We sincerely thank the reviewer for the careful reading of our manuscript and for the constructive comments, which have helped us substantially improve the clarity and rigor of the paper. We realized that several of the reviewer's concerns (in particular Comments 1, 2, 6, and 9) originated from a single omission on our part: the manuscript never stated the native horizontal resolution of the ACCESS-CM2 input fields. We have corrected this omission throughout the revised manuscript and figures. Below, the reviewer's comments are reproduced in italics, followed by our point-by-point responses. Line numbers refer to the submitted (preprint) version unless otherwise noted.

Comment 1

This is what confuses the reviewer the most: BHRR is trained with ACCESS-CM2 data as the input and PGFv3 data as the target, both at the same resolution of 0.25 degrees. It seems to the reviewer that the BHRR tool has no downscaling capabilities of improving the resolution of the coarse data from climate models?

Response

We thank the reviewer for raising this important point, and we apologize for the confusion, which we now realize stemmed from an omission in our description. The ACCESS-CM2 atmospheric output is provided on its native N96 grid of approximately 1.875° longitude × 1.25° latitude (192 × 144 global grid cells; nominal resolution ≈250 km in the CMIP6 metadata; Bi et al., 2020), which is substantially coarser than the 0.25° reference grid. Over the Oceania domain used in this study, the native grid contains only 44 × 39 cells, compared with 200 × 280 cells at 0.25°.

As described in Sect. 2.2, we first regrid this coarse field to 0.25° by linear interpolation. The resulting field lies on the fine grid, but it is spatially blurred and carries only the coarse-scale information content of the native model grid linear interpolation cannot create the fine-scale structure that the native grid never resolved. The Restormer stage then restores the fine-scale spatial structure (coastlines, islands, land-sea contrasts, and regional gradients) that is absent from the interpolated input. The input and target therefore share the same nominal grid, but not the same effective resolution: the workflow constitutes an effective resolution enhancement from ≈250 km to 0.25° (≈25 km). This is the same logic as the widely used NEX-GDDP-CMIP6 product, in which coarse native GCM fields are downscaled to 0.25° and bias-corrected against the same reference (Thrasher et al., 2022). We deliberately frame the task as "image restoration from a linearly upscaled input" rather than pure super-resolution (see the first Discussion topic), because the GCM and the reference also differ structurally, not only in resolution.

We have revised the manuscript as follows:

Sect. 2.2, Lines 194-199:

Raw ACCESS-CM2 atmospheric fields are provided on the native N96 grid (approximately 1.875° longitude × 1.25° latitude, 192 × 144 global grid cells; nominal resolution ≈250 km; Bi

et al., 2020), which is substantially coarser than the 0.25° reference grid; over the Oceania domain the native grid contains only 44 × 39 cells, compared with 200 × 280 cells at 0.25°.

Sect. 2.2:

Linear interpolation places the coarse fields on the fine grid but cannot create fine-scale structure that the native grid never resolved. The interpolated inputs are therefore spatially blurred and retain only the coarse-scale information content of the native grid. The restoration stage subsequently recovers fine-scale structure consistent with the 0.25° reference, which constitutes an effective resolution enhancement from approximately 250 km to 25 km.

Introduction:

Because the driving GCM operates on a much coarser native grid than the 0.25° reference, this restoration constitutes an effective resolution enhancement rather than a same-resolution mapping.

To strengthen the manuscript, we have added the following reference(s) to the reference list:

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Sullivan, A., Bodman, R., Law, R., Harman, I., Srbinovsky, J., Rashid, H. A., Dobrohotoff, P., Mackallah, C., Yan, H., Hirst, A., Savita, A., Dias, F. B., Woodhouse, M., Fiedler, R., and Heerdegen, A.: Configuration and spin-up of ACCESS-CM2, the new generation Australian Community Climate and Earth System Simulator Coupled Model, J. South. Hemisph. Earth Syst. Sci., 70, 225–251, https://doi.org/10.1071/ES19040, 2020.

Comment 2

Following the previous question: since current global climate models can already perform simulations at 0.25 degree resolution, why would we need BHRR which produces post-processed outputs still in 0.25 degree resolution?

Response

We respectfully clarify that most CMIP6-class GCMs, including ACCESS-CM2 (≈250 km), are not run at 0.25°. Global simulations at or near 0.25° are currently restricted to a small number of HighResMIP-type experiments covering limited periods, ensemble sizes, and scenarios, at very high computational cost (Haarsma et al., 2016). For the multi-scenario, multi-decadal projections required by impact and engineering applications (e.g., the SSP2-4.5 and SSP5-8.5 projections to 2100 used in this study), standard-resolution CMIP6 simulations remain the primary available source, and statistical post-processing of these simulations remains the standard practice. The purpose of BHRR is therefore not to duplicate native high-resolution simulation, but to perform, within a single workflow, (i) restoration of realistic fine-scale spatial structure from coarse GCM fields and (ii) distribution-aware bias correction. This is precisely the role of the widely adopted NEX-GDDP-CMIP6 product, against which BHRR shows improved RMSE, PBIAS, distributional alignment, and extreme-index fidelity (Figs. 4–6).

We have revised the manuscript as follows:

Introduction:

In particular, most CMIP6-class GCMs are not run at 0.25°. Global simulations at or near 0.25° remain restricted to a few HighResMIP-type experiments with limited periods, ensemble sizes, and scenarios at very high computational cost (Haarsma et al., 2016). Statistical post-processing of standard-resolution CMIP6 simulations therefore remains the principal route to the 0.25°-scale information required by impact and engineering applications.

To strengthen the manuscript, we have added the following reference(s) to the reference list:

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model

Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Comment 3

To improve structural clarity of the manuscript, please consider reorganizing the Introduction section. Any lines after line 122 can be moved to a later section: specifically, line 122-152 to the Methods section and line 153-169 to the Conclusions section. If the novelty of the work needs to be stressed in the Introduction section, make it concise.

Response

We agree that the Introduction was overloaded and have streamlined it as suggested. The detailed descriptions of the study domain, datasets, and scenarios (L122-152), which overlapped substantially with Sect. 2, have been condensed or moved into the Methods section. Regarding L153-169, we have condensed the statement of contributions to three concise single-sentence items; following common practice in GMD and the broader literature, we would prefer to retain this brief contributions statement and the short paper-organization paragraph in the Introduction rather than move them to the Conclusions, and we hope the reviewer finds the shortened version acceptable.

We have revised the manuscript as follows:

Introduction (domain/dataset paragraph, condensed):

The framework is implemented over a fixed 200 × 280 Oceania domain, which offers strong spatial heterogeneity including land–ocean contrasts and diverse climate zones within a tractable computational domain. A single driving GCM and three near-surface temperature variables (Tas, Tmax, and Tmin) are used. Princeton Global Forcing version 3 (PGFv3) is adopted as the observation-based reference dataset, and NEX-GDDP-CMIP6 is used as the benchmark for a like-for-like comparison. Future projections are examined under SSP2-4.5 and SSP5-8.5.

Sect. 2.2 (SSP description moved from Introduction Lines 142–152; the remainder of that paragraph was deleted because the evaluation metrics are described in Sects. 2.4–2.7):

SSP2-4.5 is an intermediate mitigation pathway reaching a radiative forcing of approximately 4.5 W/m^2 by 2100, whereas SSP5-8.5 is a high-emission pathway reaching approximately 8.5 W/m^2 by 2100 (Tebaldi et al., 2021).

Introduction (contribution 1, condensed):

(1) A two-stage Transformer post-processing framework (BHRR v1.0) that decouples Restormer-based spatial restoration from ViT-based distributional correction, addressing the inability of pixel-wise image regression to remove systematic distributional biases.

Introduction (contribution 2, condensed):

(2) A quantile-function formulation in which the ViT predicts a reference-based quantile map used as an explicit transfer function for equidistant CDF matching, so that the learned correction can be applied to future projections while preserving CDF-based ranks.

Introduction (contribution 3, condensed):

(3) A like-for-like evaluation against NEX-GDDP-CMIP6 over Oceania for three temperature variables, including climate-change signal preservation under two SSP scenarios using ETCCDI indices, Sen's slope, and spatial STDEV.

Comment 4

In line 138: It is vague what "observation-based reference dataset" means. Does it mean a high-resolution regional assimilation dataset? If so, please be explicit in description.

Response

"Observation-based reference dataset" refers to the Princeton Global Forcing dataset version 3 (PGFv3). It is not a regional assimilation product. PGFv3 is a global, 0.25°, daily meteorological forcing dataset constructed by bias-correcting reanalysis fields against gridded station-based observational products (Sheffield et al., 2006). We adopted PGFv3 because it is the same reference climatology used in the NEX-GDDP-CMIP6 BCSD procedure (Thrasher et al., 2022); using a common reference for both BHRR and the benchmark ensures that performance differences reflect the post-processing methodology rather than differences in reference climatology. The condensed Introduction paragraph (Comment 3) now also names PGFv3 explicitly at first mention.

We have revised the manuscript as follows:

Sect. 2.2:

PGFv3 is an observation-based global meteorological forcing dataset constructed by bias-correcting reanalysis fields against gridded station-based observational products (Sheffield et al., 2006). It is a global product rather than a regional assimilation dataset. It is also 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.

To strengthen the manuscript, we have added the following reference(s) to the reference list:

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.

Comment 5

In line 231: What does it mean by “bias correction in quantile-function space”. This description seems opaque to anyone not very familiar with the specific topic and practices. It will be helpful to elaborate on this for the broader audience.

Response

We thank the reviewer for this suggestion. The phrase "bias correction in quantile-function space" first appears in the workflow overview (Sect. 2.3, L231), and we agree that it needed a plain-language explanation at that point. We have added two sentences immediately after the first occurrence. The first explains the intuition, namely that the correction adjusts the statistical distribution of values at each grid cell rather than the individual daily values, and that each future day is then transferred to the value of equal rank in the corrected distribution. The second points the reader to Sect. 2.3.2, where the concept is developed in full detail. Specifically, Sect. 2.3.2 explains that (i) the quantile maps are computed post-hoc from the complete time series of Restormer-restored daily fields, with the empirical quantile function evaluated at 99 probability levels at each grid cell, (ii) the ViT is trained to translate the Restormer-derived quantile map Q_R into the reference-based quantile map Q_{ref}, and (iii) each future daily value is corrected by equidistant CDF matching (Eqs. 11–12), which preserves the CDF-based rank under the historical Restormer distribution and retrieves the value of the same rank from the predicted reference distribution. Together with the explicit definition of the quantile map added in response to Comment 9, we believe the concept is now accessible to a broader audience.

We have revised the manuscript as follows:

Sect. 2.3:

In other words, instead of correcting each daily value directly, the method adjusts the statistical distribution of values at each grid cell and then transfers each future day to the value of equal rank in the corrected distribution.

Comment 6

In Figure 1: The squares before and after the “linear interpolation” stage should have the same overall size (with different grid resolution), right? For now, the way the figure is drawn obscures the idea being conveyed.

Response

We thank the reviewer for this suggestion. The reviewer is correct. The spatial extent is identical before and after interpolation, and only the grid density changes. We have redrawn Fig. 1 so that the squares before and after the linear-interpolation stage share the same overall size and spatial extent. The native coarse grid is now shown as visibly larger cells and the 0.25° grid as a dense mesh within a square of the same size, and the two grids are annotated directly in the figure as "Native GCM grid (~1.875° × 1.25°, ≈250 km)" and "0.25° (same spatial extent, only the grid density changes)". The caption has also been revised so that it reflects the two roles of the framework, namely spatial restoration and bias correction, rather than a simple grid conversion. We believe the redrawn figure also resolves the source of confusion behind Comments 1 and 2, since the resolution gap between the native GCM grid and the 0.25° reference grid is now visually explicit.

We have revised the manuscript as follows:

Sect. 2.3, Fig. 1:

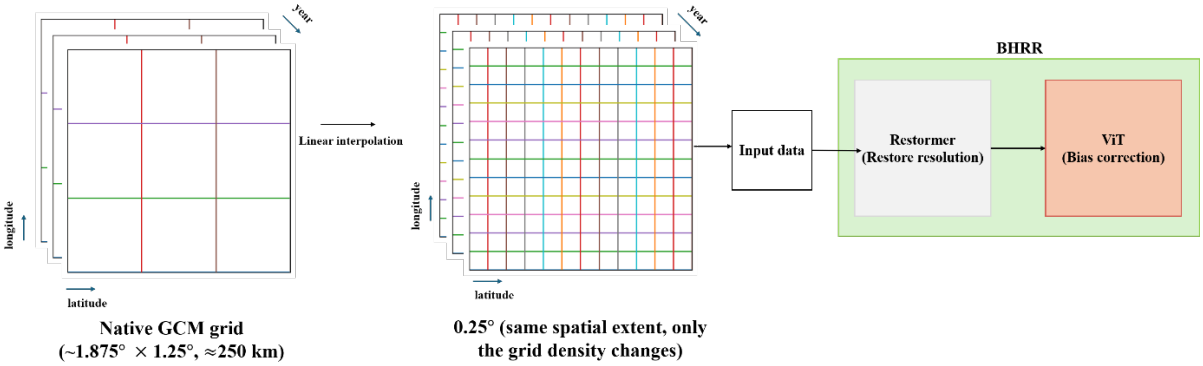

Fig. 1. Overall architecture and processing workflow of the BHRR framework for spatial restoration and bias correction of coarse-resolution GCM temperature fields

Comment 7

In Line 243: It is stated that “The restoration training workflow is summarized in Fig. 2”, but Figure 2 shows something different: Overview of ViT network architecture for bias correction.

Response

We thank the reviewer for catching this error. The cross-reference at L243 was incorrect. The sentence was intended to refer to the Restormer architecture, which is provided in the Supplement (Fig. S2), not to Fig. 2 of the main text, which shows the ViT architecture for the bias-correction stage. We have corrected the sentence so that it points to Fig. S2 and rephrased it to describe the architecture of the restoration module. We have also re-checked every figure and table citation in the manuscript against the captions to ensure full consistency.

We have revised the manuscript as follows:

Sect. 2.3.1:

The architecture of the restoration module is summarized in Fig. S2 in the Supplement.

Fig. S2. Restormer (Zamir et al., 2021) architecture used for patch-based climate field restoration

Comment 8

In Line 302: For neural network inference task on a 200x280 grid, 15 seconds for a single inference seems extremely long. What might be the speed bottleneck here?

Response

The reviewer's intuition is correct: the pure network forward pass is much faster than 15 s. The originally reported "≈15 s" was an end-to-end wall-clock time for a one-off single-field run, dominated by one-time costs rather than by the inference itself. We profiled the inference path stage by stage (NVIDIA RTX 3070, PyTorch 2.6, while the original figure was obtained on an A100, the decomposition is hardware-generic). The one-time costs are approximately 3-5 s for the Python and deep-learning library import, about 0.45 s for model construction, checkpoint loading (72.7 MB), and GPU transfer, and about 5.1 s for the first tiled inference including one-time CUDA/cuDNN kernel initialization. The recurring per-field costs are approximately 1.05 s for the steady-state sliding-tile inference per 200×280 field (24 tiles of 64×64, sequential forwards), about 0.09 s when all 24 tiles are batched into a single forward pass, and about 0.18 s per daily field for the CPU-side per-pixel CDF lookup of the ViT stage (about 0.016 s vectorized). A cold single-field run therefore totals approximately 10-12 s, consistent with the reported value of about 15 s, but more than 90% of it is one-time initialization paid once per session. In batched production runs (31,411 daily fields), sustained throughput was approximately 0.5 s per field including disk I/O.

We have revised the manuscript as follows:

Sect. 2.3.1 (Restormer training and inference):

All Restormer experiments are conducted on an NVIDIA A100 GPU, and end-to-end training requires approximately 48 h. After one-time model loading, inference takes approximately 1 s per 200 × 280 field and about 0.5 s per field in sustained batched processing.

Sect. 2.3.2 (ViT training and inference):

All ViT experiments are conducted on an NVIDIA A100 GPU, and end-to-end training requires approximately 24 h. Bias-correction inference comprises a single ViT forward pass and a pixel-wise CDF lookup, which together take approximately 0.2 s per daily field.

Discussion (second paragraph):

In this study, BHRR required approximately 72 hours for training (Restormer and ViT combined), while inference is completed in approximately one second per full-domain field, indicating rapid deployment after one-time training.

Conclusions (computational-perspective paragraph):

From a computational perspective, the combined Restormer and ViT training cost was on the order of 72 hours on an NVIDIA A100 environment, and inference is completed in approximately one second per full-domain field. The source code and trained weights are publicly available to support reproducibility and further development.

Comment 9

In Section 2.3.2: The term “quantile map” is extensively used in the text, but there lacks an explanation of it definition. Does it mean the statistical value distribution of the temperature variables on each grid of the domain? Is a “quantile-map” two-dimensional or three-dimensional, or more? Since this definition is central to the bias-correction section, it deserves a separate figure.

Response

We thank the reviewer for this suggestion, and the reviewer's reading is exactly right. A quantile map encodes the statistical value distribution of the temperature variable at every grid cell. It is a three-dimensional array of size T × H × W, in which each grid cell stores its empirical quantile function evaluated at the probability levels $\tau \in \{0.01, \cdots, 0.99\}$, with the probability level acting as the third (channel) dimension. Section 2.3.2 already described how the quantile map is constructed post-hoc from the complete Restormer-restored historical series, but the dimensionality was not stated explicitly. We have revised the corresponding sentence in Sect. 2.3.2 to state the array dimensions. Following the reviewer's suggestion, we have also added a dedicated schematic of the quantile-map construction as a new panel (a) of Fig. 2, built from the actual Restormer-restored Tas series at one grid cell, illustrating the per-cell daily time series, the empirical quantile function evaluated at 99 probability levels, and the stacking of all cells into the three-dimensional array. Panel (b) retains the ViT architecture, so no figure renumbering is required.

We have revised the manuscript as follows:

Sect. 2.3.2 (quantile-map construction, panel reference added):

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 (Fig. 2a).

Sect. 2.3.2 (explicit dimensional definition, sentence revised):

Each quantile map is therefore a three-dimensional array of size T × H × W, in which each grid cell stores its quantile function evaluated at the probability levels $\tau \in \{0.01, \cdots, 0.99\}$ and the probability level acts as the channel dimension (Fig. 2a).

Sect. 2.3.2 (ViT architecture, panel reference updated):

The ViT architecture (Fig. 2b) consists of patch embedding, Transformer encoder blocks, and an image reconstruction head.

Fig. 2 (new panel (a) added) and its caption:

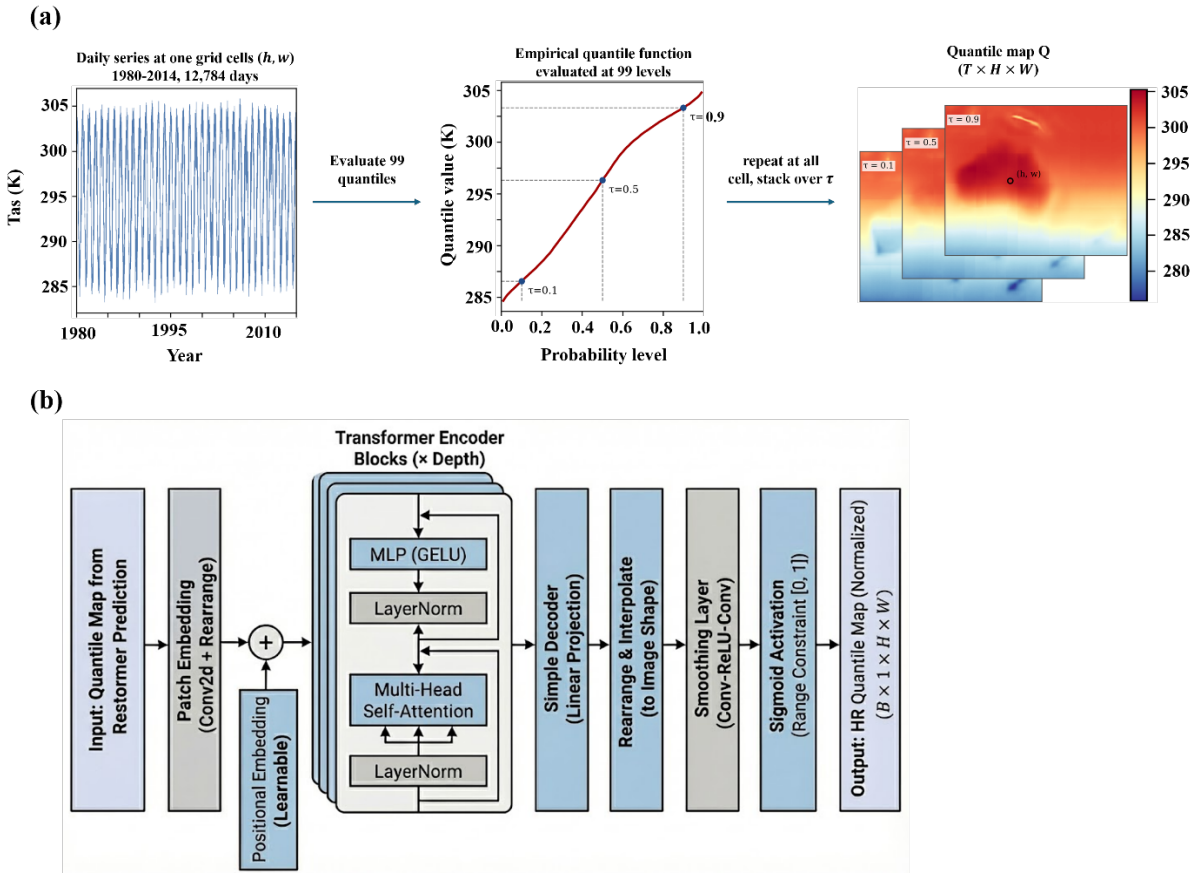

Fig. 2. (a) Construction of the quantile map from the per-cell daily time series and (b) overview of the ViT network architecture for bias correction

Comment 10

In Figure 3: Are the “Restorer” results predicted using input data not from the training dataset? This should be explicitly mentioned in the text.

Response

We thank the reviewer for this important question, and we agree that the evaluation protocol must be stated explicitly in the text. The protocol is as follows. For each variable, the 12,784 daily fields of 1980–2014 are randomly partitioned, with a fixed random seed, into a training set (80%) and a held-out validation set (20%, corresponding to 2,557 days). The validation days are excluded from all weight updates and are used only for model selection. Figure 3, as originally presented, was computed on a systematic subsample of days spanning the full historical period, which therefore includes both training and validation days.

To address the reviewer's underlying concern directly, we re-computed the Fig. 3 statistics, using the identical metric procedure, restricted to the 2,557 held-out validation days only. The results are statistically indistinguishable from the full-period values, with differences of at most 0.001 in median SSIM and 0.1 dB in median PSNR, indicating that the reported performance does not rest on memorization of training samples. The comparison is provided in the new Table S1 in the Supplement.

Table S1. Median SSIM and PSNR of the linearly interpolated input, the Restormer output, and the full BHRR output against the PGFv3 reference over Oceania, computed with the same procedure as Fig. 3 (full 200 × 280 domain, fixed data range of 50 K) for the full historical period (1980-2014, 12,784 days) and for the held-out 20% validation subset (2,557 days that contributed no weight updates during training). The close agreement between the two evaluation sets indicates that the reported performance does not rest on memorization of training samples.

Variable	Evaluation set	SSIM	PSNR
Tmin	Full 1980–2014	0.877	26.7
Tmin	Validation-only (20%)	0.877	26.7
Tas	Full 1980–2014	0.908	28.0
Tas	Validation-only (20%)	0.908	28.0
Tmax	Full 1980–2014	0.887	27.2
Tmax	Validation-only (20%)	0.886	27.0

Two further points support this evaluation design. First, BHRR is a calibration tool rather than a forecasting model. As for the BCSD procedure underlying NEX-GDDP-CMIP6, which is calibrated on the same 1980–2014 period against the same PGFv3 reference, the standard protocol is to fit the transfer on the historical calibration period, assess calibration-period skill, and verify that the correction transfers to the application period. The genuinely out-of-sample application is the scenario projection (2015–2100) assessed in Sects. 3.3–3.4. Second, because the ACCESS-CM2 historical run is a free-running coupled simulation, its daily sequence is not weather-synchronized with the observation-based reference. The mapping learned by the Restormer is therefore a quasi-stationary structural restoration rather than a memorization of specific paired days, which is consistent with the near-identical validation-only metrics in Table S1.

We have revised the manuscript as follows:

Section 2.3.1:

For each variable, the 12,784 daily fields of 1980–2014 are randomly partitioned, with a fixed random seed, into a training set (80%) and a held-out validation set (20%, corresponding to 2,557 days). The validation days are excluded from all weight updates and are used only for model selection.

Section 2.3.1:

Performance statistics in Section 3.1 are reported over the full historical period. Supplement Table S1 confirms that the corresponding metrics computed only on the held-out validation days are statistically indistinguishable, indicating that the reported performance does not rest on memorization of training samples.

Fig. 3 caption:

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 (validation-only values are provided in Table S1)

Supplement: New Table S1 added, computed with the same procedure as Fig. 3, reporting the full-period and validation-only SSIM and PSNR for the linearly interpolated input, the Restormer output, and the full BHRR output.

Comment 11

Why do Figure 3 and Figure 4 employ different metrics for comparison: in Figure 3, SSIM and PSNR; in Figure 4, RMSE and PBIAS. Since the ViT Bias-Corrector is applied on top of the Restorer, shouldn't the same metrics be employed to assess how much more improvement the ViT Bias-Corrector adds on top of the Restorer?

Response

We thank the reviewer for this suggestion. The two figures were intentionally designed with different metrics because the two stages target different aspects of quality. Figure 3 evaluates the restoration stage, whose goal is to recover fine-scale spatial structure, and therefore uses spatial-structural fidelity metrics (SSIM and PSNR). Figure 4 evaluates the bias-correction stage, whose goal is to correct the magnitude and the systematic error of the estimated values themselves, and therefore uses value-error metrics (RMSE and PBIAS).

Nevertheless, we agree that the incremental contribution of the ViT should also be traceable in a common metric space. We have therefore added a unified table as the new Table S2 in the Supplement, reporting all four metrics (SSIM, PSNR, RMSE, and PBIAS) for the linearly interpolated input, the Restormer output, and the full BHRR output, together with NEX-GDDP-CMIP6, computed against the same PGFv3 reference. SSIM and PSNR follow the Fig. 3 procedure and RMSE and PBIAS follow the Fig. 4 procedure, so the table is directly comparable with both figures. The unified table confirms the intended division of labor between the two stages. The Restormer stage provides the spatial-fidelity gain and the bulk of the RMSE reduction relative to linear interpolation. The ViT stage removes the remaining systematic bias, bringing the median PBIAS to within 0.06% of zero for all three variables, and further improves SSIM while leaving RMSE essentially at the Restormer level, which shows that the bias-correction stage does not degrade the spatial fidelity recovered by the restoration stage. The full BHRR chain attains lower median RMSE than NEX-GDDP-CMIP6 for all three variables.

We have revised the manuscript as follows:

Section 3.1.2:

To make the incremental contribution of each stage traceable in a common metric space, Table S2 in the Supplement reports all four metrics (SSIM, PSNR, RMSE, and PBIAS) for the

linearly interpolated input, the Restormer output, and the full BHRR output, together with NEX-GDDP-CMIP6, computed against the same PGFv3 reference over 1980–2014. The bias-correction stage removes the remaining systematic bias and further improves SSIM while leaving RMSE essentially at the Restormer level, and the full chain attains lower median RMSE than NEX-GDDP-CMIP6 for all three variables.

Supplement: New Table S2 added, reporting the unified four-metric evaluation of each processing level

Table S2. Unified evaluation of each processing level against the PGFv3 reference over Oceania during 1980–2014 (medians of per-day metrics, with SSIM and PSNR computed as in Fig. 3 and RMSE and PBIAS computed as in Fig. 4 over the NEX-GDDP-CMIP6 land mask)

Variable	Processing steps	SSIM	PSNR (dB)	RMSE (K)	PBIAS (%)
Tmin	Linear interpolation (input)	0.854	25.2	4.48	0.676
	Restormer (Stage 1)	0.877	26.7	3.65	0.142
	Restormer + ViT (BHRR)	0.923	26.7	3.67	−0.005
	NEX-GDDP-CMIP6	–	–	3.87	−0.019
Tas	Linear interpolation (input)	0.891	26.3	3.77	0.211
	Restormer (Stage 1)	0.908	28	3.09	0.029
	Restormer + ViT (BHRR)	0.941	27.9	3.1	0.024
	NEX-GDDP-CMIP6	–	–	3.45	0.151
Tmax	Linear interpolation (input)	0.856	24.4	4.98	−0.559
	Restormer (Stage 1)	0.887	27.2	3.59	−0.178
	Restormer + ViT (BHRR)	0.893	26.7	3.75	−0.052
	NEX-GDDP-CMIP6	–	–	4.18	−0.009

Additional minor correction

While re-computing the Fig. 3 statistics for this response, we noticed that the variable order in two sentences of Sect. 3.1.1 did not match the order of the reported values, and we have corrected the enumeration to Tmin, Tas, and Tmax. The reported values themselves are unchanged and were verified during the re-computation.