

Supplement

The Supplement contains five figures. Figure S1 shows that the surface-matched subsurface chlorophyll fraction bias is retained under both BGC-Argo chlorophyll definitions (Section 3.7). Figure S2 summarizes the per-model median biases over the full comparison (Section 3.2). Figure S3 shows the per-model vertical structure within each observed-DCM regime (Sections 3.3 and 4.5). Figure S4 follows each model across the three observed-DCM regimes and shows that all nine place their maximum deeper and express a weaker vertical contrast relative to the observations as the regime deepens, with no model reproducing the observed profile (Section 3.4). Figure S5 stratifies the surface-matched subsurface chlorophyll fraction bias by observed-DCM regime and compares two surface-match definitions, showing that its negative sign is robust to the definition, and to the choice of the absolute threshold, while its magnitude and the ordering among regimes are not (Sections 3.6 and 4.1).

The observed DCM depth is identical between the reprocessed and the earlier factor-two BGC-Argo chlorophyll definitions, confirming that the reprocessing does not alter the vertical position of chlorophyll features (Section 2.2). For the eight models that resolve deeper than 137.5 m, extending the subsurface-fraction window to 175 m leaves the fraction essentially unchanged, and in the model-minus-observation bias the window contribution cancels because it enters observations and models equally (Section 2.4).

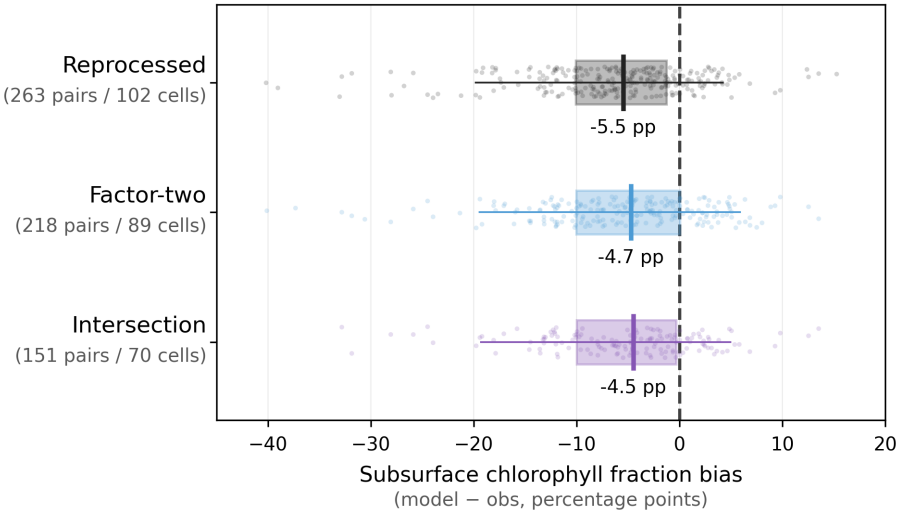
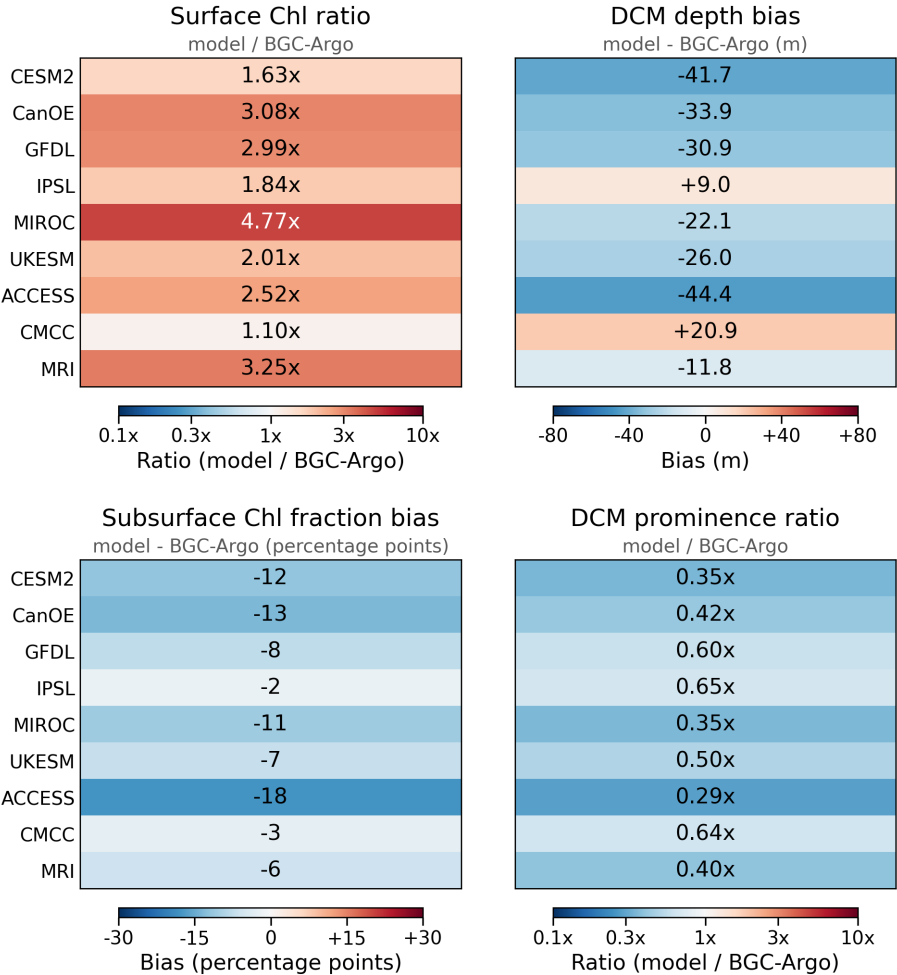


Figure S1. Reprocessing-independence of the surface-matched subsurface chlorophyll fraction bias (control figure). Subsurface chlorophyll fraction bias (model minus observation, percentage points) within the surface-matched band, recomputed under three BGC-Argo chlorophyll definitions: the reprocessed product (263 model-cell pairs / 102 cells), the earlier global factor-two product (218 / 89), and their intersection (151 / 70). Because a per-cell, depth-independent rescaling leaves the subsurface chlorophyll fraction bias of a

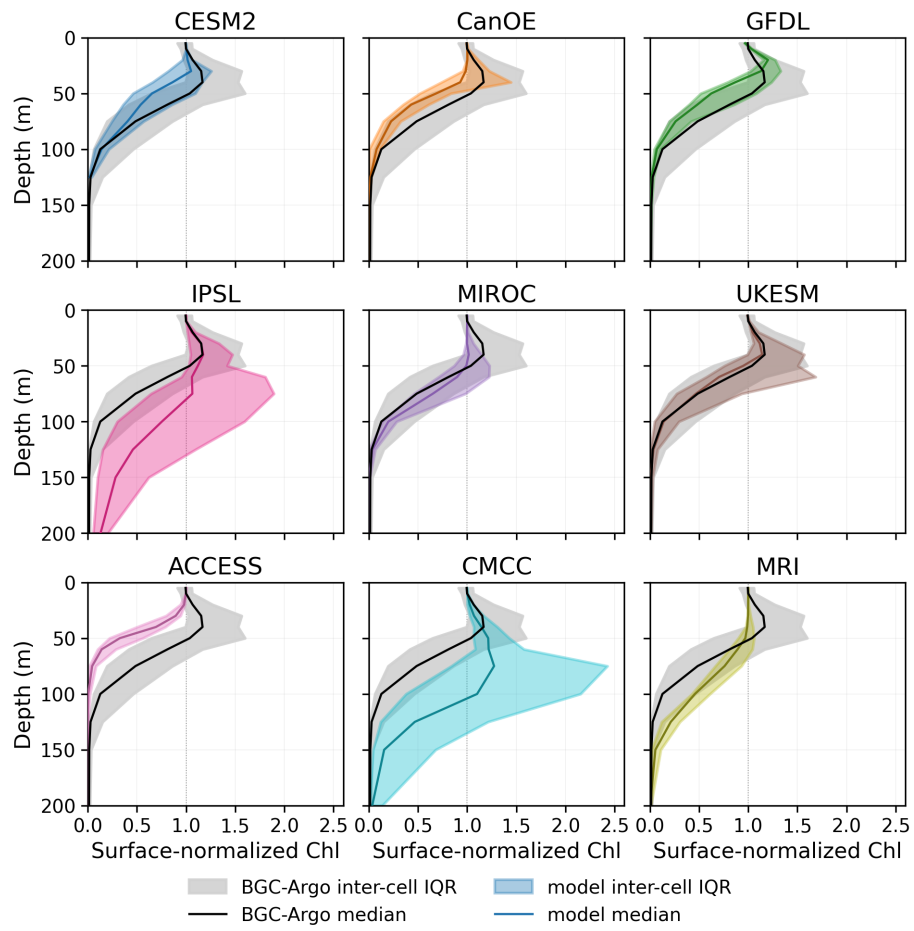
25

given pair unchanged (Section 3.7), the three definitions differ only in which pairs satisfy the surface-matching criterion, so the same pair carries the same bias value in each row. In each row, points are individual model–cell pairs (jittered), the box is the interquartile range, the heavy vertical line is the median, whiskers span the 5th–95th percentiles, and the dashed line marks zero bias. The pair-based median is -5.5 , -4.7 , and -4.5 percentage points for the reprocessed, factor-two, and intersection definitions, respectively; the median is below zero in every case, and the upper quartile is no higher than $+0.04$ percentage points — that is, at most marginally above zero — in every case. The pooled pairs are not independent, so the distributions are shown descriptively and no significance tests are applied.

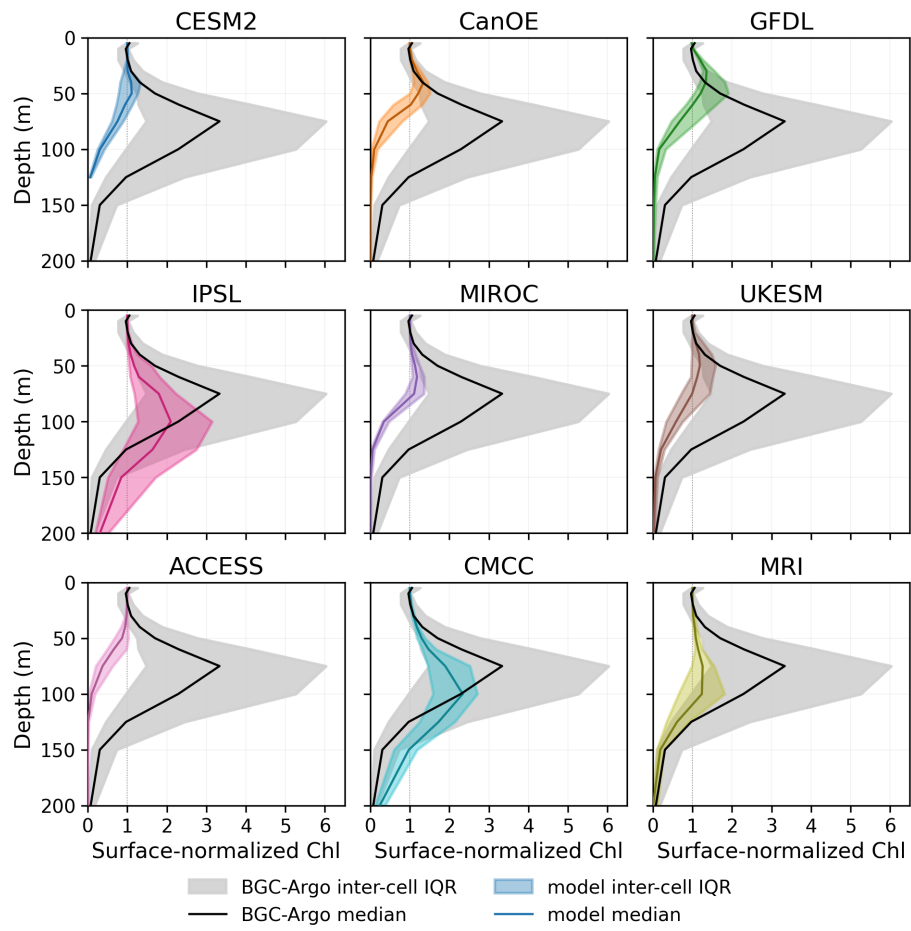


30

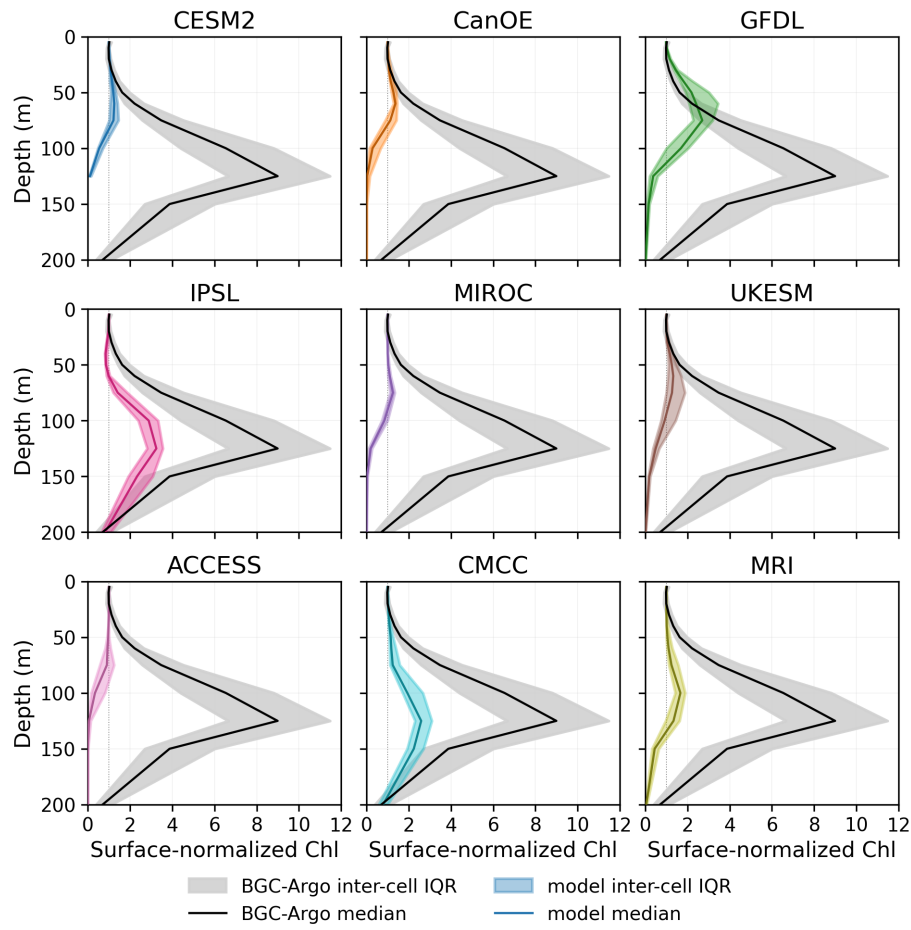
Figure S2. Model-by-model summary of vertical chlorophyll biases (median vertical-structure biases for each of the nine CMIP6 models; 1,809 pairs across 201 cells, balanced at 201 per model). Surface chlorophyll and DCM prominence are model-to-BGC-Argo ratios; DCM-depth and subsurface-fraction biases are model minus BGC-Argo. The subsurface-fraction and prominence biases are negative and below unity, respectively, in all nine models; the DCM-depth bias varies in sign. The figure is not a performance ranking.



(a)



(b)



(c)

Figure S3. Per-model vertical chlorophyll structure within observed DCM-depth regimes: (a) shallow, (b) intermediate, and (c) deep ($n = 67$ comparison cells each), each showing the nine models. As in Fig. 3, the grey band and black line are the BGC-Argo inter-cell interquartile range and median (identical in every panel of a given regime), and the coloured band and line are each model's own inter-cell interquartile range and median of chlorophyll normalized to its 0–10 m median (median across cells; full names in Table 1). Stratifying by regime isolates within-regime spatial diversity from the between-regime spread that broadens the pooled comparison in Fig. 3. The x-axis is linear, with an upper limit that differs among the regimes (a, 2.6; b, 6.5; c, 12.0) so that each regime's spread is resolved — note the different horizontal scales when comparing across (a)–(c); the dotted line marks the surface reference ($=1$). CESM2's chlorophyll is resolved only to about 137.5 m on the common grid and is therefore truncated at depth. Pooled model–cell comparisons within a regime are not independent and are summarized descriptively without significance tests.

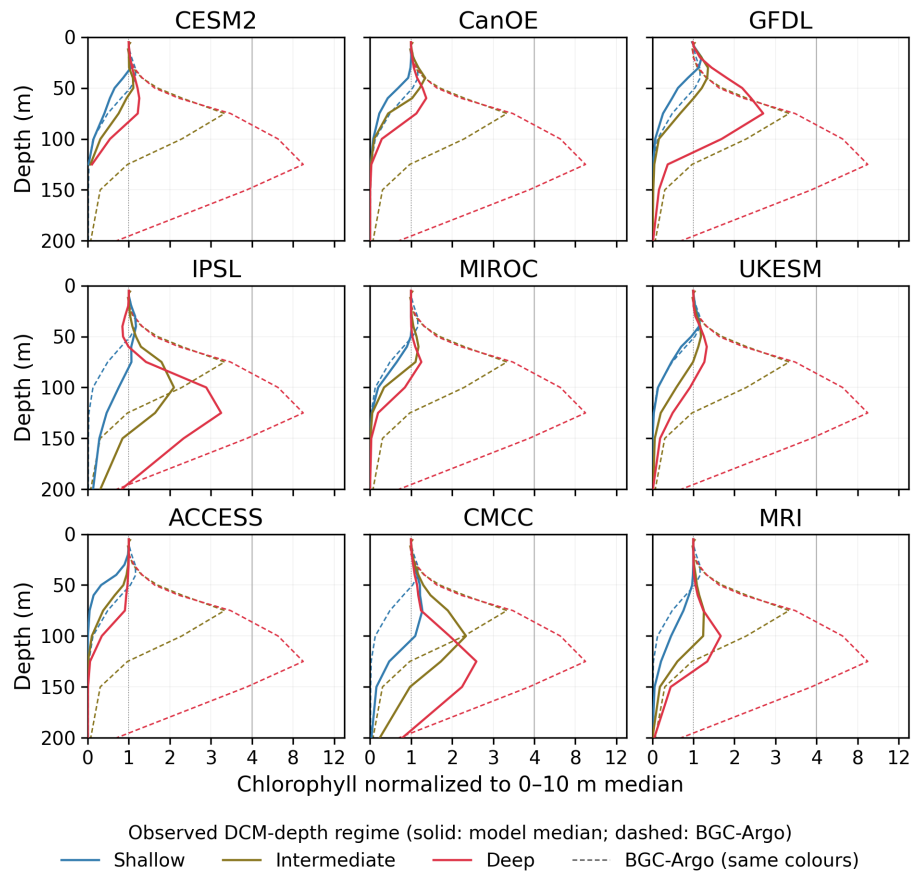
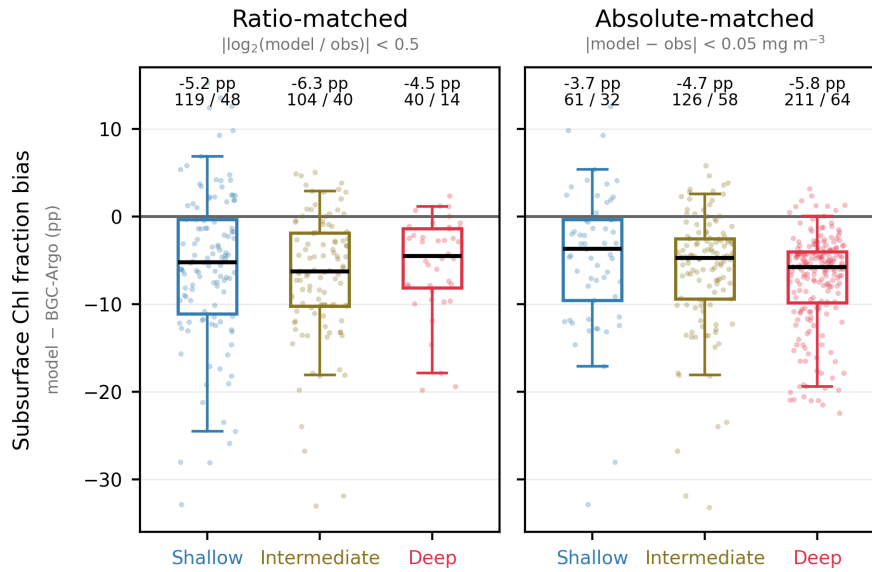


Figure S4. Per-model vertical structure across the three observed DCM-depth regimes. Each panel is one model. Within a panel, the three solid curves are that model's median normalized chlorophyll profile in the shallow, the intermediate and the deep regime (legend); the dashed curves are the BGC-Argo medians for the same three regimes, drawn in the same colours, and are identical in every panel. Chlorophyll is normalized to the 0–10 m median of each cell, the same quantity as in Figs 3 and 4, and the horizontal axis is the piecewise-linear scale of Fig. 4 (linear to 4, compressed above it; the vertical line marks the break, the dotted line marks unity). Each regime contains 67 of the 201 comparison cells (Fig. A1); the medians are taken over cells within a regime at each common depth. This figure is the transpose of Fig. 4: Fig. 4 compares the models within a regime, and this one follows one model across the regimes. Every model places its subsurface maximum deeper from the shallow to the deep regime, and in every model its vertical contrast relative to the observations weakens as the regime deepens; no model reproduces the observed profile in any regime, and the shortfall is largest where the observed DCM is deepest. CESM2's chlorophyll is resolved only to about 137.5 m on the common grid, so its curves stop there. Cells pooled within a regime are not independent, so no significance test is attached to the comparison.



60

65

70

75

Figure S5. Subsurface chlorophyll fraction bias in surface-matched cells, by observed-DCM regime and surface-match definition. For each model–cell pair whose surface chlorophyll matches BGC-Argo, the subsurface chlorophyll fraction bias (model minus observation, percentage points; negative values indicate that the model under-represents subsurface chlorophyll) is shown against the observed DCM-depth regime (shallow, intermediate, deep); here obs denotes BGC-Argo surface chlorophyll. Two surface-match definitions are compared: ratio-matched, $|\log_2(\text{model} / \text{obs})| < 0.5$ (left), and absolute-matched, $|\text{model} - \text{obs}| < 0.05 \text{ mg m}^{-3}$ (right). In each panel the box is the interquartile range, the heavy horizontal line is the median, whiskers span the 5th–95th percentiles, and jittered points are individual matched pairs; the number above each box is the median in percentage points followed by the number of matched pairs and of distinct cells. Because pairs within a cell or a model are not independent, no significance test is applied and both pair- and cell-based counts are reported. In every regime and under both definitions the median lies below zero, so the sign of the subsurface deficit is robust to the surface-match definition; its magnitude and the ordering among regimes are not — under absolute matching the deep regime shows the largest deficit, whereas under ratio matching the deep regime is captured by only a small, biased subset (40 pairs in 14 cells), because ratio matching is stringent where observed surface chlorophyll is low: in all 67 deep-regime cells the absolute criterion admits a wider range of modelled surface chlorophyll than the ratio criterion, so the ratio-matched deep pairs are contained within the absolute-matched ones (40 of 211). Summed over the three regimes, ratio matching retains 263 pairs in 102 cells and absolute matching 398 pairs in 154 cells. The threshold 0.05 mg m^{-3} is nominal: repeating the absolute match with 0.05, 0.10 and 0.20 mg m^{-3} leaves the pair-based median below zero in every regime, and in the intermediate and the deep regimes it becomes more negative as the threshold is relaxed (deep regime: -5.8 , -6.8 and -8.6 percentage points), so the values shown here are the conservative ones; in the shallow regime the median does not vary monotonically with the threshold. The subsurface chlorophyll fraction is not normalized by surface chlorophyll and therefore shares no denominator with either matching variable.