The manuscript investigates the relationship between open-ocean polynyas (OOPs) and deep convection across 49 CMIP6 pre-industrial control simulations. The topic is important, and the manuscript is generally clearly written. However, I am not quite convinced that the present analysis provides sufficiently new physical understanding or that the monthly results support the event-scale and mechanistic conclusions.

My main concerns are as follows.

1. I do not see a sufficiently clear difference between this study and Mohrmann et al. (2021). That study already examined Southern Ocean polynyas across 27 CMIP6 models, including their spatial distribution, frequency, seasonality, hydrography and relationship with deep convection. It also concluded that deep convection is closely linked to OOPs but is not sufficient for their occurrence, assessed the influence of model components and resolution, and compared the results with satellite and float observations. The present study includes more models and longer piControl simulations, but most models from the earlier study are included again. The main new element appears to be the classification into "deep" and "shallow" OOPs based on whether an OOP overlaps with MLD exceeding 2000 m. I am not sure what new physical understanding this classification provides beyond the previous conclusion that some OOPs are associated with deep convection while others are not. I would like the authors to explain more clearly which unresolved question from Mohrmann et al. (2021) is addressed and what qualitatively new result is obtained. This is particularly important because Mohrmann et al. used daily output where available and showed that monthly averaging underestimates polynya area, whereas the present manuscript uses monthly data to make stronger claims about event coupling and mechanisms.

2. I am particularly concerned about the use of monthly data. OOP formation, sea ice divergence and mixed-layer deepening occur over days to weeks. September means cannot show whether an OOP and deep convection actually occurred at the same time, which process occurred first, or whether the opening was caused by ocean heat or wind-driven divergence. A short OOP may disappear through monthly averaging, while low monthly SIC does not necessarily represent a persistent open-water event. Similarly, monthly temperature and salinity may smooth or completely miss short convective episodes. Daily CMIP6 sea ice output is available for many models. I do not think that the authors necessarily need to analyse daily data for the complete 300-500 years, but daily fields from representative models and selected events are needed to validate the monthly classification and examine the lead–lag relationship. Otherwise, the conclusions should be limited to monthly spatial associations rather than event coupling or process attribution.

3. I am also not clear whether the analysis uses the archived mlotst or MLD reconstructed from monthly temperature and salinity. The authors acknowledge substantial differences between these estimates in several models, and Appendix A indicates that the calculation method can reduce or even eliminate some diagnosed deep convection regions. This is important because the 2000 m threshold controls both the detection of convection and the classification of OOPs as deep or shallow. I would therefore like to see how sensitive the results are to the source of MLD and the selected threshold. A consistent MLD definition should be used where possible, with comparisons between archived and reconstructed MLD and alternative thresholds.

4. The fixed 2000 m criterion also interacts directly with bathymetry. It cannot be satisfied where the water column is shallower than 2000 m. Deep-convection detection is therefore structurally suppressed over shelves, near continental margins and over shallow seamounts, including topographically favoured regions such as Maud Rise. Differences in model bathymetry and horizontal resolution may consequently affect both the total convection area and the deep/shallow OOP classification. The authors should check whether the identified OOP cells are limited by bathymetry and whether a depth-relative or bottom-reaching criterion changes the results. More generally, I do not think that the present deep/shallow classification is yet process based. “Deep OOPs” are defined by MLD exceeding 2000 m, so the result that they have deeper MLD is expected from the definition. Likewise, an OOP without September-mean MLD exceeding 2000 m cannot automatically be interpreted as wind driven. It could result from shallower mixing, lateral ocean heat transport, bathymetric limitation, different timing between convection and ice opening, or monthly averaging. I think the mechanistic interpretation should be softened unless it can be supported by higher-frequency diagnostics.

4. I do not understand why the analysis is restricted to September. The justification appears to be that Southern Ocean sea ice cover is generally near its annual maximum during this month. However, maximum sea ice extent does not necessarily coincide with maximum OOP activity or the deepest convection. OOPs can open, expand and close at different times between August and November, and their seasonal timing can differ among regions and models. Using September alone may therefore miss events occurring earlier or later in winter. It may also classify an OOP and convection as disconnected simply because their peaks occur in different months. This could directly affect the diagnosed overlap and the division into deep and shallow OOPs. At minimum, I think the authors should repeat the analysis across the winter season and show how sensitive the results are to the selected month.

Specific comments

1. Line 250 says OOPs in 44 models, whereas line 553 reports 40 models. Please check these numbers.
2. The statement that deep convection is a “necessary condition” for OOP development contradicts the subsequent identification of OOPs without deep convection.
3. The latent/sensible classification around Lines 378-380 appears to be reversed. Coastal polynyas are generally associated with latent-heat processes, whereas open-ocean polynyas are commonly associated with sensible ocean heat.