https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

BHRR v1.0: a two-stage Transformer framework for simultaneous spatial restoration and quantile-function bias correction of climate model temperature fields

Young Hoon Song[1], Hyung Ju Kim [2] Eun-Sung Chung[2*]

[1] Department of Applied Artificial Intelligence, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea
[2] Faculty of Civil Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea

[*] Correspondence to: Eun-Sung Chung eschung@seoultech.ac.kr

Abstract

Bias correction of climate model temperature fields in the image domain is difficult because general circulation model (GCM) outputs and observation-based references occupy different statistical distributions at each grid cell, so pixel-wise regression can recover spatial structure while leaving distributional biases intact. This study presents a two-stage Transformer framework, bias-corrected high-resolution restoration (BHRR), that addresses this problem by decoupling spatial restoration from distribution-aware bias correction. The framework is evaluated on daily near-surface temperature fields over a fixed 200× 280 grid-point (latitude × longitude at 0.25° resolution) Oceania domain by sequentially coupling spatial restoration and distribution-aware bias correction. In the first stage, a Restormer model restores high-resolution spatial structure from linearly interpolated model fields. In the second stage, a Vision Transformer predicts a reference-based quantile map that is used as an explicit transfer function for equidistant cumulative distribution function (CDF) matching in future projections. Across daily minimum, mean, and maximum near-surface air temperature, the restoration stage improves spatial fidelity, increasing median structural similarity to 0.876-0.908 and median peak signal-to-noise ratio to 26.6-28.1 dB. The bias-correction stage further reduces systematic error, yielding near-zero median percent bias (< 0.1%) and lowering median root mean square

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

error (mean temperature by approximately 0.5 K and maximum temperature from 4.4 K to 3.7 K). To verify that the framework preserves climate-change signals rather than collapsing future projections toward historical climatology, future projections under SSP2-4.5 and SSP5-8.5 are examined using ETCCDI extreme indices and Sen's slope. The results confirm scenario-dependent differences in extreme-temperature diagnostics, and spatial-variability analysis shows patterns consistent with a standard downscaled benchmark, supporting the use of the BHRR v1.0 framework as a technical post-processing tool for distribution-aware bias correction of gridded climate fields.

Keywords: image-based bias correction, quantile-function learning, Transformer, high-resolution restoration, climate model post-processing

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

1. Introduction

Reliable high-resolution, bias-corrected daily temperature fields are increasingly required for a wide range of downstream engineering and environmental applications. For example, urban heat-stress assessment relies on spatially resolved daily maximum temperature to identify heat-island hot spots and to inform adaptation measures, particularly as regional heatwaves have shown significant increases in frequency and duration worldwide (Perkins-Kirkpatrick and Lewis, 2020). Building energy-demand models require location-specific heating and cooling degree-day series derived from daily mean temperature, and uncertainties in climate projections propagate directly into energy-demand trend estimates (Deroubaix et al., 2021). Hydrologic and wildfire risk analyses also depend on accurate representations of temperature extremes, because warming-driven reductions in fuel moisture substantially increase fire potential in projected future climates (Gergel et al., 2017). Climate-resilient infrastructure planning similarly demands high-resolution temperature projections that faithfully capture both the mean climate signal and its distributional tails, as engineering design standards are increasingly required to account for non-stationary climate conditions (Buhl and Markolf, 2023). In all these applications, the term “bias” refers to the systematic difference between a climate model’s simulated temperature field and an observation-based reference dataset; removing this bias is essential before model outputs can be used as engineering design inputs. Global climate projections produced by the Coupled Model Intercomparison Project (CMIP) general circulation models (GCMs) provide the primary scientific basis for mitigation and adaptation planning and remain a key tool for exploring future climate change (Song et al., 2024; Birkmann et al., 2022). However, practical regional assessments often demand fine-scale climate information, while limitations in spatial resolution and data quality hinder the direct use of raw GCM outputs (Themeßl et al., 2012; Song and Chung, 2025). Consequently, a broad set of statistical and empirical post-processing approaches has been developed to enhance the accuracy and utility of climate information for applications (Cannon et al., 2015; Giorgi et al., 2009; Song et al., 2022; Gudmundsson et al., 2012).

In operational practice, post-processing commonly consists of two components: downscaling to enhance spatial resolution and recover regional patterns, and bias correction to adjust systematic discrepancies relative to an observation-based reference dataset. Despite continuous advances, many workflows still implement downscaling and bias correction as separate stages, which increases repeated calibration across variables, regions, and time periods and raises

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

computational and operational costs for large ensembles and long simulation periods (Kendon et al., 2010; Vogel et al., 2023). This separation becomes a practical bottleneck when the same processing must be applied repeatedly to multiple scenarios and long time horizons.

Downscaling is commonly categorized into dynamic and statistical approaches. Dynamic downscaling drives a regional climate model using large-scale GCM fields as boundary conditions and can capture regional processes more explicitly, but its computational cost often limits applications over long periods and large domains (Liang et al., 2008). Statistical downscaling is computationally efficient, yet it may struggle to fully represent complex nonlinear processes and rare extremes that are critical for risk-oriented engineering analyses (Baño-Medina et al., 2022; Sachindra et al., 2018; Song and Chung, 2025). These limitations have motivated growing interest in deep-learning-based approaches, which can learn nonlinear mappings and enable rapid inference once trained (LeCun et al., 2015; Sun et al., 2024). In particular, Transformer-family architectures are promising for gridded climate applications because self-attention can capture long-range dependencies and multi-scale spatial patterns (Islam et al., 2023; Curran et al., 2024). Restormer has shown strong performance for high-resolution image restoration (Zamir et al., 2021; Chen et al., 2024), and Vision Transformer (ViT) models integrate global context via self-attention and have recently been reported to improve Transformer-based downscaling and bias-correction of weather and climate data (Dosovitskiy et al., 2021; Curran et al., 2024).

Even when spatial resolution is enhanced, residual mismatches with reference datasets often remain, motivating explicit bias correction. Among many approaches, quantile mapping (QM) is one of the most widely used distribution-based techniques because it corrects the full distribution rather than only mean error (Gudmundsson et al., 2012). However, practical integration should go beyond producing corrected high-resolution fields in a single forward pass; it should also preserve distributional behavior that directly affects extremes and climate indices used in impact assessment. A fundamental difficulty arises when bias correction is cast as an image-to-image learning task: because GCM and observation-based reference fields can occupy systematically different statistical distributions at every grid cell, a model that minimizes pixel-wise reconstruction error alone may recover spatial structure while leaving distributional biases largely intact. Deterministic end-to-end value regression therefore tends to regress toward the mean, underestimate extremes, and provide limited transparency in how tail biases are adjusted under climate non-stationarity (Rampal et al., 2025; Wang and Tian,

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

2024). Therefore, an integrated workflow that explicitly represents and transfers distributional information can provide clearer interpretability and stronger control for climate-information production used in engineering decision pipelines.

In addition, uncertainty must be considered because both climate simulations and learning systems embed multiple uncertainty sources. GCM projections reflect uncertainty driven by greenhouse-gas scenarios, model structure and parameterizations, and initial and boundary conditions (Smith and Forster, 2021; Tokarska et al., 2020; Brunner et al., 2020; Ribes et al., 2021). Deep learning models also involve aleatoric uncertainty from data noise and epistemic uncertainty from imperfect model structures and parameters (Kendall and Gal, 2017; Hüllermeier and Waegeman, 2020; Gawlikowski et al., 2022). In CMIP6-based studies, standard deviation (STDEV) is commonly used as an interpretable dispersion metric to characterize climate variability and projection spread (Tebaldi et al., 2021; Merrifield et al., 2023), motivating the use of spatial STDEV as a transparent proxy to summarize dispersion-related aspects of spatial variability across datasets and scenarios.

To address these challenges, this study presents the technical development of a Transformer-based bias-corrected high-resolution restoration framework, BHRR v1.0 (referred to hereafter as BHRR), and evaluates it as a post-processing tool against established benchmarks. BHRR sequentially integrates (i) Restormer-based spatial restoration and (ii) ViT-based pixel-wise bias correction formulated in quantile-function space. The first stage restores high-resolution spatial structures from linearly interpolated GCM fields. The second stage does not directly regress corrected temperatures; instead, the ViT predicts a reference-based high-resolution quantile map that functions as an explicit transfer object for equidistant cumulative distribution function (CDF) matching in future projections. This design separates spatial restoration from distributional correction, enabling modular calibration while preserving distributional control relevant to extremes.

The framework is implemented over a fixed 200 × 280 Oceania domain, which provides strong spatial heterogeneity (including land-ocean contrasts and diverse climate zones) and thus offers a practical testbed for image-to-image restoration and post-processing of gridded climate fields. For a controlled like-for-like evaluation, a single underlying GCM (ACCESS-CM2) is used, and three near-surface temperature variables commonly used in CMIP6 are considered: daily mean near-surface air temperature (Tas), daily maximum near-surface air temperature (Tmax), and daily minimum near-surface air temperature (Tmin). Princeton Global Forcing version 3

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

(PGFv3) is adopted as the observation-based reference dataset for training and evaluation and is also the reference used in the NASA Earth Exchange Global Daily Downscaled Projections for CMIP6 (NEX-GDDP-CMIP6) bias-correction and spatial disaggregation (BCSD) procedure, enabling a consistent benchmark comparison (Thrasher et al., 2022).

To assess whether the proposed workflow preserves scenario-dependent climate-change signals, future projections are examined under two Shared Socioeconomic Pathway (SSP) scenarios from the CMIP6 scenario framework: SSP2-4.5 and SSP5-8.5. SSP2-4.5 represents an intermediate mitigation pathway that reaches a radiative forcing level of approximately 4.5 W/m^2 by 2100, whereas SSP5-8.5 represents a high-emission pathway that reaches approximately 8.5 W/m^2 by 2100 (Tebaldi et al., 2021). Performance is evaluated against PGFv3 using complementary spatial-fidelity and distributional error metrics. In addition, temperature-based extreme indices recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI) and Sen's slope trend analysis are used to examine scenario-dependent changes, and spatial STDEV is employed as an interpretable proxy for spatial variability and dispersion-related uncertainty.

The main contributions of this work can be summarized as follows:

(1) A two-stage Transformer-based post-processing framework (BHRR v1.0) is developed for image-based bias correction of gridded climate fields. The framework decouples spatial restoration (Restormer) from distributional correction (ViT-based quantile-map translation), which addresses the difficulty that pixel-wise image regression alone cannot remove systematic distributional biases.

(2) The ViT does not directly regress corrected temperatures. Instead, it predicts a reference-based quantile map that serves as an explicit transfer function for equidistant CDF matching. This formulation allows the learned correction to be applied to future projections while preserving CDF-based ranks.

(3) The framework is evaluated over Oceania for three temperature variables (Tas, Tmin, and Tmax) and compared against NEX-GDDP-CMIP6 using spatial-fidelity and distributional error metrics. Furthermore, climate-change signal preservation is assessed under two SSP scenarios using ETCCDI extreme indices, Sen's slope, and spatial STDEV diagnostics.

The remainder of this paper is organized as follows. Section 2 describes the datasets, the BHRR framework, and the evaluation protocol; Section 3 presents the results; Section 4 discusses implications and limitations; and Section 5 concludes the paper.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

2. Dataset and methods

2.1 Study area

The study area is the Oceania domain, centered on the Australian continent and including surrounding oceans and selected island regions (Fig. S1). Oceania provides a favorable setting for image-to-image climate-field processing because the continental landmass is relatively compact, allowing the computational domain to remain tractable at a fixed spatial resolution. In this study, the domain is represented on a regular 0.25°-resolution grid comprising 200× 280 grid points (i.e., 200 latitude rows and 280 longitude columns), which facilitates patch-based training and tiled inference while controlling GPU memory and runtime, yet still preserves key spatial structures such as coastlines, arid interiors, and tropical rainfall belts. Climatically, Oceania spans a pronounced latitudinal gradient from tropical conditions in the north, through subtropical and arid regimes across the interior, to temperate climates in the south yielding strong spatial heterogeneity in both temperature and seasonality (Hollins et al., 2018; Peel et al., 2007; Sharmila and Hendon, 2020). The region is also strongly influenced by large-scale modes of variability, including the El Niño Southern Oscillation, Indian Ocean Dipole, the Australian monsoon, and the Southern Annular Mode, which induce pronounced interannual variability and spatially heterogeneous responses (Risbey et al., 2009). These characteristics allow diverse climate regimes to be evaluated within a single consistent domain, making Oceania an appropriate testbed for assessing the robustness of deep-learning-based workflows for high-resolution restoration and distribution-aware bias correction under historical and scenario conditions.

2.2 Datasets

Daily near-surface temperature variables (Tas, Tmax, and Tmin) were obtained from the ACCESS-CM2 CMIP6 simulations (Mackallah et al., 2022). The historical period (1980-2014) was used to train and evaluate the proposed framework, and future projections were analyzed over 2015–2100 under two SSPs (SSP2-4.5 and SSP5-8.5). To provide consistent spatial geometry across all datasets used in this study, raw ACCESS-CM2 fields were regridded to 0.25° using linear interpolation and represented on the fixed 200 × 280 Oceania domain employed throughout the paper. To benchmark BHRR against a widely used statistically downscaled product, the corresponding ACCESS-CM2 outputs from the NASA NEX-GDDP-CMIP6 dataset were also used (Thrasher et al., 2022). NEX-GDDP-CMIP6 was produced using

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

the bias correction and spatial disaggregation (BCSD) procedure and is widely adopted in climate-impact and engineering-oriented applications. Using the same underlying GCM (ACCESS-CM2) for both the raw CMIP6 simulations (processed by BHRR) and the NEX-GDDP-CMIP6 product enables a transparent like-for-like comparison, in which differences can be attributed primarily to the post-processing methodology rather than to inter-model structural variability.

A single GCM was intentionally selected for two reasons. First, it isolates the effect of the proposed Transformer-based restoration and bias correction from confounding inter-model differences. Second, because Tas, Tmin, and Tmax share comparable physical units and broadly similar value ranges across CMIP-class simulations, evaluating multiple temperature variables within one GCM provides a direct test of whether BHRR consistently handles different aspects of the temperature distribution (including means and extremes), while maintaining interpretability in the benchmark comparison (You et al., 2021).

As the reference dataset and training target, PGFv3 at 0.25° resolution was used for Tas, Tmax, and Tmin. PGFv3 is the reference dataset used in the NEX-GDDP-CMIP6 BCSD procedure and was therefore adopted to ensure a consistent baseline climatology for both BHRR and the benchmark product. During training, BHRR learns the mapping from regridded ACCESS-CM2 fields to PGFv3 on the same 0.25° grid. During evaluation, both the BHRR-processed ACCESS-CM2 outputs and the NEX-GDDP-CMIP6 outputs are compared against PGFv3, enabling a clearer assessment of incremental improvements attributable to BHRR rather than to differences in reference climatology or grid definitions.

2.3 Bias corrected high resolution restoration

This study proposes BHRR, a two-stage Transformer workflow that performs resolution enhancement and distribution-aware bias correction within a single operational pipeline (Fig. 1). First, daily GCM temperature fields are resampled to the fixed 200×280 domain using linear interpolation to construct the model inputs. Second, the Restormer module restores high-resolution spatial structure from the upscaled fields and produces daily temperature fields on the same grid as the reference dataset. Third, the ViT module performs bias correction in quantile-function space by translating the Restormer-derived high-resolution quantile map into a reference-based high-resolution quantile map; the predicted quantile map then acts as an explicit transfer function to correct future daily values via equidistant CDF matching (Eqs. 11-

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

12). By separating spatial restoration from distributional alignment while keeping both stages within one workflow, BHRR provides transparent control over distributional adjustments (including tail behavior) without sacrificing practical inference efficiency.

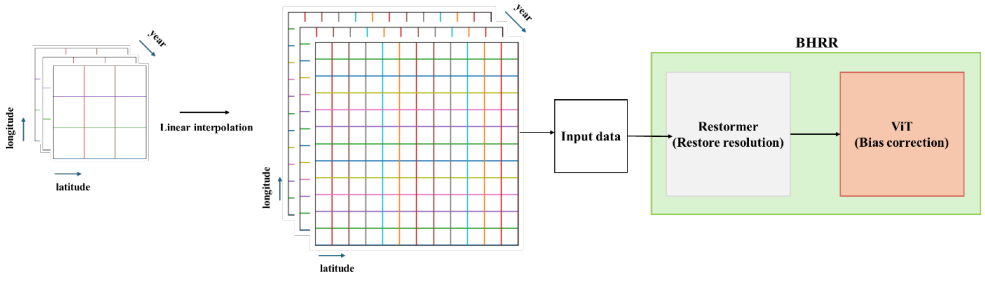

Fig. 1. Overall architecture and processing workflow of the BHRR framework

2.3.1 Restormer-based spatial restoration

The restoration training workflow is summarized in Fig. 2. The restoration module follows the Restormer architecture (Zamir et al., 2021), treating daily gridded temperature fields as single-channel images. For each variable, the input field X is the linearly regridded GCM output on the 200 × 280 domain, and the target field Y is the corresponding reference-based temperature field on the same grid. To ensure stable optimization and consistent scaling across time, both input and target are standardized using the mean (μ) and standard deviation (σ) computed from the training period, as shown in Eq. (1):

$$X^* = \frac{X-\mu}{\sigma}, Y^* = \frac{Y-\mu}{\sigma} \quad (1)$$

Given a normalized input $X^* \in \mathbb{R}^{1\times200\times280}$, Restormer estimates the normalized restored field $\hat{Y}^*$, as shown in Eq. (2).

$$\hat{Y}^* = f_\theta(X^*) \quad (2)$$

where $f_\theta(\cdot)$ denotes the Restormer network with trainable parameters θ. The restored output is then converted back to physical units (Kelvin) using the inverse transformation of Eq. (1), as shown in Eq. (3):

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

$$\hat{Y} = \hat{Y}^{*}\sigma + \mu \quad (3)$$

Training is performed using patch-based sampling to balance data diversity, GPU memory, and runtime. At each training iteration, a patch of size $P \times P$ ($P = 64$) is extracted from the full $H \times W$ domain ($H = 200, W = 280$) by uniformly sampling the top-left corner coordinates (r, c) from the ranges $r \in \{0, 1, \ldots, H - P\}$ and $c \in \{0, 1, \ldots, W - P\}$, yielding the sub-image $X^{*}[r: r + P, c: c + P]$. This uniform random cropping ensures that all spatial locations, including boundary and corner regions, have equal probability of being sampled across training epochs, thereby promoting diverse spatial coverage. Specifically, 64× 64 patches are extracted from the full 200 × 280 domain. During training, random patch cropping is applied, and standard image augmentations (horizontal/vertical flips and transposition) are used to improve generalization. For validation, a centered 64 × 64 patch is used by default. In addition, full-domain inference is executed at every epoch to provide consistent qualitative and quantitative diagnostics on the complete spatial field. The restoration objective minimizes the absolute error in normalized space using an L1 loss, as shown in Eq. (4):

$$L_{rest} = \left\| \hat{Y}^{*} - Y^{*} \right\|_{1} \quad (4)$$

Where, L_{rest} denotes the restoration loss. $\hat{Y}^{*}$ denotes the standardized Restormer prediction. Y^{*} denotes the standardized target field. $\|\cdot\|_{1}$ denotes the sum of absolute differences over the spatial domain.

To generate full-domain predictions consistent with the training patch size, Restormer inference uses sliding-tile prediction with overlap. After partitioning X^{*} into 64 × 64 tiles, an overlap of 16 pixels (stride = 48) is applied. Overlapping predictions are then aggregated by weighted averaging, as shown in Eq. (5):

$$\hat{Y}^{*}(i,j) = \frac{\sum_{k=1}^{K} \hat{Y}_{k}^{*}(i,j)}{\sum_{k=1}^{K} w_{k}(i,j)},\ w_{k}(i,j) \in \{0,1\} \quad (5)$$

Here, $\hat{Y}_{k}^{*}$ denotes the prediction from the k-th tile, and $w_{k}(i,j)$ indicates whether pixel (i,j) is covered by the k-th tile. Model performance is monitored in physical space (Kelvin). PSNR is

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

computed using the target data range as the data range, and SSIM is computed after normalizing both prediction and target to $[0, 1]$ based on the target $[min, max]$ interval. For interpretability during training, baseline PSNR/SSIM values are also computed by directly comparing the upscaled input against the target, enabling epoch-wise tracking of the incremental gain from restoration.

Training is performed for up to 150 epochs using AdamW (learning rate: 1×10^{-4}; weight decay: 0) with gradient clipping ($\|\nabla\| \leq 1.0$). At each epoch, predictions are evaluated on both the validation patch and the tiled full-domain output, and the checkpoint achieving the minimum validation loss is selected for final inference over all time steps. All Restormer experiments are conducted on an NVIDIA A100 GPU; end-to-end training requires approximately 48 h, and inference for a full 200 × 280 field takes approximately 15 s.

2.3.2 ViT-based quantile-map translation and equidistant CDF matching

The bias-correction stage is formulated as a mapping between quantile maps rather than a direct regression of instantaneous corrected temperatures. Critically, the quantile maps used as input to this stage are not internal parameters of the Restormer; instead, they are computed post-hoc from the complete time series of Restormer-restored daily fields produced during Stage 1. Specifically, after the Restormer has been trained and applied to every historical day, the resulting daily restored values at each grid cell are collected over the full historical period and the empirical quantile function is computed at a discrete set of probability levels. This post-hoc construction yields the Restormer-derived quantile map that serves as input to the ViT. LetQ_R denote the high-resolution quantile map computed from the Restormer-restored historical series at each grid cell, and let Q_{ref} denote the corresponding reference-based high-resolution quantile map. Each quantile map represents the pixel-wise quantile function evaluated at a discrete set of probability levels (e.g., $\tau \in \{0.01, \cdots, 0.99\}$). The ViT is trained to transform Q_R into Q_{ref}, and the predicted quantile map $\hat{Q}_{ref}$ is then used as an explicit distributional transfer object for correcting future values while preserving CDF-based ranks. This design improves interpretability because the learned operator is directly tied to distributional differences (including tails) and separates “learning the historical distribution mapping” from “applying the correction to future values.”. The ViT architecture (Fig. 2) consists of patch embedding, Transformer encoder blocks, and an image reconstruction head. The input quantile map is

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

partitioned into N non-overlapping $P \times P$ patches x_p, and positional embeddings E_{pos} are added to form the token sequence, as shown in Eq. (6):

$$z_0 = \left[x_p^1 E; x_p^2 E; \cdots; x_p^N E\right] + E_{pos} \quad (6)$$

Here, E denotes the patch-embedding matrix. The token sequence is propagated through L encoder blocks, each comprising multi-head self-attention (MSA) and a feed-forward multilayer perceptron (MLP). Layer normalization (LN) and residual connections are applied in a pre-norm configuration, as shown in Eqs. (7)-(8):

$$z_l' = MSA\big(LN(z_{l-1})\big) + z_{l-1} \quad (7)$$

$$z_l = MSA\big(LN(z_l')\big) + z_l' \quad (8)$$

After the encoder, the final representation is projected by an MLP head and rearranged back into the two-dimensional grid, as shown in Eqs. (9)-(10):

$$Z_{out} = Linear(Dropout(GELU(Linear(z_l)))) \quad (9)$$

$$\hat{Q}_{ref} = Rearrange(Z_{out}) \quad (10)$$

To mitigate boundary discontinuities introduced by patch-based processing, a convolutional smoothing layer is applied at the output. Quantile maps are normalized to a predefined physical range $[V_{min}, V_{max}]$used consistently during training and inference, and a sigmoid activation is applied to constrain the normalized prediction to (0, 1) before denormalization.

The predicted quantile map is then used to correct future daily values using equidistant CDF matching. Let $X_R^{fut}(t)$ denote the Restormer-restored future value at time t. The correction preserves the CDF-based rank under the historical Restormer distribution $F_{Q_R}(\cdot)$ and retrieves the value at the same rank from the predicted reference distribution $F_{\hat{Q}_{ref}}^{-1}(\cdot)$, as shown in Eq. (11):

$$X_{corr}^{fut}(t) = F_{\hat{Q}_{ref}}^{-1}\left(F_{Q_R}\left(X_R^{fut}(t)\right)\right) \quad (11)$$

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

Because quantiles are available only on discrete probability levels, the transfer is implemented via linear interpolation between adjacent quantiles, as shown in Eq. (12):

$$X_{corr}^{fut}(t) \approx Interp\,(Q_R, \hat{Q}_{ref}, X_R^{fut}(t)) \quad (12)$$

Operationally, the corrected daily field is produced by (i) one ViT forward pass to obtain $\hat{Q}_{ref}$ and (ii) a rank-preserving CDF lookup and interpolation applied pixel-wise for each future time step. Table 2 summarizes the ViT hyperparameters used in this study (input resolution 200 × 280; patch size 8 × 8; embedding dimension 512; encoder depth 12; heads 8; MLP ratio 4.0; dropout 0.1; sigmoid output activation). Under this configuration, the ViT contains approximately 38.48M trainable parameters. All ViT experiments are conducted on an NVIDIA A100 GPU; end-to-end training requires approximately 24 h, and inference for bias correction requires approximately 15 s.

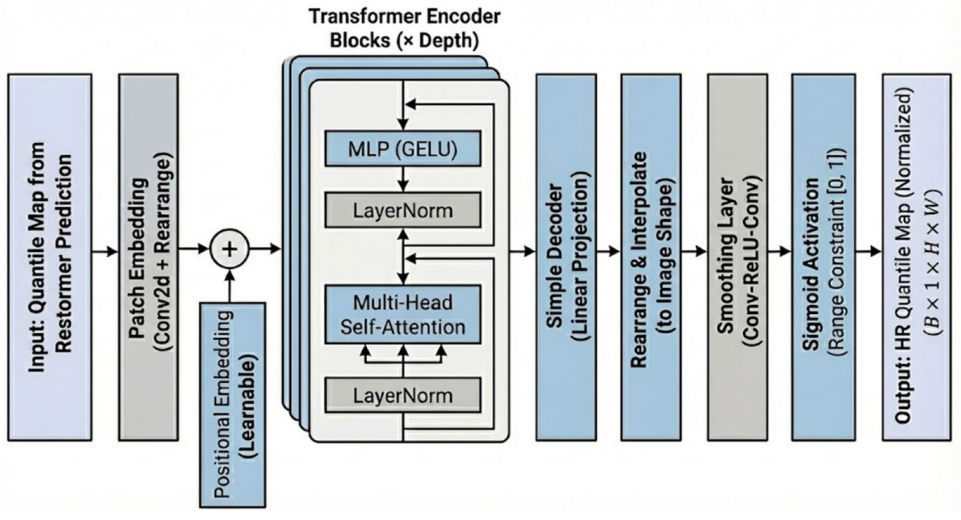

Fig. 2. Overview of ViT network architecture for bias correction

Table 2. Hyperparameter settings of the ViT

Hyperparameter	Value	Description
Input Resolution	200×280	Resolution of quantile maps (Restormer-output HR and reference-based HR)

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

Patch Size	8×8	Size of non-overlapping patches
Embedding Dimension	512	Hidden dimension size of transformer
Encoder Depth	12	Number of transformer encoder blocks
Number of Heads	8	Number of heads in Multi-Head Self-Attention
MLP Ratio	4.0	Ratio of MLP hidden dim to embedding dim
Dropout Rate	0.1	Probability of element zeroing for regularization
Output Activation	Sigmoid	Constraint for normalized range [0, 1]

2.4 Evaluation metrics

This study evaluated high-resolution temperature fields produced by the Transformer-based BHRR model and the NASA NEX-GDDP-CMIP6 against the reference dataset using PSNR, SSIM (Wang et al., 2004), Root Mean Square Error (RMSE), and Percent Bias (PBIAS). Table 1 presents the equations and a brief description of these evaluation metrics. The selected metrics quantify complementary aspects of performance, including overall magnitude error (RMSE), spatial-structural fidelity (PSNR and SSIM), and systematic bias (PBIAS), thereby providing a comprehensive assessment of the reliability of the proposed framework.

Table 1. Evaluation metrics for assessing the BHRR and NEX-GDDP-CMIP6

Metrics	**Equations**	**Factors**
PSNR	$= 10log_{10}(\frac{MAX_X^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i^{sim} - X_i^{ref})^2})$	X_i^{ref}: reference data X_i^{sim}: simulated/downscaled high-resolution temperature n: number of space–time samples
PBIAS	$= \frac{\sum_{i=1}^{n}(X_i^{ref} - X_i^{sim})}{\sum_{i=1}^{n} X_i^{ref}} \times 100$	
RMSE	$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i^{sim} - X_i^{ref})^2}$	

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

SSIM	$= \frac{(2\mu_{re}\mu_{sim} + C_1)(2\sigma_{ref,sim} + C_2)}{(\mu_{ref}^2 + \mu_{sim}^2 + C_1) + (\sigma_{ref}^2 + \sigma_{sim}^2 + C2)}$	μ_{ref}: mean of reference field X^{ref} μ_{sim}: mean of simulated field X^{sim} σ_{ref}^2: variance of X^{ref} σ_{sim}^2: variance of X^{sim} $\sigma_{ref,sim}$: covariance between X^{ref} and X^{sim} C_1, C_2: small positive constants to stabilize the division (typically related to data dynamic range)

382

383 **2.5 Expert team on climate change detection and indices**

384 The changes in the frequency and intensity of temperature extremes are quantified using
385 standardized indices recommended by the ETCCDI. Because ETCCDI indices are derived
386 from daily temperature series using consistent definitions, they provide a transparent basis for
387 like-for-like comparisons between BHRR outputs and the benchmark product.

388 Five temperature-based indices widely used in impact-oriented analyses are considered:
389 Summer Days (SU), Tropical Nights (TR), Frost Days (FD), the annual maximum of daily
390 maximum temperature (TXx), and the annual minimum of daily minimum temperature (TNn).
391 The indices are computed annually at each grid cell from daily temperatures, as shown in Eqs.
392 (13)-(17):

393

394 $SU_y = \sum_d 1(TX_{d,y} > 298.15)$ (13)

395 $TR_y = \sum_d 1(TN_{d,y} > 293.15)$ (14)

396 $FD_y = \sum_d 1(TN_{d,y} < 273.15)$ (15)

397 $TXx_y = \max(TX_{d,y})$ (16)

398 $TNn_y = \min(TN_{d,y})$ (17)

399

400 Here, y denotes the year and d denotes the day within year y. $TX_{d,y}$ and $TN_{d,y}$ represent the
401 daily maximum and daily minimum near-surface air temperature, respectively. SU, TR, and

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

FD are frequency-based indices (counts of days exceeding fixed thresholds), whereas TXx and TNn are intensity-based indices (annual extrema). All indices are calculated from daily fields and reported on an annual basis for each grid cell to support consistent spatial comparisons across datasets and scenarios.

2.6 Sen's slope

Monotonic trends are quantified using Sen's slope, a robust nonparametric estimator defined as the median of pairwise slopes. Sen's slope is computed at each grid cell for BHRR and the benchmark product under SSP2-4.5 and SSP5-8.5, and spatial patterns are compared for the near-future period, NF (2031-2065), and the far-future period, FF (2066-2100). Differences in trend magnitude between products are summarized using an inter-product difference field defined as (BHRR − NEX-GDDP-CMIP6). Sen's slope is estimated as the median of all pairwise slopes between years, as shown in Eq. (18):

$$\beta = median_{i<j}\left(\frac{I(t_j) - I(t_i)}{t_j - t_i}\right) \quad (18)$$

Here, β denotes the Sen's slope at a given grid cell, $I(t_j)$ denotes the annual value of the selected ETCCDI index (or annual temperature metric) in year t, and t_i and t_j denote two distinct years within the analysis period ($(i < j)$. Slopes are reported on a per-decade basis to support consistent interpretation across NF and FF summaries.

2.7 Uncertainty and spatial variability

Because BHRR is derived from a single-GCM workflow, conventional multi-model ensemble spread is not directly available. Therefore, variability-based diagnostics are used to compare uncertainty-related characteristics between BHRR and the benchmark product. Spatial variability is quantified using the STDEV computed at each grid cell over the analysis period, and STDEV patterns are compared across products, SSPs, and future periods. STDEV is computed as shown in Eq. (19):

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2} \quad (19)$$

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

where σ denotes the standard deviation at a given grid cell over the analysis period, X_k denotes the annual value of the temperature variable (Tmin, Tmax, or Tas) in year k, $\bar{X}$ denotes the period mean of X_k and n denotes the number of years in the period. STDEV was computed for Tmin, Tmax, and Tas for both BHRR and NEX-GDDP-CMIP6 during the NF and FF periods under SSP2-4.5 and SSP5-8.5, and spatial distributions were analyzed. Furthermore, this study applied mean $\pm 1\sigma$ envelopes to regional time-series summaries to visualize variability around the mean and support interpretation of scenario-dependent differences.

3. Result

3.1 Performance of BHRR for historical period downscaling and bias correction

3.1.1 Restormer-based resolution restoration

This study compared the ACCESS-CM2 outputs obtained using linear interpolation and Restormer in terms of SSIM and PSNR over Oceania, as shown in Fig. 3. Overall, Restormer increased the median SSIM for Tmin, Tmax, and Tas to 0.876, 0.908, and 0.886, which are higher than those of the linearly interpolated GCM by 0.028, 0.022, and 0.031, respectively. The interquartile ranges were comparable to or narrower than those of the linearly interpolated GCM, indicating more consistent spatial performance. Restormer also improved median PSNR to 26.6, 28.1, and 27.3 dB for Tmin, Tmax, and Tas, which is approximately 1-3 dB higher than the linearly interpolated GCM, with the largest gains for Tmax. Although a few low outliers are present, the overall PSNR distribution of Restormer remains higher than that of linear interpolation.

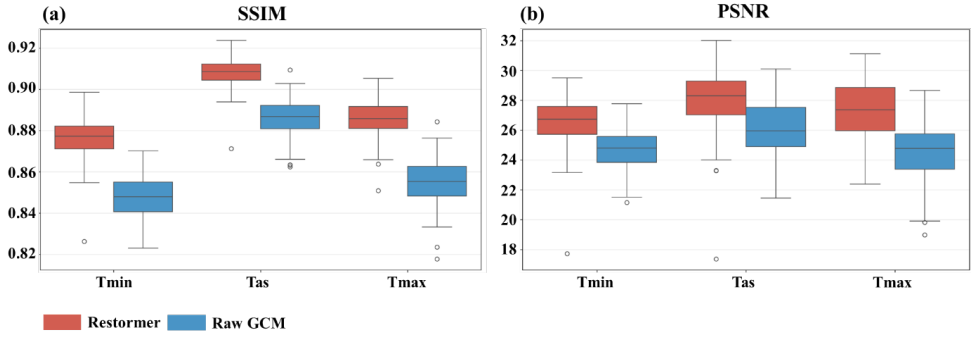

Fig. 3. Boxplot comparison of (a) SSIM and (b) PSNR between Restormer and raw GCM outputs for three temperature variables over Oceania during 1980-2014

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

Fig. S3 presents representative examples that illustrate the qualitative impact of Restormer on spatial detail. Across all three variables, Restormer better resolves small island features and sharpens land-sea contrasts relative to linear interpolation. Improvements are particularly evident over island regions such as New Zealand and Papua New Guinea, and over the Tasman Sea where linearly interpolated fields deviate from the reference while Restormer produces patterns that are more consistent. Similar enhancements are also observed for coastal structures and regional gradients, including along the southeastern coast of Australia for Tmin and in northern island regions for Tmax.

3.1.2 ViT-based bias correction

In this study, bias correction was performed by applying a ViT to GCM outputs whose spatial resolution had been enhanced using Restormer. Fig. 4 compares the BHRR-based GCM and the ACCESS-CM2 outputs from NEX-GDDP-CMIP6 using boxplots of RMSE and PBIAS.

For RMSE, the median Tmin errors were similar for the ViT and NEX-GDDP-CMIP6 at approximately 3.5-4.0 K, but the ViT showed a lower and narrower interquartile range, indicating more stable performance across Australia. For Tas, the ViT median RMSE was about 0.5 K lower than that of NEX-GDDP-CMIP6, with a narrower interquartile range. For Tmax, the ViT also reduced the median RMSE to 3.7 K compared with 4.4 K for NEX-GDDP-CMIP6, again with reduced spread.

For PBIAS, the ViT medians were consistently closer to zero than those of NEX-GDDP-CMIP6, indicating reduced systematic bias. The ViT showed near-neutral biases for all variables, including Tmin 0.08 $\pm$ 0.45%, Tas -0.02 $\pm$ 0.59%, and Tmax 0.07 $\pm$ 0.36%, while NEX-GDDP-CMIP6 exhibited positive biases for Tmin 0.23 $\pm$ 0.50%, Tas 0.38 $\pm$ 0.54%, and Tmax 0.17 $\pm$ 0.69%. The PBIAS difference was most pronounced for Tas at approximately 0.40%, and the ViT generally showed narrower interquartile ranges and comparable or smaller standard deviations, indicating less dispersed bias.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

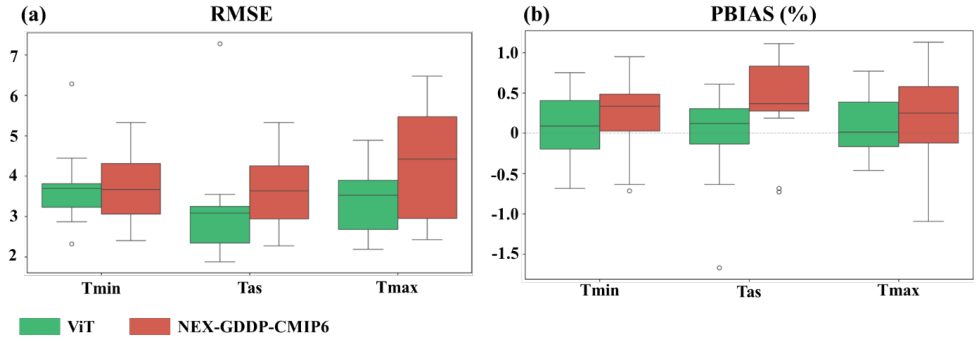

Fig. 4. Comparison of RMSE (a) and PBIAS (b) for three temperature variables using boxplots over Oceania during 1980-2014

Fig. S4 presents spatial RMSE patterns for Tmin, Tas, and Tmax over Oceania during 1980-2014 and the corresponding difference field (ViT minus NEX-GDDP-CMIP6). For Tmin, both models showed larger errors over the central and south-central interior of Australia, while the ViT exhibits slightly reduced magnitudes, including reductions of about 1-2 K over the south-central inland and along the eastern coastal regions. For Tmax, the ViT RMSE is generally 6-7 K over central and southern inland areas, whereas NEX-GDDP-CMIP6 exceeds 7 K across much of the same region, with the largest ViT improvements over north-central Australia. The RMSE difference is predominantly negative over most inland areas, indicating overall higher fidelity for the ViT, while near land–sea boundaries the differences are smaller and more mixed. In some island regions such as Papua New Guinea and parts of northern Indonesia, NEX-GDDP-CMIP6 performs slightly better, typically by less than 1-2 K.

Fig. 5 presents Q-Q plots comparing the NEX-GDDP-CMIP6 and ViT-based bias-corrected outputs with the reference data for Tmin, Tas, and Tmax over Oceania during 1980-2014, based on approximately 4.6×10^8 scalar data points (12768×200×180) that were subsampled in batches of 1,000 for the computation of R^2.

For Tmin, the R^2 of NEX-GDDP-CMIP6 was 0.714, which was lower than that of the ViT. In both models, the data were most densely concentrated in the 290-300K range, indicating relatively high predictive performance in this interval. The difference in variance between the two models was small. However, the ViT estimates were more tightly aligned with the reference data than those of NEX-GDDP-CMIP6.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

For Tas, the R^2 of the ViT was 0.788, which is 0.051 higher than the value obtained for NEX-GDDP-CMIP6, and the ViT remained closer to the reference data across the entire Tas temperature range. Moreover, for points outside the densely populated 290-300K interval, the spread of the ViT estimates was narrower than that of NEX-GDDP-CMIP6, indicating that the ViT explains the variability in Tas more effectively.

For Tmax, the performance gap between the two models was the largest among the three variables. NEX-GDDP-CMIP6 yielded an R^2 of 0.650, the lowest of the three, with a broad spread above 290K that resulted in substantial variance. By contrast, the ViT achieved an R^2 of 0.726, which is 0.076 higher than that of NEX-GDDP-CMIP6. In particular, the overestimation and underestimation evident in NEX-GDDP-CMIP6 were substantially reduced in the ViT results, which lay closer to the reference data, demonstrating a clear advantage of the ViT for predicting extreme temperatures.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

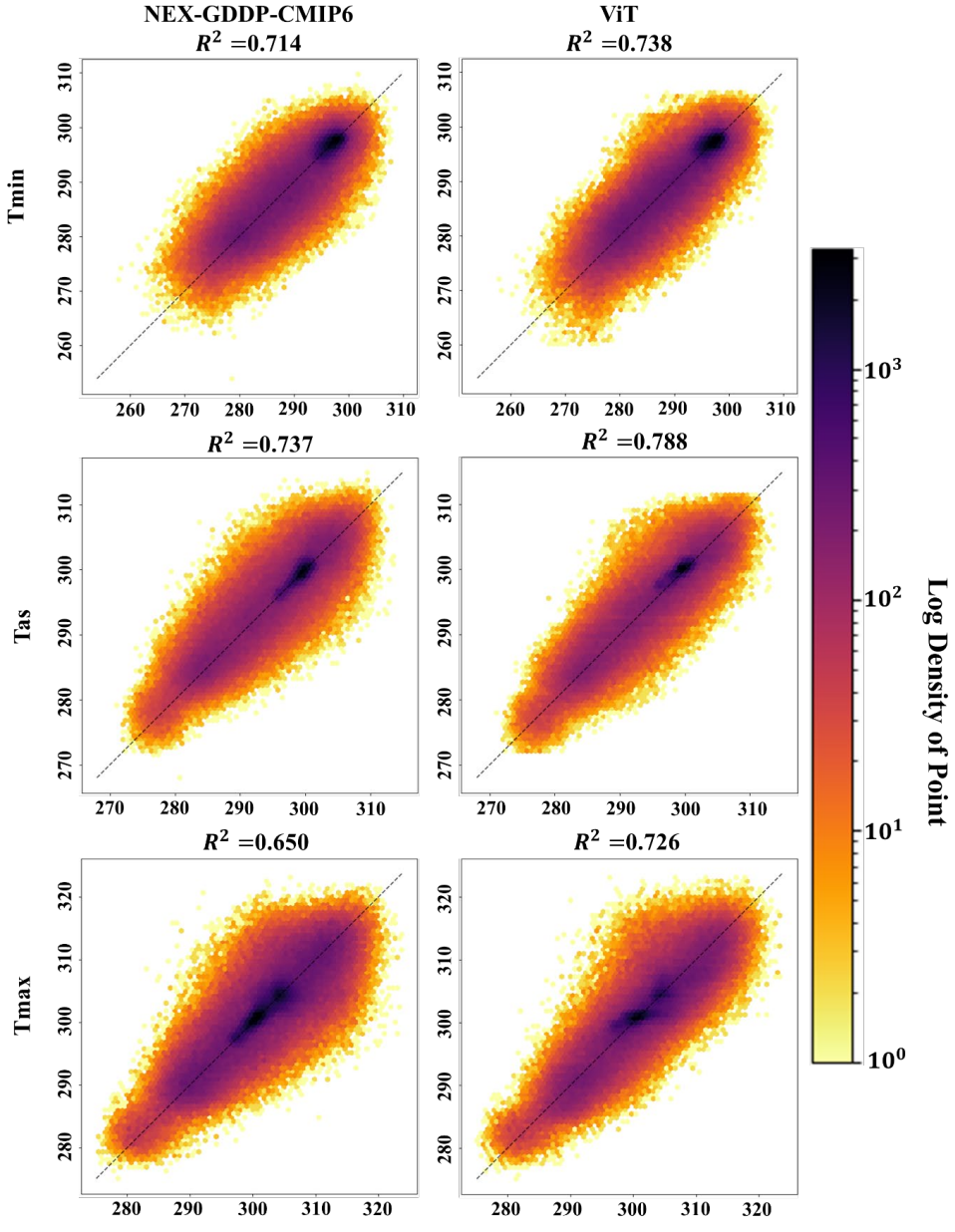

Fig. 5. Point-density scatter (R^2) of BHRR and NEX-GDDP-CMIP6 for three climate variables against the reference dataset over Oceania during 1980-2014 (x-axis: reference data, y-axis: model)

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

3.2 Historical evaluation of temperature extreme indices

This study analyzed the spatial distribution of differences between NEX-GDDP-CMIP6 and BHRR relative to the reference data to evaluate how well each model reproduces temperature extreme indices during the historical period. Fig. 6 compares NEX-GDDP-CMIP6 and the BHRR-based GCM using frequency-based ETCCDI indices SU, TR, and FD and intensity-based indices TXx and TNn.

For SU, both models showed relatively small differences from the reference data, although NEX-GDDP-CMIP6 underestimated SU by about 5-10 days over eastern Australia. By contrast, BHRR remained within approximately ± 5 days over most regions, indicating more stable performance. For TR, the discrepancies were larger. NEX-GDDP-CMIP6 substantially underestimated TR by about 20-40 days over tropical northern Australia, and the difference increased to as much as 60 days in the hottest years, indicating an underestimation of warm nights. BHRR exhibited smaller departures from the reference across the same regions. For FD, both models generally remained within ± 5 days, but NEX-GDDP-CMIP6 underestimated FD by about 5-10 days over mountainous areas of southeastern Australia.

For TXx, NEX-GDDP-CMIP6 underestimated extreme hot temperatures by 2-4 K across most of Australia, with differences reaching 2-4 K along the southern and eastern coasts during the hottest years and showing spatial discontinuities. In contrast, BHRR differed from the reference by approximately ± 1-2 K and produced spatially smoother patterns. For TNn, NEX-GDDP-CMIP6 overestimated the coldest temperatures by 2-4 K over much of inland Australia, whereas BHRR showed smaller deviations, indicating improved representation of cold extremes.

.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

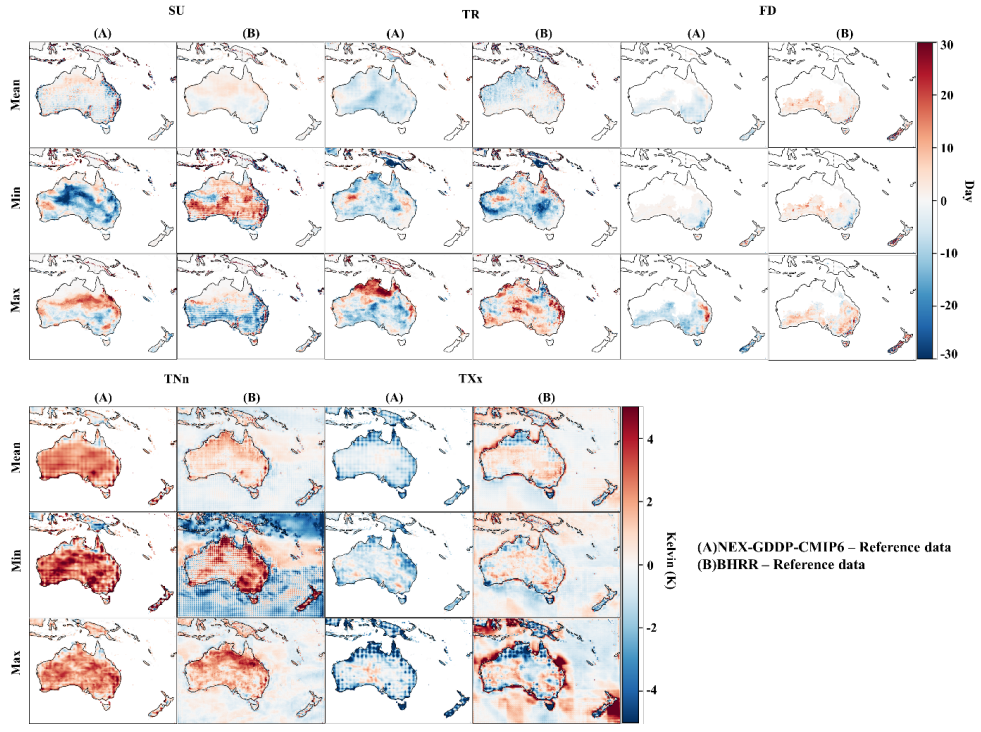

Fig. 6. Biases in ETCCDI indices relative to the reference dataset for BHRR and NEX-GDDP-CMIP6 over Oceania during 1980-2014

3.3 Validation of climate-change signal preservation under SSP scenarios

3.3.1 Scenario-dependent response of temperature variables

A critical requirement for any bias-correction framework is that the correction does not collapse future projections toward historical climatology but instead preserves the climate-change signal from the driving GCM (Cannon et al., 2015; Maurer and Pierce, 2014). To verify this property, BHRR outputs were compared with NEX-GDDP-CMIP6 under two SSP scenarios. This study projected three future temperature variables over Oceania using BHRR and NEX-GDDP-CMIP6 under SSP2-4.5 and SSP5-8.5. Fig.7 presents the temporal evolution of land-only air temperature for Tas, Tmax, and Tmin, comparing BHRR with NEX-GDDP-CMIP6 under the SSP scenarios. During 1980–2014, the reference dataset showed mean values of approximately 296 K (Tas), 302 K (Tmax), and 290 K (Tmin) with low variability, whereas both models projected marked warming under both SSPs. After about 2050, SSP5-8.5 diverged sharply from SSP2-4.5.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

For Tas, SSP2-4.5 remained near 299 K in both datasets, whereas SSP5-8.5 approached 303 K by 2100. The differences between BHRR and NEX-GDDP-CMIP6 were negligible before 2050 but increased thereafter, reaching about 1 K by 2100. For Tmax, SSP5-8.5 increased to about 309 K by 2100 while SSP2-4.5 remained near 305 K, and BHRR became 1-2 K higher than NEX-GDDP-CMIP6 after about 2070. For Tmin, SSP5-8.5 increased to about 297 K by 2100 and showed the largest warming relative to the historical baseline, consistent with amplified nighttime warming. Under SSP2-4.5, Tmin ranged from about 292 K to 294 K, and BHRR exceeded NEX-GDDP-CMIP6 by roughly 1 K after about 2070.

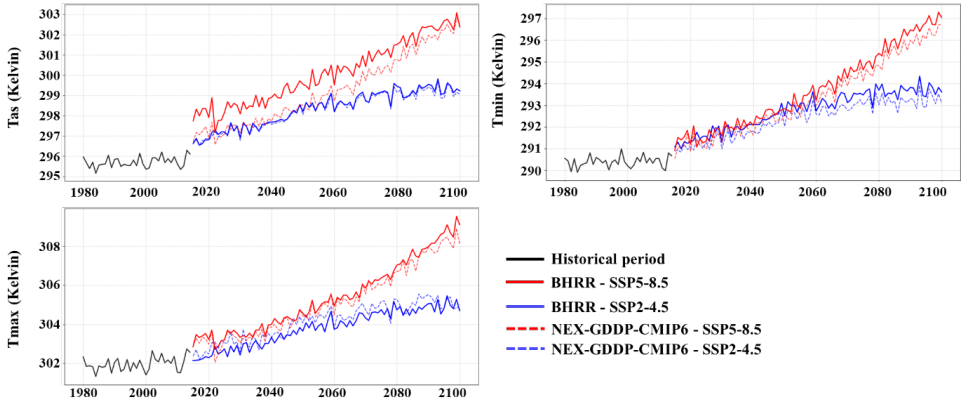

Fig. 7. Projections of three climate variables for BHRR and NEX-GDDP-CMIP6 over Oceania land-mean under SSP2-4.5 and SSP5-8.5 during 1980-2100

Fig. 8 compares spatial anomalies for Tmin, Tas, and Tmax in the NF and FF relative to 1980-2014 under SSP2-4.5 and SSP5-8.5. For Tmin, BHRR exhibited largely uniform warming of about 2-4 K under SSP2-4.5 and 2-5 K under SSP5-8.5 in NF and increased further to 3-5 K and 4-7 K in FF, whereas NEX-GDDP-CMIP6 showed smaller anomalies of about 1-3 K in NF and 2-4 K in FF. For Tas, inter-model differences were largest under SSP5-8.5, with BHRR projecting stronger warming that extended into adjacent marine areas and reached about 5-7 K in NF and 6-8 K in FF, while NEX-GDDP-CMIP6 showed more moderate increases. For Tmax, BHRR produced a more heterogeneous pattern, including enhanced warming along the southeastern coast of about 5-7 K in NF and 7-9 K in FF, whereas NEX-GDDP-CMIP6 exhibited a more spatially uniform pattern of about 2-4 K in NF and 4-6 K in FF.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

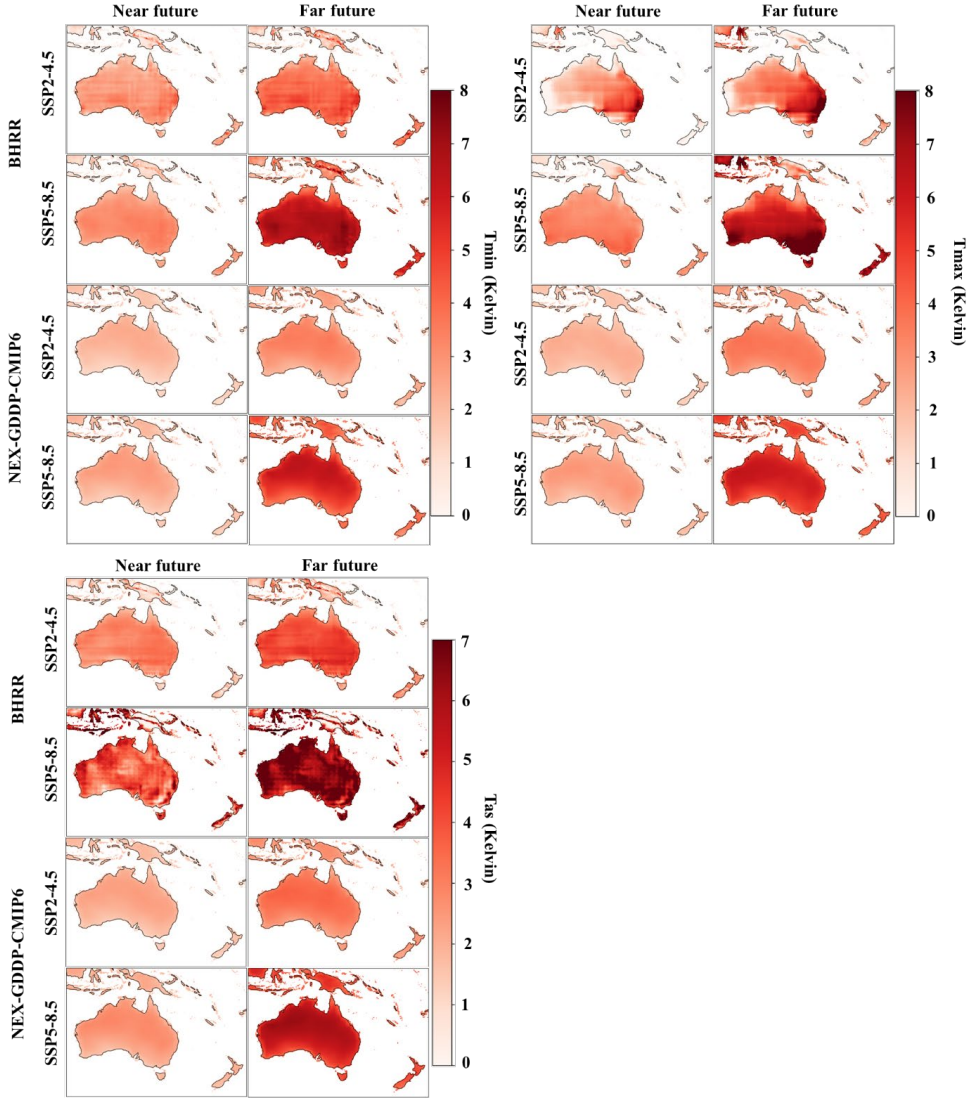

Fig. 8. Spatial anomalies of future three climate variables relative to 1980-2014 for BHRR and NEX-GDDP-CMIP6 under SSP2-4.5 and SSP5-8.5

3.3.2 Extreme-index response under future scenarios

This study calculated ETCCDI temperature extreme indices over Oceania under SSP2-4.5 and SSP5-8.5 using the BHRR projections. Fig. 9 shows projected changes in FD, SU, TR, TNn, and TXx for NF and FF relative to 1980-2014. FD decreased broadly, approaching -80 days

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

over most areas, with smaller reductions in southern Australia and New Zealand. SU increased under both scenarios, exceeding 70 days in FF under SSP2-4.5 and increasing by more than 75 days over most regions in FF under SSP5-8.5. TR also increased strongly, exceeding 70 days in NF under SSP2-4.5 and approaching about 80 days in FF, while SSP5-8.5 yielded an increase of about 80 days in FF, indicating intensified nighttime heat. TNn increased under all scenarios and exceeded 6 K over most areas in FF under SSP5-8.5, consistent with weakened cold extremes. TXx increased in both scenarios, with broader amplification under SSP5-8.5 and values approaching about 6 K across Oceania in FF.

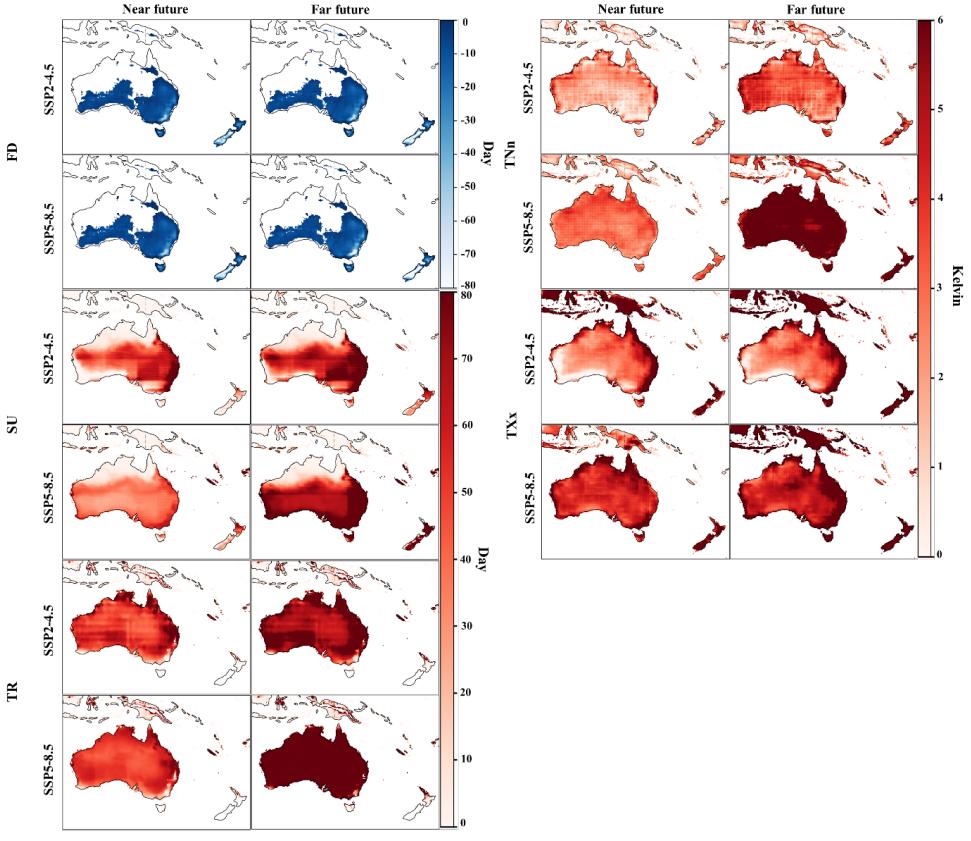

Fig. 9. Projected changes in future ETCCDI indices relative to 1980-2014 from BHRR under SSP2-4.5 and SSP5-8.5

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

Table 3 summarizes regional-average changes in ETCCDI indices for NF and FF relative to the historical baseline and includes inter-model differences. Under SSP2-4.5, SU and TR increased by 4.95 and 38.76 days in NF and by 26.45 and 39.93 days in FF. Under SSP5-8.5, SU and TR increased by 17.50 and 50.82 days in NF and by 49.11 and 79.28 days in FF. BHRR produced larger SU and TR than NEX-GDDP-CMIP6 in both NF and FF, with FF differences of 2.12 days for SU and 15.41 days for TR. FD decreased in all periods, with the largest reduction occurring in FF under SSP5-8.5 at -12.68 days, and NEX-GDDP-CMIP6 produced slightly higher FD than BHRR by 0.68 days in NF and 0.49 days in FF. TXx increased by 3.55 K in NF and 4.68 K in FF under SSP2-4.5 and by 4.44 K in NF and 7.21 K in FF under SSP5-8.5, while TNn increased in all scenarios and reached its maximum change in FF under SSP5-8.5 at 6.19 K. In FF, BHRR exceeded NEX-GDDP-CMIP6 by 3.33 K for TXx and 0.81 K for TNn.

Table 3. Differences between the historical values and the projected changes of ETCCDI indices over Oceania under the SSP2-4.5 and SSP5-8.5 scenarios

Index	Historical period	SSP2-4.5		SSP5-8.5		Difference (BHRR – NEX-GDDP-CMIP6)	
		Near	Far	Near	Far	Near	Far
SU	268.51	4.95	26.45	17.5	49.11	0.1	2.12
TR	163.51	38.76	39.93	50.82	79.28	10.62	15.41
FD	12.86	-11.74	-12.04	-12.11	-12.68	-0.68	-0.49
TXx	313.85	3.55	4.68	4.44	7.21	3.01	3.33
TNn	278.85	2.1	3.13	2.97	6.19	0.04	0.81

3.3.3 Trend analysis using Sen's slope for extreme indices

This study applied Sen's slope to compare long-term trends in ETCCDI indices from BHRR and NEX-GDDP-CMIP6 under SSP2-4.5 and SSP5-8.5. Fig. 10 shows spatial patterns for NF and FF for each model and their difference, and Table 4 summarizes changes in trend magnitude from NF to FF.

For SU under SSP2-4.5, NF slopes were similar at 5.87 and 5.93 days per decade for BHRR and NEX-GDDP-CMIP6, respectively, and decreased in FF to 2.11 and 2.10 days per decade,

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

indicating strong deceleration. Under SSP5-8.5, NF slopes were higher at 6.58 and 6.38 days per decade and increased in FF to 8.28 and 7.87 days per decade, indicating strengthened trends. For TR, positive slopes were observed under both scenarios. Under SSP2-4.5, slopes decreased from 6.80 and 6.43 days per decade in NF to 1.12 and 0.97 days per decade in FF. Under SSP5-8.5, slopes increased from 9.63 and 8.18 days per decade in NF to 13.69 and 12.00 days per decade in FF, with BHRR remaining consistently large.

For FD, slopes were negative under both scenarios. Under SSP2-4.5, NF slopes were -0.25 and -0.12 days per decade and converged in FF to -0.03 days per decade for both models. Under SSP5-8.5, NF slopes were similar at -0.22 and -0.23 days per decade, whereas in FF BHRR remained negative at -0.06 days per decade and NEX-GDDP-CMIP6 became positive at 0.15 days per decade, indicating a sign reversal. For TXx under SSP2-4.5, NF slopes were 0.35 and 0.30 K per decade and decreased in FF to 0.18 and 0.26 K per decade. Under SSP5-8.5, NF slopes were similar at 0.55 and 0.56 K per decade, while FF slopes increased to 0.59 K per decade for BHRR and 0.93 K per decade for NEX-GDDP-CMIP6. For TNn under SSP2-4.5, NF slopes were 0.72 and 0.34 K per decade and decreased in FF to 0.23 and 0.16 K per decade. Under SSP5-8.5, NF slopes were 0.73 and 0.48 K per decade and increased in FF to 1.06 and 0.80 K per decade, indicating stronger weakening of cold extremes under higher forcing. Across periods, TNn trends were generally larger in BHRR than in NEX-GDDP-CMIP6.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

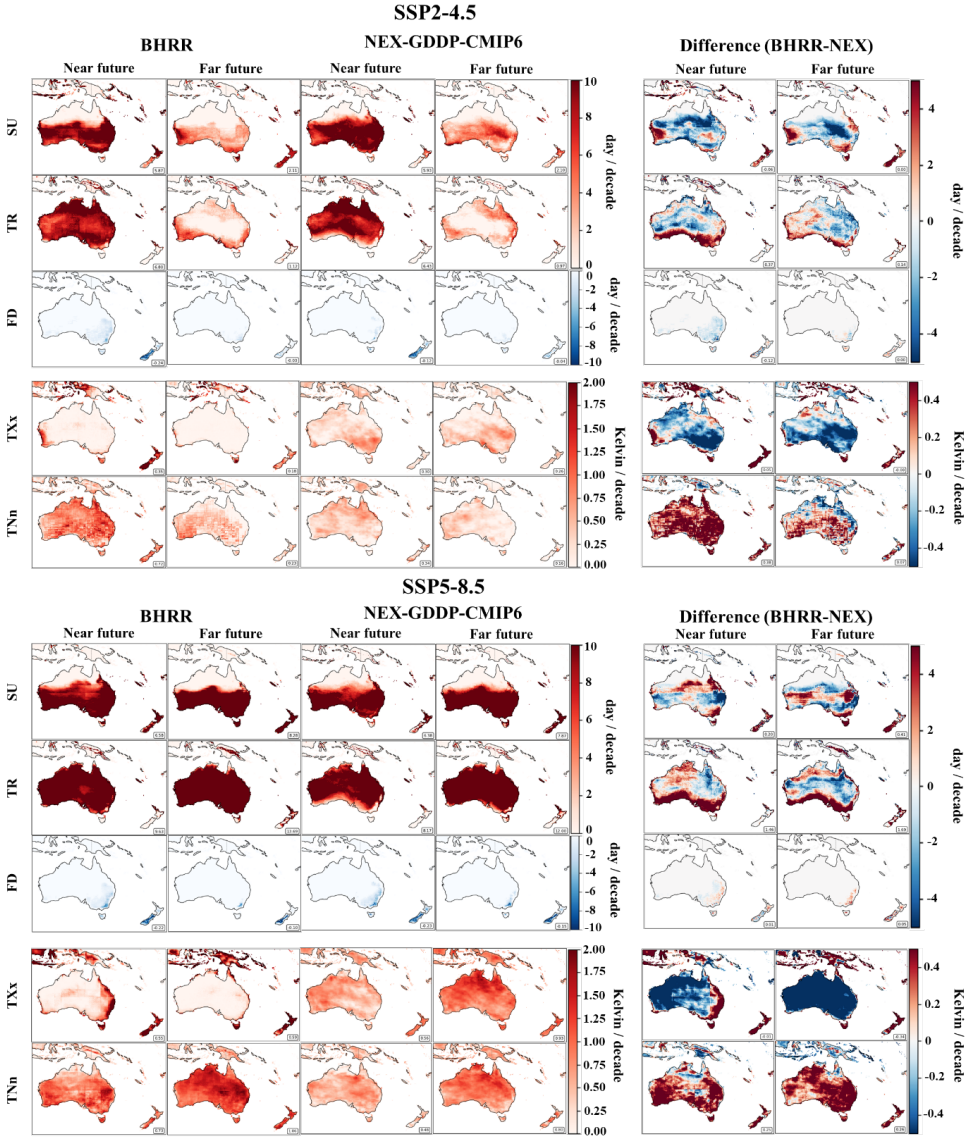

Fig. 10. Sen's slope trends of ETCCDI indices for BHRR and NEX-GDDP-CMIP6 and their differences under SSP2-4.5 and SSP5-8.5 for both futures

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

Table 4 Comparison of the magnitudes of Sen's slope trends of ETCCDI indices for BHRR and NEX-GDDP-CMIP6 under SSP2-4.5 and SSP5-8.5 for the two future periods

Index	Scenario	NF (BHRR / NEX-GDDP)	FF (BHRR / NEX-GDDP)	Change in trend magnitude from NF to FF
SU	SSP2-4.5	5.87 / 5.93	2.11 / 2.10	Deceleration (-64%)
	SSP5-8.5	6.58 / 6.38	8.28 / 7.87	Amplification (+26%)
TR	SSP2-4.5	6.80 / 6.43	1.12 / 0.97	Deceleration (-84%)
	SSP5-8.5	9.63 / 8.18	13.69 / 12.00	Acceleration (+42%)
FD	SSP2-4.5	-0.25 / -0.12	-0.03 / -0.03	Deceleration
	SSP5-8.5	-0.22 / -0.23	-0.06 / 0.15	Deceleration
TXx	SSP2-4.5	0.35 / 0.30	0.18 / 0.26	Deceleration
	SSP5-8.5	0.55 / 0.56	0.59 / 0.93	Maintained / amplification
TNn	SSP2-4.5	0.72 / 0.34	0.23 / 0.16	Deceleration
	SSP5-8.5	0.73 / 0.48	1.06 / 0.80	Acceleration (+45%)

3.4 Uncertainty and spatial variability assessed by spatial STDEV

In this study, spatial STDEV was calculated for three temperature variables (Tmin, Tmax, and Tas) produced by BHRR and NEX-GDDP-CMIP6 during the NF and FF to quantify interannual variability and related uncertainty characteristics under future climate conditions. Fig. 11 presents spatial STDEV patterns for Tmin, Tmax, and Tas in NF and FF under SSP2-4.5 and SSP5-8.5, together with inter-model differences.

For Tmax, BHRR generally showed smaller STDEV than NEX-GDDP-CMIP6 across both NF and FF under both scenarios. Differences were predominantly -1.0 to -0.5 K, indicating that Tmax variability in BHRR was about 0.5-1.0 K lower, particularly over arid inland regions. Positive differences occurred in parts of eastern and southern Oceania, where BHRR showed higher variability by about 0.5-0.8 K, and coastal-adjacent areas tended to exhibit higher STDEV than inland regions. For Tmin, both models showed similar spatial structures, with relatively large STDEV over south-central Australia in both NF and FF. Differences were generally positive at about 0.6-1.0 K, indicating higher Tmin variability in BHRR, while differences were small over the northern tropics. From NF to FF, BHRR showed increased Tmin STDEV under SSP2-4.5, and under SSP5-8.5 the NF–FF increase reached about 1.2 K

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

in some areas, whereas NEX-GDDP-CMIP6 showed more limited NF–FF change. For Tas, NEX-GDDP-CMIP6 showed only small NF–FF differences under SSP2-4.5, whereas BHRR generally produced larger STDEV, with predominantly positive differences of about 0.3–0.6 K over central and southern Australia under both scenarios.

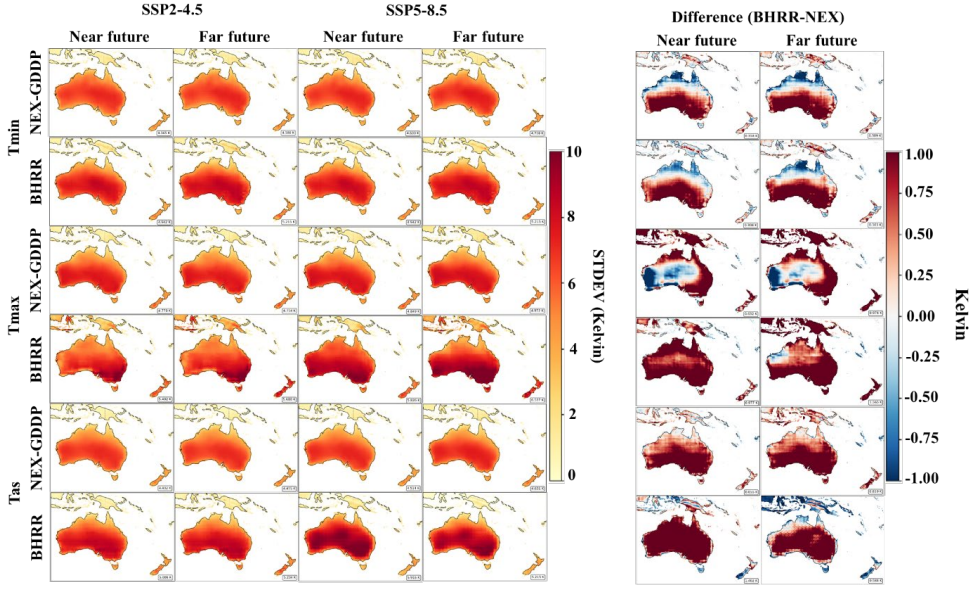

Fig. 11. Spatial variability of three climate variables quantified by spatial STDEV for BHRR and NEX-GDDP-CMIP6 and their differences under SSP2-4.5 and SSP5-8.5

Fig. 12 shows land-only area-mean time series of Tas, Tmax, and Tmin under SSP2-4.5 and SSP5-8.5 for BHRR and NEX-GDDP-CMIP6, with shaded envelopes indicating ±1σ spatial STDEV over the land domain. Under SSP2-4.5, all variables increased gradually through the century, whereas under SSP5-8.5 warming was stronger and the separation from SSP2-4.5 became more pronounced after mid-century. Across variables, the ±1σ envelopes remained substantial, indicating persistent spatial heterogeneity and regional contrasts in warming. BHRR and NEX-GDDP-CMIP6 agreed closely under SSP2-4.5, while under SSP5-8.5 differences became more apparent later in the century, with BHRR often slightly higher for Tas and Tmax and smaller differences for Tmin. Overall, inter-model differences were modest relative to the scenario-driven signal, but divergence under SSP5-8.5 suggested increased model-dependent sensitivity under stronger forcing.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

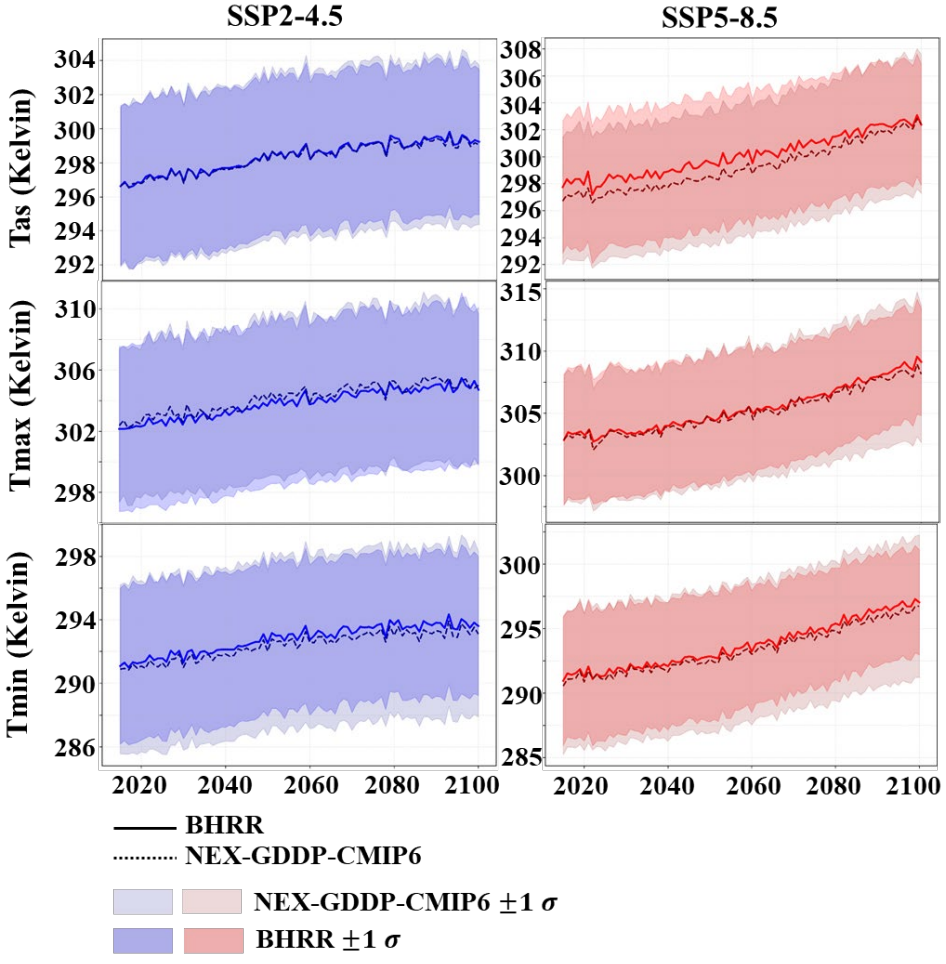

Fig. 12. Oceania land-only mean projections of three climate variables for BHRR and NEX-GDDP-CMIP6 under SSP2-4.5 and SSP5-8.5 with $\pm 1\sigma$ spatial STDEV

Fig. S5 presents boxplots of spatial standard deviation (spatial STDEV, K) for Tmin, Tas, and Tmax over Oceania. Because reference data are available only for the historical period, they are shown for two domains, land-only and land plus ocean, to isolate domain effects on spatial heterogeneity. Model results are shown for NF and FF under SSP2-4.5 and SSP5-8.5. In the historical reference data, Tas and Tmax exhibited larger spatial STDEV over the land-plus-

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

ocean domain than over land-only by about 1.5-2.0 K, reflecting the land-ocean thermal contrast, whereas Tmin showed weaker domain dependence.

Relative to these benchmarks, NEX-GDDP-CMIP6 generally lay within or near the land-only reference range, while BHRR aligned more closely with the higher-variability land-plus-ocean reference, particularly for Tas. For Tas and Tmax, BHRR frequently occupied the upper portion of the historical land-plus-ocean distribution, whereas departures were smaller for Tmin. Future projections showed clear scenario dependence. Under SSP5-8.5, Tmax exhibited the largest increase in spatial STDEV in BHRR from NF to FF and in some cases exceeded the historical land-plus-ocean reference range, consistent with an intensified land–ocean thermal disparity. By contrast, Tmin showed maintained or slightly reduced spatial STDEV in both models, consistent with partial spatial homogenization as colder regions warmed more rapidly

4. Discussion

Transformer-based architectures have recently advanced single-image restoration and super-resolution by capturing long-range dependencies (Liang et al., 2022; Liu et al., 2022). Building on these developments, BHRR introduces a two-stage Transformer workflow that separates spatial restoration and distributional bias correction. Restormer restores high-resolution spatial structures from linearly upscaled GCM fields, while the subsequent ViT does not directly regress corrected temperatures. Instead, it predicts a reference-based high-resolution quantile map from the Restormer-output quantile map, which is then used as an explicit transfer function for equidistant CDF matching in future projections. This design differs from end-to-end value-regression approaches and provides explicit control over distributional adjustments, including tail behavior relevant for extremes.

Recent diffusion-model studies also report strong restoration performance and often adopt Transformers as backbones (Peebles and Xie, 2023; Bao et al., 2023). However, operational climate-information production typically requires repeated processing across large ensembles and long simulation periods, making workflow complexity and repeated calibration practical burdens (Rampal et al., 2024). Integrating high-resolution restoration and bias correction within a single framework can therefore improve operational efficiency. In this study, BHRR required approximately 72 hours for training (Restormer and ViT combined), while inference was completed in under 30 seconds, indicating rapid deployment after one-time training. On this basis, four topics are discussed below.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

The first topic concerns deep learning-based high-resolution reconstruction of GCM outputs. A central challenge is that the spatial structures of low-resolution GCM inputs and high-resolution reference targets are not necessarily identical. Although resolution has improved in CMIP6, GCM fields can still be spatially coarse relative to reference datasets, particularly for topography, coastlines, and small islands (Evin et al., 2024; Nair et al., 2023). Where terrain and land–sea masks change abruptly, structural discrepancies can lead to learning difficulties beyond simple resolution enhancement, reflecting domain-shift effects in supervised mapping (Ben-David et al., 2010). For this reason, this study framed the task as image restoration from a linearly upscaled input rather than pure super-resolution. Consistent with this framing, linearly interpolated fields showed blurred land–ocean boundaries and weakened fine-scale structures, whereas Restormer restored sharper boundaries and improved spatial textures (Fig. 3), with higher SSIM and PSNR than linear interpolation. These results suggest that restoring missing boundary and texture information can improve spatial fidelity, but spatial restoration alone does not remove numerical biases relative to the reference dataset, motivating a subsequent bias-correction step. Conversely, if bias correction were applied directly through quantile-map translation without a preceding spatial-restoration stage, the input fields would retain the blurred spatial structures inherent in linearly interpolated GCM outputs, and the resulting corrected fields would be expected to exhibit degraded spatial-fidelity metrics (SSIM and PSNR) because the quantile-based correction adjusts distributional properties at each grid cell but does not recover missing fine-scale spatial information. This bidirectional limitation motivates the two-stage design of BHRR: the Restormer first restores high-resolution spatial structure to maximize SSIM and PSNR, and the ViT-based quantile-map translation then corrects distributional biases without compromising the spatial detail already recovered.

The second topic concerns the need for quantile learning in deep learning-based bias correction. A central challenge of image-based bias correction is that GCM and observation-based reference fields can occupy systematically different cumulative distribution functions (CDFs) at each grid cell. When an image-to-image network is trained to minimize pixel-wise reconstruction error, it can learn spatial patterns effectively, but systematic distributional offsets—including shifted means and compressed or inflated tails—tend to persist because the loss function does not explicitly penalize CDF misalignment. This makes conventional image-regression approaches insufficient for producing fields that are both spatially faithful and distributionally unbiased. To overcome this limitation, BHRR formulates bias correction as a

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

quantile-function translation task rather than a direct value-regression task. In BHRR, the ViT predicts the full reference-based quantile map Q_{ref} rather than pointwise corrected temperatures, explicitly representing the target distribution at each grid cell and enabling a rank-preserving CDF-based correction. This design supports evaluation not only of mean errors but also of distribution alignment, including tails that drive extreme indices. Distributional discrepancies can be especially challenging when the target variable has strong non-Gaussian behavior. Although not tested here, precipitation is a common example because of frequent zeros, heavy tails, and spatial intermittency (Lee and Haran, 2024), where reducing mean error alone may not ensure realistic extremes. For temperature, the ViT-based correction improved RMSE and PBIAS across Tmin, Tas, and Tmax, with PBIAS close to zero, and reduced regional errors in inland and south-central areas, indicating spatially varying corrections rather than a uniform adjustment. Q-Q diagnostics also showed closer distributional agreement with the reference than NEX-GDDP-CMIP6, supporting the value of quantile-based learning for distributional alignment. A remaining question is whether historical improvements are carried forward consistently into future projections while preserving scenario-dependent change signals.

The third topic concerns whether BHRR preserves the climate-change signal embedded in the driving GCM. This is a central requirement for any bias-correction framework, because traditional quantile mapping can distort the GCM signal during distribution alignment (Maurer and Pierce, 2014; Themeßl et al., 2012). Signal-preserving methods such as Quantile Delta Mapping were developed to address this problem by retaining quantile-specific changes (Cannon et al., 2015; Cannon, 2018). In BHRR, the ViT learns a historical quantile-function mapping that is then applied to future daily values through equidistant CDF matching. Therefore, the correction is derived from the historical distributional difference and applied to future values without overwriting the scenario-dependent shift produced by the GCM. To verify this property, this study assessed whether differences between SSP2-4.5 and SSP5-8.5 were preserved in BHRR outputs using ETCCDI indices and Sen's slope (Kim et al., 2020; Zhao et al., 2023). The results showed larger increases in SU and TR, stronger decreases in FD, and greater amplification toward FF under SSP5-8.5, while Sen's slope indicated strengthened trends for indices such as TR. These patterns are consistent with the expected radiative-forcing difference between the two scenarios and with prior evidence that extreme temperature changes over Australasia intensify with warming (IPCC, 2022; Song et al., 2024; Deng et al., 2022). Based on these results, the equidistant CDF matching formulation in BHRR appears to preserve

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

the scenario-dependent signal in a manner comparable to established signal-preserving approaches.

The final topic discusses variability in future projections from BHRR and NEX-GDDP-CMIP6. This study used spatial STDEV as a proxy for spatial variability and uncertainty characteristics. For land-only mean time series, BHRR and NEX-GDDP-CMIP6 remained similar under SSP2-4.5, whereas differences widened over time under SSP5-8.5. The ±1σ envelopes indicated sustained spatial heterogeneity and suggested larger variability under stronger forcing. Because land–ocean thermal contrast can further amplify spatial variability, interpretation benefits from comparison with previous studies of regional variability and extremes. Prior work has reported increasing variability across Oceania toward later periods and under higher emissions (Molina et al., 2025; Deng et al., 2022; Song et al., 2024). Consistent with these findings, this study indicated amplified spatial variability in BHRR projections, particularly under SSP5-8.5 and in FF, with scenario dependence broadly comparable to NEX-GDDP-CMIP6. The framework provides a transferable two-stage architecture whose components can be retrained independently for other GCMs, variables, and domains.

5. Conclusion

This study presented the development and evaluation of BHRR v1.0, a two-stage Transformer framework for simultaneous spatial restoration and quantile-function bias correction of gridded climate model temperature fields. The framework addresses the difficulty that conventional image-to-image regression cannot simultaneously recover spatial structure and correct distributional biases, by decoupling Restormer-based spatial restoration from ViT-based quantile-function translation. BHRR was evaluated against the reference dataset and benchmarked against NEX-GDDP-CMIP6 using PSNR, SSIM, RMSE, and PBIAS. Climate-change signal preservation was assessed under SSP2-4.5 and SSP5-8.5 were examined using ETCCDI temperature extreme indices and Sen's slope, and spatial STDEV was used as a proxy for spatial variability and uncertainty. The main conclusions were as follows:

1. Restormer improved the spatial fidelity of GCM temperature fields relative to linear interpolation, yielding higher median SSIM and PSNR for Tmin, Tas, and Tmax and indicating more stable spatial restoration.
2. The ViT-based bias-correction stage reduced overall errors and systematic bias, with lower RMSE and PBIAS and closer distributional alignment to the reference dataset

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

than NEX-GDDP-CMIP6, supporting the effectiveness of the pixel-wise quantile-mapping formulation.

3. ETCCDI indices and Sen's slope confirmed that BHRR preserved the climate-change signal from the driving GCM. Radiative-forcing differences between SSP2-4.5 and SSP5-8.5 were reflected in the corrected projections, with SSP5-8.5 producing larger increases in SU and TR, decreases in FD, and amplified changes toward FF.

4. Spatial STDEV showed scenario-dependent spatial variability consistent with that of the benchmark product. Future work will extend BHRR to additional GCMs, regions, and non-Gaussian variables such as precipitation.

From a computational perspective, the combined Restormer and ViT training cost was on the order of 72 hours on an NVIDIA A100 environment, while inference for a full domain was completed in under 30 seconds. The source code and trained weights are publicly available to support reproducibility and further development.

Competing interests

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

Author contributions

Young Hoon Song: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing. **Hyung Ju Kim**: Conceptualization, Data curation, Validation, Visualization, Methodology, Writing – original draft. **Eun-Sung Chung**: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Acknowledgements

This study was supported by National Research Foundation of Korea (NRF) (RS-2023-00246767).

Data availability

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

The input datasets and example BHRR v1.0 outputs are archived on Zenodo (Song et al., 2026b; https://doi.org/10.5281/zenodo.20152297) under CC BY 4.0, with the CMIP6-derived subfolders under CC BY-SA 4.0. The trained model weights and source code are in the companion code archive (Song et al., 2026a; https://doi.org/10.5281/zenodo.19441661). The upstream datasets used in this study are the CMIP6 ACCESS-CM2 r1i1p1f1 v20191108 historical (Dix et al., 2019a; https://doi.org/10.22033/ESGF/CMIP6.4271), SSP245 (Dix et al., 2019b; https://doi.org/10.22033/ESGF/CMIP6.4321), and SSP585 (Dix et al., 2019c; https://doi.org/10.22033/ESGF/CMIP6.4332) simulations from the Earth System Grid Federation; the NEX-GDDP-CMIP6 v1.0 benchmark (Thrasher et al., 2022; https://doi.org/10.7917/OFSG3345); and the Princeton Global Forcing v3 reference (Sheffield et al., 2006; https://doi.org/10.1175/JCLI3790.1). All resources last accessed 13 May 2026.

Code availability

The BHRR v1.0 framework is implemented in Python using PyTorch. The exact version used to produce all results in this paper is permanently archived on Zenodo at https://doi.org/10.5281/zenodo.19441661 (Song et al., 2026a) under the CC BY 4.0 license. The archive contains the training and inference scripts, model definitions for both the Restormer-based spatial-restoration stage and the ViT-based quantile bias-correction stage, utility routines, and an example notebook for reproducing the main experiments. The ongoing development version is hosted at https://github.com/soooynon13-sys/BHRR. The Restormer module is based on the official implementation of Zamir et al. (2022), whose backbone is available at https://github.com/swz30/Restormer and is not redistributed in this archive.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

References

1. Banõ-Medina, J., Manzanas, R., Cimadevilla, E., Fernandez, J., Gonzalez-Abad, J., Cofinõ, A. S., and Gutierrez, J. M.: Downscaling multi-model climate projection ensembles with deep learning (DeepESD): Contribution to CORDEX EUR-44, Geosci. Model Dev., 15, 6747–6758, https://doi.org/10.5194/gmd-15-6747-2022, 2022.
2. Bao, F., Nie, S., Xue, K., Cao, Y., Li, C., Su, H., and Zhu, J.: All are Worth Words: A ViT Backbone for Diffusion Models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22669–22679, https://doi.org/10.1109/CVPR52729.2023.02171, 2023.
3. Birkmann, J., Jamshed, A., McMillan, J. M., Feldmeyer, D., Totin, E., Solecki, W., Ibrahim, Z. Z., Roberts, D., Kerr, R. B., Poertner, H. O., Pelling, M., Djalante, R., Garschagen, M., Leal Filho, W., Guha-Sapir, D., and Alegría, A.: Understanding human vulnerability to climate change: A global perspective on index validation for adaptation planning, Sci. Total Environ., 803, https://doi.org/10.1016/j.scitotenv.2021.150065, 2022.
4. Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., Meehl, G. A., Predoi, V., Roberts, M. J., and Eyring, V.: Quantifying Progress Across Different CMIP Phases With the ESMValTool, J. Geophys. Res.-Atmos., 125, https://doi.org/10.1029/2019JD032321, 2020.
5. Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., and Knutti, R.: Reduced global warming from CMIP6 projections when weighting models by performance and independence, Earth Syst. Dynam., 11, 995–1012, https://doi.org/10.5194/esd-11-995-2020, 2020.
6. Buhl, M. and Markolf, S.: A review of emerging strategies for incorporating climate change considerations into infrastructure planning, design, and decision making, Sustain. Resilient Infrastruct., 8, 157–169, https://doi.org/10.1080/23789689.2022.2134646, 2023.
7. Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Clim. Dyn., 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2018.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

8. Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?, J. Climate, 28, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.
9. Chae, S. T., Chung, E. S., and Jiang, J.: Robust Siting of Permeable Pavement in Highly Urbanized Watersheds Considering Climate Change Using a Combination of Fuzzy-TOPSIS and the VIKOR Method, Water Resour. Manag., 36, 951–969, https://doi.org/10.1007/s11269-022-03062-y, 2022.
10. Chen, X., Li, Z., Pu, Y., Liu, Y., Zhou, J., Qiao, Y., and Dong, C.: A Comparative Study of Image Restoration Networks for General Backbone Network Design, in: Computer Vision – ECCV 2024, Lecture Notes in Computer Science, vol. 15129, Springer, Cham, https://doi.org/10.1007/978-3-031-73209-6_5, 2024.
11. Curran, D., Saleem, H., Hobeichi, S., and Salim, F.: Resolution-Agnostic Transformer-based Climate Downscaling, in: Tackling Climate Change with Machine Learning Workshop, NeurIPS 2024, arXiv:2411.14774, 2024.
12. Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., Mackallah, C., Marsland, S., O'Farrell, S., Rashid, H., Srbinovsky, J., Sullivan, A., Trenham, C., Vohralik, P., Watterson, I., Williams, G., Woodhouse, M., Bodman, R., Dias, F. B., Domingues, C. M., Hannah, N., Heerdegen, A., Savita, A., Wales, S., Allen, C., Druken, K., Evans, B., Richards, C., Ridzwan, S. M., Roberts, D., Smillie, J., Snow, K., Ward, M., and Yang, R.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP historical, version v20191108, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.4271, 2019a.
13. Dix, M., Bi, D., Dobrohotoff, P., et al.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 ScenarioMIP ssp245, version v20191108, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.4321, 2019b.
14. Dix, M., Bi, D., Dobrohotoff, P., et al.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 ScenarioMIP ssp585, version v20191108, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.4332, 2019c.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

15. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, in: International Conference on Learning Representations (ICLR), arXiv:2010.11929, 2021.
16. Deroubaix, A., Labuhn, I., Camredon, M., Gauber, B., Monerie, P.-A., Popp, M., Raber, J., Rue, C., Cuesta, J., and Dufresne, J.-L.: Large uncertainties in trends of energy demand for heating and cooling under climate change, Nat. Commun., 12, 5197, https://doi.org/10.1038/s41467-021-25504-8, 2021.
17. Evin, G., Ribes, A., and Corre, L.: Assessing CMIP6 uncertainties at global warming levels, Clim. Dyn., 62, 8057–8072, https://doi.org/10.1007/s00382-024-07323-x, 2024a.
18. Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Shahzad, M., Yang, W., Bamler, R., and Zhu, X. X.: A Survey of Uncertainty in Deep Neural Networks, Artif. Intell. Rev., 56, 1513–1589, https://doi.org/10.1007/s10462-023-10562-9, 2023.
19. Gergel, D. R., Nijssen, B., Abatzoglou, J. T., Lettenmaier, D. P., and Stumbaugh, M. R.: Effects of climate change on snowpack and fire potential in the western USA, Clim. Change, 141, 287–299, https://doi.org/10.1007/s10584-017-1899-y, 2017.
20. Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bull., 58, 175–183, 2009.
21. Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – A comparison of methods, Hydrol. Earth Syst. Sci., 16, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012, 2012.
22. Hollins, S. E., Hughes, C. E., Crawford, J., Cendón, D. I., and Meredith, K. M.: Rainfall isotope variations over the Australian continent – Implications for hydrology and isoscape applications, Sci. Total Environ., 645, 630–645, https://doi.org/10.1016/j.scitotenv.2018.07.082, 2018.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

23. Hüllermeier, E. and Waegeman, W.: Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods, Mach. Learn., 110, 457–506, https://doi.org/10.1007/s10994-021-05946-3, 2021.
24. IPCC: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., https://doi.org/10.1017/9781009325844, 2022.
25. Islam, S., Elmekki, H., Elsebai, A., Bentahar, J., Drawel, N., Rjoub, G., and Pedrycz, W.: A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks, Expert Syst. Appl., 241, 122666, https://doi.org/10.1016/j.eswa.2023.122666, 2024.
26. Kendall, A. and Gal, Y.: What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?, in: Advances in Neural Information Processing Systems (NeurIPS) 30, pp. 5580–5590, 2017.
27. Kendon, E. J., Jones, R. G., Kjellström, E., and Murphy, J. M.: Using and designing GCM-RCM ensemble regional climate projections, J. Climate, 23, 6485–6503, https://doi.org/10.1175/2010JCLI3502.1, 2010.
28. Kim, Y. H., Min, S. K., Zhang, X., Sillmann, J., and Sandstad, M.: Evaluation of the CMIP6 multi-model ensemble for climate extreme indices, Weather Clim. Extrem., 29, https://doi.org/10.1016/j.wace.2020.100269, 2020.
29. LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, https://doi.org/10.1038/nature14539, 2015.
30. Lee, B. S. and Haran, M.: A Class of Models for Large Zero-inflated Spatial Data, J. Agric. Biol. Environ. Stat., 30, 746–768, https://doi.org/10.1007/s13253-024-00619-9, 2025.
31. Liang, J., Cao, J., Fan, Y., Zhang, K., Ranjan, R., Li, Y., Timofte, R., and Van Gool, L.: VRT: A Video Restoration Transformer, IEEE T. Image Process., 33, 2171–2182, https://doi.org/10.1109/TIP.2024.3372454, 2024.
32. Liang, X. Z., Kunkel, K. E., Meehl, G. A., Jones, R. G., and Wang, J. X. L.: Regional climate models downscaling analysis of general circulation models present climate biases propagation into future change projections, Geophys. Res. Lett., 35, https://doi.org/10.1029/2007GL032849, 2008.

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

33. Liu, C., Yang, H., Fu, J., and Qian, X.: Learning Trajectory-Aware Transformer for Video Super-Resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5677–5686, https://doi.org/10.1109/CVPR52688.2022.00560, 2022.
34. Mackallah, C., Chamberlain, M. A., Law, R. M., Dix, M., Ziehn, T., Bi, D., Bodman, R., Brown, J. R., Dobrohotoff, P., Druken, K., Evans, B., Harman, I. N., Hayashida, H., Holmes, R., Kiss, A. E., Lenton, A., Liu, Y., Marsland, S., Meissner, K., Menviel, L., O'Farrell, S., Rashid, H. A., Ridzwan, S., Savita, A., Srbinovsky, J., Sullivan, A., Trenham, C., Vohralik, P. F., Wang, Y. P., Williams, G., Woodhouse, M. T., and Yeung, N.: ACCESS datasets for CMIP6: Methodology and idealised experiments, J. South. Hemisph. Earth Syst. Sci., 72, 93–116, https://doi.org/10.1071/ES21031, 2022.
35. Maurer, E. P. and Pierce, D. W.: Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean, Hydrol. Earth Syst. Sci., 18, 915–925, https://doi.org/10.5194/hess-18-915-2014, 2014.
36. Merrifield, A. L., Brunner, L., Lorenz, R., Humphrey, V., and Knutti, R.: Climate model Selection by Independence, Performance, and Spread (ClimSIPS v1.0.1) for regional applications, Geosci. Model Dev., 16, 4715–4747, https://doi.org/10.5194/gmd-16-4715-2023, 2023.
37. Nair, M. M., Rajesh, A. N., Sahai, A. K., and Lakshmi Kumar, T. V.: Quantification of uncertainties in projections of extreme daily precipitation simulated by CMIP6 GCMs over homogeneous regions of India, Int. J. Climatol., 43, 7365–7380, https://doi.org/10.1002/joc.8269, 2023.
38. Peebles, W. and Xie, S.: Scalable Diffusion Models with Transformers, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4195–4205, https://doi.org/10.1109/ICCV51070.2023.00387, 2023.
39. Perkins-Kirkpatrick, S. E. and Lewis, S. C.: Increasing trends in regional heatwaves, Nat. Commun., 11, 3357, https://doi.org/10.1038/s41467-020-16970-7, 2020.
40. Rampal, N., Gibson, P. B., Sherwood, S., Abramowitz, G., and Hobeichi, S.: A Reliable Generative Adversarial Network Approach for Climate Downscaling and

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

Weather Generation, J. Adv. Model. Earth Syst., 17, https://doi.org/10.1029/2024MS004668, 2025.

41. Rampal, N., Hobeichi, S., Gibson, P. B., Baño-Medina, J., Abramowitz, G., Beucler, T., González-Abad, J., Chapman, W., Harder, P., and Gutiérrez, J. M.: Enhancing Regional Climate Downscaling through Advances in Machine Learning, Artif. Intell. Earth Syst., 3, https://doi.org/10.1175/aies-d-23-0066.1, 2024.
42. Risbey, J. S., Pook, M. J., McIntosh, P. C., Wheeler, M. C., and Hendon, H. H.: On the remote drivers of rainfall variability in Australia, Mon. Weather Rev., 137, 3233–3253, https://doi.org/10.1175/2009MWR2861.1, 2009.
43. Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S., and Perera, B. J. C.: Statistical downscaling of precipitation using machine learning techniques, Atmos. Res., 212, 240–258, https://doi.org/10.1016/j.atmosres.2018.05.022, 2018.
44. Sha, Y., Gagne, D. J., West, G., and Stull, R.: Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part I: Daily maximum and minimum 2-m temperature, J. Appl. Meteorol. Clim., 59, 2057–2073, https://doi.org/10.1175/JAMC-D-20-0057.1, 2020.
45. Sharmila, S. and Hendon, H. H.: Mechanisms of multiyear variations of Northern Australia wet-season rainfall, Sci. Rep., 10, https://doi.org/10.1038/s41598-020-61482-5, 2020.
46. Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, J. Climate, 19, 3088-3111, https://doi.org/10.1175/JCLI3790.1, 2006.
47. Smith, C. J. and Forster, P. M.: Suppressed Late-20th Century Warming in CMIP6 Models Explained by Forcing and Feedbacks, Geophys. Res. Lett., 48, https://doi.org/10.1029/2021GL094948, 2021.
48. Song, Y. H. and Chung, E.-S.: Intercomparison of bias correction methods for precipitation of multiple GCMs across six continents, Geosci. Model Dev., 18, 8017–8045, https://doi.org/10.5194/gmd-18-8017-2025, 2025.
49. Song, Y. H., Kim, H. J., and Chung, E.-S.: BHRR: a two-stage Transformer framework for simultaneous spatial restoration and quantile-function bias

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

correction of climate model temperature fields (Version 1.0.0), Zenodo [code], https://doi.org/10.5281/zenodo.19441661, 2026a.

50. Song, Y. H., Kim, H. J., and Chung, E.-S.: BHRR v1.0 - Data Archive (Version v1.0.0), Zenodo [data set], https://doi.org/10.5281/zenodo.20152297, 2026b.
51. Song, Y. H., Chung, E.-S., and Shahid, S.: Global Future Climate Signal by Latitudes Using CMIP6 GCMs, Earths Future, 12, https://doi.org/10.1029/2022EF003183, 2024.
52. Song, Y. H., Chung, E.-S., and Shahid, S.: The New Bias Correction Method for Daily Extremes Precipitation over South Korea using CMIP6 GCMs, Water Resour. Manag., 36, 5977–5997, https://doi.org/10.1007/s11269-022-03338-3, 2022.
53. Song, Y. H., Nashwan, M. S., Chung, E.-S., and Shahid, S.: Advances in CMIP6 INM-CM5 over CMIP5 INM-CM4 for precipitation simulation in South Korea, Atmos. Res., 247, https://doi.org/10.1016/j.atmosres.2020.105261, 2021.
54. Sun, Y., Deng, K., Ren, K., Liu, J., Deng, C., and Jin, Y.: Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review, ISPRS J. Photogramm., 208, 14–38, https://doi.org/10.1016/j.isprsjprs.2023.12.011, 2024.
55. Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., Van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K., Hurtt, G., Kriegler, E., Meehl, G., Moss, R., Bauer, S., Boucher, O., Brovkin, V., Yhb, Y., Dix, M., Gualdi, S., Guo, H., John, J., Kharin, S., Kim, Y. H., Koshiro, T., Ma, L., Olivié, D., Panickal, S., Qiao, F., Rong, X., Rosenbloom, N., Schupfner, M., Séférian, R., Sellar, A., Semmler, T., Shi, X., Song, Z., Steger, C., Stouffer, R., Swart, N., Tachiiri, K., Tang, Q., Tatebe, H., Voldoire, A., Volodin, E., Wyser, K., Xin, X., Yang, S., Yu, Y., and Ziehn, T.: Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6, Earth Syst. Dynam., 12, 253–293, https://doi.org/10.5194/esd-12-253-2021, 2021.
56. Themeßl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate

https://doi.org/10.5194/egusphere-2026-1958
Preprint. Discussion started: 20 May 2026

change signal, Clim. Change, 112, 449–468, https://doi.org/10.1007/s10584-011-0224-4, 2012.

57. Thrasher, B., Wang, W., Michaelis, A., Melton, F., Lee, T., and Nemani, R.: NASA Global Daily Downscaled Projections, CMIP6, Sci. Data, 9, https://doi.org/10.1038/s41597-022-01393-4, 2022.
58. Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R.: Past warming trend constrains future warming in CMIP6 models, Sci. Adv., 6, eaaz9549, https://doi.org/10.1126/sciadv.aaz9549, 2020.
59. Vogel, E., Johnson, F., Marshall, L., Bende-Michl, U., Wilson, L., Peter, J. R., Wasko, C., Srikanthan, S., Sharples, W., Dowdy, A., Hope, P., Khan, Z., Mehrotra, R., Sharma, A., Matic, V., Oke, A., Turner, M., Thomas, S., Donnelly, C., and Duong, V. C.: An evaluation framework for downscaling and bias correction in climate change impact studies, J. Hydrol., 622, https://doi.org/10.1016/j.jhydrol.2023.129693, 2023.
60. Wang, F. and Tian, D.: Multivariate bias correction and downscaling of climate models with trend-preserving deep learning, Clim. Dyn., 62, 9651–9672, https://doi.org/10.1007/s00382-024-07406-9, 2024.
61. Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: From error visibility to structural similarity, IEEE T. Image Process., 13, 600–612, https://doi.org/10.1109/TIP.2003.819861, 2004.
62. You, Q., Cai, Z., Wu, F., Jiang, Z., Pepin, N., and Shen, S. S. P.: Temperature dataset of CMIP6 models over China: evaluation, trend and uncertainty, Clim. Dyn., 57, 17–35, https://doi.org/10.1007/s00382-021-05691-2, 2021.
63. Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., and Yang, M.-H.: Restormer: Efficient Transformer for High-Resolution Image Restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5718–5729, https://doi.org/10.1109/CVPR52688.2022.00564, 2022.
64. Zhao, J., Gan, T. Y., Zhang, G., and Zhang, S.: Projected changes of precipitation extremes in North America using CMIP6 multi-climate model ensembles, J. Hydrol., 621, https://doi.org/10.1016/j.jhydrol.2023.129598, 2023.