



1 **A native-grid cover–depth benchmark reveals coupled snow-state biases in**
2 **CMIP6 over the Tibetan Plateau**

3

4 Yao Xiao¹, Nan Zhou², Guojie Hu^{1,3*}, Ren Li¹, Guangyue Liu¹, Tonghua Wu¹,
5 Xiaodong Wu^{1,4}, Defu Zou¹, Erji Du¹, Yonghua Zhao¹, & Lin Zhao²

6 ¹ Cryosphere Research Station on the Qinghai-Tibet Plateau, State Key Laboratory of
7 Cryospheric Science and Frozen Soil Engineering, Northwest Institute of Eco-
8 Environment and Resources, Chinese Academy of Sciences, Lanzhou 730030, China

9 ² School of Geographical Sciences, Nanjing University of Information Science &
10 Technology, Nanjing 210044, China

11 ³ University of Chinese Academy of Sciences, Beijing 100049, China

12 ⁴ Qinghai Provincial Key Laboratory of Restoration Ecology for Cold Region,
13 Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810008,
14 China

15

16 *Correspondence to: G. Hu, huguojie123@lzb.ac.cn

17

18

19



20 **Abstract:**

21 Snow over the Tibetan Plateau is often widespread in winter but remains shallow over
22 much of the region. This makes snow-covered area and snow depth complementary, not
23 interchangeable, measures of model skill. We develop a monthly native-grid cover–depth
24 benchmark for 2002–2014, evaluating CMIP6 historical simulations and major
25 reanalysis products against MODIS-derived snow-covered area percentage (SNC) and
26 CHE snow depth (SD). The reference fields are transferred to each product’s native grid
27 before bias calculation, so that snow-state errors are diagnosed on the spatial support of
28 the evaluated product. CMIP6 shows a dominant coupled snow-state excess over the
29 Plateau. Median SNC biases reach 37.7%, 30.2% and 24.1% in January, February and
30 March, respectively, while SD bias is also positive but strongly skewed by several large
31 outliers. In the January–March cover–depth bias space, 19 of 21 CMIP6 models
32 overestimate both SNC and SD. The largest coupled biases occur over the western and
33 high-elevation Plateau, indicating a structured cold-season error rather than a domain-
34 wide offset. Reanalysis products show different forms of mismatch: ERA5-Land
35 overestimates both SNC and SD, whereas MERRA-2 combines underestimated SNC
36 with overestimated SD. Forcing diagnostics suggest that cold-biased persistence,
37 snowfall input, snowfall partitioning and snow-process errors all contribute to the
38 model spread. A cover–depth benchmark is therefore needed before Tibetan Plateau
39 snow products are used in cryospheric, hydrological or land–atmosphere studies.

40 **Short summary**

41 Snow on the Tibetan Plateau is often widespread in winter but usually shallow. We
42 developed a benchmark that evaluates snow-covered area and snow depth together on
43 each model’s native grid. Most CMIP6 models overestimate both variables in winter,
44 especially over the western and high-elevation Plateau. The benchmark helps users
45 choose snow products for cryospheric, hydrological and climate studies.

46



47 **1 Introduction**

48 Snow on the Tibetan Plateau (TP) is often shallow, patchy and short-lived, but its
49 influence on the land surface is not small. A thin snow layer can change surface albedo,
50 modify radiative and turbulent fluxes, insulate or cool the ground, and shift the timing
51 of meltwater release (Miao et al., 2022; Wang et al., 2020). These effects feed into
52 runoff generation, ecosystem activity, frozen-ground conditions and land–atmosphere
53 exchange across High Mountain Asia (Barnett et al., 2005; Brown and Mote, 2009; You
54 et al., 2020). For climate and cryosphere applications, TP snow is therefore not just a
55 seasonal background variable. Its timing, extent and depth help determine how energy
56 and water are exchanged between the atmosphere, the surface and the subsurface.

57 This snow system is not easy to benchmark. Much of the Plateau does not develop a
58 continuous, deep seasonal snowpack. Snow occurrence is intermittent, strongly
59 seasonal and spatially fragmented (Che et al., 2016; Hao et al., 2022; Wang et al., 2024;
60 You et al., 2020). During winter and early spring, snow-covered area can expand widely,
61 while snow accumulation remains limited over large parts of the Plateau. Snow-covered
62 area percentage (SNC) and snow depth (SD) may therefore give different assessments
63 of the same model (Jiang et al., 2025; Orsolini et al., 2019). A product can place snow
64 over too large an area without producing excessive depth. Another can generate too
65 much SD over a narrower snow-covered domain. Annual means further weaken the
66 signal because the main snow state is concentrated in a short cold-season window.

67 Existing evaluations have provided valuable information, but they do not yet give a
68 clear answer to a practical question: which products can be trusted for TP snow-state
69 applications? Snow-cover products, snow-depth datasets, reanalysis fields and climate-
70 model outputs are often assessed separately, against different references and at different
71 spatial supports (Chen et al., 2018; Kouki et al., 2022; Orsolini et al., 2019; Yan et al.,
72 2022). This separation is manageable for single-variable product validation, but it is
73 less useful for applications that depend on both snow extent and snow amount.
74 Hydrological modelling, snow–albedo feedback analysis, frozen-ground studies and



75 land–atmosphere coupling all require a more explicit view of the joint snow state.

76 The grid on which the benchmark is performed also matters. CMIP6 models and
77 reanalysis products differ greatly in spatial resolution, while satellite and gridded
78 reference datasets have much finer support (Lei et al., 2023; Miao et al., 2022).
79 Remapping all products to a common grid makes figures easier to compare, but it can
80 mix model resolution, reference resolution and interpolation choices into the bias
81 estimate itself (Lei et al., 2023). We therefore use a native-grid strategy: the reference
82 fields are transferred to each evaluated product’s grid, and product-level biases are
83 calculated on that native grid. Common grids are used only for display and synthesis.

84 A TP snow benchmark should then ask more than whether a product has “too much
85 snow”. Excess snow can mean an overly large snow-covered area, excessive
86 accumulation, delayed melt under cold-biased conditions, too much snowfall, biased
87 rain–snow partitioning, or errors in snow-cover conversion, density, compaction and
88 melt processes (Essery et al., 2013; Kouki et al., 2022; Raleigh et al., 2015). These
89 sources cannot be separated by a single mean snow-bias score. In this study, MODIS-
90 derived monthly SNC and CHE monthly SD are used as paired references. They do not
91 define one combined “true snow field”; instead, they constrain two related but non-
92 equivalent dimensions of the monthly snow state.

93 Here we develop a monthly, native-grid, cover–depth benchmark for TP snow during
94 2002–2014. CMIP6 historical simulations and major reanalysis products are evaluated
95 against MODIS-derived SNC and CHE SD (Che et al., 2015, 2016; Hao et al., 2022).
96 We diagnose biases through seasonal cycles, spatial patterns, elevation bands, west–
97 east contrasts and a cover–depth bias space. Temperature, precipitation and snowfall
98 are used only as forcing signatures, to help distinguish cold-biased persistence, snowfall
99 input, snowfall partitioning and residual snow-process errors. The benchmark is
100 designed to answer three questions: what paired monthly snow state is represented by
101 the references; how CMIP6 and reanalysis products depart from that state; and which
102 products are suitable, questionable or correction-dependent for TP snow-state



103 applications.

104 **2 Data**

105 **2.1 CMIP6 simulations**

106 We evaluated monthly CMIP6 historical simulations over the TP for 2002–2014. The
107 period was chosen to match the overlap between MODIS-era snow-cover observations
108 and the CMIP6 historical simulations, which end in 2014. The benchmark was built
109 around two primary snow-state variables: SNC and SD. Four additional CMIP6
110 variables were used to support interpretation: near-surface air temperature (TAS), total
111 precipitation (PR), snowfall (PRSN), and surface snow amount (SNW). TAS, PR and
112 PRSN were used for forcing-signature diagnosis, whereas SNW was retained only as
113 an auxiliary snow-mass variable and was not treated as an independent benchmark
114 target. The CMIP6 models were divided into two working subsets. The snow-state
115 benchmark subset required monthly SNC and SD and was used for the main cover–
116 depth evaluation. The forcing-signature subset additionally required TAS, PR and
117 PRSN and was used to examine whether snow-state biases were associated with cold-
118 biased persistence, snowfall input or snowfall partitioning. This subset was used only
119 for diagnostic interpretation, not for formal attribution. Model names, available
120 variables and subset membership are listed in Supplementary Table S1.

121 **2.2 Reanalysis products**

122 Three reanalysis products were included as comparison datasets: ERA5-Land,
123 MERRA-2 and JRA-55 (Gelaro et al., 2017; Kobayashi et al., 2015; Muñoz-Sabater et
124 al., 2021). They were not used as snow references. Instead, they provide an
125 observationally constrained comparison group for assessing whether snow-state biases
126 found in CMIP6 are also present in widely used reanalysis products. ERA5-Land and
127 MERRA-2 were evaluated for both SNC and SD where comparable monthly fields
128 were available. JRA-55 was used mainly for SD and forcing-signature comparison
129 because no comparable monthly SNC field was available in the processed output. This
130 asymmetric variable availability was retained explicitly in the reanalysis benchmark



131 and in the cover–depth bias-space analysis. Dataset characteristics are summarized in
132 Table 1.

133 **Table 1.** Datasets used in the native-grid Tibetan Plateau snow benchmark.

Dataset group	Dataset	Variables	Native resolution	Sources
Climate models	CMIP6 historical simulations	SNC, SD; TAS, PR, PRSN	Model dependent	(Eyring et al., 2016)
Reanalysis products	ERA5-Land	SNC, SD, TAS, PR, PRSN	0.1°	(Muñoz-Sabater et al., 2021)
Reanalysis products	MERRA-2	SNC, SD, TAS, PR, PRSN	0.5° x 0.625°	(Gelaro et al., 2017)
Reanalysis products	JRA-55	SD, TAS, PR, PRSN	1.25°	(Kobayashi et al., 2015)
Snow-cover reference	China MODIS daily 500 m snow-cover product	Binary snow cover aggregated to monthly SNC	500 m	(Hao et al., 2022)
Snow-depth reference	CHE_SND_CHINA	Daily SD aggregated to monthly SD	0.25°	(Che et al., 2015)
Topography	TP boundary, SRTM DEM	Elevation, west-east sector	1km	(Farr et al., 2007)

134 Note. The benchmark period is 2002-2014 unless otherwise stated. SNC and SD are the primary snow-
135 state variables. TAS, PR and PRSN are used for forcing-signature diagnosis. Reanalysis products are
136 comparison datasets rather than snow references.

137 **2.3 Snow-cover and snow-depth references**

138 The cover and depth dimensions of the benchmark were constrained by separate
139 reference datasets. Snow cover was referenced to the China MODIS daily 500 m snow-
140 cover product (Hao et al., 2022). Daily binary snow-cover records were converted to
141 monthly SNC, providing an observational counterpart to model and reanalysis SNC.
142 Snow depth was referenced to CHE_SND_CHINA, a gridded daily snow-depth dataset
143 for China derived from passive-microwave brightness-temperature retrievals (Che et
144 al., 2015). Daily records were aggregated to monthly mean SD and standardized to
145 metres before comparison with model and reanalysis SD. In this paired-reference
146 design, MODIS-derived SNC constrains the monthly extent of snow cover, whereas
147 CHE SD constrains snow-depth magnitude.



2.4 Topography and regional partition

Elevation was used as the primary stratification variable because TP snow climatology and model bias are strongly height dependent. Elevation bands were defined as <3000 m, 3000–4000 m, 4000–5000 m, 5000–6000 m and ≥ 6000 m. The highest band was retained for completeness but interpreted cautiously because of limited area support. Elevation stratification was based on the Shuttle Radar Topography Mission DEM (Farr et al., 2007), resampled or aggregated to each evaluated product grid. A west–east partition at 90° E was used to describe broad regional contrast rather than formal climate-zone separation.

3 Methods

3.1 Native-grid cover–depth benchmark

The benchmark was designed to evaluate monthly TP snow states on the spatial support of each product. We therefore did not interpolate all products to a common analysis grid before calculating bias. Instead, the MODIS-derived SNC and CHE SD reference fields were transferred to the native grid of each CMIP6 model or reanalysis product. Product-level biases were then calculated on those native grids. Common grids were used only for multi-model map display. The overall workflow of paired references, native-grid bias calculation, cover–depth diagnosis and product screening is summarized in Fig. 1.

166

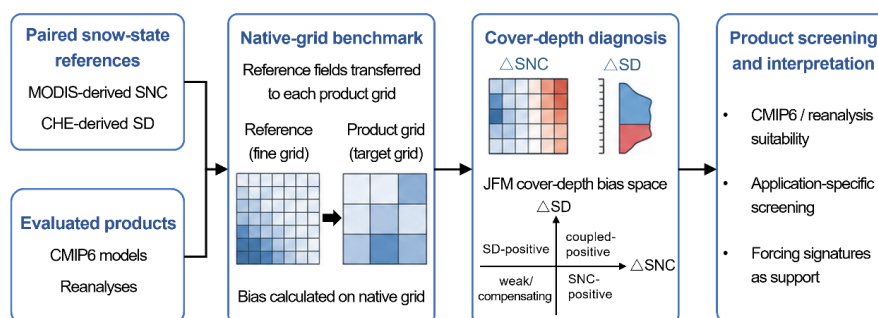




Figure 1. Workflow of the native-grid cover–depth benchmark. MODIS-derived SNC and CHE-derived SD are used as paired reference constraints. Reference fields are transferred to each CMIP6 or reanalysis product grid before bias calculation. Native-grid SNC and SD biases are then interpreted through JFM cover–depth bias space and used for product screening and forcing-signature diagnosis.

Daily MODIS snow-cover records were converted to monthly SNC, and daily CHE snow-depth records were aggregated to monthly mean SD. The reference fields were transferred to each target grid using source-cell-center assignment with cosine-latitude weighting. This approach accounts for the first-order latitude dependence of grid-cell area, but it is not exact polygon-overlap remapping. The implementation details, valid-cell screening and display-grid procedure are given in Supplementary Methods S1–S2 and Fig. S1.

3.2 Monthly snow-state bias metrics

The benchmark used two primary monthly bias metrics:

$$\Delta SNC = SNC - SNC^{ref},$$

$$\Delta SD = SD - SD^{ref}.$$

Here, SNC^{ref} is the MODIS-derived monthly SNC reference transferred to the evaluated product’s native grid, and SD^{ref} is the corresponding CHE monthly SD reference. Bias is defined as product minus reference throughout the study.

Biases were summarized as monthly climatological cycles, annual means, January–March (JFM) means, spatial maps, elevation-band statistics, west–east contrasts and model-level rankings. JFM was used as the main peak-season window because it captures the dominant snow-covered and snow-accumulation state over the Plateau. Annual means were retained only as a secondary summary, mainly to show how strongly the cold-season signal is diluted when snow-poor months are included.

The benchmark is intentionally monthly. It does not assess daily snow onset, melt timing, snow-spell persistence or event-scale continuity. Those diagnostics require daily information and a different temporal framework, which are not consistently available across the full CMIP6–reanalysis product pool.



195 **3.3 Stratified summaries and bias organization**

196 Snow-state biases were first calculated at the native-grid cell level and then summarized
197 by elevation band and by west–east sector. Elevation was used to diagnose height-
198 dependent bias amplification. The 90° E split was used as a broad regional contrast
199 rather than a formal climate-zone classification. For CMIP6, model-level statistics were
200 calculated before ensemble aggregation. Ensemble medians and interquartile ranges
201 were used as the main summaries, especially for SD, because several models produce
202 very large positive SD biases. This avoids allowing a small number of outliers to
203 dominate the ensemble interpretation.

204 We then examined whether SNC and SD biases were organized in similar places and
205 seasons. Spatial co-organization was described using the frequency of positive JFM
206 SNC and SD bias across models and by identifying coupled positive areas. Temporal
207 organization was assessed using model–month bias matrices, elevation–month
208 summaries, west–east seasonal cycles and JFM interannual variability. These
209 diagnostics are not separate benchmark targets. They link the single-variable SNC and
210 SD results to the coupled snow-state interpretation.

211 **3.4 Cover–depth bias space and model screening**

212 To separate different forms of snow-state error, JFM SNC and SD biases were evaluated
213 in a common cover–depth bias space. Two positive-bias thresholds were used: 10% for
214 SNC and 0.05 m for SD. Products were classified as SNC-positive only, SD-positive
215 only, coupled positive, or weak/compensating according to which thresholds were
216 exceeded. These classes are diagnostic categories, not universal model ranks. Their
217 purpose is to distinguish products that mainly overestimate snow-covered area,
218 products that mainly overestimate snow depth, and products that overestimate both
219 dimensions of the monthly snow state. The same bias-space framework was used for
220 CMIP6 models and reanalysis products, allowing reanalyses to be positioned relative
221 to the CMIP6 ensemble. The classification also provides the basis for the application-
222 oriented model-screening summary.



223 **3.5 Forcing signatures and robustness checks**

224 Near-surface air temperature, total precipitation and snowfall were used to interpret
225 snow-state biases in the forcing-signature subset. For CMIP6 models, forcing
226 differences were calculated relative to ERA5-Land over the same monthly period.
227 Snowfall fraction was calculated from snowfall and total precipitation where
228 precipitation was positive.

229 The analysis was kept deliberately conservative. A cold-biased persistence signature
230 was inferred when positive SNC or SD bias co-occurred with negative temperature bias,
231 especially in late winter and spring. A snowfall-input signature was inferred when
232 positive SD bias was associated with high snowfall bias. A snowfall-partitioning
233 signature was inferred when snow bias was more consistent with snowfall fraction than
234 with total precipitation. Cases without a clear forcing signature were treated as residual
235 snow-process errors, potentially linked to snow-cover conversion, density, compaction,
236 melt or other snow-scheme behavior. These diagnostics support interpretation and
237 model screening; they are not used as a formal attribution experiment. Robustness
238 checks focused on ensemble-summary sensitivity, the influence of the largest SD outlier
239 and threshold sensitivity of the cover–depth classification.

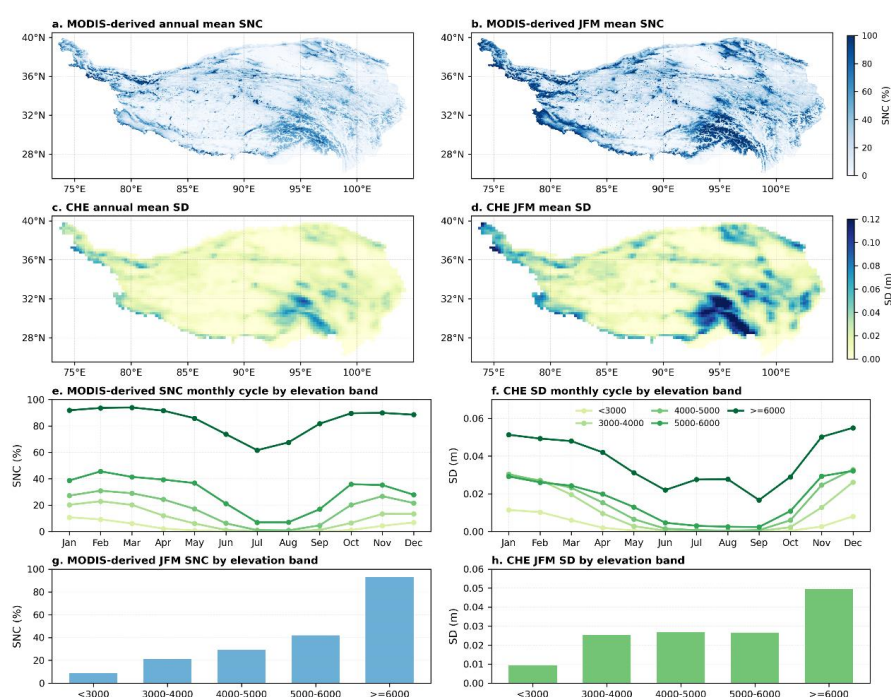
240 **4 Results**

241 **4.1 A seasonally extensive but shallow reference snow state**

242 The paired reference fields show a simple but important feature of TP snow: snow cover
243 expands widely in winter, whereas snow accumulation remains limited over most of the
244 Plateau. Annual mean SNC is modest across much of the domain, but the January–
245 March (JFM) climatology reveals a broader snow-covered area along mountain margins
246 and high-elevation terrain (Fig. 2a, b). This contrast explains why annual means are a
247 weak summary for TP snow. The main snow state is concentrated in a short cold-season
248 window. Elevation sharpens this pattern. JFM SNC increases from 8.7% below 3000 m
249 to 21.1% at 3000–4000 m, 29.0% at 4000–5000 m and 41.9% at 5000–6000 m (Fig. 2e,
250 g). The ≥ 6000 m bin reaches 93.2%, but its small area support makes it less suitable for



251 broad interpretation. CHE SD has a much flatter structure. JFM SD is about 0.009 m
252 below 3000 m and remains only 0.025–0.027 m across 3000–6000 m, rising to about
253 0.050 m in the highest band (Fig. 2f, h). Broad winter snow cover over the TP does not
254 imply deep snow accumulation. This mismatch means that SNC and SD must be
255 assessed together; annual means or single-variable snow-bias metrics would miss an
256 important part of the reference snow state.



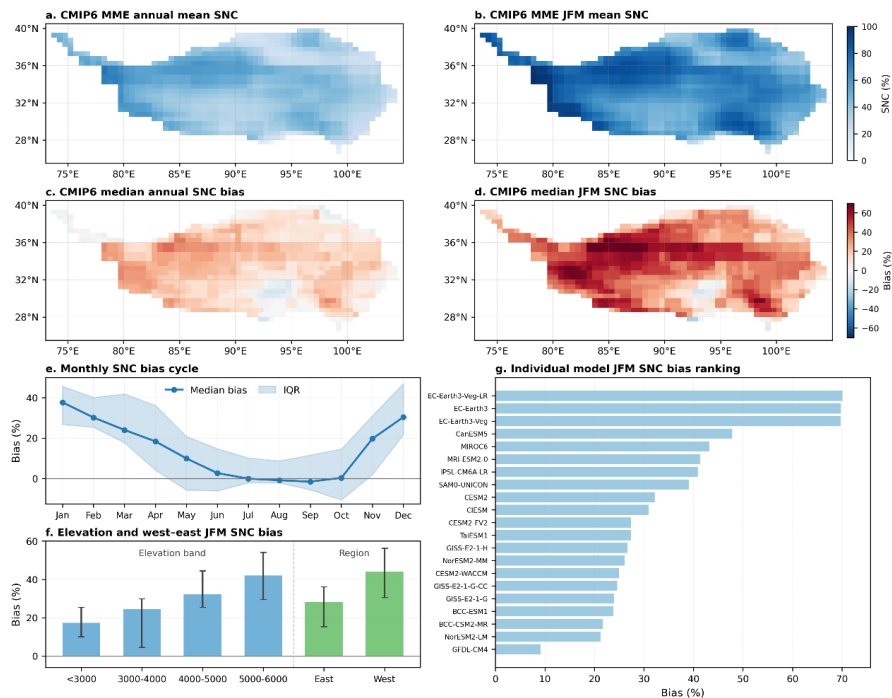
257
258 **Figure 2.** Reference snow state over the Tibetan Plateau. (a, b) MODIS-derived annual-mean and JFM-
259 mean SNC; (c, d) CHE annual-mean and JFM-mean SD; (e, f) monthly climatologies by elevation band;
260 and (g, h) JFM elevation dependence. Maps are for display; statistics are calculated on the reference grids.

261 4.2 CMIP6 shows a cold-season excess in both SNC and SD

262 CMIP6 reproduces the seasonal expansion of TP snow, but the snow-covered area is
263 too large during the core snow season. The multi-model SNC field shows widespread
264 snow-covered conditions over the western and interior Plateau in JFM (Fig. 3a, b).
265 Relative to the MODIS-derived reference, the median SNC bias is positive over most



266 snow-affected regions (Fig. 3c, d). The bias is strongly seasonal. Median TP-mean SNC
267 bias reaches 37.7% in January, 30.2% in February and 24.1% in March, remains
268 positive into April, and then weakens to near-zero or slightly negative values in July–
269 September (Fig. 3e). This is not a year-round offset. It is a cold-season error, consistent
270 with excessive snow-covered area, overly persistent snow cover, or delayed snow
271 retreat. The spatial structure is also clear. JFM SNC bias increases with elevation, from
272 17.3% below 3000 m to 24.5% at 3000–4000 m, 32.3% at 4000–5000 m and 42.0% at
273 5000–6000 m (Fig. 3f). The western Plateau has a larger median bias than the eastern
274 Plateau, 44.2% versus 28.2%. Individual models differ widely. GFDL-CM4 is closest
275 to the MODIS-derived SNC reference, with a TP-mean JFM bias of 9.1%, whereas EC-
276 Earth3, EC-Earth3-Veg and EC-Earth3-Veg-LR exceed 69% (Fig. 3g).



277



278 **Figure 3.** CMIP6 bias in monthly snow-covered area percentage. (a, b) CMIP6 ensemble annual-mean
279 and JFM-mean SNC; (c, d) corresponding median bias relative to the MODIS-derived reference; (e)
280 monthly bias cycle; (f) JFM bias by elevation band and west–east sector; (g) model ranking by TP-mean
281 JFM bias. Model-level statistics are calculated on native grids, with maps displayed on a common grid.

282 SD bias is also positive, but its ensemble structure is different. The ensemble median
283 SD remains shallow in an absolute sense, yet it is still higher than the CHE reference
284 over many snow-affected regions, especially in JFM and over high-elevation or
285 mountain-margin areas (Fig. 4a–d). The monthly SD bias peaks later than the SNC bias:
286 0.076 m in January, 0.118 m in February and 0.150 m in March, before dropping rapidly
287 toward near-zero in July–August (Fig. 4e).

288 The elevation and regional patterns resemble SNC, although the distribution is more
289 skewed. Median JFM SD bias rises from 0.048 m below 3000 m to 0.126–0.133 m
290 across 4000–6000 m, and the western Plateau again exceeds the eastern Plateau, 0.154
291 m versus 0.082 m (Fig. 4f). Most models show modest to moderate positive SD bias,
292 but a few models dominate the upper tail. IPSL-CM6A-LR is the most extreme case,
293 with a JFM SD bias of 1.832 m. CanESM5 and BCC-ESM1 also show large positive
294 SD biases of 0.502 m and 0.420 m (Fig. 4g). This skewed SD distribution justifies using
295 medians, interquartile ranges and model-level screening rather than relying on
296 ensemble means.

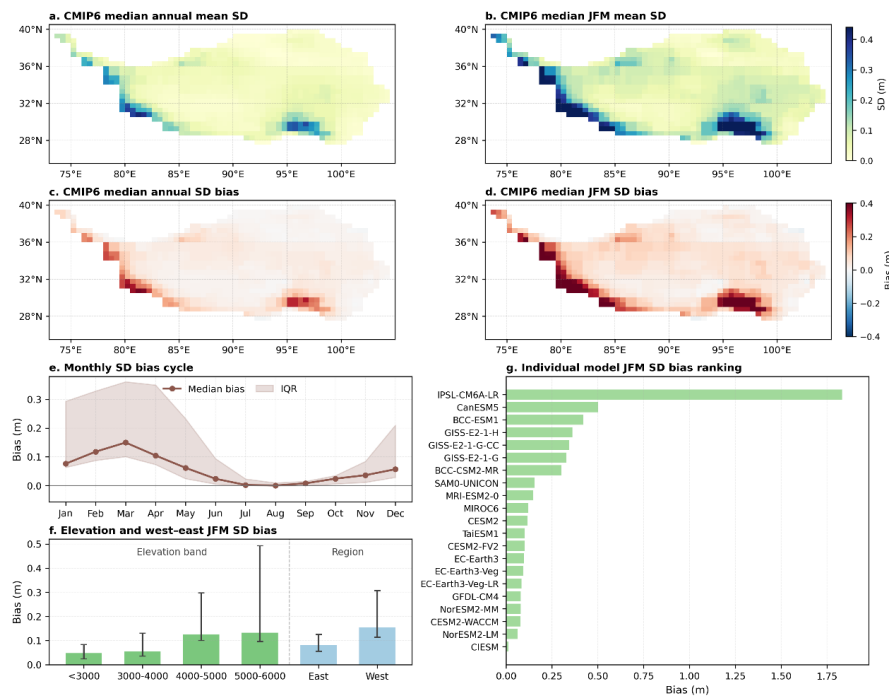
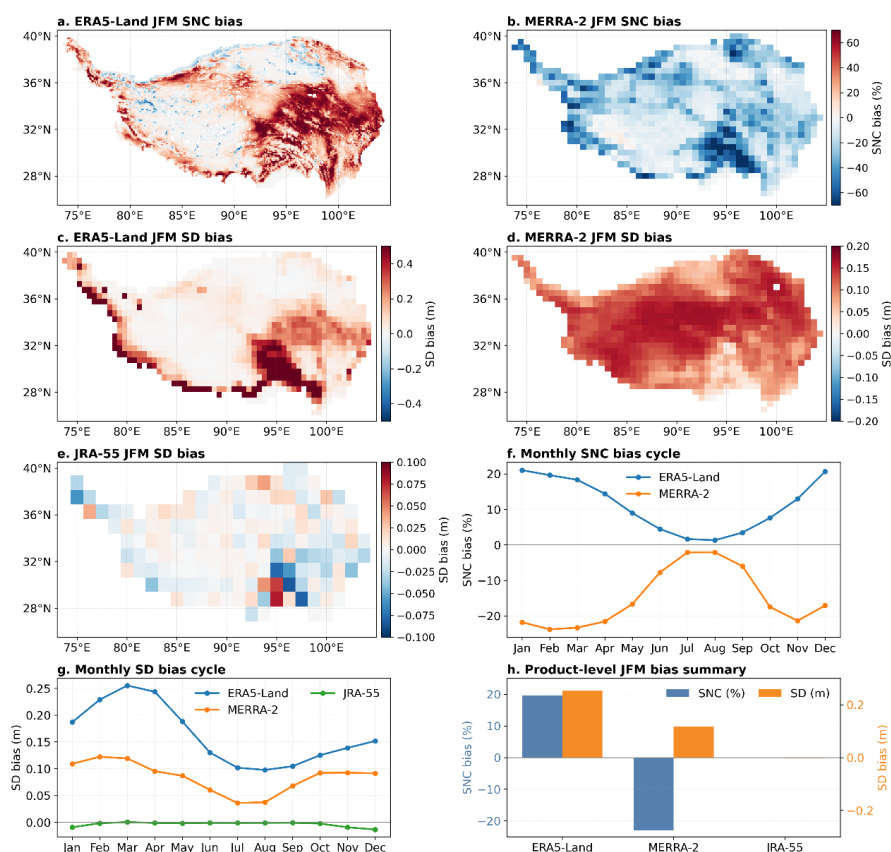


Figure 4. CMIP6 bias in monthly snow depth. (a, b) CMIP6 ensemble annual-mean and JFM-mean SD; (c, d) corresponding median bias relative to the CHE reference; (e) monthly bias cycle; (f) JFM bias by elevation band and west-east sector; and (g) model ranking by TP-mean JFM bias. Medians and interquartile ranges are used because SD biases are strongly skewed across models.

4.3 Reanalysis products show variable-specific cover–depth mismatch

The reanalysis products do not provide a uniform correction to CMIP6 snow-state errors. ERA5-Land and MERRA-2 show opposite SNC behaviour in JFM. ERA5-Land overestimates snow-covered area over large parts of the Plateau, with a TP-mean JFM SNC bias of 19.7%. MERRA-2 underestimates SNC over much of the domain, with a TP-mean bias of −23.0% (Fig. 5a, b, h). Their monthly cycles reinforce this contrast: ERA5-Land remains mostly positive, whereas MERRA-2 is persistently negative outside the summer minimum period (Fig. 5f).



310

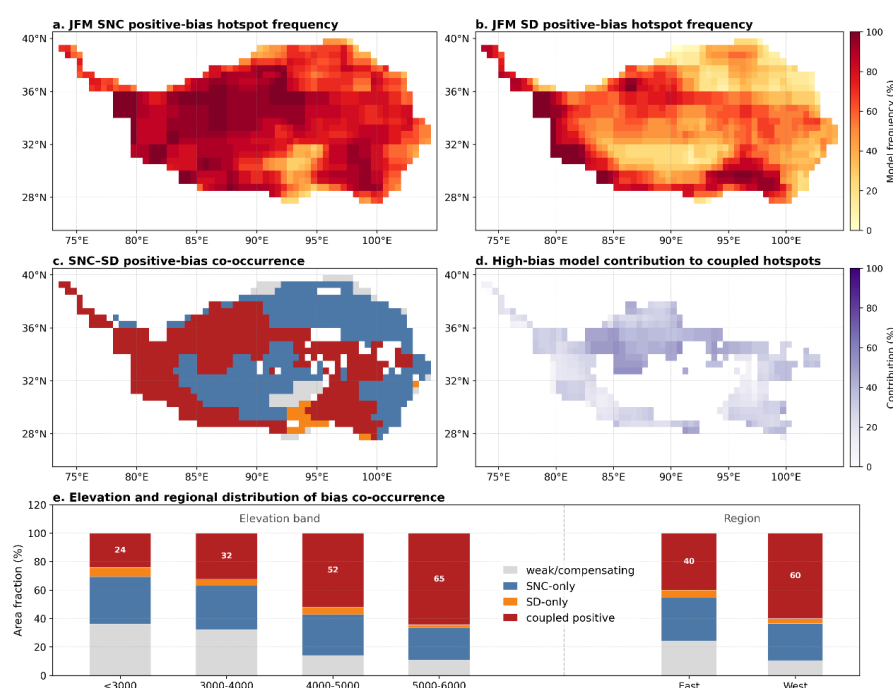
311 **Figure 5.** Reanalysis snow-state biases over the Tibetan Plateau. (a, b) JFM SNC bias for ERA5-Land
312 and MERRA-2 relative to the MODIS-derived reference; (c–e) JFM SD bias for ERA5-Land, MERRA-
313 2 and JRA-55 relative to the CHE reference; (f, g) monthly bias cycles for SNC and SD; and (h) product-
314 level JFM biases. JRA-55 is shown for SD only because no comparable monthly SNC field was available.

315 SD tells a different story. ERA5-Land and MERRA-2 both overestimate JFM SD, with
316 TP-mean biases of 0.253 m and 0.117 m, respectively (Fig. 5c, d, h). ERA5-Land has
317 localized positive SD bias along high-elevation margins and parts of the southern
318 Plateau. MERRA-2 shows a broader positive SD bias across the interior Plateau. JRA-
319 55 is close to neutral for SD, with a JFM bias of -0.004 m and an absolute bias of 0.014
320 m, but it cannot be placed in the full cover–depth space because no comparable monthly
321 SNC field is available in the current processed output (Fig. 5e, g, h). The reanalysis
322 comparison shows that snow-state skill is variable dependent. ERA5-Land is close to a



323 coupled-positive case, MERRA-2 combines negative SNC bias with positive SD bias,
324 and JRA-55 provides only an SD-side constraint in this workflow.

325 4.4 Coupled positive biases are concentrated over the western and high-elevation 326 Plateau



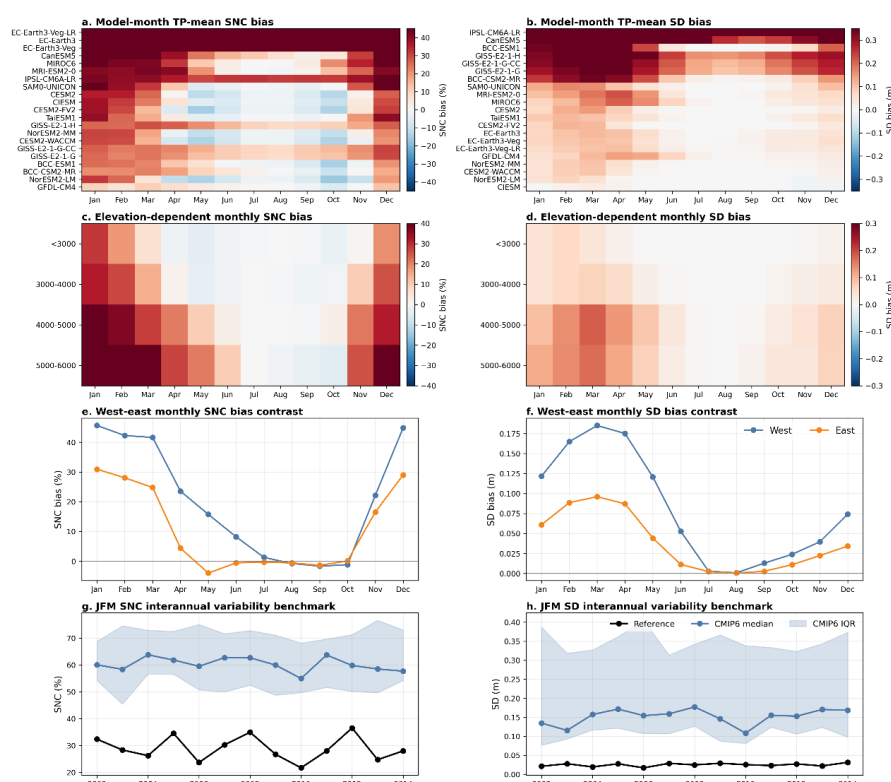
327

328 **Figure 6.** Spatial co-organization of CMIP6 SNC and SD biases. (a, b) Model frequency of positive JFM
329 SNC and SD bias; (c) SNC–SD co-occurrence classes; (d) coupled positive hotspots associated with
330 high-bias models; and (e) class fractions by elevation band and west–east sector. Maps are displayed on
331 a common grid after native-grid model calculations.

332 SNC and SD biases are related, but they are not spatially identical. Positive JFM SNC
333 bias is widespread across the CMIP6 ensemble: the mean model frequency of positive
334 SNC bias is 78.8%, with a median of 81.0% across supported display-grid cells (Fig.
335 6a). Positive SD bias is less ubiquitous but still common, with a mean frequency of
336 52.9% and a median of 52.4% (Fig. 6b). Coupled positive hotspots, where both
337 variables are positive in the model-frequency sense, occupy a substantial part of the
338 Plateau, with a mean frequency of 48.9% (Fig. 6c). The coupled pattern strengthens



sharply with elevation. Coupled positive bias accounts for 32.4% of the area in the 3000–4000 m band, 52.3% at 4000–5000 m and 64.6% at 5000–6000 m (Fig. 6e). The western Plateau also has a larger coupled-positive fraction than the eastern Plateau, 60.2% versus 40.0%. Cover-only positive bias remains common at lower and middle elevations, but the strongest errors occur over the western high Plateau, where positive SNC and SD biases co-occur.



345

Figure 7. Temporal structure of CMIP6 SNC and SD biases. (a, b) Model-month TP-mean biases in SNC and SD; (c, d) elevation-dependent monthly median biases; (e, f) west–east monthly bias cycles; and (g, h) JFM interannual variability in the references and CMIP6 ensemble.

The temporal diagnostics point to the same structure. SNC bias is concentrated in winter and early spring, with many models retaining positive bias into April and again in November–December (Fig. 7a). SD bias peaks later, with the largest ensemble-median biases in February–March (Fig. 7b). High-elevation bands carry the largest cold-season



353 SNC and SD biases, while warm-season biases are small or near zero (Fig. 7c, d). West–
354 east monthly cycles show stronger cold-season bias in the western Plateau for both
355 variables (Fig. 7e, f). Interannual variability does not remove this baseline error. The
356 CMIP6 median JFM SNC time series remains well above the MODIS-derived reference
357 throughout 2002–2014, and the model interquartile range generally does not collapse
358 onto the reference (Fig. 7g). The SD time series shows the same tendency relative to
359 CHE SD, although year-to-year variations are visible (Fig. 7h). CMIP6 therefore
360 reproduces some temporal variability, but around a biased snow-state baseline.

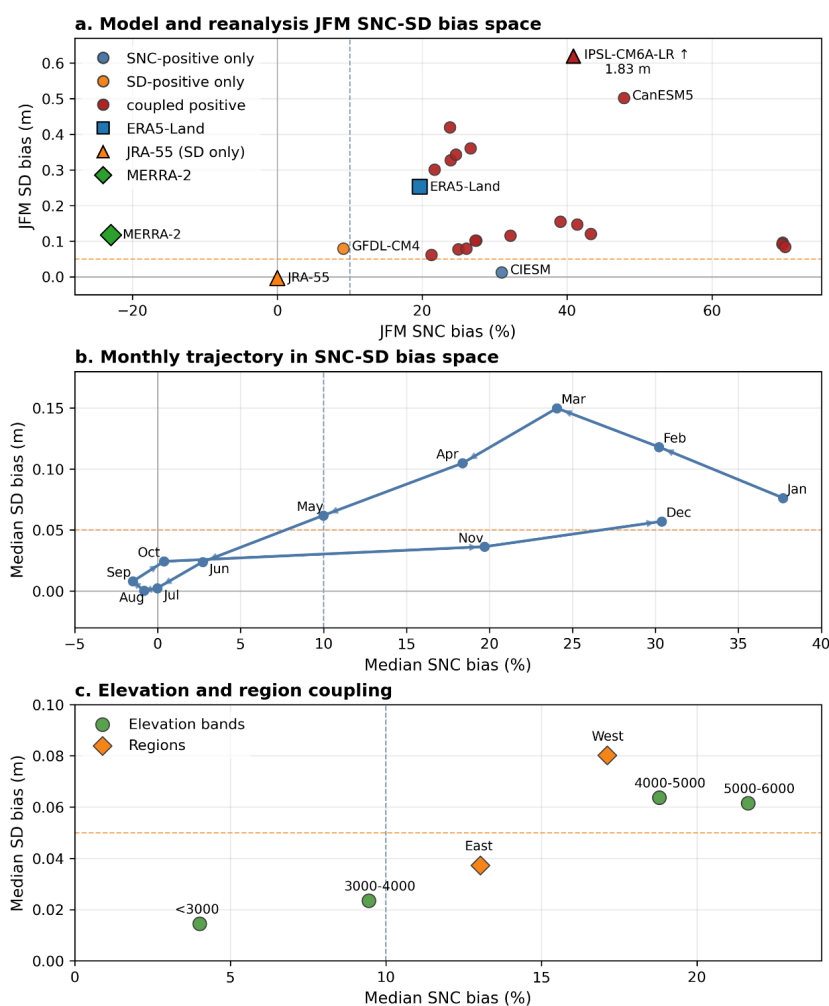
361 **4.5 Cover–depth bias space links benchmark results to model screening**

362 The JFM cover–depth bias space gives the clearest summary of the CMIP6 snow-state
363 problem. Using 10% for SNC and 0.05 m for SD as positive-bias thresholds, 19 of 21
364 CMIP6 models fall into the coupled-positive class (Fig. 8a). CIESM is classified as
365 SNC-positive only, and GFDL-CM4 as SD-positive only. No model falls into the weak
366 or compensating class under the default thresholds. The dominant CMIP6 error is
367 therefore a coupled snow-state excess, not an isolated snow-cover or snow-depth bias.
368 The monthly trajectory shows how this coupled error develops through the year (Fig.
369 8b). The CMIP6 median remains close to zero in July–August, shifts toward positive
370 SNC and SD bias in autumn, and enters the coupled-positive quadrant during winter
371 and early spring. SD bias reaches its maximum in March, while SNC bias remains
372 strongly positive. This seasonal path is consistent with a sequence in which snow-
373 covered area becomes too extensive first, followed by accumulation of excessive SD
374 later in the season.

375 Elevation and regional bias-space summaries reinforce the hotspot interpretation. The
376 4000–5000 m and 5000–6000 m bands show larger coupled SNC–SD biases than lower
377 elevations. The western Plateau is again the main centre of coupled positive bias (Fig.
378 8c). Reanalysis products occupy distinct positions in the same space: ERA5-Land is
379 coupled positive, MERRA-2 is depth-positive but SNC-negative, and JRA-55 is near
380 neutral for SD but lacks an SNC coordinate (Fig. 8d). This placement shows why a



381 single snow-bias score would be misleading.



382

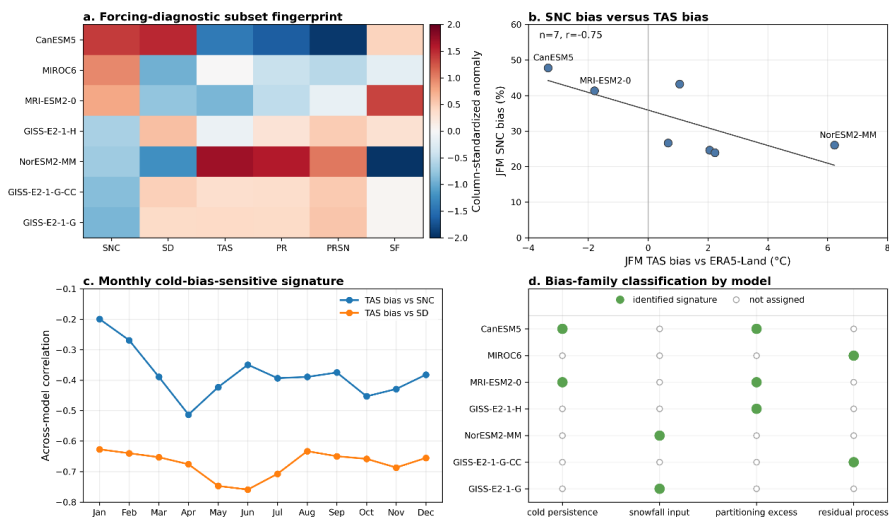
383 **Figure 8.** Cover-depth bias space for CMIP6 and reanalysis products. (a) JFM SNC–SD bias space for
 384 CMIP6 models and reanalysis products; (b) monthly trajectory of the CMIP6 median bias; and (c)
 385 elevation- and region-dependent SNC–SD coupling. Dashed lines mark the positive-bias thresholds.
 386 IP SL-CM6A-LR is clipped for display, with its full SD bias labelled.

387 Forcing signatures add interpretation but do not turn the benchmark into an attribution
 388 experiment. In the forcing-signature subset, JFM SNC bias is negatively related to TAS
 389 bias, with $r = -0.75$, consistent with a cold-biased persistence signature (Fig. 9a, b).



390 Monthly correlations between TAS bias and snow-state bias remain mostly negative,
391 and the strongest TAS–SD correlations reach about -0.75 to -0.76 in May–June (Fig.
392 9c). Snowfall input and snowfall fraction explain parts of the model spread, but their
393 signals are less uniform than the temperature relationship. The model-level fingerprint
394 separates cold persistence, snowfall-input, snowfall-partitioning and residual snow-
395 process signatures (Fig. 9d). Excessive TP snow in CMIP6 is therefore not produced by
396 one pathway alone.

397



398

399 **Figure 9.** Forcing signatures associated with CMIP6 snow-state biases. (a) Standardized JFM forcing-
400 signature matrix for models with complete TAS, PR and PRSN diagnostics; (b) JFM SNC bias versus
401 TAS bias; (c) monthly across-model correlations between TAS bias and SNC or SD bias; and (d) model-
402 level bias-family assignments. The analysis is diagnostic and is not used as formal attribution.

403 Table 2 summarizes the model-level JFM cover–depth classes and provides the basis
404 for model-use screening. GFDL-CM4 is the least biased option for SNC-focused
405 applications, although its SD remains positive. CIESM is closest for SD-focused
406 applications but still strongly overestimates SNC. No model is fully robust for coupled
407 snow-state applications under the default thresholds. Several models can support
408 forcing-signature diagnosis where TAS, PR and PRSN are available, but their snow-



state biases still require explicit screening. The main practical result is strict: TP snow products should not be used for coupled snow-state applications without first checking their position in cover–depth bias space.

Table 2. Model-level JFM snow-state biases and cover-depth classes in CMIP6.

Model	JFM SNC bias (%)	JFM SD bias (m)	Cover-depth class	Main caution
GFDL-CM4	9.09	0.080	SD-positive only	Low SNC bias; SD remains positive
CIESM	30.97	0.013	SNC-positive only	Low SD bias; SNC remains high
NorESM2-LM	21.28	0.061	Coupled positive	—
CESM2-WACCM	24.96	0.077	Coupled positive	—
NorESM2-MM	26.08	0.079	Coupled positive	—
EC-Earth3-Veg-LR	70.08	0.084	Coupled positive	Large SNC bias
EC-Earth3-Veg	69.68	0.092	Coupled positive	Large SNC bias
EC-Earth3	69.77	0.096	Coupled positive	Large SNC bias
CESM2-FV2	27.43	0.101	Coupled positive	—
TaiESM1	27.34	0.102	Coupled positive	—
CESM2	32.16	0.116	Coupled positive	—
MIROC6	43.24	0.120	Coupled positive	—
MRI-ESM2-0	41.36	0.146	Coupled positive	—
SAM0-UNICON	39.04	0.154	Coupled positive	—
BCC-CSM2-MR	21.67	0.301	Coupled positive	Large SD bias
GISS-E2-1-G	23.92	0.328	Coupled positive	Large SD bias
GISS-E2-1-G-CC	24.63	0.343	Coupled positive	Large SD bias
GISS-E2-1-H	26.66	0.361	Coupled positive	Large SD bias
BCC-ESM1	23.84	0.420	Coupled positive	Large SD bias
CanESM5	47.82	0.502	Coupled positive	Large coupled positive bias
IPSL-CM6A-LR	40.86	1.832	Coupled positive	Extreme SD outlier

Note. Biases are calculated on each model native grid and averaged over the TP for JFM. Cover-depth classes use the default thresholds of 10% for SNC and 0.05 m for SD.

5 Discussion

5.1 Why TP snow requires a coupled cover–depth benchmark

The need for a cover–depth benchmark comes from the snow state of the Tibetan Plateau itself. TP snow is often widespread in winter but remains shallow over much of the region. This combination is well documented by satellite snow-cover products and gridded snow-depth datasets, and it is consistent with the highly seasonal, fragmented and terrain-sensitive nature of Plateau snow (Che et al., 2016; Hao et al., 2022; Wang et al., 2024). In such a system, snow-covered area and snow depth cannot be treated as



423 equivalent measures of model performance. A model may produce too much snow
424 cover without much depth excess, or it may accumulate too much snow within a smaller
425 snow-covered domain.

426 This is the central reason for evaluating SNC and SD as paired but non-equivalent
427 snow-state variables (Jiang et al., 2025; Orsolini et al., 2019). Earlier snow evaluations
428 have often treated snow cover, snow depth, reanalysis products and climate-model
429 outputs in separate workflows. Those studies remain useful, but they do not directly
430 answer whether a product is suitable for applications that require both snow extent and
431 snow amount. Hydrological modelling, snow–albedo feedback, frozen-ground
432 simulations and land–atmosphere coupling all depend on the joint snow state, not on a
433 single snow variable alone (Barnett et al., 2005; Brown and Mote, 2009; Wang et al.,
434 2020).

435 Our paired-reference design makes this distinction explicit. MODIS-derived SNC
436 constrains the cover dimension, while CHE SD constrains the depth dimension. The
437 references are not merged into one artificial “true snow field”. They define two axes of
438 a monthly snow state. This matters over the TP because the reference climatology
439 already shows broad JFM snow cover but limited accumulation. Annual means weaken
440 this contrast and can hide the season when most of the model bias occurs.

441 The native-grid strategy follows the same logic. CMIP6 models, reanalysis products
442 and high-resolution reference datasets operate at different spatial supports. Remapping
443 everything to one common analysis grid is convenient, but it can mix product resolution,
444 reference resolution and interpolation choices into the bias estimate (Lei et al., 2023;
445 Miao et al., 2022). By transferring the references to each product’s native grid before
446 calculating bias, the benchmark keeps the primary comparison closer to the scale at
447 which each product represents snow. This is a practical step, but it is central to the
448 credibility of a multi-product TP benchmark.



449 **5.2 A structured CMIP6 snow-state excess over the TP**

450 The dominant CMIP6 error is not simply “too much snow” over the TP. It is a structured
451 cold-season snow-state excess. Most models overestimate both SNC and SD in JFM,
452 and 19 of 21 models fall into the coupled-positive class under the default thresholds.
453 This is the main result of the benchmark. It shows that the prevailing CMIP6 error is
454 not an isolated snow-cover bias or an isolated snow-depth bias, but a joint
455 overestimation of the two snow-state dimensions. The seasonal structure is clear. SNC
456 bias is largest in January–March and remains positive into early spring, whereas SD
457 bias peaks later, in February–March. This sequence suggests that models first maintain
458 snow-covered conditions over too large an area and then accumulate or retain excessive
459 depth later in the season. Similar issues have been noted in broader snow-model
460 evaluations, where cold bias, precipitation phase, snow-cover depletion, compaction,
461 density and melt processes all affect simulated snow states (Essery et al., 2013; Kouki
462 et al., 2022; Raleigh et al., 2015).

463 The bias is also spatially organized. Coupled positive bias strengthens with elevation
464 and is larger over the western Plateau than over the eastern Plateau. This matters
465 because the western and high-elevation TP contains much of the cold, dry and
466 topographically complex terrain where small errors in temperature, snowfall
467 partitioning or snow persistence can strongly affect monthly snow conditions (Lei et al.,
468 2023; Miao et al., 2022; Wang et al., 2024). A TP-mean bias alone is therefore
469 insufficient. It can miss the regions where snow-state errors are most persistent and
470 most relevant for downstream cryospheric and hydrological applications. The SD
471 benchmark adds another caution. Several models show moderate positive SD bias, but
472 a few produce very large errors. IPSL-CM6A-LR is the clearest outlier, and CanESM5
473 and BCC-ESM1 also show large SD excesses. Mean-based ensemble summaries are
474 therefore poorly suited to this problem. Model-level screening, medians and
475 interquartile ranges give a more defensible view of CMIP6 snow depth over the TP.

476 The reanalysis comparison supports the same conclusion from a different product class.



ERA5-Land is close to a coupled-positive case, MERRA-2 underestimates SNC while overestimating SD, and JRA-55 provides only an SD-side constraint in the current workflow. This is consistent with the dependence of reanalysis snow fields on product-specific land models, assimilation choices and precipitation forcing (Gelaro et al., 2017; Kobayashi et al., 2015; Orsolini et al., 2019). Reanalyses are valuable comparison products, but they are not automatically reliable snow-state references over the TP.

The forcing signatures give useful clues without turning the benchmark into an attribution experiment. The negative relation between TAS bias and SNC bias is consistent with cold-biased snow persistence. Other models show signatures more consistent with snowfall input, snowfall partitioning or residual snow-process behaviour. This mixture is expected, because TP snow errors arise from both atmospheric forcing and snow-scheme choices (Kouki et al., 2022; Miao et al., 2022). The key point is not that one mechanism explains all models, but that the cover–depth benchmark narrows the error families that require further process diagnosis.

5.3 Consequences for model use in TP snow-related applications

The model-use implication is direct: acceptable performance in one snow-state dimension does not guarantee acceptable performance in the other. GFDL-CM4 has the smallest JFM SNC bias among the evaluated CMIP6 models, but its SD remains positively biased. CIESM is closest to the CHE SD reference, but it still strongly overestimates SNC. These two cases show why model selection should be application-specific. Snow-cover applications and snow-depth applications may require different model choices. For coupled snow-state applications, the benchmark gives a stricter message. No CMIP6 model is fully robust under the default JFM cover–depth thresholds. This is important for TP studies because many applications depend on snow cover and snow depth together. Snow cover affects albedo and surface radiative forcing; snow depth affects insulation, water storage and melt timing. Bias in either variable can propagate into runoff, soil temperature, frozen-ground simulations and land–atmosphere exchange (Barnett et al., 2005; Brown and Mote, 2009; Wang et al., 2020;



505 Xiao et al., 2025).

506 This issue is especially relevant for process attribution and future projection studies.
507 CMIP6 snow outputs are often used as drivers, covariates or boundary conditions in
508 cryospheric and hydrological analyses. If the historical snow state already contains a
509 strong coupled positive bias, projected snow changes or snow-mediated climate
510 impacts may inherit part of that error (Jiang et al., 2025; Kouki et al., 2022). The risk is
511 highest in the western and high-elevation TP, where the present benchmark identifies
512 the strongest coupled bias. Historical snow-state screening should therefore precede the
513 use of CMIP6 snow outputs in TP hydrology, frozen-ground, glacier and land–
514 atmosphere studies.

515 The same caution applies to reanalysis products. ERA5-Land, MERRA-2 and JRA-55
516 should be treated as product-specific estimates rather than as uniformly better
517 alternatives to CMIP6. Their strengths differ by variable. For applications requiring
518 both cover and depth, reanalyses should be placed in the same cover–depth bias space
519 as climate models (Orsolini et al., 2019). This is one of the practical contributions of
520 the benchmark: it provides a common diagnostic language for models and reanalyses.
521 Forcing diagnostics have a narrower role. Complete TAS, PR and PRSN fields help
522 identify whether a model error is more consistent with cold persistence, snowfall input
523 or snowfall partitioning. They do not make a biased snow product suitable for direct
524 use. In this study, the cover–depth benchmark is the screening layer; forcing signatures
525 are the interpretive layer.

526 **5.4 Limitations and future benchmark extensions**

527 This benchmark is monthly by design and does not resolve daily snow onset, melt
528 timing, snow-spell persistence or event-scale snowfall continuity. These diagnostics are
529 important over the TP but are difficult to evaluate consistently because daily snow-state
530 outputs are limited across CMIP6. Future model intercomparison projects, including
531 CMIP7, would benefit from standardized daily snow-cover fraction, snow depth,
532 snowfall and snow water equivalent outputs, which would allow snow persistence and



533 melt-season errors to be tested directly. The references also carry uncertainty. MODIS-
534 derived SNC can be affected by cloud, terrain and fractional-snow retrieval issues. CHE
535 SD provides a consistent gridded depth constraint, but passive-microwave snow-depth
536 retrievals remain sensitive to snow grain size, liquid water, forest cover, terrain and
537 shallow-snow conditions, which should be considered when interpreting small SD
538 differences among low-bias models. These uncertainties should be kept in mind when
539 comparing low-bias models. They do not, however, explain away the main benchmark
540 result: the dominant CMIP6 biases are large, coherent and concentrated in specific
541 seasons and regions.

542 The forcing analysis is deliberately conservative. TAS, PR and PRSN are used to
543 identify signatures, not causal pathways. Formal attribution would require controlled
544 experiments, snow-scheme diagnostics, precipitation-phase tests or offline land-surface
545 simulations. The present benchmark is better viewed as a screening and diagnosis tool.
546 It shows which models are biased, where the biases occur, and which error families
547 deserve closer process-level investigation. Future work can extend the framework in
548 three directions. A daily benchmark could evaluate onset, melt timing and persistence.
549 A storage-focused benchmark could include snow water equivalent where reliable
550 references are available. Projection studies could use the historical cover–depth
551 position of each model as a prior screen before analysing future TP snow change. The
552 value of the present benchmark is that it makes this screening explicit: TP snow
553 products should be selected according to their cover–depth behaviour, not by a single
554 snow-bias score or by product category alone.

555 **6 Conclusions**

556 This study developed a monthly native-grid cover–depth benchmark for Tibetan Plateau
557 snow states. Its main value is to evaluate snow-covered area percentage and snow depth
558 as paired but non-equivalent variables, rather than reducing model performance to a
559 single snow-bias metric. By transferring MODIS-derived SNC and CHE SD references
560 to each product’s native grid, the benchmark reduces scale mixing and makes the



561 comparison closer to the spatial support of the evaluated product. The benchmark shows
562 that CMIP6 snow errors over the Tibetan Plateau are dominated by a coupled cold-
563 season excess. Most models overestimate both SNC and SD during January–March,
564 and the strongest biases occur over the western and high-elevation Plateau. This pattern
565 is not a uniform domain-wide offset; it is a structured snow-state error that would be
566 partly hidden by annual means, single-variable evaluation or ensemble-mean
567 summaries.

568 The main contribution of this work is to make Tibetan Plateau snow-product use more
569 selective and application specific. GFDL-CM4 is the least biased option for SNC-
570 focused applications, and CIESM is closest for SD-focused applications, but no
571 evaluated CMIP6 model is fully robust for coupled snow-state use under the default
572 January–March thresholds. Reanalysis products also show variable-dependent
573 behaviour and should not be treated as uniformly reliable alternatives. Historical cover-
574 depth performance should therefore be checked before CMIP6 or reanalysis snow
575 products are used in Tibetan Plateau hydrological, cryospheric, land–atmosphere or
576 projection studies.



577 **Code/data availability.**

578 The CMIP6 historical simulations used in this study are available through the Earth
579 System Grid Federation (ESGF; <https://esgf-node.llnl.gov/search/cmip6/>). ERA5-Land
580 data are available from the Copernicus Climate Data Store. MERRA-2 data are
581 available from NASA Goddard Earth Sciences Data and Information Services Center
582 (GES DISC). JRA-55 data are available from the Japan Meteorological Agency. The
583 MODIS-derived snow-cover product is available from the National Glacial, Permafrost
584 and Desert Data Center (<http://www.ncdc.ac.cn>). The CHE_SND_CHINA snow-depth
585 dataset are distributed by the National Tibetan Plateau Data Center
586 (<https://data.tpdc.ac.cn>). Essential Python scripts documenting the native-grid reference
587 aggregation and benchmark calculations, together with selected source data for key
588 diagnostic results, will be deposited on Zenodo, a private reviewer-access link is
589 provided during peer review:
590 https://zenodo.org/records/20454026?token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjVhNzljOTc1LTljYmMtNDE1OS1hNDIhLTdlMmZjNDA0MmZiMSIsImRhdGEiOnt9LCJyZW5kb20iOiI3NDQyZGUzMzk0M2U1MjRlYTk2ZTllODgxYjIxNWlyNCJ9.jaBO82o6M0kVVvu4jIji4JLUoStB2k_5IAkwUIQVkvzQOLmTOlOAq7ULVxzWOLq5ZHY4qCeWu3H5PSOp8qJ1-w.

595 **Author contributions.**

596 YX and GH designed the study. YX performed data processing, benchmark analysis,
597 visualization and manuscript drafting. NZ contributed to data preparation and
598 diagnostic analysis. GH, RL and LZ supervised the study and refined the scientific
599 framing. RL, TW, XW, GL, DZ, ED, YZ and LZ contributed to the interpretation of the
600 results and manuscript revision. All authors discussed the results, reviewed the
601 manuscript and approved the submitted version.

602 **Competing interests.**

603 The authors declare that they have no conflict of interest.



604 **Acknowledgements**

605 The authors thank the CMIP6 modelling groups, the Earth System Grid Federation, the
606 Copernicus Climate Data Store, NASA GES DISC and the Japan Meteorological
607 Agency for providing the model and reanalysis datasets used in this study. The authors
608 also acknowledge the National Glacial, Permafrost and Desert Data Center and the
609 National Tibetan Plateau Data Center for providing the MODIS-derived snow-cover
610 product and the CHE_SND_CHINA snow-depth dataset. We thank the scientific data
611 providers and platform teams who made these datasets publicly available. The authors
612 used ChatGPT-5.5 only for language refinement and readability improvement. All
613 scientific analyses, interpretations and conclusions were developed and verified by the
614 authors.

615 **Financial support.**

616 This research was supported by the National Natural Science Foundation of China
617 (42322608, 42471168, W2412013, U23A2062), and the program of the Key Laboratory
618 of Cryospheric Science and Frozen Soil Engineering, CAS (CSFSE-ZQ-2407 and
619 CSFSE-ZZ-2408).

620

621



622 References

- 623 Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential impacts of a warming climate on
624 water availability in snow-dominated regions, *Nature*, 438, 303–309,
625 <https://doi.org/10.1038/nature04141>, 2005.
- 626 Brown, R. D. and Mote, P. W.: The response of northern hemisphere snow cover to a
627 changing climate*, *Journal of Climate*, 22, 2124–2145,
628 <https://doi.org/10.1175/2008JCLI2665.1>, 2009.
- 629 Che, T., Dai, L., and Li, X.: Long-term series of daily snow depth dataset in China (1979–
630 2024), National Tibetan Plateau/Third Pole Environment Data Center,
631 <https://doi.org/10.11888/Geogra.tpd.c.270194>, 2015.
- 632 Che, T., Dai, L., Zheng, X., Li, X., and Zhao, K.: Estimation of snow depth from passive
633 microwave brightness temperature data in forest regions of northeast China, *Remote*
634 *Sensing of Environment*, 183, 334–349, <https://doi.org/10.1016/j.rse.2016.06.005>, 2016.
- 635 Chen, X., Long, D., Liang, S., He, L., Zeng, C., Hao, X., and Hong, Y.: Developing a
636 composite daily snow cover extent record over the tibetan plateau from 1981 to 2016
637 using multisource data, *Remote Sens. Environ.*, 215, 284–299,
638 <https://doi.org/10.1016/j.rse.2018.06.021>, 2018.
- 639 Essery, R., Morin, S., Lejeune, Y., and B Ménard, C.: A comparison of 1701 snow models
640 using observations from an alpine site, *Adv. Water Resour.*, 55, 131–148,
641 <https://doi.org/10.1016/j.advwatres.2012.07.013>, 2013.
- 642 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K.
643 E.: Overview of the coupled model intercomparison project phase 6 (CMIP6)
644 experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958,
645 <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- 646 Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller,
647 M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M.,
648 Oskin, M., Burbank, D., and Alsdorf, D.: The shuttle radar topography mission, *Reviews*
649 *of Geophysics*, 45, 2005RG000183, <https://doi.org/10.1029/2005RG000183>, 2007.
- 650 Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A.,
651 Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R.,
652 Draper, C., Akella, S., Buchard, V., Conaty, A., Da Silva, A. M., Gu, W., Kim, G.-K.,
653 Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman,
654 W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The modern-era
655 retrospective analysis for research and applications, version 2 (MERRA-2), *J. Climate*,
656 30, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>, 2017.
- 657 Hao, X., Huang, G., Zheng, Z., Sun, X., Ji, W., Zhao, H., Wang, J., Li, H., and Wang, X.:
658 Development and validation of a new MODIS snow-cover-extent product over China,
659 *Hydrol. Earth Syst. Sci.*, 26, 1937–1952, <https://doi.org/10.5194/hess-26-1937-2022>,
660 2022.
- 661 Jiang, Y., Che, T., and Dai, L.-Y.: Inconsistent response patterns of snow cover duration and
662 snow depth over the Tibetan Plateau to global warming, *Adv. Clim. Change Res.*, 16,
663 501–511, <https://doi.org/10.1016/j.accre.2025.03.007>, 2025.



- 664 Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M., Onoda, H., Onogi, K., Kamahori,
665 H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 reanalysis:
666 General specifications and basic characteristics, *J. Meteorol. Soc. Jpn.*, II, 93, 5–48,
667 <https://doi.org/10.2151/jmsj.2015-001>, 2015.
- 668 Kouki, K., Räisänen, P., Luojus, K., Luomaranta, A., and Riihelä, A.: Evaluation of northern
669 hemisphere snow water equivalent in CMIP6 models during 1982–2014, *The*
670 *Cryosphere*, 16, 1007–1030, <https://doi.org/10.5194/tc-16-1007-2022>, 2022.
- 671 Lei, Y., Pan, J., Xiong, C., Jiang, L., and Shi, J.: Snow depth and snow cover over the tibetan
672 plateau observed from space in against ERA5: Matters of scale, *Clim. Dyn.*, 60, 1523–
673 1541, <https://doi.org/10.1007/s00382-022-06376-0>, 2023.
- 674 Miao, X., Guo, W., Qiu, B., Lu, S., Zhang, Y., Xue, Y., and Sun, S.: Accounting for
675 topographic effects on snow cover fraction and surface albedo simulations over the
676 tibetan plateau in winter, *JAMES*, 14, e2022MS003035,
677 <https://doi.org/10.1029/2022MS003035>, 2022.
- 678 Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G.,
679 Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G.,
680 Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.:
681 ERA5-land: A state-of-the-art global reanalysis dataset for land applications, *Earth Syst.*
682 *Sci. Data*, 13, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021>, 2021.
- 683 Orsolini, Y., Wegmann, M., Dutra, E., Liu, B., Balsamo, G., Yang, K., De Rosnay, P., Zhu, C.,
684 Wang, W., Senan, R., and Arduini, G.: Evaluation of snow depth and snow cover over
685 the tibetan plateau in global reanalyses using in situ and satellite remote sensing
686 observations, *The Cryosphere*, 13, 2221–2239, <https://doi.org/10.5194/tc-13-2221-2019>,
687 2019.
- 688 Raleigh, M. S., Lundquist, J. D., and Clark, M. P.: Exploring the impact of forcing error
689 characteristics on physically based snow simulations within a global sensitivity analysis
690 framework, *Hydrol. Earth Syst. Sci.*, 19, 3153–3179, [https://doi.org/10.5194/hess-19-](https://doi.org/10.5194/hess-19-3153-2015)
691 [3153-2015](https://doi.org/10.5194/hess-19-3153-2015), 2015.
- 692 Wang, J., Tang, L., and Lu, H.: The new indices to describe temporal discontinuity of snow
693 cover on the qinghai-tibet plateau, *npj Clim. Atmos. Sci.*, 7, 189,
694 <https://doi.org/10.1038/s41612-024-00733-y>, 2024.
- 695 Wang, W., Yang, K., Zhao, L., Zheng, Z., Lu, H., Mamtimin, A., Ding, B., Li, X., Zhao, L.,
696 Li, H., Che, T., and Moore, J. C.: Characterizing Surface Albedo of Shallow Fresh Snow
697 and Its Importance for Snow Ablation on the Interior of the Tibetan Plateau, *J.*
698 *Hydrometeorol.*, 21, 815–827, <https://doi.org/10.1175/JHM-D-19-0193.1>, 2020.
- 699 Xiao, Y., Hu, G., Li, R., Xiao, M., Li, Z., Zhao, J., Xing, Z., Wu, T., Wu, X., Liu, G., and
700 Zhao, L.: Beyond carbon: multi-scale thermal and hydrological feedback of permafrost
701 on the Tibetan Plateau, *Earth Sci. Rev.*, 270, 105248,
702 <https://doi.org/10.1016/j.earscirev.2025.105248>, 2025.
- 703 Yan, D., Ma, N., and Zhang, Y.: Development of a fine-resolution snow depth product based
704 on the snow cover probability for the Tibetan Plateau: validation and spatial–temporal
705 analyses, *Journal of Hydrology*, 604, 127027,
706 <https://doi.org/10.1016/j.jhydrol.2021.127027>, 2022.



707 You, Q., Wu, T., Shen, L., Pepin, N., Zhang, L., Jiang, Z., Wu, Z., Kang, S., and
708 AghaKouchak, A.: Review of snow cover variation over the Tibetan Plateau and its
709 influence on the broad climate system, Earth-Science Reviews, 201, 103043,
710 <https://doi.org/10.1016/j.earscirev.2019.103043>, 2020.

711

712