Contents

The supplementary material to this paper consists of:

1. **Individual model results**; a collection of figures presenting results from the individual PlioMIP2 ensemble members, using 100 years of data,
2. **Analysis of Niño3.4 s.d. using 500 years of data** instead of 100 years; figures and explanation on Niño3.4 s.d. variation with 500 years of data from CESM2 and MIROC4m,
3. **Spectal peak-counting robustness**; results with different settings in spectral analyses of ENSO time-series, in order to test robustness,
4. **selected_data.xlsx**; Selection of data of scatter plots presented in the main paper, specifically Figures 2, 6, 9 and 10(a), collected in an Microsoft Excel file (attached). More PlioMIP2 data is available, see the section 'Code and data availability' in the main paper for details.

1 Individual model results

For the sake of completeness, we include here figures showing several individual model results using 100 years of data. In the order of treatment in the paper, we present:

- the Niño3.4 index in Figure S1;
- the power spectra of the Niño3.4 index, including 95% confidence interval, in Figure S2;
- the leading EOF of the SST anomalies in the tropical Pacific (23°S - 23°N, 140°E - 80°W) in Figure S3;
- and the meridional mean (5°S - 5°N), zonal SSTs in the Pacific in Figure S4.

2 Analysis of Niño3.4 s.d. using 500 years of data

Previous studies from Wittenberg (2009) and Tindall et al. (2016) show that ENSO variability can exhibit significant changes on the interdecadal and centennial scale. While a time-series length of 100 years as used here is clearly sufficient to study properties such as amplitude and spectra of ENSO on its typical interannual time scales, the presence of centennial scale modulations on ENSO properties could imply that the choice of the 100-year segment affects our conclusions on interannual variability differences between pre-industrial and mid-Pliocene. To study the robustness of our ENSO amplitude analysis method with 100 years of data, we repeated the calculation of the Niño3.4 standard deviation (s.d.) with 500 years of data. For this we used data of two PlioMIP2 ensemble members, MIROC4m and CESM2. Specifically, we calculated the Niño3.4 index for the last 500 years of the pre-industrial E^{280} and mid-Pliocene Eoi^{400} simulations. The time-series are normalised and detrended as mentioned in the Methods section of the main paper. We look at how the Niño3.4 s.d. changes when using different 100 year segments, as well as the full 500 years. We expect variations, but the main question is whether the conclusions regarding the differences between the pre-industrial and mid-Pliocene ensembles change.

The Niño3.4 s.d. using the five 100 year segments as well as the full 500 years of data is shown in Figure S5 for CESM2 and Figure S6 for MIROC4m. Figure S5 (a) shows the Niño3.4 s.d. for the CESM2 E^{280} and Eoi^{400} simulations. It can be seen that the absolute values show a significant spread around the one-to-one line. Figure S5 (b) shows the relative change of Niño3.4 s.d. in the mid-Pliocene. Depending on the length of the time-series, there is either a decrease or increase of the Niño3.4 s.d. in the mid-Pliocene. Furthermore, it shows that the value of the Niño3.4 s.d. varies within a ±9% standard deviation of the value using 500 years of data, indicating the presence of centennial scale variation of ENSO amplitude in CESM2.

Figure S6 (a) shows the Niño3.4 s.d. for the MIROC4m E^{280} and Eoi^{400} simulations. It shows that the absolute values of the Niño3.4 s.d. show little change when using different 100 years segments, and that it is clear that the ENSO amplitude is reduced in the mid-Pliocene simulations. Figure S6 (b) shows the relative change of the Niño3.4 s.d. in the mid-Pliocene. The MIROC4m 100 year results show a variation of $\pm 4\%$ around the 500 year value, suggesting a much weaker centennial scale modulation of ENSO amplitude compared to CESM2. Consequently, the significant reduction of the ENSO amplitude in the mid-Pliocene in MIROC4m is independent of the choice of 100 year time-series segment.

From these results, it can be concluded that in some models, there are centennial scale variations in the Niño3.4 s.d., which become apparent when looking at individual ensemble members. However, the reported variations for CESM2 ($\pm 9\%$) and MIROC4m ($\pm 4\%$) are not enough to affect the PlioMIP2 ensemble mean reduction in Niño3.4 s.d. (24%, with 15 out of 17 models agreeing). When considering the individual model results, a variation of $+9\%$ would still result in 12 out of 17 individual ensemble members showing a reduction in Niño3.4 s.d., while a variation of -9% would result in all individual ensemble members showing a reduction in Niño3.4 s.d. In conclusion, the 100 year time series we use here for analysis of ENSO amplitude changes in the PlioMIP2 ensemble are sufficient to draw robust conclusions.

3 Spectral peak-counting robustness

The spectral peak-counting procedure (explained in the main paper) is performed using the multi-taper spectra of the Niño3.4 index of 100 years of data. In this section we will repeat the analysis using the Niño3 index, using different settings for the multi-taper (MT) method and using the fast-fourier transform (FFT) instead of the multi-taper spectral method:

1. Niño3.4 index, multi-taper method with 3 tapers and a bandwidth (bw) parameter of 2 (as in the main paper);
2. **Niño3 index**, multi-taper method with 3 tapers and a bandwidth parameter of 2;
3. Niño3.4 index, multi-taper method with **5 tapers** and a bandwidth parameter of 2;
4. Niño3.4 index, multi-taper method with 3 tapers and a **bandwidth parameter of 4**;
5. Niño3.4 index, **fast-fourier transform**.

The results of these five different cases are shown in Figure S7. The results show clear variations, but also feature the same trends in all the cases: an increase of significant peaks in the 1.5 - 2 year period, a decrease in the 2.5 - 4.5 year period and a similar number of significant peaks in the 7 - 10 year period. It can be concluded that the peak-counting procedure is robust against number of tapers, bandwidth parameter, Niño index and spectral method.

References

Tindall, J. C., Haywood, A. M., and Howell, F. W.: Accounting for centennial-scale variability when detecting changes in ENSO: A study of the Pliocene, Paleoceanography, 31, 1330–1349, https://doi.org/10.1002/2016PA002951, 2016.

Wittenberg, A. T.: Are historical records sufficient to constrain ENSO simulations?, Geophysical Research Letters, 36, 1–5, https://doi.org/10.1029/2009GL038710, 2009.

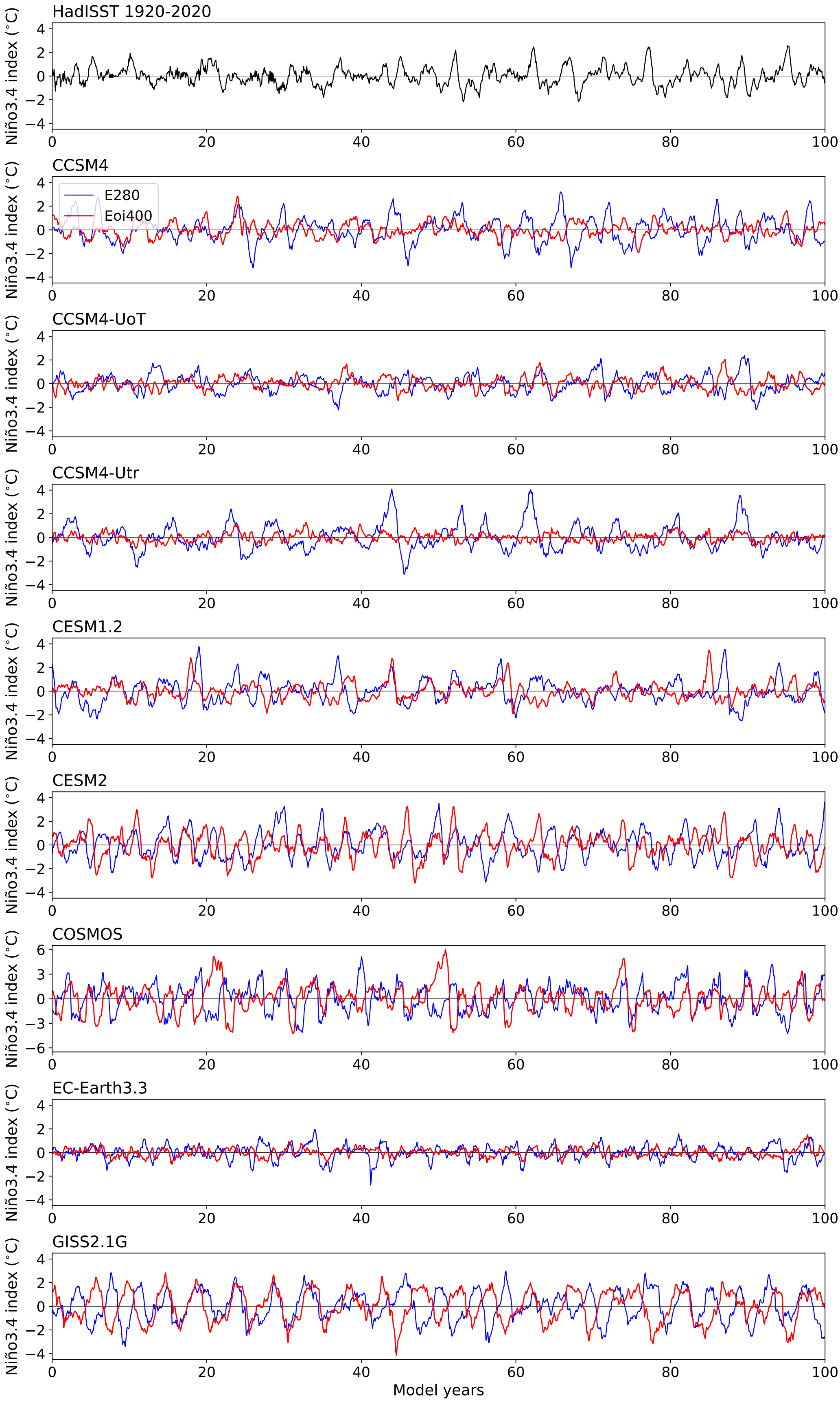

Figure S1: Niño3.4 index for all PlioMIP2 models, both pre-industrial E^{280} simulation (blue), mid-Pliocene Eoi^{400} simulation (red), as well as HadISST 1920-2020 results.

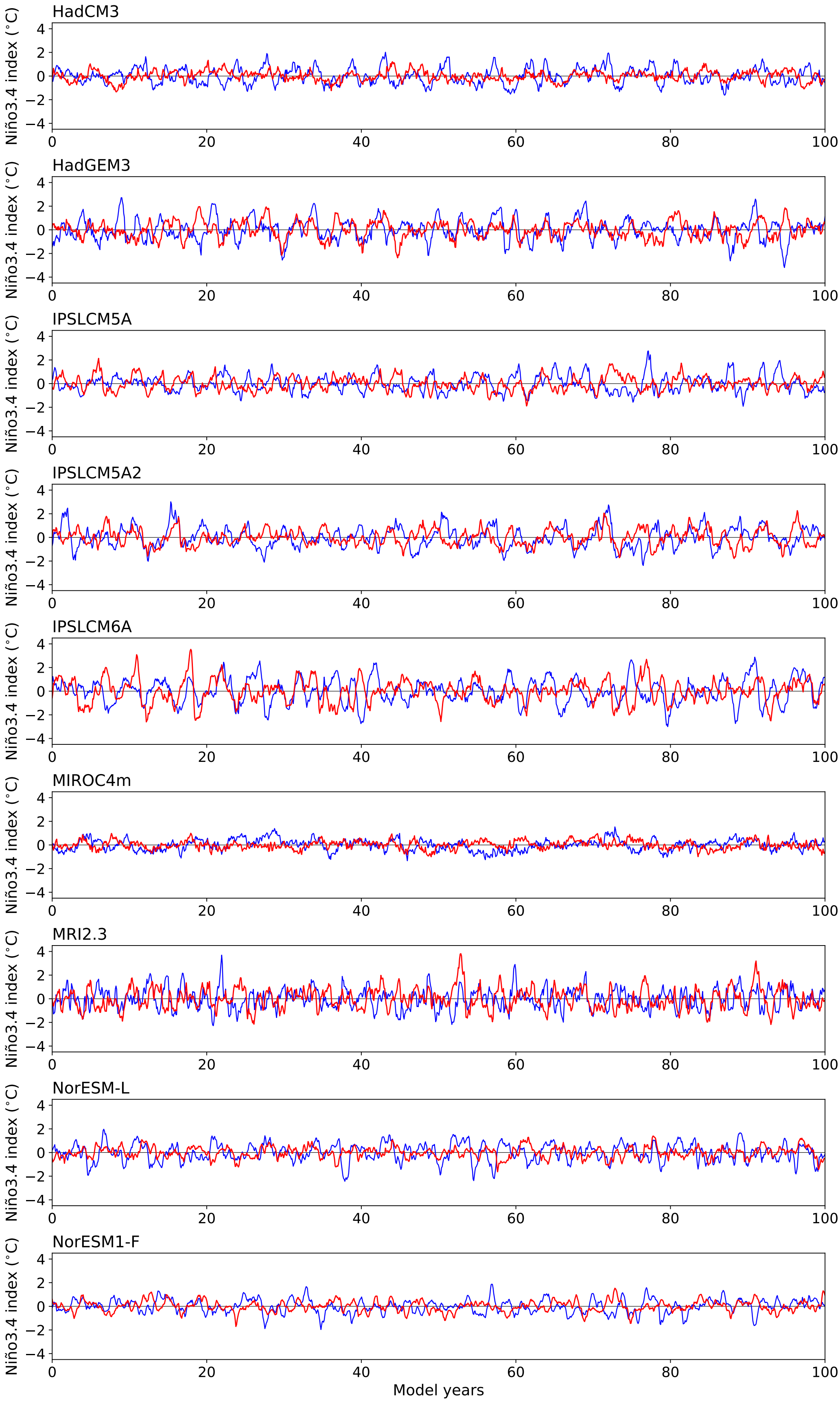

Figure S1: (cont.) Niño3.4 index for all PlioMIP2 models, both pre-industrial E^{280} simulation (blue), mid-Pliocene Eoi^{400} simulation (red), as well as HadISST 1920-2020 results.

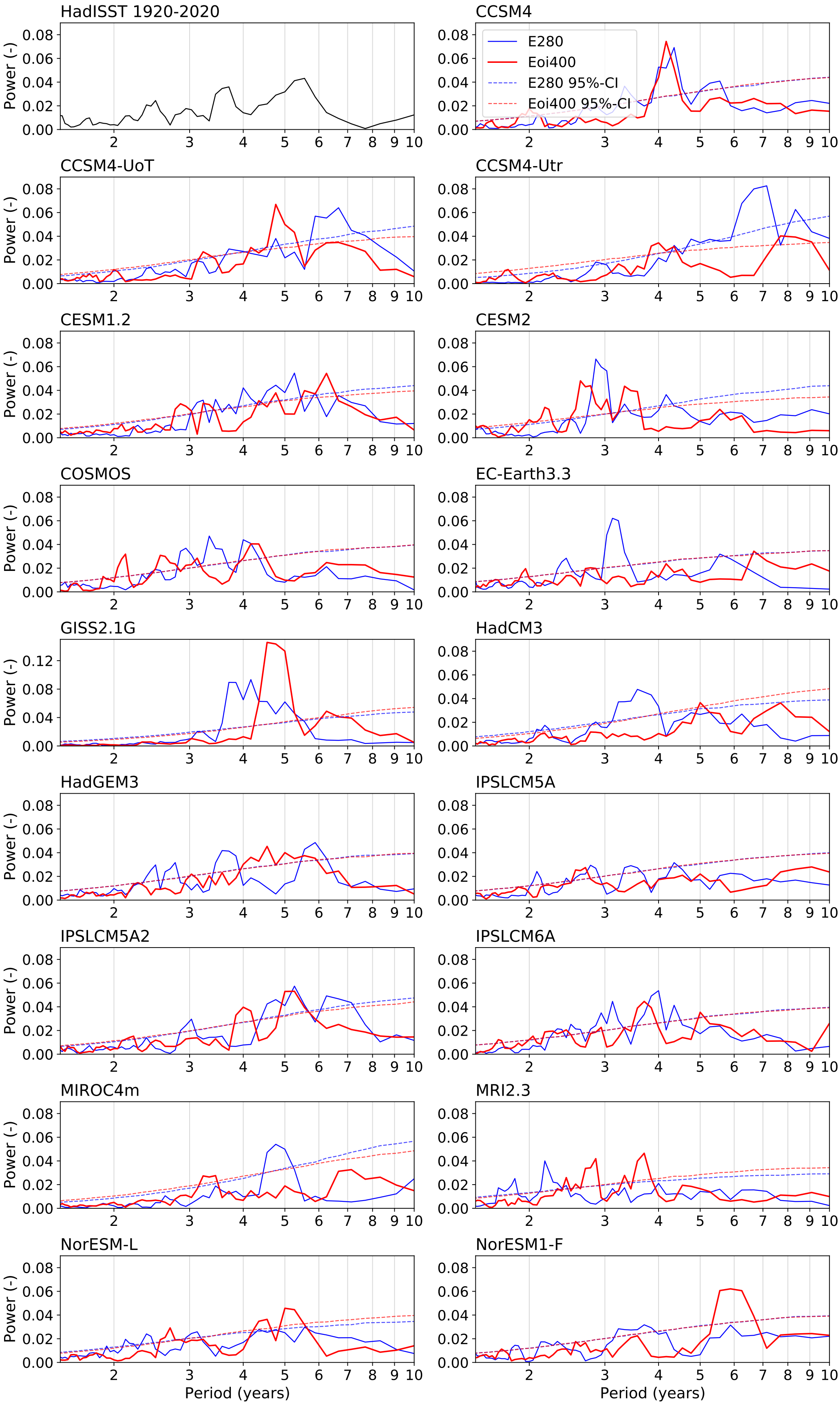

Figure S2: Multitaper power spectrum of the Niño3.4 index for all PlioMIP2 models, both pre-industrial E^{280} simulation (blue), mid-Pliocene Eoi^{400} simulation (red) as well as HadISST 1920-2020 results. Also including the 95% confidence interval for both simulations (coloured dashed line).

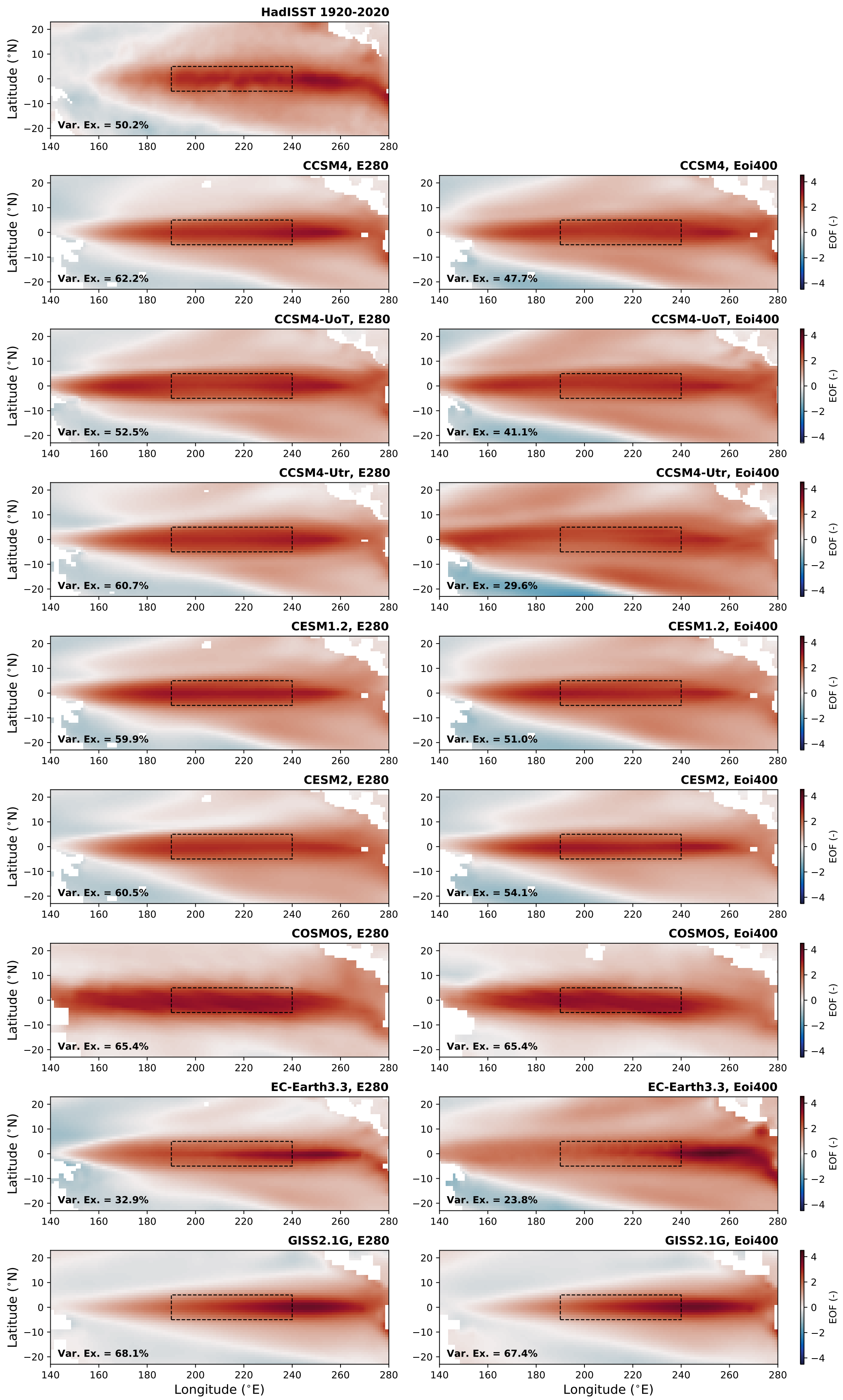

Figure S3: Leading EOF of the tropical Pacific SST anomalies for all PlioMIP2 models, as well as HadISST 1920-2020 results. Percentage of variance explained indicated in left bottom corner.

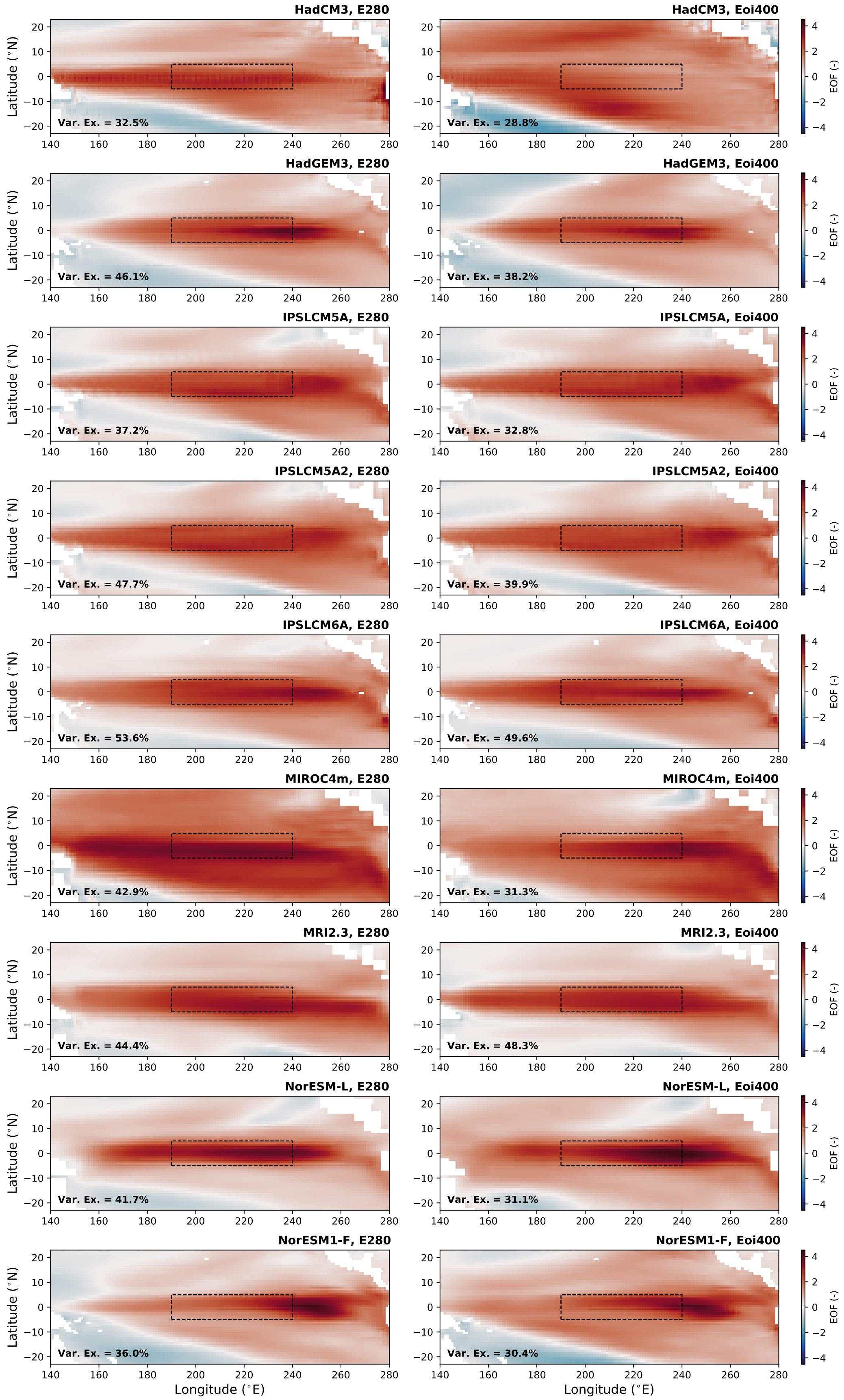

Figure S3: (cont.) Leading EOF of the tropical Pacific SST anomalies for all PlioMIP2 models, as well as HadISST 1920-2020 results. Percentage of variance explained indicated in left bottom corner.

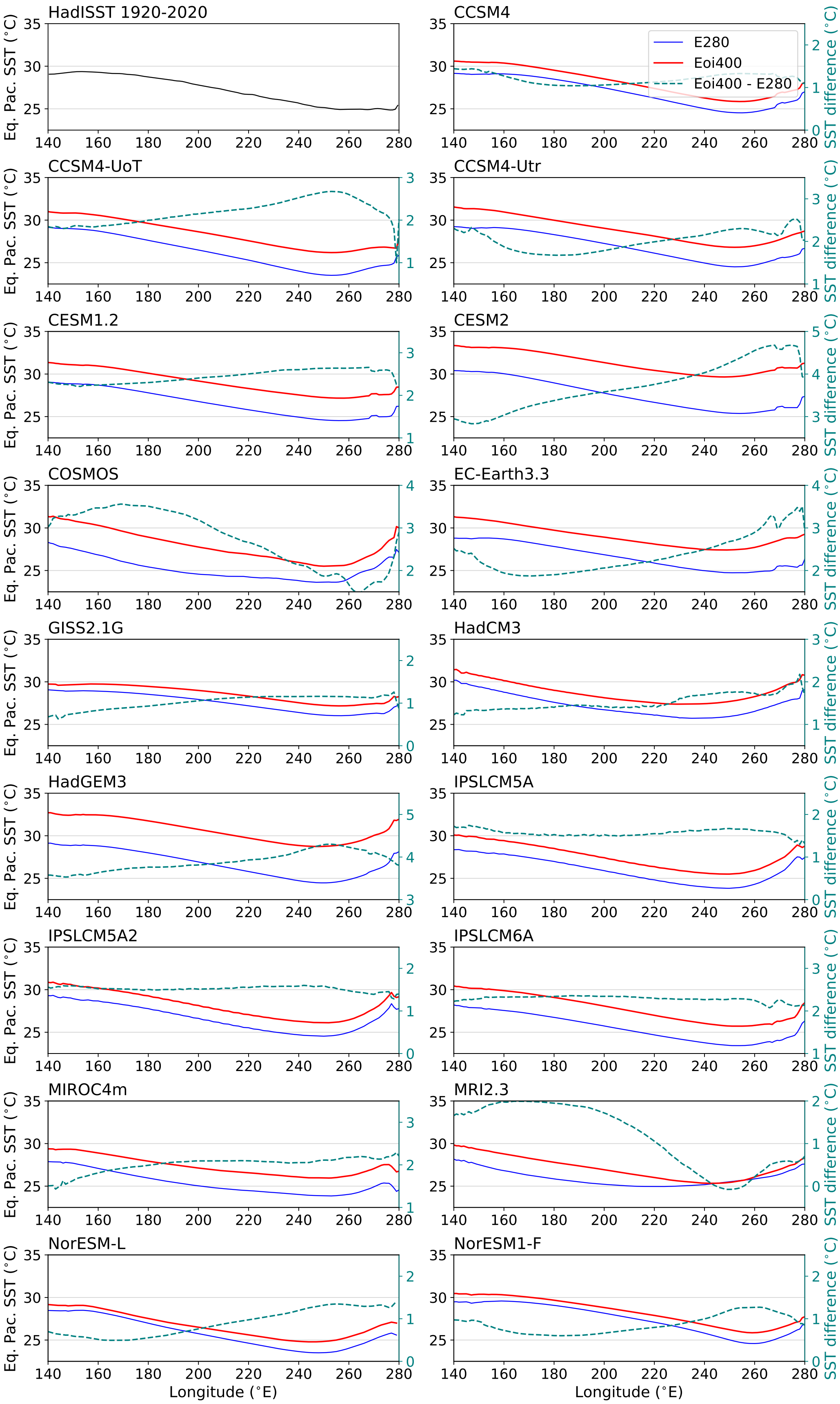

Figure S4: 5°S-5°N mean SSTs in the Pacific for all the PlioMIP2 models, as well as HadISST results. Pre-industrial E^{280} simulation (blue), mid-Pliocene Eoi^{400} simulation (red), Eoi^{400}-E^{280} difference (teal dashed line, right-hand vertical axis). Note that the range of the right-hand vertical axis shifts per model.

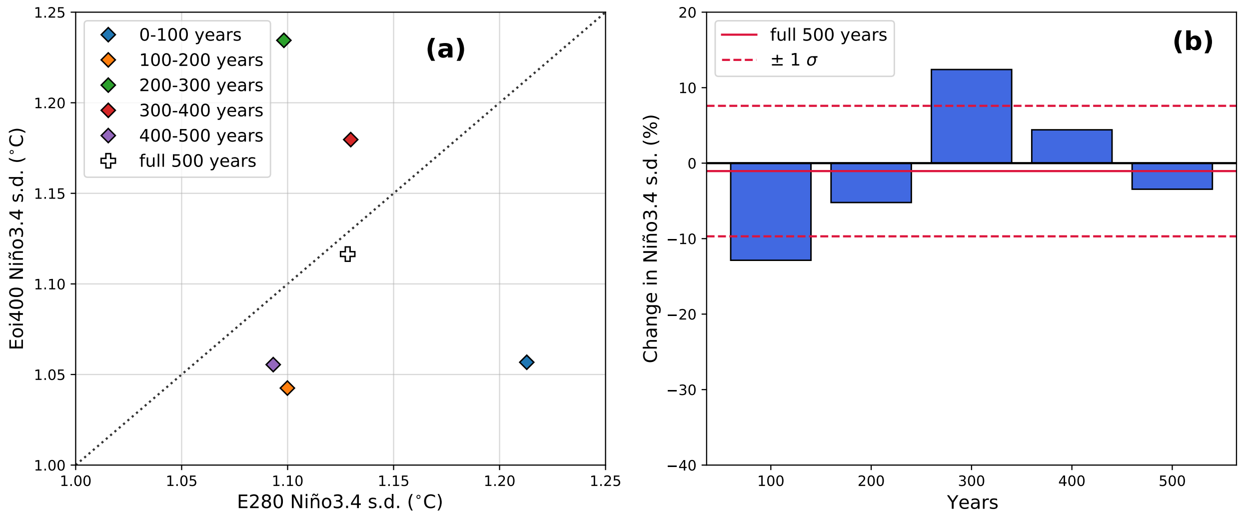

Figure S5: CESM2 Niño3.4 index standard deviation (s.d.); **(a)** E^{280} versus Eoi^{400} for different 100 years segments as well as full 500 years and **(b)** relative change in the mid-Pliocene, including the standard deviation around the 500 year value based on the five different 100 year values.

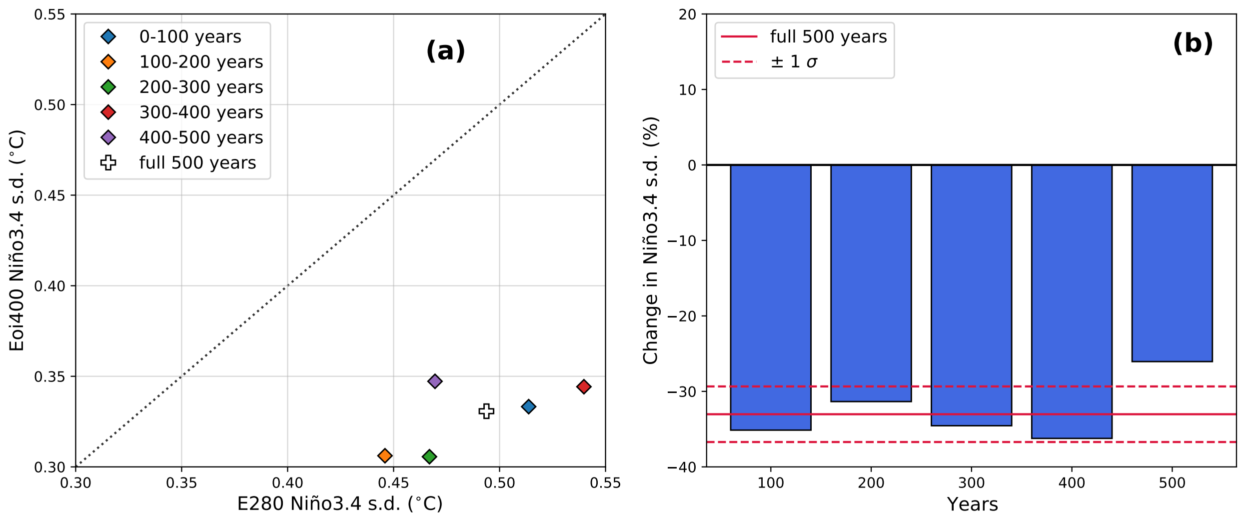

Figure S6: MIROC4m Niño3.4 index standard deviation (s.d.); **(a)** E^{280} versus Eoi^{400} for different 100 years segments as well as full 500 years and **(b)** relative change in the mid-Pliocene, including the standard deviation around the 500 year value, based on the five different 100 year values.

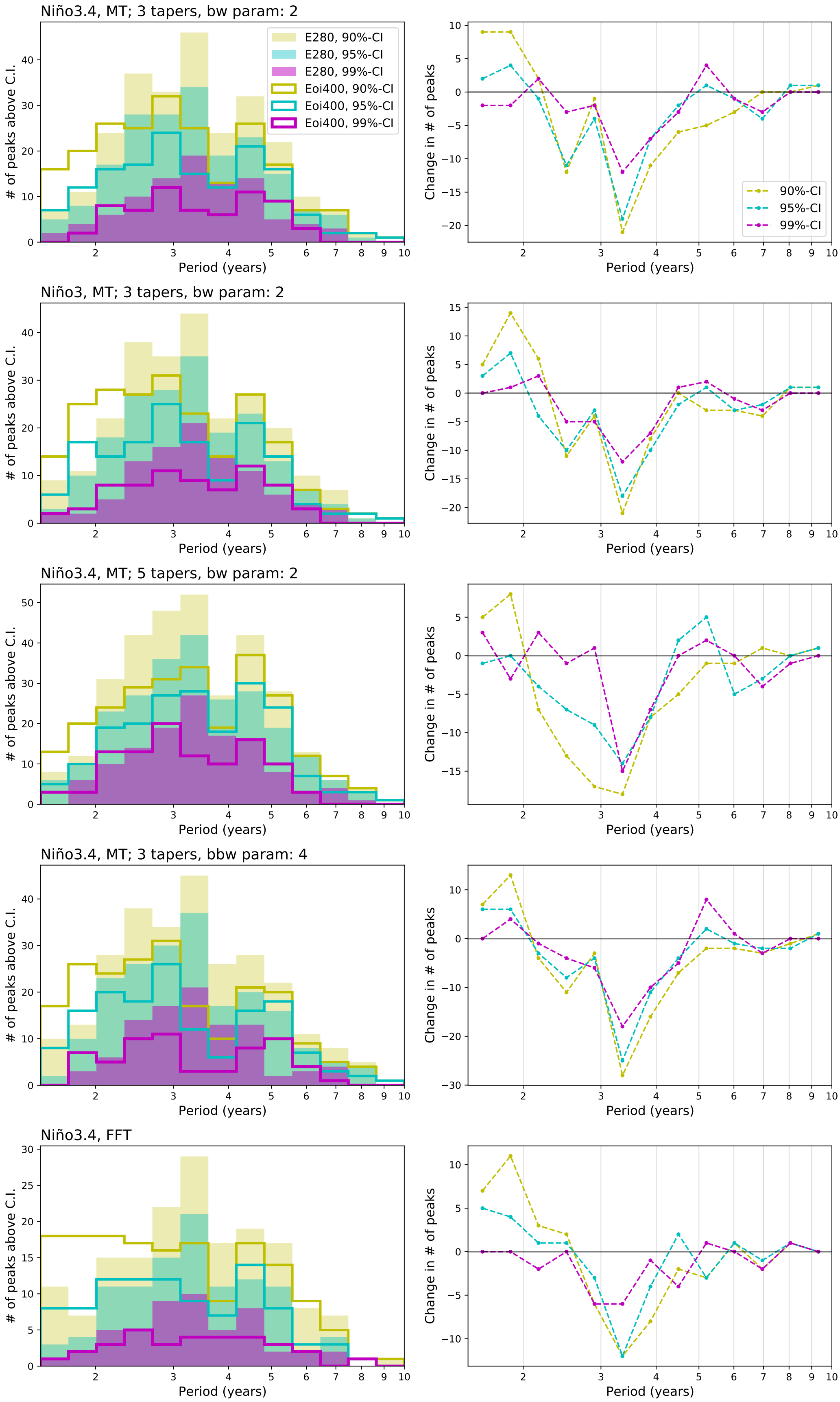

Figure S7: Ensemble sum number of spectral peaks that are above the 90%, 95% and 99% confidence levels, binned in the 1.5-10 year period range. Histogram of the E^{280} and Eoi^{400} peaks (left) and difference between the two (right), for different spectral methods and settings (top to bottom).