Supplementary methods

S1. Native-grid transfer of reference fields

Primary bias calculations were made on the native grid of each evaluated CMIP6 model or reanalysis product. For each target grid, cell boundaries were inferred from neighbouring latitude–longitude grid centres. A source pixel was assigned to a target cell when its centre fell within the inferred boundary of that target cell. For a target cell j, the transferred reference value was calculated as

$$R_j^{ref} = \frac{\sum_{i \in S_j} R_i \cos(\phi_i)}{\sum_{i \in S_j} \cos(\phi_i)},$$

where R_i is the source-pixel value, S_j is the set of valid source pixels assigned to target cell j, and ϕ_i is the latitude of source pixel i. The cosine-latitude weight accounts for the first-order area variation of regular latitude–longitude grids. This procedure is source-cell-centre assignment with cosine-latitude weighting, not exact polygon-overlap remapping.

The two reference variables were prepared with slightly different temporal orders. For MODIS snow cover, daily binary records were first converted to target-cell daily SNC, and the daily target-cell values were then averaged to monthly SNC. For CHE SD, daily source-grid records were first aggregated to monthly mean SD and then transferred to each target grid. This order preserves the monthly snow-cover fraction for SNC while avoiding unnecessary daily remapping of SD.

S2. Spatial support, masks and ancillary fields

A valid-fraction field was retained for each target grid to indicate the proportion of valid source pixels contributing to the transferred reference value. Target cells with insufficient reference support were excluded from product-level statistics. The same support logic was applied to model, reanalysis and reference comparisons.

Continuous or fractional ancillary fields, including elevation, lake fraction, glacier fraction and land-support fraction, were transferred with the same cosine-latitude

weighting used for the snow references. Categorical fields were assigned by cosine-weighted mode. Elevation bands and the west–east sector label were then derived on each product's native grid.

Lake influence was included in the default land-analysis mask. Glacier fraction was not removed from the baseline benchmark; it was retained as a diagnostic field and tested separately. Multi-model display maps were generated only after product-level native-grid statistics had been calculated.

S3. Monthly summaries and ensemble handling

Monthly SNC and SD biases were calculated on the native grid of each evaluated product using the definitions given in the main text. January–March (JFM) means were calculated from monthly biases and used as the main peak-season summary. Annual means were calculated as secondary summaries. Monthly climatological cycles were calculated from the 2002–2014 monthly records.

Spatial summaries were aggregated by elevation band and by west–east sector after native-grid bias calculation. The elevation bands were <3000 m, 3000–4000 m, 4000–5000 m, 5000–6000 m and ≥6000 m. The ≥6000 m band was retained for completeness but interpreted cautiously because of its small area. The west–east split was fixed at 90° E and used only as a broad regional contrast.

For CMIP6, statistics were first calculated for each model. Ensemble medians and interquartile ranges were then derived from model-level values. Ensemble means were not used as the default summary because several models produced very large SD biases. Mean-based summaries were kept only for sensitivity checks.

S4. Cover–depth classification and forcing-signature implementation

The main text defines the cover–depth bias space using JFM SNC and SD biases. The default positive-bias thresholds were

$$T_{SNC} = 10\%, T_{SD} = 0.05\ \text{m}.$$

Products were classified as

SNC-positive only: $\Delta SNC \geq T_{SNC},\ \Delta SD < T_{SD},$

SD-positive only: $\Delta SNC < T_{SNC},\ \Delta SD \geq T_{SD},$

coupled positive: $\Delta SNC \geq T_{SNC},\ \Delta SD \geq T_{SD},$

All remaining products were treated as weak or compensating cases. Threshold sensitivity was tested by varying T_{SNC} and T_{SD} around the default values. The check focused on whether three conclusions remained stable: the dominance of coupled-positive CMIP6 behaviour, the relative positions of the reanalysis products, and the application-oriented screening outcome.

Forcing diagnostics were calculated only for the subset with available TAS, PR and PRSN. CMIP6 forcing biases were calculated relative to ERA5-Land for the same months and spatial support. Snowfall fraction was defined as $SF = \frac{PRSN}{PR}$, where $PR > 0$. Cases with zero or invalid total precipitation were excluded from snowfall-fraction calculations.

A compact fingerprint matrix was built from JFM SNC bias, SD bias, TAS bias, PR bias, PRSN bias and snowfall-fraction bias. Each column was standardized across the forcing-signature subset for display. These standardized values were used only to compare relative contrasts among diagnostic models and were not interpreted as absolute bias magnitudes. Monthly cold-bias sensitivity was assessed using across-model correlations between TAS bias and SNC bias, and between TAS bias and SD bias, for each calendar month. The correlation analysis was used as a diagnostic check on the seasonal consistency of the temperature–snow relationship, not as causal attribution.

Supplementary tables

Table S1. CMIP6 model availability and forcing-signature subset.

Model	SNC	SD	TAS	PR	PRSN	SNW
BCC-CSM2-MR	Yes	Yes	No	No	No	Yes
BCC-ESM1	Yes	Yes	No	No	No	Yes
CESM2	Yes	Yes	No	No	Yes	Yes
CESM2-FV2	Yes	Yes	No	No	Yes	Yes
CESM2-WACCM	Yes	Yes	No	No	No	Yes
CIESM	Yes	Yes	Yes	No	No	Yes
CanESM5	Yes	Yes	Yes	Yes	Yes	Yes
EC-Earth3	Yes	Yes	Yes	No	No	Yes
EC-Earth3-Veg	Yes	Yes	Yes	No	No	Yes
EC-Earth3-Veg-LR	Yes	Yes	Yes	No	No	Yes
GFDL-CM4	Yes	Yes	Yes	No	No	Yes
GISS-E2-1-G	Yes	Yes	Yes	Yes	Yes	Yes
GISS-E2-1-G-CC	Yes	Yes	Yes	Yes	Yes	Yes
GISS-E2-1-H	Yes	Yes	Yes	Yes	Yes	Yes
IPSL-CM6A-LR	Yes	Yes	Yes	No	No	Yes
MIROC6	Yes	Yes	Yes	Yes	Yes	Yes
MRI-ESM2-0	Yes	Yes	Yes	Yes	Yes	Yes
NorESM2-LM	Yes	Yes	No	No	No	Yes
NorESM2-MM	Yes	Yes	Yes	Yes	Yes	Yes
SAM0-UNICON	Yes	Yes	No	No	No	Yes
TaiESM1	Yes	Yes	No	No	No	Yes

Note. Yes indicates complete processed monthly data for 2002-2014 in the current benchmark files.

Supplementary figures

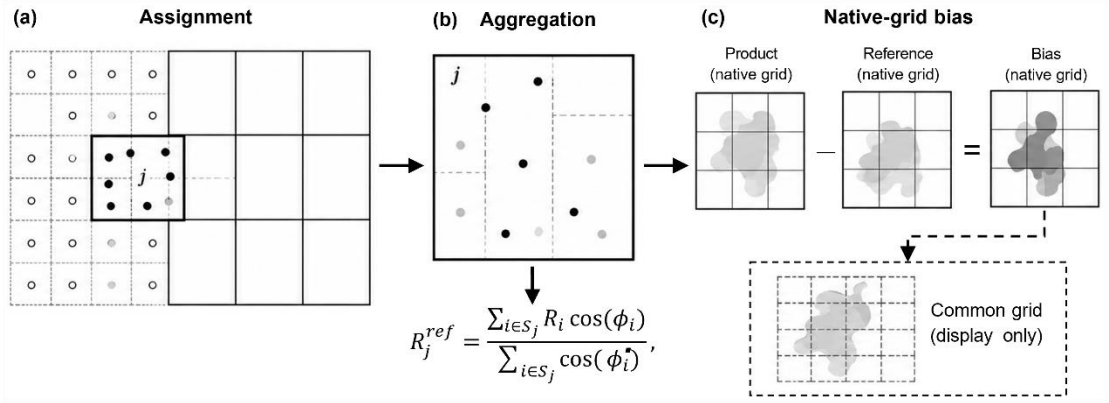

Fig. S1. Native-grid reference aggregation and bias calculation. High-resolution reference data are first assigned to each product native-grid cell and aggregated using latitude-area weights. Biases are then calculated on the product native grid as product minus aggregated reference. A common grid is used only for displaying spatial patterns.