The referee's comments below are in italics, our answer in plain font in blue

This is an important data paper describing a valuable product that enriches the existing toolkit. The main question is whether we need another dataset if it is generated using a different model but a more or less similar method. In other words, how much can this new dataset add to the current toolkit? I think it does offer something new, especially the nudging experiment. Although some issues related to the nudging approach are not yet fully understood, I consider it the main contribution of this new dataset. This is also a very challenging task because uncertainties from different sources, including psm, climate models, and original records, will all accumulate and leave a strong imprint on the final product. It is difficult to determine which source degrades the quality of the dataset. Users often have little choice but to use all data they can obtain. This is a common situation in current paleoclimate research and probably reflects the nature of reconstruction work itself. Another difficulty is that we do not know the truth. At present, perhaps the best situation we can expect is that a new dataset falls somewhere within the range of estimates from existing datasets. Based on this paper and other recent work on similar topics, this seems to be how the field is moving forward.

First of all, we thank the reviewer for their thorough read of our manuscript and the overall positive comment on our work.

This study is part of a project aiming to produce a fully coupled ocean–atmosphere reanalysis of the past centuries using a state-of-the-art Earth system model. The first step was to equip NorCPM, which has so far been used for instrumental reanalyses and climate predictions, with the ability to assimilate paleoclimate records. This paper therefore constitutes the foundation towards a fully dynamical coupled paleo reanalysis, and describes in detail the methods and the adaptations made to the system. As noted by the reviewer, such a detailed description is particularly important given that uncertainties arise from many different sources. Following the wind-nudging approach used to initialise the atmosphere in NorCPM predictions (Garcia-Oliva et al., 2024), here applied on multi-centennial timescales, our first step towards a coupled reanalysis was to reconstruct the atmosphere, in particular the historical changes in the atmospheric circulation, and to use it to constrain the coupled model through wind nudging. To generate the target winds, we use an offline approach, since the benefit of an online method is limited by the short predictability of the atmosphere at the annual assimilation timescale. Producing the offline reanalysis with NorESM itself ensures consistency between the reconstructed fields and the model used in the wind-nudged simulations.

Although our system and methodological choices differ from existing products, we agree that an offline reanalysis alone would add limited value. This is why the wind-nudged simulations, which represent a first step towards a fully coupled reanalysis, are presented as the main contribution of this work. In addition, paleo-NorCPM1 extends a modelling framework that is already used for instrumental reanalyses and predictions, providing a consistent continuum from paleo (with foreseen deep past applications) to present-day reconstructions and future predictions within the same modelling system.

Finally, we agree that the absence of a true reference and the growing number of paleo reanalyses over the Common Era raise important questions about how these products differ

and how robust they are. To date, no systematic intercomparison of the available products exists. To address this, a new PAGES working group, CFRAME (Climate Field Reconstructions and Methods Intercomparison; https://pastglobalchanges.org/cframe), has recently been launched, in which the first author is involved. It aims to systematically compare data-assimilation and statistical reconstruction methods, assess their robustness and identify sources of uncertainty. Paleo-NorCPM1 will contribute to this effort.

I do not have any major concerns about the paper. Given its scope, depth, and the details required to explain how the dataset was generated, the authors did a very good job of including the necessary information (so it is a very long paper actually). The work mainly includes two parts, the methodology and the evaluation. I cannot comment too much on the methodology, but I think the nudging component is the most novel part of the study. The authors may need to discuss more clearly what they plan to do along this direction in the future. From what I can see, the other method, offline DA, is already quite well developed. No matter how it is redone or retested, it may not lead to a major improvement. The nudging approach may have more room for further development. The evaluation was also conducted well, although it is difficult because no true ground truth is available.

We thank the reviewer for this positive assessment of our work. We agree that the wind-nudging simulations constitute the main novelty of this study. To our knowledge, this is the first time that the atmospheric circulation reconstructed by a paleo reanalysis has been used to constrain a fully coupled Earth system model over several centuries, allowing the ocean to respond dynamically to the reconstructed circulation while preserving atmosphere–ocean feedbacks.

Our next step is to implement spectral nudging (as discussed later in this document). Since the information provided by the proxies lies mainly in the large-scale circulation, constraining only the large scales would reduce the constraint imposed on the coupled system, remove the need for artificially generated sub-monthly winds, and leave more freedom for the thermodynamic feedbacks to operate. In the longer term, the wind-nudging framework provides a basis for a fully coupled ocean–atmosphere paleo reanalysis, in which oceanic proxy records, such as corals, are assimilated directly so that the ocean state is updated at each assimilation step as already done with instrumental ocean observations in NorCPM (Garcia-Oliva et al., 2024). These perspectives will be described more clearly in the conclusions of the revised manuscript.

The most valuable part of this paper is in lines 420–490. The other parts basically represent what we already know or what has been discussed in LMR and other reconstruction papers. I still question how reliable the wind fields generated by offline DA are when they are further used to force a fully coupled model. This is an interesting step, but it seems that monthly winds +artificially generated 6-h winds+ q-flux are all used. The authors go over these steps very quickly without discussing the advantages and disadvantages of each one. For example, are the artificially generated 6-h winds too strong or too weak? Should these 6h winds also reflect some changes in the monthly wind? I also wonder why the authors do not nudge u, v, &t above the PBL if the goal is to produce realistic climate changes. All these issues should be discussed extensively because the ocean response to realistic daily wind variability can be very different from its response to artificially generated 6-h winds. How are freshwater fluxes, particularly precipitation, related to these wind patterns? Freshwater

forcing may be critical for explaining some of the biases that are now probably compensated for by the q-flux. This may be especially important over the Southern Ocean, where wind stress curl/fresh water can play an important role in determining the ocean circulation and thermal structure.

We are pleased that the reviewer considers the wind-nudging experiments as the most valuable part of the manuscript. Regarding the reliability of the reconstructed winds, paleo-NorCPM1 offline shows significant skill compared with ERA5 over 1950–2000, particularly for the zonal winds and in regions where proxy records are available (Fig. R1). Since there are no proxy records directly reflecting the winds, the skill for the winds therefore arises from the model covariances between the assimilated variables and the wind fields. In this context, the agreement with ERA5 is remarkably good. Negative correlations with ERA5 are found only in a few regions. For the zonal winds, this is the case in the southern tropical Atlantic, where almost no records are available, so the reconstruction depends heavily on the model covariances, and therefore on the model biases, which are known to be large in this region (Richter and Tokinaga, 2020). Low correlations are also found in the eastern tropical Pacific, although proxy records are available there. We attribute this mainly to model biases in the coupling between the winds and the sea surface conditions, a common bias in CMIP models (Wills et al., 2022). Importantly, the regions with low correlations are associated with an ensemble spread that covers the reconstruction errors. Figure R2 shows the reliability of the 850-hPa zonal and meridional winds over 1950–2000, computed from annual means as $1 - \mathrm{RMSE} / \sqrt{\sigma^2_{ens} + \sigma^2_{obs}}$, where RMSE is the root-mean-square error of the ensemble mean with respect to ERA5, σ^2_{ens} the time-averaged ensemble variance, and σ_{obs} the uncertainty of ERA5. As this uncertainty is not known, we consider three values: 0, 10 % and 20 % of the temporal standard deviation of ERA5. A reliability close to zero indicates that the ensemble spread is consistent with the actual errors, positive values indicate an overdispersive (conservative) ensemble, and negative values an underdispersive (overconfident) ensemble. The reliability is positive almost everywhere, including in regions with low correlations, and the results are almost insensitive to the assumed uncertainty of ERA5. The ensemble is thus generally overdispersive, which is consistent with the conservative design of the system and with the ensemble calibration ratios generally below one in Table 3. An overdispersive ensemble is generally preferable to an underdispersive one, as it avoids overconfident estimates, which is particularly important here since the reconstructed winds are subsequently used to constrain the coupled model. Overdispersion is also expected to be more pronounced for variables that are not directly constrained by the proxies, such as the winds, which are only updated through their covariances with the assimilated variables, so that the reduction of the ensemble spread remains limited. Negative values are only found in the eastern tropical Pacific, particularly for the meridional winds, where the low correlations, combined with a relatively small spread, likely reflect the model bias in the coupling between winds and sea surface conditions discussed above. A detailed evaluation of the reconstructed winds will be included in the revised manuscript.

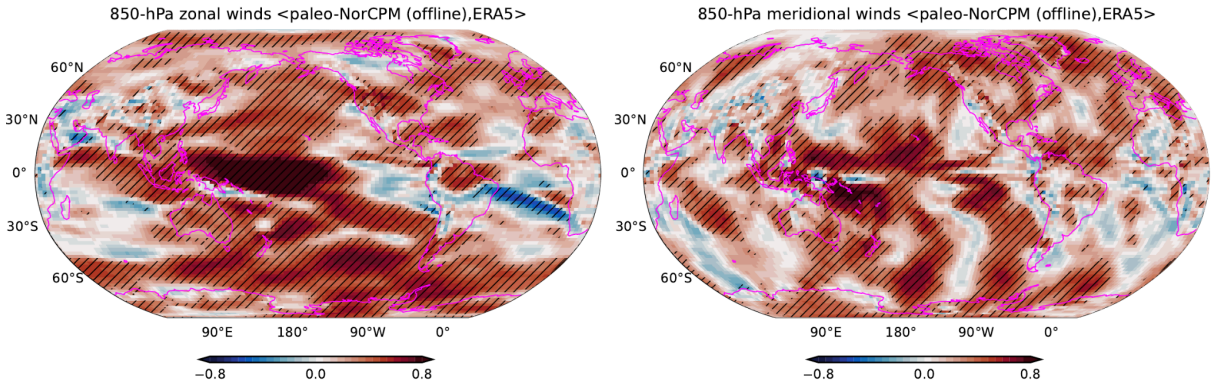


Figure R1 – Correlation coefficients between paleo-NorCPM1 (ensemble mean) and ERA5 for the 850-hPa zonal and meridional winds over the 1950–2000 period (annual averages). Hatching indicates statistically significant positive correlations (95% level).

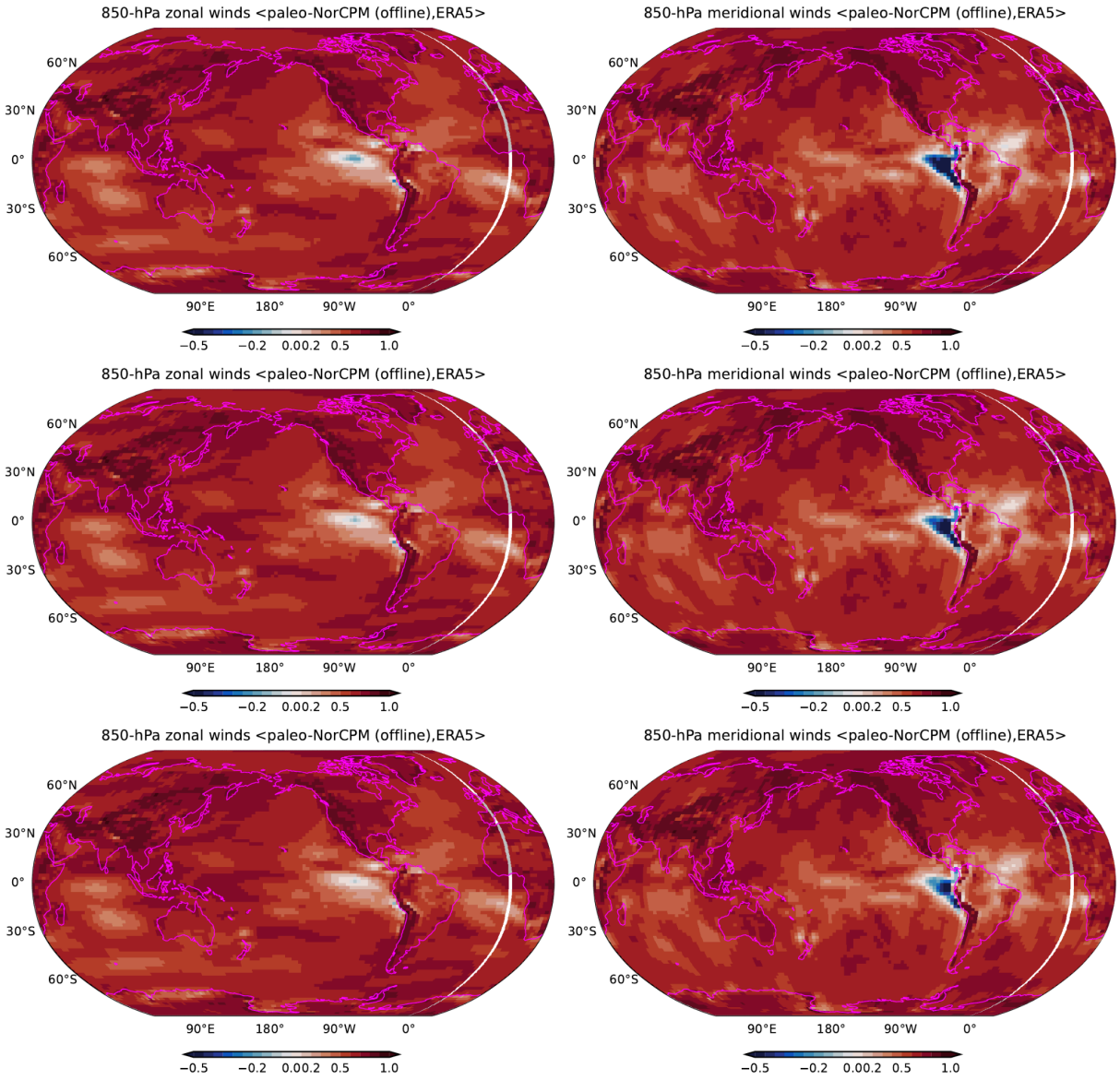


Figure R2 – Reliability of paleo-NorCPM1 offline with respect to ERA5 for the 850-hPa zonal (left) and meridional (right) winds over 1950–2000 (annual means). The uncertainty of ERA5

is assumed to be zero (top row), and 10 % (middle row) and 20 % (bottom row) of the temporal standard deviation of ERA5.

As for the wind nudging setup, we thank the reviewer for this detailed and constructive comment. We agree that the description was too brief in the original manuscript. Each element of the setup results from deliberate choices that were tested during the development of the framework. We will substantially expand Sect. 3 to explain the rationale, advantages and limitations of each of them, based on the following points.

Paleo-NorCPM1 offline provides winds at monthly resolution only. Using a twin experiment setup, we first tried to nudge the model directly toward the monthly winds with restoring timescales of one month, 10 days and one week. All of these configurations gave weak synchronisation with the target (e.g., atmospheric and ocean tropical modes) and a growing time lag, consistent with previous studies using CAM (Kruse et al., 2022). Effective synchronisation requires a short restoring timescale, typically 6 hours. This is the reason sub-monthly variability is added to the monthly target fields. The sub-monthly variability is taken from a 30-year NorESM1-ME pre-industrial control simulation, from which the monthly mean is removed, following the approach used to force ocean models (Naughten et al., 2018; Dalaiden et al., 2025). By construction, its amplitude matches the own synoptic variability from the model. We fully agree with the reviewer that this high-frequency variability is not necessarily consistent with the reconstructed monthly circulation. This will be further detailed in the updated manuscript. Because the proxies provide no information on the actual sub-monthly variability, following a probabilistic approach, we treat it as a source of uncertainty and sample it across the ensemble: each member uses a different starting year within the 30-year segment, and therefore a different realisation of the high-frequency variability. This prevents the reconstruction from depending on a single, arbitrary realisation that may be inconsistent with the monthly state. The ensemble spread thus includes the uncertainty arising from the unknown synoptic variability, while the ensemble-mean response is mainly constrained by the reconstructed monthly winds. We assessed this approach in a twin experiment (Fig. S5), in which six members are nudged toward the monthly winds of a reference NorESM1 simulation, each combined with a different realisation of the high-frequency variability (and starting from different oceanic initial conditions). The nudged members closely reproduce the monthly variability of the reference simulation for the main atmospheric modes (SAM, Amundsen Sea Low, NAO, NPO and East Atlantic Pattern), as well as for the coupled ocean–atmosphere modes, including Niño3.4, the PDO, the IPO and the IOD. The good agreement across members shows that the response is constrained by the monthly winds and is not sensitive to the particular realisation of the high-frequency variability. However, we acknowledge that the lack of covariance between high-frequency and monthly variability remains a limitation, which motivates the development of spectral nudging (see below).

Regarding nudging thermodynamical variables. Only nudging meridional and zonal winds is a deliberate choice. The main motivation of the wind-nudging framework is to let the ocean respond dynamically to the reconstructed circulation while keeping its ability to feed back on the atmosphere, through changes in the thermodynamics. Nudging temperature (and humidity) in addition to the winds would strongly constrain the surface heat and freshwater fluxes (e.g., precipitation as mentioned by the reviewer), bringing the configuration close to an ocean model forced by prescribed atmospheric conditions, as in Dalaiden et al. (2025),

and removing the coupled feedbacks that we aim to represent. Therefore, constraining only the atmospheric circulation lets the thermodynamic state evolve through atmosphere-ocean interactions, which is desired in our case. In the same spirit, the boundary layer is excluded from the nudging so that air–sea fluxes can adjust freely. We acknowledge that this freedom comes at a cost. Because the thermodynamic state and the surface fluxes are not directly constrained, the coupled system can develop enhanced variability and some undesired behaviours, such as the episodic deep convection events in the Southern Ocean. This trade-off between physical freedom and constraint is inherent to the approach, and we will discuss it explicitly in the revised manuscript.

As for the ad hoc heat-flux correction, the wind-nudged simulations exhibit a slow ocean cooling, predominantly in the deep layers. This cooling also appears when nudging toward the winds of a free NorESM1 simulation (twin experiment), showing that it is inherent to the nudging procedure rather than to the high-frequency reconstructed wind fields. As mentioned in the original manuscript, a similar drift has recently been reported in wind-nudged CESM simulations (Gilbert et al., 2025). As in forced ocean simulations, where the atmospheric forcing does not respond to the simulated ocean, the target wind fields are prescribed and do not adjust to the evolving sea surface conditions. The nudging therefore continuously pulls the atmospheric circulation away from the own dynamics of the model, while temperature and humidity evolve freely, so that the coupled system never reaches its own equilibrium, resulting in a small net heat loss from the ocean. However, identifying a clear mechanism proved highly challenging despite considerable effort, as the imbalance is very small compared with the individual surface flux terms and their respective internal variability. Because this drift could strongly affect the simulated state after a few centuries, we apply a small (0.22 W m^{-2}), spatially uniform and constant heat-flux correction estimated from the twin experiment (lines 478–488). Such corrections (e.g., salinity restoring) are common in forced ocean simulations to compensate for imperfect forcing or missing feedback. Here we restrict it to a uniform and constant heat flux to avoid damping the coupled variability. However, we recognise that this correction is far from perfect as it compensates for the effect of the drift on the global heat budget but may not address its root cause, and in particular not the associated changes in the surface freshwater fluxes. This will be clearly stated in the revised manuscript.

Because only the winds are nudged, precipitation, evaporation and sea-ice freshwater fluxes are not directly constrained, but evolve as the model response to the imposed circulation. They are linked to the winds through the position and intensity of the storm tracks, which control precipitation, the wind-speed dependence of evaporation, and wind-driven sea-ice changes. We agree with the reviewer that freshwater forcing is likely an important contributor to the biases, especially in the Southern Ocean, where stratification is mainly controlled by salinity. Consistent with this, precipitation is less well reconstructed in paleo-NorCPM1 wind nudging than in paleo-NorCPM1 offline (Fig. S7), and the wind-nudged simulations show pronounced changes in the mixed-layer depth in the Southern Ocean (Fig. S12), suggesting changes in the ocean density structure. We also considered applying a freshwater correction. However, a spatially dependent correction, or restoring sea-surface salinity as commonly done in forced ocean models, would damp salinity anomalies and the associated feedbacks, and thus counteract the coupled response that the framework aims to represent.

As discussed, identifying a single root cause of the drift is far from trivial, and we are not convinced that such a single cause exists. Rather, we interpret the drift as the result of a persistent imbalance inherent to the nudging approach. Because the atmospheric circulation is constrained, it cannot adjust to the evolving ocean state, for instance to changes in sea surface temperature, as it would in a free coupled simulation. This imbalance is intrinsic to any nudging of the circulation in a coupled model and cannot be fully removed, but it can be reduced by limiting the constraint to the large-scale information, and by leaving the model to generate its own synoptic variability, consistent with the large-scale atmospheric circulation information and its thermodynamic state (including the evolving ocean conditions). The current grid-point nudging constrains the winds at all spatial scales. Therefore, the natural next step is spectral nudging (e.g., Feser et al., 2025), which is particular relevant in our case since the information provided by the paleo reanalysis lies mainly in the large-scale circulation, as reflected in the skill for the major modes of variability (Table 3). This would remove the need for artificially generated sub-monthly winds and should reduce the imbalance, and thereby the drift and the associated biases, as much as possible. This limitation and the planned developments will be discussed explicitly in the revised manuscript.

When introducing the nudging approach, the authors should cite the original work that implemented this method in CAM5. That paper discusses in detail how the nudging method was implemented in CAM5. Many follow-up studies have not given enough credit to this relatively early and original application. https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2012JD018588

We thank the reviewer for pointing out the original work on the implementation of the atmospheric nudging in CAM. We will cite this paper when introducing atmospheric nudging.

The literature review in the intro is quite narrow and does not include much recent work. I feel that the authors should make effort to improve this part. Many relevant studies have been published, and previous work should be properly recognized and credited.

We thank the reviewer for this comment and agree that the literature review in the introduction should be widened. In the revised manuscript, we will substantially extend the introduction to better acknowledge recent developments in paleoclimate data assimilation, in particular, the online paleoclimate data assimilation using statistical emulators of climate dynamics: the recent seasonal-resolution reanalysis of LMR (Meng et al., 2025), the emerging use of machine learning in paleoclimate data assimilation, in particular deep-learning-based emulators used as forecast models in online frameworks (Meng and Hakim, 2024; Sun et al., 2025).