Supplement of Geosci. Model Dev., 19, 11225–11242, 2026
https://doi.org/10.5194/gmd-19-11225-2026-supplement
© Author(s) 2026. CC BY 4.0 License.



Geoscientific Model Development
Open Access
EGU

Supplement of

A differentiable framework for general circulation model precipitation bias correction

Kamlesh Sawadekar et al.

Correspondence to: Chaopeng Shen (cshen@engr.psu.edu)

The copyright of individual parts of the supplement might differ from the article licence.

Ablation experiments supporting the δCLIMBA configuration

S1 Design of the ablation experiments

δCLIMBA combines a temporal encoder (NN_1), a spatial encoder (NN_2), and a spatial-correlation term in the training loss, static physiographic attributes as auxiliary predictors, and a monotone parametric transform of the raw climate-model series. To test whether these elements are necessary, we removed them in turn and retrained the model from scratch for each of the six CMIP6 GCMs (ACCESS-CM2, GFDL-ESM4, IPSL-CM6A-LR, MIROC6, MPI-ESM1-2-LR, MRI-ESM2-0) against the Livneh reference (Table S1). Removing NN_2 along with spatial correlation loss reduces each spatio-temporal block to a temporal encoder acting independently at each grid point. Removing the static attributes deletes elevation, slope, aspect, and land cover from the input vector. Replacing the monotone transform with direct output makes the network emit the corrected value itself, removing the structural guarantee that the correction preserves the rank ordering of the climate-model series. All other settings are identical within each test.

Two generalization tests are used similar to the ones mentioned in Section 2.3 and 2.4. The held-out-period test trains on 1979–2000, validates in 1965–1978, and tests in 2001–2014 over the conterminous United States. The held-out-region test trains on 16 HUC-02 regions, validates on region 07 (Upper Mississippi), and tests on region 05 (Ohio), which is never seen during training; the period is 1990–2014 throughout, so the only extrapolation is spatial. We report the median absolute percentage bias of nine ETCCDI precipitation indices, both with the six GCMs pooled as additional grid points and with the indices first averaged across GCMs (multi-model mean), together with the box-counting fractal dimension of the time-mean field, which measures how realistically the spatial structure of the corrected field is organized.

Table S1. Ablation configurations. Bold entries differ from the baseline. All remaining settings are identical within each test (Conv1D temporal encoder, two spatio-temporal blocks, hidden size 64, 16 seasonal neighbours, three autoregressive lags, degree-8 monotone basis). The quantile emphasis of the distributional loss is 0.9, except for the direct-output run and the held-out-region no-attention run, which use 0.5.

Experiment	Spatial attention	Static attributes	Transform
Baseline (full)	yes	yes	monotone
No spatial attention (NN_2)	no	yes	monotone
No static attributes	yes	no	monotone
Direct output	yes	yes	direct

S2 Results

Every configuration removes most of the raw climate-model bias in the held-out period. Averaged over the nine indices, the median absolute bias falls from 42.0 % for the raw GCMs to between 14.6 % and 18.6 % after correction (Table S2), and from 43.3 % to between 12.1 % and 18.3 % in the multi-model mean. The static attributes are the component whose contribution the indices resolve directly: removing them raises the index-averaged bias from 14.7 % to 18.6 % (14.8 % to 18.3 % in the multi-model mean) and degrades six of the nine indices, confirming that physiographic predictors carry information that the precipitation series alone does not supply. The indices do not resolve the remaining components: the configuration without NN_2 attains 14.6% (12.1 % in the multi-model mean), nominally matching the baseline, and direct output is best on R10mm and R20mm. R99pTOT is degraded relative to the raw GCMs by every configuration (31.8 % against 44.5–53.2 %), so the far tail remains difficult.

The fractal dimension of the time-mean field separates the configurations that the indices cannot. The mean absolute deviation of the fractal-dimension curve from the reference is 0.021 for the baseline, against 0.029 without spatial attention, 0.035 without static attributes, and 0.030 for the raw GCMs. Removing spatial attention (NN_2) and spatial correlation loss therefore returns the spatial organization of the corrected field to approximately that of the uncorrected model, and removing the static attributes degrades it below the uncorrected model, even though both configurations match or exceed the baseline on several marginal indices. Both components therefore act on the spatial structure of the field rather than on the point-wise distribution. Direct output attains a marginally lower deviation (0.019), as expected of a network that is free to redraw the field without regard to the input series.

In the held-out region, no configuration improves on the raw GCMs in the index view (Table S3): the index-averaged bias is 27.7 % for the raw GCMs against 33.4 % for the baseline in the pooled view, and 28.0 % against 28.8 % in the multi-model mean. A transform learned on 16 HUC-02 regions thus does not yet transfer reliably to an unseen region, and we do not claim that it does. Spatial attention nevertheless governs how severely the model degrades. The two configurations are nearly indistinguishable when the GCMs are pooled (33.4 % against 33.2 %), but in the multi-model mean, which is how bias-adjusted ensembles are normally used, the baseline is clearly better (28.8 % against 33.4 %), and it reproduces the daily spatial variability (Wasserstein distance of the daily spatial standard deviation, 25.8 % against 52.1 %) and the 99.9th percentile (38.2 % against 48.0 %) more closely. Removing attention also permits outright failure: for ACCESS-CM2 the corrected field is almost never dry, with a mean wet-spell length error of 2.2×10^4 % against 56 % for the baseline. Neighbourhood context keeps the learned transform physically plausible where the model has no local training data.

Table S2. Median absolute bias (%) of nine ETCCDI precipitation indices over the held-out period 2001–2014, with the six GCMs pooled as additional grid points. Lower is better; bold marks the best corrected configuration for each index.

Index	Raw GCM	Baseline (full)	No spatial attention	No static attributes	Direct output
SDII (Monthly)	22.8	31.0	**19.2**	25.7	33.5
CDD (Yearly)	22.8	**16.1**	23.4	19.2	20.2

CWD (Yearly)	59.1	**18.7**	25.4	21.1	26.3
Rx1day	75.4	51.5	**35.3**	50.0	70.1
Rx5day	93.8	43.0	**37.4**	48.6	57.0
R10mm	32.6	16.5	20.1	23.8	**15.4**
R20mm	39.0	30.0	36.6	33.9	**25.0**
R95pTOT	73.6	**21.5**	23.5	27.0	23.9
R99pTOT	31.8	49.7	53.2	**44.5**	47.3
Average	**42.0**	**14.7**	**14.6**	**18.6**	**18.2**

Table S3. Median absolute bias (%) of the same indices in the held-out region (HUC-02 region 05, Ohio) over 1990–2014. Lower is better; bold marks the better of the two corrected configurations in each view.

Index	**Pooled across GCMs**			**Multi-model mean**		
	Raw GCM	**Baseline**	**No attention**	**Raw GCM**	**Baseline**	**No attention**
SDII (Monthly)	9.5	**24.5**	24.7	14.0	**19.7**	33.7
CDD (Yearly)	13.6	**14.8**	16.8	8.9	**10.4**	12.9
CWD (Yearly)	32.4	**27.9**	28.9	33.4	**33.0**	34.5
Rx1day	42.0	**41.5**	43.6	48.1	**45.1**	55.8
Rx5day	43.2	**38.9**	43.3	42.3	**28.9**	34.3
R10mm	13.2	44.2	**40.5**	14.5	**23.5**	39.5
R20mm	31.4	68.1	**62.1**	33.8	**56.1**	67.7
R95pTOT	61.3	47.9	**43.1**	67.7	**45.2**	57.2
R99pTOT	19.5	47.9	**47.5**	23.3	34.2	**32.8**
Average	**27.7**	**33.4**	**33.2**	**28.0**	**28.8**	**33.4**

S3 Implications for the δCLIMBA configuration

The ablations support retaining all components, with evidence of differing kinds. The static attributes are supported by the index-based results alone, where their removal costs about four percentage points of index-averaged bias. Spatial attention is supported by the fractal-dimension diagnostic in the held-out period and, more decisively, by the held-out-region test, where its removal degrades the multi-model-mean bias and the daily spatial variability and permits a catastrophic failure for one GCM. The monotone transform is retained on structural grounds: direct output fits the historical climatology slightly better, but it provides no guarantee that the sequencing and the forced response of the driving GCM survive the correction, which is a prerequisite for applying the method to future scenarios. These results also suggest that marginal index biases, the standard currency of the bias-adjustment literature, under-reward design choices that improve the structure of corrected fields: they did not distinguish the baseline from the configuration without spatial attention, yet the two differ clearly in spatial organization and in their behaviour under spatial extrapolation.