We thank the reviewer for the insightful comments. Our point-to-point responses are placed below each major comment and highlighted in blue.

This manuscript presents multiyear predictions of global air-sea CO2 fluxes by integrating the NOAA GFDL SPEAR model with the COBALT ocean biogeochemical model. The manuscript is well structured, provides a detailed methodology, and offers a balanced discussion on the limitation and future directions of carbon flux prediction using Earth system models. I only have minor comments, which are shown below:

1. I'm wondering whether the first leading predictable component accounts for the majority of the total variance, potentially affecting the extraction of 2nd and 3rd components. Would the ENSO-like patterns emerge more prominently if the analysis were conducted after removing the linear trend? Would the multi decadal pattern more closely resemble standard modes such as the PDO or AMO?

The principal strength of APT analysis is that it decomposes multivariate data into orthogonal components, each maximizing a distinct portion of total predictability. By design, the leading component represents the most predictable signal. When applied to raw time series, APT isolates a component dominated by externally forced variability (driven by CO_2 and aerosol forcing) that remains consistent across ensemble members and lead times. This represents the primary predictable signal identified by APT, exhibiting the longest predictability horizon (theoretically reaching the maximum forecast length in the absence of ensemble noise). However, forced variability does not represent skill gained specifically from initialization, as evidenced by Figure 6c, where ACC values are comparably high across all simulation types. Because APT components are orthogonal, sequentially removing leading components leaves subsequent components unaffected, enabling a clean separation between externally forced signals (the first component) and internally generated natural variability (subsequent components).

During manuscript preparation, we also evaluated APT using detrended time series as the reviewer asked. APT reliably extracted the remaining components from the detrended dataset (i.e., new Component 1 aligned with original Component 2, new Component 2 aligned with original Component 3, etc.). However, standard linear detrending failed to fully remove nonlinear forced signals (such as volcanic impacts), causing residual forced variability to contaminate the first component. Conversely, polynomial detrending introduced spurious multidecadal artifacts. Consequently, performing APT on the raw time series proved to be the most robust approach for isolating forced predictability from initialized natural predictability. We will expand our explanation of these methodological choices in the revised text and include the detrended APT analysis in the Supplemental Material.

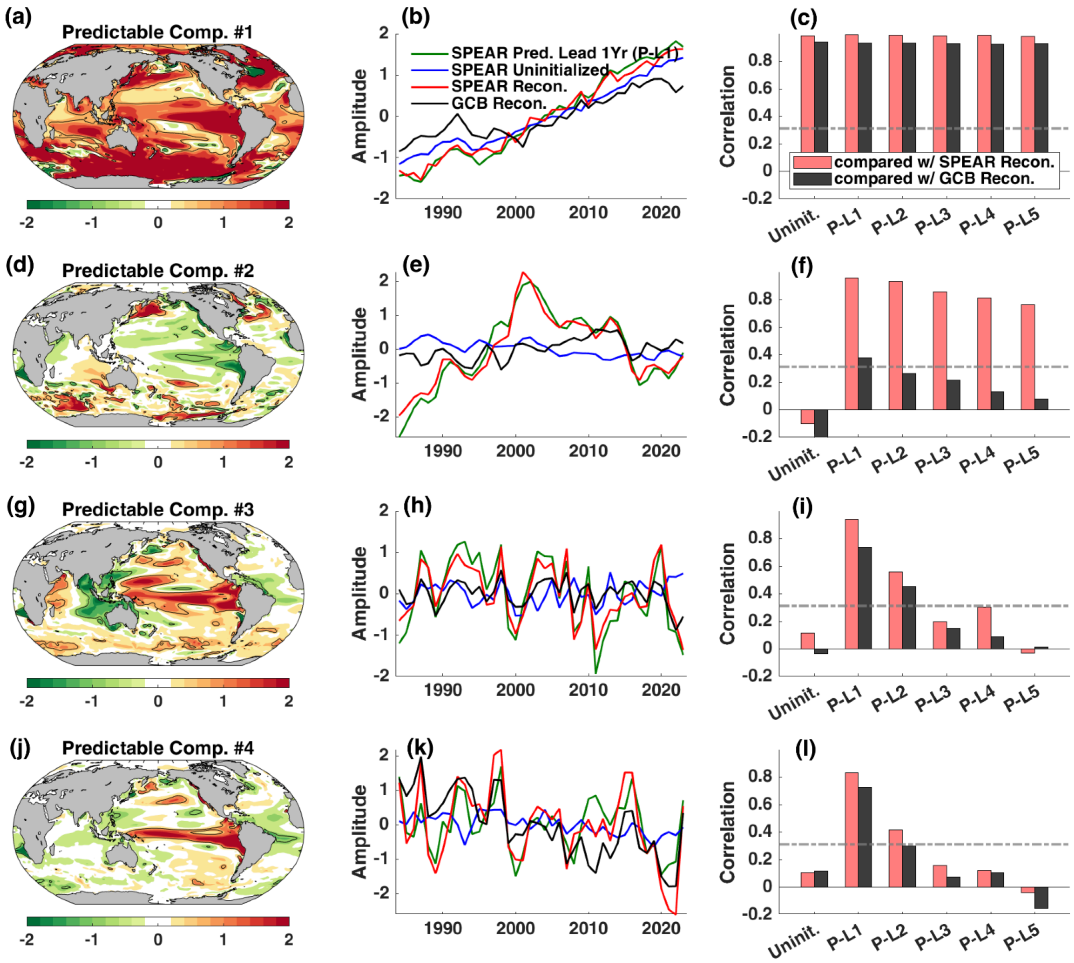
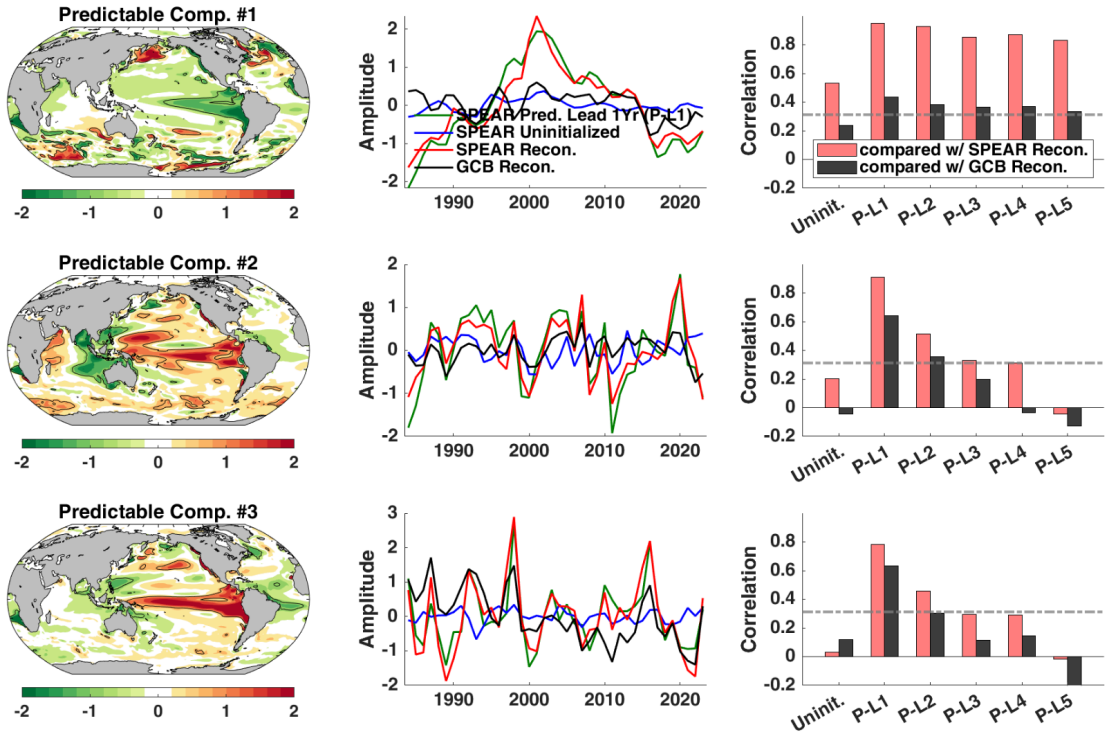


Figure 6 (this study) showing the four leading predictable components from APT analysis using the original or raw anomalies.



Additional figures showing the three leading predictable components from APT analysis using the detrended anomalies.

2. Are observed aerosols and CO_2 from the target years used after initialization?

Correct - the observed aerosols and CO_2 are indeed used to force the predictions while the atmospheric restoring was stopped at initialization. That was one of the reasons we used the APT analysis to isolate the forced skill (mostly addressed by component 1) from the true skill driven by predictability of climate variability, the information of which is contained in the initial conditions but not in the forcing. There are new opportunities with our ongoing efforts to explore the predictability in the emission-driven simulations where CO_2 between the ocean, land, and atmosphere is fully interactive.

3. There are some errors in writing. Examples are as follows:

- Line 275: The equation should be numbered (4).

- Line 367: Fig. 1b --> Fig. 1e

- Figure 3 caption: The current figure shows six panels, but the caption has the description of eight panels.

- Figure 4 caption: (e-j) --> (f-j).

- Figure 6c,f,i,l: The color description in the legend is opposite to the caption.

- Figure 7-8 caption: The time-series panels are (a,c,e), rather than (a,b,c).

We thank the reviewer for pointing out these editorial errors and will carefully address them in the revised manuscript.