

## Response to Editor

In addition to the reviewers' comments, I have several additional suggestions for further improving this manuscript.

We thank the reviewer for his/her valuable and insightful comments and suggestions that help improve the manuscript. Following are the point-to-point replies to the comments (blue indicates original comment, and black indicates our reply).

1. I strongly suggest performing a thorough proofreading of the manuscript, as there are numerous grammatical errors throughout. Examples include: (L45) one of the most prominent "modes" of interannual variability, (L54) "overly regular ENSO oscillation"?, (L134) "uses", (L211) the first term "are", among many others not listed here.

**Reply:** We sincerely thank the editor for the careful reading and for highlighting these grammatical issues. We have thoroughly proofread the entire manuscript and corrected all identified errors. The specific corrections for the examples noted by the editor are as follows:

(L45) "one of the most prominent interannual variabilities" → "one of the most prominent modes of interannual variability"

(L54) "overly regular ENSO oscillation" → "an overly regular ENSO oscillation"

(L134) "the model use hybrid coordinates" → "the model uses hybrid coordinates"

(L211) "the first term on the right-hand side are the damping process" → "the first term on the right-hand side is the damping process"

In addition, we have carefully reviewed the entire manuscript for similar grammatical issues and corrected them accordingly. We believe the revised manuscript has been substantially improved in terms of language quality.

2. For Figure 4, I suggest also including the observed counterparts to allow for a direct validation of the model results.

**Reply:** We thank the editor for this valuable suggestion. In the revised manuscript, we have added the reanalysis counterpart (ORAS5) to Figure 4 to allow for a direct comparison with the model results. The reanalysis bars are displayed alongside the f3-L, f3-H, and their difference (f3-L minus f3-H), enabling readers to assess the model performance against the reanalysis benchmark. A detailed discussion of this point has been added to section 4.1 of the revised manuscript.

3. While the source code related to the diagnostics is provided in the data archive (<https://zenodo.org/records/17778266>), it needs to be well-documented with README files. Specifically, explicit step-by-step instructions for the calculations and figure plotting, along with the sample data, must be provided for each figure shown in the manuscript. This will ensure that readers can reproduce the results of this study and easily apply the approach to similar analyses. Furthermore, the current organization of the data structure (e.g., "BJ index", "TC Detection") could be improved, for instance, by sorting the files/folders according to the figure numbers in the paper.

**Reply:** We thank the editor for this valuable suggestion, which has significantly improved the transparency and reproducibility of our work. The files and folders have been reorganized according to the figure numbers in the manuscript, making the structure clearer and easier to navigate. In addition, we have prepared the README files that provide explicit step-by-step instructions for the calculations and figure plotting, together with the data required to reproduce each figure. We believe that these revisions substantially improve the usability and accessibility of our shared resources and will facilitate both reproducibility of the present study and application of the diagnostic framework to similar analyses. The updated source code and analysis data are now available at Zenodo: <https://doi.org/10.5281/zenodo.19552337>

## Response to Reviewer #1

This manuscript presents a process-based evaluation of ENSO simulation sensitivity to horizontal resolution in the CAS FGOALS-f3 climate system model. By comparing low-resolution (~100 km) and high-resolution (~25 km) configurations, the authors diagnose differences in ENSO amplitude, oscillation regularity, and underlying air–sea feedback processes using a reproducible framework including BJ index decomposition and high-frequency wind diagnostics. The study is well structured, methodologically transparent, and aligns with the scope of *Geoscientific Model Development*, particularly under the “Model Evaluation Papers” category. The process-oriented approach and the explicit tracing of resolution-sensitive feedback pathways are meaningful for both model developers and modeler users. However, several issues still need clarification or strengthening before publication. Overall, I find the manuscript suitable for publication after minor revision. Below I outline specific comments and suggestions.

We thank the reviewer for his/her valuable and insightful comments and suggestions that help improve the manuscript. Following are the point-to-point replies to the comments (blue indicates original comment, and black indicates our reply).

1. One of the central arguments follows the logical chain “TC->HF westerlies->stochastic forcing-> ENSO irregularity”, which is physically plausible and well motivated. However, the manuscript does not quantify the relative magnitude of HF wind variance versus ENSO growth rate. It would be helpful to evaluate whether the stochastic forcing amplitude differs significantly relative to the linear growth rate (e.g., using a simple signal-to-noise ratio metric). Even a simple variance ratio metric or growth rate comparison would further strengthen this section.

**Reply:** We thank the reviewer for this helpful suggestion. Following this comment, we have introduced a noise-to-signal ratio (NSR) metric to quantify the relative magnitude of stochastic atmospheric forcing compared to the ENSO signal. The NSR is defined as:

$$NSR = \frac{\sigma(u'_{HF})}{\sigma(SSTA_{Ni\tilde{n}o3.4})} \quad (1)$$

where  $\sigma(u'_{HF})$  is the standard deviation of 90-day high-pass-filtered zonal wind anomalies averaged over the western equatorial Pacific (5°S–5°N, 120°E–180°), and  $\sigma(SSTA_{Ni\tilde{n}o3.4})$  denotes the standard deviation of SSTA averaged over the Niño3.4 region (5°S–5°N, 170°–120°W). A larger *NSR* indicates stronger stochastic forcing relative to

the ENSO signal, implying a greater potential for HF atmospheric activity to disrupt the regularity of the ENSO cycle.

The *NSR* values are 2.67, 1.98, and 4.73 ( $\text{m s}^{-1} \text{K}^{-1}$ ) for the observation, f3-L, and f3-H, respectively. The substantially smaller *NSR* in f3-L reflects the combination of its weaker high-frequency wind activity and stronger ENSO amplitude, confirming that the stochastic forcing in f3-L is insufficient to disrupt its overly intense ENSO oscillation. In contrast, the larger *NSR* in f3-H indicates that stronger stochastic forcing acts on a weaker ENSO signal, facilitating the irregular oscillation that more closely resembles the observation. These quantitative results further support our interpretation in the manuscript.

To respond to the reviewer's concern, the corresponding context have added in the revised manuscript.

2. The BJ framework in Section 2.3.2 should be presented more clearly to meet GMD's reproducibility standards. Specifically, every symbol should be defined explicitly, units of each term should be provided, and the areas used for the eastern and western box regions in the BJ index calculation need to be specified. The full formulation can be provided either in the main text (with complete equations) or in an Appendix with a clean, self-contained mathematical definition.

**Reply:** Thanks for the comment. We agree that a self-contained and explicit mathematical definition is crucial for GMD standards.

In the revised manuscript, we have thoroughly updated Section 2.3.2 to explicitly define all symbols, specify units for every term, and clearly state the eastern and western box regions. These changes ensure the methodology is clear and fully reproducible as requested.

3. For a GMD audience, it would be helpful to briefly discuss the computational cost increase from f3-L to f3-H and provide the implications for CMIP7 model development strategy. This would enhance model-development relevance of the manuscript.

**Reply:** We thank the reviewer for this constructive comment, which enhances the practical relevance of this study for the model development community. We have added a brief discussion of computational costs and implications for future model development in the last section of the revised manuscript. The key information is summarized below.

The computational cost increases substantially from f3-L to f3-H. The low-resolution version (f3-L) runs on 384 processor cores and achieves a throughput of approximately 15–20 model years per wall-clock day, whereas the high-resolution version (f3-H) requires 6,144 processor cores and achieves only ~0.25 model years per wall-clock day. This represents a vast increase in computational cost (accounting for both the reduced throughput and the 16-fold increase in processor usage), making century-scale ensemble simulations with f3-H considerably more demanding.

This cost–benefit trade-off carries important implications for CMIP7 model development strategy. First, our results demonstrate that the atmospheric resolution of ~25 km is a critical threshold for realistically simulating TC activity and its associated HF wind forcing on ENSO. This finding lends support to the emerging development of variable-resolution modeling frameworks, such as the Model for Prediction Across Scales (MPAS) and the ICOSahedral Nonhydrostatic (ICON) model, which can selectively refine the grid over the tropical Pacific to capture these resolution-sensitive processes while maintaining coarser resolution elsewhere, thereby achieving a favorable balance between physical fidelity and computational affordability. Second, the process-based diagnostic framework employed in this study provides a model-agnostic and reproducible toolkit that can be readily applied to systematically evaluate ENSO simulation across different models and resolutions participating in CMIP7 and HighResMIP2. We encourage the community to adopt such process-oriented diagnostics as a complement to conventional statistical metrics, so that resolution-induced improvements (or degradations) in ENSO simulation can be traced back to specific physical mechanisms rather than assessed solely by outcome-based indices.

4. Consider adding a short graphical summary (schematic) figure illustrating the two key pathways:

(1) “resolution->wind stress structure->feedback->amplitude”,

(2) “resolution->TC->HF noise-> irregularity”).

Such a conceptual figure would help readers quickly grasp the paper’s main messages.

**Reply:** Thanks for the comment. As suggested, we have added a schematic figure to synthesize the two pathways discussed in the manuscript. For the reviewer’s convenience, this figure (Figure 15 in the revised manuscript) is also copied below.

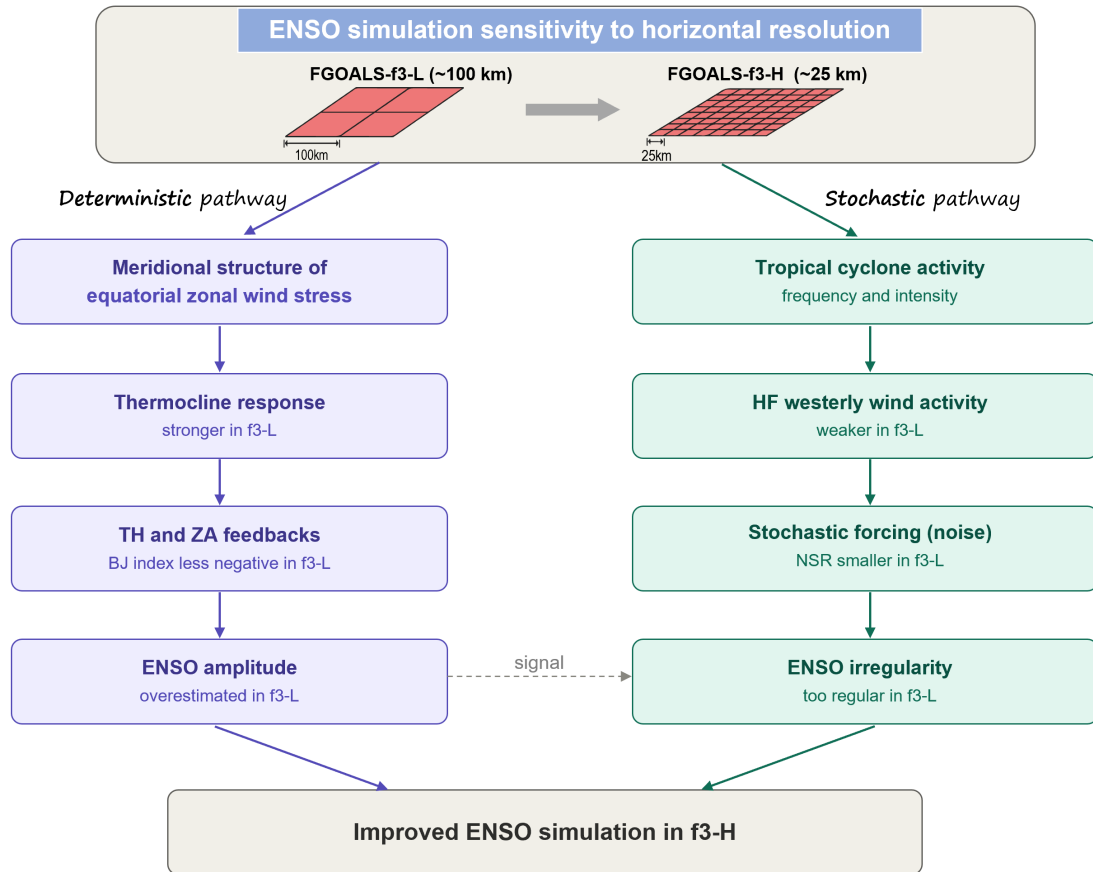


Figure A1 Schematic diagram illustrating how increased horizontal resolution (~100 km to ~25 km) improves ENSO simulation in FGOALS-f3 via both the deterministic feedback processes and the stochastic atmospheric forcing pathways.

5. Line 271-274: ENSO regularity is currently discussed mainly based on qualitative inspection of the Niño3.4 time series. It would be helpful to complement this with a simple quantitative metric of regularity (e.g., spectral peak sharpness/width, autocorrelation-based periodicity, coefficient of variation of event intervals, or an “irregularity index”). This would make the comparison more objective.

**Reply:** Thanks for this valuable suggestion. As suggested, we have proposed an ENSO irregularity index to quantitatively evaluate the irregularity of ENSO oscillation. This metric is based on the coefficient of variation of inter-event time intervals (CVT), and is computed through the following steps.

First, ENSO events are identified using the 3-month running mean of the Niño3.4 index (monthly SSTA averaged in 5°S–5°N, 170°–120°W). A warm (cold) event is defined when the 3-month running mean Niño3.4 index exceeds 0.5 (falls below –0.5) standard deviation of the Niño3.4 index, and the event is considered to terminate when the

Niño3.4 index returns to within the  $\pm 0.5$  standard deviation range for at two consecutive months.

Second, the time interval between two successive events of the same sign is defined as the time separation between adjacent event peaks (i.e., the month of maximum warming for warm events or maximum cooling for cold events).

Third, the CVT is computed as the ratio of the standard deviation to the mean of these inter-event intervals:

$$CVT = \frac{\sigma_T}{\mu_T} \quad (1)$$

where  $T$  denotes the set of all inter-event intervals, and  $\mu_T$  and  $\sigma_T$  denote the mean and the standard deviation of these intervals, respectively.

Finally, the CVT is calculate separately for warm events ( $CVT_{warm}$ ) and cold events ( $CVT_{cold}$ ), and their average is taken as the final ENSO irregularity index used in this study. A larger  $CVT$  indicates more irregular ENSO oscillation with highly variable inter-event spacing, whereas a smaller  $CVT$  (approaching zero) indicates a more periodic and regular oscillation.

Based on the proposed ENSO irregularity index, we calculated the  $CVT$  for the observation, f3-L and f3-H. As shown in Table A1, the  $CVT$  values are 0.61, 0.17, and 0.53 for the observation, f3-L and f3-H, respectively. The results clearly indicate that ENSO oscillation in f3-L is excessively regular compared to the observation ( $CVT$  of 0.17 vs. 0.61), whereas f3-H produces a degree of irregularity much closes to the observation.

Table A1. The ENSO irregularity index (CVT) for the observation, f3-L and f3-H.

	<b>Observation</b>	<b>f3-L</b>	<b>f3-H</b>
<i>CVT</i>	0.61	0.17	0.53

In summary, these quantitative results corroborate our previous qualitative assessment that ENSO variability in f3-L is overly regular and self-sustained compared to that in f3-H and the observation. To respond to the reviewer’s concern, the corresponding analysis and description have been added into the revised manuscript.

## 6. Several typos and grammatical refinements are still needed.

6.1 Line 184-185: Consider removing the full name after the abbreviation “HF” if it has already been defined earlier.

6.2 Replace “key influencing ENSO simulation” with “a key factor influencing ENSO simulation”.

6.3 Line 134: “use” should be “uses”.

6.4 Line 160: In Table 1, the land component abbreviation should be “CLM4.0”, not “CLIM4.0”.

6.5 Line 174: “are” should be “is”.

6.6 Data availability section contains a duplicated DOI string: <https://doi.org/https://doi.org/...> Please correct it.

**Reply:** We are very grateful to the reviewer for your careful and detailed reading of this manuscript. We have corrected all of the listed errors in the revised manuscript. We appreciate your great help in improving the manuscript’s overall quality and readability.

## Response to Reviewer #2

This manuscript examines the impact of horizontal resolution on ENSO amplitude in the FGOALS-f3 model. The authors show that differences in the meridional structure of zonal wind stress anomalies lead to changes in thermocline and zonal advective feedbacks within the BJ framework, thereby modulating ENSO amplitude and regularity.

Overall, the manuscript is well-organized and mostly clear. The conclusions drawn from the analysis are generally sound and suggest important implications. However, further refinements in the presentation and result interpretation would significantly enhance its overall impact and persuasiveness. Please see the following comments:

We thank the reviewer for his/her valuable and insightful comments and suggestions that help improve the manuscript. Following are the point-to-point replies to the comments (blue indicates original comment, and black indicates our reply).

1. The manuscript contains numerous equations and symbolic representations. It would improve clarity if the notation were made more consistent throughout the paper. For example, overbars and primes are both used to denote anomalies in different places, and in some places, primes seem to indicate filtered anomalies. A clearer and more unified notation system would help readers follow the derivations more easily.

**Reply:** We thank the reviewer for this valuable suggestion. We agree that the notation in the previous manuscript was not sufficiently unified, particularly regarding the meanings of overbars, primes, and filtered anomalies.

In the revised manuscript, we have carefully standardized the notation throughout the paper as follows:

- (1) Overbars consistently denote the climatological mean (or basic-state) quantities;
- (2) Primes consistently denote the interannual anomalies obtained by removing the climatological seasonal cycle.
- (3) For the high-frequency (HF) wind diagnostics, we now explicitly use the subscript 'HF' to denote the HF-filtered anomalies (e.g.,  $u'_{HF}$ ), so as to distinguish them from the interannual anomalies used in the BJ index framework.

In addition, we have revised the text in the Methods section to define these notations explicitly when they first appear, and we have checked the equations and surrounding descriptions throughout the manuscript to ensure consistency.

For the reviewer's convenience, some notations are copied below.

*“Throughout this study, an overbar  $\overline{(\ )}$  denotes the climatological mean field, and*

*a prime ( )' denotes the interannual anomaly obtained by removing the climatological seasonal cycle. The subscript 'HF' indicates the HF (sub-90-day) filtered field. For example,  $u'_{HF}$  denotes the HF component of daily zonal wind anomaly, obtained by applying a 90-day high-pass filter to the daily anomaly field.*

*All symbols are used consistently throughout the paper unless otherwise specified.”*

We believe these revisions have substantially improved the clarity and readability of the manuscript.

2. The BJ index calculation assumes a fixed mixed layer depth of 65 m. Since model resolution may affect vertical stratification and mixed layer structure, it would be helpful if the authors could discuss the potential sensitivity of their results to this assumption. For example, would the use of model-specific mixed layer depths change the quantitative estimates of the feedback components?

**Reply:** We thank the reviewer for this insightful comment. As suggested, we have examined the sensitivity of the BJ index results to the mixed layer depth (MLD) assumption.

We first examined the spatial distribution of the climatological MLD in f3-L and f3-H. Here the MLD is defined as the depth at which the ocean temperature is 0.8°C lower than the SST, following Wang et al. (2012) and Chen et al. (2016). As shown in Figure B1, the climatological MLD exhibits a pronounced zonal variation along the equatorial Pacific: it is relatively shallow in the far eastern Pacific and gradually deepens toward the central equatorial Pacific. Moreover, the MLD differs between the two model versions, with the eastern box-mean value of approximately 65 m in f3-L and 50 m in f3-H over the eastern equatorial Pacific (i.e., the BJ-index eastern box region). Given this zonal and inter-model variability, we adopted two complementary strategies for the BJ index diagnostic framework.

**Strategy 1 (Constant MLD):** A fixed MLD of a constant value is applied to both f3-L and f3-H when computing the BJ index. This approach follows the conventional practice in previous BJ index studies (e.g., Kim and Jin, 2011a, 2011b; Chen et al., 2019a, 2019b) and facilitates a direct comparison between the two simulations under an identical diagnostic framework. From the perspective of the BJ-index eastern-box average, the climatological mean MLD over the eastern equatorial Pacific is approximately 65 m in f3-L and 50 m in f3-H. Therefore, in this first approach, we use a constant value of 65 m for both simulations.

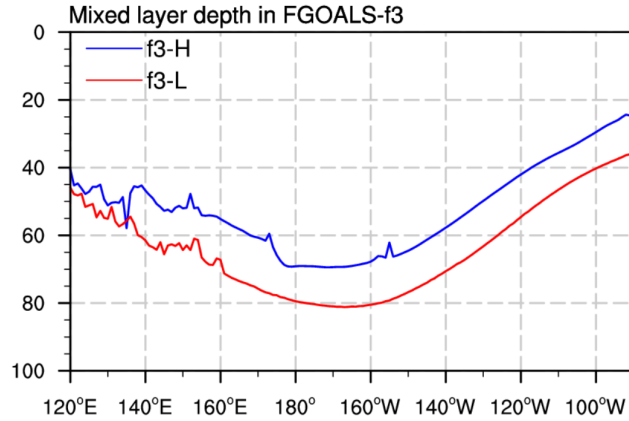
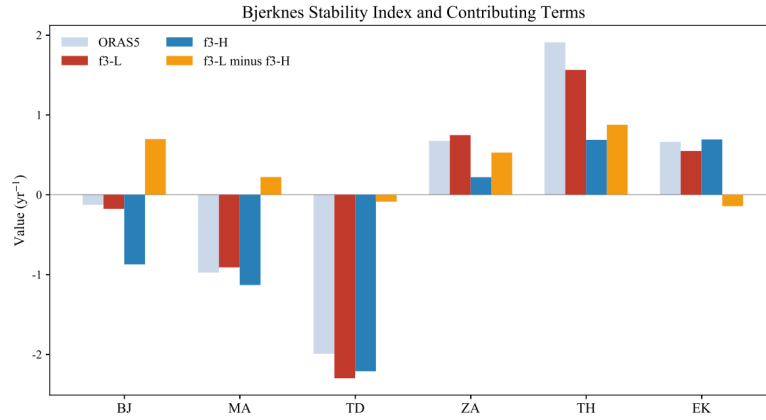


Figure B1. The longitudinally varying climatological mixed layer depth (unit: m) averaged over the equatorial Pacific (5°S–5°N) in f3-L (red line) and f3-H (blue line).

The BJ index results under Strategy 1 are shown in Figure B2. Figure B2 presents the BJ index calculated using a fixed MLD for the reanalysis, f3-L, and f3-H, as well as their difference (f3-L minus f3-H). The results demonstrate that although both models yield negative BJ indices, the value for f3-L is significantly larger (i.e., less negative) than that for f3-H. According to the physical interpretation of the BJ index (Jin et al., 2006; Kim and Jin, 2011a; 2011b), the less negative BJ index in f3-L indicates that the coupled air-sea system in f3-L is more unstable than that in f3-H. This more unstable coupled system is more favorable for ENSO growth, thereby leading to a stronger ENSO amplitude in f3-L than that in f3-H.

It is worth noting that the BJ index derived from ORAS5 is not more negative than that of f3-L, despite the observed ENSO amplitude being smaller than that simulated by f3-L. This apparent inconsistency can be attributed to two factors. First, reanalysis products carry inherent uncertainties, and direct comparison between reanalysis-derived and model-derived BJ indices should be interpreted with caution. Second, the BJ index is a linear diagnostic framework that does not account for nonlinear processes, including nonlinear atmospheric responses, semi-stochastic atmospheric noise (i.e., the high-frequency zonal wind anomalies discussed in this study), and oceanic nonlinear processes such as nonlinear dynamical heating. In other words, while the BJ index is a useful tool for assessing whether the linear air–sea coupling framework favors ENSO growth, the actual ENSO amplitude is jointly determined by linear coupling, nonlinear processes, and stochastic forcing. This represents an inherent limitation of the BJ index framework. Therefore, a comprehensive evaluation of ENSO simulation requires not only the BJ index analysis of linear feedback processes but also diagnostics beyond the linear framework to assess the roles of nonlinear processes and stochastic forcing—as is addressed in Section 5 of this study. In the following, our primary focus is on

examining the differences in the BJ index and its contributing terms between f3-L and f3-H.



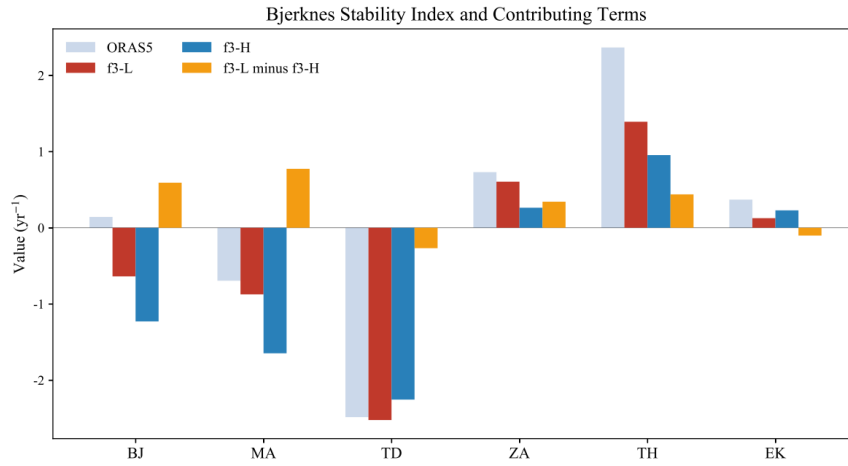
**Figure B2.** BJ index and the corresponding main contributing terms for the reanalysis (grey bars; ORAS-5), f3-L (red bars), f3-H (blue bars) and their difference (f3-L minus f3-H, orange bars). The BJ index is calculated using a fixed MLD of 65 m. The five contributing terms include dynamic damping by mean advection (MA), thermodynamic damping feedback (TD), zonal advection feedback (ZA), thermocline feedback (TH) and Ekman feedback (EK).

A further question arises: which physical processes contribute to the more unstable coupled system in f3-L? By examining the differences in the five contributing terms of the BJ index (orange bars in Fig. B2), we find that the thermocline feedback (TH) term and the zonal advection feedback (ZA) term are the decisive factors driving the BJ index difference between the two model versions. Therefore, the subsequent analysis focuses on the physical mechanisms responsible for the stronger TH and ZA terms in f3-L relative to f3-H.

**Strategy 2 (Longitude-varying MLD):** Each model version uses its own longitude-dependent climatological MLD (averaged over the equatorial band, 5°S–5°N) when computing the BJ index. This approach accounts for zonal variations and inter-model differences in stratification, providing a more physically realistic diagnostic. In this approach, when calculating the BJ index, we diagnose the mixed-layer temperature anomaly above the longitudinally varying climatological mixed layer depth (Figure B1) within the three-dimensional eastern equatorial Pacific box.

Figure B3 presents the BJ index and its contributing terms calculated using the longitude-varying MLD for the reanalysis, f3-L, and f3-H, as well as their difference (f3-L minus f3-H). The main results are broadly consistent with those obtained under Strategy 1. Specifically, the BJ index of f3-H remains more negative than that of f3-L, which largely explains the weaker ENSO amplitude in f3-H; and the BJ index difference between the two models is still primarily attributable to the TH and ZA terms.

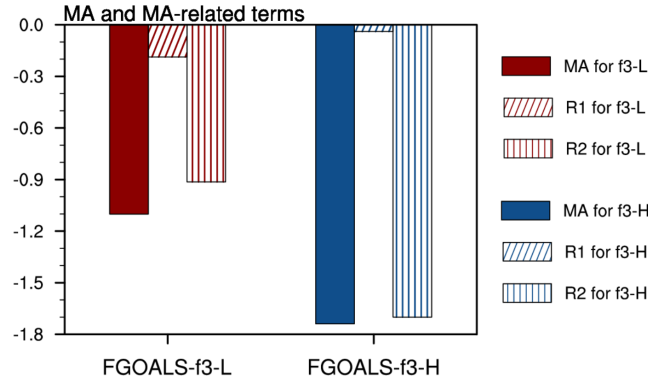
The only notable discrepancy between the two strategies lies in the mean advection (MA) term, which represents the dynamic damping by mean zonal and meridional advection. Under Strategy 2, the MA term exhibits a considerably larger difference between f3-L and f3-H than under Strategy 1. To understand this, we further decomposed the MA term into its two components: the damping by mean zonal current ( $R1 = -a_1 \frac{\langle \Delta \bar{u} \rangle_E}{L_x}$ ) and the damping by mean meridional current ( $R2 = -a_2 \frac{\langle \Delta \bar{v} \rangle_E}{L_y}$ ). As shown in Figure B4, the MA term difference primarily arises from the meridional component (R2).



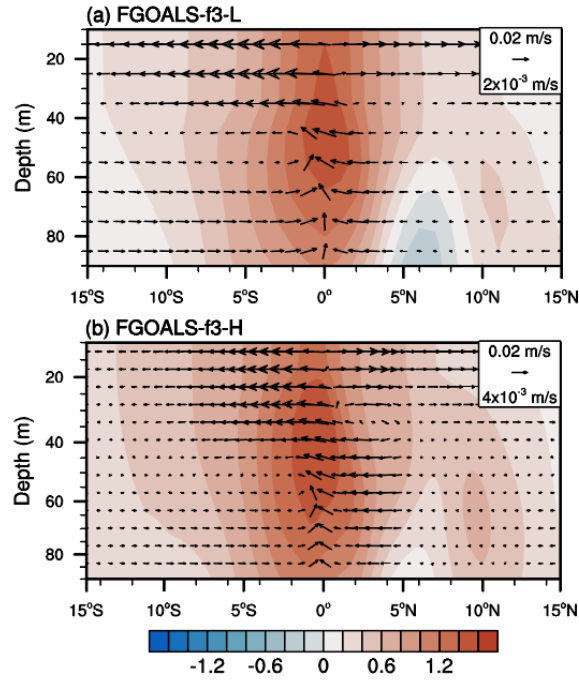
**Figure B3.** Same as Figure B2, but the BJ index is calculated using the longitude-varying MLD.

This meridional damping component is closely related to the mean subtropical cell (STC) circulation in the central-eastern equatorial Pacific. Figure B5 shows the latitude–depth distribution of the mean ocean currents and ENSO-related ocean temperature anomalies averaged over 150°W–90°W. The ENSO-related temperature anomalies are centered at the equator from the surface down to approximately 100 m, and are advected poleward by the mean meridional currents, constituting a dynamical damping effect. In both models, the mean meridional flow associated with the STC is directed poleward above approximately 40 m but equatorward below 40 m. Consequently, the advection of temperature anomalies by the mean meridional current reverses sign at approximately 40 m, so that the contributions from the upper and lower portions partially cancel each other. When a constant MLD of 65 m is used for vertical averaging, this cancellation occurs similarly in both models, yielding comparable MA terms. However, under the longitude-varying MLD strategy, the shallower MLD in f3-H means that the vertical averaging captures a thinner layer in which the poleward (damping) branch dominates, resulting in a more negative MA term in f3-H compared

to f3-L. This MLD sensitivity of the MA term partially explains why the BJ index in f3-H becomes even more negative under Strategy 2.



**Figure B4.** The MA term and its two components: dynamic damping by mean zonal current ( $R1 = -a_1 \frac{\langle \Delta \bar{u} \rangle_E}{L_x}$ ) and dynamic damping by mean meridional current ( $R2 = -a_2 \frac{\langle \Delta \bar{v} \rangle_E}{L_y}$ ), for f3-L (red) and f3-H (blue).



**Figure B5.** Oceanic mean currents (vectors, units:  $\text{m s}^{-1}$ ) and ENSO-related ocean temperature anomalies (contours, units:  $^{\circ}\text{C}$ ) averaged over  $150^{\circ}\text{W}$ – $90^{\circ}\text{W}$  for (a) f3-L and (b) f3-H. Here the ENSO-related ocean temperature anomalies are obtained by regressing the ocean temperature anomaly field onto the Niño3.4 index.

In summary, the sensitivity test demonstrates that the core conclusions of the BJ index analysis, that is namely, the more unstable coupled system in f3-L and the dominant roles of the TH and ZA terms are robust across both MLD strategies.

Jin, F. F., Kim, S. T., and Bejarano, L.: A coupled-stability index for ENSO,

Geophysical Research Letters, 33, <https://doi.org/10.1029/2006gl027221>, 2006.

Wang, L., Li, T., and Zhou, T. J.: Intraseasonal SST variability and air-sea interaction over the Kuroshio extension region during boreal summer. *Journal of Climate*, 25, 1619–1634, <https://doi.org/10.1175/JCLI-D-11-00109.1>, 2012.

Chen, L., Li, T., Behera, S. K., and Doi, T.: Distinctive precursory air–sea signals between regular and super El Niños, *Advances in Atmospheric Sciences*, 33, 996–1004, <https://doi.org/10.1007/s00376-016-5250-8>, 2016.

Kim, S. T. and Jin, F.-F.: An ENSO stability analysis. Part II: Results from the twentieth and twenty-first century simulations of the CMIP3 models, *Climate Dynamics*, 36, 1609–1627, <https://doi.org/10.1007/s00382-010-0872-5>, 2011a.

Kim, S. T. and Jin, F.-F.: An ENSO stability analysis. Part I: results from a hybrid coupled model, *Climate Dynamics*, 36, 1593–1607, <https://doi.org/10.1007/s00382-010-0796-0>, 2011b. Chen, L., Wang, L., Li, T., and Liu, J.: Drivers of reduced ENSO variability in mid-Holocene in a coupled model, *Climate Dynamics*, 52, 5999–6014, <https://doi.org/10.1007/s00382-018-4496-5>, 2019a.

Chen, L., Zheng, W., and Braconnot, P.: Towards understanding the suppressed ENSO activity during mid-Holocene in PMIP2 and PMIP3 simulations, *Climate Dynamics*, 53, 1095–1110, <https://doi.org/10.1007/s00382-019-04637-z>, 2019b.

3. The comparison between the CMIP experiments and the OMIP experiments is interesting. However, it seems that the response of thermocline depth anomalies to zonal wind stress differs between OMIP and CMIP in a resolution-dependent manner (i.e., OMIP shows a stronger response than CMIP at high resolution, but a weaker response at low resolution). What are the physical reasons for this contrasting behavior?

**Reply:** We thank the reviewer for this excellent and insightful comment. We fully agree that the resolution-dependent contrast between the OMIP experiment and the CMIP experiment (i.e., the historical experiment) is a noteworthy phenomenon that warrants explanation. We have carefully analyzed the physical reasons behind this behavior, and our findings are summarized as follows:

As established in Section 4.2, equatorial thermocline depth anomalies are primarily driven by surface wind stress forcing, and the meridional structure of the zonal wind stress anomalies ( $\tau_x'$ ) plays a critical role in determining the efficiency of this forcing.

To understand the contrasting OMIP–CMIP behavior, we compared the meridional structures of the “normalized  $\tau_x'$ ” between the CMIP and OMIP experiments for both model versions, as shown in Figure B6. Here the normalized  $\tau_x'$  is obtained by regressing the  $\tau_x'$  field onto Nino4-region-averaged  $\tau_x'$ ; and then averaging over the Nino4 longitude range (160°E–150°W). This normalization procedure enables a fair comparison of the meridional structure of  $\tau_x'$  across different experiments and model versions. In f3-L, the CMIP experiment produces stronger  $\tau_x'$  near the equator compared to the OMIP forcing, which is derived from the JRA55-do reanalysis (red lines in Fig. B6). This leads to an enhanced thermocline response in CMIP relative to OMIP. Conversely, in f3-H, the CMIP experiment yields weaker equatorial  $\tau_x'$  than the OMIP forcing (blue lines in Fig. B6), resulting in a weaker thermocline response in CMIP relative to OMIP. This explains the resolution-dependent contrast identified by the reviewer.

To further quantify these structural differences in  $\tau_x'$  between the CMIP and OMIP experiments, the meridional distribution index (*MDI*) that was proposed by Chen et al. (2015), is employed. The MDI is defined as:

$$MDI = \frac{\int_{-10^\circ}^{10^\circ} \tau_x'(y) |y| dy}{\int_{-10^\circ}^{10^\circ} \tau_x'(y) dy} \quad (7)$$

where  $y$  denotes latitude, and  $\tau_x'(y)$  represents the meridional profile of the normalized  $\tau_x'$  shown in Fig. B6. The *MDI* provides a quantitative measure of the meridional concentration of ENSO-related  $\tau_x'$  within the equatorial band. Specifically, a smaller MDI indicates that  $\tau_x'$  is more concentrated near the equator, whereas a larger MDI indicates a broader meridional distribution.

The qualitative differences in  $\tau_x'$  distributions shown in Fig. B6 are corroborated by the MDI results (Table B1). In f3-L, the CMIP experiment yields a notably smaller MDI (2.55°) than the OMIP experiment (2.68°), indicating a more equatorially concentrated  $\tau_x'$  structure that can more efficiently drive thermocline variability, and hence produces a larger  $\beta_h$  in the CMIP experiment. Conversely, in f3-H, the CMIP experiment exhibits a larger MDI (2.71°) than the OMIP experiment (2.64°), corresponding to a broader  $\tau_x'$  distribution that drives a more moderate thermocline response and a smaller  $\beta_h$ .

In summary, this resolution-dependent contrast between the CMIP and OMIP experiments further reinforces our core conclusion: the meridional structure of  $\tau_x'$ , particularly its degree of equatorial concentration, is the primary factor governing the

strength of the thermocline feedback. These results also highlight the sensitivity of ENSO-related air-sea coupling to even subtle differences in equatorial wind stress meridional distribution: a small resolution-induced change in the meridional structure of  $\tau_x'$  can produce substantially different thermocline responses, which in turn fundamentally modulate the associated air-sea coupling processes. Consequently, an accurate representation of the wind stress meridional structure is critical for improving ENSO simulations in climate models. To respond to this concern, we have incorporated the relevant analysis and the corresponding results into the revised manuscript.

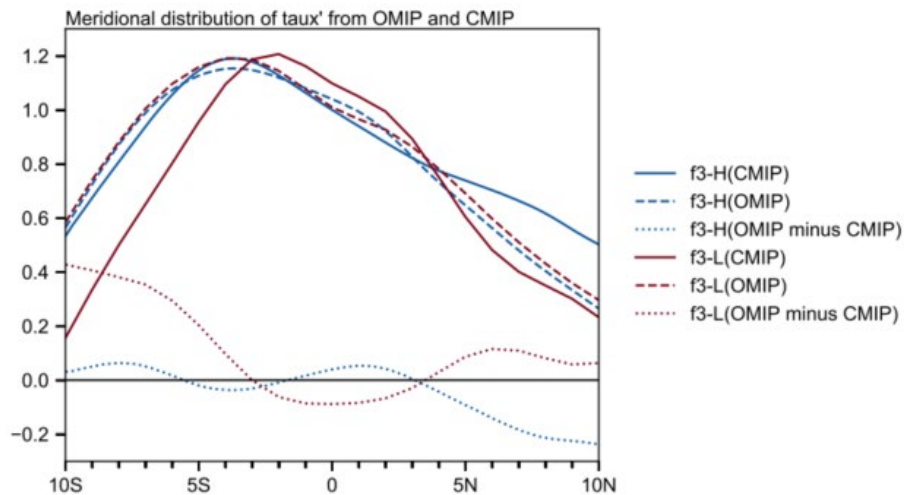


Figure B6. Meridional structure of normalized zonal wind stress anomalies [units:  $\text{N m}^{-2} (\text{N m}^{-2})^{-1}$ ] averaged over  $160^\circ\text{E}$ – $150^\circ\text{W}$  for f3-L in CMIP (red solid line), f3-L in OMIP (red dashed line), and their difference (red dotted line, OMIP minus CMIP); f3-H in CMIP (blue solid line), f3-H in OMIP (blue dashed line), and their difference (blue dotted line, OMIP minus CMIP). The normalized zonal wind stress anomalies are obtained by regressing the zonal wind stress anomaly field onto the Niño4 region ( $5^\circ\text{S}$ – $5^\circ\text{N}$ ,  $160^\circ\text{E}$ – $150^\circ\text{W}$ ) averaged zonal wind stress anomalies and then averaged over the Niño4 longitude range ( $160^\circ\text{E}$ – $150^\circ\text{W}$ ).

Table B1. Meridional distribution index (MDI; units:  $^\circ$ ) of  $\tau_x'$ , calculated from the meridional structure of normalized  $\tau_x'$  as shown in Fig. B6.

	<b>f3-L</b>	<b>f3-H</b>
CMIP	2.55	2.71
OMIP	2.68	2.64

Chen, L., Li, T., and Yu, Y. Q.: Causes of strengthening and weakening of ENSO amplitude under global warming in four CMIP5 models, *Journal of Climate*, 28, 3250–3274, <https://doi.org/10.1175/jcli-d-14-00439.1>, 2015.

4. The BJ index provides a useful linear stability framework for interpreting ENSO amplitude changes. However, given the relatively limited simulation length (~65 years)

and the analysis being based on a single model family, it would be helpful for the authors to briefly acknowledge the potential limitations of applying a linear BJ framework to interpret resolution-dependent changes in ENSO dynamics in the discussion, particularly considering the role of nonlinear and stochastic processes.

**Reply:** We thank the reviewer for this thoughtful suggestion. We fully agree that the limitations of the linear BJ framework should be explicitly acknowledged, particularly in the context of resolution-dependent ENSO dynamics.

In fact, as discussed in our response to Comment 2, we have already added a dedicated discussion of this issue in the revised manuscript (Section 4.1). Specifically, we note that the BJ index is a linear diagnostic framework that does not account for nonlinear processes, including nonlinear atmospheric responses, semi-stochastic atmospheric noise (i.e., the high-frequency zonal wind anomalies discussed in Section 5), and oceanic nonlinear processes such as nonlinear dynamical heating (Wei et al., 2026). While the BJ index is a useful tool for assessing whether the linear air–sea coupling framework favors ENSO growth, the actual ENSO amplitude is jointly determined by linear coupling, nonlinear processes, and stochastic forcing. Therefore, a comprehensive evaluation of ENSO simulation requires not only the BJ index analysis of linear feedback processes but also diagnostics beyond the linear framework—as is addressed in Section 5, where we examine the role of high-frequency stochastic wind forcing in shaping ENSO irregularity.

Regarding the relatively limited simulation length (~ 65 years), we acknowledge that this may induce sampling uncertainty in the BJ index estimates. However, the historical experiment of f3-H (highresSST-present) only provides simulation outputs covering 1950–2014, which is comparable to the analysis periods adopted in previous BJ index studies (e.g., Kim and Jin, 2011a, 2011b).

We have incorporated some relevant discussions into the revised manuscript to explicitly acknowledge these limitations. We thank the reviewer for helping us strengthen this aspect of the paper.

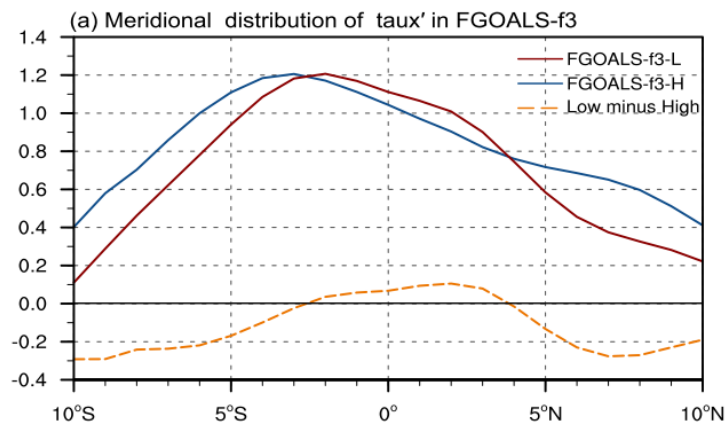
Kim, S. T. and Jin, F.-F.: An ENSO stability analysis. Part II: Results from the twentieth and twenty-first century simulations of the CMIP3 models, *Climate Dynamics*, 36, 1609–1627, <https://doi.org/10.1007/s00382-010-0872-5>, 2011a.

Kim, S. T. and Jin, F.-F.: An ENSO stability analysis. Part I: results from a hybrid coupled model, *Climate Dynamics*, 36, 1593–1607, <https://doi.org/10.1007/s00382-010-0796-0>, 2011b.

Wei, X. J. Chen, L., and Sun, M.: Fine-tuning Atmospheric Parameters for Improving ENSO Simulation in the Zebiak–Cane Model, *Advances in Atmospheric Science*, 43, 420–435, <https://doi.org/10.1007/s00376-025-4423-8>, 2026.

5. In Figure 7a, there seems to be a horizontal black line around 0.2, but it is not described in the caption. Please clarify what this line represents.

**Reply:** We thank the reviewer for pointing this out. We apologize for the confusion. The horizontal black line at 0.2 in the original Figure 7a was added solely as a visual guide to separate the difference curve (orange dashed line) from the individual model curves (red and blue solid lines) and does not represent any physical threshold. In the revised manuscript, we have removed this line and replaced it with a horizontal reference line at 0.0, which serves as baseline for assessing the sign of the resolution-induced differences. For the reviewer’s convenience, the updated plot is copied below.



**Figure B7.** Meridional structure of normalized zonal wind stress anomalies [units:  $\text{N m}^{-2} (\text{N m}^{-2})^{-1}$ ] averaged over the Niño4 longitude range ( $160^\circ\text{E}$ – $150^\circ\text{W}$ ) for f3-L (red solid line), f3-H (blue solid line) and their difference (orange dash line, f3-L minus f3-H). In this plot, all the data are interpolated onto a  $1^\circ \times 1^\circ$  grid to facilitate the comparison.

6. Lines 29-30. The sentence “The low-resolution severely overestimates ENSO amplitude” lacks a noun after “low-resolution.” Please revise (e.g., “low-resolution version”).

**Reply:** Corrected as suggested.

7. Line 38: The word “feeble” sounds somewhat informal in this context. Please consider replacing it with “weak”.

**Reply:** Done as suggested.

8. Line 422: “may be not the primary driver” contains a word order issue. Please revise to “may not be the primary driver.”

**Reply:** Corrected as suggested.

9. Line 500: “yielding a more realistic characteristics of TC activity” contains a number agreement issue. Please revise to either “yielding more realistic characteristics of TC activity” or “yielding a more realistic representation of TC activity.”

**Reply:** Corrected as suggested.