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.

Level–shape decomposition delivers accurate multihorizon temperature-field forecasts across the warming Lanzhou–Xining urban agglomeration

Xiao Wang

sxwx@scun.edu.cn

Qinghai Normal University

Ailin He

Sichuan Minzu College

Donghui Zhang

China Academy of Space Technology

Article

Keywords: regional temperature forecasting, level–shape decomposition, rolling-origin validation, thermal-state transition, ERA5-Land reanalysis, spatial field prediction, warm footprint

Posted Date: September 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10708209/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on November 5th, 2026.
See the published version at https://doi.org/10.1038/s41598-026-78313-8.

Level–shape decomposition delivers accurate multihorizon temperature-field forecasts across the warming Lanzhou–Xining urban agglomeration

Xiao Wang [1,2,*], Ailin He [2] and Donghui Zhang [3]

[1]College of Geographical Science, Qinghai Normal University, Xining 810008, Qinghai, China; sxwx@scun.edu.cn (X.W.)

[2]Sichuan Minzu College, Kangding 626001, Sichuan, China; 150025@scun.edu.cn (A.H.)

[3]Institute of Remote Sensing Satellite, China Academy of Space Technology, Beijing 100094, China; zhangdonghui@alu.cdut.edu.cn (D.Z.)

*** Correspondence:**
Xiao Wang
sxwx@scun.edu.cn; Tel.: +86-153-1398-8271

Abstract: Accurate monthly reanalysis temperature-field forecasts must recover both regional displacement and the spatial organization of anomalies, particularly when a region changes thermal state. We developed RegimeTransfer-DCSF (Deep–Causal Shape Fusion), a multihorizon framework that separates a forecast-causal regional level from median-centred spatial shape, learns short-lead shape through a compact fine–coarse attention hierarchy, and transfers shape responsibility from the deep branch to a stable anchor as lead increases. Using 12-month histories of 12 ERA5-Land variables, the model directly forecast future ERA5-Land 2-m air temperature at one-, three-, and six-month leads (h1, h3, and h6) over 39 county-level administrative units of the Lanzhou–Xining urban agglomeration. Within two rolling-origin validation folds, 114 horizon-specific cases yielded 114,114 county-domain grid-cell predictions. The selected DCSF fit achieved a pooled MAE of 1.131 °C, a root-mean-square error of 1.494 °C, and an R^2 of 0.9767, with 83.68% of absolute errors within 2 °C; a fold-wise two-refit county ensemble retained a pooled MAE of 1.135 °C. Averaged across the 114 county-area-weighted cases, the selected fit reduced MAE by 9.95% relative to seasonal climatology and by 20.35% relative to TimeFilter. On the complete 3,655-cell support, the four-refit DCSF mean ranked first in the ten-method rolling-validation benchmark at 1.174 °C, with an across-refit standard deviation of 0.0024 °C. During 25 h1 thermal-state changes, DCSF reduced area-weighted MAE by 0.301 °C (19.79%) relative to TimeFilter, with a fold-stratified three-month moving-block bootstrap interval of 0.073–0.526 °C. In a separate retrospective ERA5-Land analysis, annual mean warm-area fraction increased by 9.29 percentage points per decade during 2000–2024; the autocorrelation-aware three-year block interval was 2.72–15.52 percentage points per decade. Six of the 48 validation target months met the widespread-warm definition and formed highly connected warm footprints. Within this rolling-validation task, explicit level–shape responsibility delivered high absolute field accuracy and its clearest short-lead value during thermal-state changes.

Keywords: regional temperature forecasting; level–shape decomposition; rolling-origin validation; thermal-state transition; ERA5-Land reanalysis; spatial field prediction; warm footprint

1 Introduction

Monthly regional temperature fields support environmental monitoring, ecosystem and agricultural planning, energy management, and climate-risk assessment. Persistent warming over the Tibetan

Plateau is accompanied by marked spatial and seasonal heterogeneity, especially across its northern sector and elevation gradients [1–3]. Useful prediction therefore requires more than estimating a regional mean. A field can change through a domain-wide displacement, a reorganization of warm and cold subregions, or both. These changes are especially important near thermal-state boundaries, where a modest mean shift can recruit a large, spatially connected fraction of the region into an anomalously warm or cold condition. Forecast quality must therefore be judged from the continuous field, its regional level, and its centred spatial shape.

ERA5-Land provides a physically consistent, multidecadal description of land-surface and near-surface atmospheric conditions at a resolution suitable for regional environmental analysis [4]. Gridded data-driven forecasting now spans common weather-field benchmarks, seasonal surface-air-temperature hindcasts, and global neural weather systems, demonstrating the value of explicit spatial targets, lead-dependent evaluation, and strong baselines [5–9]. Convolutional, hybrid-dynamical, and adaptive-Fourier systems provide complementary high-resolution field-forecasting precedents [10–12]. Deep time-series and Earth-system models have expanded the available toolkit. Earthformer introduced cuboid attention for multidimensional Earth-system forecasting [13], while iTransformer, TimeMixer, ModernTCN, TimeXer, MoSSL, TimeFilter, and Timer-XL developed complementary strategies for variate-token attention, multiscale mixing, convolutional modelling, exogenous-variable use [14–17], self-supervised spatiotemporal learning, patch filtering, and long-context forecasting [18–20]. Patch tokens, transformed two-dimensional periodic representations, and sparse-attention decoding broaden the long-horizon sequence-forecasting design space [21–23]. Yet a common structural tension remains: when regional level and within-region shape are learned through one undifferentiated path, the model must preserve stable seasonal organization and recover transient spatial change with the same representation at every lead.

The tension becomes sharper as lead time increases. Spatiotemporal recurrence and lead-specialized cascades illustrate complementary ways to represent field evolution and horizon dependence [7,24]. Flexible short-lead corrections can exploit recent spatial information, whereas stable seasonal and forecast-causal anchors become increasingly valuable at longer leads. TimeMixer, for example, assigns distinct roles to decomposed multiscale components in time-series forecasting [15]. Decomposition, frequency-enhanced representation, and strong linear seasonal-trend models further show why component responsibility and demanding baselines matter [25–27]. The present field task introduces a spatially explicit division: regional displacement and within-domain organization are modelled separately, and their responsibilities are fixed by forecast horizon. This separation also provides an interpretable evaluation language because a forecast may miss the regional level while retaining the spatial shape, or match the level while displacing that shape.

The Lanzhou–Xining urban agglomeration is a demanding setting for this problem. It spans the transition between the northeastern Tibetan Plateau and adjoining lower-elevation terrain, with pronounced west–east thermal contrasts, large seasonal amplitude, and an evolving warm footprint. Remote-sensing land-surface-temperature studies identify multiscale thermal structure associated with urban expansion, vegetation, and valley–hill topography, while the regional settlement system forms a connected Lanzhou–Xining corridor with two principal cores [28,29]. Earlier work in the region used temperature cubes and interpretable machine learning to attribute compound dry-heat exposure and wet-bulb-temperature drivers [30,31]. The present study addresses a different scientific task: rolling-origin prediction of future monthly ERA5-Land T2M fields, followed by a separate retrospective analysis of where forecast advantage occurs and how the regional thermal field is changing.

We introduce RegimeTransfer-DCSF (Deep–Causal Shape Fusion). Throughout this study, *causal* denotes strict forecast-origin availability: every model input is dated at or before the origin, enforcing temporally valid information flow. The framework combines a forecast-causal regional level, a median-centred anchor shape, a compact fine–coarse deep-shape expert, horizon-conditioned decoding, fixed horizon responsibility, and smooth proxy-to-target loss migration. We make three contributions. First, we formulate multihorizon temperature-field forecasting as explicit regional-level and centred-shape recovery. Second, we establish high absolute accuracy, complete benchmark leadership, and optimization-seed stability within two rolling-origin validation folds. Third, we show that the short-lead advantage is concentrated in changing thermal states and place that result within a separately quantified, spatially coherent expansion of the ERA5-Land warm footprint.

2 Materials and methods

2.1 Study area and data foundation

2.1.1 County-domain study area

The analysis covers the Lanzhou–Xining urban agglomeration in Gansu and Qinghai Provinces, China. The authoritative county polygon layer spans 98.92–105.43°E and 34.84–37.63°N, while the 1,001 included grid centres span 99.0–105.3°E and 34.9–37.5°N. The layer comprises 39 administrative polygons, and grid centres were included when they fell strictly within the valid county polygons. The resulting WGS84 0.1° lattice defined the primary maps, thermal-state analyses, and county-domain performance summaries (Figure 1).

The forecasting model generated fields on a larger 44 × 86 computational grid containing 3,655 valid cells and 129 cells outside the fixed valid-domain mask. This support retained the spatial context needed by the model, whereas all geographical interpretation in the paper uses the 1,001-cell county domain. Annual climatology, seasonal amplitude, and the 2000–2024 local warming trend reveal the strong west–east and topographic thermal structure that the forecasts must recover (Figure 1).

2.1.2 ERA5-Land predictors, targets, and fold-specific climatology

Monthly ERA5-Land data covered January 2000 through December 2024, providing 300 complete months. ERA5-Land supplies a temporally consistent land reanalysis at approximately 9-km native resolution, with monthly means derived from its hourly production [4]. ERA5 supplies the global atmospheric reanalysis used to force ERA5-Land [32]. The Copernicus Climate Data Store distributes the product on a regular 0.1° latitude–longitude grid [33]. Each forecast used the preceding 12 months of 12 variables: 2-m air temperature (T2M), 2-m dew-point temperature (D2M), surface pressure (SP), evaporation (EVP), surface sensible heat flux (SHF), skin temperature (SKT), surface solar radiation (SSR), total precipitation (TP), 10-m zonal and meridional wind (U10 and V10), and high- and low-vegetation leaf-area indices (LAI_HighVeg and LAI_LowVeg). The two LAI inputs are prescribed seasonal climatological covariates in ERA5-Land rather than independent interannual vegetation observations.

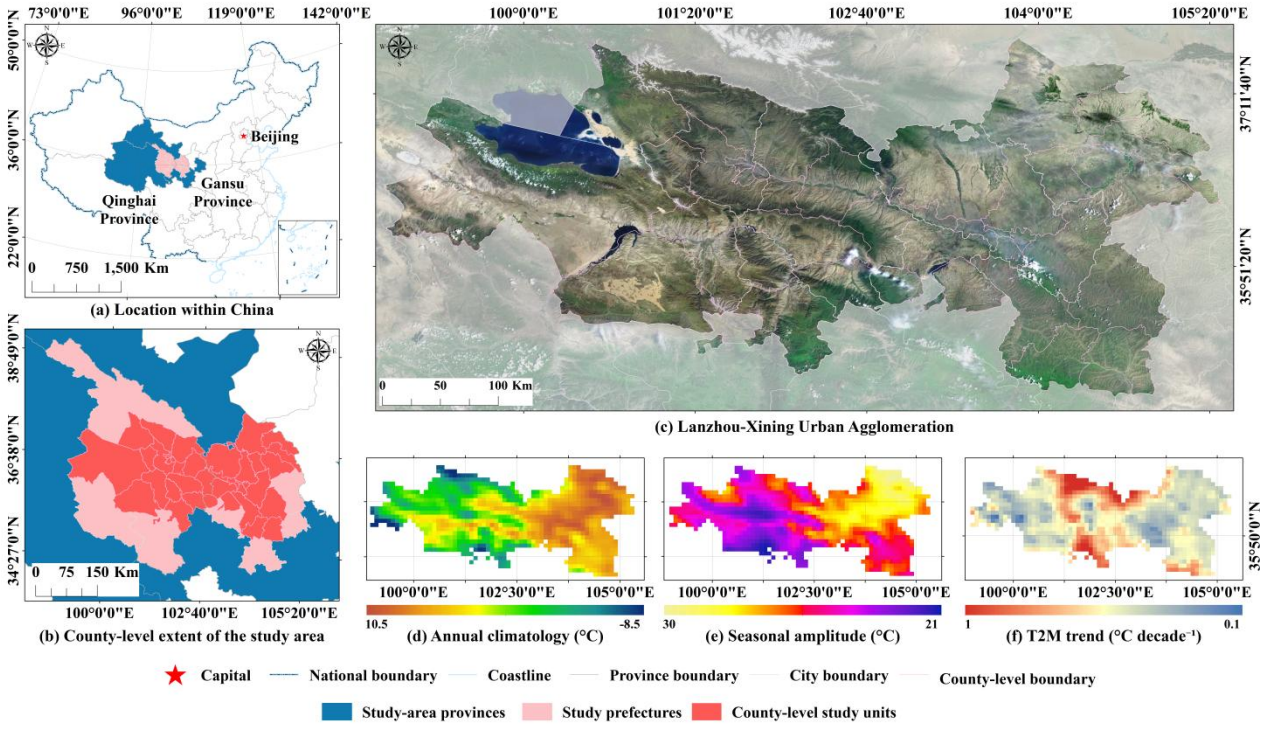


Figure 1. China location, 39-county study extent, Tianditu imagery, and ERA5-Land climatology, seasonality, and warming across the Lanzhou–Xining urban agglomeration. Panels are (a) China context, (b) county extent, (c) Tianditu imagery (accessed 4 August 2026), and (d–f) annual climatology, seasonal amplitude, and 2000–2024 trend at 1,001 grid centres.

A fixed mask identified the same 129 boundary cells in every variable and month; the 3,655 valid support cells contained no time-varying missing values. Predictor values were paired with their validity masks before the convolutional stem, so masked cells could not enter the learned representation. All fold-dependent normalization and climatological quantities were estimated from the corresponding training period. For rolling fold f and calendar month m, let $T_{f,m}$ contain the fold-training indices in month m, let $Y_j(s)$ be T2M at time index j and support cell s, and define the training-only climatological field as

$$C_{f,m}(s) = \frac{1}{|\mathrm{T}_{f,m}|} \sum_{j \in \mathrm{T}_{f,m}} Y_j(s) \tag{1}$$

Equation (1) makes the seasonal background fold-specific and prevents validation months from entering any climatological term. Within one fold, the shorter notation C_m denotes $C_{f,m}$. The forecast target was future ERA5-Land T2M in degrees Celsius at h1, h3, and h6. No station observations or target-period covariates entered model fitting or inference. Complete fields were first produced on the 3,655-cell support and then cropped by grid-centre membership to create the 1,001-cell county analysis archive without refitting or additional inference (Table 1).

Table 1. County study domain, data, and rolling-origin evaluation design. The county analysis uses 1,001 grid centres on a common 3,655-cell model support.

Component	Specification	Quantity
Historical record	Monthly ERA5-Land record	2000–2024; 300 months; 25 complete years
County study domain	County-level study boundary on the WGS84 0.1° grid-centre lattice	39 county-level administrative units; 1,001 county grid centres
Computational support	WGS84 0.1° model grid over the Lanzhou–Xining urban agglomeration	44 × 86 grid; 3,655 valid cells; 129 cells outside the valid-domain mask
Forecast predictors	T2M, D2M, SP, EVP, SHF, SKT, SSR, TP, U10, V10, LAI high vegetation, and LAI low vegetation	Preceding 12 months; both LAI fields are prescribed monthly climatological covariates
Forecast target	Direct future ERA5-Land T2M	Lead months: h1, h3, and h6
Historical warm footprint	Cell- and calendar-month q90 relative to the fixed 2000–2018 reference	Cosine-latitude-weighted county-domain warm-area fraction
Rolling A	Training: 2000–2014; target months: 2016–2017	24 unique target months; 57 forecast cases
Rolling B	Training: 2000–2016; target months: 2018–2019	24 unique target months; 57 forecast cases
Forecast horizons	Direct h1, h3, and h6 predictions	38 cases per horizon; 114 cases in total
Thermal-state definition	Fold-training cell/calendar-month q10 and q90; regional/widespread thresholds 20%/50%	Six mutually exclusive states; complete 3 × 3 threshold sensitivity
Evaluation and inference	Pooled unweighted grid metrics; cosine-area-weighted case metrics; fold-target-month resampling	114 × 1,001 = 114,114 county predictions; model generation on 3,655-cell support

2.1.3 Rolling-origin folds, horizons, and thermal states

Evaluation used two temporally separated rolling-origin folds (Table 1) [34–36]. Rolling A trained on 2000–2014 and evaluated targets in 2016–2017; rolling B trained on 2000–2016 and evaluated targets in 2018–2019. Each fold contained 19 forecast origins. Direct predictions at h1, h3, and h6 yielded 57 horizon-specific cases per fold, while the union of their target dates covered 24 distinct months. Across folds, the evaluation comprised 38 cases per horizon, 114 horizon-specific cases, 48 distinct fold–target months, and 114,114 county grid-cell predictions. Months after the training cutoff could enter only as predictor history once they were available at the forecast origin; they never entered parameter fitting, normalization, or climatology estimation.

Forecast-period thermal states were defined from cell- and calendar-month-specific q10 and q90 thresholds of training-period T2M anomalies relative to the Equation (1) climatology. Let Ω_c be the 1,001-cell county domain, let m_t be the target calendar month, let φ_s be the latitude of cell s, and let $a_s = \cos\varphi_s$ be its area weight. The area-weighted warm fraction for target month t was

$$F^{+}_{f,t} = \frac{\sum_{s\in\Omega_c} a_s \mathbf{1}\left[Y_t(s) - C_{f,m_t}(s) > q_{0.90,f,m_t}(s)\right]}{\sum_{s\in\Omega_c} a_s} \tag{2}$$

Equation (2) supplies the warm-coverage coordinate used by the mutually exclusive thermal-state rule. The corresponding cold fraction was

$$F^{-}_{f,t} = \frac{\sum_{s\in\Omega_c} a_s \mathbf{1}\left[Y_t(s) - C_{f,m_t}(s) < q_{0.10,f,m_t}(s)\right]}{\sum_{s\in\Omega_c} a_s} \tag{3}$$

Equation (3) supplies the cold-coverage coordinate used jointly with Equation (2) to assign the target state. Here, $q_{p,f,m}(s)$ denotes the cell- and calendar-month-specific p-quantile estimated from fold f training anomalies, and $1[\cdot]$ is the indicator function. Labels were assigned in priority order: spatially polarized when both fractions were at least 0.20; widespread warm or cold when the corresponding fraction was at least 0.50; regional warm or cold when it was at least 0.20; and near climatology otherwise. A state change occurred when adjacent within-fold h1 target months had different labels. The principal h1 transition analysis contained 25 changing target months. A prespecified 3 × 3 sensitivity grid crossed regional thresholds of 0.15, 0.20, and 0.25 with widespread thresholds of 0.40, 0.50, and 0.60; forecast performance did not determine the alternative definitions.

2.2 RegimeTransfer-DCSF

RegimeTransfer-DCSF separates the future field into a regional level and a median-centred spatial shape (Figure 2). The level path represents the seasonal field and forecast-causal regional displacement. Two shape paths then represent a stable horizon-specific anchor and a learned short-lead correction. Their responsibility is fixed by horizon rather than estimated separately for each forecast case. The selected 48-channel sigmoid-migration specification contains 239,809 trainable neural parameters and uses the fold-specific XGBoost h6 anchor fitted with maximum depth 6, 600 trees, and learning rate 0.03.

2.2.1 Causal regional level and centred anchor

Let Ω denote the 3,655-cell valid computational support and let $Z(s)$ be any spatial field. The median-centring operator is

$$\mathrm{C}_{\Omega}[Z](s) = Z(s) - \mathrm{med}_{u\in\Omega} Z(u) \tag{4}$$

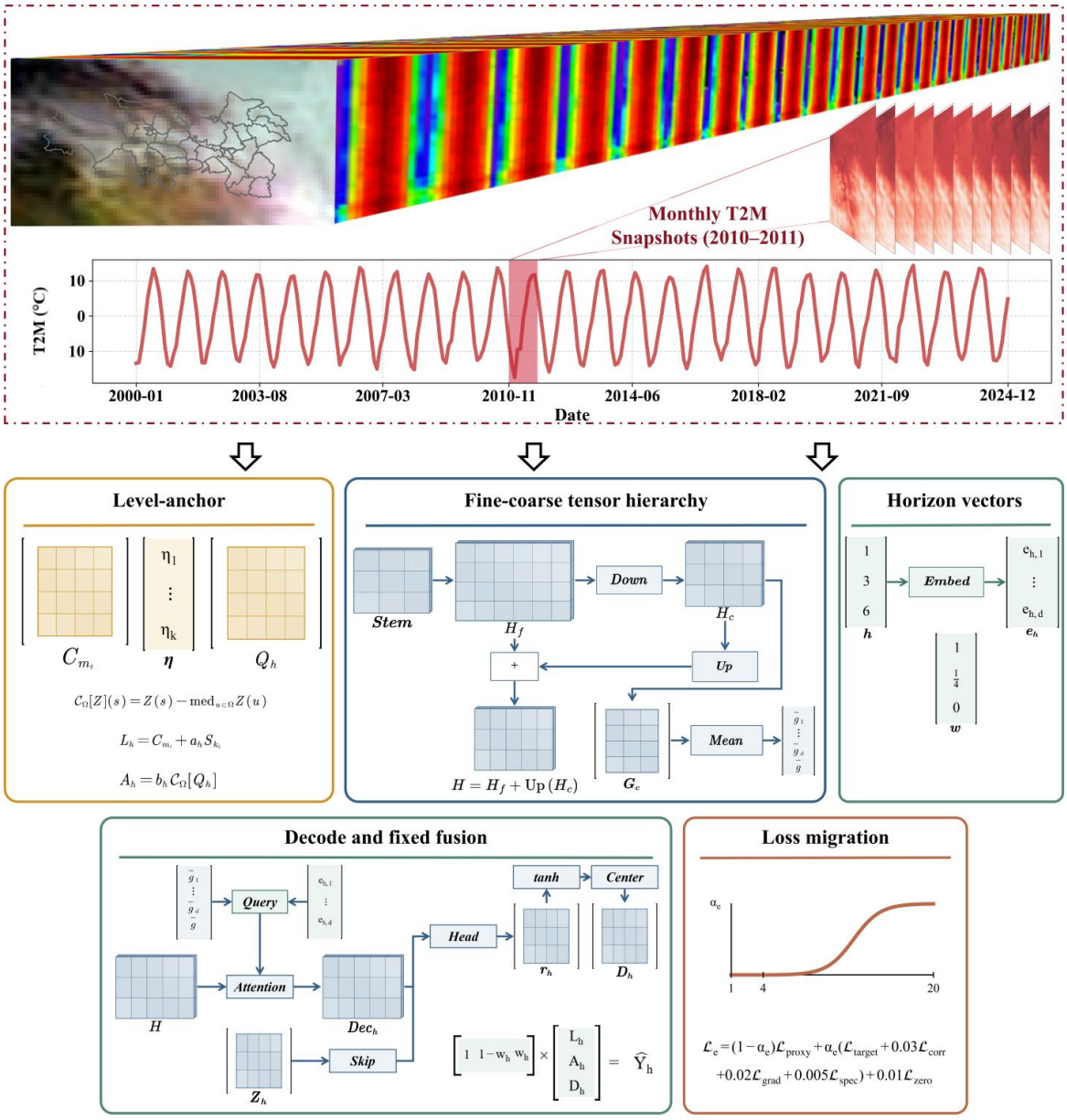
Equation (4) fixes every shape contribution to zero support-wide median before fusion. Centring is performed on the 3,655-cell support; the cropped 1,001-cell county subset is not re-centred.

For a forecast origin o, target calendar month m_t, and lead $h \in \{1,3,6\}$, the horizon-specific causal history set is

$$\mathcal{H}_{k_h}(o, m_t) = \mathrm{Last}_{k_h}\{j \le o : m_j = m_t\} \tag{5}$$

Equation (5) admits only same-calendar-month observations dated no later than the forecast origin. The operator Last_{k_h} returns the k_h most recent eligible indices; m_j is the calendar month at observation index j, and k_h is the number of years retained for horizon h. For each selected history field, its regional median anomaly is

$$\eta_j = \mathrm{med}_{s\in\Omega}\left[Y_j(s) - C_{f,m_j}(s)\right] \tag{6}$$



183

184 **Figure 2.** RegimeTransfer-DCSF architecture and training schedule. Panels show anchoring, tensor
185 decoding, fusion, and loss migration.

186 Equation (6) reduces each causal history field to its displacement from the fold-training seasonal
187 background. The causal regional state carried into the forecast is

$$S_{k_h}(o, m_t) = \frac{1}{k_h} \sum_{j \in \mathcal{H}_{k_h}(o, m_t)} \eta_j \tag{7}$$

188 Equation (7) averages the most recent same-month regional anomalies selected by Equation (5). The
189 regional-level field is then

$$L_h(s) = C_{f,m_t}(s) + a_h S_{k_h}(o, m_t) \tag{8}$$

Equation (8) assigns the seasonal field and domain-wide displacement to the level path; a_h is a fixed horizon coefficient.

Anchor evidence is separately defined for the three leads. At h1, the origin-available raw anchor evidence is

$$Q_1(s) = Y_o(s) - C_{f,m_o}(s) \tag{9}$$

Equation (9) uses the forecast-origin anomaly relative to its own calendar-month background, where m_o is the calendar month of forecast origin o. At h3, the origin field is expressed against the target-month background as

$$Q_3(s) = Y_o(s) - C_{f,m_t}(s) \tag{10}$$

Equation (10) converts persistence at the origin into a target-season-referenced shape. At h6, the held-fixed fold-specific XGBoost anchor supplies

$$Q_6(s) = \hat{Y}_6^{\mathrm{XGB}}(o, s) - C_{f,m_t}(s) \tag{11}$$

Equation (11) expresses the depth-6, 600-tree h6 forecast as an anomaly around the target-month background; its learning rate is 0.03. The centred anchor shape at each horizon is

$$A_h(s) = b_h \mathrm{C}_{\Omega}[Q_h](s) \tag{12}$$

Equation (12) removes the support-wide median of the horizon-specific evidence and applies the fixed anchor coefficient b_h. The held-fixed coefficient triples (k_h, a_h, b_h) were $(2, -0.85, 0.40)$, $(2, -0.65, -0.10)$, and $(5, -1.25, 0.30)$ for h1, h3, and h6, respectively. Every term in Equations (5)–(12) is dated at or before the forecast origin.

2.2.2 Fine–coarse deep-shape expert and horizon-conditioned decoding

The deep branch learns a centred shape correction from the 12-month multivariate history. The input tensor has dimensions batch × 12 months × 12 variables × 44 × 86 cells; its boolean validity mask is concatenated with the predictor channels at each month. Two stride-2 convolutional stem layers reduce each monthly field to an 11 × 22 fine grid with 48 hidden channels. Two four-head fine cuboid-attention blocks use 4 × 6 spatial windows; the second block shifts its window by 2 × 3 tokens. Four learned global vectors exchange information bidirectionally with the fine tokens. A stride-2 hierarchy then produces a 6 × 11 coarse grid, which passes through one four-head cuboid block with a 3 × 5 window. This compact hierarchy adapts Earthformer-style local and global space–time interaction to the centred shape task [13].

Let H_f be the 12-step fine memory and H_c the 12-step coarse memory after their respective cuboid blocks. The operator Up denotes bilinear return to 11 × 22 followed by the learned 3 × 3 channel projection. The fused multiscale memory is

$$H = H_f + \mathrm{Up}(H_c) \tag{13}$$

Equation (13) preserves the fine representation while adding domain-scale context from the coarse branch. Let e_h be the learned embedding of h1, h3, or h6, let $\bar{g}$ be the arithmetic mean of the four final global vectors, and let W_g be their learned linear projection. The decoder query is

$$q_h = e_h + W_g \bar{g} \tag{14}$$

Equation (14) combines the horizon identity with a domain-wide summary of the encoded history. The query is expanded to each fine-grid position and reads the 12-step memory through learned multi-head attention [37]. If H_T denotes the final time slice of the fused memory, the decoded fine-grid feature is

$$Z_h = \mathrm{Attn}(q_h, H, H) + H_T \tag{15}$$

Equation (15) forms the horizon-conditioned feature and retains the latest encoded state through H_T. The attended feature is bilinearly returned to 44 × 86 and passed through a convolutional decoder.

The full-resolution local skip contains only forecast-causal fields. Let P_h be the centred field from the same calendar month one year before the target, expressed relative to the target-month climatology, and let O_h be the centred forecast-origin anomaly. The five-channel descriptor is

$$U_h = \mathrm{Concat}\left(0, P_h, O_h, \nabla_y P_h, \nabla_x P_h\right) \tag{16}$$

Equation (16) gives the decoder direct access to the centred seasonal field, the origin anomaly, and target-season spatial gradients without target-period input; ∇_x and ∇_y are the implemented horizontal and vertical adjacent-cell finite differences, and the first channel is the zero bridge used by the selected centred-shape model. Denote the convolutional decoder by Dec, the two-layer local projection by Skip, group normalization by Norm, and the final 1 × 1 convolution by Head. The unconstrained residual field is

$$r_h = \mathrm{Head}\left(\mathrm{Norm}\left[\mathrm{Dec}(Z_h) + \mathrm{Skip}(U_h)\right]\right) \tag{17}$$

Equation (17) combines the attention feature and causal local skip before the one-channel residual head [38]. The deep shape used at inference is bounded and re-centred as

$$D_h = \mathrm{C}_\Omega[3\tanh(r_h/3)] \tag{18}$$

Equation (18) limits the pointwise correction before enforcing the zero-median shape constraint of Equation (4). The bound is 3 °C in the centred-shape space.

2.2.3 Fixed horizon responsibility and final field

The final field prediction is

$$\hat{Y}_h(s) = L_h(s) + (1 - w_h)A_h(s) + w_h D_h(s) \tag{19}$$

Equation (19) assigns fixed deep-shape weights $(w_1, w_3, w_6) = (1, 0.25, 0)$. The h1 forecast is $L_1 + D_1$; h3 is $L_3 + 0.75A_3 + 0.25D_3$; and h6 is exactly $L_6 + A_6$. These weights are shared by all samples and are not input-dependent gates. The schedule gives the learned expert complete short-lead shape responsibility, blends it with the causal anchor at h3, and transfers full h6 shape responsibility to the anchor.

2.2.4 Proxy-to-target loss migration

The selected model was optimized for 20 fixed epochs. Epochs 1–4 used the centred forecast-origin proxy shape alone. For epochs 5–20, let p_e increase linearly from 0 to 1, and let $g(x) = (1 + \exp[-x])^{-1}$ be the logistic function. The target-loss weight is

$$\alpha_e = \frac{g\left(12[p_e - \frac{1}{2}]\right) - g(-6)}{g(6) - g(-6)} \tag{20}$$

Equation (20) increases target responsibility smoothly from zero at epoch 5 to one at epoch 20; $\alpha_e = 0$ during epochs 1–4.

Value errors were standardized separately by horizon. Let V_h contain all valid training target-shape values at lead h. The robust scale is

$$\sigma_h = \max\left\{1.4826\,\mathrm{med}_{v \in V_h}\left|v - \mathrm{med}_{u \in V_h} u\right|, 0.1\right\} \tag{21}$$

Equation (21) uses a median-absolute-deviation scale with a 0.1 °C floor. For residual x and transition parameter β, the Smooth-L1 penalty is

$$\varphi_\beta(x) = \begin{cases} \dfrac{x^2}{2\beta}, & |x| < \beta \\ |x| - \dfrac{\beta}{2}, & |x| \geq \beta \end{cases} \tag{22}$$

Equation (22) supplies $\beta = 1$ for value losses and $\beta = 0.5$ for adjacent-gradient losses.

For batch row i, let $M_i(s)$ be the valid-cell mask, $n_i = \sum_s M_i(s)$, h_i its lead, $\tilde{D}_i$ the pre-centring bounded model shape, and R_i either the centred origin proxy or centred future target shape. With batch size B, the horizon-scaled shape-value loss is

$$\mathcal{L}_{\text{val}}(R) = \frac{1}{B}\sum_{i=1}^{B} \frac{\sigma_{h_i}}{n_i} \sum_{s\in\Omega} M_i(s)\varphi_1\left(\frac{\tilde{D}_i(s) - R_i(s)}{\sigma_{h_i}}\right) \tag{23}$$

Equation (23) defines L_{proxy} when R_i is the origin shape and L_{target} when $R_i = T_i$, the future target shape. All selected-model training rows have equal weight.

Three target-shape terms retain spatial organization. Spatial co-organization is encouraged by

$$\mathcal{L}_{\text{corr}} = \frac{1}{B}\sum_{i=1}^{B}\left[1 - \text{corr}_{\Omega}\left(\tilde{D}_i, T_i\right)\right] \tag{24}$$

Equation (24) penalizes loss of masked Pearson correlation between the predicted and future centred shapes. Let Ω_x and Ω_y be the valid horizontal and vertical adjacent-cell pairs, and let angle brackets denote the arithmetic mean over the indicated pairs. Local differences are preserved through

$$\mathcal{L}_{\text{grad}} = \frac{1}{2B}\sum_{i=1}^{B}\left[\left\langle\varphi_{0.5}(\nabla_x\tilde{D}_i - \nabla_x T_i)\right\rangle_{\Omega_x} + \left\langle\varphi_{0.5}(\nabla_y\tilde{D}_i - \nabla_y T_i)\right\rangle_{\Omega_y}\right] \tag{25}$$

Equation (25) compares adjacent valid-cell differences in both grid directions. Let A_4 denote 4 × 4 average pooling with stride 4, let F be the orthonormal two-dimensional real Fourier transform, and let $\odot$ denote elementwise multiplication. Coarse spatial-scale amplitude is constrained by

$$\mathcal{L}_{\text{spec}} = \frac{1}{B}\sum_{i=1}^{B}\left\langle\left|\left|\mathcal{F}\left[\text{A}_4(M_i \odot \tilde{D}_i)\right]\right| - \left|\mathcal{F}\left[\text{A}_4(M_i \odot T_i)\right]\right|\right|\right\rangle \tag{26}$$

Equation (26) compares pooled Fourier amplitudes rather than phase. Frequency-domain reconstruction losses and decomposed frequency representations provide precedents for complementing pointwise objectives with spectral structure [26,39]. The outer angle brackets denote the arithmetic mean over retained spectral coefficients. Residual leakage into the regional level is penalized by

$$\mathcal{L}_{\text{zero}} = \frac{1}{B}\sum_{i=1}^{B}\left|\frac{1}{n_i}\sum_{s\in\Omega} M_i(s)\tilde{D}_i(s)\right| \tag{27}$$

Equation (27) keeps the learned shape close to zero arithmetic mean during optimization, complementing the inference-time median centring in Equation (18). The complete epoch-dependent objective is

$$\begin{aligned}\mathcal{L}_e &= (1-\alpha_e)\mathcal{L}_{\text{proxy}} \\ &+\alpha_e\left(\mathcal{L}_{\text{target}} + 0.03\mathcal{L}_{\text{corr}} + 0.02\mathcal{L}_{\text{grad}} + 0.005\mathcal{L}_{\text{spec}}\right) + 0.01\mathcal{L}_{\text{zero}}\end{aligned} \tag{28}$$

Equation (28) transfers value responsibility from the causal proxy to the future target while retaining the three structural terms and zero-level constraint. Optimization used AdamW with a learning rate of

$4 \times 10{-}4$, weight decay of $10{-}4$, batch size 4, and gradient clipping at 1. No early stopping or within-schedule checkpoint selection was applied.

2.3 Comparators and evaluation

2.3.1 Comparator hierarchy and training protocol

The benchmark contained DCSF, seasonal climatology, XGBoost, and seven learned time-series models: iTransformer, TimeMixer, ModernTCN, TimeXer, MoSSL, TimeFilter, and Timer-XL. XGBoost represented scalable tree boosting [40]. The learned set represented variate-token attention, decomposable multiscale mixing, modern temporal convolution, exogenous-variable attention [14–17], multimodal spatiotemporal self-supervision, patch-specific graph filtration, and long-context causal forecasting [18–20]. Seven deep baselines were evaluated over two folds and five seeds, producing 70 fold–seed runs. The complete ten-method leaderboard used full-support scalar evidence, with fold-specific scores and run-to-run dispersion retained in Supplementary Table S1. Verified implementations and configurations are reported in Supplementary Table S2. Each learned baseline used one prespecified architecture-specific configuration under the shared five-seed rolling evaluation; DCSF selection compared the ten candidates in Supplementary Table S3. County-domain distributional analyses used the selected DCSF fit, seasonal climatology, and the available learned-model grid ensembles; ModernTCN and TimeFilter figures use their five-seed ensemble forecasts.

The two rolling-origin folds formed the development-validation setting for the anchor specification, candidate selection, and rolling-validation scores. Ten DCSF candidates spanned target-only, abrupt, linear, cosine, sigmoid, fixed-mixture, no-structure, recency-weighted, compact, and larger variants (Supplementary Table S3). The 48-channel sigmoid-migration model achieved the lowest validation MAE and defined the final specification. Four separately trained fold–seed refits, using seeds 2026 and 2027 in each fold, reused the same temporal folds to measure optimization-seed stability. A separate 50-configuration sensitivity matrix perturbed level, anchor, sign, component, and horizon responsibility around the selected specification (Supplementary Table S4).

2.3.2 Absolute accuracy and paired transition evaluation

Absolute accuracy was summarized by MAE, root-mean-square error (RMSE), bias, R^2, median absolute error, 90th-percentile absolute error, and the fractions within 1 and 2 °C. MAE directly reports average absolute deviation on the target scale, while the complementary metrics retain distributional and comparative information [41,42]. Table 2A reports pooled unweighted metrics across all 114,114 county grid-cell predictions. Case-wise regional metrics first applied cosine-latitude weights over the 1,001 cells and were then averaged across the 114 horizon-specific cases. These estimands remain separate. For the h1 component diagnostics, regional-level absolute error was the absolute cosine-latitude-weighted county-mean prediction error; centred-shape correlation was calculated after removing each county field's weighted mean; and grid MAE used the same county-area weights.

Comparator gain was defined as comparator MAE minus DCSF MAE, so positive values favour DCSF. Transition comparisons matched models within fold, target month, and horizon. Paired uncertainty used 20,000 fold-stratified circular moving-block bootstrap replicates over each complete 18-month classifiable h1 sequence, retaining temporal order and each target month's complete spatial field [43,44]. The primary block length was three months; two- and six-month blocks provided sensitivity checks (Supplementary Table S5). The prespecified 3×3 thermal-state threshold grid supplied a parallel definition check (Supplementary Table S6). Win counts indicate target months in which DCSF had lower area-weighted MAE.

To characterize the spatial setting of transition gains using only origin-available climate history, grid centres were divided within each fold into equal-count tertiles by the ordinary-least-squares slope of annual-mean ERA5-Land T2M over that fold's training period: 2000–2014 for rolling A and 2000–2016 for rolling B. TimeFilter-minus-DCSF gains were recalculated within each fold-specific tertile for the same 25 h1 changing target months. The high-minus-low contrast used the same fold-stratified moving-block bootstrap. A fixed 2000–2014 trend and the retrospective 2000–2024 trend supplied sensitivity definitions (Supplementary Table S7). Cellwise correlations and quantile-bin medians were treated as spatial descriptions over fold–cell memberships.

2.3.3 Reproducibility evidence

The reproducibility record comprises the four fold–seed refits, the ten-candidate selection study, and the 50-configuration level–anchor sensitivity matrix. Machine-readable source data accompany every display object. The 114,114-row county comparison dataset retains targets, predictions, comparator errors, geography, thermal state, and warming strata; code, configurations, trained parameter states, and intermediate analyses are identified in the Data availability statement.

2.4 Retrospective warm-footprint and physical-context analyses

2.4.1 Warm and cold footprints

The retrospective footprint analysis used all 300 ERA5-Land months from 2000–2024. For each county grid centre and calendar month, the fixed 2000–2018 reference period defined its own climatological mean and the q10 and q90 of anomalies around that mean. Warm and cold fractions followed Equations (2) and (3), with these fixed-reference quantities replacing their fold-training counterparts. Annual statistics included mean and maximum warm-area fraction and counts of regional-or-widespread (at least 0.20) and widespread (at least 0.50) warm months.

Trend magnitudes were estimated with Theil–Sen slopes [45]. For the four annual series, monotonic association was evaluated with the Hamed–Rao modified Mann–Kendall test, which adjusts the variance for significant autocorrelation in the detrended rank series [46,47]. Autocorrelation-aware uncertainty used 20,000 circular block-bootstrap replicates of annual residuals around the Theil–Sen line [44]. The primary block length was three years; two- and five-year blocks supplied sensitivity checks (Supplementary Table S8). Seasonal rows retained the conventional Mann–Kendall test and Sen interval. Fractions were converted to percentage points per decade. Seasonal summaries used DJF, MAM, JJA, and SON; early and recent periods were 2000–2009 and 2015–2024.

2.4.2 Thermal-state chronology and widespread-warm episodes

The 48 distinct validation target months were ordered within each fold to describe residence and switching. Warm connected components used eight-neighbour adjacency. For every widespread-warm target month, the analysis recorded the incoming state, absolute regional-anomaly jump, warm-area fraction, largest-component share, centroid, spatial texture, and outgoing state when available. The six months were identified from the ERA5-Land county-domain state definition before forecast errors were examined.

2.4.3 Physical context and selected spatial realizations

For each validation target month, D2M, SKT, SSR, TP, EVP, and wind speed were standardized using fold-training cell- and calendar-month statistics and then averaged over the county domain. Regional- and widespread-warm months (n=11) were compared descriptively with the remaining months (n=37). Four spatial realizations followed one prespecified display rule: h1 warming transitions in which DCSF

had lower county-area-weighted MAE than both ModernTCN and TimeFilter. These cases show the spatial form of the aggregate transition advantage.

3 Results

3.1 DCSF achieves high absolute accuracy and benchmark leadership

The selected DCSF fit delivered high absolute field accuracy. Across 114,114 county-domain grid-cell predictions, its pooled MAE was 1.131 °C, RMSE was 1.494 °C, bias was −0.277 °C, and R^2 was 0.9767 (Table 2A). Median and P90 absolute errors were 0.872 and 2.420 °C, respectively; 55.39% of predictions were within 1 °C and 83.68% were within 2 °C. The fold-wise two-refit county ensemble retained a pooled MAE of 1.135 °C, RMSE of 1.501 °C, R^2 of 0.9765, and 83.42% of errors within 2 °C, placing the county-domain absolute result on a seed-aggregated footing. Fold-wise predictions remained concentrated near the ERA5-Land target 1:1 line at h1, h3, and h6 (Figure 3).

Table 2. County absolute accuracy, complete ten-method benchmark leadership, parameter counts, and fold–seed DCSF refit stability. Panel A separates pooled-grid and area-weighted case MAE, panel B uses equal-fold means, panel C reports refits, and the DCSF count excludes the fixed XGBoost h6 anchor.

A. County-domain absolute accuracy

Model	Grid n	Pooled MAE (°C)	Case-mean MAE (°C)	RMSE (°C)	R^2	Bias (°C)	Median AE (°C)	P90 AE (°C)	Within 1°C	Within 2°C
DCSF selected	114,114	1.131	1.131	1.494	0.9767	-0.277	0.872	2.420	55.39%	83.68%
DCSF 2-seed	114,114	1.135	1.135	1.501	0.9765	-0.279	0.873	2.429	55.25%	83.42%
Seasonal climatology	114,114	1.256	1.256	1.669	0.9709	-0.093	0.929	2.780	53.17%	79.05%
TimeFilter 5-seed	114,114	1.420	1.420	1.860	0.9639	+0.127	1.110	3.074	45.50%	76.01%
ModernTCN 5-seed	114,114	1.423	1.423	1.863	0.9637	+0.067	1.122	3.023	45.11%	75.92%

B. Complete 3,655-cell rolling-validation leaderboard

Rank	Method	Equal-fold MAE (°C)	Trainable neural parameters	Runs	DCSF MAE reduction (%)
1	DCSF	1.1741	239,809	4	—
2	Seasonal climatology	1.2587	—	1	6.72
3	TimeFilter	1.4965	4,206	10	21.55
4	iTransformer	1.5031	9,222	10	21.89
5	TimeXer	1.5043	17,446	10	21.95
6	TimeMixer	1.5182	2,049	10	22.67
7	ModernTCN	1.5293	1,332,746	10	23.23
8	XGBoost	1.5369	—	1	23.61
9	Timer-XL	1.5777	8,942	10	25.58

Rank	Method	Equal-fold MAE (°C)	Trainable neural parameters	Runs	DCSF MAE reduction (%)
10	MoSSL	2.2985	112,881	10	48.92

C. DCSF fold–seed refit stability

Fold	Seed / summary	MAE (°C)	Range (°C)
Rolling A	2026	1.1755	—
Rolling A	2027	1.1722	—
Rolling B	2026	1.1719	—
Rolling B	2027	1.1768	—
Across folds	4 refits	1.1741 ± 0.0024	1.1719–1.1768

Benchmark context reinforced the absolute result. Across the 114 county-area-weighted forecast cases, mean MAE was 1.1307 °C for the selected DCSF fit, 1.2557 °C for seasonal climatology, 1.4195 °C for TimeFilter, and 1.4230 °C for ModernTCN. DCSF therefore reduced area-weighted mean case MAE by 9.95% relative to seasonal climatology and by 20.35% relative to TimeFilter. The separately pooled unweighted grid MAEs reported in Table 2A were 1.1308, 1.2558, 1.4197, and 1.4231 °C, respectively. On the complete ten-method, 3,655-cell rolling-validation leaderboard (Table 2B), the four-refit DCSF mean of 1.1741 °C ranked first and was 21.55% lower than TimeFilter's 1.4965 °C; fold-specific means and run-to-run dispersion are retained in Supplementary Table S1.

Optimization-seed stability was exceptionally tight. The four separately trained full-support refits ranged from 1.1719 to 1.1768 °C, with a mean of 1.1741 °C and a standard deviation of 0.0024 °C (Table 2C). The selected DCSF fit also retained the lowest P90 absolute error in all six fold-by-horizon combinations against both ModernTCN and TimeFilter (Figure 4). The complete ECDFs localize the most consistent separation to the upper-error tail.

3.2 The advantage concentrates in changing thermal states

3.2.1 Matched gains during h1 state changes

The clearest application advantage occurred when the ERA5-Land target thermal state changed. Across 25 h1 changing target months, DCSF achieved an area-weighted MAE of 1.219 °C compared with 1.520 °C for TimeFilter, a reduction of 0.301 °C (19.79%); DCSF won 17 of 25 months, and the three-month moving-block bootstrap interval was 0.073–0.526 °C (Table 3A). Relative to seasonal climatology, the gain was 0.218 °C (15.16%; 18/25 wins; interval 0.081–0.355 °C). The mean gain over ModernTCN was 0.221 °C (15.37%; 14/25 wins), with an interval of 0.019–0.405 °C. The h1 advantage was specifically concentrated in change: the mean TimeFilter-minus-DCSF MAE difference was 0.301 °C across the 25 state changes but 0.002 °C across the 11 classifiable non-changing months. Their difference was +0.299 °C, with 92.7% of moving-block replicates above zero. The first h1 target in each fold had no preceding target-month state and was not assigned to either group.

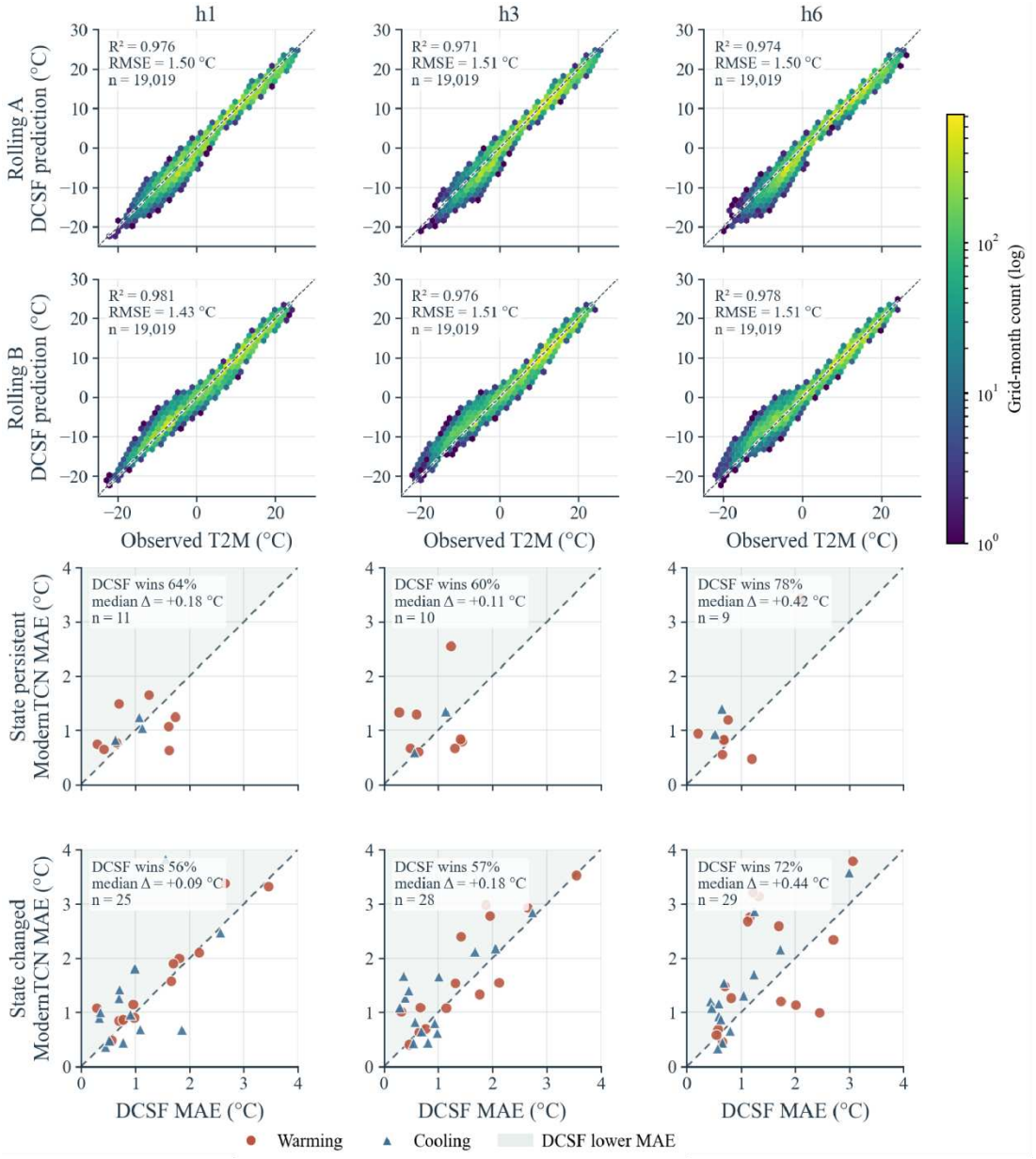


Figure 3. Absolute DCSF agreement and paired gains over ModernTCN across horizons and thermal-state conditions. Columns are h1, h3, and h6; rows show fold-wise predicted-versus-target ERA5-Land T2M and persistent/changing target-month MAE pairs.

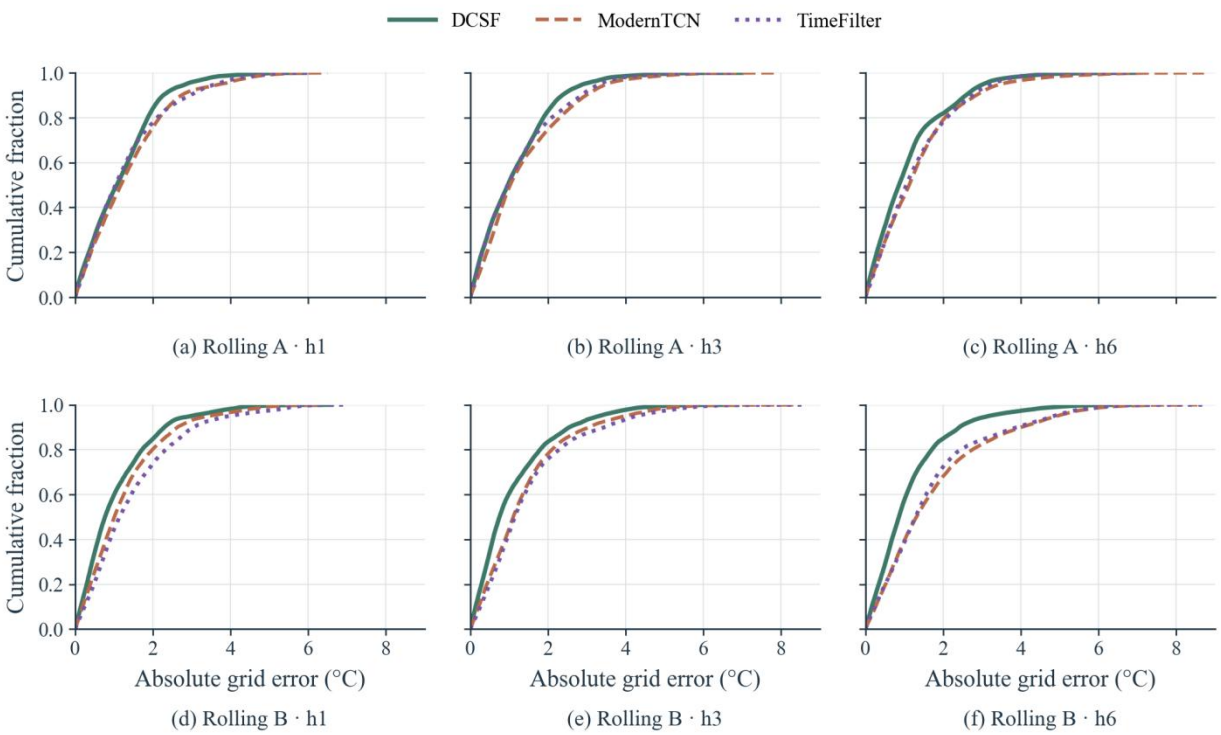


Figure 4. Complete absolute grid-error distributions across folds and forecast horizons. ECDFs compare DCSF, ModernTCN, and TimeFilter through the full ERA5-Land target-error tail.

Both transition directions retained substantial gains. Relative to TimeFilter, DCSF reduced MAE by 0.294 °C across 12 warming transitions (16.61%; 8/12 wins) and by 0.307 °C across 13 cooling transitions (23.82%; 9/13 wins). Across the full 3 × 3 state-threshold grid, the number of h1 changing months ranged from 23 to 27 and every TimeFilter-relative all-changing gain remained positive, spanning 0.235–0.316 °C (Supplementary Table S6).

Table 3. Matched DCSF gains during h1 transitions and across training-period warming strata. Panel A uses fold-stratified block bootstrap; panel B reports 2,002 fold–cell memberships and the high-minus-low contrast.

A. Matched h1 transition gains

Condition	Comparator	n	DCSF MAE	Comp. MAE	Gain (°C)	Reduction (%)	Wins	95% block CI (°C)
All state-changing months	ModernTCN 5-seed ensemble	25	1.219	1.440	+0.221	+15.37	14/25	[+0.019, +0.405]
All state-changing months	TimeFilter 5-seed ensemble	25	1.219	1.520	+0.301	+19.79	17/25	[+0.073, +0.526]
All state-changing months	Seasonal climatology	25	1.219	1.437	+0.218	+15.16	18/25	[+0.081, +0.355]
Warming transitions	TimeFilter 5-seed ensemble	12	1.474	1.768	+0.294	+16.61	8/12	[+0.043, +0.694]
Cooling transitions	TimeFilter 5-seed ensemble	13	0.983	1.291	+0.307	+23.82	9/13	[-0.039, +0.706]

B. Fold-training-only local-warming gradient

Warming stratum	Mean trend	Fold–cell n	Unique cells	n	DCSF MAE	TF MAE	Gain (°C)	Reduction (%)	Cell-month wins / contrast CI
Low	0.031	668	410	25	1.223	1.493	+0.270	+18.10	61.28%
Middle	0.169	666	473	25	1.223	1.527	+0.303	+19.87	63.68%
High	0.319	668	407	25	1.210	1.539	+0.329	+21.36	68.19%
High − low	—	—	—	25	—	—	+0.058	—	95% CI [-0.068, +0.179]

3.2.2 Level–shape diagnostics during h1 state changes

The component diagnostics localized the selected fit's transition advantage (Figure 5). Across the same 25 h1 state changes, DCSF had lower regional-level absolute error than ModernTCN in 14 months, with a median paired difference of −0.192 °C. Centred-shape correlation was higher in 22 months, with a median gain of 0.0035, while grid MAE was lower in 14 months, with a median paired difference of −0.0916 °C. The frequent shape-correlation gain, together with the aligned median level and grid-error differences, is consistent with the intended separation of regional displacement and centred spatial organization.

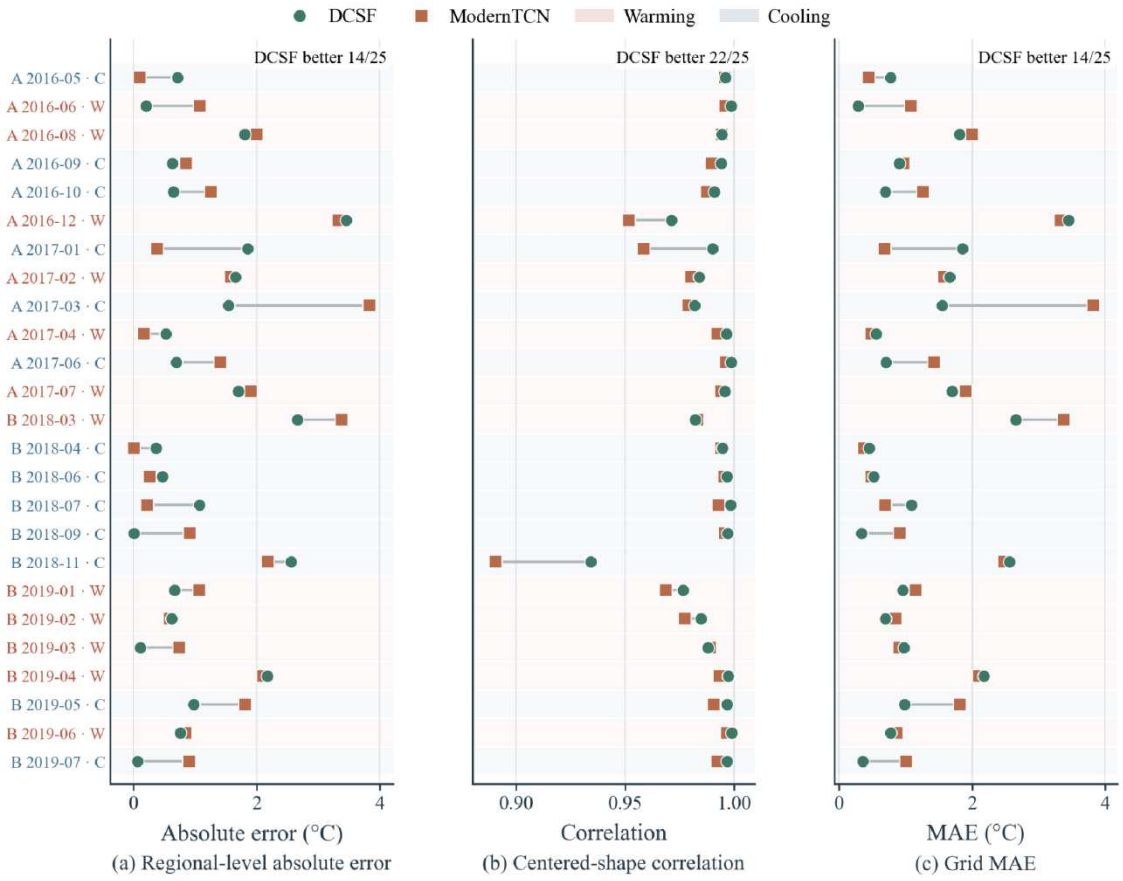


Figure 5. Regional-level, centred-shape, and gridwise recovery during h1 state changes. Paired dumbbells compare DCSF and ModernTCN for each changing target month; W and C denote warming and cooling.

3.3 The county-domain warm footprint is expanding

3.3.1 Long-term expansion of warm extent

The ERA5-Land regional thermal field changed markedly during 2000–2024. Relative to the fixed 2000–2018 reference thresholds, annual mean warm-area fraction increased by 9.29 percentage points per decade, with a three-year circular block-bootstrap interval of 2.72–15.52 percentage points per decade and a Hamed–Rao modified Mann–Kendall p value of 0.00545. The Hamed–Rao variance factor was 1.000 because no detrended-rank autocorrelation lag exceeded the implemented 95% bound. Its mean rose from 9.99% in 2000–2009 to 22.46% in 2015–2024 (Figure 6; Table 4A). The positive mean-footprint trend persisted with two- and five-year block lengths, whose interval lower bounds were 3.22 and 1.92 percentage points per decade. Annual maximum warm-area fraction increased by 15.49 percentage points per decade. The frequency signal moved in the same direction: widespread-warm months increased by 0.91 months per decade, while regional-or-widespread warm months increased by 1.43 months per decade. The strongest seasonal warm-footprint slopes occurred in SON (+11.49 percentage points per decade) and DJF (+9.03 percentage points per decade).

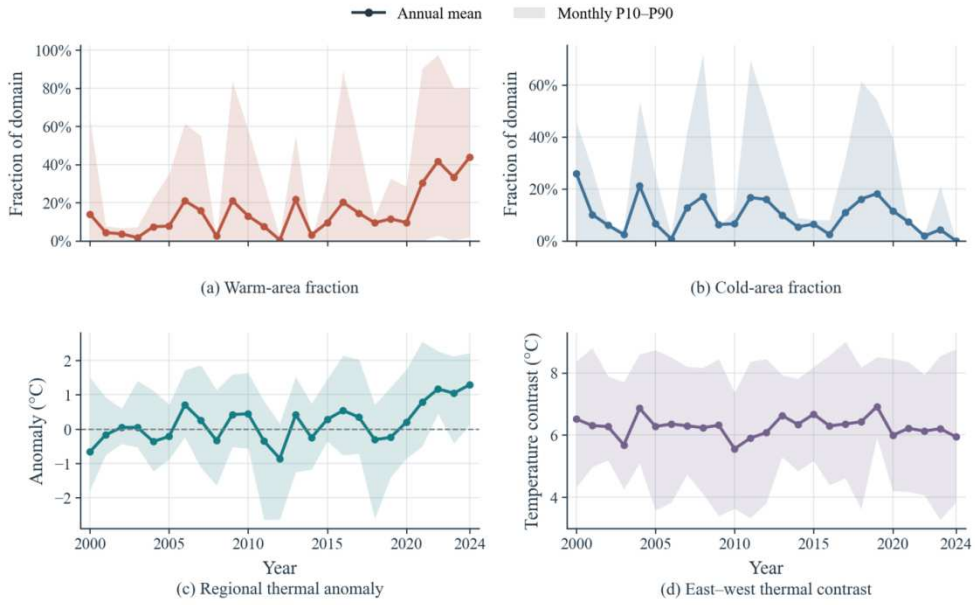


Figure 6. Evolution of ERA5-Land regional thermal structure from 2000 to 2024. Annual means and monthly P10–P90 intervals show warm and cold footprints, regional anomaly, and east–west contrast.

Table 4. County-domain warm-footprint trends during 2000 – 2024 and six widespread-warm validation episodes. Annual rows use Sen/Hamed – Rao; seasonal rows use Mann – Kendall/Sen. Early/Recent denote 2000 – 2009/2015 – 2024.

A. County-domain warm-footprint trends

Scope	Metric	Trend / decade	95% CI	τ	p	**Early (2000–2009)**	**Recent (2015–2024)**
Annual	Mean warm-area fraction	+9.29 pp	[2.72, 15.52]	0.400	0.00545	9.99%	22.46%
Annual	Annual maximum warm area	+15.49 pp	[1.94, 28.09]	0.317	0.028	58.21%	84.59%

Scope	Metric	Trend / decade	95% CI	τ	p	**Early (2000–2009)**	**Recent (2015–2024)**
Annual	Widespread-warm months	+0.91 mo	[0.00, 1.71]	0.377	0.015	1.00 mo	2.40 mo
Annual	Regional-or-widespread warm months	+1.43 mo	[0.17, 2.58]	0.384	0.012	2.00 mo	4.00 mo
DJF	Mean warm-area fraction	+9.03 pp	[2.06, 17.57]	0.372	0.00946	8.51%	20.88%
MAM	Mean warm-area fraction	+1.89 pp	[-0.08, 14.33]	0.204	0.163	12.83%	19.87%
JJA	Mean warm-area fraction	+2.93 pp	[0.00, 13.73]	0.218	0.136	7.73%	20.07%
SON	Mean warm-area fraction	+11.49 pp	[1.30, 23.80]	0.343	0.017	10.89%	29.02%

B. County-domain widespread-warm episode dynamics

Target month	Fold	Incoming state	Incoming jump (°C)	Warm area (%)	Largest component (%)	Next state
2016-08	Rolling A	Near climatology	1.838	98.13	100.00	Regional warm
2016-12	Rolling A	Near climatology	3.101	98.48	100.00	Regional warm
2017-07	Rolling A	Regional cold	2.686	95.22	100.00	Near climatology
2017-12	Rolling A	Near climatology	1.451	74.48	100.00	End of fold
2018-03	Rolling B	Near climatology	4.821	95.96	100.00	Near climatology
2019-04	Rolling B	Regional cold	3.009	90.21	99.66	Widespread cold

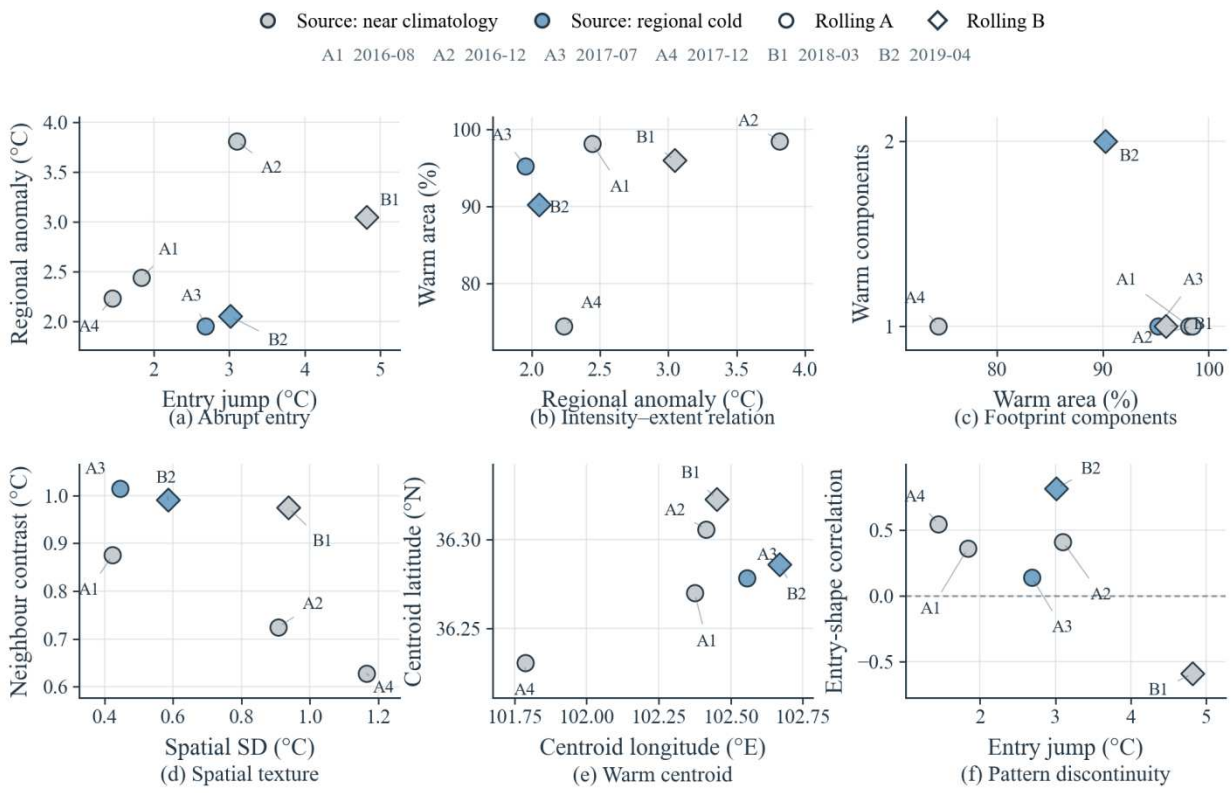
3.3.2 Widespread-warm target months are spatially coherent

Six of the 48 distinct validation target months met the county-domain widespread-warm definition: August and December 2016, July and December 2017, March 2018, and April 2019 (Table 4B). Their mean incoming regional-anomaly jump was 2.818 °C, and warm area covered an average of 92.08% of the county domain. On average, 99.94% of the warm area belonged to the largest eight-neighbour component. The monthly fingerprints therefore describe connected regional expansion rather than scattered isolated pixels (Figure 7).

The six target months entered from different background states—four from near climatology and two from regional cold—and spanned distinct spatial textures, centroids, and pattern discontinuities. All five with a following ERA5-Land target month exited the widespread-warm state, returning to regional warm, near climatology, or widespread cold. This entry–expansion–exit structure identifies widespread warmth as a monthly transition pulse embedded in a dynamic state sequence.

3.3.3 Thermal-state residence and switching

Among the 48 distinct validation target months, 25 were near climatology, five regional warm, six widespread warm, seven regional cold, and five widespread cold. State labels changed in 32 of 46 within-fold month-to-month transitions (69.57%), and the median absolute regional-anomaly jump was 0.959 °C. Figure 8 retains the complete residence and switching order, while Supplementary Figure S1 provides the full transition-count matrix.



478

479 **Figure 7.** Entry and spatial fingerprints of six widespread-warm validation target months. Matched
480 scatter panels encode entry jump, warm extent, footprint structure, texture, centroid, and pattern
481 continuity.

482 **3.4 Warming-transition gains track training-period local-warming gradients**

483 Table 3B summarizes all 25 h1 state changes. The TimeFilter-relative transition gain retained a
484 monotonic low-to-high increase when the warming strata were defined only from each fold's training
485 period. Mean gain rose from 0.270 °C in the low tertile to 0.303 °C in the middle tertile and 0.329 °C
486 in the high tertile; the corresponding relative reductions were 18.10%, 19.87%, and 21.36%. The paired
487 high-minus-low estimate was +0.058 °C, with 81.7% of three-month moving-block replicates above
488 zero and an interval of −0.068–0.179 °C. The same monotonic ordering was retained with one fixed
489 pre-validation 2000–2014 trend definition (0.267, 0.307, and 0.328 °C).

490 Figure 9 isolates the 12 warming transitions and reveals the clearest fold-replicated spatial gradient.
491 The correlation between training-period annual trend and TimeFilter-minus-DCSF gain was positive
492 in rolling A (r=0.378; 1,001 cells) and rolling B (r=0.333; 1,001 cells). After removing each fold's
493 mean trend and mean gain, the combined 2,002 fold–cell memberships retained a positive correlation
494 (r=0.355), and their quantile-bin medians traced the same upward relation.



Figure 8. ERA5-Land thermal-state residence and switching across rolling folds. Coloured runs retain monthly residence and connectors show the complete within-fold switching order.

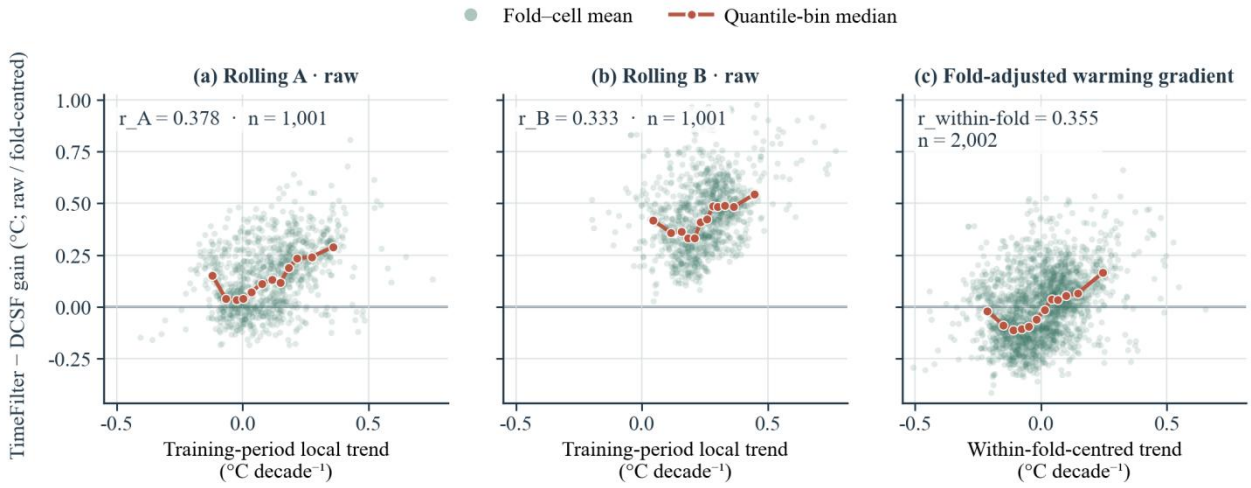


Figure 9. Fold-replicated association between training-period local warming and DCSF advantage during h1 warming transitions. Panels show Rolling A, Rolling B, and pooled within-fold-centred gains; n = 1,001, 1,001, and 2,002.

3.5 Warm states occupy a coherent physical context

ERA5-Land warm target months occupied a distinct multivariate context (Figure 10). Relative to the other 37 target months, the 11 regional- or widespread-warm months showed the largest median standardized contrasts in SKT (+1.540), SSR (+1.035), TP (−1.052), and D2M (+0.722). EVP was modestly higher (+0.250), whereas wind speed differed less (−0.159). These fold-standardized values describe contemporaneous ERA5-Land context and were not supplied to the forecast at the target month.

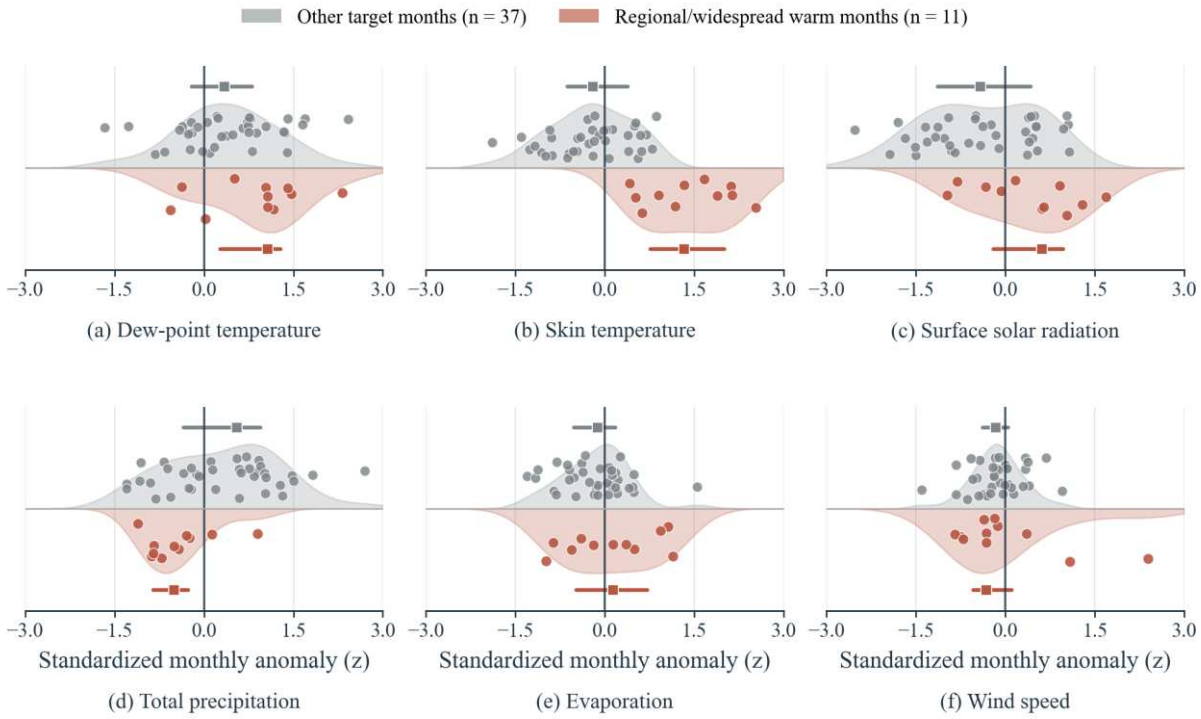


Figure 10. Contemporaneous ERA5-Land context of regional and widespread warm target months. Split distributions compare standardized dew-point, skin-temperature, radiation, precipitation, evaporation, and wind anomalies with other months.

3.6 Spatial realizations show where short-lead gains occur

The selected h1 transitions show how the scalar advantage appears in space (Figure 11). Each column begins with the previous and target ERA5-Land anomaly fields, followed by DCSF, ModernTCN, and TimeFilter forecasts and the local ModernTCN-minus-DCSF absolute-error gain. The four targets—June and August 2016, July 2017, and March 2018—span rapid warming and entry into widespread warmth. DCSF more closely retained both the regional displacement and the county-scale arrangement of positive and negative anomalies, producing coherent local error reductions across major parts of the domain.

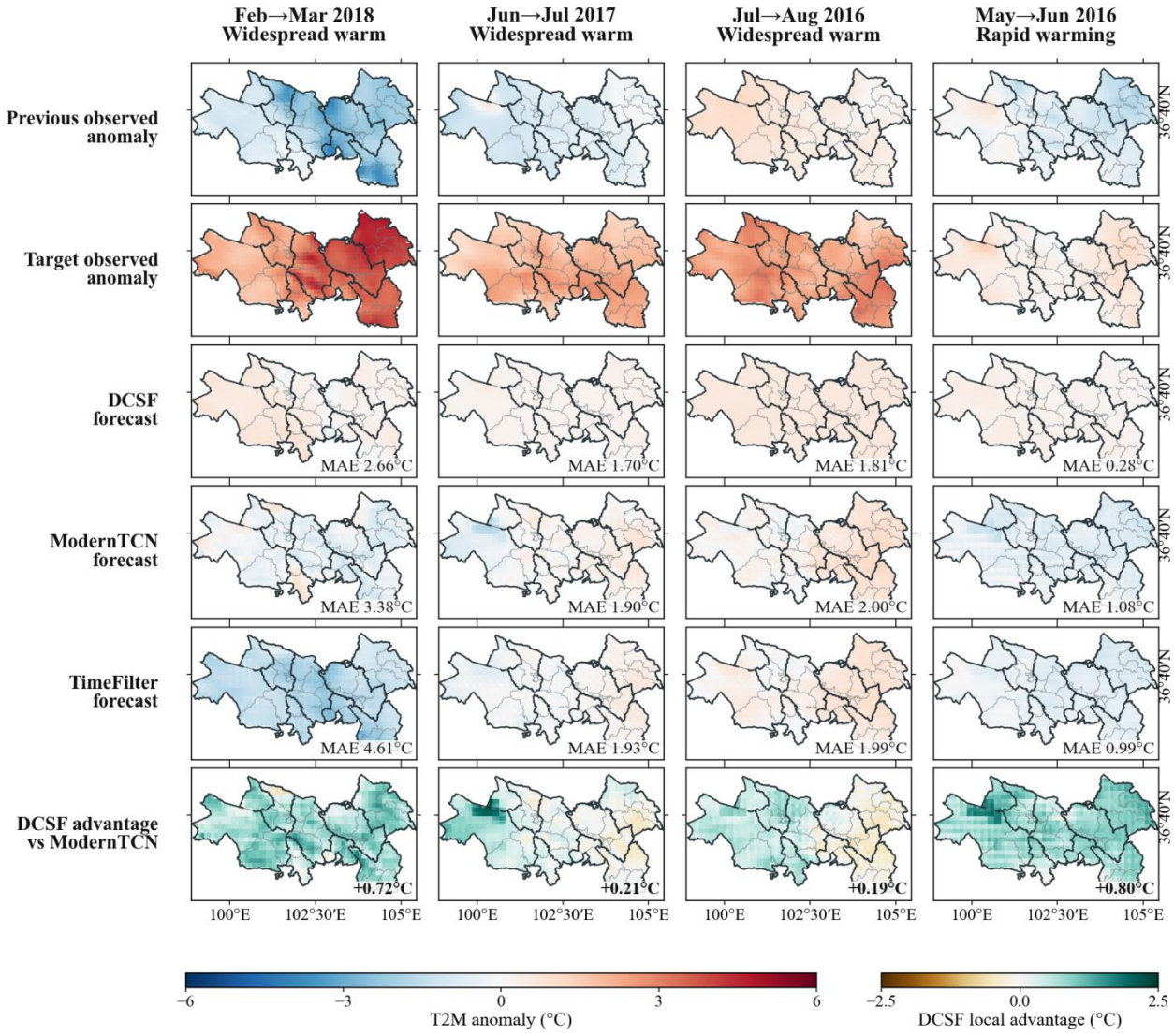


Figure 11. Spatial realization of DCSF advantage across four h1 warming transitions. Columns show four cases; rows show preceding and target anomalies, three forecasts, and local ModernTCN-minus-DCSF error gain.

4 Discussion

4.1 Why separating level and centred shape improves multihorizon forecasts

The central methodological result is that regional displacement and within-domain spatial organization become more predictable when they are assigned distinct responsibilities. Seasonal climatology was the strongest compatible non-DCSF county-domain comparator, which demonstrates how much of monthly T2M is carried by a stable annual spatial field. A single flexible path must preserve that field, estimate a changing regional offset, and reconstruct local departures simultaneously. DCSF instead treats the seasonal field and regional state as the level term in Equation (8), then confines both anchor and deep corrections to centred shape. This separation protects the recurring seasonal backbone while reserving learned capacity for the part of the field that genuinely changes.

Median centring provides the identifiability constraint that makes this division operational. Because Equations (4), (12), and (18) remove the support-wide median from both shape paths, neither shape contribution can reproduce the regional displacement already represented by the level term. The h1 diagnostics align with this mechanism: DCSF achieved higher centred-shape correlation than ModernTCN in 22 of 25 thermal-state changes, while median paired regional-level error and final grid MAE also favoured DCSF. The method therefore gains accuracy through coordinated recovery of two interpretable field properties, not through a scalar correction that leaves the spatial anomaly unspecified.

The fine–coarse hierarchy supplies complementary spatial context within the deep-shape responsibility. Fine 4×6 cuboids retain local organization, the shifted second block exchanges information across the initial window boundaries, and the 6×11 coarse branch broadens the receptive context. Four global vectors then connect the local memories to a domain-wide summary used by the horizon query. This is a focused use of Earthformer-style cuboid attention [13]: the attention hierarchy models the centred residual, while the seasonal field and regional displacement remain explicit outside it. The resulting division concentrates the deep representation on spatial shape and gives each internal scale a clear forecasting role.

The fixed horizon schedule adds a second layer of specialization. At h1, recent spatial information supports complete use of the deep shape; at h3, the prediction blends one-quarter deep shape with three-quarters anchor shape; at h6, the level-plus-anchor path is exact. The schedule therefore changes which evidence carries shape responsibility as lead increases, instead of applying one learned correction uniformly at all leads. This principle is related to the component specialization used by TimeMixer across temporal scales [15], but here the decomposition is spatial and the responsibility shift is tied directly to forecast horizon. Consequently, h1 component diagnostics describe the learned expert, whereas h6 accuracy tests the causal level–anchor construction.

Proxy-to-target migration supports the same allocation during optimization. The first four epochs teach the deep branch an origin-available centred spatial proxy; the normalized sigmoid then transfers value responsibility to the future target while correlation, adjacent-gradient, and pooled-spectrum terms retain spatial organization. Among ten candidates, the 48-channel sigmoid specification attained the minimum validation MAE at 1.169501 °C. This result selected the sigmoid schedule within the prespecified candidate set and supported the full level–shape formulation, fixed horizon fusion, and rolling-origin comparisons. The four independent fold–seed refits, whose MAEs spanned 1.1719–1.1768 °C with a standard deviation of 0.0024 °C, show that this formulation produces a stable optimization outcome.

Together, the component diagnostics, candidate evaluation, seed refits, and 50-configuration sensitivity matrix establish a coherent mechanism: the seasonal background remains explicit, regional displacement is handled causally, spatial shape is learned in a centred representation, and the trusted shape source changes with lead. This mechanism explains why DCSF can improve both absolute field accuracy and transition recovery without giving the deep branch unrestricted control of every horizon.

4.2 Thermal-state changes reveal the strongest practical advantage

Thermal-state changes are the scene in which the level–shape formulation becomes most valuable. Across 25 h1 changing target months, DCSF reduced area-weighted MAE by 0.301 °C (19.79%) relative to TimeFilter, won 17 of 25 matched months, and retained a three-month moving-block interval of 0.073–0.526 °C. It also reduced MAE by 0.218 °C (15.16%) relative to seasonal climatology and by 0.221 °C (15.37%) relative to ModernTCN. These are comparisons within the same fold, target

month, and horizon, so the transition advantage represents better reconstruction of the same future field rather than a change in sample composition.

A state change places simultaneous pressure on both forecast responsibilities. The regional anomaly must move far enough to cross a thermal-state boundary, while the locations above q90 or below q10 must reorganize into the new warm–cold footprint. Seasonal climatology represents the recurring spatial background well, but it carries no origin-specific correction for this displacement. TimeFilter and ModernTCN supply flexible learned forecasts, yet their undifferentiated field representations do not explicitly prevent regional level and local shape from competing. DCSF directly matches the structure of the task: the causal level moves the field, and the centred expert reconstructs its short-lead organization.

Several independent views of the same cases support this interpretation. During the 25 h1 changes, DCSF had lower regional-level absolute error than ModernTCN in 14 months and higher centred-shape correlation in 22 months. It retained the lowest 90th-percentile absolute error in every fold-by-horizon comparison against both ModernTCN and TimeFilter, and the complete ECDFs show that the advantage reaches the upper error tail. The selected maps then locate those gains within the county domain: DCSF more closely retains both regional displacement and the arrangement of positive and negative anomalies. Paired months, error distributions, component diagnostics, and spatial realizations therefore converge on one field-recovery advantage.

The persistent-month comparison sharpens the application meaning. TimeFilter-minus-DCSF gain averaged 0.301 °C across changing h1 months and 0.002 °C across the 11 classifiable non-changing months. All nine prespecified state-threshold definitions preserved a positive all-changing gain, with estimates from 0.235 to 0.316 °C. DCSF's incremental value is thus concentrated at thermal boundaries, where a monthly forecast must revise the established seasonal field rather than simply continue it. For regional environmental monitoring, agricultural scheduling, and energy planning, this is the part of the monthly sequence in which spatially explicit guidance changes most rapidly.

The training-period warming gradient adds a spatial dimension to this result. TimeFilter-relative transition gain increased from 0.270 to 0.303 and 0.329 °C across low-, middle-, and high-warming tertiles defined separately within each training fold. The same positive relation appeared in rolling A ($r=0.378$), rolling B ($r=0.333$), and the pooled within-fold-centred memberships ($r=0.355$); a fixed 2000–2014 definition retained the same low-to-high ordering. These concordant patterns place the largest realized DCSF benefit where the local thermal background changed most rapidly during the information period available to the model. They support spatial targeting of forecast evaluation toward fast-changing parts of the corridor while retaining the complete-domain accuracy result as the primary foundation.

4.3 Forecasting value under an expanding warm footprint

The retrospective analysis shows that the regional thermal field itself is becoming more dominated by broad warm fields. Relative to fixed 2000–2018 cell- and calendar-month thresholds, annual mean warm-area fraction increased by 9.29 percentage points per decade during 2000–2024 and rose from 9.99% in 2000–2009 to 22.46% in 2015–2024. The three-year circular block interval was 2.72–15.52 percentage points per decade, and the Hamed–Rao modified Mann–Kendall p value was 0.00545. Two- and five-year block definitions retained positive lower bounds. Annual maximum warm coverage and the frequencies of regional-or-widespread and widespread warmth increased in the same direction, showing expansion in mean extent, peak reach, and recurrence.

This footprint view extends regional-mean warming into a spatially explicit object. Long-term warming over the Tibetan Plateau and its northern sector is spatially and seasonally heterogeneous [1–3]. Across the Lanzhou–Xining corridor, elevation, valley–hill structure, urban development, and the west–east climatic gradient further organize the thermal field [28,29]. A rise in mean temperature describes domain-wide displacement; the warm-area fraction additionally measures how many locally calibrated cells enter unusual warmth together. Its marked increase demonstrates that the future field problem concerns the spatial reach of anomalous conditions as well as their average intensity.

The six widespread-warm validation target months reveal the short-term geometry of that expansion. Warm coverage averaged 92.08%, and 99.94% of the covered area belonged to the largest eight-neighbour component. These months entered widespread warmth through regional-anomaly jumps averaging 2.818 °C and, for all five months with a following target observation, moved to another state at the next step. Widespread warmth therefore appears as a connected monthly pulse within a rapidly switching sequence, rather than as a collection of isolated warm cells. This geometry explains why continuous-field recovery is scientifically informative: the forecast must locate a near-domain-wide footprint while still resolving the internal anomaly pattern.

The physical-context distributions further characterize these warm fields. Relative to the other 37 validation target months, the 11 regional- or widespread-warm months showed the largest median standardized contrasts in SKT (+1.540), SSR (+1.035), D2M (+0.722), and TP (−1.052), with a smaller positive EVP contrast (+0.250) and a modest wind-speed contrast (−0.159). In ERA5-Land, the warm state is thus embedded in a coherent surface-radiative and near-surface atmospheric setting. These contemporaneous variables describe the environmental state of the target month; DCSF reaches that month from the preceding 12-month history. Their role in the paper is to connect the forecast target to a physically recognizable regional condition.

The level–shape formulation is well matched to this evolving target. Warm-footprint expansion changes the amount of the county domain recruited into unusual warmth, which is strongly influenced by regional displacement, while connected-component geometry and local anomaly arrangement remain shape properties. DCSF's strongest h1 gain during state changes therefore occurs in the same type of monthly dynamics that generates widespread warm pulses. The gain translates into improved reconstruction of a field whose anomalously warm coverage is expanding and whose broadest states arise through abrupt spatial transitions.

The present evidence establishes this capability for monthly ERA5-Land T2M on the 0.1° county grid. The complete retained archives—target and forecast fields, cell thermal backgrounds, transition labels, connected components, uncertainty resamples, and physical-context summaries—also preserve the evidence needed for county-scale thermal-coverage screening, alternative state-threshold analyses, and horizon-specific footprint diagnostics without changing the selected forecasting formulation.

4.4 Relationship to earlier work and transferable implications

This study advances the Lanzhou–Xining evidence base from retrospective attribution to prospective field reconstruction. Earlier regional studies quantified compound dry-heat exposure with XGBoost and SHAP and examined the temporal attribution of wet-bulb-temperature drivers [30,31]. The present target is future monthly T2M itself. Its core evidence is rolling-origin field accuracy, level–shape recovery, transition-conditioned gain, and warm-footprint geometry. The three studies therefore share a regional data foundation while answering complementary questions: how heat-related conditions evolved, which variables covaried with them, and how accurately the future continuous temperature field can be reconstructed from origin-available information.

The method also occupies a distinct position within data-driven weather and time-series forecasting. WeatherBench made common spatial targets and transparent evaluation central to learned weather prediction, while FuXi and GraphCast demonstrated global neural field forecasting at medium range [5,7–9,11]. Regional monthly prediction presents a distinctive field structure: an exceptionally strong annual spatial pattern is accompanied by lead-dependent regional displacement and local reorganization. DCSF translates the full-field principle to this setting by preserving the seasonal field explicitly and learning a centred correction under a rolling-origin protocol.

Relative to Earthformer, DCSF confines the fine–coarse cuboid-attention hierarchy to centred short-lead shape rather than asking it to represent the complete temperature field [13]. Relative to decomposition-oriented models such as TimeMixer, the components have spatial scientific meaning—regional displacement and within-domain organization—and their responsibility is fixed across h1, h3, and h6 [15]. Relative to strong generic sequence alternatives including ModernTCN and TimeFilter, the method adds a forecast-causal seasonal–regional scaffold around the learned representation [16,19]. These comparisons locate the contribution in responsibility design around a compact attention hierarchy.

The same design principle is relevant to gridded environmental variables that combine a stable calendar background, a domain-wide shift, and a reorganizing spatial anomaly. Land-surface temperature, soil moisture, evapotranspiration, snow cover, and vegetation-condition fields can exhibit this structure, although each variable requires its own anchor, centring choice, loss scaling, and rolling-origin evaluation. The transferable object is therefore the decomposition logic: identify an explicit background, assign regional displacement to a causal level, learn a centred shape, and vary shape responsibility with lead according to evidence.

Transfer also benefits from preserving the evaluation discipline used here. Fold-specific normalization and climatology, forecast-origin availability, separately reported absolute accuracy, matched transition comparisons, multi-seed refits, and hierarchy-aware bootstrap intervals make the advantage interpretable. For transfer to another region or variable, selecting horizon weights and anchor sources within the target task preserves this evidence-aligned responsibility design. The present fixed values provide a validated formulation for the Lanzhou–Xining monthly T2M task and a strong initialization for related field-forecasting tasks.

Within this regional forecasting task, the result is both methodologically and environmentally coherent. DCSF achieved a pooled county-domain MAE of 1.131 °C and R^2 of 0.9767, reduced mean county-area-weighted case MAE by 20.35% relative to TimeFilter, and ranked first in the complete ten-method full-support benchmark. Its clearest h1 advantage occurred during thermal-state changes, while the independent retrospective analysis documented a rapidly expanding, highly connected warm footprint. Together, these findings establish level–shape responsibility as an effective strategy for forecasting the monthly spatial temperature dynamics that increasingly characterize the region.

5 Conclusions

RegimeTransfer-DCSF achieved high absolute accuracy for monthly ERA5-Land T2M forecasting over the Lanzhou–Xining county domain, with a pooled MAE of 1.131 °C and an R^2 of 0.9767 across 114 horizon-specific cases and 114,114 grid-cell predictions. Across county-area-weighted cases, its MAE was 20.35% lower than TimeFilter, while the four-refit mean ranked first in the complete ten-method, 3,655-cell rolling-validation benchmark. The clearest short-lead advantage occurred during 25 h1 thermal-state changes, where DCSF reduced MAE by 19.79% relative to TimeFilter. A separate 2000–2024 ERA5-Land analysis showed an autocorrelation-robust expansion of the county-domain

warm footprint by 9.29 percentage points per decade, defining an environmentally consequential regional target for continuous-field prediction. Within the evaluated Lanzhou–Xining rolling-validation task, these results support explicit level–shape responsibility as an effective multihorizon forecasting strategy.

Author contributions

Conceptualization, X.W.; methodology, X.W. and D.Z.; software, A.H.; validation, A.H.; formal analysis, A.H.; investigation, A.H.; data curation, A.H.; visualization, A.H.; writing—original draft preparation, X.W.; writing—review and editing, D.Z.; supervision, X.W.; funding acquisition, X.W. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Key Laboratory of Social Computing and Intelligent Governance in the Western Sichuan Plateau at Sichuan Minzu College (grant no. CXGY202612) and the Project of the Science and Technology Department of Sichuan Province (grant no. 2025JDKP0080).

Data availability statement

ERA5-Land monthly averaged data are publicly available from the Copernicus Climate Data Store (https://doi.org/10.24381/cds.68d2bb30). A versioned reproducibility archive containing the county-domain analysis tables, figure source data, code, configurations, trained parameter states, run inventories, intermediate analyses, and the authoritative county_studyarea.shp boundary layer with provenance hashes is available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare no conflict of interest.

Supplementary material

The Supplementary Material for this article is provided in a separate file.

References

1. Liu, X. & Chen, B. Climatic warming in the Tibetan Plateau during recent decades. *International Journal of Climatology* **20**, 1729–1742 (2000).
2. Duan, A. & Xiao, Z. Does the climate warming hiatus exist over the Tibetan Plateau? *Scientific Reports* **5**, 13711 (2015).
3. Duan, J., Li, L. & Fang, Y. Seasonal spatial heterogeneity of warming rates on the Tibetan Plateau over the past 30 years. *Scientific Reports* **5**, 11725 (2015).
4. Munoz-Sabater, J. *et al.* ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *EARTH SYSTEM SCIENCE DATA* vol. 13 4349–4383 (2021).
5. Rasp, S. *et al.* WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting. *Journal of Advances in Modeling Earth Systems* **12**, e2020MS002203 (2020).
6. Qian, Q. F., Jia, X. J. & Lin, H. Machine Learning Models for the Seasonal Forecast of Winter Surface Air Temperature in North America. *Earth and Space Science* **7**, e2020EA001140 (2020).

7. Chen, L. *et al.* FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science* **6**, 190 (2023).
8. Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).
9. Rasp, S. *et al.* WeatherBench 2: A Benchmark for the Next Generation of Data-Driven Global Weather Models. *Journal of Advances in Modeling Earth Systems* **16**, e2023MS004019 (2024).
10. Weyn, J. A., Durran, D. R. & Caruana, R. Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere. *Journal of Advances in Modeling Earth Systems* **12**, e2020MS002109 (2020).
11. Kochkov, D. *et al.* Neural general circulation models for weather and climate. *Nature* **632**, 1060–1066 (2024).
12. Kurth, T. *et al.* FourCastNet: Accelerating Global High-Resolution Weather Forecasting Using Adaptive Fourier Neural Operators. in *Proceedings of the Platform for Advanced Scientific Computing Conference* 1–11 (ACM, 2023). doi:10.1145/3592979.3593412.
13. Gao, Z. *et al.* Earthformer: Exploring Space-Time Transformers for Earth System Forecasting. in *Advances in Neural Information Processing Systems 35* 25390–25403 (Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2022). doi:10.52202/068431-1841.
14. Liu, Y. *et al.* iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. in *International Conference on Learning Representations* vol. 2024 11116–11140 (2024).
15. Wang, S. *et al.* TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting. in *International Conference on Learning Representations* vol. 2024 38626–38652 (2024).
16. Luo, D. & Wang, X. ModernTCN: A Modern Pure Convolution Structure for General Time Series Analysis. in *International Conference on Learning Representations* vol. 2024 31728–31770 (2024).
17. Wang, Y. *et al.* TimeXer: Empowering Transformers for Time Series Forecasting with Exogenous Variables. in *Advances in Neural Information Processing Systems 37* vol. 37 469–498 (2024).
18. Deng, J., Jiang, R., Zhang, J. & Song, X. Multi-Modality Spatio-Temporal Forecasting via Self-Supervised Learning. in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24* 2018–2026 (International Joint Conferences on Artificial Intelligence Organization, 2024).
19. Hu, Y. *et al.* TimeFilter: Patch-Specific Spatial-Temporal Graph Filtration for Time Series Forecasting. in *Proceedings of the 42nd International Conference on Machine Learning* vol. 267 24893–24911 (2025).
20. Liu, Y., Qin, G., Huang, X., Wang, J. & Long, M. Timer-XL: Long-Context Transformers for Unified Time Series Forecasting. in *International Conference on Learning Representations* 83982–84006 (2025).
21. Nie, Y., Nguyen, N., Sinthong, P. & Kalagnanam, J. A Time Series Is Worth 64 Words: Long-term Forecasting with Transformers. in *The Eleventh International Conference on Learning Representations* (Kigali, Rwanda, 2023).
22. Wu, H. *et al.* TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. in *The Eleventh International Conference on Learning Representations* (Kigali, Rwanda, 2023).
23. Zhou, H. *et al.* Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* **35**, 11106–11115 (2021).
24. Shi, X. *et al.* Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. in *Advances in Neural Information Processing Systems* vol. 28 802–810 (2015).

25. Wu, H., Xu, J., Wang, J. & Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. in *Advances in Neural Information Processing Systems* vol. 34 22419–22430 (2021).
26. Zhou, T. *et al.* FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. in *Proceedings of the 39th International Conference on Machine Learning* vol. 162 27268–27286 (2022).
27. Zeng, A., Chen, M., Zhang, L. & Xu, Q. Are Transformers Effective for Time Series Forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence* **37**, 11121–11128 (2023).
28. Pan, J., Dong, L., Wang, N. & Yang, Z. A multiscale study of thermal environment pattern in Lanzhou-Xining agglomeration. *Remote Sensing for Land & Resources* **30**, 138–146 (2018).
29. Tong, H., Shi, P., Luo, J. & Liu, X. The Structure and Pattern of Urban Network in the Lanzhou-Xining Urban Agglomeration. *CHINESE GEOGRAPHICAL SCIENCE* vol. 30 59–74 (2020).
30. Wang, X. *et al.* Tensor based spatiotemporal attribution of compound dry heat exposure in the Lanzhou-Xining urban agglomeration from 2000 to 2025. *FRONTIERS IN ENVIRONMENTAL SCIENCE* **13**, 1684244 (2025).
31. Wang, X., He, A. & Zhang, D. Decoding urban humid heat: temporal attribution of wet-bulb temperature drivers in the Lanzhou–Xining corridor. *Frontiers in Earth Science* **14**, 1860008 (2026).
32. Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049 (2020).
33. Copernicus Climate Change Service (C3S). ERA5-Land monthly averaged data from 1950 to present. Copernicus Climate Data Store https://doi.org/10.24381/cds.68d2bb30 (2019).
34. Tashman, L. J. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* **16**, 437–450 (2000).
35. Roberts, D. R. *et al.* Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017).
36. Bergmeir, C., Hyndman, R. J. & Koo, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis* **120**, 70–83 (2018).
37. Vaswani, A. *et al.* Attention Is All You Need. in *Advances in Neural Information Processing Systems* vol. 30 5998–6008 (2017).
38. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Lecture Notes in Computer Science* 234–241 (Springer International Publishing, 2015). doi:10.1007/978-3-319-24574-4_28.
39. Jiang, L., Dai, B., Wu, W. & Loy, C. C. Focal Frequency Loss for Image Reconstruction and Synthesis. in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* 13899–13909 (IEEE, 2021). doi:10.1109/ICCV48922.2021.01366.
40. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, 2016). doi:10.1145/2939672.2939785.
41. Willmott, C. & Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* **30**, 79–82 (2005).
42. Hyndman, R. J. & Koehler, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting* **22**, 679–688 (2006).
43. Künsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* **17**, 1217–1241 (1989).

44. Politis, D. N. & Romano, J. P. A circular block-resampling procedure for stationary data. in *Exploring the Limits of Bootstrap* 263–270 (John Wiley & Sons, New York, NY, 1992).
45. Sen, P. K. Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association* **63**, 1379–1389 (1968).
46. Mann, H. B. Nonparametric Tests Against Trend. *Econometrica* **13**, 245 (1945).
47. Hamed, K. H. & Ramachandra Rao, A. A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology* **204**, 182–196 (1998).

Supplementary Files

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

- WangXiaoSupplementaryMaterialv1.0.docx