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.

Dual-Stream Downscaling and Flow-Constrained Correction of Meteorological Fields over Complex Terrain

Ao Shen

shenao@nudt.edu.cn

College of Advanced Interdisciplinary Studies, National University of Defense Technology

Ying Yuan

College of Advanced Interdisciplinary Studies, National University of Defense Technology

Wei Zhong

Hongrang He

College of Advanced Interdisciplinary Studies, National University of Defense Technology

Article

Keywords: meteorological-field downscaling, dual-stream fusion, residual correction, flow- constrained correction

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11134908/v1

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

Additional Declarations: There is **NO** Competing Interest.

Dual-Stream Downscaling and Flow-Constrained Correction of Meteorological Fields over Complex Terrain

Abstract

High-resolution near-surface meteorological fields are essential, yet conventional numerical simulation is computationally expensive and data-driven methods often lack physical consistency. Here, a lightweight framework for meteorological-field downscaling and observation-driven correction is presented as a cost-effective alternative. In the downscaling stage, multivariate synoptic-scale information is fused with high-resolution terrain by a dual-stream generative adversarial network, and the fields are refined in a reconstruction-driven cascade. In the correction stage, a simplified UNet learns additive corrections supervised directly by station observations, and a flow-constrained regularizer enforces physical consistency. In independent tests, the downscaling stage reduces the 0.25° four-variable error by 54.0–73.5% relative to bicubic interpolation, and the full-field correction reduces the station MAE of T2m, V10, and U10 by 17.81%, 27.59%, and 21.06%, respectively. The full training-to-inference workflow can be completed within a few days on an ordinary personal computer, making it practical for operational fine-scale meteorological services.

Keywords: meteorological-field downscaling; dual-stream fusion; residual correction; flow-constrained correction

Introduction

High-spatial-resolution meteorological data are an important basis for fine-grained weather forecasting, meteorological disaster prevention, and wind-energy resource assessment[1,2,3]. Owing to observational costs and the computational complexity of physical models, the raw fields produced by numerical weather prediction (NWP) and general circulation models are often too coarse[4]. Statistical downscaling achieves scale transfer through empirical relationships between large-scale predictors and local variables, but its linear-regression assumptions are difficult to justify under strongly nonlinear atmospheric conditions, and the determination of weather-type numbers and the treatment of transitions between types remain highly subjective[5,6,7]. Dynamical downscaling is built on regional climate models, has a well-defined physical basis, and produces physically self-consistent fine-scale fields that satisfy the dynamical and thermodynamic constraints among variables, but its computational cost is prohibitive: the CPU time of a typical high-resolution regional run can reach tens to hundreds of times that of the driving coarse-resolution global simulation[8,9,10,11].

Rapid progress in artificial intelligence (AI) has motivated new approaches to meteorological-field downscaling[12,13,14,15,16]. Field downscaling has increasingly been treated as a special class of image super-resolution task[17,18,19,20], in which the strong feature-extraction and nonlinear-mapping capabilities of deep learning are exploited to convert low-resolution grids into high-resolution fields. Vandal et al. first applied a super-resolution convolutional network to climate-field downscaling, demonstrating the feasibility of deep learning for statistical downscaling[21,22]. Various improved architectures have since been proposed: joint spatiotemporal convolutions were introduced to capture the temporal evolution of meteorological fields[23,24,25], and several meteorological variables were downscaled simultaneously within multi-task learning frameworks, using the physical coupling between variables to strengthen reconstruction consistency[26,27,28]. To address the over-smoothing issue, generative adversarial networks (GANs) were introduced into super-resolution[29,30,31,32]; in climate-model downscaling, Groenke et al. showed that GAN-based reconstruction preserves spatial statistics better than deterministically trained

models[33,34], and Harris et al. used conditional GANs to generate high-resolution precipitation fields, indicating the potential of generative methods for reconstructing extreme precipitation events[35,36]. Diffusion models have more recently attracted wide attention as a new generative paradigm[37,38,39] and have been applied to meteorological-field downscaling, with demonstrated advantages in statistical uncertainty quantification[40].

Despite this progress, the practical application of generative models to meteorological downscaling still faces three challenges. First, existing generative downscaling models usually take a single variable, or a small number of variables at the same resolution, as input. A coarse-resolution field cannot uniquely determine the configuration of the missing high-frequency components, so the model tends to regress toward the conditional mean; local gradients and extreme amplitudes are attenuated, and over-smoothing arises[41]. Auxiliary data such as terrain have been introduced to provide additional prior information[42,43,44], but simply concatenating low-resolution features does not fully capture the constraints imposed by high-resolution static geographic information on subgrid spatial patterns. Second, even when downscaled fields display plausible spatial texture, systematic biases may persist at station locations. Station-wise statistical correction can reduce single-point errors but does not naturally translate into a spatially continuous correction field[45,46], and correction driven solely by station observations tends to overfit sparse station data, introducing non-physical perturbations that undermine the dynamical coherence and physical consistency of the overall field[47,48]. Third, the term "lightweight" in the existing literature typically refers only to the inference stage[49]; mainstream Transformer architectures and diffusion models exhibit strong feature-extraction capabilities, but their complex network structures and large parameter counts make training highly dependent on high-performance computing, which limits the efficiency of model iteration and the flexibility of deployment in production environments[50].

To address these problems, a meteorological-field downscaling and correction framework is proposed in which the downscaling and correction modules are trained in stages but chained together during inference. The downscaling module is a lightweight GAN-based cascaded network with dual-stream fusion of dynamic variables and static terrain. The dynamic branch extracts shared synoptic-scale features from mean sea-level pressure (MSLP), 2m temperature (T2m), and the 10m wind components (V10 and U10), and the static branch extracts terrain structure from DEM, slope, and aspect on the target grid. Each generator performs supervised 4× downscaling of native-resolution inputs, after which the full reconstructed multivariate state is iteratively refined to produce successively finer grids; given terrain data at the corresponding resolution, the cascade design can be extended to still higher resolutions. The correction module is an observation-driven full-field residual network that takes the early-fused downscaled fields and normalized DEM as input and jointly reconstructs the signed residuals of T2m, V10, and U10. Training follows a two-stage strategy that combines uniformly sampled supervised cores with complete-field fine-tuning. Station estimates are obtained from the gridded field through differentiable spatial interpolation, so that the observational the error of the corrected field at station locations is minimized directly, and a second directional derivative of the residual along the local wind direction is introduced as a regularizer, so that bias correction and the physical-consistency constraint are enforced simultaneously. At inference, a single forward pass outputs a continuous residual field over the whole region, and station observations are used only in training, never as inference inputs.

The main contributions of this paper are as follows:

1. A reconstruction-driven dual-stream multivariate cascaded downscaling scheme is proposed. The second stage receives the complete 0.25° reconstructed meteorological fields together with 0.0625° terrain and does not rely on any intermediate meteorological truth. The same compact generators

are reused across scales, achieving 16× spatial refinement with about 8.44 million parameters, and the 0.25° four-variable RMSE is reduced by 54.0%–73.5% relative to bicubic interpolation. The feasibility of extending the cascade toward finer resolutions is also discussed.

2. An observation-driven, flow-constrained residual correction scheme with a two-stage random-context to complete-field strategy is proposed. Uniform random cropping allows the model to cover different terrain and weather regimes, supervision on the core region avoids contamination from crop boundaries, and complete-field fine-tuning makes the training state consistent with one-pass full-field inference, which eliminates fixed-tile stitching by construction. On the fully reconstructed background, the correction reduces the station MAE of T2m, V10, and U10 by 17.81%, 27.59%, and 21.06%, respectively.

3. Lightweight operation is achieved across the complete workflow, including training. By streamlining the number of residual dense blocks, compressing the feature channels, and removing batch-normalization layers, the complete system contains only about 8.72 million parameters. The average single-timestamp forward time for two-stage downscaling plus full-field correction is about 87.7 ms, and the entire workflow from training to inference can be completed on a consumer-grade personal computer.

Results

The study region spans 105°–120°E and 25°–40°N, covering the eastern Tibetan Plateau, the Sichuan Basin, the North China Plain, and coastal areas. Climate gradients and terrain forcing are both strong, making the region a demanding and representative testbed for downscaling and correction methods. Training uses more than 10 million station observation records in total, with about 1,200 station observations per timestamp. Model training and testing are performed on a single NVIDIA GeForce RTX 4060 Laptop GPU (8 GB of memory). The data are split chronologically: downscaling training uses hourly data from 2021 to 2023, the validation set extends to 15 February 2024, and the independent test covers 7,689 hourly timestamps from 15:00 UTC on 15 February 2024 to 23:00 UTC on 31 December 2024. The split is fixed before training, and test timestamps do not participate in any parameter optimization. The evaluation proceeds in two stages, covering the downscaling performance and the subsequent observation-driven correction. The 0.25° stage-one downscaling output is evaluated on the grid against the reanalysis reference field and the bicubic-interpolation baseline. Because dense 0.0625° gridded truth is unavailable, the stage-two downscaling output is evaluated through a combination of station-observation comparison, spatial-spectrum and gradient diagnostics, and a scale-closure test after aggregation back to 0.25°. Correction training uses 5,556 complete timestamps immediately preceding the test period. The full-field correction is evaluated on 2,133 independent held-out timestamps driven entirely by reconstructions, with no station observations entered at inference. Significance is tested by block resampling over UTC days (10,000 replicates), and a difference is considered significant when the 95% confidence interval of the error difference does not contain zero.

Dual-stream cascaded downscaling recovers fine spatial structure

To quantify stage-one downscaling accuracy, Table 1 summarizes the gridded evaluation of the 0.25° four variables relative to bicubic interpolation on the fully held-out timestamps, including RMSE, correlation coefficient, and gradient RMSE.

Table 1 | Gridded evaluation of the 0.25° stage-one downscaling relative to bicubic interpolation

Variable	Bicubic RMSE	Model RMSE	RMSE gain	Correlation	Bicubic grad. RMSE	Model grad. RMSE	Gradient gain
MSLP	76.531	20.284	73.50%	0.9998	27.673	12.842	53.59%
T2m	1.605	0.573	64.28%	0.9983	0.683	0.365	46.55%
V10	1.068	0.393	63.17%	0.9857	0.448	0.242	45.99%
U10	0.805	0.370	53.98%	0.9774	0.355	0.229	35.36%

Relative to bicubic interpolation, the four-variable (MSLP, T2m, V10, and U10) RMSE is reduced by 73.5%, 64.3%, 63.2%, and 54.0%, and the gradient RMSE by 53.6%, 46.6%, 46.0%, and 35.4% (Table 1). The advantage of the dual-stream multivariate model therefore comes not only from overall bias adjustment but also from the recovery of spatial gradients at 0.25°.

Fig. 1 presents a spatial example of stage-one downscaling.

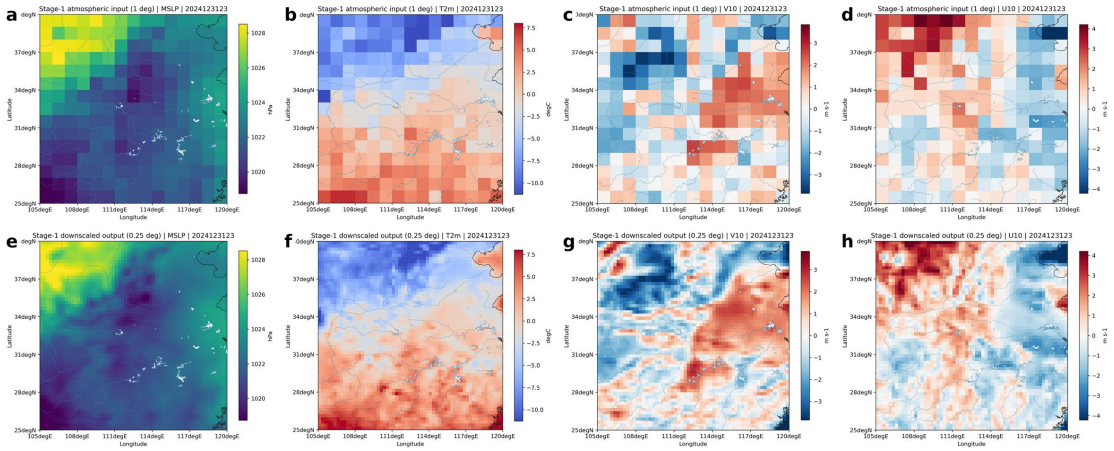


Fig. 1 | Spatial example of 0.25° stage-one downscaling (23:00 UTC on 31 December 2024). a–d, 1° input fields (MSLP, T2m, V10, U10); e–h, corresponding 0.25° outputs of the dual-stream multivariate generators. This timestamp lies within the fully held-out test period. MSLP, T2m, and wind components are expressed in hPa, °C, and m/s, respectively.

The 0.25° outputs preserve the large-scale synoptic phase of the 1° inputs and recover fine spatial gradients that are missing from the inputs (Fig. 1). The high- and low-pressure centres and the ridge–trough structures of MSLP become clearer (Fig. 1e), T2m develops warm–cold contrasts corresponding to the terrain (Fig. 1f), and the wind components show spatially coherent acceleration and channelling features over complex terrain (Fig. 1g, h).

An “oracle cascade”, in which the stage-two meteorological inputs use 0.25° truth, is set up as a diagnostic upper bound on error propagation. The operational method instead uses the four stage-one reconstructed fields throughout. Fig. 2 presents the fully reconstructed cascade case at the same held-out timestamp as Fig. 1.

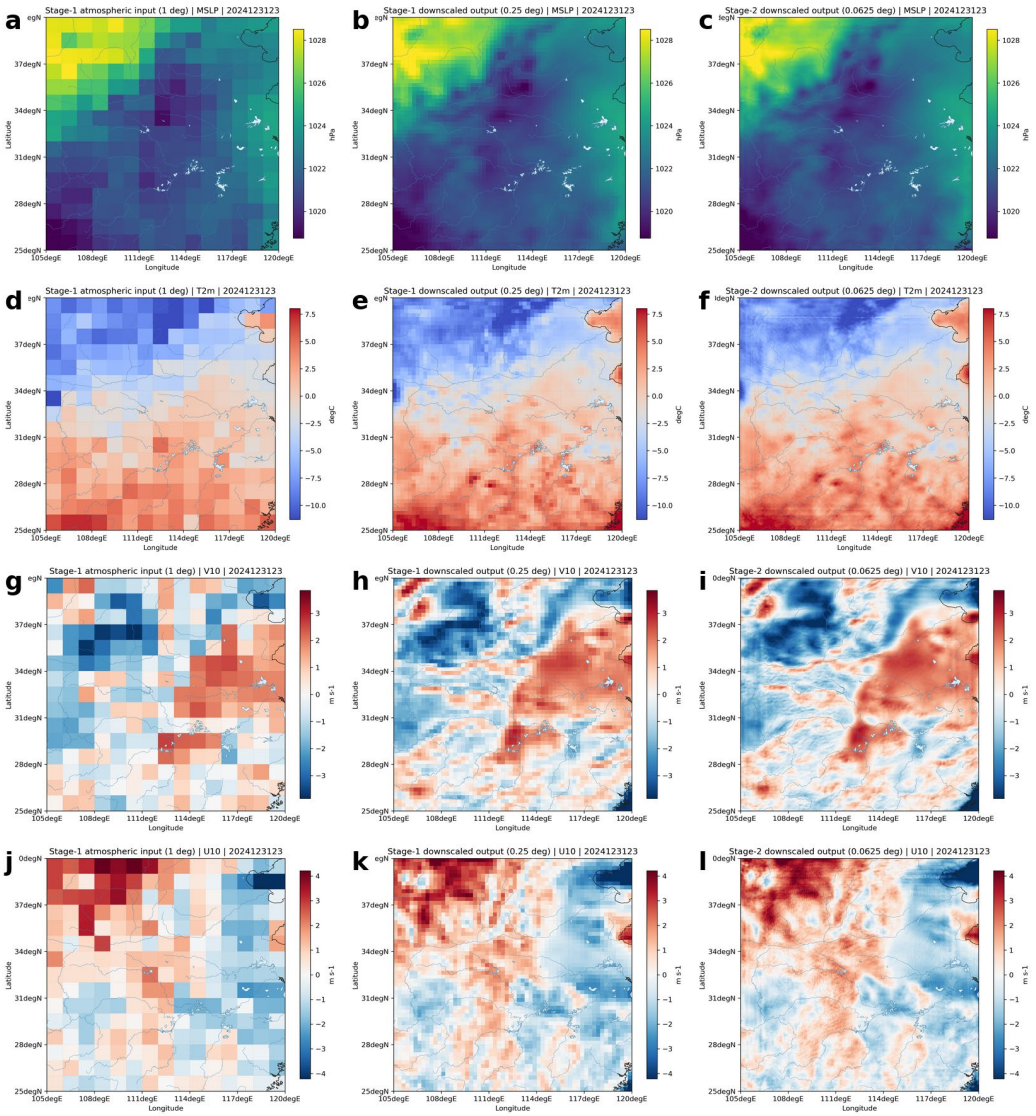


Fig. 2 | Fully reconstructed cascade case (23:00 UTC on 31 December 2024). The four rows correspond to MSLP, T2m, V10, and U10 (a–c, d–f, g–i, j–l), and the three columns correspond to the 1° input, the 0.25° stage-one reconstruction, and the 0.0625° stage-two reconstruction. All stage-two meteorological inputs are stage-one reconstructed fields, and no 0.25° meteorological truth is read. Units: hPa for MSLP, °C for T2m, and m/s for wind components.

The stage-two T2m field preserves the north–south temperature contrast of the stage-one reconstruction and adds fine-scale texture that follows the terrain (Fig. 2e, f). The stage-two U10 field develops narrow acceleration and lee-sheltering structures along terrain corridors that are present in neither the 1° input nor the 0.25° reconstruction (Fig. 2j–l). MSLP maintains a stable large-scale circulation phase across the cascade (Fig. 2a–c), and the channelized acceleration structures of V10 are consistent with those of U10 (Fig. 2g–i).

To examine the reconstruction skill of the recursive second stage, Table 2 reports scale-closure and fine-scale-detail statistics for the cascade.

Table 2 | Scale closure and newly added detail of the second stage

Variable	Stage 1 vs HR RMSE	Aggregated vs stage 1 RMSE	Aggregation corr.	Aggregated vs HR RMSE	Fine-detail std.
MSLP	20.284	23.674	0.9997	30.126	24.713
T2m	0.573	0.553	0.9986	0.761	0.610

Variable	Stage 1 vs HR RMSE	Aggregated vs stage 1 RMSE	Aggregation corr.	Aggregated vs HR RMSE	Fine-detail std.
V10	0.393	0.360	0.9883	0.517	0.415
U10	0.370	0.308	0.9850	0.444	0.349

Stage-one RMSE and aggregated stage-two RMSE both use 0.25° ERA5 as the reference. The standard deviation of the newly added detail is the residual standard deviation of the 0.0625° field relative to the bicubic-upscaled stage-one field.

After aggregation back to 0.25° by 4×4 area averaging, the RMSE relative to the stage-one reconstruction is 23.67 Pa for MSLP, 0.553 °C for T2m, 0.360 m/s for V10, and 0.308 m/s for U10, with correlation coefficients all above 0.9849; relative to 0.25° ERA5, the aggregated RMSE values are 30.13 Pa, 0.761 °C, 0.517 m/s, and 0.444 m/s, respectively (Table 2). The second stage therefore maintains the large-scale phase and spatial distribution, but the recursively generated detail is accompanied by measurable error propagation.

Table 3 further compares the MAE of the 1° input, the 0.25° stage-one product, and the 0.0625° stage-two product on the common set of valid stations.

Table 3 | MAE comparison on the common set of valid stations

Variable	Count	1° MAE	0.25° MAE	0.0625° MAE	vs 1°	vs 0.25°
T2m	8,385,362	1.557	1.468	1.451	+6.86%	+1.19%
V10	8,370,718	1.281	1.277	1.304	-1.80%	-2.16%
U10	8,370,718	1.140	1.130	1.133	+0.54%	-0.33%

Positive values indicate improvement of the second stage relative to the baseline, and negative values indicate degradation.

On the common station set, 0.0625° T2m continues to improve relative to both the 1° and 0.25° products (Table 3; MAE of 1.451, compared with 1.557 and 1.468 for the 1° and 0.25° products), while V10 and U10 remain close to the 0.25° product, with differences within about 2%. Cascade-strategy experiments further show that the station errors of the fully reconstructed cascade and the oracle cascade are very close. The operational-cascade MAE values of T2m, V10, and U10 are 1.335, 1.202, and 1.069, while the oracle results are 1.332, 1.194, and 1.073. Compared with a second stage that uses only bicubic upscaling, the learned second stage improves T2m and U10 by 6.2% and 2.3%, respectively, and the difference for V10 is small. The stage-two detail should therefore be interpreted jointly through stations, cross-scale closure, gradients, and spectra, rather than judged by visual sharpness alone.

Even 0.0625° (about 6 km) is still not truly “fine” for many local applications. The cascade design in principle allows further recursion toward finer resolutions: as long as static terrain data at the corresponding resolution are available on the target grid, another refinement stage can be stacked on top of the existing complete reconstructed state[51,52]. Because weights are reused across scales, an additional stage requires no new model parameters. At finer scales, however, the lack of dense truth and station constraints would make any fine-tuning and, above all, quantitative evaluation considerably more difficult, and this extension remains to be tested systematically.

Without dense 0.0625° labels, the second stage produces variable-dependent fine-scale increments while maintaining high cross-scale correlation. Recursive downscaling alone, however, cannot remove the local biases of the meteorological variables relative to the stations, so a separate observation-driven correction stage is required.

Flow-constrained full-field residual correction improves station credibility

The correction module learns only additive residuals rather than regenerating the complete meteorological field, and training follows the two-stage strategy in which uniformly sampled supervised cores are combined with complete-field fine-tuning.

To evaluate the improvement of the correction model at the station scale, Table 4 compares the MAE, RMSE, mean bias, and correlation coefficient of the fully reconstructed 0.0625° background and a single complete-field correction on the 2,133 independent held-out timestamps. Fig. 3 further contrasts the station errors before and after correction and compares them with linear model output statistics (MOS)45.

Table 4 | Independent station evaluation of the fully reconstructed 0.0625° background and the corrected field

Method	Variable (unit)	Count	MAE	RMSE	Bias	Corr.	MAE gain
Background	T2m (°C)	2,658,371	1.377	1.844	-0.055	0.974	—
Background	V10 (m/s)	2,652,683	1.225	1.591	-0.272	0.629	—
Background	U10 (m/s)	2,652,683	1.087	1.449	-0.050	0.613	—
NetIC	T2m (°C)	2,658,371	1.131	1.498	-0.209	0.983	17.81%
NetIC	V10 (m/s)	2,652,683	0.887	1.198	-0.026	0.712	27.59%
NetIC	U10 (m/s)	2,652,683	0.858	1.169	+0.015	0.707	21.06%

Both use identical valid station–time masks, and the correction model receives no station observations at inference.

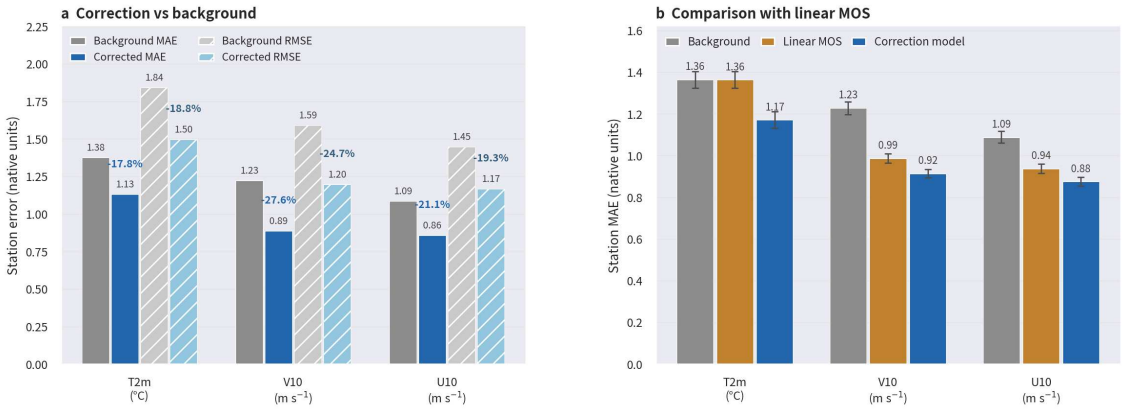


Fig. 3 | Station-scale evaluation of the complete-field correction. a, station MAE (solid) and RMSE (hatched) before and after correction on the same independent station–time samples, with blue numbers giving the reduction relative to the background (values as in Table 4). b, station MAE of the correction model and linear model output statistics (MOS); error bars show 95% confidence intervals from UTC-day block resampling (10,000 replicates). MOS is fitted on pairs of model output and station observations from the training period, whereas the correction model uses no station observations at inference. Units: °C for T2m and m/s for wind components.

Relative to the uncorrected background, the correction model reduces the station MAE of T2m, V10, and U10 from 1.377, 1.225, and 1.087 to 1.131, 0.887, and 0.858, corresponding to improvements of 17.81%, 27.59%, and 21.06%, respectively (Table 4, Fig. 3). The RMSE is improved by 18.76%, 24.72%, and 19.30%. The correlation coefficients of V10 and U10 increase from 0.629 and 0.613 to 0.712 and 0.707, and that of T2m increases from 0.974 to 0.983. The comparison with linear MOS (Fig. 3b) shows that the station MAE of the correction model is lower than that of MOS for all three variables (T2m, 1.17 versus 1.36 °C; V10, 0.92 versus 0.99 m/s; U10, 0.88 versus 0.94 m/s).

Table 5 summarizes the random-context training, complete-field fine-tuning, and independent-test configurations of the correction model; complete-field fine-tuning places the model in the real spatial extent of operational inference.

Table 5 | Two-stage training and one-pass complete-field inference configuration of the correction model

Stage	Input	Output	Times	Epochs	Learning rate	Spatial protocol	Batch
Random context	4×112×112	3×112×112 residual	5,556	32	1×10^{-4}	80 core + 16 context	16
Full-field tune	4×272×272	3×272×272 residual	5,556	8	2.5×10^{-5}	240 field + 16 reflection	4
Independent test	4×240×240	3×240×240 corrected	2,133	—	—	One full-field pass	1

The cascade products still retain local errors at the station scale (Table 3). Fig. 4 presents a complete-field correction case to illustrate how the correction model compensates for these local biases in a spatially continuous manner.

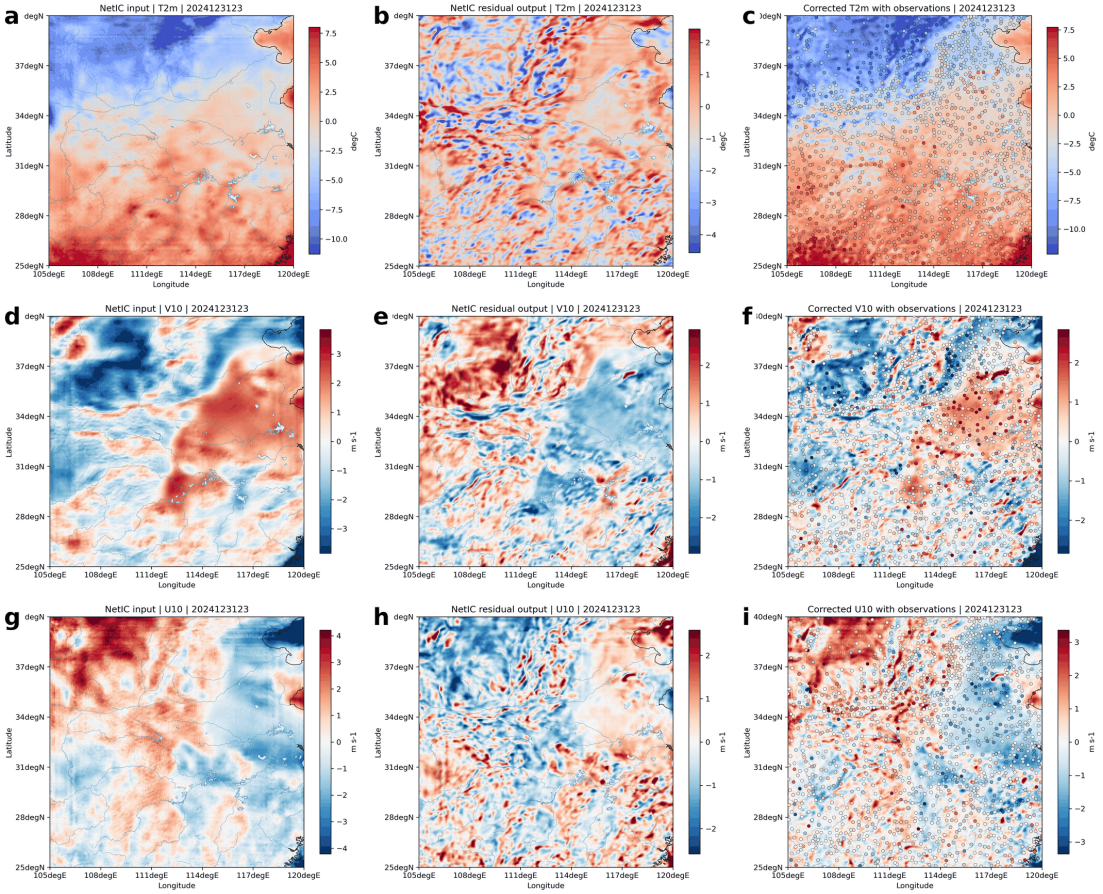


Fig. 4 | Complete-field correction case (23:00 UTC on 31 December 2024). a–c, fully reconstructed background, signed residual, and corrected field of T2m; d–f, V10; g–i, U10. Simultaneous valid station observations are overlaid on the corrected fields in the right column, with the scatter points sharing the colour scale of the field. Units: °C for T2m and m/s for wind components.

The correction residuals form continuous positive and negative adjustments in terrain-transition zones and in regions with significant background bias (Fig. 4b, e, h), while the large-scale structure after correction is still governed by the downscaled background. Because inference covers the whole region directly, the residual contains no stitching boundaries between fixed 80×80 blocks. The values of the corrected fields at station locations agree generally with the overlaid observations, providing a pointwise confirmation of the gridded correction (Fig. 4, right column). The consistent improvement at the station scale and the spatial continuity together indicate that observation-driven residual learning can compensate for the local biases that are not removed in the cascaded reconstruction. The residual fields also illustrate the effect of the flow constraint: the adjustments are organized into continuous bands along the background flow direction rather than isolated spots scattered between stations, and the

residual amplitude varies smoothly along streamlines without grid-scale non-physical patches. This spatial morphology is consistent with the design of the flow-constrained regularizer. Current limitations include station representativeness error, a residual negative bias of T2m after correction, the inability of the deterministic output to express multiple plausible solutions, and the absence of internal ablations for the complete-field correction; future work will add station-elevation and exposure metadata, spatial leave-station-out validation, probabilistic residual modelling, and a direction weight that decays with wind speed in weak-wind regions.

Ablation experiments and computational efficiency

The contribution of each model component is evaluated through ablation experiments, and computational efficiency is assessed separately. The downscaling ablations compare the full multivariate model with single-variable input, no-terrain, no-GAN, and several stage-two recursion strategies, and the correction ablations compare the full correction model with four variants (station-loss-only, isotropic regularization, no-DEM input, and independent inference) on the same fully reconstructed background (Table 6). All groups share the same data split, model capacity, and evaluation window, so that the difference in each ablation can be traced to a single design choice.

Table 6 | Design matrix of the main and ablation experiments (rows grouped by Fig. 5a–c)

Experiment	Multivariate	Terrain	Stage-two input	Station loss	Flow loss	Purpose
0.25° gridded ablations (Fig. 5a)						
Bicubic	No	No	—	—	—	Interpolation baseline
Full downscaling	Yes	DEM/slope/aspect	Fully reconstructed	—	—	Main method
Single-variable	No	Yes	Fully reconstructed	—	—	Multivariate contribution
No terrain	Yes	No	Fully reconstructed	—	—	Terrain contribution
No GAN	Yes	Yes	Fully reconstructed	—	—	Adversarial contribution
0.0625° cascade station ablations (Fig. 5b)						
No learned stage 2	Yes	Yes	Bicubic enlargement	—	—	Stage-two contribution
Oracle cascade	Yes	Yes	Partial truth	—	—	Propagation diagnostic
No weight sharing	Yes	Yes	Fully reconstructed (independent weights)	—	—	Weight-reuse contribution
Correction ablations (Fig. 5c)						
NetIC	T2m/V10/U10	Normalized DEM	Fully reconstructed	Yes	Yes	Full-field correction
Station loss only	T2m/V10/U10	Normalized DEM	Fully reconstructed	Yes	No	Flow-loss contribution
Isotropic regularization	T2m/V10/U10	Normalized DEM	Fully reconstructed	Yes	Isotropic	Flow-loss form
No DEM input	T2m/V10/U10	No	Fully reconstructed	Yes	Yes	DEM contribution
Independent inference	Per-variable	Normalized DEM	Fully reconstructed	Yes	Yes	Joint-correction contribution

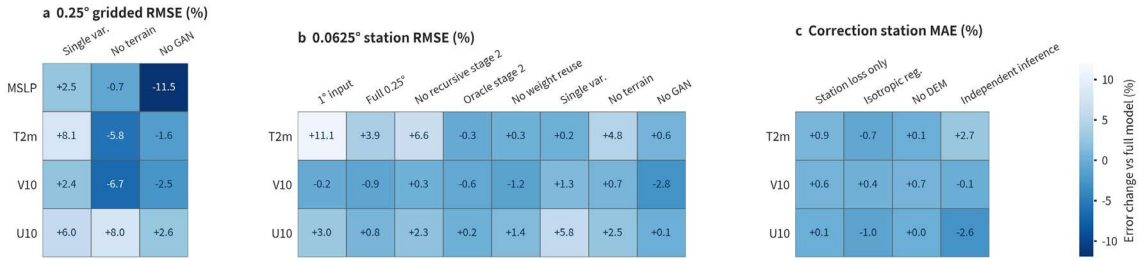


Fig. 5 | Relative error changes of the downscaling and correction ablations. a, change in 0.25° gridded RMSE relative to the full model (%); b, change in RMSE on the common station samples at 0.0625° relative to the full model (%); c, change in station MAE of the correction-ablation variants relative to the full model on the fully reconstructed background (%).

The full model serves as the reference for all ablation comparisons (Fig. 5). On the 0.25° grid, the full downscaling model reduces the RMSE of MSLP, T2m, V10, and U10 by 59.7%, 57.5%, 59.8%, and 55.2% relative to bicubic interpolation. Single-variable input, which removes cross-variable information, consistently increases the 0.25° RMSE of all four variables by 2.4%–8.1%, with T2m the most sensitive (+8.1%). At the 0.0625° station scale, it degrades U10 by 5.8% and V10 by 1.3%. Among the tested components, removing the joint encoding of meteorological variables produces the most consistent error increase across variables.

Removing terrain does not consistently degrade the 0.25° grid-mean RMSE. It even improves the metric slightly for some variables (T2m −5.8%, V10 −6.7%, MSLP −0.7%), and only U10 degrades (+7.9%). At the 0.0625° station scale, removing terrain consistently degrades the station RMSE of all three variables (T2m +4.8%, U10 +2.5%, V10 +0.7%; all significant under the 95% paired date-block bootstrap). The influence of terrain input therefore shows clear dependence on scale and metric. The weak terrain benefit in the 0.25° grid-mean RMSE should not be interpreted as redundancy of the terrain stream, because domain-wide grid-mean metrics dilute local and conditional terrain forcing.

Removing adversarial training shows a typical accuracy–detail trade-off. The no-GAN model attains lower 0.25° gridded RMSE for MSLP, T2m, and V10 (−11.5%, −1.6%, and −2.5%), and only U10 degrades (+2.6%). Its station-scale effects are inconsistent (T2m +0.6%, V10 −2.8%, U10 +0.1%), and the spatial gradients and fine-scale variance are weakened at the same time. The adversarial objective therefore mainly preserves fine spatial structure rather than grid-mean accuracy.

The recursion-strategy controls further decompose the contribution of the second stage. Relative to a second stage with only bicubic upscaling, learned recursion improves the station RMSE of T2m and U10 by 6.6% and 2.3%, respectively, while V10 is close to neutral (+0.3%). The station errors of the oracle cascade, in which non-target stage-two variables use 0.25° truth, differ from those of the operational fully reconstructed cascade by no more than 0.6%. This indicates that the stage-two uncertainty mainly comes from the scale recursion itself rather than from missing intermediate truth. Removing cross-scale weight sharing changes the errors by no more than ±1.4%, indicating that parameter reuse is a safe lightweight default.

The correction ablations (Fig. 5c) give the station-MAE changes of each variant relative to the full configuration. Retaining only the station loss increases the station MAE of T2m, V10, and U10 by 0.87%, 0.60%, and 0.07%, respectively. Removing the DEM input changes the three by +0.14%, +0.65%, and +0.04%. The isotropic-regularization and independent-inference variants attain slightly better pointwise MAE for individual variables (independent inference reduces U10 by 2.64%) but degrade the others (independent inference increases T2m by 2.73%). The effect of each variant is therefore variable dependent.

Taken together, these results indicate that the full configuration provides a robust compromise across

metrics and is retained as the operational default.

To assess the cost of operational deployment, Table 7 measures the single-timestamp forward efficiency of each component on unified hardware.

Table 7 | Single-timestamp forward efficiency on unified hardware

Component	Parameters	Mean	Median	P95	Batch	Protocol
Stage-one four variables	8.44 M	41.65 ms	41.04 ms	58.64 ms	1	15×15→60×60
Two-stage downscaling	8.44 M	85.70 ms	84.59 ms	108.19 ms	1	15×15→240×240
Full-field NetIC	0.276 M	1.96 ms	1.89 ms	2.37 ms	1	One 240×240 pass

Pure forward efficiency on an NVIDIA RTX 4060 Laptop GPU with FP32 and batch size one. The statistics exclude data reading, preprocessing, denormalization, file I/O, and plotting.

Table 7 shows that the average forward times of stage-one four-variable downscaling and two-stage four-variable downscaling are 41.65 and 85.70 ms, respectively. A single complete-field forward pass of the correction model takes only 1.96 ms on average, with a median of 1.89 ms and a P95 of 2.37 ms. Adding the independently measured averages, the full chain of two-stage downscaling plus complete-field correction takes about 87.7 ms per timestamp, well below 0.2 s. Most of the computation comes from the four variable generators, and the complete-field correction adds only about 2.3% of the forward time. In terms of parameters, each variable generator contains about 2.11 million parameters, the four generators together contain 8.44 million and are reused at both scales, and the correction model contains only 276,483 parameters. The complete system therefore has about 8.72 million parameters, and the entire workflow from training to inference can be completed on a consumer-grade personal computer.

Methods

Overall framework: reconstruction-driven cascaded downscaling and observation-driven residual correction

The overall framework consists of reconstruction-driven cascaded downscaling followed by observation-driven full-field residual correction (Fig. 6).

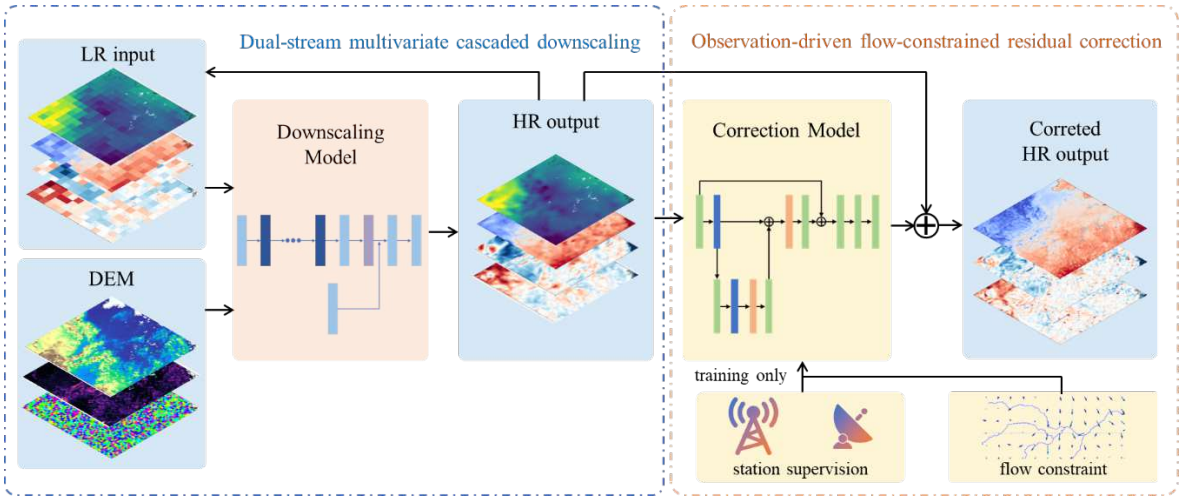


Fig. 6 | Overall scientific framework. Left, terrain-guided dual-stream multivariate cascaded downscaling; right, complete-field residual correction by the correction model.

The hourly inputs are the ERA5 fields of mean sea-level pressure, 2 m temperature, and 10 m zonal and

meridional winds on the study domain described in the Results, with 4× refinement at each stage.

The static terrain inputs are DEM, slope, and aspect. Meteorological inputs are standardized using training-set statistics, and terrain inputs are normalized to unit range.

On the basis of the above data, the complete input-to-output process of the framework is defined formally by Eqs. (1)–(3). Let the 1° background field be $X^{(0)}$=[P,T,V,U], where P, T, V, and U are mean sea-level pressure, 2 m temperature, and the 10 m meridional and zonal wind components, respectively, and let $Z^{(s)}$=[dem,slope,aspect] be the terrain at the target scale. The first stage maps $X^{(0)}$ to the complete reconstructed state through four variable-specific generators conditioned on the 0.25° terrain $Z^{(1)}$:

$$\hat{X}^{(1)} = G_1\left(X^{(0)}, Z^{(1)}\right),\ X^{(0)} = [P,T,V,U],\ Z^{(1)} = [dem, slope, aspect] \quad (1)$$

The second stage refines the fields recursively on the 0.0625° terrain $Z^{(2)}$, using only stage-one reconstructions as meteorological inputs:

$$\hat{X}^{(2)} = G_2\left(\hat{X}^{(1)}, Z^{(2)}\right) \quad (2)$$

The correction stage takes the reconstructed temperature, wind components, and normalized terrain as input and learns an additive residual Δ=[ΔT,ΔV,ΔU] constrained by station observations. The final product is the sum of the background and the residual. Station observations supervise Y only through the training objective and are not inference inputs:

$$Y = \hat{X}^{(2)} + \Delta,\ \Delta = F(Q),\ Q = [\hat{T}^{(2)}, \hat{V}^{(2)}, \hat{U}^{(2)}, DEM_{norm}] \quad (3)$$

The gridded evaluation includes MAE, RMSE, mean bias, correlation coefficient, spatial-gradient RMSE, structural similarity, radial power-spectrum error, and variograms. Temperature extremes are assessed using the 5th, 95th, and 99th percentile errors. For the wind field, besides the errors of the U10 and V10 components, wind-speed MAE, absolute wind-direction error, and vector RMSE are also reported. The 0.0625° results additionally report the scale-consistency error after aggregation back to 0.25°.

The station evaluation reports MAE, RMSE, mean bias, and correlation coefficient. The uncorrected background serves as the primary equal-information baseline, and linear MOS, which likewise uses no station observations at inference, is included as a statistical-correction reference. Spatial objective-analysis methods that directly use simultaneous neighbouring-station observations are not included in the numerical ranking of the correction model, because their inference information conditions differ from those of a neural correction that requires no station input.

The primary downscaling baseline is bicubic interpolation. The internal ablations keep the training data and evaluation window consistent, and compare single-variable input, no terrain, no GAN, no learned second stage, the oracle cascade, and independent stage-two weights. Deep-downscaling results from the literature are used for discussion and positioning, but simple numerical ranking is not performed when datasets and resolutions are inconsistent.

A lightweight dual-stream multivariate downscaling network

Let the multivariate meteorological field at stage s be $X^{(s)} \in R^{4\times H^s \times W^s}$, let k∈{P,T,V,U} be the target-variable index, and let $Z^{(s+1)} \in R^{3\times 4H^s \times 4W^s}$ be the terrain at the target scale. The structure of the generator for variable k is shown in Fig. 7, and its mapping is defined as:

$$\hat{X}_k^{(s+1)} = G_{\theta_k}\left(\hat{X}^{(s)}, Z^{(s+1)}\right) \quad (4)$$

The dynamic branch extracts synoptic-scale features from the four-channel input through five residual-in-residual dense blocks (RRDBs) and a global residual connection, and the deep features are upsampled 4× to the target grid.

The terrain branch encodes DEM, slope, and aspect independently on the target grid. The dynamic and terrain features are concatenated and fused to generate a single target variable. This design keeps the high-cost RRDB computation in the low-resolution space and avoids compressing the high-resolution terrain to the coarse grid.

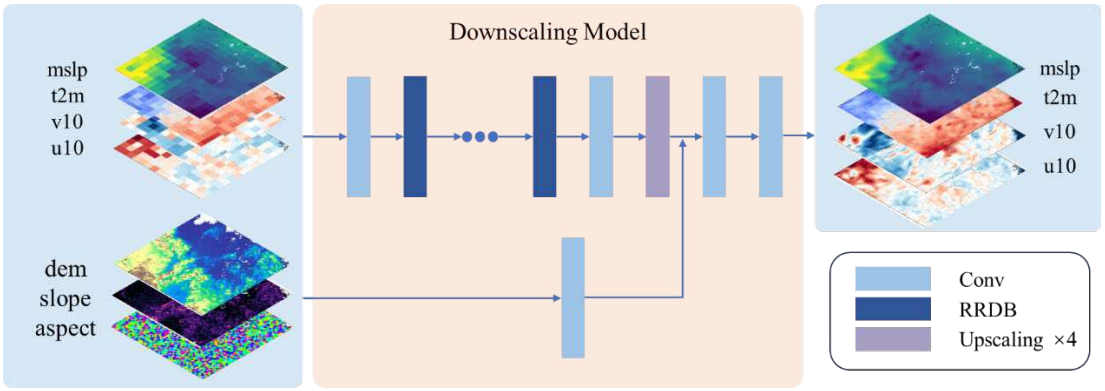


Fig. 7 | Dual-stream downscaling network and fully reconstructed recursion. The dynamic branch jointly encodes the four meteorological variables on the coarse grid and extracts synoptic-scale features through RRDBs. The static branch encodes DEM, slope, and aspect on the target grid. The upsampled dynamic features are fused with the terrain features to generate the target variable. The four variable generators run in parallel to obtain the complete reconstructed state, which is passed to the next scale.

At the first stage, the input is $X^{(0)}$=[P,T,V,U] at 1°, and the four generators output:

$$\hat{X}^{(1)} = concat\left(\hat{X}_P^{(1)}, \hat{X}_T^{(1)}, \hat{X}_V^{(1)}, \hat{X}_U^{(1)}\right) \tag{5}$$

The second stage uses only reconstructed fields:

$$\hat{X}_k^{(2)} = G_{\theta_k}\left(\hat{X}^{(1)}, Z^{(2)}\right), \quad \hat{X}^{(2)} = concat_k\left(\hat{X}_k^{(2)}\right) \tag{6}$$

The parameters θk of the same variable generator are shared between the two scales. This design keeps the model size from growing with cascade depth, and it explicitly defines the second stage as a recursive mapping of the complete reconstructed atmospheric state.

Dense gridded reference fields are available for the 1°-to-0.25° stage. The generator loss consists of an L1 reconstruction loss and an adversarial loss:

$$L_G = L_1\left(\hat{X}_k^{(1)}, X_k^{(1)}\right) + \lambda_{adv} L_{adv} \tag{7}$$

The L1 loss constrains the numerical amplitude, and the adversarial loss encourages the generated results to have local statistics similar to those of the 0.25° reference field. The discriminator is a compact U-shaped encoder–decoder that outputs a gridwise real/fake logit. Both networks are optimized with Adam (learning rates 1×10^{-4} and 5×10^{-5} for the generator and discriminator, respectively, batch size 64) for 80 epochs, and the adversarial weight λadv is 0.2 initially and 0.3 after 1,000 iterations.

Because dense 0.0625° gridded truth is unavailable, the second stage by default reuses the first-stage weights for scale recursion. Training is therefore performed in stages, while the inference chain remains a single continuous pass.

Observation-driven flow-constrained correction network

The correction input is Q=[$\hat{T}^{(2)}$,$\hat{V}^{(2)}$,$\hat{U}^{(2)}$,DEM], where the dynamic fields keep their physical units and the DEM is normalized. The four channels are concatenated at the input end and jointly encoded from the first convolutional layer; the correction network has no separate dynamic and terrain branches.

The correction model is a lightweight encoder–decoder without batch normalization (Fig. 8): two

strided convolutions reduce the spatial size to one quarter, two bilinear-interpolation steps restore the resolution, and additive skip connections reuse the encoder features. There is no additional bottleneck block at the lowest resolution, and the total number of parameters is 276,483. Random-context training ensures that the derivative stencil always uses the real neighbourhood, and complete-field fine-tuning makes the spatial boundary conditions of training consistent with those of operational inference.

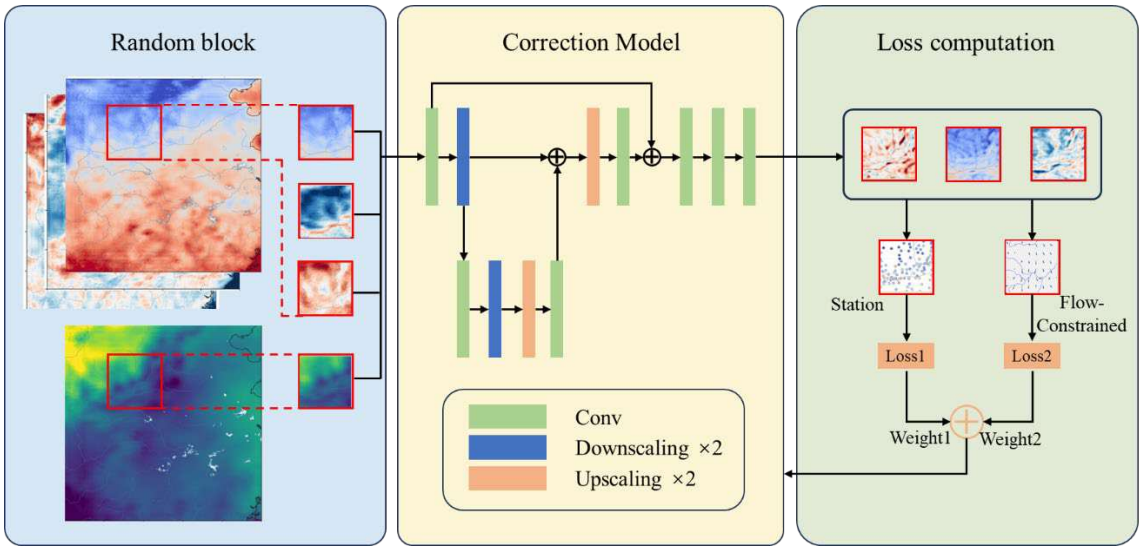


Fig. 8 | Architecture of the complete-field residual correction network. Left, supervised cores and their context are sampled uniformly at random with complete timestamps as units during training; centre, a lightweight encoder–decoder reconstructs three signed residual channels through strided-convolution downsampling and bilinear upsampling; right, the station-supervision loss and the flow-constrained regularizer are combined into the total loss with weights.

The correction mapping is defined as:

$$\Delta = C_{\varphi}(Q),\ Y = Q_{1:3} + \Delta \tag{8}$$

The output-convolution weights are initialized from $N(0,10^{-5})$ and the biases are set to 0, so that the initial corrected field is close to the background. The output channels are ordered ΔT2m, ΔV10, and ΔU10.

In the first stage of correction training, complete timestamps are used as sampling units, supervised cores are selected uniformly at random over the whole region with a margin of real context, and the station loss and the flow-constraint loss are computed only inside the core region.

In the second stage, the model is fine-tuned on complete fields. Inference uses the same boundary treatment: the reflect-padded field is processed in a single pass and cropped back, without tile stitching.

For the ith station inside the core region, the longitude and latitude are mapped to differentiable sampling coordinates, and the three-variable reconstructed values are extracted from the corrected field Y. Missing variables are excluded through masks. The loss of each variable is averaged separately, and the per-variable averages are then combined with equal weights, so that a variable with more valid observations does not dominate the training. The station-supervision loss is defined as:

$$L_{station} = mean_c mean_{i\in valid(c)}\left[S(Y_c;\ r_i) - y_{i,c}\right]^2 \tag{9}$$

Here c denotes T2m, V10, and U10 in turn, and M is the valid-observation mask. Station observations do not enter the network input, so the model does not require complete station data at every moment during inference.

The local unit wind direction is defined from the downscaled wind field:

$$e = (e_x, e_y) = \frac{(\bar{U},\bar{V})}{\sqrt{\bar{U}^2 + \bar{V}^2 + \varepsilon}} \quad (10)$$

For each corrected residual channel Δc, the second directional derivative along the e direction is computed:

$$D_e^2 \Delta_c = e_x^2 \left(\frac{\partial^2 \Delta_c}{\partial x^2}\right) + 2e_x e_y \left(\frac{\partial^2 \Delta_c}{\partial x \partial y}\right) + e_y^2 \left(\frac{\partial^2 \Delta_c}{\partial y^2}\right) \quad (11)$$

The flow-constrained regularizer and the total correction loss are defined as:

$$L_{flow} = \Sigma_c \, mean[(D_e^2 \Delta_c)^2] \quad (12)$$

$$L_{IC} = L_{station} + \lambda_{flow} L_{flow}, \ \lambda_{flow} = 0.05 \quad (13)$$

This regularizer only constrains the along-flow curvature of the correction residual. It is not equivalent to mass conservation, a divergence constraint, or the full momentum equations.

Station supervision constrains the corrected field at the observed locations, and the flow-constrained regularizer prevents non-physical patches that oscillate rapidly along the flow direction between stations.

The background wind field in the flow constraint is used only to define the local directional coordinate; the penalized quantity is the second-order curvature of the residual of each output variable along that direction, not the wind field itself. Structures such as frontal zones and advection plumes in near-surface temperature are organized by the same atmospheric flow as the wind field, so the constraint carries the same physical meaning for T2m as for the wind components. The correction ablations also support this design: the full flow-constrained model attains lower station errors than the station-loss-only variant for all three variables.

References

1. Veers, P. et al. Grand challenges in the science of wind energy. Science 366, eaau2027 (2019).
2. Clifton, A., Barber, S., Stökl, A., Frank, H. & Karlsson, T. Research challenges and needs for the deployment of wind energy in hilly and mountainous regions. Wind Energy Sci. 7, 2231–2254 (2022).
3. Benton, B. N. et al. Super-resolution for renewable energy resource data with wind from reanalysis data and application to Ukraine. Energies 18, 3769 (2025).
4. Goger, B., Rotach, M. W., Gohm, A., Stiperski, I. & Fuhrer, O. Current challenges for numerical weather prediction in complex terrain: topography representation and parameterizations. In 2016 International Conference on High Performance Computing & Simulation, 890–894 (2016).
5. Hewitson, B. C. & Crane, R. G. Climate downscaling: techniques and application. Clim. Res. 7, 85–95 (1996).
6. von Storch, H., Zorita, E. & Cubasch, U. Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J. Clim. 6, 1344–1351 (1993).
7. Wilby, R. L., Dawson, C. W. & Barlow, E. M. SDSM—a decision support tool for the assessment of regional climate change impacts. Environ. Model. Softw. 13, 1–13 (1998).
8. Giorgi, F. Simulation of regional climate using a limited area model: case study of a European summer. J. Clim. 3, 941–963 (1990).
9. Racherla, P. N., Shindell, D. T. & Faluvegi, G. The added value of global climate model information in quantifying the uncertainty of regional-scale climate projections. J. Geophys. Res. Atmos. 117, D3 (2012).
10. Soares, P. M. et al. WRF high-resolution dynamical downscaling of ERA-Interim for Portugal. Clim. Dyn. 39, 2497–2522 (2012).
11. Fernández-González, S. et al. Sensitivity analysis of the WRF model: wind-resource assessment for complex terrain. J. Appl. Meteorol. Climatol. 57, 733–753 (2018).

12. Rasp, S. & Lerch, S. Neural networks for postprocessing ensemble weather forecasts: a review of recent results and future directions. Mon. Weather Rev. 150, 1–26 (2022).
13. Buster, G., Benton, B. N., Glaws, A. & King, R. N. High-resolution meteorology with climate change impacts from global climate model data using generative machine learning. Nat. Energy 9, 894–906 (2024).
14. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204 (2019).
15. Bracco, A. et al. Machine learning for the physics of climate. Preprint (2024).
16. Cui, Y. et al. Forecasting the eddying ocean with a deep neural network. Nat. Commun. 16, 2268 (2025).
17. Dong, C., Loy, C. C., He, K. & Tang, X. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38, 295–307 (2016).
18. Lim, B., Son, S., Kim, H., Nah, S. & Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In CVPR Workshops, 136–144 (2017).
19. Shi, W. et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In CVPR, 1874–1883 (2016).
20. Zhang, L., Yang, A., Amor, R. A., Zhang, B. & Rao, D. Super resolution on global weather forecasts. Preprint at https://arxiv.org/abs/2409.11502 (2024).
21. Vandal, T., DeCastro, A., Wang, S. et al. Statistical downscaling of climate models using deep learning. In AAAI Workshop: AI for Social Good (2017).
22. Vandal, T. et al. DeepSD: generating high resolution climate change projections through single image super-resolution. In KDD 2017, 1663–1672 (2017).
23. Shi, X., Gao, Z., Lausen, L. et al. Deep learning for precipitation nowcasting: a benchmark from high-resolution radar data. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 31 (2017).
24. Mei, P., Yang, Z., Liu, C, Wang, L. & Yin, Z. ST-MSRN: an enhanced spatio-temporal super-resolution model for complex meteorological data reconstruction. Big Earth Data https://doi.org/10.1080/20964471.2025.2540712 (2025).
25. Lei, W., Qin, H., Chen, H. & Hou, X. A spatio-temporal downscaling model based on multi-meteorological feature fusion. Radio Engineering 54, 483–496 (2024). (in Chinese)
26. Groenke, D., Krasnopolsky, V., Mulholland, M. et al. Machine learning for climate model downscaling and emulation. In AGU Fall Meeting Abstracts (2020).
27. Hong, Y., Wang, X., Wang, B., Li, W. & Han, G. A fusion method based on physical modes and satellite remote sensing for 3D ocean state reconstruction. Remote Sens. 17, 1468 (2025).
28. Kuang, Q., Shen, C., Yu, T. & Liu, J. Weather statistical downscaling using a 3D multi-scale residual Laplacian pyramid network. Trans. Atmos. Sci. 45, 660–673 (2022). (in Chinese)
29. Goodfellow, I. et al. Generative adversarial nets. In Advances in Neural Information Processing Systems 27 (2014).
30. Ledig, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. In CVPR, 4681–4690 (2017).
31. Wang, X. et al. ESRGAN: enhanced super-resolution generative adversarial networks. In ECCV Workshops (2018).
32. Wang, X., Xie, L., Dong, C. & Shan, Y. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data. In ICCV Workshops, 1905–1914 (2021).
33. Groenke, D., Krasnopolsky, V., Mulholland, M. et al. Climate model downscaling with GANs. In AGU Fall Meeting Abstracts (2020).
34. Li, H., Wang, Y., Huang, G., Tao, W. & Lin, P. Generative downscaling and bias correction of multivariable Earth system model simulations. Geophys. Res. Lett. 52, e2025GL117397 (2025).

35. Harris, D., Foufoula-Georgiou, E., Droegemeier, K. K. et al. Multiscale analogs for precipitation downscaling via GANs. J. Adv. Model. Earth Syst. (2022).
36. Sun, S. et al. Fusion of multi-source precipitation records via coordinate-based generative models. Nat. Commun. https://doi.org/10.1038/s41467-025-67987-9 (2025).
37. Mardani, M. et al. Residual corrective diffusion modeling for km-scale atmospheric downscaling. Commun. Earth Environ. 6, 124 (2025).
38. Sun, H. et al. China regional 3km downscaling based on residual corrective diffusion model. Preprint at https://arxiv.org/abs/2512.05377 (2025).
39. Chakraborty, D. et al. Multimodal atmospheric super-resolution with deep generative models. Preprint at https://arxiv.org/abs/2506.22780 (2025).
40. Mardaneh, M., McGinn, R., Gharabaghi, B. et al. Diffusion models for climate downscaling. Environ. Data Sci. (2023).
41. Blau, Y. & Michaeli, T. The perception-distortion tradeoff. In CVPR, 6228–6237 (2018).
42. Lin, C. et al. Reconstructing fine-scale 3D wind fields with terrain-informed machine learning. Nat. Commun. 17, 3713 (2026).
43. Lian, J. et al. TerraWind: a deep learning-based near-surface winds downscaling model for complex terrain region. Geophys. Res. Lett. 51, e2024GL112124 (2024).
44. Le Toumelin, L. et al. Emulating the adaptation of wind fields to complex terrain with deep learning. Artif. Intell. Earth Syst. 2, e220034 (2023).
45. Glahn, H. R. & Lowry, D. A. The use of model output statistics (MOS) in objective weather forecasting. J. Appl. Meteorol. 11, 1203–1211 (1972).
46. Rasp, S. & Lerch, S. Neural networks for postprocessing ensemble weather forecasts. Mon. Weather Rev. 146, 3885–3900 (2018).
47. Karniadakis, G. E. et al. Physics-informed machine learning. Nat. Rev. Phys. 3, 422–440 (2021).
48. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707 (2019).
49. Jiang, P. et al. Efficient super-resolution of near-surface climate modeling using the Fourier neural operator. J. Adv. Model. Earth Syst. 15, e2023MS003800 (2023).
50. Dosovitskiy, A. et al. An image is worth 16×16 words: transformers for image recognition at scale. In ICLR (2021).
51. Li, M. et al. Reconstructing high-resolution DEMs from 3D terrain features using conditional generative adversarial networks. Int. J. Appl. Earth Obs. Geoinf. 133, 104115 (2024).
52. Zhang, B. et al. Super-resolution reconstruction of a 3 arc-second global DEM dataset. Sci. Bull. 67, 2526–2530 (2022).
53. Hersbach, H. et al. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049 (2020).
54. GEBCO Compilation Group. GEBCO 2025 Grid (2025).

Competing interests

The authors declare no competing interests.