

Review of Fresh Eyes on CMIP Model Biases: Diagnosing Biases Across CMIP Generations and Implications for CMIP7; Brown et al.,

Decision: major revisions needed.

The submitted manuscript presents an overview of common climate model biases that have been documented and persistent across multiple CMIP generations. Specifically, the authors focus on variables (diagnostics) that have been selected for assessment by the Rapid Evaluation Framework (REF). The paper aims to provide a literature review of the performance across CMIP generations for the REF diagnostics and is aimed at end-users who may be new to CMIP or are looking to connect realms (e.g., understand consequences or relevance of biases in the ocean and their connection to ice, atmosphere, land, etc.). I recognize that what the authors are trying to achieve is a large and valuable task, so I commend their efforts taking this on. Specifically, it is very important to clearly summarize model biases and the consequences of them to end users who might not have domain expertise but wish to use CMIP results for various applications. **Therefore, I think the topic is important to the community.**

However, in its current state and given its goals and intended audience, I believe this manuscript needs substantial revision prior to being ready for publication. I have made many comments below in my review (I have reviewed the entire manuscript, but my comments are mostly focused on the ocean and ice portions, where my expertise lies). Overall, given that the authors state one of their goals is to present a literature review, I felt that several topics were incomplete in their assessment of relevant literature and would greatly benefit from more discussion. The entire paper would also greatly benefit from attempting to quantify and discuss biases from CMIP3 to CMIP6. Much of the literature discussed, I felt was focused on CMIP6 and there was limited discussion of cross-generational performance / improvements. There was also limited discussion / examples of “success stories” in bias reduction across generations. Can the authors identify any within these realms?

There are also issues throughout the text regarding accuracy and clarity of some statements, which I have noted in the line-comments below.

I also strongly believe that the manuscript would **greatly benefit** from the inclusion of a set of figures for each diagnostic discussed to illustrate the biases across the CMIP ensemble. This should be feasible, especially since most of these diagnostics are relatively lightweight to compute and the authors could even leverage REF software / code if necessary. This would greatly increase the impact and readability of the paper, as right now it is very hard to follow the discussion without maps and visualizations to illustrate the issues at hand.

Line-by-line and minor comments can be found below:

Abstract:

“The Coupled Model Intercomparison Project (CMIP) **simulates** standardized experiments spanning historical, future, and hypothetical conditions, to better understand the Earth system’s evolution.” → **I suggest using another word than “simulates” here as CMIP itself does not “simulate” anything. CMIP “provides the framework to standardize, organize, and archive results from ...”.**

“Systematic biases remain” → **Systematic biases of climatically relevant processes ... all models will have biases ... the limitation is when they are in regions / for processes that are important to the climate system (e.g., by controlling rates of ocean heat uptake, carbon budgets, etc).**

“AMOC’s large scale transport biases” → **apostrophe is off in AMOC. Also, since the sentence already starts with “biases”, you can remove “biases” here after “transport”.**

L12: “enhanced coupling” → “improved coupling” ?

L13: “the current state of understanding” → drop the “s” after “understandings”

L18: “almost 30 years”, → It is officially over 30 years as the framework began in 1995 (AMIP started in 1989). See <https://climate.esa.int/en/news-events/cmip-the-evolution-of-climate-modelling/>.

L22: “historical, future, and experimental conditions,” → I understand what is trying to be conveyed by “experimental conditions”, but it is a bit confusing to a reader not familiar with CMIP Future scenarios are also “experimental conditions”. Can you please reword and be more explicit here by what you mean?

L23: “experience as a function of different climate forcings.” → I suggest “as a result of different possible future emission pathways” ... it really is emission pathways (which come with a lot of other societal assumptions, etc ... not isolated forcings, in the case of SSPs. Can you also please add a citation here?

L25: “Through ~~all the phases of~~ CMIP phases, there have been systematic model biases, where the behavior of **climate phenomena simulated in** historical experiments deviate from observations of the Earth system.” → It is not that the behavior of the historical experiments themselves differ It’s that certain climate processes within the simulations differ. You can have a model that does a good job at the trajectory of global ave SST or SAT (which we do well quite well!), but still has very wrong sea ice representation, for example.

L27: “higher extension” → please add “vertical”.

L33: “The focus of CMIP7 on emission-driven rather than concentration-driven future scenarios will **increase the effect** of how well the carbon cycle is represented in coupled climate models.” → **what does “increase the effect of” imply here? Running in emission driven vs. concentration driven doesn’t necessarily improve *how the carbon cycle is represented*.**

Please reword for accuracy. More comprehensive ESMs are what improve carbon-cycle representation (e.g., more biogeochemical tracers in the ocean for example, or atmospheric chemistry (prescribed vs. interactive ozone) ...). If models have biases in their physical components, running in “emissions-driven” mode and allowing the feedbacks between simulated climate components can actually make things worse! The ocean and atmosphere carbon stocks can end up being out of sink with the observed rate of atmospheric CO₂ increase.

L38: “The REF platform will enable comparison between models and across idealized and historical experiments against observational evidences.” → **I believe it is only historical experiments, not idealized. Please check. Also, change “observational evidences” to “observational estimates” ...**

L40: The Hoffman et al., reference can now be updated to the most recent.

L44: “across all components of the” → Across “multiple components” would be more accurate here as CMIP models do not include *all* components (e.g., ice sheets).

L79: Please fix the apostrophe before S in Sea and elsewhere throughout this section.

L122: “Temperature and salinity variations affect the density of seawater, controlling the ocean’s stratification, lateral pressure gradients **which drive ocean transport**, and **the formation of water masses that fill the ocean interior**. This impacts ocean overturning and circulation at a range of scales.” → See my edits. I think it is very important to mention that stratification is essentially what controls water mass transformation / formation / subduction / transport **this is what is relevant for ocean heat and carbon uptake**

L126: Add “processes” after “climate”. Again ... I think it is important to state that air-sea fluxes set the interior ocean water mass structure and influence ocean heat uptake and ocean carbon uptake ...

L130: Region AND model dependent! Some models have improved quite a bit!!!

L138: “particularly in regions **where observations are sparse** and model constraints are weak especially in the deep ocean **where observations are sparse**”. **Please elaborate, a reader might not understand why biases in the deep ocean lead to SST biases?**

L140: “can introduce initial shocks and model drift”. → This is a very jargony / technical sentence for the target audience. I assume you are referring to “initialization shocks” that lead to “numerical / spurious drift”. Please explain how everything you listed above can lead to model drift, I am not following. This sentence also needs a citation.

L144: “The persisting Southern Ocean warm bias in several generations of CMIP models is mainly attributed to inaccurate representation of AMOC (Wang et al., 2014), and clouds (Hyder et al., 2018).” **This is much too strong of a statement and relies on the results from only the two cited studies. The cause of the warm SST bias in the Southern Ocean is a topic of on-going research and other studies suggest it may also be due to errors in the surface freshwater budget potentially linked to the lack of glacial meltwater inclusion. Please provide more references for this and nuance.**

Figure 6 in Beadling et al., 2020 shows Southern Ocean upper-ocean T and S biases. A key take away: *“The upper ocean remains biased too fresh and too warm relative to observed. There has not been a clear improvement in ensemble performance relative to CMIP3 and CMIP5. There are still several models with exceptionally warm upper SOs with errors that translate into very poor representations of Antarctic SIE.”*

<https://journals.ametsoc.org/view/journals/clim/33/15/jcliD190970.xml>

L151: “distribution, circulation and”. Are you referring to atmosphere or ocean circulation here?

L159 – 160: “Ocean density and related thermodynamics are especially sensitive to salinity variations in polar regions where water is cold, compared to more temperate regions where density fluctuations are largely driven by temperature changes.” → **This is due to the non-linearity in the equation of state of seawater. Please add a citation. A reader from the target audience may not understand why cold temperatures mean a region is more sensitive to salinity (it is because the thermal expansion coefficient depends on temperature).**

L166-167: “, and the Southern Ocean freshwater fluxes are underestimated in CMIP6 (Kaufman et al., 2025).”. → **Indeed, but more explanation is needed** – lack of glacial meltwater fluxes!! Also, a reader might be very confused by this discussion, the fact that you state the ocean is too fresh (also see Figure 6, in Beadling et al., 2020 linked above) but also that the freshwater fluxes are underestimated. Please draw the connection for the reader.

L167: “run-off errors,” → what runoff errors are being referred to here? Too much runoff from Antarctica / Greenland? Please provide a citation that has linked the surface fresh biases to runoff and precip errors.

L168: “The Indian Ocean is another region where SSS biases are important, with salinity stratification forming barrier layers and thin freshwater lenses near the ocean surface, mediating air-sea fluxes, and upper-ocean dynamics and thermodynamics.” → **the use of “barrier layers” is a bit awkward here, I suggest rewording. This statement also is not unique to the Indian Ocean.**

L171: “with thick observed barrier layers, driven”. **Please avoid using the term “barrier layers”**

L174: “summarize” → I am not sure that summarize is the right word here. Integrate?

L176: “(warmed faster than observations)”. → Since you are talking about OHC, which is not just temperature, **but volume**, this should state “(has taken up heat faster than observed)”. Since OHC is dependent on both temperature change and volume, where the heat goes matters for the full layer integral. Same comment on “cooling” used later in the sentence.

L179 – 180: “This bias is attributable to inaccurate SST and biased mixing/surface fluxes into the warmest regions of the ocean” → **What is “this bias” that you are referring to? The study by Sohail et al., referenced here indicates that the COLD bias is traced to SST errors in the warmest regions of the ocean. Please be precise.**

L176 – 178: “near the surface (0 - 200 m) and below 2000 m depth, and a cooling bias between approximately 400 - 2000 m depth (Sohail et al., 2021). The opposing biases compensate and disappear when averaged over 0 - 700 m depth (Takano et al., 2023).” --> This sentence does not make logical sense as written. You mention that the opposing biases are in the upper 200 m (too much OHC) and over 400 to 2000 m (too little OHC), but then the compensation you mention is just within the upper 700 m. Please reword for accuracy.

L180 – 183: “In the Arctic, CMIP6 models exhibit a warm and fresh bias at depth through 1979-2014, with effectively no improvement in Arctic deep hydrography from CMIP5 to CMIP6. It was suggested that these Arctic biases could result from inaccurate mixing schemes, insufficient model resolution, and propagation of biases from adjacent ocean regions (Khosravi et al., 2022).” → This sentence seems very out of place; this does not discuss OHC and instead discusses temperature and salinity biases.

OHC Section overall: I felt that this was a very incomplete discussion of the literature regarding OHC and understanding where sources of these biases may come from. The Southern Ocean is the most important basin for the oceanic uptake of anthropogenic heat, but none of the relevant literature assessing OHC / OH uptake in this region was reviewed. It is very hard to follow the discussion without actual figures in the paper, the

reader is required to go look at all the individual papers cited to understand where in the water column these biases are and whether or not they are significant.

Figure 3. An arrow connecting freshwater fluxes to “Sea Ice & Coastal Processes” should be drawn as freshwater flux biases impacts ocean stratification which influences sea ice formation processes and coastal heat budgets. Freshwater flux biases also impacts the formation of dense waters (e.g., dense shelf water formation along the Antarctic coast).

L193: “It also defines the albedo (reflectivity) of the ocean surface ...” → I would reword for accuracy, **sea ice influences planetary albedo**, it is not part of the ocean surface, it sits on top of the ocean surface.

L193: “short-wave” should not be hyphenated here.

L194: “A loss of sea ice decreases **the** albedo, increases the heat uptake of the ocean” → please be specific about what components of the heat budget you are referring to and that it increases “the planetary” or “the surface” albedo, not just “the albedo”.

L204 – 217: There are more recent studies that can be discussed and referenced here which discuss the persistence of biases and potential sources / remaining uncertainties. E.g., Roach and Polvani 2026

(<https://journals.ametsoc.org/view/journals/clim/39/7/JCLI-D-25-0477.1.xml>), Bushuk et al., 2026: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL124043>.

L245 – 248: “When this trend component is included in comparisons with historical CMIP simulations, the long-term observed trend lies within the spread of modeled SIA trends, leading some to argue that models do in fact reasonably capture multidecadal Antarctic SIA trends overall (Holmes et al., 2024).” → **Please read Roach and Polvani, 2026 and Bushuk et al., 2026 linked above which make a clear case against this statement (using a much larger ensemble of models). It is important that we do not propagate this narrative that models are “better at capturing” an observed trend a trend that is very non-linear!**

L259: “AMOC remains one of the most biased large-scale ocean diagnostics in coupled models.” **Relative to what? Boundary currents? The ACC? Gyres?**

L286 – 288: “Heuzé (2021) also showed that CMIP6 models still struggle to represent Antarctic Bottom Water formation, with overly warm and fresh deep waters and insufficient bottom-water export, highlighting persistent global deficiencies in deep-water mass formation and circulation (Fig. 3).” → I suggest removing this sentence as AABW is driven by very different mechanisms and this sentence loses the focus on the AMOC.

L257: Please add “northern” before “latitudes”.

Section 4 in general: I recognize that this section is supposed to serve as a bringing together of everything that was discussed prior, but there are several large claims being made without any citations.

E.g., “Parameterized mixing affects ocean stratification and deep-ocean heat uptake, contributing to widespread SST and salinity biases (Fig. 3), including the equatorial Pacific cold-tongue bias and Southern Ocean warm bias seen across CMIP6 models.”.

This sentence holds a lot of weight and requires a reader to understand why parameterized mixing might be linked to the Southern Ocean being too warm But in fact, even models that are “eddy permitting” in the mid-latitudes / latitudes of the ACC still show warm SSTs in the Southern Ocean so that is not the full story.

E.g., “Systematic biases in sea ice area (SIA) in both the Arctic and Antarctic arise from the spatiotemporal distribution of ice cover, and the response of sea ice to anthropogenic warming. These discrepancies are associated with deficiencies in the representation of mass balance components, ice-albedo feedbacks, and the underlying physical processes that control sea ice cover. The omission or oversimplification of key processes in some models, such as ice-sheet-ocean interactions and freshwater fluxes, further reduces confidence in simulated sea ice dynamics and projections, and feed into biases in the ocean (Fig. 3).”

E.g., “Inadequate representation of cloud fraction, phase partitioning, and optical depth affects both incoming solar radiation and the trapping of outgoing longwave radiation, contributing to biases in land and ocean, such as warm biases in the Southern Ocean (Fig. 3).”

E.g., “Convection schemes often misrepresent entrainment rates, trigger thresholds, and convective organization, contributing to the double ITCZ problem, unrealistic monsoon variability, and shifted storm tracks.”

Please also break up this discussion (Sections 4.1 & 4.2) into distinct paragraphs for readability.

Section 4.2: You might want to consider referencing this recent paper that discusses what is being summarized:

<https://journals.ametsoc.org/view/journals/bams/107/4/BAMS-D-25-0079.1.xml>

L1065: “Systematic biases arise from interacting uncertainties and parameterizations governing the exchange of energy, water, and carbon across the Earth system.” → I suggest adding “in Earth System Models and Climate Models” or something after systematic biases to indicate you are referring to biases within climate models, or “CMIP models”.

L1066 – 68: I found this sentence very confusing ... you can have a bias that is sourced from the ocean (i.e., you would see it in an OMIP run) ... not everything has to “propagate down through the atmosphere”.

L1069: “and simplified land and ocean biogeochemistry, many”. **Also simplified in their atmospheric chemistry!** Not all models have interactive ozone for example. And have different aerosol schemes.

L1070: “are persistent across” → “persist across”.

L1071: “Ocean processes are a significant source of persistent CMIP6 bias because they control long-term heat and carbon uptake (Fig. 6).” → The reasoning is incomplete as this is written. The ocean processes are not a significant source **because** they control long-term heat and carbon uptake. The fact that they play a large role in the planetary heat and carbon budget does not necessarily mean that they will cause a substantial amount of model biases.

You might want to reword to “*Given that the ocean plays a large role in the transient climate response via its control on planetary heat and carbon uptake, biases in ocean processes that persist across model generations can propagate into other realms of the climate system*”

L1073: “, which regulate heat and salt.” → This feels incomplete. Which regulates heat and salt transport? Is that what you mean?

L1076: “thereby affecting **water mass formation**, including deep-water formation and the strength of the global overturning circulation, **which influences the ocean’s capacity to sequester heat and carbon.**” ← See text suggestions. Most of the anthropogenic heat and carbon is going in within water masses that we find in the upper ocean and at intermediate depths (AAIW and mode waters; these are critical for heat and carbon storage).

L1077: “AMOC” is not a “climate mode”. **Please reword for accuracy.**

L1078: “Systematic biases in sea ice area (SIA) in both the Arctic and Antarctic arise from the spatiotemporal distribution of ice cover, and the response of sea ice to anthropogenic warming.” → **Please reword for accuracy These are not what CAUSE the systematic bias.**

L1083: I suggest transitioning to a new paragraph when you transition to the discussion of land biases.

L1093: “permafronst” → typo.

L 1128: “affect whether warming stabilizes, continues, or reverses.” → “whether current and future warming ...”

L1130: “long-term committed **climate** warming”.

L1133: “or the most relevant diagnostics of the global climate system,” → I would lighten the wording here as this is subjective, one can argue that other diagnostics are more relevant.

L1146 – 47: “Where observations are available, there is not necessarily high confidence surrounding the bias as some diagnostics have uncertainty in their reference datasets.”

→ this feels vague, can you provide an example?

L1186 – 88: “However, persistent biases in key modes of variability such as ENSO and AMOC are unlikely to be fully resolved, as they emerge from coupled interactions across scales rather than a single model component.” → **AMOC is not a “mode of variability”**.

L1195: “We also anticipate substantial progress in cryosphere and Earth system coupling. Improved sea dynamics and thermodynamics, together with more realistic ice-sheet-ocean interactions, are expected to reduce biases in sea ice area and freshwater fluxes (Trivedi et al., 2025).” → It would be really useful for the authors to provide an indication of how many CMIP7 modeling centers are planning to submit simulations with ice-sheet coupling.

L1215: “This review synthesises current understanding of systematic biases across the Earth system,” → I would rephrase to “current understanding of key systematic ...” as it is not fully comprehensive.

Section 5: Since this paper is targeting toward end-users who may not be as familiar with CMIP / coupled modeling. It would be **really useful** to end on a positive note here, regardless of systematic biases (and we have improved in key areas!!) climate models are still useful tools we just need to choose how to use their output appropriately.