Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

IceCast: A hybrid post-processing framework integrating historical observations and physical forecasts for transmission-line icing prediction

Shaohui Zhou
China Southern Power Grid Co., Ltd.
Bo Gong
China Southern Power Grid Co., Ltd.
Qi Guo
China Southern Power Grid Co., Ltd.
Xingya Xi
China Meteorological Administration (CMA)
Zhengqi Lu
Zhejiang Institute of Meteorological Sciences
Haipeng Zhang
China Southern Power Grid Co., Ltd.
Yucheng Su
China Southern Power Grid Co., Ltd.
Haofei Sun
Shanghai Typhoon Institute, China Meteorological Administration
Zhiqiu Gao
zgao@nuist.edu.cn

Nanjing University of Information Science and Technology

Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11618806/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

IceCast: A hybrid post-processing framework integrating historical observations and physical forecasts for transmission-line icing prediction

Shaohui Zhou[1], Bo Gong[1], Qi Guo[1], Xingya Xi[2], Zhengqi Lu[3], Haipeng Zhang[1], Yucheng Su[1], Haofei Sun[4] and Zhiqiu Gao[5]*

[1] Electric Power Research Institute, China Southern Power Grid Co., Ltd., Guangzhou 510700, China

[2] Guangdong Provincial Key Laboratory of Regional Numerical Weather Prediction, Guangzhou Institute of Tropical and Marine Meteorology, China Meteorological Administration (CMA), Guangzhou, 510080, China

[3] Zhejiang Institute of Meteorological Sciences, Hangzhou 310008, China

[4] Shanghai Typhoon Institute, China Meteorological Administration, Shanghai 200030, China

[5] School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing 210044, China

* Correspondence: zgao@nuist.edu.cn

Abstract

Icing on transmission lines threatens power grids, demanding accurate ice-thickness forecasts for early warning. Physical forecasts represent future conditions but contain errors, whereas observations capture recent icing evolution. Here we introduce IceCast, a hybrid post-processing model integrating 48 h of hourly pylon ice-thickness observations with a 24 h physical icing forecast calculated using corrected forecast temperature. IceCast learns hourly residual corrections to predict equivalent ice thickness at a reference height of 10 m above ground. We trained on ten cold waves during 2022–2025, tested on three others from that period, and validated on six events during winter 2025–2026. For icing occurrence (ice thickness exceeding 2 mm), IceCast increased the critical success index (CSI) by 31.9% and 16.8% relative to the physical forecast in the testing and validation sets, respectively. Compared with mean squared error (MSE) training, a composite loss further improved CSI for ice thickness exceeding 20 mm. Sensitivity analyses showed that historical icing changes and future physical-forecast changes provided complementary predictive information beyond the initial icing state. These findings support hybrid post-processing for predicting icing occurrence and heavy icing.

Introduction

Ice accretion on transmission lines increases the mechanical loads on conductors and pylons and may cause galloping, flashover, line breakage and pylon collapse, threatening the safe operation of power grids[1–3]. In the complex terrain of southern China, cold-air intrusions, moisture transport, topography and local micrometeorological conditions jointly affect the formation and distribution of freezing rain and in-cloud supercooled liquid water, resulting in substantial differences in icing severity among transmission lines and pylons[2]. De-icing operations, line inspections and emergency dispatch require not only determining whether icing occurs, but also anticipating when ice thickness will increase, persist or decrease. Accurate prediction of future ice thickness and its hourly changes is therefore important for icing warnings and power-grid disaster prevention and emergency operations.

Existing icing prediction approaches mainly include physical-driven models (PDMs) and data-driven models (DDMs)[4]. Common physical-driven icing models include Makkonen-type models and the Jones model. Makkonen-type models calculate ice growth from meteorological conditions and the collision, collection and freezing of liquid water droplets on conductors, whereas the Jones model mainly estimates freezing-rain ice growth from precipitation rate and wind speed[5,6]. When coupled with numerical weather prediction (NWP), PDMs can use future meteorological conditions provided by NWP to calculate changes in ice thickness during the forecast period[7–10], although their predictions are still affected by errors in meteorological forcing, the initial icing state and process parameterizations[7,8,11–14]. DDMs do not explicitly solve the above icing processes, but instead use historical meteorological and icing data to learn nonlinear relationships among meteorological conditions, existing icing states and subsequent icing[15–17]. In addition to historical monitoring data, future NWP meteorological variables can also be used as inputs to DDMs, allowing data-driven forecasts to account for weather changes during the forecast period[17,18]. However, the predictive capability of DDMs depends on the coverage of the training samples, and their generalization to rare icing processes and unseen events may remain limited[15]. Existing studies have begun to explore hybrid forecasting approaches that combine the icing processes represented by PDMs, or their model outputs, with the data-learning capability of DDMs, so as to exploit the respective strengths of the two approaches[17,19].

Hybrid forecasting can combine physical-driven and data-driven methods at different stages. When a physical forecast is already available, one common approach is to use paired forecasts and observations to learn

systematic forecast errors and then apply statistical or machine-learning methods for post-processing[20–22]. In short-term precipitation forecasting, recent observation sequences can also be used together with future atmospheric forecasts, allowing the model to use both observational information before forecast initialization and atmospheric information during the forecast period[23]. These studies suggest that historical observations can serve not only as a reference for error correction but also as time-series inputs describing recent process changes. However, in post-processing of physical forecasts of continuous ice thickness, whether recent observed icing sequences can provide information beyond the ice-thickness state at forecast initialization still lacks targeted investigation. Online monitoring records the actual changes in ice thickness before forecast initialization, whereas PDMs provide changes in ice thickness during the forecast period according to future NWP meteorological conditions. It is therefore necessary to further investigate whether the continuous changes in historical observations provide predictive information beyond the last observed ice thickness, whether the future physical icing sequence provides information beyond its initial ice thickness, and whether jointly using the two time series can further improve future ice-thickness prediction.

The limited number of severe icing samples is another challenge for data-driven post-processing[15]. In the monitoring dataset used in this study, no-icing and light-icing samples account for the majority, whereas larger ice thicknesses are mainly concentrated in a small number of pylons and cold-wave events, resulting in an imbalanced distribution of training samples across different icing intensities. Under these conditions, using an overall continuous-error metric such as mean squared error (MSE) as the sole training objective does not ensure simultaneous improvement in the prediction of icing occurrence, hourly changes and moderate-to-heavy icing. Similar studies of heavy-precipitation forecasting have shown that sample imbalance and the design of the training loss function can affect model learning and predictive performance for the relatively small number of heavy-precipitation events[23–25]. Therefore, in addition to reducing numerical errors in ice thickness, icing post-processing also needs to consider icing occurrence, ice-thickness magnitude, hourly changes and predictive performance across different icing intensities.

Based on these issues, we propose IceCast, a hybrid post-processing model that combines historical icing observations with future physical icing forecasts. IceCast takes 48 h of hourly observed ice thickness and a 24 h physical icing forecast for the same pylon as inputs to a Transformer and learns hourly residuals to correct the physical forecast. The physical icing calculation uses corrected forecast temperature, and the final prediction variable is standardized as equivalent ice thickness at a reference height of 10 m above ground. To address the different prediction objectives and the imbalance of samples across icing intensities, we further construct a composite training loss that considers icing occurrence, ice-thickness magnitude, hourly changes and icing intensity. By comparing separate and combined inputs of historical observations and physical forecasts, and by conducting sensitivity experiments targeting the temporal changes in the two sequences, we examine whether the historical observation sequence and the future physical forecast sequence can provide predictive information beyond their respective end value and initial value, whether jointly using the two types of information can further improve future ice-thickness prediction, and how different training loss functions affect predictive performance for general and heavy icing.

Results

Figure 1 shows the study region and the distribution of monitored pylons. After quality control, 2,082 pylons were included in the analysis. Ten cold-wave events during 2022–2025 were used for model training, three additional independent events from the same period were used for testing, and six events during winter 2025–2026 were used for validation (Fig. 1c–f). Each forecast sample represents one 24 h forecast initialized at 08:00 for one pylon. Icing occurrence was defined as observed or predicted ice thickness exceeding 2 mm, and observed-icing errors were calculated for hours when observed ice thickness exceeded 2 mm. The complete workflow from icing-monitoring and meteorological-forecast data preparation, through generation of the temperature-corrected physical icing forecast, to IceCast inputs, corrected outputs and evaluation is shown in Supplementary Fig. S1. Dataset construction, model architecture, training settings and evaluation methods are described in Methods.

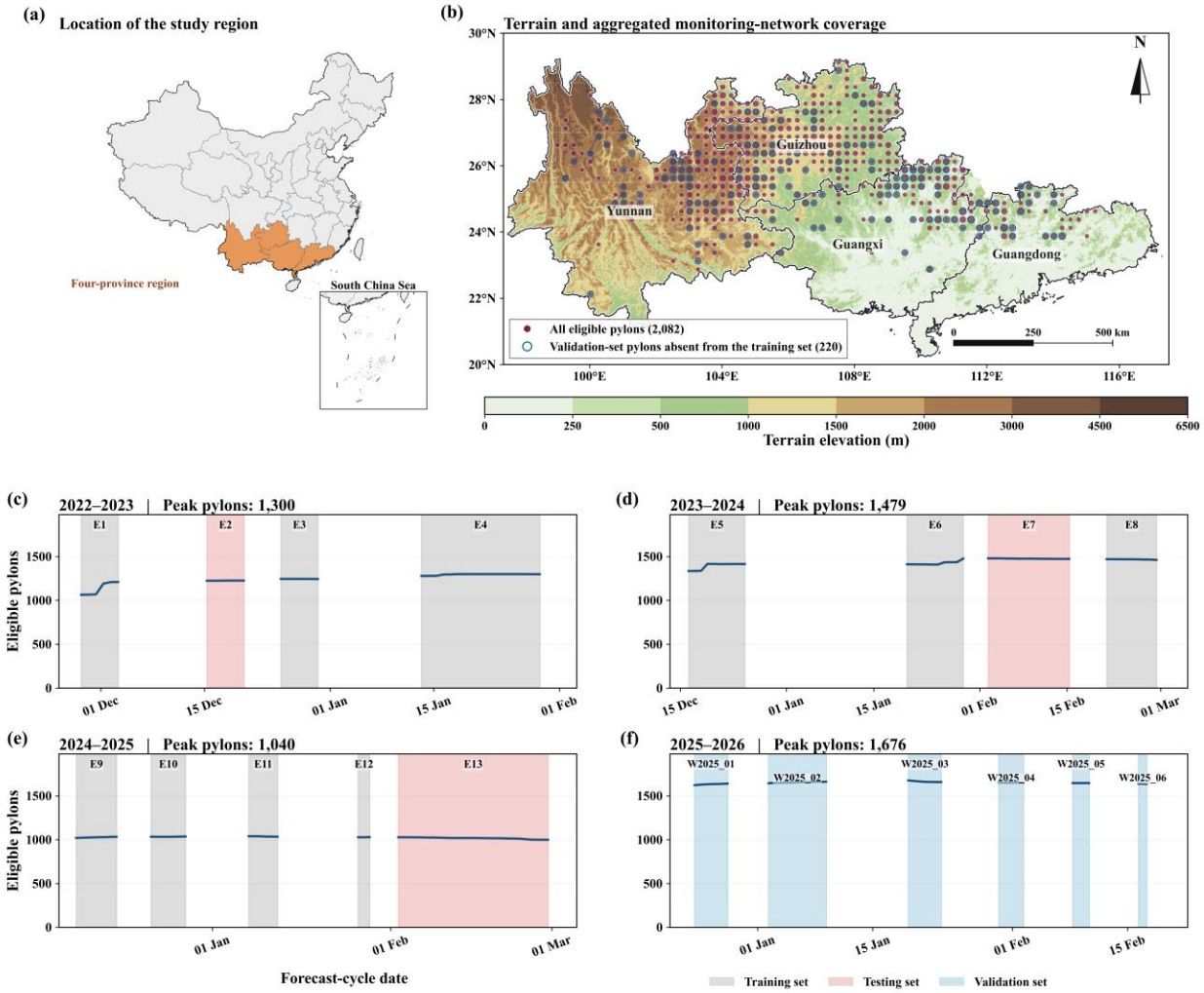

Fig. 1. Study region and monitored pylons after quality control. (a) Location of the four-province study region within China; the South China Sea is shown in a separate inset. **(b)** Regional topography and monitored-pylon distribution. Filled symbols represent all 2,082 pylons included in the analysis, aggregated on a 0.25° grid; blue outlines indicate grid cells containing the 220 validation-set pylons that were not present in the training set. Exact pylon coordinates and identifiers are not shown for security reasons. **(c)–(f)** Number of pylons included at each daily 08:00 initialization during the four winters from 2022–2023 to 2025–2026. Grey, red and blue shading denote training, testing and validation events, respectively.

Combining historical observations with temperature-corrected physical forecasts

We first tested whether jointly using historical observations and future temperature-corrected physical forecasts improved icing-occurrence prediction, and separated the effects of input information, output formulation, training loss and model architecture. Figure 2a,b compares the critical success index (CSI) at 2 mm across different inputs, output formulations and training losses in the testing and validation sets, whereas Fig. 2c tests whether the gain from adding the temperature-corrected physical forecast depends on model architecture. Six Transformer configurations were evaluated in Fig. 2a,b. Each was trained and evaluated with three random seeds (42, 2025 and 3407), giving 18 model-training runs. The History-only Transformer used only 48 h of historical observations, the Physics-only Transformer used only the 24 h temperature-corrected physical forecast, and the Aligned-fusion Transformer used the correctly paired historical observations and temperature-corrected physical forecast. In the Shuffled-physics Transformer, the correctly paired temperature-corrected physical forecast was replaced by a temperature-corrected physical forecast from another sample within the same cold-wave event. These four models directly predicted future ice thickness and were trained with MSE. The Residual-fusion Transformer and IceCast used the same joint inputs and learned residual corrections to the temperature-corrected physical forecast; the former used MSE and the latter used the occurrence–magnitude–trajectory–intensity (OMTI) loss.

Both input sources provided additional information for icing-occurrence prediction. In the testing set, 2 mm CSI was 0.546 ± 0.015 for the History-only Transformer and 0.605 ± 0.007 for the Physics-only Transformer, whereas the Aligned-fusion Transformer reached 0.621 ± 0.011. The corresponding validation-set values were 0.490 ± 0.011, 0.537 ± 0.013 and 0.565 ± 0.004. Relative to the History-only Transformer, adding the temperature-corrected physical forecast increased CSI by 0.075 in both datasets, corresponding to relative

improvements of 13.6% and 15.4% in the testing and validation sets. Relative to the Physics-only Transformer, adding historical observations increased CSI by 0.016 and 0.028, corresponding to relative improvements of 2.6% and 5.3%. Historical icing observations and future temperature-corrected physical icing forecasts therefore provided complementary information, and their joint use improved icing-occurrence prediction.

When the temperature-corrected physical forecast corresponding to the current pylon and forecast period was replaced by another forecast sequence from the same cold wave, 2 mm CSI of the Shuffled-physics Transformer decreased to 0.551 ± 0.009 in the testing set and 0.489 ± 0.006 in the validation set, close to the History-only Transformer. By contrast, the Aligned-fusion Transformer using the correctly paired temperature-corrected physical forecast was higher by 0.070 and 0.076, respectively. Thus, the performance gain depended on future temperature-corrected physical-forecast information corresponding to the current pylon and forecast period.

After establishing the complementarity of the two input sources, we compared the effects of output formulation and training loss on 2 mm icing-occurrence prediction (Fig. 2a,b). With identical inputs and MSE training, changing from direct ice-thickness prediction to residual correction did not improve CSI. The Residual-fusion Transformer achieved CSI values of 0.618 and 0.560 in the testing and validation sets, slightly below the 0.621 and 0.565 obtained by the Aligned-fusion Transformer. Replacing MSE with OMTI increased CSI to 0.627 and 0.572, the highest values in both datasets. Thus, for 2 mm icing-occurrence prediction, residual correction alone did not provide a clear gain, whereas the OMTI loss improved icing detection. The effects of OMTI on moderate-to-heavy icing and ice-thickness errors are examined in the next section.

To test whether the benefit of joint inputs depended on a particular architecture, we also compared Transformer, long short-term memory (LSTM) networks with attention, and decomposition-linear (DLinear) models (Fig. 2c). Under MSE training, Aligned-fusion LSTM-attention increased 2 mm CSI by 0.047 and 0.051 relative to History-only LSTM-attention in the testing and validation sets, respectively. The corresponding gains for Aligned-fusion DLinear over History-only DLinear were 0.059 and 0.081. The Transformer showed the same direction of improvement. The benefit of adding the temperature-corrected physical icing forecast therefore occurred across all three architectures and was not specific to the Transformer.

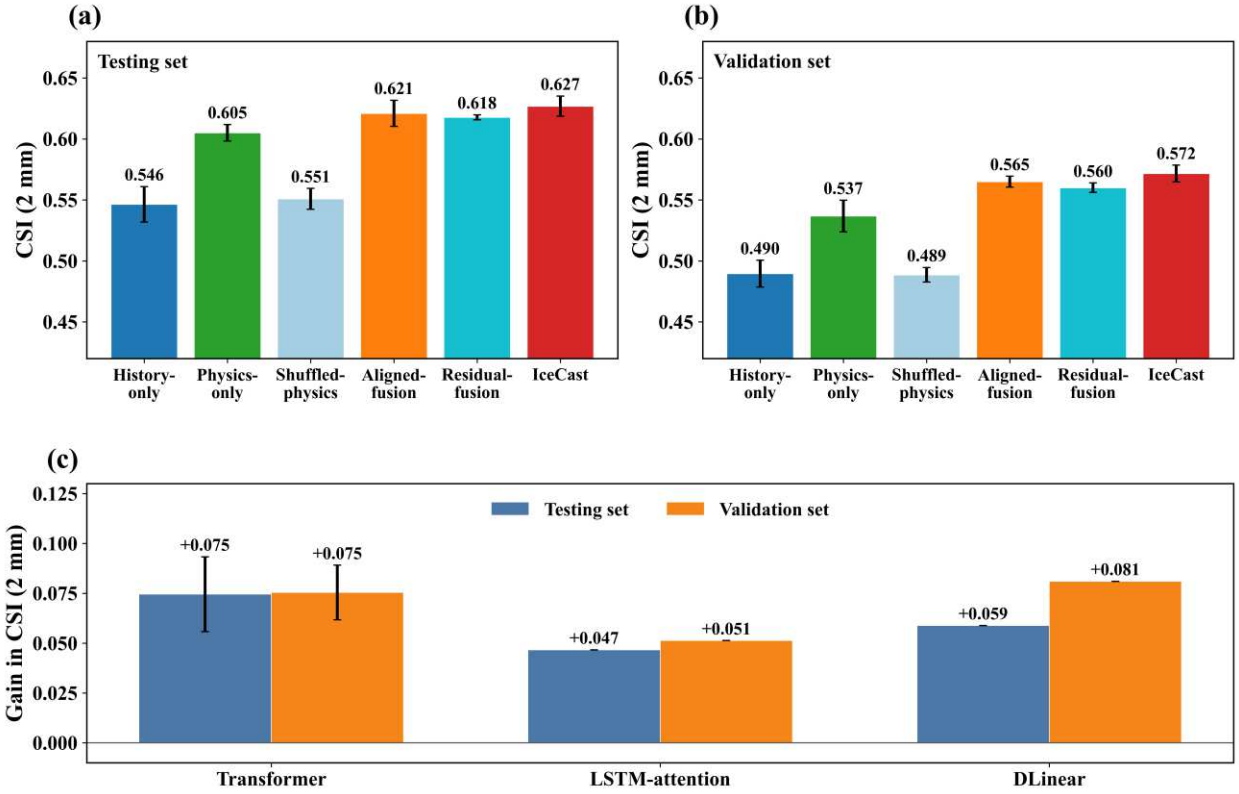

Fig. 2. Effect of jointly using historical observations and temperature-corrected physical icing forecasts. (a), (b) 2 mm CSI in the testing and validation sets for different inputs, output formulations and training losses. The History-only Transformer uses only 48 h of historical observations, whereas the Physics-only Transformer uses only the 24 h temperature-corrected physical forecast. The Aligned-fusion Transformer uses correctly paired historical observations and temperature-corrected physical forecasts, whereas the Shuffled-physics Transformer replaces the temperature-corrected physical forecast with another forecast sequence from the same cold wave. These models directly predict ice thickness and are trained with MSE. The Residual-fusion Transformer uses the same inputs as the Aligned-fusion Transformer but learns residual corrections to the temperature-corrected physical forecast and is trained with MSE. IceCast uses the same residual formulation and is trained with OMTI. Values are means across random seeds 42, 2025 and 3407; error bars indicate one standard deviation (SD). **(c)** Change in 2 mm CSI after adding the correctly paired temperature-corrected physical forecast to historical observations across different model architectures. LSTM-attention and DLinear compare Aligned-fusion with History-only configurations. Transformer error bars show the SD of paired differences across three random seeds; LSTM-attention and DLinear use the fixed random seed 42.

Prediction skill across icing intensities and lead times

We next evaluated the prediction performance of different methods across icing intensities and forecast lead times. The comparison included the temperature-corrected physical forecast, Persistence as a baseline for maintaining the current icing state, and three hybrid correction methods: extreme gradient boosting (XGBoost), the Aligned-fusion Transformer and IceCast. XGBoost and the Aligned-fusion Transformer both use 48 h of historical observations and the 24 h temperature-corrected physical forecast and are trained with MSE, enabling comparison of different models under the same joint inputs. IceCast uses the same joint inputs but applies residual correction and is trained with OMTI. Performance was evaluated using icing detection at multiple ice-thickness thresholds (Fig. 3a,b), ice-thickness prediction within observed intensity intervals (Fig. 3c,d), and lead-time-dependent 2 mm CSI and observed-icing mean absolute error (MAE) (Fig. 3e–h).

Overall, IceCast improved 2 mm icing-occurrence prediction while reducing ice-thickness errors during observed icing (Supplementary Tables S1 and S2; Supplementary Fig. S2). In the testing set, IceCast increased 2 mm CSI from 0.475 for the temperature-corrected physical forecast to 0.627 ± 0.008, a relative improvement of 31.9%, while reducing the false alarm ratio (FAR) and miss alarm ratio (MAR) by 42.7% and 19.2%, respectively. Observed-icing root mean square error (RMSE) and MAE decreased by 25.8% and 28.5%. In the validation set, CSI increased from 0.489 to 0.572 ± 0.007, a relative improvement of 16.8%, while observed-icing RMSE and MAE decreased by 22.1% and 20.6%. When errors were calculated only for hit hours in which both observed and predicted ice thickness exceeded 2 mm (Supplementary Tables S1 and S2), IceCast reduced RMSE and MAE in the testing set, whereas its validation-set errors were close to those of the temperature-corrected physical forecast.

CSI at different ice-thickness thresholds was used to evaluate detection across icing intensities (Fig. 3a,b). CSI was calculated at thresholds of 2, 5, 10, 20 and 30 mm. At 2, 5, 10 and 20 mm, XGBoost and the Aligned-fusion Transformer produced similar CSI values, indicating that changing the data-driven model architecture

alone did not generate a clear performance difference when inputs and training objectives were similar. IceCast, by contrast, achieved higher CSI than XGBoost and the Aligned-fusion Transformer at every threshold, with the separation from the other methods increasing at higher thresholds. At 30 mm, CSI for XGBoost and IceCast was 0.143 and 0.457 in the testing set and 0.052 and 0.351 in the validation set, respectively. IceCast also reduced observed-icing RMSE and MAE relative to the temperature-corrected physical forecast by 25.8% and 28.5% in the testing set and by 22.1% and 20.6% in the validation set. Thus, the advantage of IceCast extended beyond 2 mm icing occurrence to the detection of moderate-to-heavy icing and the prediction of ice thickness when icing was observed.

To determine whether the improvement at higher ice-thickness thresholds arose from the training loss, we held the inputs, Transformer architecture and residual output formulation fixed and compared only MSE and OMTI. With OMTI, 20 mm CSI increased from 0.384 ± 0.062 to 0.543 ± 0.016 in the testing set and from 0.362 ± 0.159 to 0.527 ± 0.036 in the validation set, corresponding to relative gains of 41.6% and 45.6%. At the same time, observed-icing RMSE decreased by 15.6% and 13.6%, whereas all-sample RMSE increased by 12.1% and 9.9%, and the 2 mm threshold showed fewer misses but more false alarms (Supplementary Tables S1 and S2). OMTI therefore did not primarily reduce mean error across all samples; instead, it increased sensitivity to moderate-to-heavy icing and improved ice-thickness prediction during observed icing.

Threshold-based CSI only evaluates whether predicted ice thickness exceeds a specified threshold, whereas the interval score compares predictions with the observed ice-thickness range within seven intensity intervals (Fig. 3c,d). In the testing set, IceCast outperformed the temperature-corrected physical forecast in every interval with available samples through (30, 40] mm. In the validation set, IceCast outperformed both the temperature-corrected physical forecast and XGBoost in every observed interval within (5,40] mm. Persistence performed better over 15–40 mm and reached 0.853 in the (30, 40] mm interval, compared with 0.377 for IceCast. In the validation set, the 75 pylon-hours above 30 mm came from only two pylons and three events, including only seven pylon-hours above 40 mm (Supplementary Table S3). IceCast therefore improved over the temperature-corrected physical forecast and other post-processing methods across most moderate-to-heavy icing intervals, but Persistence remained competitive for long-lasting heavy icing represented by very few samples.

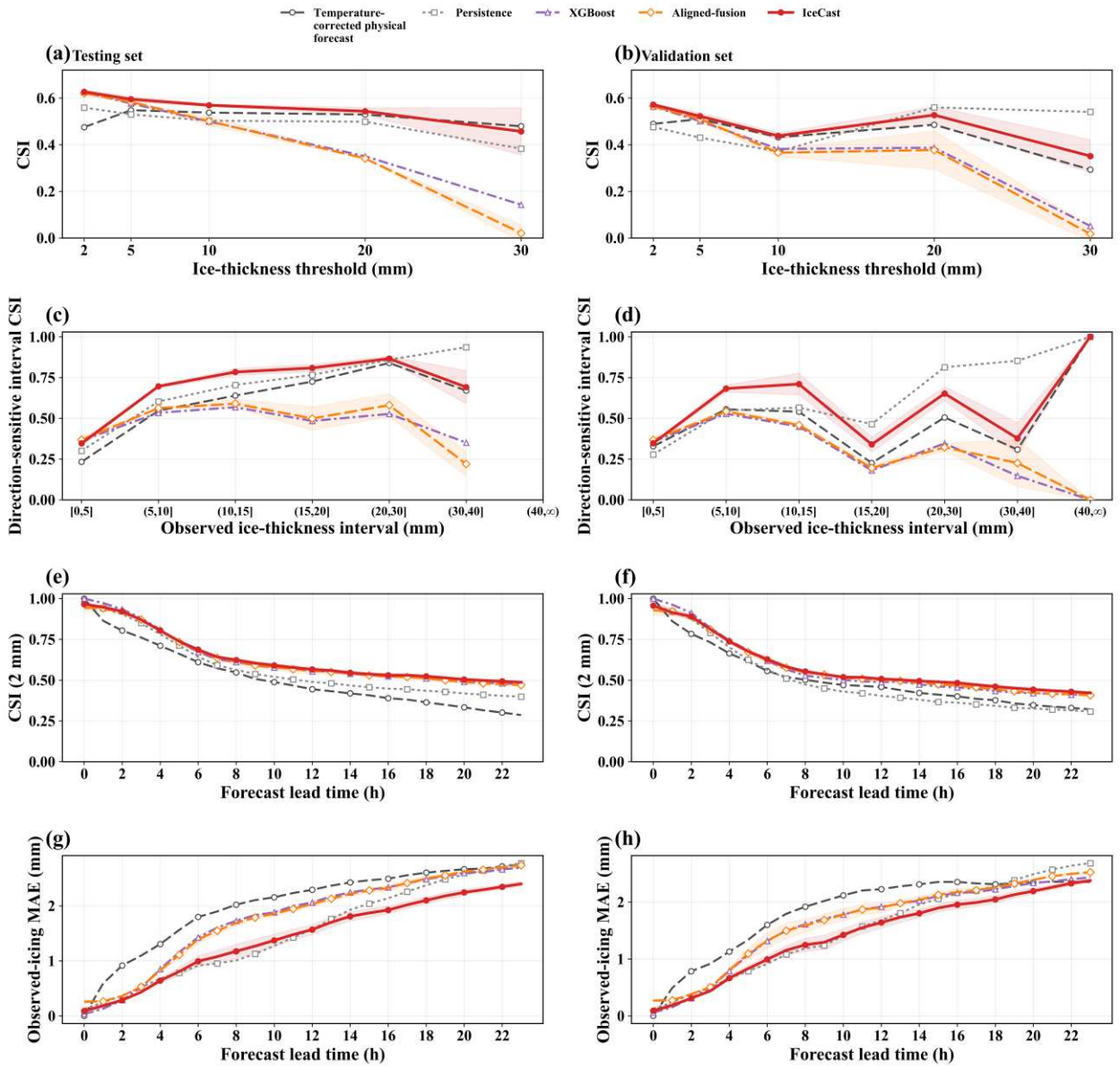

Fig. 3. Prediction skill across icing intensities and forecast lead times. (a), (b) CSI at different icing-occurrence thresholds. **(c), (d)** Direction-sensitive interval CSI; interval definitions, tolerances and sample counts are given in Methods and Supplementary Table S3. Scores are not shown for intervals with no observations. **(e), (f)** 2 mm CSI as a function of forecast lead time. **(g), (h)** Observed-icing MAE as a function of forecast lead time, including missed icing hours. Columns show the testing and validation sets, respectively. Neural-network curves are means across three random seeds and shading denotes ± one SD; deterministic baselines are evaluated once.

After comparing performance across icing intensities, we examined forecast lead time (Fig. 3e–h). At lead 23, IceCast CSI in the testing set was 70.4% higher than that of the temperature-corrected physical forecast, with values of 0.487 and 0.286, respectively. In the validation set, CSI increased by 31.3%, from 0.321 to 0.421.

For ice-thickness errors, IceCast reduced observed-icing MAE relative to the temperature-corrected physical forecast at every lead from 1 to 23 in the testing set and at leads 1–22 in the validation set. At lead 23, MAE was 2.400 mm for IceCast and 2.750 mm for the temperature-corrected physical forecast in the testing set; in the validation set, the two values were nearly identical at 2.377 and 2.352 mm. Persistence had lower observed-icing MAE than IceCast at testing leads 5–11 and validation leads 4–10. Overall, IceCast maintained improved icing-occurrence prediction throughout the forecast period and reduced observed-icing MAE at almost all lead times, although Persistence remained competitive at some earlier leads.

Finally, to exclude the possibility that use of the known initial ice thickness at lead 0 inflated pooled scores, we compared CSI with and without lead 0. Lead 0 was initialized from the 08:00 observed ice thickness available at forecast issuance. After excluding this lead, the 2 mm CSI advantage of IceCast over the temperature-corrected physical forecast increased from 0.152 to 0.155 in the testing set and from 0.082 to 0.085 in the validation set. IceCast therefore retained a stable CSI advantage over leads 1–23, showing that the improvement was not caused by including the known initial ice thickness in the score but arose from correction of subsequent forecast hours.

Prediction performance across cold-wave events and regions

We next evaluated IceCast across individual cold-wave events, pylons not represented in training, and different regions, and examined representative successful and unsuccessful cases. Across the nine cold waves in the testing and validation sets, IceCast achieved higher 2 mm CSI than the temperature-corrected physical forecast in eight events (88.9%): all three testing events and five of six validation events (Supplementary Table S4). The only exception was the short and weak event on 16–17 February 2026. During the 30 January–2 February 2026 event, Persistence outperformed IceCast, although it remained below IceCast in the pooled validation results. The pooled CSI improvement therefore reflected gains across most cold-wave events rather than being dominated by a single event.

The validation set contained 220 pylons that were not present in the training set, contributing 5,779 forecast samples and 281 observed-icing pylon-hours. On these unseen pylons, IceCast achieved a 2 mm CSI of 0.548 ± 0.040 and an observed-icing MAE of 1.201 ± 0.023 mm, compared with 0.325 and 1.496 mm for the temperature-corrected physical forecast. IceCast therefore improved both icing-occurrence prediction and ice-thickness accuracy during observed icing on pylons not seen during training.

We then compared performance across prefecture-level administrative regions (Fig. 4). Among regions for which CSI was defined and at least 20 observed-icing pylon-hours were available, CSI improved in 91.7% of testing-set regions and 90.5% of validation-set regions. Median CSI increments were 0.204 and 0.092, respectively. The evaluated regions and magnitude of local improvements varied among cold-wave events (Supplementary Figs. S3 and S4). Overall, IceCast improved performance not only across most cold-wave events but also across most evaluated regions in both datasets.

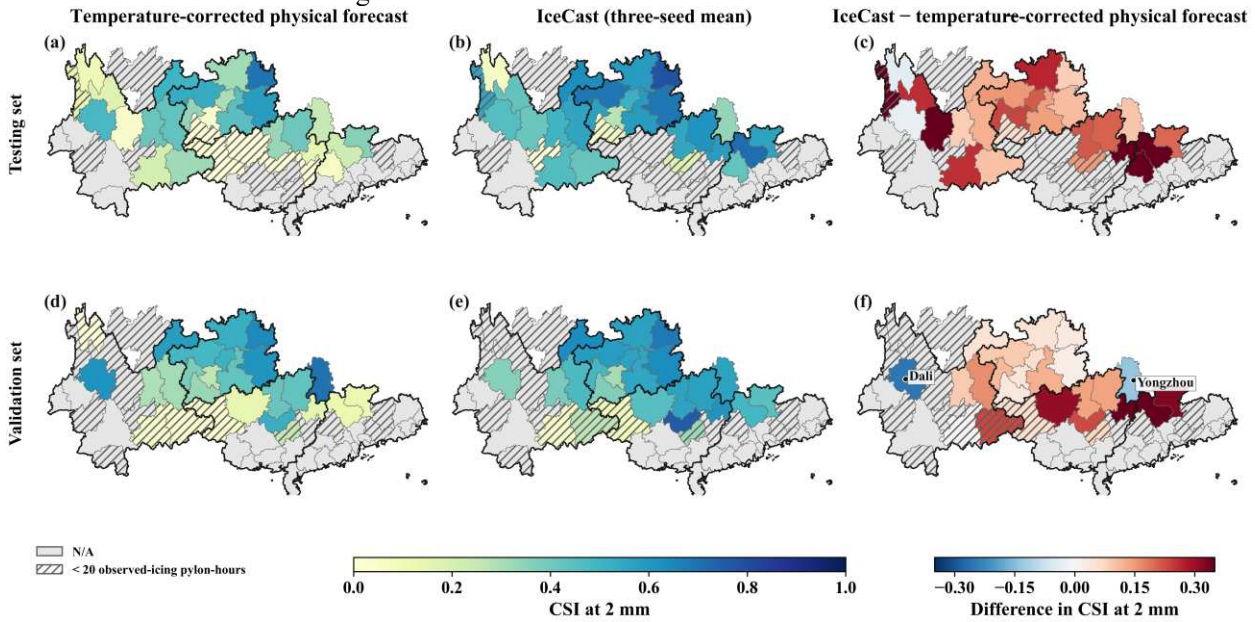

Fig. 4. Spatial prediction skill at the prefecture level. (a)–(c) Testing set; **(d)–(f)** validation set. **(a), (d)** Temperature-corrected physical forecast; **(b), (e)** mean 2 mm CSI of IceCast across three random seeds; **(c), (f)** difference between the two. Dali and Yongzhou are labelled in **(f)** and examined in Supplementary Fig. S5. Scores for each prefecture-level region are pooled over all aligned pylon-hours within that region. Hatching denotes regions with fewer than 20 observed-icing pylon-hours, grey denotes undefined scores, and thick outlines indicate the core study region.

To illustrate model behavior during individual icing processes, we selected three complete events in which IceCast performed well (Fig. 5). During the 15–21 December 2022 event, the temperature-corrected physical forecast delayed the initial growth and predicted ice decay too early, whereas IceCast followed the observed growth more closely and maintained icing for longer. During 2–16 February 2024, successive temperature-corrected physical forecasts first overestimated peak ice thickness and later underestimated persistent ice; IceCast reduced both errors and better followed changes above 25 mm. During 24–29 December 2025, IceCast reproduced the onset, peak and decay of icing more closely. These examples show that IceCast can correct biases in the temperature-corrected physical forecast during growth, peak icing and subsequent decay.

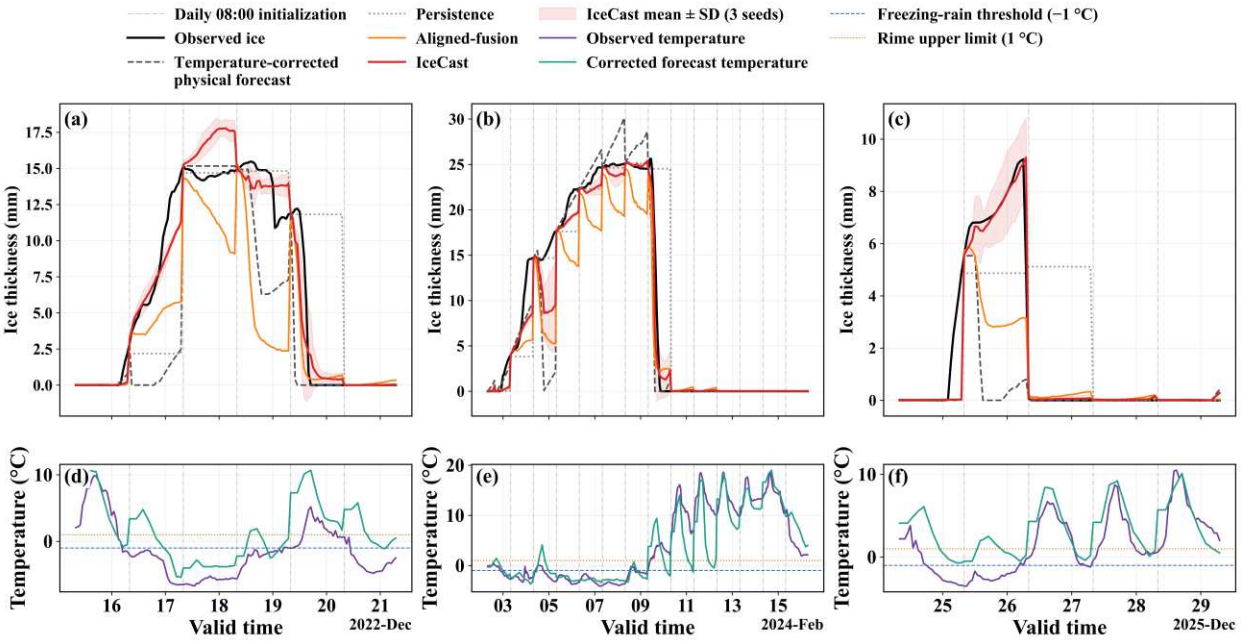

Fig. 5. Three complete icing processes in which IceCast performed well. (a), (d) 15 December 2022 08:00 to 21 December 2022 07:00; process RMSE values for the temperature-corrected physical forecast, Persistence, Aligned-fusion Transformer and IceCast are 4.56, 4.99, 4.57 and 1.69 mm, respectively. **(b), (e)** 2 February 2024 08:00 to 16 February 2024 07:00; corresponding RMSE values are 3.78, 5.43, 4.02 and 2.12 mm. **(c), (f)** 24 December 2025 08:00 to 29 December 2025 07:00; corresponding RMSE values are 2.96, 2.68, 1.98 and 0.80 mm. **(a)–(c)** Observed and predicted ice thickness. **(d)–(f)** Monitored temperature and corrected forecast temperature. Horizontal reference lines denote the −1 °C freezing-rain threshold and the 1 °C upper rime-temperature limit. Forecasts from 08:00 to 07:00 the following day are concatenated by valid time; vertical dash-dotted lines denote successive daily 08:00 initializations. IceCast curves are means across three random seeds, and red envelope shading denotes one SD.

Despite the overall improvement, failures remained in some local regions and heavy-icing processes. Dali (−0.253) and Yongzhou (−0.135) showed the largest validation-set CSI decreases, mainly because of increased false alarms. During one Dali cold wave, after recent observed growth, IceCast continued to predict increasing ice even as the temperature-corrected physical forecast began to decrease and observed icing had disappeared. During one Yongzhou event, IceCast maintained large ice thickness after both the observations and the temperature-corrected physical forecast had declined (Supplementary Fig. S5), indicating that the model can sometimes prolong an earlier icing state incorrectly. In the two heavy-icing processes shown in Supplementary Fig. S6, IceCast delayed the premature decrease in the temperature-corrected physical forecasts but still underestimated persistent or renewed icing. Thus, although IceCast improved prediction across most events and regions, substantial errors remained for local ice decay and for heavy icing that persisted or began growing again.

Robustness across temperature-correction errors

The preceding failure cases indicate that IceCast can still produce large correction errors when the temperature-corrected physical forecast does not adequately represent icing onset, growth or decay. To examine such cases, we diagnosed validation hours with observed ice growth but little or no growth in the temperature-corrected physical forecast (Supplementary Fig. S7 and Supplementary Table S5). The sample contained 315 hours from 130 pylons. Of these, 107 hours from 59 pylons (34.0%) satisfied the other rime conditions but failed the relative-humidity (RH) condition. A further eight hours had freezing efficiency no greater than 0.1 despite being classified as rime. These results suggest that icing-onset criteria and growth-efficiency parameters in the physical model may contribute to some missed or underestimated growth.

Independent observations corresponding to RH, liquid water content (LWC) and related microphysical quantities were not available, preventing quantitative separation of input errors from parameterization errors in the temperature-corrected physical icing forecast and the subsequent IceCast correction. Temperature, by contrast, was available from concurrent monitoring. We therefore defined the temperature-correction error as the absolute difference between corrected forecast temperature and monitored temperature and stratified samples by this quantity to test whether IceCast retained an advantage over the temperature-corrected physical forecast under different correction-error magnitudes (Fig. 6).

We calculated 2 mm CSI for temperature-correction-error bins of 0–1.5, >1.5–3, >3–5 and >5 °C (Fig. 6a,b). Relative to the temperature-corrected physical forecast, IceCast increased CSI by 27.2%, 36.0%, 66.0% and 239.2% across these bins in the testing set and by 9.1%, 13.2%, 44.6% and 67.1% in the validation set. When the temperature-correction error exceeded 5 °C, CSI increased from 0.131 to 0.444 in the testing set and from 0.267 to

0.446 in the validation set. IceCast therefore retained a clear improvement in icing-occurrence prediction even for samples with large residual temperature errors after correction.

We then compared observed-icing MAE across the same temperature-correction-error bins (Fig. 6c,d). IceCast reduced ice-thickness error in all four bins, by 19.6–45.7% in the testing set and 8.7–39.1% in the validation set. For temperature-correction errors greater than 5 °C, MAE decreased from 4.81 to 2.61 mm in the testing set and from 3.49 to 2.13 mm in the validation set. Thus, IceCast improved both icing detection and ice-thickness prediction during observed icing across the full range of temperature-correction errors considered, indicating robustness in icing correction across the evaluated temperature-correction-error bins.

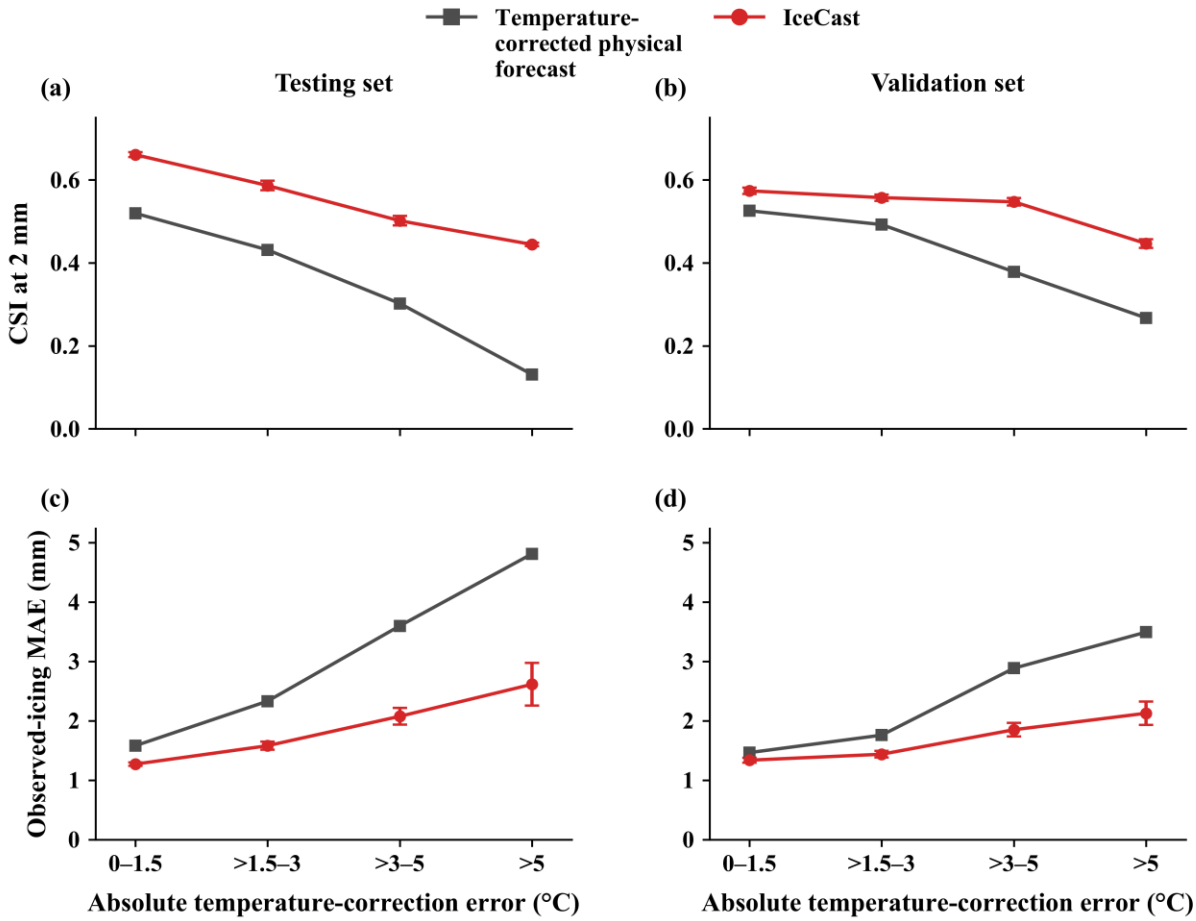

Fig. 6. Prediction performance across temperature-correction errors. (a), (b) 2 mm CSI over all common valid pylon-hours. **(c), (d)** MAE for the subset with observed ice thickness >2 mm. Columns show the testing and validation sets, respectively. Temperature-correction error is defined as the absolute difference between corrected forecast temperature and monitored temperature and is grouped into 0–1.5, >1.5–3, >3–5 and >5 °C. Grey squares denote the deterministic temperature-corrected physical forecast; red circles and error bars denote the mean ± one SD of IceCast across three random seeds. Within each metric and temperature-correction-error bin, both methods are evaluated on the same lead 1–23 hours.

Roles of historical observations and temperature-corrected physical forecasts

We next examined how IceCast uses historical icing observations and future temperature-corrected physical forecasts. Normalized absolute gradients of the learned residual correction with respect to each input time were used to quantify the model response to individual input positions (Fig. 7a). The 24 physical-forecast positions accounted for 62.9% of the total gradient and the 48 historical-observation positions for 37.1%. Physical lead 0 alone contributed 30.8%, leads 1–23 together contributed 32.1%, the final 6 historical hours contributed 22.9%, and the preceding 42 historical hours contributed 14.2%. The final observation at 07:00 and the initial ice thickness of the 08:00 temperature-corrected physical forecast were generally similar, with somewhat larger differences during observed icing (Supplementary Table S6). Because the temperature-corrected physical icing forecast itself is initialized from the current 08:00 ice thickness, a strong model response near forecast initialization is expected. The later physical-forecast leads nonetheless retained substantial gradient contribution, indicating that the model did not respond only to the initial state but continued to use information from the future temperature-corrected physical forecast.

We then used conditional nonlinear optimal perturbation (CNOP) analysis to test whether historical icing changes and future temperature-corrected physical-forecast changes still affected the correction when the initial

icing state was fixed (Fig. 7b,c). The final 07:00 observation and the 08:00 physical lead 0 were held fixed, while the remaining historical observations and later temperature-corrected physical forecasts were perturbed. Perturbing historical observations alone produced a root mean square (RMS) change of 0.625 ± 0.227 mm in the 24 h residual output; perturbing future temperature-corrected physical forecasts alone produced 1.167 ± 0.205 mm; and perturbing both produced 1.937 ± 0.387 mm. For the joint CNOP, 63.9% of the normalized absolute perturbation amplitude occurred in the temperature-corrected physical-forecast channel, and all 10 largest free perturbation positions belonged to physical-forecast leads (Fig. 7b). Thus, even with the initial icing state fixed, both historical icing changes and future temperature-corrected physical-forecast changes substantially altered the IceCast correction.

Finally, we directly tested the predictive contribution of the two sequences using matched sequence replacement (Fig. 7d,e). Donor sequences were preferentially selected from the same cold wave and physical lead-0 ice-thickness group, while preserving the original history endpoint and physical initial value. Replacing the historical observations reduced 2 mm CSI from 0.645 ± 0.005 to 0.599 ± 0.006 and increased observed-icing MAE from 1.384 ± 0.027 to 1.671 ± 0.034 mm. Replacing changes in the temperature-corrected physical forecast over leads 1–23 reduced CSI to 0.581 ± 0.016 and increased MAE to 1.581 ± 0.062 mm. Replacing historical observations had a larger effect on ice-thickness MAE, whereas replacing future temperature-corrected physical-forecast changes had a larger effect on 2 mm CSI. Together, the three analyses show that the current icing state is an important predictive basis, but IceCast also uses preceding icing changes and future temperature-corrected physical-forecast changes, both of which provide predictive information beyond the initial state.

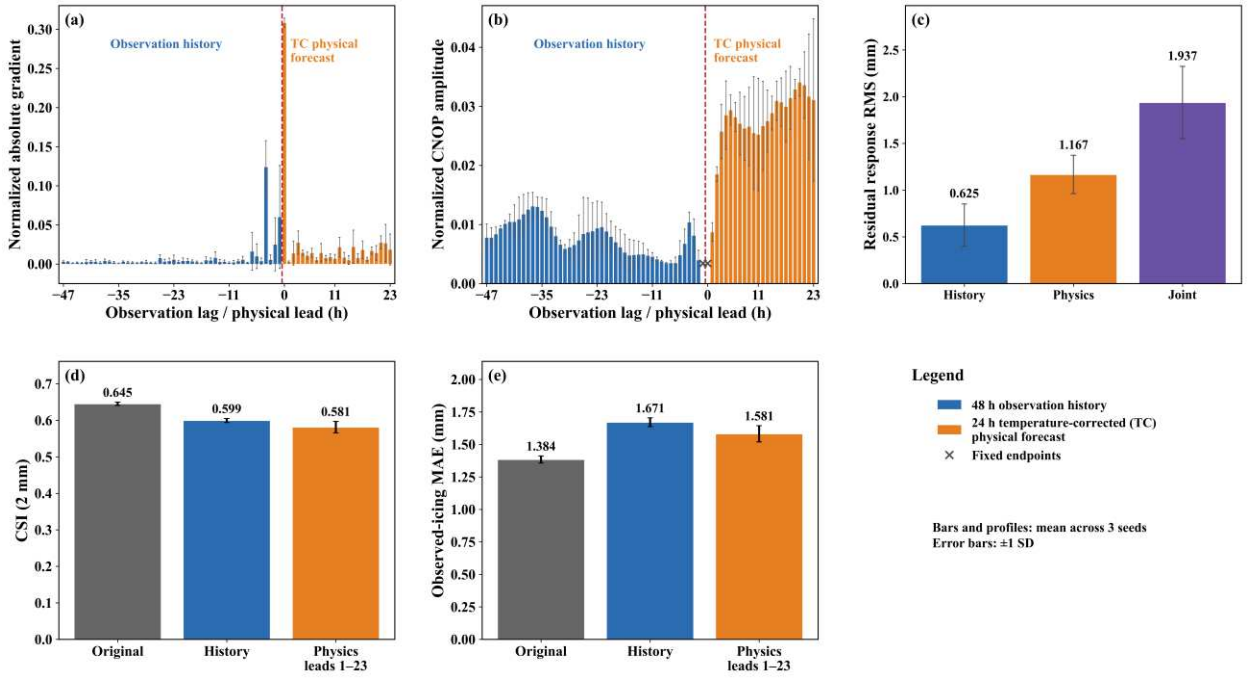

Fig. 7. Input sensitivity and icing-sequence replacement analyses for historical observations and temperature-corrected physical forecasts. (a) Normalized absolute gradient of the learned residual correction with respect to each input time, representing local first-order sensitivity around the current inputs. **(b)** Normalized joint CNOP; crosses denote the fixed final observation and physical lead 0. Historical lags −47 to 0 h are relative to the end of the historical sequence at 07:00, whereas physical leads 0–23 h are relative to the 08:00 initialization. **(c)** RMS changes in residual output when perturbing historical observations, temperature-corrected physical forecasts, or both; this quantity is not a forecast error relative to observations. **(d), (e)** 2 mm CSI and observed-icing MAE after matched icing-sequence replacement. All panels use the same 2,000 internal-validation forecast samples. Values are means across three random seeds; error bars denote one SD.

Discussion

IceCast combines changes in historically observed ice thickness with future ice thickness predicted by the temperature-corrected physical model, thereby incorporating the current icing state, preceding changes and physically based information about future weather forcing. Previous machine-learning studies have mainly used NWP output, meteorological observations or monitoring sequences for icing prediction[15–17]. IceCast instead introduces recent icing changes into machine-learning post-processing of a temperature-corrected physical icing forecast. Our results show that the complementarity of these two information sources is an important source of model improvement. Historical observations constrain the current icing state and its preceding growth, persistence and decay, whereas the temperature-corrected physical forecast provides possible icing changes under future weather conditions. The sensitivity and sequence-replacement analyses show that IceCast uses not only the initial ice thickness but also additional information from historical icing changes and future temperature-corrected

physical-forecast changes. Even when the physical ice thickness itself is biased, a forecast that retains information about icing timing or the direction of growth can still contribute useful information for post-processing. The advantage of IceCast therefore does not arise from simply extrapolating historical observations or independently correcting a temperature-corrected physical forecast, but from jointly using past icing changes and future temperature-corrected physical forecasts to correct subsequent ice thickness.

Moderate-to-heavy icing prediction cannot be optimized solely using mean ice-thickness error across all hours. When no-icing and light-icing samples dominate, predictions concentrated near zero can achieve low average error while missing the relatively small number of moderate-to-heavy icing samples[26]. Studies of extreme-precipitation nowcasting similarly show that event-detection skill and precipitation-amount accuracy are not equivalent[24,27], and bias-aware precipitation methods also address uneven error distributions[25]. The OMTI loss combines focal supervision for icing occurrence[28] with ice-thickness error, adjacent-hour change error and errors across icing-intensity intervals. This increases attention to moderate-to-heavy icing and ice-thickness prediction during observed icing, but relative to MSE training it also produces more false alarms and higher all-sample RMSE. Training objectives therefore require a trade-off among missed moderate-to-heavy icing, false alarms and overall mean error. Persistence remained competitive during persistent icing but cannot anticipate renewed growth, further indicating that icing evaluation should distinguish growth, persistence and decay rather than relying only on overall mean error.

Several limitations remain, and future work could address them in the following ways. First, meteorological forcing remains an important source of icing-forecast error[29]. A 3 km atmospheric model grid cannot fully represent local conditions near individual pylons. The temperature-correction-error analysis also shows that, although IceCast consistently improves the temperature-corrected physical forecast across correction-error bins, its absolute error remains constrained by the quality of upstream meteorological inputs. The pylon coverage of temperature-corrected physical icing forecasts was 99.8%, 81.2%, 61.4% and 51.0% across the four winters, where coverage is defined as the proportion of all forecast pylons for which a temperature-corrected physical forecast was available. Expanding terminal coverage for temperature correction and strengthening local meteorological verification at icing-prone pylons therefore remain important. Machine-learning weather prediction models[30–32] and recent work downscaling forecast temperature to 1 km[33] may provide alternative meteorological inputs, although their value for transmission-line icing prediction requires direct evaluation. Second, observations above 30 mm remain concentrated in a small number of pylons and cold-wave events; we will expand independent validation of moderate-to-heavy icing processes to evaluate model stability at higher icing intensities. Overall, this study shows that jointly post-processing historical observed ice-thickness sequences and temperature-corrected physical icing forecasts performs better than using either information source alone and provides a feasible hybrid framework for transmission-line icing prediction under moderate-to-heavy icing and complex terrain.

Methods

Dataset construction

This study used icing and temperature monitoring data and meteorological forecasts. Icing and temperature monitoring data were obtained from the icing monitoring system of the Electric Power Research Institute of China Southern Power Grid. The study region covers Yunnan, Guizhou, Guangxi and Guangdong and includes interprovincial transmission corridors (Fig. 1a,b). The monitoring system records tension, inclination angle, images and meteorological variables every 10 min. Equivalent ice thickness is retrieved from mechanical-load information, including tension and inclination angle, and was used as the prediction target in this study[3,34–36]. Both icing and temperature monitoring data underwent time standardization, multi-source consistency checks, anomaly identification and correction, and were then arithmetically averaged to hourly values for modeling. After quality control, ice thickness from different voltage classes was standardized to equivalent ice thickness at a reference height of 10 m above ground[37],

$$x_{10} = x_h \left(\frac{10\ \mathrm{m}}{h}\right)^{0.22}, \tag{1}$$

where x_h is the equivalent ice thickness at the representative monitoring height h, and x_{10} is the equivalent ice thickness at the 10 m reference height. Here, x_h and x_{10} are in millimetres, and h is in metres. Representative monitoring heights of 10, 15, 20, 30, 40 and 50 m were used for 10, 35, 110, 220, 500 and 800 kV lines, respectively[36]. Monitored ice thickness and the temperature-corrected physical icing forecast were both standardized to the 10 m reference height. After quality control, 2,082 pylons were retained for analysis.

Meteorological forecasts were obtained from the 3 km NWP product of the China Meteorological Administration (CMA) Public Meteorological Service Centre[38], including hourly forecasts of temperature, wind speed, RH, pressure and cloud water mixing ratio at 10 m above ground, together with surface precipitation and precipitation type. Forecast LWC was calculated from the cloud water mixing ratio as

$$w = 10^3 \rho q_{\text{cloud}}, \tag{2}$$

where w is in $\mathrm{g\ m^{-3}}$, q_{cloud} is the forecast cloud water mixing ratio in $\mathrm{kg\ kg^{-1}}$, and ρ is air density in $\mathrm{kg\ m^{-3}}$. Corrected forecast temperature was obtained using relationships developed from pylon temperature observations and forecast temperature at nearby national meteorological stations, based on approximately 2,288 terminals during the winters of 2020–2022[36,39]. Only pylons for which a temperature-corrected physical icing forecast was available were retained.

After the meteorological inputs and initial icing state were defined, hourly conditions were used to classify freezing-rain icing, rime icing, melting or unchanged ice thickness. The thresholds followed the previously calibrated regional operational parameterization[36,40–42]. Freezing-rain icing required precipitation of 0.01–3 $\mathrm{mm\ h^{-1}}$, corrected temperature below −1 °C, and surface wet-bulb temperature between −6.6 and 0 °C. Rime icing required precipitation below 0.1 $\mathrm{mm\ h^{-1}}$, RH above 80%, wind speed no greater than 5 $\mathrm{m\ s^{-1}}$, corrected temperature between −10 and 1 °C, and forecast LWC above 0.025 $\mathrm{g\ m^{-3}}$. When icing conditions were not met, melting or unchanged ice thickness was determined from temperature and precipitation conditions.

Rime accretion was calculated using the Makkonen model[5], whose classical mass-growth form is

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \alpha_1 \alpha_2 \alpha_3\, wVA, \tag{3}$$

where M is ice mass; t is time; α_1, α_2 and α_3 are the collision, sticking and freezing efficiencies, respectively; w is forecast LWC ($\mathrm{g\ m^{-3}}$); V is wind speed ($\mathrm{m\ s^{-1}}$); and A is the projected windward area of the conductor ($\mathrm{m^2}$). The sticking efficiency was set to 1. Collision efficiency was calculated from droplet collection, and freezing efficiency from the surface energy balance.

Freezing-rain accretion was calculated using the Jones model[6], which combines vertical precipitation flux and wind-driven horizontal water flux to estimate equivalent glaze-ice growth,

$$R_{\text{eq}} = \sum_i \frac{1}{\pi \rho_i} [(P_i \rho_w)^2 + (3.6 V_i w_i)^2]^{1/2}, \tag{4}$$

where R_{eq} is equivalent glaze-ice growth (mm), P_i is hourly precipitation rate ($\mathrm{mm\ h^{-1}}$), ρ_w and ρ_i are the densities of water and ice ($\mathrm{g\ cm^{-3}}$), respectively, V_i is wind speed ($\mathrm{m\ s^{-1}}$), and w_i is forecast LWC at the same hour ($\mathrm{g\ m^{-3}}$). Ice density was set to $0.9\ \mathrm{g\ cm^{-3}}$. Ice-mass increments from the Makkonen model were likewise converted to radial ice thickness using the ice density and a bare-conductor radius of 14.5 mm. Each daily temperature-corrected physical icing forecast was initialized with the observed pylon ice thickness available at 08:00.

The IceCast dataset was constructed from these products. Each forecast sample corresponds to one 24 h forecast initialized at 08:00 for one pylon and contains 48 h of historical observations, a 24 h temperature-corrected physical icing forecast and a 24 h observed target. We define

$$\mathbf{x}^{\text{hist}} = [y_{-48}, \ldots, y_{-1}],\ \mathbf{x}^{\text{phys}} = [p_0, \ldots, p_{23}],\ \mathbf{y} = [y_0, \ldots, y_{23}], \tag{5}$$

where the historical observations end at 07:00 and the temperature-corrected physical forecast and target span 08:00 to 07:00 the following day. The 24 h target $\mathbf{y}$ is used only for supervised training and evaluation and is not included in the model inputs. Physical lead 0 is initialized from the observed 08:00 ice thickness and therefore satisfies $p_0 = y_0$ to storage precision; this value is not included in the 48 h historical observation sequence. The forecasting task is

$$[\mathbf{x}^{\text{hist}} \parallel \mathbf{x}^{\text{phys}}] \in \mathbb{R}^{72} \rightarrow \mathbf{y} \in \mathbb{R}^{24}, \tag{6}$$

where $\parallel$ denotes concatenation along the time dimension. All observed and temperature-corrected physical ice thicknesses were mapped to a common dimensionless model space using the fixed scale $s_y = 10$ mm (y and s_y in millimetres; z dimensionless),

$$z = \frac{y}{s_y},\ y = s_y z. \tag{7}$$

The training set comprised 10 cold waves during 2022–2025, the testing set comprised three additional independent cold waves from the same period, and the validation set comprised six cold waves during winter 2025–2026 (Fig. 1c–f). Training samples were jointly stratified by cold-wave event and maximum 24 h target ice thickness into seven classes: ≤2, (2,5], (5,10], (10,20], (20,30], (30,40] and >40 mm. Within each stratum, 90% of

samples were assigned to the training subset and 10% to the internal validation subset. The internal validation subset was used for training-design decisions and checkpoint selection. The testing and validation sets were not used for model training. All models used the same data partitions.

IceCast architecture

IceCast uses a Transformer encoder–decoder architecture with $d_{\text{model}} = 64$, four attention heads, three encoder layers, three decoder layers and a 128-dimensional feed-forward network[43] (Fig. 8). Each layer uses rectified linear unit (ReLU) activation, dropout of 0.1 and layer normalization after residual blocks. The 48 h historical observations and 24 h temperature-corrected physical icing forecast are concatenated along the time dimension to form 72 input positions, which are linearly projected, combined with learnable positional encodings, and passed to the Transformer encoder. The decoder receives 24 zero-valued queries with positional encodings, reads the encoded historical and physical-forecast information through cross-attention, and outputs all 24 forecast hours in parallel.

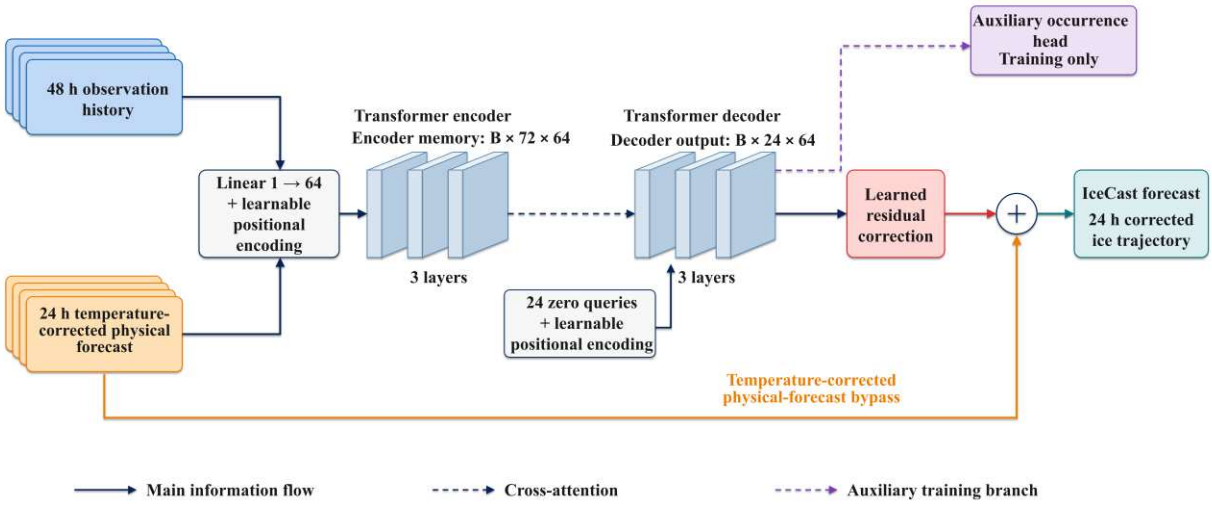

Fig. 8. IceCast architecture. The 48 h historical observations and 24 h temperature-corrected physical icing forecast are jointly embedded and passed through a three-layer Transformer encoder. The encoder output forms the decoder memory, and 24 queries correspond to the 24 forecast leads. The decoder is connected to an auxiliary icing-occurrence classification head and a residual regression head. The classification head is used only during training, whereas final ice-thickness predictions are produced by the residual regression head. The temperature-corrected physical forecast is added through a bypass connection to the learned residuals to form the corrected 24 h ice-thickness forecast.

IceCast uses residual correction. In model space, the prediction is

$$\hat{\mathbf{z}} = \mathbf{z}^{\text{phys}} + \Delta_\theta(\mathbf{z}^{\text{hist}}, \mathbf{z}^{\text{phys}}), \tag{8}$$

where $\mathbf{z}^{\text{hist}}$ and $\mathbf{z}^{\text{phys}}$ are the normalized historical observations and temperature-corrected physical forecast, respectively, and Δ_θ is the learned residual correction. The weights and bias of the residual regression head are initialized to zero, so the model output initially equals the temperature-corrected physical forecast. After conversion back to millimetres, negative predicted ice thicknesses are clipped to zero. The model contains 256,002 trainable parameters.

IceCast is trained with the OMTI loss, which jointly considers icing occurrence, ice-thickness magnitude, adjacent-hour changes and errors across icing-intensity intervals. The four components are defined as follows.

The occurrence component uses focal supervision to emphasize more difficult classification samples. Let $q_{b,t}$ be the binary target at the 2 mm threshold and $\pi_{b,t}$ the predicted icing probability,

$$\mathcal{L}_{\text{occ}} = \frac{1}{24B}\sum_{b=1}^{B}\sum_{t=0}^{23}\alpha_{b,t}\left(1 - q_{b,t}\pi_{b,t} - (1-q_{b,t})(1-\pi_{b,t})\right)^2 c_{b,t}, \tag{9}$$

where $c_{b,t}$ is the binary cross-entropy and $\alpha_{b,t}$ is a class weight. The class weights were 0.75 for icing and 0.25 for non-icing.

The magnitude component constrains hourly ice-thickness errors. Let $e_{b,t} = \hat{z}_{b,t} - z_{b,t}$ and let $\mathcal{V}$ denote the set of valid targets,

$$\mathcal{L}_{\text{mag}} = \frac{1}{N_{\mathcal{V}}} \sum_{(b,t)\in\mathcal{V}} h_{0.25}\,(e_{b,t}), \tag{10}$$

where h_β is the SmoothL1 function,

$$h_\beta(e) = \begin{cases} \frac{e^2}{2\beta}, & |e| < \beta, \\ |e| - \frac{\beta}{2}, & |e| \geq \beta. \end{cases} \tag{11}$$

The adjacent-hour change component constrains differences between predicted and observed hourly ice-thickness changes. Define

$$d_{b,t} = (\hat{z}_{b,t} - \hat{z}_{b,t-1}) - (z_{b,t} - z_{b,t-1}), \tag{12}$$

then

$$\mathcal{L}_{\text{traj}} = \frac{1}{\max(1, N_{\mathcal{D}})} \sum_{(b,t)\in\mathcal{D}} h_{0.10}\,(d_{b,t}), \tag{13}$$

where $\mathcal{D}$ is the set of valid adjacent time pairs.

The intensity component divides target ice thickness into seven intervals, [0,5], (5,10], (10,15], (15,20], (20,30], (30,40] and (40,∞) mm, and normalizes errors by the tolerance τ_k assigned to each interval,

$$\mathcal{L}_{\text{int}} = \frac{1}{N_K} \sum_{k\in\mathcal{K}_B} \frac{1}{n_k} \sum_{(b,t)\in\mathcal{I}_k} h_1\left(\frac{\hat{z}_{b,t} - z_{b,t}}{\tau_k}\right). \tag{14}$$

The dimensionless tolerances were $\tau_k = (0.10, 0.15, 0.25, 0.35, 0.50, 0.70, 1.00)$, obtained by dividing the interval tolerances in millimetres by $s_y = 10\ \text{mm}$.

The final OMTI loss is

$$\mathcal{L}_{\text{OMTI}} = \mathcal{L}_{\text{occ}} + \mathcal{L}_{\text{mag}} + 2\mathcal{L}_{\text{traj}} + 0.05\mathcal{L}_{\text{int}}. \tag{15}$$

Target changes with hourly decreases greater than 5 mm were excluded from the change-related loss because they may represent de-icing operations or abrupt anomalous changes. Component weights were determined from the training set and internal validation subset and were fixed for all IceCast experiments.

Model training and comparison experiments

Four categories of model-training and comparison experiments were conducted. **Input-information comparisons** included the History-only Transformer, Physics-only Transformer, Shuffled-physics Transformer and Aligned-fusion Transformer. **Output-formulation and loss comparisons** included the Aligned-fusion Transformer, Residual-fusion Transformer and IceCast. **Architecture comparisons** used Transformer, LSTM-attention[44] and DLinear[45] under History-only and Aligned-fusion configurations. **Baseline comparisons** included the temperature-corrected physical forecast, Persistence, model output statistics (MOS) baselines (Physics-only linear MOS and Blended linear MOS), XGBoost and Light Gradient Boosting Machine (LightGBM). Inputs, prediction formulations, training losses, random seeds and parameter settings for all models are given in Supplementary Table S7.

Evaluation metrics

Model performance was evaluated using classification metrics at multiple ice-thickness thresholds, direction-sensitive interval CSI and continuous error metrics. For thresholds

$$u \in \{2, 5, 10, 20, 30\}\ \text{mm}, \tag{16}$$

hits H, false alarms F, misses M and correct negatives C were defined according to whether observed and predicted ice thickness exceeded threshold u. The 2 mm threshold was used as the primary icing-occurrence threshold[13,36]. CSI, FAR, MAR and bias are

$$\text{CSI} = \frac{H}{H+F+M},\ \text{FAR} = \frac{F}{H+F},\ \text{MAR} = \frac{M}{H+M},\ \text{Bias} = \frac{H+F}{H+M}. \tag{17}$$

To evaluate ice-thickness prediction across icing intensities, we used a direction-sensitive interval CSI[36]. Observed ice thickness was divided into [0,5], (5,10], (10,15], (15,20], (20,30], (30,40] and (40,∞) mm, with corresponding allowable errors of 1.0, 1.5, 2.5, 3.5, 5.0, 7.0 and 10.0 mm. For interval k, a hit was counted when both observation and prediction exceeded 2 mm and the prediction error fell within the allowable range. An

underprediction exceeding the allowable error, or missed icing, was counted as a miss; an overprediction exceeding the allowable error, or a false icing prediction, was counted as a false alarm. Interval CSI is

$$\mathrm{CSI}_k = \frac{H_k}{H_k + F_k + M_k}, \tag{18}$$

where H_k, F_k and M_k are the numbers of hits, false alarms and misses for the kth observed ice-thickness interval.

Continuous ice-thickness errors were evaluated using RMSE and MAE. For N valid prediction–observation pairs, with p_i and o_i denoting predicted and observed ice thickness, both in millimetres,

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2}, \ \mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|p_i - o_i|. \tag{19}$$

All-sample RMSE and MAE use all valid pylon-hours. Observed-icing RMSE and MAE use only hours with observed ice thickness >2 mm and therefore include missed icing. Hit-only RMSE and MAE use only hours for which both observed and predicted ice thickness exceed the corresponding threshold. Neural-network results are reported as the mean and one SD across three random seeds.

Model interpretability analyses

Gradient sensitivity analysis was used to evaluate the response of IceCast to different input times[46–48]. A fixed set of 2,000 forecast samples was selected from the internal validation subset. For each of the three IceCast models trained with different random seeds, we calculated the absolute gradient of the sum of the 24 h residual corrections with respect to each of the 72 input positions. For input position j,

$$g_j = \left|\frac{\partial}{\partial x_j}\sum_{t=0}^{23}\Delta_{\theta,t}(\mathbf{x})\right|, \ \tilde{g}_j = \frac{g_j}{\sum_{k=1}^{72} g_k}, \tag{20}$$

where $\Delta_{\theta,t}$ is the learned residual at forecast lead t and $\tilde{g}_j$ is the normalized absolute gradient. The analysis targeted the learned residual correction rather than the final ice-thickness forecast, avoiding direct sensitivity to the additive bypass of the temperature-corrected physical forecast. Normalized gradients were averaged across the 2,000 samples and three models, with SDs also calculated, to compare the relative contributions of individual times in the 48 h historical observations and 24 h temperature-corrected physical forecast (Fig. 7a).

CNOP analysis was used to test how much finite-amplitude perturbations to historical icing changes and future temperature-corrected physical-forecast changes could alter the IceCast correction[48]. Let the 72-dimensional input in millimetres be $\mathbf{x}_{\mathrm{mm}}$ and define the residual response as $R_\theta(\mathbf{x}_{\mathrm{mm}}) = s_y\Delta_\theta(\mathbf{x}_{\mathrm{mm}}/s_y)$. The optimal perturbation under condition c is

$$\boldsymbol{\delta}_c^{\mathrm{opt}} = \arg\max_{\boldsymbol{\delta}\in\mathcal{C}_c}\frac{\|R_\theta(\mathbf{x}_{\mathrm{mm}}+\boldsymbol{\delta}) - R_\theta(\mathbf{x}_{\mathrm{mm}})\|_2}{\sqrt{24}}. \tag{21}$$

Here, c denotes perturbations applied only to historical observations, only to the temperature-corrected physical forecast, or to both. The final historical-observation time and physical lead 0 were fixed to preserve the current icing state at initialization. For all other perturbable positions, RMS perturbation amplitude was constrained to no more than 1 mm, absolute perturbation at any single position to no more than 2 mm, and adjacent-time perturbation differences to no more than 0.5 mm. Perturbed ice thickness was kept nonnegative. CNOP was approximated using projected gradient ascent for 60 iterations from two random initializations, with a step size of 0.25 mm and one three-point smoothing pass per update. The analysis used the same 2,000 samples and three IceCast models as the gradient analysis. RMS changes in the 24 h residual output under the different perturbation conditions were used to compare the influence of historical observations and future temperature-corrected physical forecasts, and normalized absolute perturbation amplitudes from the joint CNOP were used to assess relative sensitivity across input times (Fig. 7b,c).

Data availability

The training, testing and validation datasets used in this study are available from the corresponding author upon reasonable request. The CMA NWP product is available through its official data-service catalogue[38]. Administrative-boundary data used in Fig. 4 are publicly available from Tianditu (https://cloudcenter.tianditu.gov.cn/administrativeDivision).

Code availability

The IceCast model-training and inference code is available from the corresponding author upon reasonable request. Third-party software used in this study is available from the respective official repositories, including PyTorch (https://github.com/pytorch/pytorch), Time-Series-Library, which includes DLinear

(https://github.com/thuml/Time-Series-Library), XGBoost (https://github.com/dmlc/xgboost), and LightGBM (https://github.com/microsoft/LightGBM).

Author contributions statement

S. Z. conceived the study, designed the research, performed the data analysis and visualization, and wrote the original draft. H. Z., B. G. and Q. G. provided technical guidance. Z. G. provided computational resources and contributed to research supervision. Y. S., H. S., Z. G., X. X. and Z. L. reviewed and edited the manuscript. All authors discussed the results and reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Acknowledgements

This study was funded by the Electric Power Research Institute, China Southern Power Grid Co., Ltd., through the project "Research on initial-condition optimization and deep learning prediction techniques for transmission-line icing" (Subproject 1: "Research on initial-field modeling techniques for icing prediction based on multi-source data"; Project No. ZBKJXM20240154). The funder contributed by providing the icing-monitoring data. The authors thank the China Meteorological Administration for providing the NWP products.

References

1. Farzaneh, M. & Chisholm, W. A. *Insulators for Icing and Polluted Environments* (IEEE Press/Wiley, 2009).
2. Wang, L. et al. The Causes and Forecasting of Icing Events on Power Transmission Lines in Southern China: A Review and Perspective. *Atmosphere* **14,** 1815 (2023).
3. Dong, B., Jiang, X. & Yin, F. Development and prospect of monitoring and prevention methods of icing disaster in China power grid. *IET Gener. Transm. Distrib.* **16,** 4480–4493 (2022).
4. Hou, H. et al. Modelling icing growth on overhead transmission lines: Current advances and future directions. *Energy Convers. Econ.* **5,** 343–357 (2024).
5. Makkonen, L. Models for the growth of rime, glaze, icicles and wet snow on structures. *Philos. Trans. R. Soc. A* **358,** 2913–2939 (2000).
6. Jones, K. F. A simple model for freezing rain ice loads. *Atmos. Res.* **46,** 87–97 (1998).
7. Kringlebotn Nygaard, B. E., Kristjánsson, J. E. & Makkonen, L. Prediction of In-Cloud Icing Conditions at Ground Level Using the WRF Model. *J. Appl. Meteorol. Climatol.* **50,** 2445–2459 (2011).
8. Engdahl, B. J. K., Nygaard, B. E. K., Losnedal, V., Thompson, G. & Bengtsson, L. Effects of the ICE-T microphysics scheme in HARMONIE-AROME on estimated ice loads on transmission lines. *Cold Reg. Sci. Technol.* **179,** 103139 (2020).
9. Podolskiy, E. A., Nygaard, B. E. K., Nishimura, K., Makkonen, L. & Lozowski, E. P. Study of unusual atmospheric icing at Mount Zao, Japan, using the Weather Research and Forecasting model. *J. Geophys. Res. Atmos.* **117,** D12106 (2012).
10. DeGaetano, A. T., Belcher, B. N. & Spier, P. L. Short-Term Ice Accretion Forecasts for Electric Utilities Using the Weather Research and Forecasting Model and a Modified Precipitation-Type Algorithm. *Weather Forecast.* **23,** 838–853 (2008).
11. Luo, J. et al. Performance Evaluation of the Makkonen Model Driven by the Output of the WRF Model in Simulating Power Line Icing Thickness in the Mountainous Regions of Southwest Sichuan, China. *J. Geophys. Res. Atmos.* **131,** e2026JD046315 (2026).
12. Punde, P., Birkelund, Y., Virk, M. S., Han, X. & Eidhammer, T. Kilometer-scale simulation of atmospheric in-cloud ground icing over complex terrain in Norwegian Arctic. *Cold Reg. Sci. Technol.* **245,** 104851 (2026).
13. Zhou, S. et al. Ensemble Learning for Spatial Modeling of Icing Fields from Multi-Source Remote Sensing Data. *Remote Sens.* **17,** 2155 (2025).
14. Zhang, P. et al. Investigation of supercooled water droplet sticking efficiency during power transmission line icing using digital holography. *Atmos. Meas. Tech.* **19,** 4491–4503 (2026).
15. Chen, Y., Li, P., Ren, W., Shen, X. & Cao, M. Field data–driven online prediction model for icing load on power transmission lines. *Meas. Control* **53,** 126–140 (2020).
16. Zhou, R. et al. Reliable monitoring and prediction method for transmission lines based on FBG and LSTM. *Adv. Eng. Inform.* **62,** 102603 (2024).
17. Zarnani, A. et al. Learning to predict ice accretion on electric power lines. *Eng. Appl. Artif. Intell.* **25,** 609–617 (2012).
18. Wang, G. et al. Prediction model for transmission line icing based on data assimilation and model integration. *Front. Environ. Sci.* **12,** 1403426 (2024).
19. Zhou, F. et al. Research on wire icing simulation technology considering the weights of meteorological elements. *Front. Energy Res.* **12,** 1346480 (2024).
20. Vannitsem, S. et al. Statistical Postprocessing for Weather Forecasts: Review, Challenges, and Avenues in a Big Data World. *Bull. Am. Meteorol. Soc.* **102,** E681–E699 (2021).
21. Zhou, S., Gao, C. Y., Duan, Z., Xi, X. & Li, Y. A robust error correction method for numerical weather prediction wind speed based on Bayesian optimization, variational mode decomposition, principal component analysis, and random forest: VMD-PCA-RF (version 1.0.0). *Geosci. Model Dev.* **16,** 6247–6266 (2023).
22. Noh, G.-H. & Ahn, K.-H. Deep learning with spatio-temporal conditioning improves global subseasonal precipitation forecasts. *npj Clim. Atmos. Sci.* **9,** 156 (2026).
23. Chen, L. et al. Environment-conditioned deep learning for severe convection nowcasting. *npj Clim. Atmos. Sci.* (in press, 2026).
24. Sun, D. et al. Deep Learning Improves GFS Wintertime Precipitation Forecast Over Southeastern China. *Geophys. Res. Lett.* **50,** e2023GL104406 (2023).

25. Tang, T. et al. Bias-targeted deep learning enhances short-range heavy rainfall forecasts. *npj Clim. Atmos. Sci.* **9,** 78 (2026).
26. de Burgh-Day, C. O. & Leeuwenburg, T. Machine learning for numerical weather and climate modelling: a review. *Geosci. Model Dev.* **16,** 6433–6477 (2023).
27. Das, P. et al. Hybrid physics-AI outperforms numerical weather prediction for extreme precipitation nowcasting. *npj Clim. Atmos. Sci.* **7,** 282 (2024).
28. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollar, P. Focal Loss for Dense Object Detection. In *2017 IEEE International Conference on Computer Vision (ICCV)* 2999–3007 (2017).
29. Hosek, J., Musilek, P., Lozowski, E. & Pytlak, P. Forecasting severe ice storms using numerical weather prediction: the March 2010 Newfoundland event. *Nat. Hazards Earth Syst. Sci.* **11,** 587–595 (2011).
30. Chen, L. et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Clim. Atmos. Sci.* **6,** 190 (2023).
31. Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382,** 1416–1421 (2023).
32. Bi, K. et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619,** 533–538 (2023).
33. Park, H., Park, S., Kang, D. & Kim, J.-H. A super-resolution framework for downscaling machine learning weather prediction toward 1-km air temperature. *npj Clim. Atmos. Sci.* **9,** 56 (2026).
34. Jiang, X. et al. Predictive Model for Equivalent Ice Thickness Load on Overhead Transmission Lines Based on Measured Insulator String Deviations. *IEEE Trans. Power Deliv.* **29,** 1659–1665 (2014).
35. Dong, B., Jiang, X. & Xiang, Z. Calculation model and experimental verification of equivalent ice thickness on overhead lines with tangent tower considering ice and wind loads. *Cold Reg. Sci. Technol.* **200,** 103588 (2022).
36. Zhou, S. *Meteorological correction and ice-accretion model optimization through multi-source data fusion in the complex terrain of southern China*. PhD thesis, Nanjing University of Information Science and Technology (2025) (in Chinese).
37. Fan, Y., Chen, B., Wu, W. & Lin, H. Method for Optimal Selection of Power and Capacity of a Mobile Battery Energy Storage DC De-icing System. Chinese patent CN105406433B (2017) (in Chinese).
38. China Meteorological Administration. CMA-MESO 3-km numerical weather-prediction product catalogue. China Meteorological Data Service Centre, https://data.cma.cn/data/detail/dataCode/F.0003.0008.S002.html (accessed 19 August 2026).
39. Zhang, H. et al. Tower air-temperature correction method and device based on weather-station forecasts, and computer equipment. Chinese patent application no. 202411598294.2 (2024).
40. Deng, F., Kang, L., Jiang, Y., Chu, J. & Liu, Y. An hourly standard ice thickness model using conventional meteorological data with its validation. *J. Appl. Meteorol. Sci.* **28,** 142–156 (2017) (in Chinese).
41. Lu, Z. *Optimization and Simulation of Conductor-Icing Parameterization Schemes*. Master's thesis, Nanjing University of Information Science and Technology (2018) (in Chinese).
42. Ramer, J. An empirical technique for diagnosing precipitation type from model output. In *Preprints, Fifth International Conference on Aviation Weather Systems* 227–230 (American Meteorological Society, 1993).
43. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30,** 5998–6008 (2017).
44. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **9,** 1735–1780 (1997).
45. Zeng, A., Chen, M., Zhang, L. & Xu, Q. Are Transformers Effective for Time Series Forecasting? *Proc. AAAI Conf. Artif. Intell.* **37,** 11121–11128 (2023).
46. Simonyan, K., Vedaldi, A. & Zisserman, A. Deep inside convolutional networks: visualising image classification models and saliency maps. Preprint at arXiv:1312.6034v2 (2014).
47. Baño-Medina, J. et al. Are AI weather models learning atmospheric physics? A sensitivity analysis of cyclone Xynthia. *npj Clim. Atmos. Sci.* **8,** 92 (2025).
48. Guo, A., Mu, M., Qin, B., Pu, J. & Li, H. A nonlinear explainable method for data-driven forecasting models in Earth system science. *npj Clim. Atmos. Sci.* **9,** 190 (2026).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- IceCastsupply.docx