire dataset. However, there is a notable scarcity of studies specifically focused on ramp forecasting. The detection and prediction of solar ramp events, which are crucial yet underexplored aspects of this field, can be briefly summarized as follows.

The early work of Florita et al. [14] employed the Swinging Door Algorithm (SDA), a data compression technique, to detect ramp events. Cui et al. [15] later enhanced the SDA through dynamic programming optimization. However, these methods are limited to the detection of ramp events and fail to provide forewarning. Existing research on the prediction of solar ramp events is primarily based on multi-modal input features, with data sources varying according to prediction timescales. Intra-hour solar forecasting utilizes cloud motion information derived from all-sky imagery [10,16–19], while intra-day solar forecasting relies on satellite imagery to analyze the dynamics of large-scale cloud systems [6,20–23]. For day-ahead and longer forecasting horizons, numerical weather prediction (NWP) data serve as the primary input [24,25]. Various prediction models have been developed, employing both statistical and deep learning methods to capture the complicated relationships among meteorological variables and solar ramp events to improve prediction accuracy.

From the perspective of statistical methods, Zhu et al. [26] developed an approach that skillfully combines Credal Networks and the Imprecise Dirichlet Model to capture the fuzzy correlation between solar ramp events and meteorological variables, thereby improving the reliability and accuracy of ramp event forewarning. The Cooperative Institute for Research in the Atmosphere (CIRA) developed CIRACAST [27], an advanced solar irradiance prediction system that incorporates high-resolution satellite imagery and wind field forecasts to predict solar irradiance and ramp events up to three hours ahead. Zhang et al. [28] developed a spatio-temporal categorical point process model that characterizes the spatial correlations of solar ramp effects and their interaction mechanisms, using meteorological variables including temperature, humidity, and cloud density. From the perspective of the deep learning model, Hirata and Aihara [29] systematically identified and quantified the uncertainty factors affecting the prediction of ramp events, then employed a deep learning technique to predict the up-ramp and down-ramp events in solar irradiance by integrating the quantification. Chu et al. [10] integrated cloud tracking techniques with artificial neural networks to predict 10-min advance ramp events with a time resolution of one minute. Wen et al. [17] proposed a CNN-based multi-step strategy that enhances the prediction of solar power ramp events by integrating spatio-temporal cloud motion information from sequential all-sky imagery.

1.2. Challenges of ramp prediction

Although several studies have advanced the prediction of solar ramp events by proposing various methodologies, the primary focus has been on optimizing deep learning architectures and integrating multiple data sources. Kumari and Toshniwal [30] and Ajith and Martínez-Ramón [31] have provided comprehensive reviews of deep learning-based models to predict solar irradiance. These reviews underscore significant advances yet also point out a critical limitation: the prevalent reliance on Euclidean Distance (ED)-based loss functions. Given that the use of RMSE skill scores has been strongly advocated to assess the accuracy of deterministic predictions of solar irradiance [32], most deep learning models for solar prediction employ RMSE or Mean Square Error (MSE) as their primary loss functions [30,31]. However, Vallance et al. [33] has conducted qualitative and quantitative analyzes of ED-based metrics, revealing their insufficiency in assessing predictions of solar ramp events. Their research indicates that ED-based metrics struggle to accurately reflect the models' capability to capture actual fluctuations, particularly in the context of ramp events. Further investigations by Le Guen and Thome [34] and Xu et al. [35] have elucidated additional shortcomings of the ED-based loss function in deep learning applications.

Nomenclature

Variables

α	Shape parameter of the Fréchet distribution
β	Weight parameter in the RSTLoss function
$\Delta(\mathbf{y},\hat{\mathbf{y}})^q_{ij}$	Euclidean Distance between true and predicted time series
γ	Smoothing parameter
$\mathbb{E}_\gamma[W_t]$	Expected temporal distortion
$\hat{\mathbf{y}}$	Predicted time series
$\mathbf{A} \in \{0,1\}^{n\times n}$	A binary matrix for describing the alignment state between time series
$\mathbf{y}$	True time series
$\mathcal{R}_{ij}$	Accumulated cost of the optimal alignment path
$\nabla_y \Delta(\mathbf{y},\hat{\mathbf{y}})$	Gradient of the distance matrix
π	Alignment path
$\mathbf{y}_r$	Ramp event boundary magnitude
$\tilde{\mu}$	Position parameter of extreme value distribution
$\tilde{\sigma}$	Scale parameter of extreme value distribution
g	Empirical constant
m	Positional parameter of the Fréchet distribution
$\mathcal{A}_{n,n}$	Set of all admissible paths

Abbreviations

BNI	Beam Normal Irradiance
BSRN	Baseline Surface Radiation Network
CIRA	Cooperative Institute for Research in the Atmosphere
CSI	Clear Sky Index
DDTW	Derivative Dynamic Time Warping
DeepTCN	Deep Temporal Convolutional neural Network
DHI	Diffuse Horizontal Irradiance
DP	Dynamic Programming
DTW	Dynamic Time Warping
ED	Euclidean Distance
GEV	Generalized Extreme Value
GHI	Global Horizontal Irradiance
GRU	Gated Recurrent Unit
MBE	Mean Bias Error
MLP	Multilayer Perceptrons
MSE	Mean Square Error
NSRDB	National Solar Radiation Data Base
NWP	Numerical Weather Prediction
PDFs	Probability Density Functions
PV	Photovoltaic
QC	Quality Control
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RS	Ramp Score
RSTLoss	Relaxed Shape Alignment and Temporal Distortion Loss
SDA	Swinging Door Algorithm
soft-DTW	soft-Dynamic Time Warping
soft-WDDTW	soft-Weighted Derivative Dynamic Time Warping
TDI	Temporal Distortion Index
TSMixer	Time-Series Mixer

Nevertheless, the development of specialized loss functions tailored to predict ramp events remains a significant yet unaddressed research gap. Conventional ED-based loss functions do not adequately capture the three key parameters of ramp events. Given the temporal sensitivity and rapid rate of fluctuation in ramp events, loss functions must penalize prediction errors differentially across these characteristics while maintaining appropriate sensitivity to non-ramp conditions.

1.3. Contributions of this work

Given the limitations of ED-based loss functions, this work introduces a novel differentiable loss function, termed Relaxed Shape Alignment and Temporal Distortion Loss (RSTLoss), for deep learning-based solar ramp-event forecasting. Unlike existing methods that utilize multimodal inputs such as all-sky imagery, satellite imagery, or NWP data, RSTLoss leverages the intrinsic capabilities of deep learning frameworks to effectively characterize solar ramp events using only widely available meteorological data time series. The approach enhances the sensitivity of deep learning predictors to challenging ramp events under conditions of limited data availability. RSTLoss facilitates a comprehensive characterization of ramp events by simultaneously optimizing the prediction of ramp duration, ramp rate, and ramp boundary magnitude. The multi-dimensional loss function optimization strategy allows deep learning predictors to concurrently focus on the temporal dynamics, morphological features, and amplitude variations of the ramp events. This endows the predictor with the ability to capture the complex features of ramp events comprehensively and accurately, thus significantly enhancing prediction accuracy. Specifically, the scientific and practical contributions of the current work are as follows:

- The primary contribution of this work is the development of RSTLoss, a novel differentiable loss function designed specifically to overcome the limitations of conventional ED-based metrics (like MSE) in solar forecasting. Unlike standard loss functions that often fail to capture sharp fluctuations, RSTLoss allows deep learning predictors to simultaneously optimize for the three essential parameters of solar ramp events: ramp duration, ramp rate, and ramp boundary magnitude, thereby enabling a more accurate and comprehensive characterization of complex irradiance dynamics.
- From an algorithmic design standpoint, we innovated by integrating a soft-Weighted Derivative Dynamic Time Warping (soft-WDDTW) module directly into the loss calculation pipeline. This component modifies the traditional soft-Dynamic Time Warping (soft-DTW) approach by incorporating derivative information and a temporal distortion penalty weight, which forces predictors to prioritize the alignment of trends and slopes rather than just absolute values. This architectural choice specifically targets the shape and temporal aspects of the prediction, allowing the model to elastically align predicted sequences with ground truth data to mitigate the temporal delays often seen in standard regression models.
- To address the inherent weakness of warping methods in capturing peak values, we introduced a novel magnitude constraint module based on Extreme Value Theory (EVT) within the loss architecture. By embedding a Fréchet distribution-based kernel function into the loss objective, the architecture explicitly penalizes the underestimation of extreme fluctuations and ramp

boundaries. This design choice creates a dual-focus optimization landscape where the model is simultaneously constrained by temporal alignment (via soft-WDDTW) and magnitude severity (via the EVT kernel), ensuring that the generated forecasts respect the physical intensity of solar ramp events.

In summary, existing differentiable time-series losses mainly improve sequence alignment or temporal matching. In contrast, RSTLoss extends soft-DTW by integrating derivative-based alignment, temporal-distortion-aware weighting, and a Fréchet-distribution-based extreme-value loss into a unified differentiable objective. This design explicitly constrains ramp duration, ramp rate, and ramp boundary magnitude, making RSTLoss specifically tailored for solar ramp-event forecasting. The remainder of the paper is organized as follows. Section 2 illustrates the theoretical limitations of ED-based and soft-DTW-based loss functions in ramp event prediction. Section 3 presents the proposed RSTLoss. Section 4 details the case study, including the dataset, selected prediction models, assessment metrics for ramp prediction, and the training procedure used. The forecasting results are presented and discussed in Section 5. Finally, Section 6 summarizes the key findings.

2. Preliminaries

In this Section, the inherent limitations of conventional loss functions in deep learning for ramp event prediction are first analyzed from theoretical perspectives.

2.1. Conventional loss functions and their limitations

Take univariate time series as an example. The true series is denoted as $\mathbf{y} = (y_0, \dots, y_{n-1}) \in \mathbb{R}^n$, and the predicted series is denoted as $\hat{\mathbf{y}} = (\hat{y}_0, \dots, \hat{y}_{n-1}) \in \mathbb{R}^n$. The accuracy of multi-step time series forecasting is typically assessed by computing similarity metrics between the predicted series $\hat{\mathbf{y}}$ and the actual series $\mathbf{y}$. These similarity metrics for time series can be divided into two primary categories: ED-based metrics and those based on DTW [36].

2.1.1. Euclidean distance loss

While ED metrics such as MSE are commonly employed to assess prediction accuracy in time series models, they exhibit significant limitations when used in the backpropagation process of deep learning models. This is particularly problematic in the context of predicting ramp events, where sub-optimal or even inappropriate parameter tuning decisions may result. For normally distributed data $\mathbf{y} \sim \mathcal{N}(\mu, \sigma^2)$, employing MSE as a loss function equates to using Maximum Likelihood Estimation (MLE) to predict the mean μ of $\mathbf{y}$. This relationship is demonstrated in Eq. (2):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{1}$$

$$\begin{aligned} \operatorname{argmin}_\theta \mathrm{MSE}_{(\mathbf{y},\hat{\mathbf{y}})} &\Leftrightarrow \operatorname{argmin}_\theta -\log\left(P\left(\mathbf{y}|\mu,\sigma^2\right)\right) \\ &= \operatorname{argmin}_\theta -\log\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\mu-\mathbf{y})^2}{2\sigma^2}\right)\right) \\ &= \operatorname{argmin}_\theta \left(\frac{(\mu-\mathbf{y})^2}{2\sigma^2} + \log\left(\sigma\sqrt{2\pi}\right)\right). \end{aligned} \tag{2}$$

Note that this Gaussian assumption is introduced only to explain the statistical foundation of conventional ED-based loss functions, rather than to characterize the statistical distribution of solar ramp events.

However, this symmetric conditional-mean objective does not explicitly distinguish ordinary errors from errors occurring at abrupt ramp boundaries. The boundary magnitudes of solar ramp events ($\mathbf{y}_r$) represent local extrema of rapid irradiance variations and therefore exhibit statistical characteristics that differ from ordinary point-wise prediction errors. The EVT provides a principled framework for characterizing the tails and extrema of stochastic processes [37]. Recent deep-learning studies have also incorporated extreme value distributions or extreme mixture models to improve the prediction of rare and high-impact observations in time series [38,39]. Motivated by this theoretical connection, we adopt an EVT-informed asymmetric formulation to emphasize prediction errors associated with ramp-boundary magnitudes. The extreme value distribution is used as a loss design principle rather than as a strict distributional assumption for the complete ramp event sequence. For a representative extreme value likelihood with location parameter $\tilde{\mu}$ and scale parameter $\tilde{\sigma}$, the corresponding negative log-likelihood can be expressed as

$$\operatorname{argmin}_\theta -\log(P(\mathbf{y}_r|\tilde{\mu},\tilde{\sigma})) = \operatorname{argmin}_\theta\left(\frac{\mathbf{y}_r-\tilde{\mu}}{\tilde{\sigma}} + \exp\left(-\frac{\mathbf{y}_r-\tilde{\mu}}{\tilde{\sigma}}\right) + \log(\tilde{\sigma})\right), \tag{3}$$

where $\tilde{\mu}$ and $\tilde{\sigma}$ denote the location and scale parameters, respectively.

Unlike the symmetric quadratic objective induced by the Gaussian likelihood, the extreme value objective is asymmetric and assigns different penalties to deviations on the two sides of the location parameter. This distinction motivates an asymmetric weighting mechanism for ramp-boundary errors. Nevertheless, the proposed method does not require all ramp-boundary magnitudes to exactly follow a specific extreme-value distribution; instead, the extreme-value formulation serves as an EVT-informed kernel for enhancing the model's sensitivity to critical ramp boundaries.

When using ED-based loss functions, the gradient of the backpropagation is expressed as:

$$\frac{\partial \mathcal{L}_{\mathrm{ED}}}{\partial y_t} = \frac{\hat{y}_t - y_t}{n \cdot \mathcal{L}_{\mathrm{ED}}}. \tag{4}$$

The distinct nature of the optimization objectives in Eqs. (2) and (3) highlights the inherent bias introduced by ED-based loss functions in the prediction of ramp event boundary magnitudes. Further complicating this issue, the use of ED-based loss functions in ramp event predictions can lead to the averaging out of significant prediction errors at the ramp boundaries. This averaging effect weakens the gradients at these critical points, potentially causing adjacent time points to generate opposing gradients. Such opposing gradients can severely counteract each other, reducing the model's sensitivity to learn critical ramp event boundary magnitudes.

2.1.2. Dynamic time warping loss

Dynamic Time Warping (DTW) is an algorithm that identifies the optimal alignment path between two time series by allowing for local temporal distortions, thus facilitating the best match in their shapes. This adjustment is achieved by seeking a path that minimizes the alignment cost [40,41]. DTW represents the alignment between two sequences, $\mathbf{y}$ and $\hat{\mathbf{y}}$, using a binary matrix $\mathbf{A} \in \{0,1\}^{n\times n}$. In this matrix, an element $\mathbf{A}(i,j) = 1$ indicates that element y_i is aligned with $\hat{y}_j$, and $\mathbf{A}(i,j) = 0$ otherwise. The set of all admissible paths, $\mathcal{A}_{n,n}$, consists of paths that begin at the top-left corner (1,1) and end at the bottom-right corner (n,n), moving through the matrix using downward (↓), rightward (→), or diagonal (↘) steps, as visualized in Fig. 1. The optimization problem of DTW is formulated as:

$$\mathrm{DTW}(\mathbf{y},\hat{\mathbf{y}}) = \min_{\pi\in\mathcal{A}_{n,n}}\left(\sum_{(i,j)\in\pi}\Delta(\mathbf{y},\hat{\mathbf{y}})^q\right)^{1/q}, \tag{5}$$

where the alignment path π is a sequence of index pairs $((i_0,j_0),\dots,(i_{n-1},j_{n-1}))$, and $\Delta(\mathbf{y},\hat{\mathbf{y}})^q_{ij} = |y_i - \hat{y}_j|^q$ represents the qth power of the ED between $\mathbf{y}$ and $\hat{\mathbf{y}}$. Dynamic Programming (DP) algorithms solve the optimization problem by minimizing the alignment cost with a complexity of $\mathcal{O}(n^2)$, grounded in Bellman's principle of optimality [42]. The recursive formulation of this process is:

$$\mathcal{R}_{i,j} = \Delta(\mathbf{y},\hat{\mathbf{y}})^q_{ij} + \min(\mathcal{R}_{i-1,j},\mathcal{R}_{i,j-1},\mathcal{R}_{i-1,j-1}), \tag{6}$$

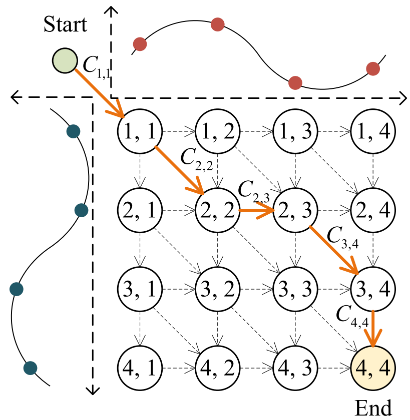


Fig. 1. Alignment of two time series $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ shown as a path through a directed acyclic graph. The alignment is encoded in a binary matrix $\mathbf{A} \in \{0,1\}^{n\times n}$, where $\mathbf{A}(i,j) = 1$ indicates pairing of $\hat{y}_i$ with y_j. The matrix $\mathcal{R}_{i,j}$ denotes the total cost of alignment, which DTW seeks to minimize.

where $\mathcal{R}_{ij}$ denotes the accumulated cost along the optimal path. This cost metric can also serve as a similarity measure between two time series, which is useful for assessing the prediction accuracy of events such as ramps [33].

Despite DTW's efficacy in measuring time series similarity, its non-differentiable nature poses a significant challenge when used as a loss function in deep learning models that rely on gradient-based optimization. To address this, Cuturi and Blondel [36] introduced a differentiable version called soft-DTW, which utilizes the log-sum-exp trick to smooth the minimum operator [36]:

$$\overset{\gamma}{\min}\{a_1, \ldots, a_n\} := \begin{cases} \min_{i\le n} a_i, & \gamma = 0, \\ -\gamma \log \sum_{i=1}^{n} \exp\left\{-a_i/\gamma\right\}, & \gamma > 0, \end{cases} \tag{7}$$

where γ is a smoothing parameter. The soft-DTW allows for a Gibbs distribution of possible alignment paths rather than a single optimal path and modifies the recursive process accordingly:

$$\mathcal{R}_{i,j} = \Delta(\mathbf{y}, \hat{\mathbf{y}})^q_{ij} + \overset{\gamma}{\min}(\mathcal{R}_{i-1,j}, \mathcal{R}_{i,j-1}, \mathcal{R}_{i-1,j-1}). \tag{8}$$

However, while soft-DTW significantly enhances the model's sensitivity to ramp events, it also introduces challenges during gradient backpropagation by excessively warping time to align the predicted and target series shapes. This can distort the temporal aspects of predictions, particularly affecting the accuracy of timing the ramp events. We will explore the implications of this temporal distortion in the context of soft-DTW's gradient updates in the subsequent sections.

Proposition. The gradient of soft-DTW during backpropagation is expressed as the expected alignment $\mathbb{E}_\gamma[\mathbf{A}_\pi]$, which can be derived as (see Appendix A.1 for detailed derivations):

$$\nabla \mathrm{DTW}^{\Delta}_{\gamma}(\mathbf{y}, \hat{\mathbf{y}}) = \mathbb{E}_\gamma[\mathbf{A}_\pi], \tag{9}$$

$$\mathbb{E}_\gamma[\mathbf{A}_\pi] = \sum_{\mathbf{A}_\pi \in \mathcal{A}_{n,n}} \mathbf{A}_\pi \cdot p_\gamma(\mathbf{A}_\pi), \tag{10}$$

$$p_\gamma(\mathbf{A}_\pi) = \frac{\exp\left\{-\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}}{\sum_{\mathbf{A}'_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}'_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}}, \tag{11}$$

where $\nabla \mathrm{DTW}^{\Delta}_{\gamma}(\mathbf{y}, \hat{\mathbf{y}})$ is the derivative with respect to the alignment cost, $\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle$ represents the scalar product of the pairwise matrix $\Delta(\mathbf{y}, \hat{\mathbf{y}})$ and the warping path $\mathbf{A}_\pi$. As $\gamma \to 0$, the exponential function in $p_\gamma(\mathbf{A}_\pi)$ becomes highly sensitive to $\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle$, emphasizing the alignment matrix $\mathbf{A}_\pi$ that minimizes this value (i.e. the optimal alignment), similar to classical DTW. Conversely, as $\gamma \to \infty$, all alignment matrices become equally probable due to the normalization effect of the exponential function, leading to a uniform distribution and an expected value biased towards linear alignment. In the soft-DTW optimization using gradient descent, the predicted series $\hat{\mathbf{y}}$ is updated in the direction:

$$\Delta y = -\eta \mathbb{E}_\gamma[\mathbf{A}_\pi] \cdot \nabla_y \Delta(\mathbf{y}, \hat{\mathbf{y}}), \tag{12}$$

where $\nabla_y \Delta(\mathbf{y}, \hat{\mathbf{y}})$ refers to the gradient of the distance matrix. Each gradient update thus modifies the alignment matrix. According to the definition of the Temporal Distortion Index (TDI) [43], this modification affects the temporal distortion metric as defined by:

$$W_t = \sum_{i,j} |i - j| \mathbf{A}^{(t)}_{ij}, \tag{13}$$

where $\mathbf{A}^{(t)}_{ij}$ represents the alignment matrix after t gradient updates. The cumulative temporal distortion of the predicted sequence is defined as:

$$W_T = \sum_{t=1}^{T} \mathbb{E}_\gamma[W_t], \tag{14}$$

where $\mathbb{E}_\gamma[W_t]$ represents the expected temporal distortion per update. Therefore, utilizing soft-DTW as the loss function leads to a progressive accumulation of temporal distortion effects during long-term training. In the absence of penalties for temporal distortion, soft-DTW facilitates time-axis warping to achieve sequence alignment, significantly impacting the accuracy of ramp duration predictions in the long term.

3. Methodology

To address the limitations of conventional loss functions in ramp prediction, we propose the novel RSTLoss, comprising two distinct components, as shown in Fig. 2. The first, termed soft-WDDTW, evaluates shape similarity and phase differences between the predicted and target time series. This component excels in capturing the overall trends of ramp events and effectively quantifies time axis offsets, thus minimizing temporal distortion and enhancing the accuracy of ramp rate predictions. The second component employs a kernel function based on the extreme value distribution, specifically tailored to enhance the prediction of boundary magnitudes in ramp events. The optimal convex combination of these two components forms RSTLoss, which significantly improves the prediction accuracy regarding the duration, rate, and boundary magnitude of ramp events.

3.1. Soft-weighted derivative dynamic time warping

The soft-DTW loss function is pivotal for forecasting ramp events as it enhances shape similarity between the predicted and target series [34,36]. Its inherently differentiable nature allows for seamless integration with gradient-based optimization processes, facilitating end-to-end learning. However, as outlined in Section 2.1.2, soft-DTW may not adequately address phase differences between the time series. To overcome this limitation, we introduce an enhanced method, soft-WDDTW, which optimizes both shape matching and time alignment, significantly improving the accuracy of ramp event predictions.

The foundation of our approach uses Derivative Dynamic Time Warping (DDTW) [44], which transforms the original data into high-level features that encapsulate sequence shape information. The transformation equation for a data point y_i is [44]:

$$G(y_i) = \frac{1}{2}\left[(y_i - y_{i-1}) + \frac{1}{2}(y_{i+1} - y_{i-1})\right], \quad 1 < i < n, \tag{15}$$

where n represents the series length. Notably, initial and final values are extrapolated as $G(y_1) = G(y_2)$ and $G(y_n) = G(y_{n-1})$, respectively, due to undefined boundaries. By applying this transformation (Eq. (15)), both predicted and actual series are converted to their respective slope features. These features are then utilized as inputs to the soft-DTW algorithm to enhance the learning of slope similarity, thereby enabling a more precise capture of trends and patterns in ramp events.

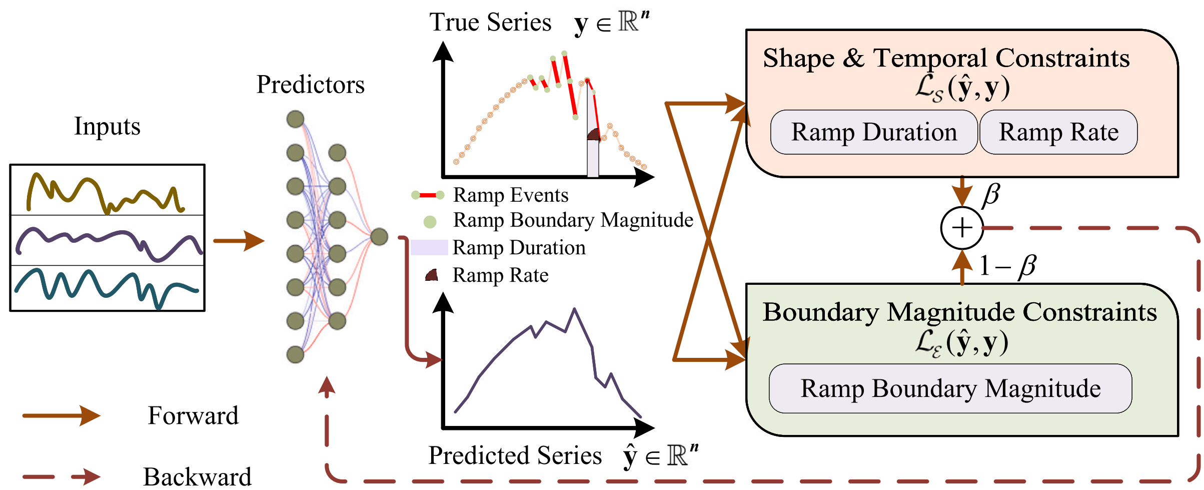


Fig. 2. The RSTLoss, designed for training deep learning models to predict ramp events, incorporates two distinct components: shape similarity and temporal distortion constraints $\mathcal{L}_{\mathcal{S}}(\hat{\mathbf{y}}, \mathbf{y})$, based on soft-Weighted Derivative Dynamic Time Warping (soft-WDDTW), and boundary magnitude constraints $\mathcal{L}_{\mathcal{E}}(\hat{\mathbf{y}}, \mathbf{y})$, which utilize extreme value distributions. The loss function is fully differentiable, enabling efficient computation through both forward and backward passes, which are implemented recursively for computational efficiency.

Addressing temporal distortion, a dynamic penalty mechanism based on phase differences between predicted and actual series is introduced. This mechanism employs a logistic weight function that assigns decreasing penalty weights as the phase difference diminishes, described by:

$$w_i = \frac{1}{1 + \exp(-g(i - n/2))}, \tag{16}$$

where $i \in \{0, 1, \ldots, n-1\}$ denotes the series length, and g is a parameter that controls the curvature of the function as well as the steepness of the weight function, thereby regulating sensitivity to phase differences. Fig. 3 demonstrates how various g values influence the weight distribution, illustrating the function's adaptability to temporal distortions. When $g = 0$, the weight function forms a straight horizontal line, indicating uniform weighting at all time points and effectively disregarding any temporal distortion. When $g = 0.03$, the weight function exhibits an approximately linear relationship. As g increases, the nonlinear characteristics of the weight allocation become increasingly pronounced. To ensure the stability of soft-WDDTW recursion during training, a parameter g is designed based on the effective interval width of the weight function, which typically refers to the range of i over which w_i transitions from low to high weight values. The variation of weights in this interval is sensitive to temporal distortion and has a significant moderating effect on losses. For any weight range $[w_1, w_2]$, the effective interval width Δi can be calculated by:

$$i = n/2 - \frac{1}{g}\ln\left(\frac{1-w}{w}\right), \tag{17}$$

$$\Delta i = i_1 - i_2 = \frac{1}{g}\left[\ln\left(\frac{1-w_1}{w_1}\right) - \ln\left(\frac{1-w_2}{w_2}\right)\right]. \tag{18}$$

In this study, we empirically set the weight range between 0.05 and 0.95. To effectively penalize temporal distortion, the effective interval width Δi is an integer smaller than the prediction horizon L. In this study, with a prediction horizon of 24 (6-h ahead with 15-min resolution), the calculated g based on Eq. (18) is greater than 0.256. Therefore, for our study, g is set to 0.3 to optimize the distinction between different levels of temporal distortion while maintaining gradient smoothness during training.

We then seamlessly integrate the aforementioned slope calculation method (Eq. (15)) and time-distortion penalty weights (Eq. (16)) into the recursive process of soft-WDDTW:

$$\mathcal{D}_{i,j} = \Delta(G(\mathbf{y}), G(\hat{\mathbf{y}}))_{i,j} \times \mathbf{W}[|i-j|] \tag{19}$$

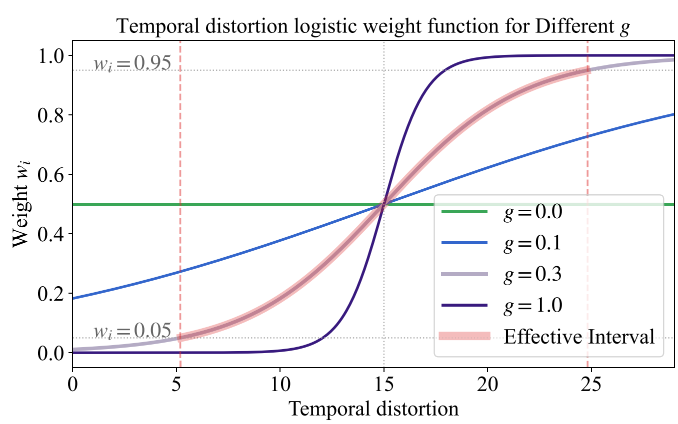


Fig. 3. Illustration of the logistic weight function and effective interval width for temporal distortion with varying parameter g.

$$\mathcal{R}_{i,j} = \mathcal{D}_{i,j} + \overset{\gamma}{\min}\{\mathcal{R}_{i-1,j-1}, \mathcal{R}_{i-1,j}, \mathcal{R}_{i,j-1}\}. \tag{20}$$

The algorithm for the forward recursion processes of soft-WDDTW is detailed in Algorithm 1.

The value of soft-WDDTW$^{\Delta}_{\gamma}(\hat{\mathbf{y}}, \mathbf{y})$ stored in $\mathcal{R}_{n,n}$ at the end of the forward recursion. The backward recursion process leverages the chain rule of differentiation to compute the gradient of $\mathcal{R}_{i,j}$ by back-propagating from $\mathcal{R}_{n,n}$:

$$\delta_{i,j} = \delta_{i+1,j}\frac{\partial \mathcal{R}_{i+1,j}}{\partial \mathcal{R}_{i,j}} + \delta_{i,j+1}\frac{\partial \mathcal{R}_{i,j+1}}{\partial \mathcal{R}_{i,j}} + \delta_{i+1,j+1}\frac{\partial \mathcal{R}_{i+1,j+1}}{\partial \mathcal{R}_{i,j}}, \tag{21}$$

where $\delta_{i,j} := \partial \mathcal{R}_{n,m} / \partial \mathcal{R}_{i,j}$ denotes the gradient. Based on Eqs. (7) and (20), each partial derivative in Eq. (21) is denoted as:

$$\begin{aligned}
\frac{\partial \mathcal{R}_{i+1,j}}{\partial \mathcal{R}_{i,j}} &= \frac{e^{-\mathcal{R}_{i,j}/\gamma}}{e^{-\mathcal{R}_{i,j-1}/\gamma} + e^{-\mathcal{R}_{i,j}/\gamma} + e^{-\mathcal{R}_{i+1,j-1}/\gamma}},\\
\frac{\partial \mathcal{R}_{i,j+1}}{\partial \mathcal{R}_{i,j}} &= \frac{e^{-\mathcal{R}_{i,j}/\gamma}}{e^{-\mathcal{R}_{i-1,j}/\gamma} + e^{-\mathcal{R}_{i,j}/\gamma} + e^{-\mathcal{R}_{i-1,j+1}/\gamma}},\\
\frac{\partial \mathcal{R}_{i+1,j+1}}{\partial \mathcal{R}_{i,j}} &= \frac{e^{-\mathcal{R}_{i,j}/\gamma}}{e^{-\mathcal{R}_{i,j}/\gamma} + e^{-\mathcal{R}_{i+1,j}/\gamma} + e^{-\mathcal{R}_{i,j+1}/\gamma}}.
\end{aligned} \tag{22}$$

The logarithmic form of Eq. (22) can be elegantly reformulated using the soft-min operations obtained from the forward recursion step:

$$\begin{aligned}\gamma \log \frac{\partial \mathcal{R}_{i+1,j}}{\partial \mathcal{R}_{i,j}} &= \min^{\gamma}\{\mathcal{R}_{i,j-1}, \mathcal{R}_{i,j}, \mathcal{R}_{i+1,j-1}\} - \mathcal{R}_{i,j} \\ &= \mathcal{R}_{i+1,j} - \mathcal{R}_{i,j} - \mathcal{D}_{i+1,j},\end{aligned}$$

$$\frac{\partial \mathcal{R}_{i+1,j}}{\partial \mathcal{R}_{i,j}} = e^{\frac{1}{\gamma}(\mathcal{R}_{i+1,j} - \mathcal{R}_{i,j} - \mathcal{D}_{i+1,j})}. \tag{23}$$

Similarly, the following relationships can also be obtained:

$$\frac{\partial \mathcal{R}_{i,j+1}}{\partial \mathcal{R}_{i,j}} = e^{\frac{1}{\gamma}(\mathcal{R}_{i,j+1} - \mathcal{R}_{i,j} - \mathcal{D}_{i,j+1})}, \tag{24}$$

$$\frac{\partial \mathcal{R}_{i+1,j+1}}{\partial \mathcal{R}_{i,j}} = e^{\frac{1}{\gamma}(\mathcal{R}_{i+1,j+1} - \mathcal{R}_{i,j} - \mathcal{D}_{i+1,j+1})}. \tag{25}$$

The Algorithm 2 illustrates the detailed process of backward recursion of soft-WDDTW.

While the introduction of the smoothing parameter γ enhances the algorithm's differentiability, it also presents challenges, as γ prevents the loss function from converging effectively to zero, which contradicts the optimization goal of minimizing loss [36,45]. Thus, similar challenges are encountered with soft-WDDTW, and we employ divergences to address these issues:

$$\mathcal{L}_{\mathcal{S}}(\hat{\mathbf{y}}, \mathbf{y}) = \text{soft-WDDTW}_{\gamma}^{\Delta}(\hat{\mathbf{y}}, \mathbf{y}) - \frac{1}{2}\left(\text{soft-WDDTW}_{\gamma}^{\Delta}(\hat{\mathbf{y}}, \hat{\mathbf{y}}) + \text{soft-WDDTW}_{\gamma}^{\Delta}(\mathbf{y}, \mathbf{y})\right). \tag{26}$$

Algorithm 1 Forward recursion to calculate soft-WDDTW.

Require:
Dataset: $\hat{\mathbf{y}} = (\hat{y}_0, \cdots, \hat{y}_{n-1}) \in \mathbb{R}^n$, $\mathbf{y} = (y_0, \cdots, y_{n-1}) \in \mathbb{R}^n$
Smoothing parameter: $\gamma \geq 0$
Distance function: $\Delta(\hat{\mathbf{y}}, \mathbf{y})_{ij} = \|\hat{y}_i - y_j\|_2^2$
Ensure:
Initialize $\mathcal{R}_{0,0} = 0$, $\mathcal{R}_{i,0} = \infty$, $\mathcal{R}_{0,j} = \infty$, $i, j \in [\![n]\!]$;
Calculate $G(\hat{\mathbf{y}})$ and $G(\mathbf{y})$ based on Eq. (15)
for $j = 1$ to n **do**
  for $i = 1$ to n **do**
    Calculate weight value $\mathbf{W}$ based on Eq. (16)
    $\mathcal{R}_{i,j} = \Delta(G(\hat{\mathbf{y}}), G(\mathbf{y}))_{i,j} \times \mathbf{W}[|i-j|] + \min_{\gamma}\{\mathcal{R}_{i-1,j-1}, \mathcal{R}_{i-1,j}, \mathcal{R}_{i,j-1}\}$
  end for
end for
return $\mathcal{R}_{n,n}$

3.2. Extreme value distribution based kernel function

To improve the prediction accuracy of the ramp boundary magnitude, the Fréchet kernel function is introduced, which is based on heavy-tailed kernel density estimation and serves as the loss function kernel for modeling the boundary magnitude. The PDF of the Fréchet distribution is defined as [37]:

$$f(u; \alpha, s, m) = \frac{\alpha}{s}\left(\frac{u-m}{s}\right)^{-1-\alpha} \exp\left[-\left(\frac{u-m}{s}\right)^{-\alpha}\right], \tag{27}$$

where u represents the values of the random variable, $\alpha > 0$ is the shape parameter, $s > 0$ is the scale parameter, and m is the positional parameter. Zhang et al. [46] modified this by employing a normalization factor to eliminate the positional parameter m, thereby constructing a Fréchet-based kernel function:

$$\mathcal{K}_{\text{Fréchet}}(u) = \frac{\alpha}{s}\left(\frac{u + s\left(\frac{\alpha}{1+\alpha}\right)^{1/\alpha}}{s}\right)^{-1-\alpha} \exp - \left(\frac{u + s\left(\frac{\alpha}{1+\alpha}\right)^{1/\alpha}}{s}\right)^{-\alpha}. \tag{28}$$

Algorithm 2 Backward recursion to calculate soft-WDDTW.

Require:
Dataset: $\hat{\mathbf{y}} = (\hat{y}_0, \cdots, \hat{y}_{n-1}) \in \mathbb{R}^n$, $\mathbf{y} = (y_0, \cdots, y_{n-1}) \in \mathbb{R}^n$
Smoothing parameter: $\gamma \geq 0$
Distance function: $\Delta(\hat{\mathbf{y}}, \mathbf{y})_{ij} = \|\hat{y}_i - y_j\|_2^2$
Ensure:
Initialize:
$\mathcal{R}_{i,n+1} = \infty$, $\mathcal{R}_{n+1,j} = \infty$, $i, j \in [\![n]\!]$
$\Delta(G(\hat{\mathbf{y}}), G(\mathbf{y}))_{i,n+1} = \Delta(G(\hat{\mathbf{y}}), G(\mathbf{y}))_{n+1,j} = 0$, $i, j \in [\![n]\!]$
$e_{i,n+1} = e_{n+1,j} = 0, i, j \in [\![n]\!]$
$\Delta(G(\hat{\mathbf{y}}), G(\mathbf{y}))_{n+1,n+1} = 0, e_{n+1,n+1} = 1, \mathcal{R}_{n+1,n+1} = \mathcal{R}_{n,n}$
for $j = n$ to 1 **do**
  for $i = n$ to 1 **do**
    Calculate weight value $\mathbf{W}$ based on Eq. (16)
    Calculate $\frac{\partial \mathcal{R}_{i+1,j}}{\partial \mathcal{R}_{i,j}}$, $\frac{\partial \mathcal{R}_{i,j+1}}{\partial \mathcal{R}_{i,j}}$ and $\frac{\partial \mathcal{R}_{i+1,j+1}}{\partial \mathcal{R}_{i,j}}$ based on Eq. (23), Eq. (24) and Eq. (25), respectively
    Calculate $\delta_{i,j}$ based on Eq. (21)
  end for
end for
return $e_{n,n}$

Then, a negative log-likelihood function based on the Fréchet kernel function is adopted as the loss function to predict the boundary magnitude of ramp events:

$$\mathcal{L}_{\mathcal{E}}(\hat{\mathbf{y}}, \mathbf{y}) = (-1-\alpha)\left(-\frac{\delta + s(\frac{\alpha}{1+\alpha})^{1/\alpha}}{s}\right)^{-\alpha} + \log \frac{\delta + s(\frac{\alpha}{1+\alpha})^{1/\alpha}}{s}, \tag{29}$$

where $\delta = |\hat{\mathbf{y}} - \mathbf{y}|$.

Using solar ramp events as a case study, the Swinging Door Algorithm (SDA) is employed to identify the starting and ending points of these events at four BSRN sites: CAB, CNR, PAL, and PAY. Subsequently, the Generalized Extreme Value (GEV) theory is utilized to fit the Probability Density Functions (PDFs) for the ramp events at each site, as depicted in Fig. 4. The PDFs demonstrate pronounced characteristics of long-tailed and heavy-tailed distributions in the boundary magnitudes of the ramp events. Notably, all the PDFs exhibit a shape parameter $\alpha > 0$, indicating conformity to the Fréchet distribution. Consequently, employing the Fréchet distribution to model the prediction error of ramp boundary magnitudes has proven to be effective.

Overall, the optimal convex combination of shape loss and extreme value loss, as defined in Eqs. (26) and (29), is encapsulated in the RSTLoss:

$$\mathcal{L}_{total}(\hat{\mathbf{y}}, \mathbf{y}) = \beta\mathcal{L}_{\mathcal{S}}(\hat{\mathbf{y}}, \mathbf{y}) + (1-\beta)\mathcal{L}_{\mathcal{E}}(\hat{\mathbf{y}}, \mathbf{y}), \tag{30}$$

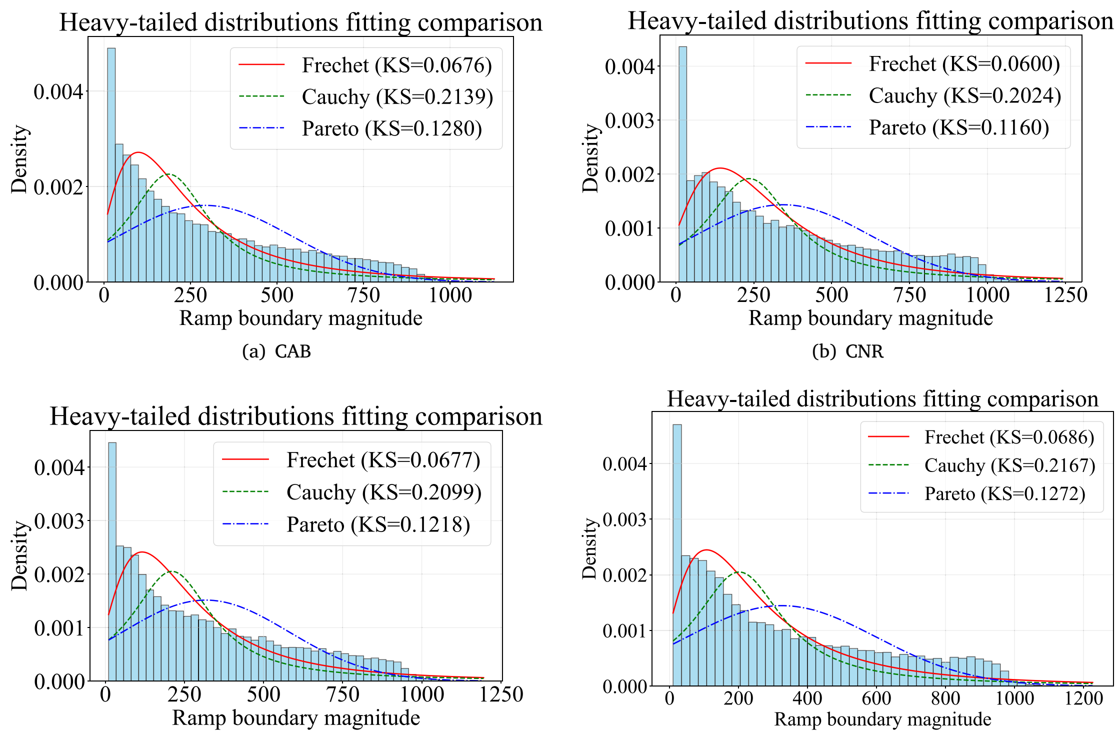
where β is the weighting parameter that balances the contributions of the shape and extreme value losses.

4. Case study

4.1. Data preparation

This study utilizes experimental data sourced from four observation sites within BSRN [49], as shown in Fig. 5. The dataset comprises time series measurements of Global Horizontal Irradiance (GHI), Beam Normal Irradiance (BNI), Diffuse Horizontal Irradiance (DHI), and key meteorological parameters such as temperature, relative humidity, and atmospheric pressure. These data are resampled to have a 15-min resolution over a period of six years (2017–2022). The primary focus of this research is to predict ramp events in GHI. To mitigate the influence of inherent seasonal variations on the analysis, the Clear Sky Index (CSI) is adopted as the main variable for prediction:

$$\text{CSI} = \frac{\text{GHI}}{\text{GHI}_{\text{cs}}}, \tag{31}$$



(a) CAB (b) CNR (c) PAL (d) PAY

Fig. 4. Fitting of GEV distribution PDFs at BSRN Sites: CAB, CNR, PAL, and PAY.

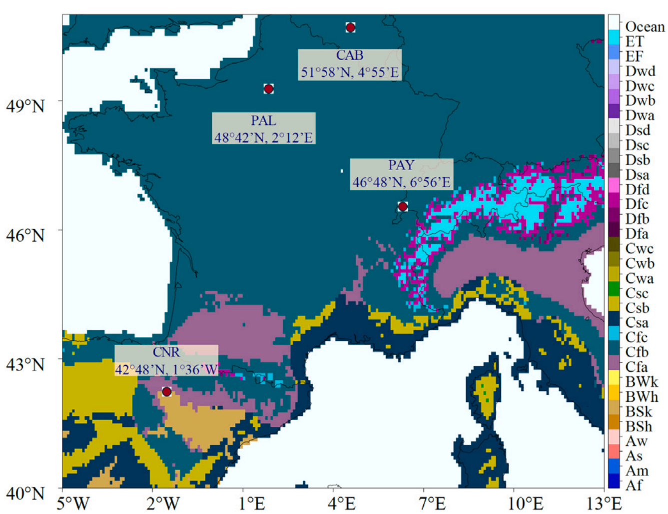


Fig. 5. Geographical distribution of the four selected BSRN stations (circular symbols) in the updated Köppen–Geiger climate classification system [47,48].

where GHI_{cs} represents the GHI under clear sky conditions at time t, calculated using either a physical or empirical clear sky model. For this study, the McClear model is employed to estimate GHI_{cs} [50], with CSI values set to 0 during nighttime periods when $GHI_{cs} = 0$. The predicted CSI is then converted back to GHI using Eq. (31) in order to compute the assessment metrics. To ensure the reliability and accuracy of the experimental results, stringent quality control (QC) measures are applied to the collected data. This involves using two specific filters:

Table 1
Model hyperparameters.

Parameter	Value
Number of encoder/decoder blocks	3
Initial learning rate	0.001
Hidden size	512
Attention heads	4
Input window size	16 (4-h)
Prediction horizons	2-h, 4-h, 6-h
Weighting parameter (β)	0.7
Shape parameter (α)	1.2
Scale parameter (s)	10
Smoothing parameter (γ)	0.1

Extremely Rare Limits and Closure Equation [48,51]:

$$\begin{bmatrix} -2 < GHI < 1.2E_{0n}\cos^{1.2}(Z) + 50 \\ -2 < DHI < 0.75E_{0n}\cos^{1.2}(Z) + 30 \\ -2 < BNI < 0.95E_{0n}\cos^{0.2}(Z) + 10 \end{bmatrix}, \tag{32}$$

$$\begin{bmatrix} |\text{closr}| < 8\% & \text{for } Z < 75^\circ \text{ and } GHI > 50 \\ |\text{closr}| < 15\% & \text{for } 75^\circ < Z < 93^\circ \text{ and } GHI > 50 \end{bmatrix}. \tag{33}$$

In the above QC procedure, Z is the solar zenith angle, E_{0n} is the extraterrestrial irradiance on a surface normal to the solar ray, and $|\text{closr}| = |\text{GHI} - (\text{DNI}\cos Z + \text{DHI})|$ is the difference between measured and computed GHI. In cases where QC procedures result in missing values, these gaps are filled using interpolation from the National Solar Radiation Database (NSRDB) [52]. In addition, Min-Max normalization is applied to all meteorological variables (except GHI) to standardize the data for further analysis. This approach ensures both the integrity of the data and the robustness of the predictive models developed in this study.

4.2. Deep learning based prediction models

To validate the effectiveness of the proposed RSTLoss for predicting solar ramp events and to assess its adaptability to different deep learning architectures, we selected five representative deep learning models. These models span the four principal architectural categories, showcasing the current diversity and advancements in deep learning for time series prediction. The selection includes one Multilayer Perceptrons (MLP)-based model: Time-Series Mixer (TSMixer) [53]; two attention mechanism-based models: Informer [54] and vanilla Transformer [55]; one Recurrent Neural Network (RNN)-based model: Gated Recurrent Unit (GRU) [56]; and one temporal Convolutional Neural Network (CNN)-based model: Deep Temporal Convolutional neural Network (DeepTCN) [57]. Further details on these prediction models can be found in Appendix A.2.

4.3. Assessment metrics of ramp prediction

To enhance the evaluation of the prediction accuracy of solar ramp events, this study employs a comprehensive suite of metrics, including two Euclidean distance (ED)-based metrics and three specialized metrics designed explicitly for assessing the accuracy of solar ramp events predictions and their temporal distortion. The source code for these metrics is publicly available for further exploration and validation.[1] The two ED-based metrics, Mean Bias Error (MBE) and RMSE, are defined as:

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i), \tag{34}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{35}$$

where y_i and $\hat{y}_i$ represent the actual and predicted values, respectively.

While ED-based metrics may not comprehensively capture the nuances of actual irradiance fluctuations, a novel metric, the Ramp Score (RS), has been introduced [33]. The RS employs SDA to segment both observed $\mathbf{y}$ and predicted $\hat{\mathbf{y}}$ time series into piecewise linear functions, enabling the detection of trends such as upward and downward ramps. SDA approximates the signal with linear segments over different time steps, constrained by a sensitivity parameter τ. A ramp is identified if a new piecewise linear function can approximate the signal with an absolute error smaller than a predefined threshold ϵ. For GHI time series, the error threshold ϵ_i is defined as:

$$\epsilon_i = \tau \cdot \max(\mathrm{GHI}_{w_i}), \tag{36}$$

where $\max(\mathrm{GHI}_{w_i})$ represents the maximum GHI value in the ith time window. Following the recommendations of Vallance et al. [33], we set $\tau = 0.18$.

To ensure the precise identification of ramp events, the compression time window parameter w of SDA is set to 30 min, allowing for the analysis of three consecutive time steps. This configuration optimizes the detection of ramp features and thereby enhances the computational accuracy of the RS. The slope of each piecewise function, which represents the magnitude and direction of ramp events, is then calculated as a function of time [14]. The RS is quantified by averaging the daily differences in the area between the predicted and actual ramp events:

$$\mathrm{RS} = \frac{1}{n_{\mathrm{day}}}\sum_{i=1}^{n_{\mathrm{day}}}\frac{1}{t^i_{\max} - t^i_{\min}}\int_{t^i_{\min}}^{t^i_{\max}}\left|\mathrm{SDA}^i(\hat{y}_t) - \mathrm{SDA}^i(y_t)\right| dt, \tag{37}$$

where $t^i_{\min}$ and $t^i_{\max}$ denote the time indices for sunrise and sunset of the ith day, respectively, and n_{day} is the number of days in the dataset. A lower RS suggests that the predicted ramp events are more closely aligned with the actual events.

Additionally, to quantify the shape accuracy of the predicted ramp events, we use the DTW Score:

$$\text{DTW Score} = \sqrt{\mathcal{R}_{i,j}}, \tag{38}$$

where $\mathcal{R}_{i,j}$ is calculated by Eq. (6). A lower DTW Score indicates a smaller shape dissimilarity between the predicted and the actual time series.

Evaluating the performance of predictive models should encompass not only the accuracy of predicted ramp magnitudes but also their temporal precision. Accurate timing is crucial for grid operators for effective load balancing, dispatch planning, and contingency responses, thereby enhancing power system stability and reliability [14,43]. Frías-Paredes et al. [43] proposed the Temporal Distortion Index (TDI) based on the DTW algorithm to analyze the temporal distortion in renewable energy forecasts. The TDI is defined as the area between the optimally aligned path $\mathcal{A}^*$, as solved by DTW, and the identity path between two time series. The identity path is defined as the series of points $(1,1),(2,2),\dots,(n-1,n-1),(n,n)$. A generalized version of the TDI can be expressed as [58]:

$$P_l = \int_{i_l}^{i_{l+1}}\left|x - \left(\frac{(x-i_l)(j_{l+1}-j_l)}{(i_{l+1}-i_l)} + j_l\right)\right| dx, \tag{39}$$

$$\mathrm{TDI} = \frac{2\sum_{l=1}^{k-1}|P_l|}{(n-1)^2}, (i,j=1,\dots,n), (l=1,\dots,k), \tag{40}$$

where i and j denote the time indices of the actual and predicted series, respectively, and k denotes the number of segments in the optimally aligned path $\mathcal{A}^*$. The TDI is a dimensionless index ranging from 0 to 1, where 0 indicates no temporal distortion and 1 indicates maximum temporal distortion. Thus, a lower TDI value characterizes a lesser degree of time series distortion.

4.4. Model training procedure

In this study, deep learning models were trained on data spanning four years, from 2017 to 2020. The models were then validated using data from 2021, and their performance was subsequently evaluated using data from 2022. All experiments were conducted on a high-performance computing platform equipped with an NVIDIA RTX 4090 GPU with 24 GB of memory. The experimental code was implemented using the PyTorch framework. Unless otherwise specified, a batch size of 32 was used for all experiments, and all models were initialized with a fixed random seed of 42 to ensure reproducibility. To mitigate overfitting, an early-stopping mechanism with a patience of five epochs was employed. All deep learning models were trained using the Adam optimizer, with an initial learning rate of 0.001 and an exponential decay rate of 0.999 [59].

The primary aim of this study was to assess the efficacy and robustness of RSTLoss relative to soft-DTW and MSE for solar ramp event prediction. To ensure that the observed performance differences were attributable primarily to the loss functions rather than architecture-specific tuning, no separate hyperparameter optimization was performed for each forecasting model. The model configurations were kept unchanged across MSE, soft-DTW, and RSTLoss, and the corresponding settings are reported in Table 1. This design does not imply that the parameters of RSTLoss were selected without analysis. The effects of the weighting parameter β, the Fréchet shape parameter α, and the scale parameter s on RS and TDI were systematically examined using a control-variable sensitivity analysis, as presented in Section 5.5. Based on this analysis, $\beta = 0.7$, $\alpha = 1.2$, and $s = 10$ were selected. The smoothing parameter was fixed at $\gamma = 0.1$ for both soft-DTW and the soft-WDDTW component of RSTLoss, following the established soft-DTW formulation [36], thereby ensuring a controlled comparison between the two alignment-based losses.

[1] https://github.com/JOTYtao/Ramp-Score.git

5. Results and discussion

5.1. Overall performance under different sky conditions

Comprehensive experimental evaluations, as presented in Appendix A.3, demonstrate the superior efficacy of RSTLoss across assessment metrics (RS, DTW Score, and TDI) for ramp event prediction. Additionally, to assess performance under diverse sky conditions, we employed the Bright-Sun clear-sky detection algorithm [60] for categorizing sky conditions into clear and cloudy. Daily classifications were based on the predominant sky condition. Performance metrics (RS and TDI) were computed on a daily basis and aggregated through statistical averaging over the testing set. This approach enables a nuanced understanding of the model's effectiveness under varied sky conditions, further emphasizing the practical applicability of RSTLoss.

Under cloudy conditions, RSTLoss demonstrates exceptional performance, as detailed in Appendix A.3 and illustrated in Fig. 6. Overall, RSTLoss holds a distinct advantage over benchmark loss functions, particularly in longer prediction horizons. Among all five prediction models, the prediction results using RSTLoss consistently exhibit the lowest RS and TDI values. Compared to MSE, RSTLoss achieves a significant improvement in the RS metric, with a reduction of up to 7.48% (4-h-ahead forecast, Transformer model) and a maximum reduction of 13.17% in the TDI metric (4-h-ahead forecast, Transformer model). When compared to soft-DTW, RSTLoss demonstrates even more substantial improvements in RS metrics, particularly for 6-h-ahead forecasts, with a reduction of 11.47% (Transformer model). For TDI metrics, the greatest reduction is 15.08% (6-h-ahead forecasts, TSMixer model). These results underscore the superior capability of RSTLoss in capturing short-term fluctuations and reducing temporal distortions, especially in longer time-scale forecasting tasks. Notably, these improvements in RS and TDI metrics occur without any deterioration in RMSE performance. The results demonstrate the unique and comprehensive benefits of RSTLoss in addressing the challenges of forecasting solar ramp events under cloudy conditions. Furthermore, RSTLoss consistently exhibits superior performance across all five benchmark models, using data from four BSRN sites. This consistency substantiates its robust generalization capabilities and architectural compatibility.

To further evaluate the robustness of the proposed loss function with respect to stochastic optimization, the training procedure was independently repeated using five random seeds, $\{42, 123, 1234, 2025, 2026\}$, while keeping the chronological training/testing partition, model architecture, optimization settings, and all other hyperparameters unchanged. Fig. 7 summarizes the statistical robustness analysis, where each circle denotes one independent training run, the diamond represents the mean RMSE, and the error bars correspond to the 95% confidence intervals. RSTLoss consistently achieves the lowest RMSE across all five independent runs while exhibiting the smallest run-to-run variation among the compared loss functions. The narrow confidence interval further indicates stable optimization behavior and low sensitivity to random initialization. In contrast, both MSE and soft-DTW consistently produce larger prediction errors and greater variability across repeated runs. These results demonstrate that the improvements obtained by RSTLoss are robust to stochastic model initialization rather than resulting from a favorable random seed.

5.2. Statistical significance analysis

Statistical robustness is evaluated using paired daily RS and TDI values. For the representative Transformer configuration under the 4-h-ahead cloudy-weather setting at the CAB station, the metrics obtained using RSTLoss are paired with those obtained using MSE and soft-DTW over the same test days. To quantify the uncertainty in the observed improvements without imposing a parametric distributional assumption, 95% confidence intervals for the mean paired differences are estimated using 10,000 paired bootstrap resamples with a fixed random seed 42. In addition, a two-sided Wilcoxon signed-rank test is performed for each paired comparison to test the null hypothesis that the median paired difference between RSTLoss and the corresponding baseline is zero. Since four related comparisons are conducted, covering RS and TDI against MSE and soft-DTW, respectively, the resulting p-values are jointly adjusted using the Benjamini–Hochberg procedure to control the false discovery rate at 0.05. The paired difference is calculated by subtracting the corresponding baseline metric from the RSTLoss metric for each test day; therefore, a negative difference indicates better performance for RSTLoss.

As reported in Table 2, RSTLoss reduces the mean RS from 72.988 with MSE to 67.529, corresponding to a relative reduction of 7.48%. The mean paired difference is −5.460, with a 95% bootstrap confidence interval of $[-5.707, -5.216]$. Compared with soft-DTW, RSTLoss reduces RS by 3.42%, with a mean paired difference of −2.394 and a 95% confidence interval of $[-2.780, -2.006]$. For TDI, RSTLoss reduces the daily mean from 19.507% with MSE to 16.938%, corresponding to a relative reduction of 13.17%. The mean paired difference is −2.569 percentage points, with a 95% bootstrap confidence interval of $[-2.693, -2.450]$. Compared with soft-DTW, RSTLoss reduces TDI by 7.13%, with a mean paired difference of −1.301 percentage points and a 95% confidence interval of $[-1.465, -1.136]$. The complementary Wilcoxon signed-rank tests yield Benjamini–Hochberg-adjusted p-values of 2.94×10^{-11} and 4.02×10^{-11} for the RS comparisons with MSE and soft-DTW, respectively, and 2.94×10^{-11} for both corresponding TDI comparisons. All adjusted p-values are substantially below the significance level of 0.05. Together with the bootstrap confidence intervals, which remain strictly below zero for all four comparisons, these results provide consistent non-parametric evidence that the reductions in RS and TDI obtained with RSTLoss are statistically significant across the paired test days.

5.3. Visualization analysis of ramp event predictions

To further evaluate the efficacy of RSTLoss in enhancing the accuracy of ramp event predictions, we randomly selected two days of 6-h-ahead prediction results from the TSMixer model for detailed visual analysis. As illustrated in Fig. 8, the results clearly highlight the advantages of RSTLoss in enhancing the similarity of prediction shapes and reducing temporal distortion in forecasting ramp events. Predictions generated by the MSE loss function show excessive smoothing, which hinders the model's ability to accurately capture and predict critical ramp events. In contrast, the predictions from the soft-DTW loss function display overly sensitive fluctuation characteristics, leading to frequent false alarms regarding ramp events. This comparative analysis underscores the superiority of RSTLoss in maintaining fidelity to actual ramp event dynamics while minimizing false predictions.

Fig. 9 provides a multi-dimensional graphical representation that illustrates the optimal alignment paths between the predicted results and the actual values (indicated by the black line). It also highlights the temporal distortion (visualized by the shaded area) and presents the distance matrix (depicted as a heat map). RSTLoss demonstrates significant advantages in reducing temporal distortion and enhancing shape similarity. Specifically, on 2022-07-01, RSTLoss reduces the RS by 15.1% and the TDI by 64.5% when compared to MSE loss. On 2022-08-27, RSTLoss achieves a 38.21% reduction in RS and a 61.7% decrease in TDI while maintaining RMSE values almost equal to those of the MSE predictions. These results strongly affirm the substantial potential of RSTLoss to enhance the accuracy of ramp event predictions and minimize temporal distortion.

5.4. Comparison of ramp boundary magnitude prediction

To rigorously assess the effectiveness of the extreme value distribution-based kernel function in RSTLoss for enhancing the accuracy of ramp event boundary magnitude predictions, we used the

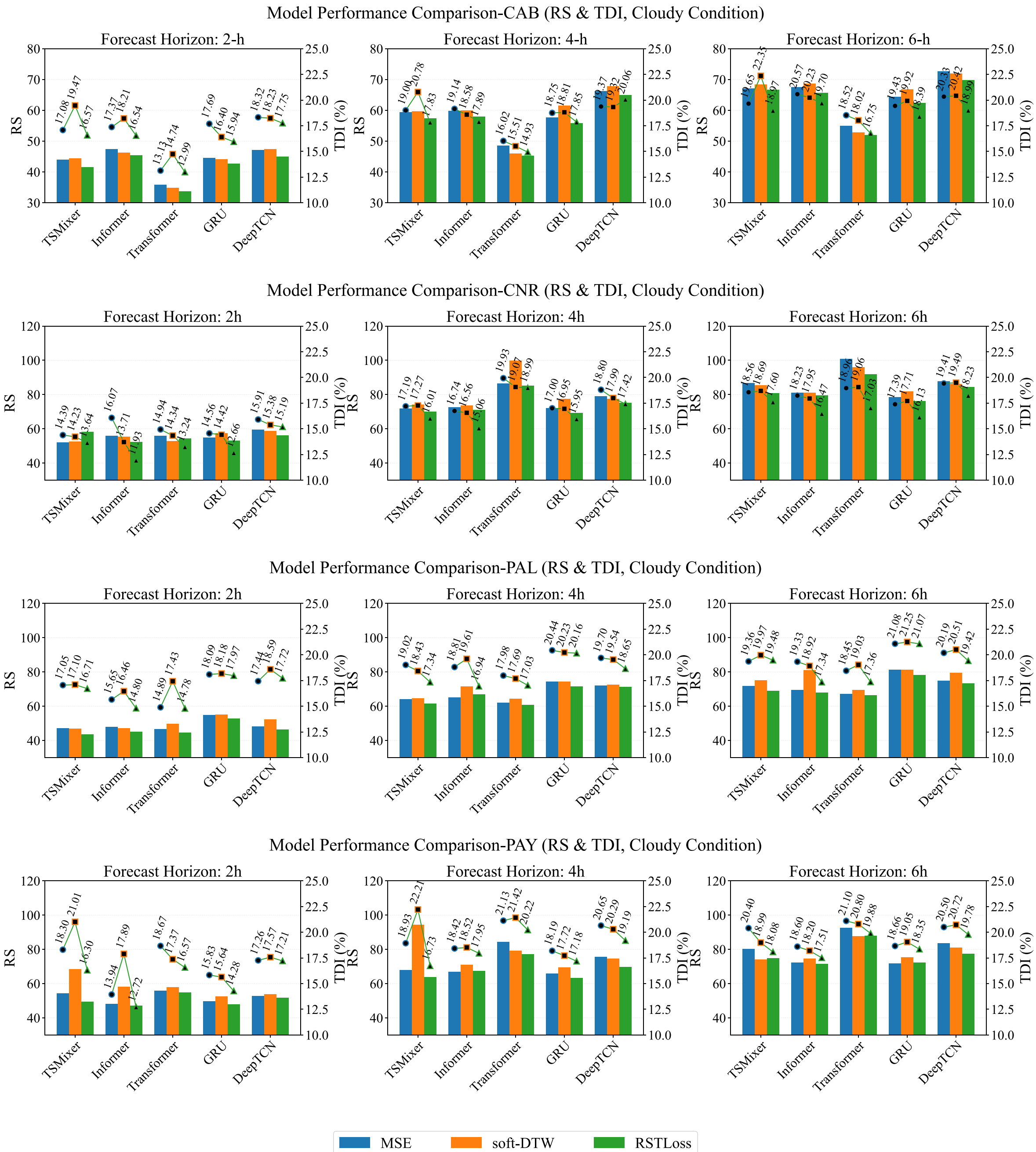


Fig. 6. The five benchmark models are compared in terms of RS (represented by colored bars) and TDI (indicated by annotated symbols) across the four BSRN sites using three loss functions: MSE, soft-DTW, and RSTLoss.

Table 2
Paired day-level comparison of RSTLoss with MSE and soft-DTW for the Transformer model under the 4-h-ahead cloudy-weather setting at the CAB station. Values before and after ± denote the mean and standard deviation, respectively. Δ denotes the paired difference calculated as RSTLoss minus the corresponding baseline, and CI denotes the 95% paired-bootstrap confidence interval. W denotes the two-sided Wilcoxon signed-rank statistic, and p_{BH} denotes the corresponding p-value after Benjamini–Hochberg correction.

Metric	Baseline	Forecast performance			Statistical significance		
		Baseline mean ± SD	RSTLoss mean ± SD	Reduction	Δ [95% CI]	W	p_{BH}
RS	MSE	72.988 ± 6.816	67.529 ± 6.295	7.48%	−5.460 [−5.707, −5.216]	0	2.94×10^{-11}
	soft-DTW	69.923 ± 6.745		3.42%	−2.394 [−2.780, −2.006]	18	4.02×10^{-11}
TDI (%)	MSE	19.507 ± 2.553	16.938 ± 2.273	13.17%	−2.569 [−2.693, −2.450]	0	2.94×10^{-11}
	soft-DTW	18.239 ± 2.514		7.13%	−1.301 [−1.465, −1.136]	6	2.94×10^{-11}

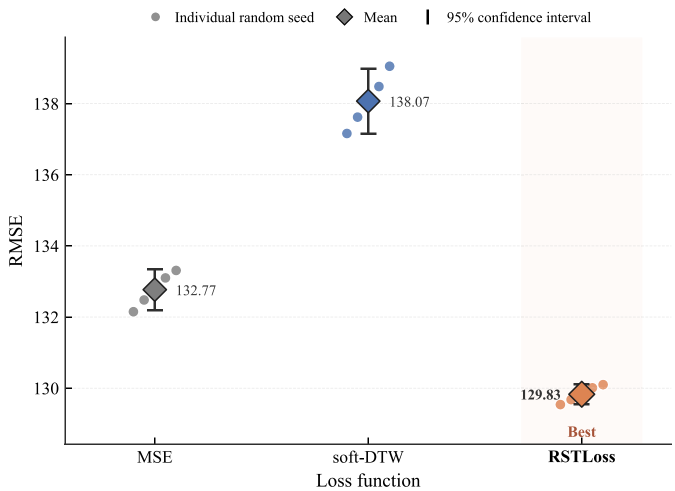


Fig. 7. Statistical robustness comparison of different loss functions over five independent random seeds. Each circle represents one independent training run, the diamond denotes the mean RMSE, and the error bars indicate the corresponding 95% confidence interval. RSTLoss consistently achieves the lowest RMSE with the smallest run-to-run variation.

SDA to precisely identify the start and end points of ramp events within the CAB dataset. We then calculated the RMSE of the boundary magnitude prediction results from the prediction models employing three different loss functions, as depicted in Fig. 10. By analyzing the performance of these models across three prediction horizons, the significant advantage of RSTLoss in improving the prediction accuracy of the boundary magnitude becomes evident. Specifically, RSTLoss reduces RMSE compared to soft-DTW in all models and prediction time ranges by up to 18.2% (Informer model, 6-h prediction). Although the improvement in MSE is more modest, RSTLoss consistently shows benefits, generally achieving RMSE reductions ranging from 1% to 6%. Performance enhancement is particularly notable in long-term forecasts (e.g., 6-h ahead), underscoring the robustness of RSTLoss in addressing the increased complexity of forecasts over time.

5.5. Hyperparameter sensitivity analysis

To investigate the impact of hyperparameters in the proposed RSTLoss on RS and TDI, we conducted a sensitivity analysis using the control variable approach for three key hyperparameters: the weighting parameter β in Eq. (30), the shape parameter α, and the scale parameter s in Eq. (27). We selected TSMixer as the benchmark predictor and performed a detailed study using the CAB site as an example. The experimental results, shown in Fig. 11, indicate that all three hyperparameters significantly affect the RS of the prediction results. The weighting parameter β delicately balances the predictive performance of the model. When β is small, the model overly focuses on the magnitude prediction of the starting and ending points of ramp events, leading to reduced sensitivity in terms of temporal duration fidelity and dynamic rate-of-change accuracy, consequently resulting in higher RS and TDI values. Conversely, when β is too large, the model becomes insufficiently sensitive to the prediction of the starting and ending points' amplitudes, which also results in higher RS metrics. The scale parameter s exhibits a clear threshold effect: when $s < 10$, RS shows only slight fluctuations, but when $s > 10$, RS increases significantly, from 66.92 to 72.35. This increase is mainly attributed to the effect of s on the shape of the Fréchet kernel, where larger s causes the Fréchet kernel to converge to a Gaussian kernel, which is detrimental to predicting boundary amplitude. Regarding the shape parameter α, a smaller α makes the loss function in Eq. (3) highly sensitive to small errors, resulting in excessive gradients during the optimization process and affecting the convergence stability of the model parameters, thus reducing prediction accuracy. Conversely, a larger α makes the loss function more tolerant of small errors while being more responsive to larger errors, producing smaller gradients and contributing to the stability of the optimization process. In contrast, the effects of s and α on TDI are relatively weak, with TDI maintaining a more stable level under different parameter configurations. Given this analysis, we have chosen $\beta = 0.7$, $\alpha = 1.2$, and $s = 10$ for reporting results in this work, as listed in Table 1.

5.6. Wind speed ramp events forecast

The proposed loss function, RSTLoss, is designed to capture three fundamental characteristics of ramp events: ramp boundary magnitude, ramp duration, and ramp rate. These features are not unique to solar power systems; they are also observed in other renewable energy systems, such as wind power, where sudden fluctuations in power output. Wind ramps present significant challenges to grid stability and forecasting. To further validate the applicability of RSTLoss in wind energy forecasting, we conducted experiments on wind speed prediction at four stations. The TSMixer model is employed to perform 6-h-ahead forecasting with 15-min temporal resolution. All other experimental parameters remained consistent with those detailed in Section-III-D. As shown in Fig. 12, RSTLoss demonstrated consistently excellent wind speed forecasting performance across the different forecast horizons for all stations. In the 2-h-ahead forecast, RSTLoss achieves the lowest RS while maintaining competitive TDI values, particularly at PAY, where it improves RS by 5.2% over MSE, and this advantage becomes even more pronounced as the forecast time is extended to 4-h-ahead and 6-h-ahead, where RSTLoss effectively balances the trade-off between RS and TDI metrics. These results convincingly demonstrate that RSTLoss not only improves the overall prediction accuracy, but also maintains robust temporal consistency in wind speed prediction over different time scales.

5.7. Training efficiency and complexity

Quantifying the computational overhead of different loss functions is important, particularly for real-time prediction scenarios. Therefore, a detailed comparison of the computational costs of MSE, soft-DTW, and RSTLoss is conducted using the TSMixer model as an example. Specifically, under identical GPU and CPU conditions, the average training time per epoch, GPU throughput, early-stopping epoch, and time complexity of each loss function are compared, following the settings described in the model training procedure. Table 3 summarizes the computational overhead of the different loss functions during TSMixer training at the CAB station, where L denotes the number of forecast steps.

RSTLoss is similar to soft-DTW in terms of training time per epoch and GPU throughput, while both require greater computational cost than MSE. Theoretically, the time complexity of RSTLoss is $O(L^2)$, which is higher than that of MSE, $O(L)$. The dominant quadratic cost arises from the $L \times L$ dynamic-programming alignment matrix in soft-WDDTW. Accordingly, the standard implementations of soft-DTW and RSTLoss also require $O(L^2)$ memory, whereas the additional temporal-distortion and Fréchet-based magnitude terms require only $O(L)$ operations and do not change the dominant complexity. The empirical peak GPU-memory measurements are consistent with this analysis. As shown in Table 3, RSTLoss introduces only approximately 2.9% additional peak-memory consumption relative to soft-DTW, indicating that the temporal-distortion weighting and Fréchet-based magnitude constraint add only modest memory overhead beyond the dynamic-programming alignment operation. However, the prediction horizon considered in short-term solar irradiance forecasting is generally 6 h, corresponding to $L = 24$ at a 15-min resolution; therefore, the quadratic computational and memory costs remain manageable. RSTLoss reaches

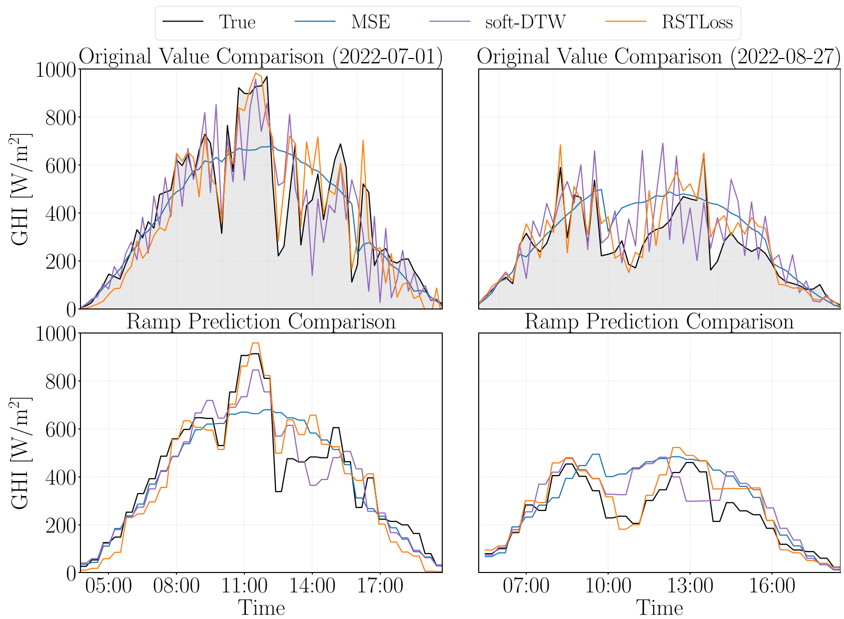


Fig. 8. Visualization of ramp event prediction: the first row compares the 6-h-ahead prediction results of the TSMixer model using MSE, soft-DTW, and RSTLoss as the loss functions at the CAB site, respectively; the second row compares the prediction accuracy of the ramp events identified by using the SDA.

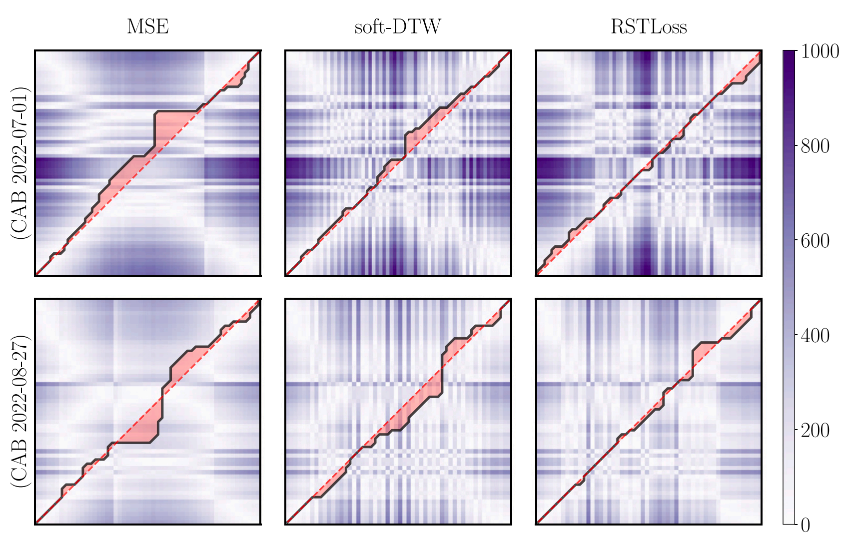


Fig. 9. Visual analysis of alignment paths and temporal distortion using MSE, soft-DTW, and RSTLoss for 6-h-ahead predictions with the TSMixer model at the CAB site. The heat maps illustrate the distance matrix between the true and predicted values. Red diagonal lines denote the ideal alignment paths, black lines show the actual alignment paths of the prediction results, and the red shaded areas highlight temporal distortions.

Table 3
Comparison of training efficiency, peak GPU-memory consumption, and computational complexity for different loss functions using TSMixer for 6-h-ahead forecasting at the CAB station.

Loss function	Time (s/epoch)	Throughput (samples/s)	Early-stopping epoch	Peak GPU memory (GB)	Time complexity
MSE	55.44	2506	32	2.43	$O(L)$
soft-DTW	66.62	2106	21	3.12	$O(L^2)$
RSTLoss	64.21	2185	20	3.21	$O(L^2)$

the early-stopping criterion in fewer training epochs, resulting in a comparable total training cost relative to MSE. Thus, although RSTLoss has a higher per-epoch computational complexity than MSE, it exhibits a favorable training-efficiency trade-off under the present experimental setting. Moreover, RSTLoss is used only during model training and does not alter the forecasting architecture. Models trained with MSE, soft-DTW, and RSTLoss therefore have identical inference complexity and latency, meaning that RSTLoss introduces no additional computational overhead during operational forecasting.

6. Conclusion

Conventional ED-based loss functions in deep learning techniques exhibit fundamental limitations in accurately capturing crucial ramp event features such as duration, rate, and boundary magnitude. To

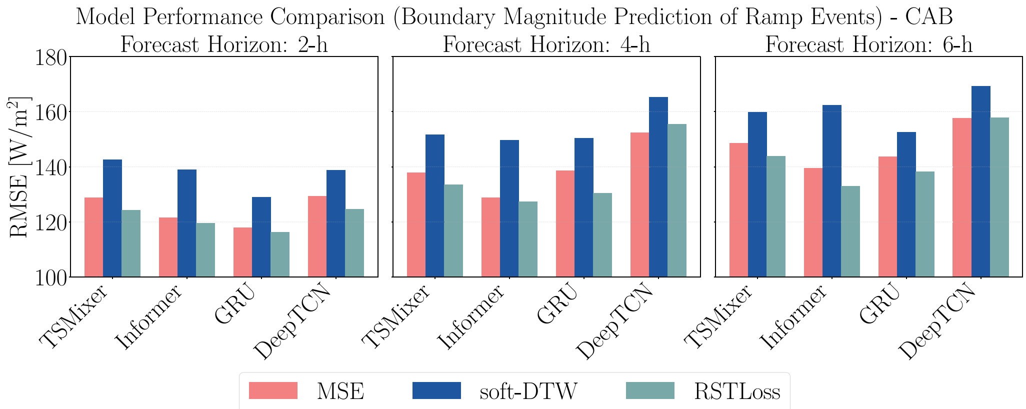


Fig. 10. Comparison of model performance in predicting the boundary magnitude of ramp events at the CAB site, where the boundary points of the ramp events are extracted by the SDA.

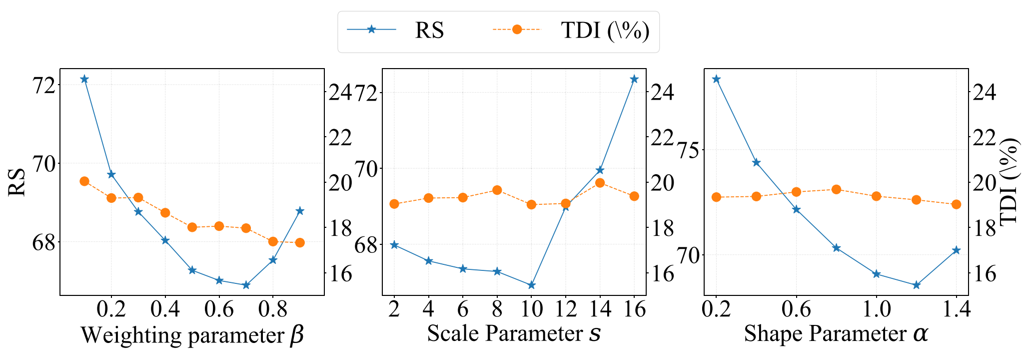


Fig. 11. Impact of hyperparameters in RSTLoss on the RS and TDI on cloudy weather condition.

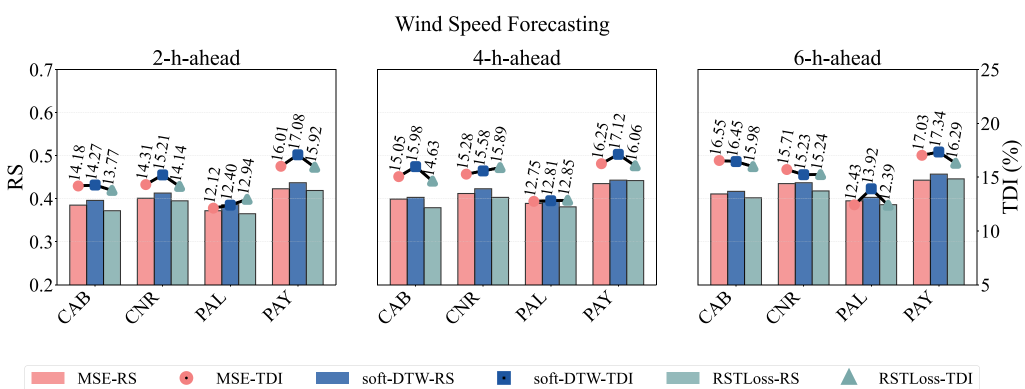


Fig. 12. Comparison of RS and TDI metrics for wind speed forecasting based on TSMixer at different prediction horizons (2-h-ahead, 4-h-ahead, and 6-h-ahead) across four BSRN stations using MSE, soft-DTW, and RSTLoss.

overcome these limitations, this study introduces RSTLoss, a novel loss function based on differentiable optimization. RSTLoss innovatively employs soft-WDDTW to effectively constrain ramp rate and duration and incorporates an extreme value distribution-based kernel function to accurately constrain boundary magnitude. Compared to the MSE loss function, RSTLoss achieves significant performance improvements in five representative deep learning architectures, evidenced by a maximum reduction of 7.48% in RS and 13.17% in TDI. Additionally, RSTLoss not only offers excellent compatibility with various deep learning architectures but also enhances model generalization capabilities. This work confirms that RSTLoss significantly improves the accuracy of solar ramp predictions while maintaining the integrity of the original model architecture. This enhancement effectively reduces the need for spare capacity allocation due to prediction inaccuracies, thereby substantially improving the economic efficiency and market competitiveness of solar power generation.

CRediT authorship contribution statement

Tao Jing: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Shanlin Chen:** Writing – review & editing, Validation. **Yinghao Chu:** Writing – review & editing, Validation, Supervision. **Mengying Li:** Writing – review & editing, Visualization, Resources, Project administration, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used GPT-5.6 to improve the language. After using this tool/service, the authors reviewed and edited the content as needed and assume full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is substantially supported by the Research Grants Council (Project No. 25213022), and the Innovation and Technology Commission (Project No. ITS/188/23) of the Hong Kong Special Administrative Region, China. This work is also partially supported by the City University of Hong Kong (Project No. 7006112 and 7020105).

Appendix. Supplemental materials

A.1. Derivation of the gradient of soft-DTW

Proposition. The gradient of soft-DTW during backpropagation is expressed as the expected alignment $\mathbb{E}_\gamma[\mathbf{A}_\pi]$, which can be derived as:

$$\nabla \mathrm{DTW}_\gamma^\Delta(\mathbf{y}, \hat{\mathbf{y}}) = \mathbb{E}_\gamma[\mathbf{A}_\pi], \tag{A.1}$$

$$\mathbb{E}_\gamma[\mathbf{A}_\pi] = \sum_{\mathbf{A}_\pi \in \mathcal{A}_{n,n}} \mathbf{A}_\pi \cdot p_\gamma(\mathbf{A}_\pi), \tag{A.2}$$

$$p_\gamma(\mathbf{A}_\pi) = \frac{\exp\left\{-\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}}{\sum_{\mathbf{A}'_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}'_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}}, \tag{A.3}$$

Proof. Let us remind that given a pairwise distance matrix $\Delta(\mathbf{y}, \hat{\mathbf{y}})$, the DTW dissimilarity is defined as Eq. (5), so the gradient can be defined as:

$$\begin{aligned}
\nabla \mathrm{DTW}_\gamma^\Delta(\mathbf{y}, \hat{\mathbf{y}}) &= -\gamma \cdot \nabla \log\left(\sum_{\mathbf{A}_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}\right) \\
&= -\gamma \frac{\nabla\left(\sum_{\mathbf{A}_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}\right)}{\sum_{\mathbf{A}'_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}'_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}} \\
&= -\gamma \frac{\sum_{\mathbf{A}_\pi \in \mathcal{A}_{n,n}} \nabla \exp\left\{-\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}}{\sum_{\mathbf{A}'_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}'_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}} \\
&= -\gamma \frac{\sum_{\mathbf{A}_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\} \cdot \left(-\frac{1}{\gamma} \cdot \sum_{\mathbf{A}_\pi \in \mathcal{A}_{n,n}} \nabla\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle\right)}{\sum_{\mathbf{A}'_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}'_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}} \\
&= \sum_{\mathbf{A}_\pi \in \mathcal{A}_{n,n}} \nabla\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle \cdot \frac{\exp\left\{-\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}}{\sum_{\mathbf{A}'_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}'_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}} \\
&= \sum_{i,j} a_{ij} \nabla \frac{\partial \Delta(\mathbf{y}, \hat{\mathbf{y}})}{\partial \mathbf{y}} \cdot \frac{\exp -\langle \mathbf{A}_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma}{\sum_{\mathbf{A}'_\pi \in \mathcal{A}_{n,n}} \exp\left\{-\langle \mathbf{A}'_\pi, \Delta(\mathbf{y}, \hat{\mathbf{y}})\rangle/\gamma\right\}} \\
&= \sum_{\mathbf{A}_\pi \in \mathcal{A}_{n,n}} \mathbf{A}_\pi \cdot p_\gamma(\mathbf{A}_\pi) \\
&= \mathbb{E}_\gamma[\mathbf{A}_\pi]
\end{aligned}$$

A.2. The prediction models

The frameworks of each prediction model are briefly described as follows.

- **TSMixer**
  Chen et al. [53] proposed a novel architecture designed by stacking MLPs to investigate the capabilities of linear models for time-series forecasting.[2] TSMixer is an innovative time series forecasting model that skilfully incorporates the power of MLP and the properties of time series analysis. With a stacked MLP structure, TSMixer is able to effectively capture complex linear and non-linear temporal patterns, and is able to efficiently utilize information between multiple correlated variables to improve the accuracy and stability of forecasts. A distinctive feature of TSMixer is its ability to incorporates future information to enhance forecast accuracy. Such a capability is particularly valuable in the field of solar forecasting, as it makes full use of the future meteorological data provided by NWP. The TSMixer model demonstrated excellent performance in the field of time series forecasting, consistently achieving state-of-the-art forecasting results on diverse datasets, including weather, electricity, solar and traffic.
- **Transformer**
  The vanilla Transformer [55] model was originally proposed for natural language processing tasks and quickly became a landmark innovation in the field of deep learning.[3] The core of the architecture is a sequence-to-sequence structure consisting of an encoder and decoder. Both the encoder and decoder modules of the Transformer are meticulously constructed from several key components: position-wise feed-forward networks, layer normalization, and a self-attention mechanism. A distinctive feature of Transformer is that it completely abandons the recurrent structure and instead relies on a multi-head self-attention mechanism to capture long-term dependencies between input features and output features. The core strength of the self-attention mechanism is its ability to create a global representation of each input feature that takes into account the context of the entire sequence. In recent years, a large number of Transformer-based time series forecasting models have emerged, highlighting the superiority of the Transformer architecture.
- **Informer**
  Zhou et al. [54] proposed a novel attention mechanism-based architecture called Informer, which designed a sparse self-attention mechanism to halve the input dimension of the cascade layer, effectively highlighting the dominant attentional pattern while significantly improving the model's ability to handle long input sequences.[4] Finally, the inference is completed by a single forward operation with a customized generative decoder. The superiority of Informer was verified on a large number of datasets.
- **Gated Recurrent Unit (GRU)**
  GRU is a variant of RNN, proposed by Cho [56]. GRU aims to solve the problem of gradient vanishing in traditional RNN while maintaining a relatively simple structure. As a simplified version of LSTM, GRU exhibits comparable performance to LSTM in several tasks but with fewer parameters and higher computational efficiency. The core innovation of GRU lies in its gating mechanism, which mainly consists of two gates: the reset gate and the update gate, which jointly control the flow of information and the updating of memories, enabling the network to effectively capture long-term dependencies. There has been a large amount of literature in the field of solar forecasting that employs GRU and their variants to improve prediction accuracy [30].
- **DeepTCN**

[2] https://github.com/ditschuk/pytorch-tsmixer
[3] https://github.com/jadore801120/attention-is-all-you-need-pytorch
[4] https://github.com/zhouhaoyi/Informer2020

Table A.4
Performance comparison of prediction models trained with RSTLoss, soft-DTW and MSE as loss functions for solar irradiance prediction at CAB site.

Models	Loss Function	Weather type	Metrics														
			RMSE [W/m²]			MBE [W/m²]			RS			DTW Score			TDI (%)		
			2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h
TSMixer	MSE	Over all	106.1	119.3	127.1	**0.725**	**6.378**	7.203	42.10	57.50	65.53	86.92	92.96	93.60	15.83	17.79	18.61
		Cloudy	111.0	124.6	132.1	2.803	**9.778**	12.92	43.90	59.31	67.10	86.66	92.62	93.03	17.08	19.00	19.65
		Clear Sky	**37.79**	49.88	67.42	**−18.21**	**−24.61**	−44.92	**26.96**	34.80	48.08	**28.44**	32.07	37.19	**3.738**	5.971	8.469
	soft-DTW	Over all	115.5	128.9	137.9	−1.423	10.16	15.29	43.14	57.42	67.18	88.54	91.83	93.61	18.23	19.56	21.36
		Cloudy	120.4	134.4	142.8	**0.768**	14.33	22.17	44.39	59.69	68.35	88.29	91.50	92.97	19.47	20.78	22.35
		Clear Sky	39.52	**45.82**	70.00	−21.40	−27.87	−47.40	29.40	**33.57**	50.89	29.04	**31.62**	38.35	4.067	**5.701**	9.659
	RSTLoss	Over all	**103.5**	**116.6**	**125.4**	4.328	12.28	**4.278**	**40.14**	**55.22**	**63.83**	**84.75**	**87.17**	**90.58**	**14.32**	**16.97**	**18.53**
		Cloudy	**107.4**	**121.5**	**130.7**	7.438	10.97	**8.601**	**41.50**	**57.35**	**66.59**	**85.36**	**88.94**	**91.14**	**16.57**	**17.83**	**18.97**
		Clear Sky	44.26	50.36	**58.55**	−24.01	−27.82	**−35.12**	37.18	39.86	**39.51**	31.37	35.96	**34.79**	5.097	6.262	**6.766**
Informer	MSE	Over all	106.1	118.5	129.0	18.27	−5.366	−16.40	45.42	58.23	67.31	88.08	93.82	99.00	16.05	18.03	19.87
		Cloudy	111.3	123.5	132.0	21.36	−2.527	−11.25	47.44	59.82	67.51	87.87	93.57	98.19	17.37	19.14	20.57
		Clear Sky	32.23	56.11	97.48	−9.829	−31.23	−63.31	26.30	38.57	65.04	27.46	**29.82**	42.86	3.243	7.224	13.04
	soft-DTW	Over all	117.4	131.31	144.1	4.984	**4.999**	−27.19	44.96	58.13	69.91	88.12	90.55	99.17	16.77	17.49	19.76
		Cloudy	122.2	136.5	145.7	6.061	8.846	−21.61	46.29	59.95	68.67	87.92	90.25	97.99	18.21	18.58	20.23
		Clear Sky	31.70	**53.55**	118.5	**−4.836**	**−30.06**	−78.05	**23.32**	**36.36**	75.97	**27.13**	30.52	45.21	**2.698**	**6.902**	15.22
	RSTLoss	Over all	**105.5**	**114.5**	**124.8**	**−1.322**	**−5.423**	**−12.48**	**43.63**	**57.16**	**65.11**	**86.85**	**88.44**	**94.32**	**15.17**	**17.20**	**18.89**
		Cloudy	**110.6**	**118.0**	**128.9**	**−0.570**	**0.397**	**−7.842**	**45.43**	**58.01**	**65.65**	**87.65**	**88.82**	**95.66**	**16.54**	**17.89**	**19.70**
		Clear Sky	**31.54**	66.74	**78.01**	−8.174	−58.47	**−54.79**	23.34	61.11	**55.19**	27.46	37.63	**41.31**	2.814	10.93	**11.04**
Transformer	MSE	Over all	95.25	108.1	116.5	−3.777	−6.229	−9.598	34.85	47.57	54.52	87.61	91.22	96.35	12.16	14.98	17.46
		Cloudy	99.78	112.5	119.8	−3.040	−4.310	−5.803	35.87	48.53	54.96	87.38	90.91	95.69	13.13	16.02	18.52
		Clear Sky	31.90	52.75	80.09	−10.49	−23.71	−44.18	23.82	28.80	44.93	27.81	31.18	**36.55**	2.773	4.928	7.106
	soft-DTW	Over all	**94.57**	107.2	116.4	**−2.401**	−4.022	−7.730	33.62	44.81	52.08	87.66	90.90	94.39	12.76	14.42	16.86
		Cloudy	**99.14**	111.8	120.1	**−1.990**	**−2.712**	**−4.722**	34.77	46.01	52.86	85.80	90.59	93.62	14.74	15.51	18.02
		Clear Sky	**29.41**	**47.51**	**74.21**	**−6.146**	**−15.95**	**−35.14**	**20.89**	28.23	41.41	**27.49**	**31.14**	37.43	**2.069**	**3.838**	**5.549**
	RSTLoss	Over all	95.01	**106.3**	**112.9**	2.700	**2.544**	**2.190**	**32.08**	**44.77**	**51.05**	**84.55**	**87.59**	**91.16**	**12.03**	**13.89**	**15.75**
		Cloudy	99.88	**111.2**	**116.8**	3.910	4.866	6.379	**33.63**	**45.28**	**51.62**	**84.10**	**87.00**	**90.15**	**12.99**	**14.93**	**16.75**
		Clear Sky	38.42	54.72	79.31	−8.332	−18.61	−35.98	22.59	**25.87**	**40.04**	32.18	35.25	39.25	2.653	4.693	6.049
GRU	MSE	Over all	102.4	117.1	124.6	3.403	7.635	9.120	42.12	55.42	62.76	89.81	93.15	**94.54**	16.33	17.49	18.23
		Cloudy	107.2	122.3	129.8	4.989	10.87	13.40	44.50	57.59	64.39	89.61	92.80	93.91	17.69	18.75	19.43
		Clear Sky	**33.75**	46.09	**58.36**	−11.05	−21.88	**−29.92**	27.83	31.96	**39.12**	27.56	32.39	37.93	3.129	5.289	**6.567**
	soft-DTW	Over all	106.4	123.5	128.3	2.745	**3.142**	**2.250**	41.68	58.82	65.10	85.00	94.91	95.47	15.16	17.69	18.90
		Cloudy	111.5	128.4	132.9	4.247	**6.758**	**7.553**	44.12	61.55	66.78	84.76	94.20	94.65	16.40	18.81	19.92
		Clear Sky	35.14	62.03	74.78	**−10.94**	−29.81	−46.09	**25.82**	39.05	49.81	**27.52**	39.22	40.55	**3.012**	6.782	8.970
	RSTLoss	Over all	**101.4**	**115.6**	**123.4**	**1.293**	10.33	3.645	**40.87**	**53.26**	**60.54**	**84.78**	**92.49**	94.97	**14.55**	**16.51**	**17.24**
		Cloudy	**106.3**	**120.1**	**128.3**	**2.714**	13.61	8.170	**42.68**	**55.82**	**62.39**	**83.57**	**92.17**	**92.37**	**15.94**	**17.85**	**18.39**
		Clear Sky	34.60	**43.31**	61.99	−11.64	**−19.53**	−37.59	27.31	**30.06**	41.89	27.60	**31.53**	**37.38**	3.020	**4.500**	7.370
DeepTCN	MSE	Over all	112.5	138.4	142.9	5.557	**21.39**	19.48	45.83	65.14	72.21	85.16	90.17	91.61	17.14	18.46	19.61
		Cloudy	117.5	144.1	147.3	8.429	28.83	28.72	47.15	66.17	72.65	84.83	89.56	90.77	18.32	19.37	20.33
		Clear Sky	44.94	67.13	93.35	−20.53	−46.34	−64.75	35.24	43.31	63.28	29.73	35.79	39.59	5.635	9.604	12.58
	soft-DTW	Over all	111.7	139.0	141.2	6.988	23.84	28.32	45.87	65.87	**71.25**	85.58	90.97	91.24	17.00	18.19	19.51
		Cloudy	116.8	145.3	146.4	9.216	29.84	37.35	47.41	67.81	71.97	85.28	90.32	90.29	18.23	**19.32**	20.42
		Clear Sky	44.32	53.98	**80.11**	**−13.31**	−30.80	**−54.05**	31.10	**36.54**	**51.53**	**29.24**	37.31	**38.55**	5.040	**7.178**	**10.64**
	RSTLoss	Over all	**110.5**	**134.6**	**136.1**	**4.469**	22.08	**−16.07**	**44.39**	**63.58**	71.31	**84.43**	**88.73**	**89.15**	**16.48**	**17.23**	**18.47**
		Cloudy	**113.2**	**140.5**	**137.9**	**8.171**	**27.82**	**−8.714**	**45.02**	**64.88**	**69.81**	**84.05**	**88.24**	**89.93**	**17.75**	20.06	**18.99**
		Clear Sky	**38.49**	**57.60**	118.1	−14.88	**−30.19**	−83.19	**28.34**	39.54	77.63	30.83	**32.81**	44.24	**4.193**	7.926	15.43

Chen et al. [57] proposed an advanced time series forecasting framework based on CNN, which provides both parametric and nonparametric methods for probability density estimation, called DeepTCN.[5] By utilizing the feature extraction capability of CNN, DeepTCN can automatically identify and utilize patterns and dependencies on different time scales, which in turn can effectively capture and learn potential correlations between series. With the parallel computing advantage of CNN, DeepTCN shows excellent computational efficiency when dealing with long sequences. There are a large number of research works proposing DeepTCN-based variants for solar forecasting [61,62].

A.3. Additional results

See Tables A.4–A.7.

Data availability

Data will be made available on request.

[5] https://github.com/oneday88/deepTCN

Table A.5
Performance comparison of benchmark models trained with RSTLoss, soft-DTW and MSE as loss functions for solar irradiance prediction at CNR site.

Models	Loss Function	Weather type	Metrics														
			RMSE [W/m^2]			MBE [W/m^2]			Ramp Score			DTW Score			TDI (%)		
			2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h
TSMixer	MSE	Over all	112.7	128.2	146.9	**0.475**	**13.22**	28.07	48.35	64.74	78.19	85.91	89.80	92.36	12.25	14.84	16.34
		Cloudy	124.5	141.1	161.1	**5.255**	**24.16**	44.88	52.11	71.82	86.62	85.54	89.29	90.88	14.39	17.19	18.56
		Clear Sky	39.72	51.06	**64.67**	−18.60	−30.45	**−39.04**	30.58	34.70	**46.05**	**31.04**	34.27	**46.26**	3.397	5.142	**7.146**
	soft-DTW	Over all	**112.3**	128.5	144.3	7.630	19.52	**14.72**	47.93	65.58	79.99	85.78	89.70	93.65	11.94	14.87	16.96
		Cloudy	124.3	141.7	155.9	12.92	31.41	**32.70**	52.59	74.09	85.68	85.42	89.16	91.82	14.23	17.27	18.69
		Clear Sky	**35.23**	**47.34**	80.79	**−13.54**	**−27.93**	−57.02	**27.39**	**34.08**	57.19	31.28	**34.11**	48.53	**2.479**	**4.925**	9.519
	RSTLoss	Over all	112.9	**126.9**	**142.2**	0.703	26.55	16.91	**46.95**	**62.81**	**76.31**	**83.06**	**87.3**	**90.35**	**10.68**	**13.58**	**15.77**
		Cloudy	**123.7**	**138.2**	**154.3**	11.43	52.36	33.76	**50.18**	**69.50**	**84.28**	**83.99**	**86.53**	**90.73**	**13.64**	**16.01**	**17.60**
		Clear Sky	63.26	107.1	76.57	−42.16	−36.49	−50.35	48.00	53.98	56.75	42.55	44.15	46.52	7.577	12.63	9.151
Informer	MSE	Over all	115.6	130.3	141.2	−9.735	**−1.517**	−15.00	51.36	65.59	76.42	89.82	91.43	99.09	13.63	14.66	16.60
		Cloudy	127.4	142.6	151.1	−7.303	6.414	−3.482	55.72	72.59	81.08	89.44	91.01	97.57	16.07	16.74	18.23
		Clear Sky	44.07	60.03	91.92	−19.44	−33.18	−61.00	32.43	38.31	59.23	32.41	33.26	47.58	3.547	6.045	9.860
	soft-DTW	Over all	110.3	132.6	141.4	**2.174**	13.59	−14.41	50.53	65.44	77.17	87.37	89.30	99.44	11.52	14.15	16.61
		Cloudy	122.0	146.6	151.3	5.234	22.63	**−1.518**	55.32	73.61	81.10	87.01	88.92	96.76	13.71	16.56	17.95
		Clear Sky	**36.17**	**45.59**	92.29	**−10.04**	**−22.48**	−65.88	**27.02**	**32.18**	63.70	**31.33**	**31.49**	52.68	**2.468**	**4.189**	11.07
	RSTLoss	Over all	**108.2**	**129.2**	**139.0**	−7.315	−2.753	**−0.347**	**48.91**	**63.13**	**74.33**	**85.87**	**87.25**	**94.27**	**10.94**	**12.71**	**15.90**
		Cloudy	**120.2**	**141.4**	**150.4**	**−3.347**	**4.366**	12.48	**52.19**	**70.95**	**79.29**	**85.40**	**87.63**	**93.00**	**11.93**	**15.06**	**16.47**
		Clear Sky	45.50	57.26	**78.10**	−23.16	−31.17	**−51.58**	32.12	37.66	**53.49**	33.84	33.87	**44.00**	3.692	4.982	**9.424**
Transformer	MSE	Over all	120.5	158.7	163.3	−27.73	−47.69	54.15	53.21	84.59	89.66	90.43	102.1	92.77	13.24	18.25	16.66
		Cloudy	131.5	167.4	180.1	−24.56	**−39.38**	78.90	55.93	86.40	100.9	89.48	99.87	91.68	14.94	19.93	18.96
		Clear Sky	59.29	118.1	60.72	−40.36	−80.90	−44.69	43.49	77.98	45.51	40.36	54.83	**41.96**	6.210	11.33	7.115
	soft-DTW	Over all	117.7	161.3	157.7	−19.06	45.11	**43.03**	49.34	90.15	88.10	87.76	95.10	94.41	12.43	**17.06**	17.19
		Cloudy	129.5	176.5	172.4	−16.77	70.02	**67.83**	52.70	99.81	95.93	87.32	92.78	92.30	14.34	19.07	19.06
		Clear Sky	46.49	**74.08**	73.65	−28.21	**−54.31**	−55.99	34.44	**53.36**	55.76	32.91	**51.67**	52.21	4.493	**8.748**	9.467
	RSTLoss	Over all	**113.4**	**157.7**	**162.4**	**−1.603**	**40.14**	61.29	**48.04**	**83.43**	**87.80**	**86.80**	**94.55**	**92.29**	**11.95**	17.20	**15.67**
		Cloudy	**125.3**	**162.0**	**169.0**	**2.819**	65.53	86.39	**50.25**	**85.04**	**91.84**	**86.46**	**91.37**	**91.36**	**13.24**	**18.99**	**17.03**
		Clear Sky	**39.17**	77.64	**53.75**	**−19.26**	−61.20	**−38.93**	**29.66**	58.04	**41.85**	**30.81**	52.77	42.31	**3.456**	9.778	**6.902**
GRU	MSE	Over all	**111.6**	**128.1**	136.9	5.790	7.029	**8.039**	49.17	63.40	70.80	88.24	91.35	92.62	12.14	14.44	15.06
		Cloudy	**123.7**	**141.2**	150.0	8.961	14.06	17.08	54.71	71.90	78.51	87.95	90.95	91.93	14.56	17.00	17.39
		Clear Sky	**32.14**	**47.92**	**61.22**	**−6.871**	−21.03	**−28.07**	25.60	32.20	41.60	**29.70**	33.35	**36.78**	**2.130**	3.848	**5.424**
	soft-DTW	Over all	114.7	138.2	143.4	8.468	**1.955**	−11.45	51.32	68.94	74.21	88.26	90.43	94.69	12.05	14.58	15.70
		Cloudy	127.3	152.0	154.8	12.51	**7.695**	**−2.430**	57.51	77.40	81.66	87.56	89.47	93.53	14.42	16.95	17.71
		Clear Sky	32.76	56.51	83.10	−7.654	**−20.97**	−47.46	25.31	34.60	50.11	29.89	33.69	44.33	2.232	4.757	7.379
	RSTLoss	Over all	112.4	128.4	**135.8**	**4.785**	3.101	11.40	**47.06**	**60.91**	**68.91**	**86.65**	**90.11**	**91.10**	**11.22**	**12.38**	**13.86**
		Cloudy	124.7	141.4	**148.5**	**7.715**	9.226	22.49	**52.97**	**69.32**	**76.24**	**86.33**	**89.72**	**91.30**	**12.66**	**15.95**	**16.13**
		Clear Sky	32.27	50.04	63.05	−6.911	−21.35	−32.88	26.05	31.56	40.46	29.83	**33.06**	37.13	2.142	**3.760**	5.452
DeepTCN	MSE	Over all	124.6	139.1	153.8	5.507	7.978	22.32	56.24	74.00	81.66	85.19	94.47	94.45	13.89	16.87	17.68
		Cloudy	136.8	151.1	166.7	12.75	20.66	40.44	59.46	79.07	87.70	84.29	92.69	92.92	15.91	18.80	19.41
		Clear Sky	52.59	74.15	**84.94**	−23.40	−42.67	**−50.01**	37.70	52.67	**57.80**	38.47	49.11	**47.86**	5.509	8.877	**10.49**
	soft-DTW	Over all	128.7	135.7	**150.4**	15.49	14.08	**−8.569**	53.38	70.44	83.35	85.32	**91.02**	97.12	13.20	15.90	18.46
		Cloudy	142.0	148.7	**158.8**	22.95	26.57	**8.803**	58.78	77.30	86.19	84.37	**90.35**	95.36	15.38	17.99	19.49
		Clear Sky	47.26	**60.61**	110.5	−14.27	**−35.79**	−77.93	31.11	**41.73**	76.63	39.05	**37.74**	49.81	4.161	**7.256**	14.17
	RSTLoss	Over all	**119.3**	**134.1**	155.0	**5.248**	**−0.987**	13.66	**51.76**	**69.07**	**77.58**	**83.14**	91.69	**92.28**	**12.99**	**15.52**	**16.89**
		Cloudy	**131.7**	**145.3**	166.0	**8.483**	**9.015**	33.81	**56.20**	**75.06**	**84.38**	**82.66**	90.89	**91.27**	**15.19**	**17.42**	**18.23**
		Clear Sky	**42.39**	77.73	99.59	**−7.668**	−40.92	−66.79	**29.02**	47.10	67.81	**33.22**	39.57	50.75	**3.876**	7.671	12.34

Table A.6
Performance comparison of benchmark models trained with RSTLoss, soft-DTW and MSE as loss functions for solar irradiance prediction at PAL sites.

Models	Loss Function	Weather type	Metrics														
			RMSE [W/m^2]			MBE [W/m^2]			Ramp Score			DTW Score			TDI (%)		
			2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h
TSMixer	MSE	Over all	107.3	123.1	130.5	−1.603	**6.053**	19.85	44.37	60.19	65.04	89.83	93.84	**92.89**	15.12	17.02	**17.38**
		Cloudy	116.1	132.9	140.9	**0.741**	11.30	28.17	46.99	64.09	71.80	89.41	93.15	**92.14**	17.05	19.02	**19.36**
		Clear Sky	31.10	**39.28**	**41.77**	**−14.25**	**−22.29**	**−25.11**	**29.20**	**33.21**	34.55	**29.78**	34.28	33.70	3.273	**4.764**	**5.163**
	soft-DTW	Over all	107.3	122.4	131.4	−3.915	9.557	22.15	44.43	61.15	68.50	89.57	92.87	95.00	15.14	16.67	17.98
		Cloudy	116.1	131.8	141.6	−1.793	17.24	31.49	46.76	64.53	74.97	89.28	91.96	94.26	17.10	18.43	19.97
		Clear Sky	**30.07**	46.02	47.11	−15.37	−31.92	−28.25	29.73	40.51	38.23	30.05	37.02	35.42	**3.039**	5.851	5.680
	RSTLoss	Over all	**105.4**	**116.0**	**125.9**	**0.538**	6.751	**13.26**	**42.27**	**58.60**	**63.14**	**87.65**	**91.35**	94.81	**14.57**	**16.18**	17.55
		Cloudy	**115.1**	**126.8**	**139.9**	4.157	**9.925**	**20.41**	**43.44**	**61.42**	**68.92**	**87.35**	**90.88**	94.00	**16.71**	**17.34**	19.48
		Clear Sky	39.25	45.87	48.16	−18.99	−22.97	−25.34	36.95	36.04	38.00	32.98	35.03	37.50	4.398	5.842	5.694
Informer	MSE	Over all	**106.0**	124.8	125.7	−11.41	−22.63	−9.173	44.98	62.05	65.50	91.98	97.74	99.32	13.90	17.08	17.62
		Cloudy	**114.6**	133.4	134.4	−11.27	−20.73	−4.174	47.84	**65.16**	69.50	91.70	97.13	98.21	15.65	18.81	19.33
		Clear Sky	30.64	60.36	60.05	−12.17	−32.93	−36.16	27.87	44.91	43.92	30.63	35.43	41.78	3.159	6.452	7.060
	soft-DTW	Over all	107.2	130.3	126.4	−7.451	19.29	**8.473**	44.41	65.60	65.06	90.89	**92.68**	95.62	14.52	17.47	17.08
		Cloudy	116.0	140.8	136.1	−7.127	25.27	15.01	47.68	71.40	69.96	90.66	92.10	94.85	16.46	19.61	18.92
		Clear Sky	**29.13**	**39.72**	**47.22**	−9.202	**−12.99**	**−26.81**	**27.49**	**32.31**	**36.88**	**28.66**	**31.28**	**35.33**	**2.629**	**4.310**	**5.765**
	RSTLoss	Over all	109.4	**122.3**	**123.6**	**0.821**	**−2.044**	−9.624	**43.18**	**61.59**	**62.89**	**90.75**	92.95	**94.25**	**12.85**	**15.04**	**16.68**
		Cloudy	118.4	**131.9**	**131.6**	**2.435**	**1.514**	**−3.827**	**45.01**	66.97	**67.89**	**90.35**	**91.21**	**93.54**	**14.80**	**16.94**	**17.34**
		Clear Sky	29.24	42.21	69.59	**−7.892**	−21.25	−40.91	28.06	33.59	45.70	29.27	34.39	42.55	2.809	5.312	7.448
Transformer	MSE	Over all	105.5	120.2	124.5	−5.630	−8.858	−8.272	43.74	57.63	62.50	90.86	96.96	97.47	13.13	16.14	16.52
		Cloudy	114.3	129.7	134.0	−5.371	−6.194	−4.913	46.50	61.89	67.19	90.61	96.44	96.93	14.89	17.98	18.45
		Clear Sky	**26.06**	40.59	46.94	−7.030	−23.23	−26.40	25.44	34.87	36.26	27.73	33.13	31.31	**2.340**	4.827	**4.631**
	soft-DTW	Over all	107.5	122.4	125.3	−6.915	13.63	−5.686	46.62	58.63	63.55	92.71	94.28	98.37	15.38	15.75	17.06
		Cloudy	116.4	132.2	135.1	−8.269	17.79	−2.820	49.50	64.22	69.29	92.50	93.49	97.58	17.43	17.69	19.03
		Clear Sky	26.94	39.20	44.91	**0.389**	**−8.833**	**−21.15**	28.71	30.65	33.71	28.44	31.96	35.92	2.742	3.840	4.913
	RSTLoss	Over all	**104.7**	**118.3**	**122.5**	**−5.294**	**−2.964**	**−4.844**	**41.90**	**55.38**	**61.27**	**88.53**	**92.14**	**96.50**	**13.07**	**15.12**	**16.43**
		Cloudy	**111.4**	**128.0**	**132.2**	**−4.900**	**−0.674**	**−1.410**	**44.52**	**60.37**	**66.26**	**88.25**	**91.63**	**95.91**	**14.78**	**17.03**	**17.36**
		Clear Sky	27.94	**33.19**	**41.72**	−7.422	−15.32	−23.38	**25.32**	**30.61**	**32.63**	**27.36**	**29.64**	**30.90**	2.571	**3.384**	5.525
GRU	MSE	Over all	**115.9**	**138.8**	146.1	**21.55**	**22.94**	21.29	50.57	68.67	78.90	86.64	92.08	94.07	16.06	18.54	19.78
		Cloudy	125.2	148.8	153.5	**27.17**	**34.79**	36.77	54.72	74.22	81.19	86.06	90.97	92.47	18.09	20.44	21.08
		Clear Sky	36.28	62.05	**96.48**	−8.794	−29.63	**−62.28**	31.27	**41.89**	**67.64**	31.15	40.20	45.63	3.554	**6.800**	**11.75**
	soft-DTW	Over all	116.3	142.7	146.5	22.71	28.68	21.45	51.07	69.13	79.53	86.86	92.11	94.77	16.12	18.36	20.06
		Cloudy	125.7	148.6	153.5	28.61	39.50	37.09	55.16	74.35	81.29	86.24	90.93	92.91	18.18	20.23	21.25
		Clear Sky	36.35	62.09	100.4	−9.136	−29.73	−62.93	30.16	43.14	69.70	31.86	40.50	47.82	3.466	6.826	12.75
	RSTLoss	Over all	116.4	140.1	**145.7**	23.13	29.79	**18.28**	**48.25**	**66.65**	**74.45**	**84.14**	**90.68**	**92.12**	**15.94**	**18.29**	**19.01**
		Cloudy	**122.8**	**145.3**	**152.8**	28.34	39.98	**33.93**	**52.70**	**71.50**	**78.16**	**85.68**	**88.83**	**90.57**	**17.97**	**20.16**	**21.07**
		Clear Sky	**35.32**	**59.67**	98.44	**−5.030**	**−25.26**	−66.18	**30.26**	42.58	68.39	**30.69**	**38.28**	**45.37**	**3.442**	6.830	12.05
DeepTCN	MSE	Over all	109.3	141.1	141.2	−0.561	24.61	**−7.062**	44.59	66.30	75.36	87.39	90.18	95.62	**15.55**	17.82	19.14
		Cloudy	117.9	**151.9**	145.5	**1.777**	**33.31**	7.547	47.98	72.03	74.88	86.94	89.30	94.12	**17.44**	19.70	20.19
		Clear Sky	36.00	53.45	115.2	−13.18	**−22.37**	−85.92	29.79	36.05	82.97	30.24	36.58	47.52	3.922	6.232	12.71
	soft-DTW	Over all	116.2	**140.9**	147.0	**−0.376**	26.55	25.54	50.96	66.52	76.23	**87.08**	90.54	93.06	16.77	17.53	19.23
		Cloudy	125.1	152.1	155.6	4.276	36.19	40.21	52.13	72.47	79.36	86.19	89.67	91.78	18.59	19.54	20.51
		Clear Sky	44.70	**46.75**	**86.76**	−25.49	−25.47	**−53.65**	41.59	**33.29**	**59.93**	38.83	**36.09**	**44.36**	5.532	**5.151**	**11.30**
	RSTLoss	Over all	**108.3**	143.3	**137.5**	1.277	**23.57**	−10.47	**42.11**	**64.11**	**73.68**	87.68	**88.70**	**92.43**	15.68	**15.94**	**18.33**
		Cloudy	**117.1**	154.4	**142.7**	3.612	34.01	**2.746**	**46.37**	**71.21**	**73.37**	**85.21**	**88.51**	**90.89**	17.72	**18.65**	**19.42**
		Clear Sky	**31.79**	51.71	104.8	**−11.33**	−32.78	−81.80	**28.38**	39.67	75.39	28.95	41.24	46.69	**3.135**	7.362	12.60

Table A.7
Performance comparison of benchmark models trained with RSTLoss, soft-DTW and MSE as loss functions for solar irradiance prediction at PAY sites.

Models	Loss Function	Weather type	Metrics														
			RMSE [W/m^2]			MBE [W/m^2]			Ramp Score			DTW Score			TDI (%)		
			2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h	2-h	4-h	6-h
TSMixer	MSE	Over all	114.5	**120.6**	135.8	−17.86	**9.510**	4.685	52.08	59.81	78.20	88.25	90.40	98.74	15.63	15.84	18.42
		Cloudy	127.7	**135.7**	145.5	−12.51	**21.26**	29.03	54.29	67.78	80.14	86.55	89.12	93.83	18.30	18.93	20.40
		Clear Sky	56.26	49.95	99.89	−34.48	**−27.01**	−70.95	47.19	38.78	74.18	45.80	41.09	63.43	6.256	5.031	11.46
	soft-DTW	Over all	128.7	154.0	**130.5**	−22.21	28.24	**−5.391**	66.01	89.64	**69.86**	95.09	103.1	**93.34**	18.29	20.24	16.92
		Cloudy	140.4	165.2	142.2	−12.28	60.79	**11.32**	68.31	94.36	**74.06**	91.82	95.15	**90.73**	21.01	22.21	18.99
		Clear Sky	82.63	112.5	**84.36**	−53.06	−72.90	**−57.31**	60.68	75.04	**60.13**	57.14	75.43	**51.89**	8.769	13.34	**9.662**
	RSTLoss	Over all	**107.5**	122.8	135.1	**−4.677**	14.64	−9.545	**45.82**	**58.14**	72.51	**85.47**	**89.86**	95.03	**13.64**	**14.83**	**16.27**
		Cloudy	**121.0**	138.5	**140.8**	**0.963**	28.91	13.29	**49.47**	**65.19**	74.92	**84.33**	**88.49**	92.45	**16.30**	**16.73**	**18.08**
		Clear Sky	**43.71**	**48.55**	107.3	**−22.21**	−29.73	−80.50	**36.51**	**38.11**	76.83	**40.89**	**42.39**	53.05	**4.304**	**4.638**	12.28
Informer	MSE	Over all	106.4	123.0	**128.4**	−11.85	−12.76	−15.10	44.90	62.10	67.94	86.75	91.57	95.21	11.65	15.95	16.40
		Cloudy	119.5	134.7	**141.0**	−10.22	−2.268	**−4.353**	48.17	**66.90**	72.26	85.72	90.45	92.90	13.94	18.42	18.60
		Clear Sky	45.61	75.78	77.23	−16.90	−45.38	−48.50	33.88	49.10	53.06	40.36	**41.12**	50.82	3.614	7.307	8.687
	soft-DTW	Over all	113.6	128.8	131.8	**−0.985**	−6.727	10.18	52.10	66.97	68.82	88.22	91.51	91.88	14.87	16.33	16.05
		Cloudy	128.1	139.2	146.1	**5.361**	7.507	27.76	58.06	70.97	74.64	87.31	89.66	89.50	17.89	18.52	18.20
		Clear Sky	44.84	88.74	70.81	−20.71	−50.96	−44.44	34.93	56.95	49.49	38.66	47.35	50.03	4.271	8.677	8.521
	RSTLoss	Over all	**105.8**	**122.4**	128.5	2.182	**−3.303**	**3.598**	**42.48**	**60.97**	**64.47**	**84.94**	**90.29**	**89.69**	**10.25**	**15.27**	**15.23**
		Cloudy	**117.3**	**132.9**	143.6	7.975	**5.523**	16.62	**47.06**	67.39	**71.35**	**84.06**	**88.96**	**88.06**	**12.72**	**17.95**	**17.51**
		Clear Sky	**41.60**	**66.44**	**61.40**	**−15.82**	**−30.73**	**−36.85**	**31.85**	**41.44**	**44.04**	**38.20**	41.37	**44.54**	**3.597**	**5.868**	**7.262**
Transformer	MSE	Over all	113.6	144.4	151.6	**−9.380**	**4.332**	33.72	49.59	81.93	84.97	87.50	95.67	95.83	15.30	18.95	18.56
		Cloudy	128.6	**153.4**	167.6	**−7.272**	**31.24**	65.76	55.71	84.61	92.57	86.49	91.48	91.95	18.67	21.13	21.10
		Clear Sky	41.16	112.0	**84.20**	−15.93	−79.28	−65.83	31.18	77.16	62.39	40.44	60.46	59.17	3.497	11.31	**9.638**
	soft-DTW	Over all	112.0	156.0	**148.6**	13.51	41.97	27.44	52.19	68.93	79.56	85.19	93.29	95.98	13.94	19.00	**18.39**
		Cloudy	127.4	172.4	163.1	19.19	73.38	57.58	57.84	79.23	**87.66**	84.50	90.00	92.07	17.37	21.42	20.80
		Clear Sky	**32.22**	86.65	86.04	**−4.123**	−55.62	−66.22	41.22	**48.07**	**48.12**	**36.08**	56.67	59.21	**1.912**	10.51	9.933
	RSTLoss	Over all	**111.8**	**138.8**	**145.6**	19.27	29.15	**25.89**	**47.35**	**66.93**	**81.79**	**84.40**	**91.33**	**92.06**	**12.62**	**17.02**	**18.39**
		Cloudy	**126.5**	155.5	**160.3**	28.38	44.20	**47.38**	**54.68**	**77.03**	87.99	**84.04**	**89.45**	**89.62**	**16.57**	**20.22**	**19.88**
		Clear Sky	46.32	**62.86**	88.43	−9.061	**−17.61**	**−40.89**	**25.82**	57.00	62.08	41.22	**48.07**	**48.12**	4.292	**5.774**	10.34
GRU	MSE	Over all	**106.2**	**119.7**	126.6	5.225	9.833	**9.496**	45.53	57.95	64.11	86.48	90.25	91.11	12.97	15.13	15.82
		Cloudy	120.3	134.9	141.4	11.18	19.70	21.96	49.55	65.83	**71.70**	85.68	88.81	89.44	15.83	18.19	18.66
		Clear Sky	**36.38**	**47.14**	60.55	−13.28	−20.82	**−29.25**	**31.48**	34.79	41.29	37.81	41.63	44.15	2.961	4.416	**5.884**
	soft-DTW	Over all	109.9	127.4	131.8	5.944	**2.884**	−11.19	48.13	62.21	69.85	86.21	88.60	93.57	12.89	15.00	16.69
		Cloudy	124.4	142.5	143.7	13.14	**12.68**	**1.476**	52.51	69.41	75.26	85.35	87.49	91.81	15.64	17.72	19.05
		Clear Sky	39.17	59.82	84.57	−16.43	−27.55	−50.53	32.69	38.92	54.79	38.36	**39.66**	47.42	3.257	5.441	8.400
	RSTLoss	Over all	106.3	120.2	**126.1**	**5.051**	7.973	10.62	**43.97**	**55.54**	**63.90**	**84.15**	**86.30**	**89.71**	**11.30**	**14.10**	**15.63**
		Cloudy	**118.5**	**132.6**	**139.2**	**10.40**	17.13	23.95	**47.96**	**63.25**	72.15	**83.36**	**85.82**	**88.37**	**14.28**	**17.18**	**18.35**
		Clear Sky	38.27	49.06	**58.14**	**−11.56**	**−20.47**	−30.82	32.06	**34.35**	**41.14**	**37.58**	40.96	**41.46**	**2.869**	**4.283**	6.079
DeepTCN	MSE	Over all	**114.7**	136.5	145.9	4.396	16.91	**9.807**	47.46	66.52	79.31	83.82	88.92	91.65	14.23	17.65	18.42
		Cloudy	**129.8**	152.9	157.5	9.770	33.19	34.06	52.83	75.51	83.50	82.85	87.59	89.00	17.26	20.65	20.50
		Clear Sky	40.90	**61.92**	101.7	**−12.30**	**−33.69**	−65.54	**30.06**	**42.66**	69.05	38.09	**42.99**	**50.21**	3.630	7.117	11.12
	soft-DTW	Over all	115.0	135.8	149.6	6.391	13.82	−31.07	48.53	67.35	82.09	**82.79**	89.59	96.03	14.51	17.48	19.21
		Cloudy	**129.8**	151.8	154.1	13.80	30.34	−12.26	53.87	74.49	81.04	**81.79**	87.64	93.53	17.57	20.29	20.72
		Clear Sky	43.64	63.93	135.0	−16.63	−37.52	−89.52	33.73	47.08	86.07	38.44	48.08	51.93	3.787	7.623	13.92
	RSTLoss	Over all	114.8	**130.7**	**138.2**	**−1.960**	**−5.362**	−14.30	**46.58**	**62.66**	**72.85**	84.80	**87.42**	**90.48**	**14.10**	**16.28**	**17.66**
		Cloudy	130.1	**145.0**	**150.3**	**2.682**	**4.774**	**0.332**	**51.78**	**69.67**	**77.31**	83.92	**86.80**	**87.87**	**17.23**	**19.17**	**19.78**
		Clear Sky	**40.03**	69.56	**91.10**	−16.39	−36.86	**−59.78**	33.43	43.31	**60.29**	**37.64**	44.90	53.23	**3.128**	**6.149**	**10.24**

References

[1] Wiese F, Taillard N, Balembois E, Best B, Bourgeois S, Campos J, Cordroch L, Djelali M, Gabert A, Jacob A, et al. The key role of sufficiency for low demand-based carbon neutrality and energy security across europe. Nat Commun 2024;15(1):9043.

[2] Yang D, Wang W, Gueymard CA, Hong T, Kleissl J, Huang J, Perez MJ, Perez R, Bright JM, Xia X, et al. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. Renew Sustain Energy Rev 2022;161:112348.

[3] Yang D. A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES). J Renew Sustain Energy 2019;11(2).

[4] Jing T, Chen S, Fang J, Porte-Agel F, Lam MC, Li M. Weakly informative priors flow matching for probabilistic regional solar irradiance forecasting. IEEE Trans Sustain Energy 2026;1–14.

[5] Xie Y, Deng N, Jing T, Li M. Real-time estimation of cloud properties and surface downwelling longwave radiation from satellite imagery: A two-stage surrogate modeling approach. Appl Energy 2026;423:128373.

[6] Chen S, Li C, Xie Y, Li M. Global and direct solar irradiance estimation using deep learning and selected spectral satellite images. Appl Energy 2023;352:121979.

[7] Ferreira C, Gama J, Matias L, Botterud A, Wang J. A survey on wind power ramp forecasting. Technical Report, Argonne National Lab.(ANL), Argonne, IL (United States); 2011.

[8] Gonzalez N, Serna-Torre P, Sánchez-Pérez PA, Davidson R, Murray B, Staadecker M, Szinai J, Wei R, Kammen DM, Sunter DA, et al. Offshore wind and wave energy can reduce total installed capacity required in zero-emissions grids. Nat Commun 2024;15(1):6826.

[9] Chen S, Jing T, Yang D, Chen Y, Li M, Chu Y. The role of spatio-temporal satellite-derived products and numerical weather forecasts in data-driven intra-day solar forecasting. Energy AI 2026;100718.

[10] Chu Y, Pedro HT, Li M, Coimbra CF. Real-time forecasting of solar irradiance ramps with smart image processing. Sol Energy 2015;114:91–104.

[11] Truewind A. AWS Truewind's final report for the Alberta forecasting pilot project. Wind Power Forecast PILOT Proj 2008;66.

[12] Kamath C. Understanding wind ramp events through analysis of historical data. In: IEEE PES t&d 2010. IEEE; 2010, p. 1–6.

[13] Kamath C. Associating weather conditions with ramp events in wind power generation. In: 2011 IEEE/PES power systems conference and exposition. IEEE; 2011, p. 1–8.

[14] Florita A, Hodge B-M, Orwig K. Identifying wind and solar ramping events. In: 2013 IEEE green technologies conference (greenTech). IEEE; 2013, p. 147–52.

[15] Cui M, Zhang J, Florita AR, Hodge B-M, Ke D, Sun Y. An optimized swinging door algorithm for wind power ramp event detection. In: 2015 IEEE power & energy society general meeting. IEEE; 2015, p. 1–5.

[16] Nie Y, Li X, Paletta Q, Aragon M, Scott A, Brandt A. Open-source sky image datasets for solar forecasting with deep learning: A comprehensive survey. Renew Sustain Energy Rev 2024;189:113977.

[17] Wen H, Du Y, Chen X, Lim E, Wen H, Jiang L, Xiang W. Deep learning based multistep solar forecasting for PV ramp-rate control using sky images. IEEE Trans Ind Inform. 2020;17(2):1397–406.

[18] Feng C, Zhang J, Zhang W, Hodge B-M. Convolutional neural networks for intra-hour solar forecasting based on sky image sequences. Appl Energy 2022;310:118438.

[19] Feng C, Zhang J. SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. Sol Energy 2020;204:71–8.

[20] Chen S, Li C, Stull R, Li M. Improved satellite-based intra-day solar forecasting with a chain of deep learning models. Energy Convers Manage 2024;313:118598.

[21] Chu Y, Wang Y, Yang D, Chen S, Li M. A review of distributed solar forecasting with remote sensing and deep learning. Renew Sustain Energy Rev 2024;198:114391.

[22] Qin J, Jiang H, Lu N, Yao L, Zhou C. Enhancing solar PV output forecast by integrating ground and satellite observations with deep learning. Renew Sustain Energy Rev 2022;167:112680.

[23] Pérez E, Pérez J, Segarra-Tamarit J, Beltran H. A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant. Sol Energy 2021;218:652–60.

[24] Benavides Cesar L, Amaro e Silva R, Manso Callejo MÁ, Cira C-I. Review on spatio-temporal solar forecasting methods driven by in situ measurements or their combination with satellite and numerical weather prediction (NWP) estimates. Energies 2022;15(12):4341.

[25] Markovics D, Mayer MJ. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew Sustain Energy Rev 2022;161:112364.

[26] Zhu W, Zhang L, Yang M, Wang B. Solar power ramp event forewarning with limited historical observations. IEEE Trans Ind Appl 2019;55(6):5621–30.

[27] Haupt SE, Kosovic B, Jensen T, Lee J, Jimenez P, Lazo J, Cowie J, McCandless T, Pearson J, Weiner G, et al. The SunCast solar-power forecasting system: the results of the public-private-academic partnership to advance solar power forecasting. Natl Cent Atmospheric Res (NCAR), Boulder (CO): Res Appl Lab Weather Syst Assess Program (US) 2016.

[28] Zhang M, Xu C, Sun A, Qiu F, Xie Y. Solar radiation ramping events modeling using spatio-temporal point processes. 2021, arXiv preprint arXiv:2101.11179.

[29] Hirata Y, Aihara K. Predicting ramps by integrating different sorts of information. Eur Phys J Spec Top 2016;225:513–25.

[30] Kumari P, Toshniwal D. Deep learning models for solar irradiance forecasting: A comprehensive review. J Clean Prod 2021;318:128566.

[31] Ajith M, Martínez-Ramón M. Deep learning algorithms for very short term solar irradiance forecasting: A survey. Renew Sustain Energy Rev 2023;182:113362.

[32] Yang D, Alessandrini S, Antonanzas J, Antonanzas-Torres F, Badescu V, Beyer HG, Blaga R, Boland J, Bright JM, Coimbra CF, David M, Frimane A, Gueymard CA, Hong T, Kay MJ, Killinger S, Kleissl J, Lauret P, Lorenz E, van der Meer D, Paulescu M, Perez R, Perpiñán-Lamigueiro O, Peters IM, Reikard G, Renné D, Saint-Drenan Y-M, Shuai Y, Urraca R, Verbois H, Vignola F, Voyant C, Zhang J. Verification of deterministic solar forecasts. Sol Energy 2020;210:20–37.

[33] Vallance L, Charbonnier B, Paul N, Dubost S, Blanc P. Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric. Sol Energy 2017;150:408–22.

[34] Le Guen V, Thome N. Deep time series forecasting with shape and temporal criteria. IEEE Trans Pattern Anal Mach Intell 2022;45(1):342–55.

[35] Xu W, Chen K, Han T, Chen H, Ouyang W, Bai L. Extremecast: Boosting extreme value prediction for global weather forecast. 2024, arXiv preprint arXiv:2402.01295.

[36] Cuturi M, Blondel M. Soft-DTW: a differentiable loss function for time-series. In: International conference on machine learning. PMLR; 2017, p. 894–903.

[37] Kotz S, Nadarajah S. Extreme Value Distributions: Theory and Applications. world scientific; 2000.

[38] Galib AH, McDonald A, Wilson T, Luo L, Tan P. DeepExtrema: A deep learning approach for forecasting block maxima in time series data. In: Proceedings of the thirty-first international joint conference on artificial intelligence. IJCAI, 2022, p. 2980–6. http://dx.doi.org/10.24963/ijcai.2022/413.

[39] S. A, Bhadra S, Dadarkar AZ, P. D. Deep extreme mixture model for time series forecasting. In: Proceedings of the 31st ACM international conference on information and knowledge management. CIKM, ACM; 2022, p. 1726–35. http://dx.doi.org/10.1145/3511808.3557282.

[40] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 1978;26(1):43–9.

[41] Cuturi M, Vert J-P, Birkenes O, Matsui T. A kernel for time series based on global alignments. In: 2007 IEEE international conference on acoustics, speech and signal processing-iCASSP'07, vol. 2, IEEE; 2007, p. II–413.

[42] Bellman R. Dynamic programming. Science 1966;153(3731):34–7.

[43] Frías-Paredes L, Mallor F, León T, Gastón-Romeo M. Introducing the Temporal Distortion Index to perform a bidimensional analysis of renewable energy forecast. Energy 2016;94:180–94.

[44] Jeong Y-S, Jeong MK, Omitaomu OA. Weighted dynamic time warping for time series classification. Pattern Recognit 2011;44(9):2231–40.

[45] Blondel M, Mensch A, Vert J-P. Differentiable divergences between time series. In: International conference on artificial intelligence and statistics. PMLR; 2021, p. 3853–61.

[46] Zhang M, Ding D, Pan X, Yang M. Enhancing time series predictors with generalized extreme value loss. IEEE Trans Knowl Data Eng 2021;35(2):1473–87.

[47] Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data 2018;5(1):1–12.

[48] Jing T, Chen S, Navarro-Alarcon D, Chu Y, Li M. SolarFusionNet: Enhanced solar irradiance forecasting via automated multi-modal feature selection and cross-modal fusion. IEEE Trans Sustain Energy 2024;1–13.

[49] Driemel A, Augustine J, Behrens K, Colle S, Cox C, Cuevas-Agulló E, Denn FM, Duprat T, Fukuda M, Grobe H, et al. Baseline surface radiation network (BSRN): structure and data description (1992–2017). Earth Syst Sci Data 2018;10(3):1491–501.

[50] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, Wald L, Schroedter-Homscheidt M, Hoyer-Klick C, Arola A, et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmospheric Meas Tech 2013;6(9):2403–18.

[51] Yang D. SolarData package update v1. 1: R functions for easy access of baseline surface radiation network (BSRN). Sol Energy 2019;188:970–5.

[52] Sengupta M, Xie Y, Lopez A, Habte A, Maclaurin G, Shelby J. The national solar radiation data base (NSRDB). Renew Sustain Energy Rev 2018;89:51–60.

[53] Chen S-A, Li C-L, Yoder N, Arik SO, Pfister T. TSMixer: An all-MLP architecture for time series forecasting. 2023, arXiv preprint arXiv:2303.06053.

[54] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35, 2021, p. 11106–15.

[55] Vaswani A. Attention is all you need. 2017, arXiv preprint arXiv:1706.03762.

[56] Cho K. Learning phrase representations using RNN encoder–decoder for statistical machine translation. 2014, arXiv preprint arXiv:1406.1078.

[57] Chen Y, Kang Y, Chen Y, Wang Z. Probabilistic forecasting with temporal convolutional neural network. Neurocomputing 2020;399:491–501.

[58] Frías-Paredes L, Mallor F, Gastón-Romeo M, León T. Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors. Energy Convers Manage 2017;142:533–46.

[59] Kingma DP. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980.

[60] Bright JM, Sun X, Gueymard CA, Acord B, Wang P, Engerer NA. Bright-sun: A globally applicable 1-min irradiance clear-sky detection model. Renew Sustain Energy Rev 2020;121:109706.

[61] Li Y, Song L, Zhang S, Kraus L, Adcox T, Willardson R, Komandur A, Lu N. A TCN-based hybrid forecasting framework for hours-ahead utility-scale PV forecasting. IEEE Trans Smart Grid 2023;14(5):4073–85.

[62] Tong J, Xie L, Fang S, Yang W, Zhang K. Hourly solar irradiance forecasting based on encoder–decoder model using series decomposition and dynamic error compensation. Energy Convers Manage 2022;270:116049.