

## **Reply to reviewers for the manuscript egosphere-2025-1924:"Constraining the time of emergence of anthropogenic signal in the global land carbon sink"**

**Na Li, Sebastian Sippel , Nora Linscheid , Miguel D. Mahecha , Markus Reichstein , and Ana Bastos**

We thank the reviewers for the productive and constructive comments. Here we reply to the comments in detail, and we suggest how to adjust the manuscript according to the comments. In addition, based on the comments, we made some extra changes in the manuscript for easier reading and understanding (at the last of each reply).

Below we address the changes according to the comments **(the line numbers in replies are according to the new manuscript that shows changes)**. The blue colored lines are my replies, and the red words are locations changed.

## Changes according to the reviewer #1:

### Specific comments:

This paper considers the time of emergence of 3 key variables in the global land carbon sink and shows how dynamical adjustment can be used to shorten the detection time. While the paper is interesting and provides an important study in the literature I have a few major concerns around the methodology and lack of observational evidence. As such, I recommend major revisions.

We thank the reviewer for raising these important points, below we address the manuscript according to the comments.

1. The most major issue with the paper is the methods, which use linear trends to calculate ToE. This is not the typical method found in other studies such as:  
<https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019GB006453> for large ensembles and  
<https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011gl050087> for the standard method.

It is clear from the Figure time series that a linear fit is not appropriate for this data.

Additionally presenting ToE of 800 years, when the timeseries is only 100-years is confusing and extrapolation. If the signal has not emerged by the end of the timeseries I suggest presenting this as not emerged rather than extrapolating.

We acknowledge the reviewer's concern regarding the validity using a linear regression approach. Below we answer your concerns one by one:

1) Has the linear trend approach been applied before?

Yes, our approach is based on Bonan et al., 2021. Using fitted linear trend of ensemble mean as signal trend has been applied in recent published papers. The first paper mentioned (Schlunegger et al., 2020) also used fitted linear trend of ensemble mean as signal change, the similar approach as we did. They wrote (Schlunegger et al., 2020):

"For each LE, the signal is the ensemble average trend, computed as the average of the ~30 ensemble members' trends (linear, least squares trend)."

The second paper (Hawkins et al., 2012) mentioned used a nonlinear approach to represent signal trend: fitted fourth order polynomial regression of global mean SAT change. However, their study involves only around 10 simulations, so the ensemble mean might not reflect the full anthropogenic signal (Maher et al., 2019, Deser et al., 2020, Milinski et al., 2020). A fourth order polynomial might overfit the trend and miscapture some internal variability as signal trend, particularly in regional signal trend, which is subject to larger variability. In addition, their approach fits each simulation separately, so the fitted anthropogenic signals among all simulations might not be consistent within one single model. At last, the year to year calculation of ToE with actual fitted NBP value as signal might be more sensitive to anomalies.

To address the comment, we compared the Hawkins et al., (2012) with our method. We replicated almost exactly the steps described in Hawkins et al. (2012):

(i) For each Earth system model used in this study, we defined the signal and noise period of NBP in the same way as in our analysis: signal in the period 1960-2009 and noise 1930-1959.

(ii) At the global scale, for each model simulation, we fit a fourth order polynomial to calculate the fitted global signal trend of NBP. Noise for each model and each simulation is defined as the standard deviations of interannual variations.

(iii) At the regional scale, we used the 10 RECCAP domains. The signal trend for each domain is calculated as a scaled value of the corresponding global signal trend for each model simulation. The scaling procedure is applied exactly as in Hawkins et al. (2012):

$$S(t) = \alpha \tilde{T}_{\text{global}}(t) + \beta$$

$S(t)$  represents the regional signal trend, and  $\tilde{T}_{\text{global}}(t)$  is the fourth-order polynomial fitted over 1960-2009. Regional noise is defined as the standard deviations of interannual NBP of each model simulation.

(iiii) Time of Emergence (ToE) is defined as  $\text{ToE} = S/N$ . We identified the first year in which that ToE exceeded 2. This is equivalent to our definition,  $\text{ToE} = 2N/S$ .

(iiiii) For each model, we take the median ToE across all simulations.

For example, in the model CESM2-LE, we have 90 simulations. We define the signal period as 1960-2009 and the noise period as 1930-1959. On the global scale, we fit a four-order polynomial to each simulation to estimate the signal trend, with noise defined as the standard deviation. At the regional scale (e.g., North America), the domain signal trend is scaled to the corresponding global trend, and noise is the standard deviations within each domain. For each simulation, ToE is calculated year by year, and the first year in which ToE exceeded 2 was recorded. The median ToE across the 90 simulations is reported as the final emergence time.

Below we show (Fig. S1) the ToE for NBP (global and 10 RECCAP domains) using the approach in Hawkins et al. (2012):



Figure S1. ToE of global and 10 RECCAP domain NBP. ToE is calculated according to the approach in Hawkins et al. (2020). ToE for each model each region is the median of all simulations.

We compare this with our approach in Figure 2 Panel b of our manuscript.

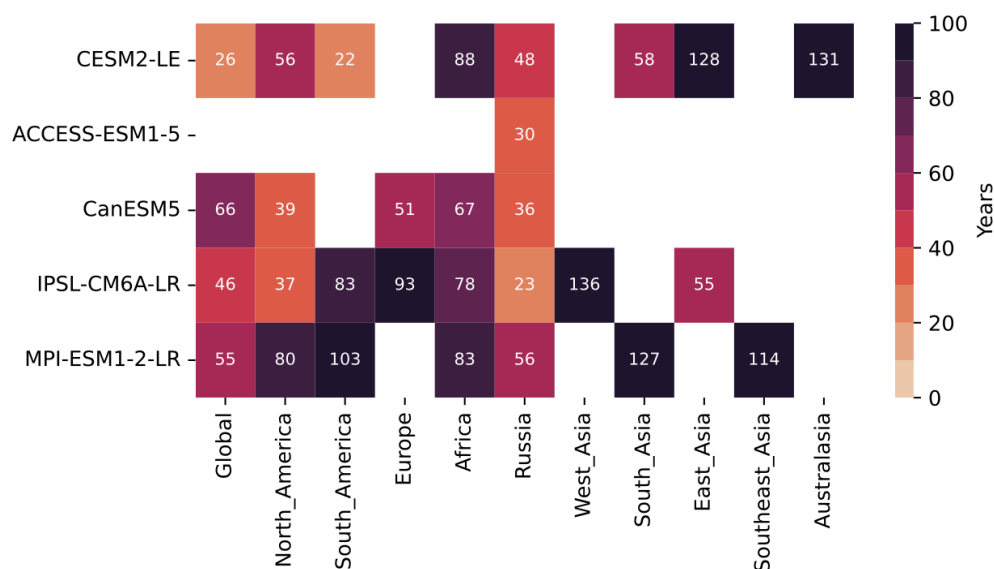


Figure 2: ToE of NBP on a global scale and across 10 RECCAP-2 regions, under historical simulations of five ESM large ensembles. Note that ToE is the years detectable after 1960, and is calculated with signal period of 1960-2009 relative to the noise period of 1930-1959, details please check Sect. 2.4. Heat map of the ToE in global and each spatial domain of NBP. Domains with no significant signal ( $P > 0.05$ ) are shown as empty squares.

We found the Hawkins et al. (2012) approach (Fig. S1) generally yields shorter ToEs than our linear approach (Fig. 2). However, in their approach, two regions show a ToE of only one year (CESM2-LE Europe and ACCESS Russia). This might be due to large signal values in 1960, or small noise in the period 1930-1959. For example, CESM2-LE

Europe with a median ToE of 1 year, across all 90 ensemble simulations. The ensemble mean since 1960 is apparently larger than most other 10 RECCAP domains (Fig. A1 in the manuscript shown below, a panel of ensemble mean as suggested by the reviewer). There might be an elevated value of fitted NBP signal trend since 1960, across most simulations in this region. Thus their approach is sensitive to the reference period selected and year by year ToE calculation makes their approach more sensitive to anomalies. Instead, we calculate the signal trend of the European ensemble mean in the period 1960-2009, the fitted linear slope shows no significant trend in this period (see Fig. A4 in the manuscript). Thus the emergence time is not detectable in this region.

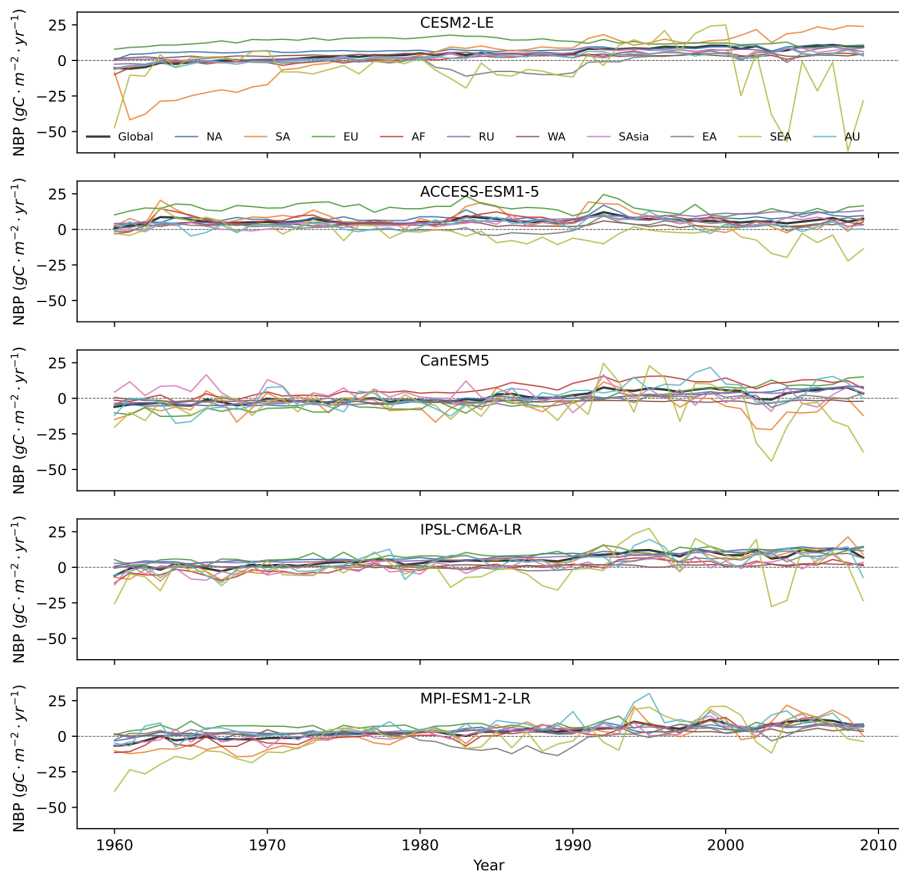


Figure A.1. Time series of NBP ensemble mean from five Earth system models. The thick black line is the global ensemble mean, and the colored lines represent ensemble means for the 10 RECCAP-2 regions.

Because ToE in the nonlinear approach is defined as the median across simulations, the exact signal or noise value cannot be directly identified. Instead, we calculated the number of simulations in each model with ToE >2 (Fig. S2).

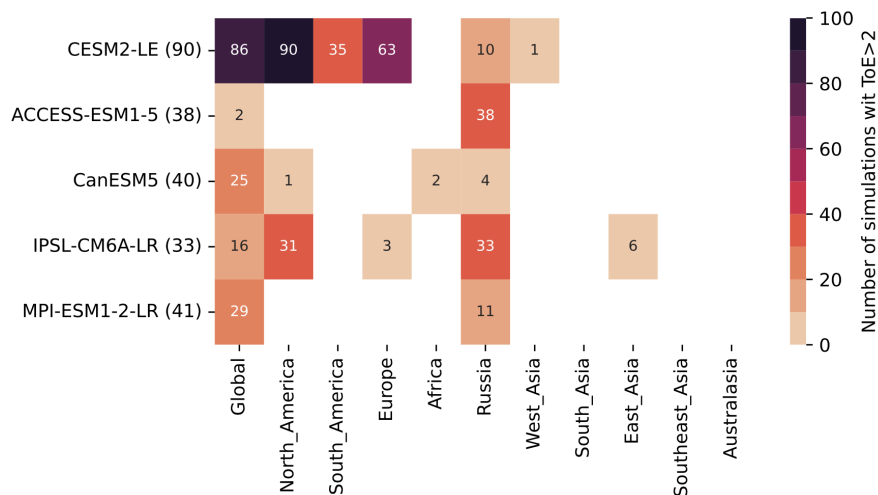


Figure S2. In the Hawkins et al (2012) approach, the number of simulations in each domain that show ToE > 2. The number of simulations included in this study are listed after each model name.

Fig. S2 shows that in many regions there are fewer than 10 simulations that exceed this threshold, meaning that less than one-third of simulations detect emergence. This indicates that the anthropogenic signals captured across simulations are not consistent.

In summary, although the nonlinear approach of Hawkins et al. (2012) generally produces shorter ToEs than our approach, it is less robust. Their approach shows: 1) unrealistic short detection times (1 year); and 2) in most regions, more than two-thirds of simulations fail to detect the anthropogenic signal. 3) the anthropogenic signals captured across models are not consistent.

This might be due to: instead of defining the signal trend as the ensemble mean, Hawkins et al. (2012) defined the signal trend separately for each simulation and then took the medians across simulations. This makes the signal trends vulnerable to anomalies in individual simulations, blurring the separation between anthropogenic signal and internal variability. Furthermore, with fewer than one-third of simulations detecting the signal, the reported median of ToEs are less reliable. Their method also calculates ToE year by year (e.g., one year ToE in Fig. A1), whereas we calculate the signal trend over the entire signal period.

This shows that the nonlinear approach used by Hawkins et al. (2012) is less robust on capturing the consistent anthropogenic signal and more sensitive to anomalies, particularly with net land carbon sink time-series, which is subject to large interannual variations.

## 2) Is linear fit for signal trend appropriate in this study?

Yes. As concerns about the validity of linear regression slope, rather than a nonlinear approach, to represent the signal trend, we have below main reason:

Capturing the dominant signal (first-order degree variability). The ensemble mean reflects the forced signal in the Earth system, including anthropogenic forcing (both

long-term consistent and short-term surges), short period natural forcings (e.g., volcanic eruptions), and even decadal internal variability (Deser et al., 2012b; Canadell et al., 2021; Eyring et al., 2021; Mercado et al., 2009; Zhang et al., 2021). The linear fitted trend of the ensemble mean captures the most consistent and incremental influence of the anthropogenic signal, while minor deviations from the linear trend slope may be ignored. A nonlinear regression might capture these minor fluctuations but at the risk of overfitting and misattributing natural forcing or decadal internal variability to anthropogenic signals. This risk is particularly pronounced at regional scales in the carbon cycle, where the ensemble mean exhibits larger variability (see Figure A.1 and Figure 3). Additionally, ToE comparisons across models is better possible with linear trends with two consistently interpretable parameters. In addition, the linearized trend is only applied over five decades (not over the whole period), and for a limited time period the linearization of the trends is more justified than for the whole period.

We have added one small paragraph explaining why linear regression is suitable for this purpose. In line 140-145:

“Here we use a linear regression slope rather than a nonlinear approach to represent the signal trend, this is to capture the dominant forced signal in the selected signal period. The ensemble mean of NBP, GPP and TER reflects the forced ecosystem response, including anthropogenic forcing, short-period natural forcings (e.g., volcanic eruptions), and decadal internal variability (Deser et al., 2012b; Canadell et al., 2021; Eyring et al., 2021; Mercado et al., 2009; Zhang et al., 2021). The linear trend captures the first-order long-term anthropogenic influence, whereas nonlinear methods could risk overfitting and misattributing natural or internal variability to anthropogenic signals, especially at regional scales where variability is larger (see Figure A.1 and Figure 3).”

### 3) Why include ToEs beyond the years of the selected study period?

We thank the reviewer for this question. Including Time of Emergence (ToE) values that extend beyond the study period provides important information about regions which might be dominated by natural variability, or are weakly influenced by anthropogenic forcing. The timing of these extended ToEs indicate how severe such limitations are in particular regions and climate situations. Since most future scenarios show ToEs below 150 years, and after reducing the circulation-induced noise in NBP, the ToEs show apparent reduced years. Here, to show the filter effect by using Ridge regression, we cut the ToEs to 150 years.

Additionally, we add this information in line 188-190:

“Note that only the calculated signal (regression slope) with significance value  $P < 0.05$  is selected. If the calculated signal (regression slope) is negative, we then take the absolute signal value to get a positive ToE. Here we select to show the ToEs less than 150 years.”

2. This method is used typically to understand when we might expect to observe changes outside the noise. However, the paper has no observational evidence in it. I suggest comparing to any observations we have.

We thank the reviewer for pointing out this critical issue, we added the evidence of using observations below, also as Figure 7:

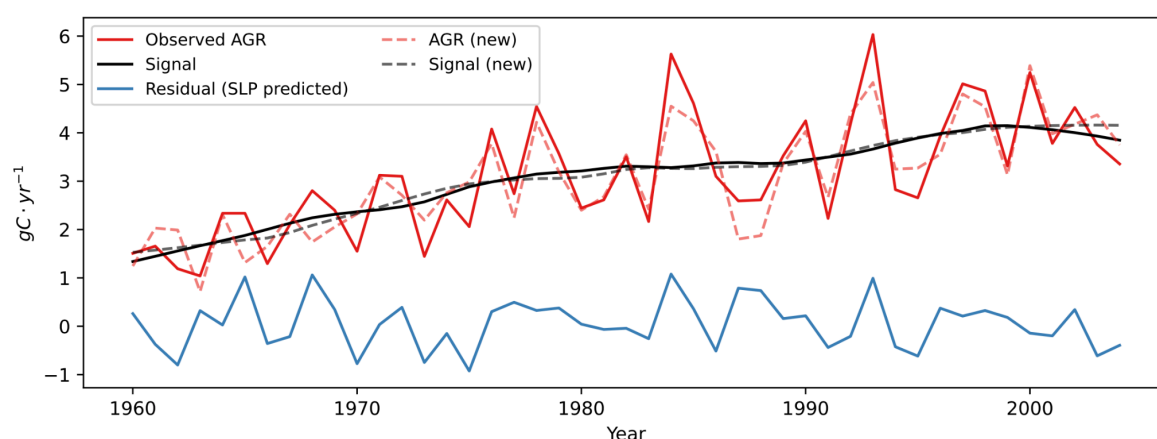


Figure 7. Time series of the atmospheric CO<sub>2</sub> growth rate AGR at Mauna Loa from 1960 to 2009 (Lan et al., 2025). Five volcanic years (1963, 1982, 1983, 1991, and 1992) are removed. The red line is the observed AGR. The black line is the long-term trend fitted with a locally weighted scatterplot smoothing (Cleveland et al., 1991, LOWESS) (signal). The residual (AGR - fitted long-term trend) was predicted using SLP through ridge regression with leave-one-out cross validation (blue line). This SLP predicted residual is then subtracted from the observed AGR to obtain a new AGR time series with circulation-induced variations removed (observed AGR - SLP predicted residual). The dashed black line is the new long-term trend. Data pretreatment and the ridge regression model follow paper Li et al. (2021). Note that the signal period is the same as in models (1960—2009). Due to limited records of CO<sub>2</sub> observations before 1958, here we calculate the noise also in the period 1960—2009.”

Accordingly, we add a paragraph to line 328 :

“We then test the same approach on observations of Atmospheric CO<sub>2</sub> growth rate (AGR) from Mauna Loa (Lan et al., 2025) for the period 1960—2009, matching the period of signal for the Earth system model analysis (Fig. 7). The ToE of the observed AGR is 33 years, with noise value of 0.87 gC · yr<sup>-1</sup> and signal value of 0.05 gC · yr<sup>-2</sup> (Fig. 7). After removing circulation-induced variations through dynamical adjustment, the ToE of the adjusted new AGR is reduced to 23 years, with noise of 0.70 gC · yr<sup>-1</sup> and signal of 0.06 gC · yr<sup>-2</sup> (Fig. 7). This represents an overall reduction of about 30%, contributed by 19% reduction in noise and 20% increase in signal. The results show that this approach can be applied in observations, enabling earlier detection of anthropogenic signals in global carbon cycle variability.”

We also added the information to abstract, Methods section “Data” and “Data pretreatment”:

In line abstract line 17-19:



“This approach substantially shortens the detection time for the global net land carbon sink: between 34–39% for the historical period and 27–54% for the future simulations. This approach ~~can, in principle, be applied to~~ can also shorten the detection time for observational based datasets (30% reduction), thereby improving our ability to detect long-term trends of land carbon sink variability.”

In line 111-112:

"We also included the observations of atmospheric CO<sub>2</sub> growth rate (AGR) at Mauna Loa from 1960 to 2009 (Lan et al., 2025), downloaded from [https://gml.noaa.gov/webdata/ccgg/trends/co2/co2\\_gr\\_gl.txt](https://gml.noaa.gov/webdata/ccgg/trends/co2/co2_gr_gl.txt), last accessed on August 18th."

And line 125-127:

"The pretreatment steps of atmospheric CO<sub>2</sub> growth rate (AGR) at Mauna Loa from 1960 to 2009 (Lan et al., 2025) follows Li et al. (2022). We first have five volcanic years (1963, 1982, 1983, 1991, and 1992) removed, then remove the long-term trend by fitting with a locally weighted scatterplot smoothing (Cleveland et al., 1991, LOWESS)."

3. The % reductions from the dynamical adjustment might be good to include as years as well as % as this is easier to understand. For the dynamical adjustment it is unclear whether this method could be applied to observations to reduce the TOE or just models. I suggest adding this information in the text.

Thanks for pointing out, we added years that are equivalent to % reduction in Table A.1 and A.3. However, it should be noted that higher relative reduction % do not imply longer year reductions among models and study periods. Table A.1 and A.3 now looks like below (note that we changed the noise period from 2020-2050 to 2020-2070, so the values below are slightly different):

Relative change % (Years)	CESM2-LE	ACCESS-ESM1-5	CanESM5	IPSL-CM6A-LR	MPI-ESM1-2-LR
Historical	34% (9)	-	39% (26)	35% (16)	36% (19)
SSP1-2.6	-	2% (2)	43% (64)	34% (29)	31% (30)
SSP2-4.5	-	-	53% (32)	-	-
SSP3-7.0	35% (15)	34% (34)	55% (19)	-	-
SSP5-8.5	-	-	52% (10)	29% (42)	37% (26)

**Table A.1.** ToE reduction in NBP, calculated according to Eq. (4).

Relative change % (Years)	CESM2-LE	ACCESS-ESM1-5	CanESM5	IPSL-CM6A-LR	MPI-ESM1-2-LR
Historical	25% (2)	32% (4)	13% (1)	26% (2)	25% (2)
SSP1-2.6	-	60% (67)	34% (3)	26% (4)	38% (7)
SSP2-4.5	-	45% (17)	34% (2)	28% (3)	38% (4)
SSP3-7.0	21% (1)	41% (11)	41% (2)	30% (2)	36% (3)
SSP5-8.5	-	43% (9)	34% (1)	19% (1)	37% (2)

**Table A.3.** ToE reduction in GPP, calculated according to Eq. (4).

We also added the equivalent years to % reduction in line 320:

“ToE ranges from ~~34.0%~~ (CESM2-LE) to ~~3938.6%~~ (CanESM5), corresponding to 9 and 26 years, respectively.”

Line 322:

“reduction ranges from ~~2927.1%~~ to ~~5554.3%~~ (42 and 19 years reduction, respectively)”

Line 324:

“it ranges from ~~1342.7%~~ (CanESM5) to ~~32.4%~~ (ACCESS-ESM1-5), corresponding to 1 and 4 years, respectively. For future scenarios, the relative reduction ranges from ~~19.2%~~ to ~~6059.4%~~ (1 and 67 years reduction, respectively).”

We also added experiment results of observations as Figure. 7 (as shown above).

4. I am confused about the use of 2020-2050 for N and 2020-2070 for S. What makes this choice? This seems very arbitrary

For the signal period, the trend of S after 2020 is more sensitive to the period selected. Across the four future scenarios, NBP shows much larger variations than in the historical runs, and almost all the signals are mixed before 2050. In addition, all four future scenarios define key CO<sub>2</sub> emission milestones by 2050 (see section 2.1). To reduce the influence of large annual variability (by selecting a period of at least 30 years) and to fully capture the impact of these policy milestones, we selected 2070 as the end year. According to Figure 4, the standard deviation of N is not sensitive to the specific period selected. We therefore choose 2020-2070, the same period with signal. The results have minor changes accordingly.

5. Figures

Figure 1: You could easily include all NBP, GPP and TER on one figure – I would recommend this rather than push some to the supplementary. I also suggest adding a multi-ensemble mean panel. The ensemble sizes seem rather arbitrary. For example ACCESS has 40 members and

CESM2 100 members – so why are they only 38 and 90? Additionally CESM2 biomass burning runs have been shown to be different to the cmip6 runs have you checked these differences?

1. Include all NBP, GPP, and Ter in one figure.

Thanks, we combined NBP, GPP and TER in Figure 1.

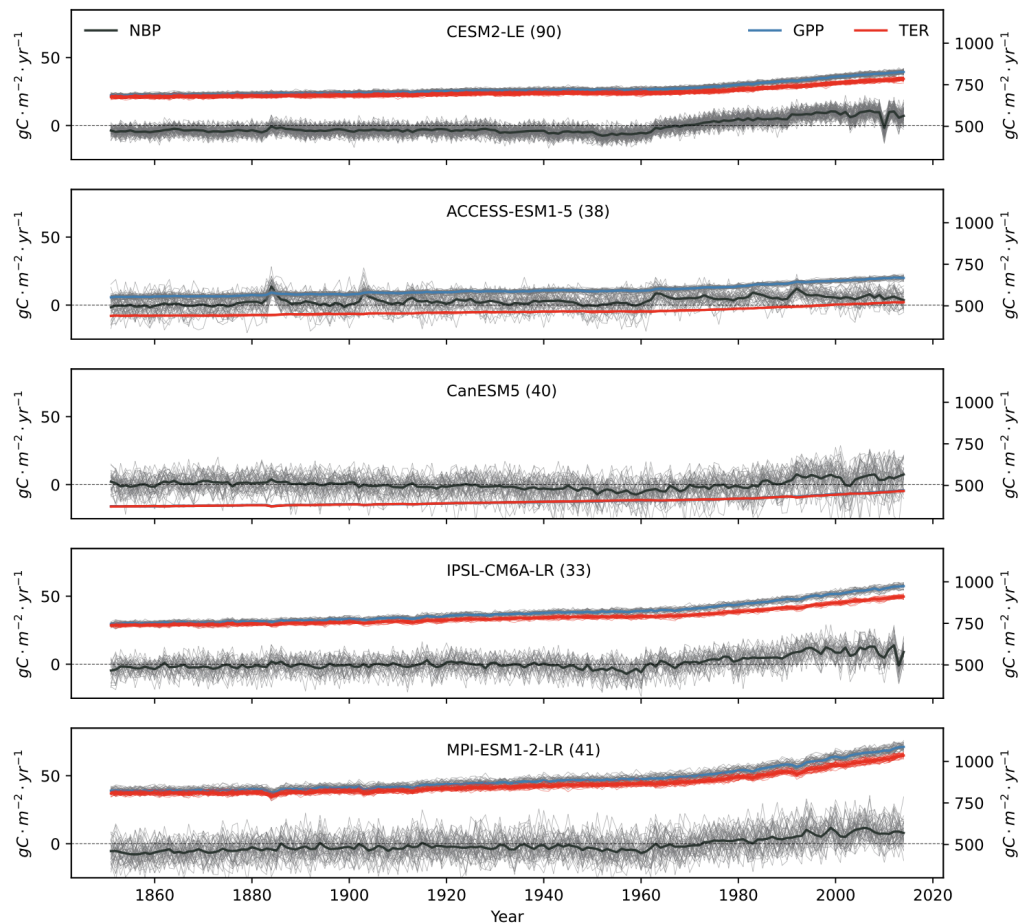


Figure 1. Time series of NBP, GPP, and TER from 1851 to 2014 in five ESM large ensembles. The thin lines represent individual simulations, while the bold lines represent the ensemble mean. The gray lines are NBP follows the left y-axis. The blue and red lines follow the right y-axis and are GPP and TER, respectively. The number of simulations for each model is listed in the legend next to the model name. Note that TER in model ACCESS-ESM1-5 only included 24 simulations, due to limited data availability.

2. Adding a multiple ensemble panel

We included three figures of multi-ensemble mean (global scale and 10 RECCAP-2 Regions) in supplementary Figure A.1 to A.3, for NBP, GPP, and TER respectively. Below are three Figures:

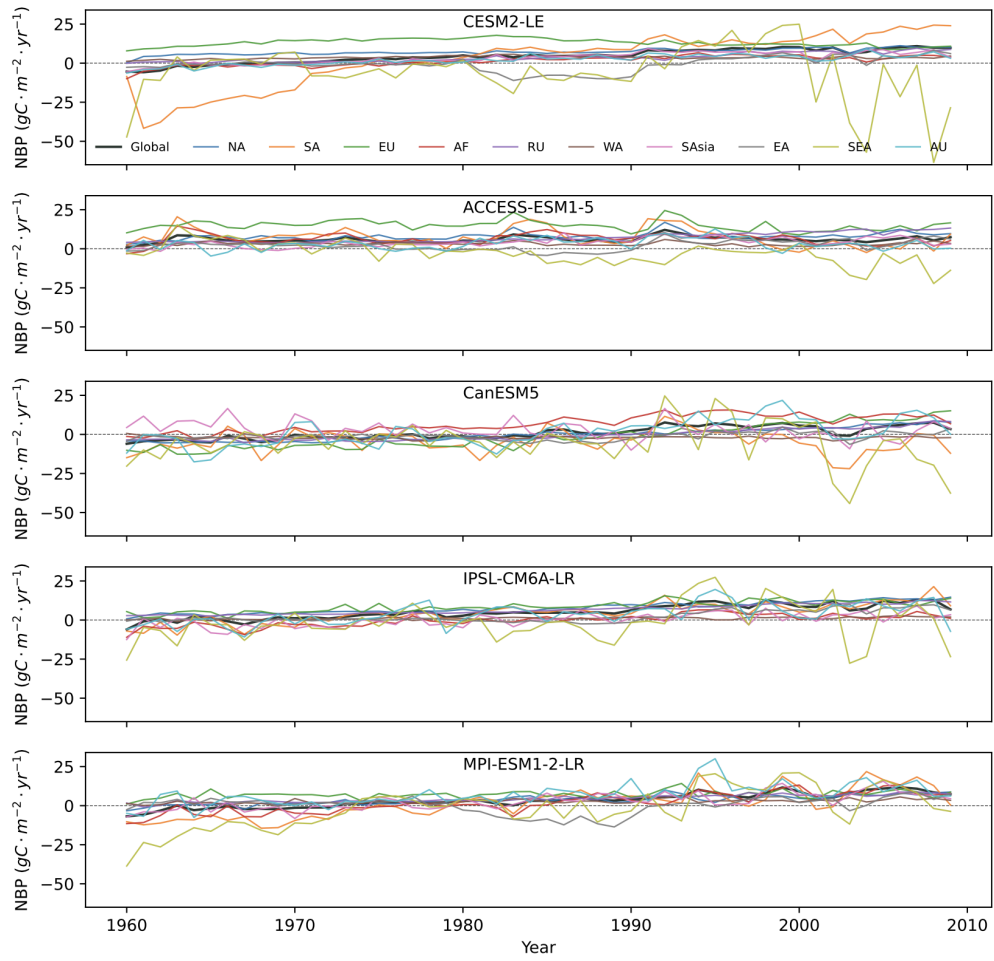


Figure A.1. Time series of NBP ensemble mean from five Earth system models. The thick black line is the global ensemble mean, and the colored lines represent ensemble means for the 10 RECCAP-2 regions.

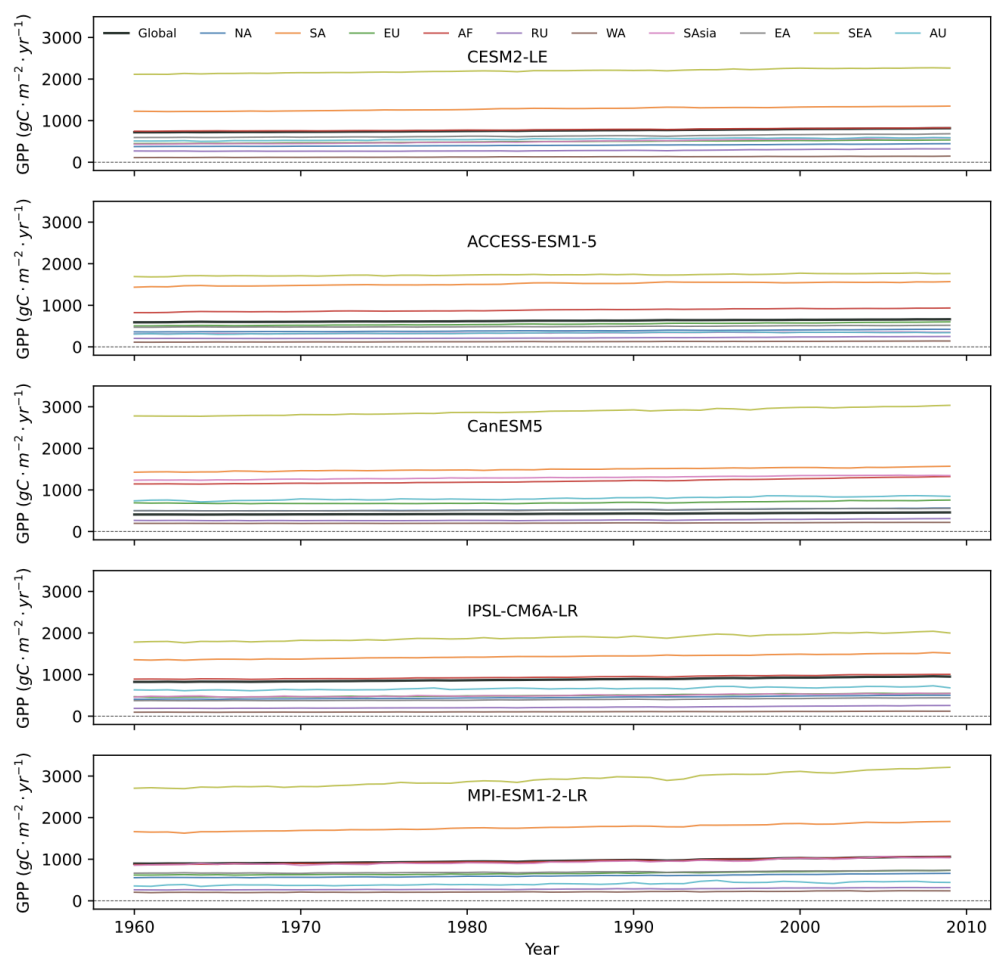


Figure A.2. Time series of GPP ensemble mean from five Earth system models. The thick black line is the global ensemble mean, and the colored lines represent ensemble means for the 10 RECCAP-2 regions.

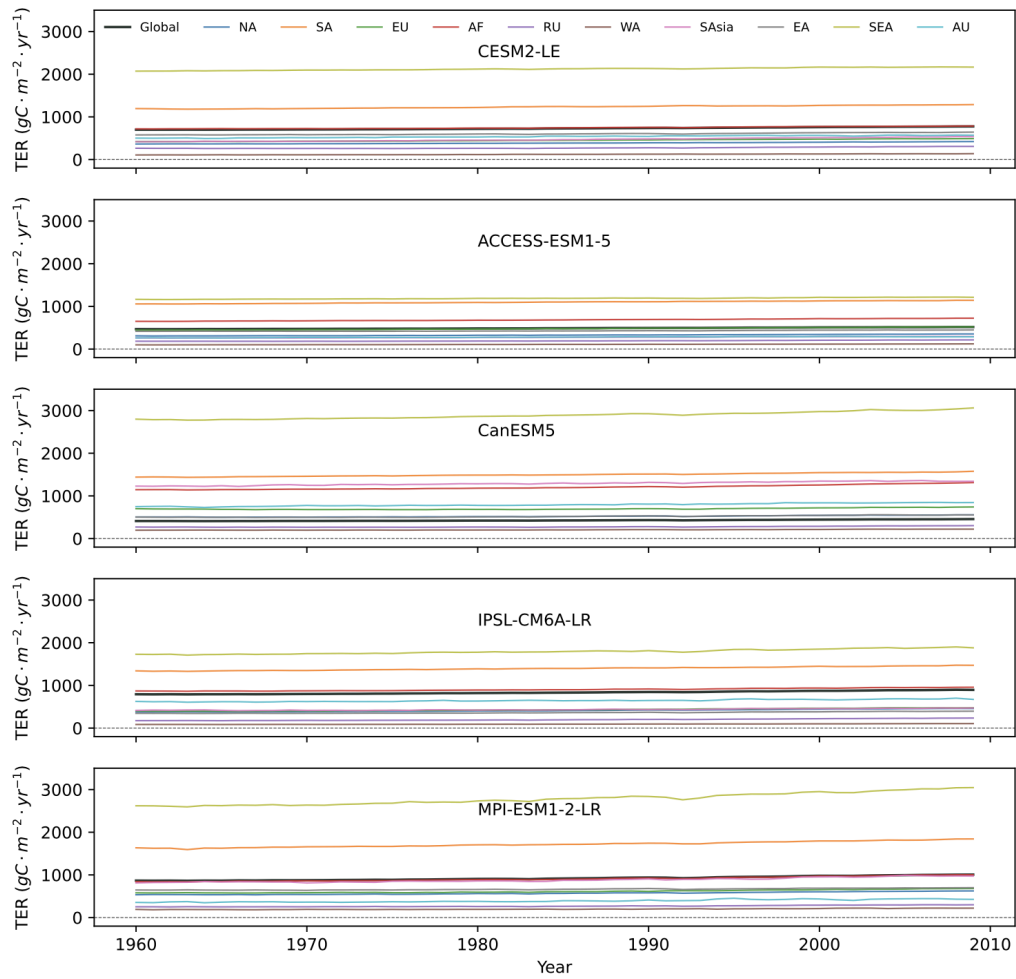


Figure A.3. Time series of TER ensemble mean from five Earth system models. The thick black line is the global ensemble mean, and the colored lines represent ensemble means for the 10 RECCAP-2 regions.

Accordingly, we add two sentences to line 221:

“We first check the ensemble mean of NBP in global scale and 10 RECCAP-2 regions (Appendix A Fig. A1). The trends of ensemble mean in NBP subjects to larger variability in regional than in global scales, particularly in Southeast Asia and Africa.”

### 3. The ensemble size looks arbitrary.

The four CMIP6 simulations are originally retrieved from <https://esgf-node.llnl.gov/projects/cmip6/>, and pre-processed by Brunner et. al. (2020), including aggregate to resolution of  $2.5^{\circ} \times 2.5^{\circ}$  and resampled to annual mean. The number of simulations for each model from pre-processed data from Brunner et al. (2020). CESM2-LE is downloaded from <https://rda.ucar.edu/datasets/d651056/#>. They include 90 simulations of NBP. We agree with the reviewer: The different large ensembles come in different sample sizes and therefore look somewhat arbitrary, as they are provided by different modeling centers who make their own decision regarding sample sizes; yet this is unfortunately nothing we could change or streamline for the paper.

#### 4. CESM2-LE burned area influenced ensemble mean?

CESM2-LE includes 90 simulations: the first 50 are CMIP6 simulations and the remaining 40 are SMBB (biomass burning) simulations. We compared the ensemble mean of the first 50 CMIP6 simulations with that of the 40 SMBB simulations, and found the difference in global scale negligible. The figure below shows that the two types of simulations are well mixed, with similar ensemble means.

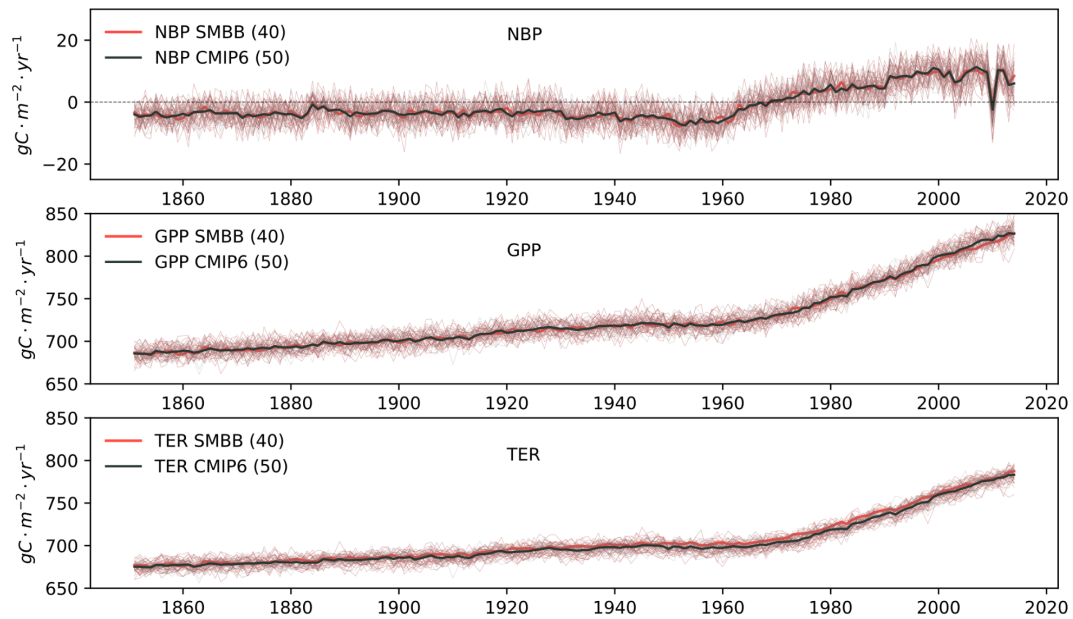


Figure 2:

Suggest adding a multi-ensemble mean to each panel – this suggestion is for all Figures. For this figure is it showing TOE = years post 1960? can this be clear in the caption please. Additionally what does ‘the spatial domains are slightly changed’ mean?

We included three figures of multi-ensemble mean (global scale and 10 RECCAP-2 Regions) to supplementary Figure A.1 to A.3, for NBP, GPP, and TER respectively. As is shown above.

Figure 2 shows ToE years after 1960. We added this to the caption:

“Figure 2. ToE of NBP on a global scale and across 10 RECCAP-2 regions, under historical simulations of five ESM large ensembles. Note that ToE is the years detectable after 1960, and is calculated with signal period of 1960-2009...”

“The spatial domain are slightly changed” means that after we aggregated the RECCAP-2 map from  $0.5^\circ \times 0.5^\circ$  to  $2.5^\circ \times 2.5^\circ$ , the aggregated map looks a bit different as in panel a, Figure 2.

Figure 3 – Is this really resolution? i.e. is it the same area regridded to a different resolution or is it a different spatial domain?

Yes, they are all spatial resolutions. We aggregated the global data per pixel to different resolutions, then calculated the ToE per pixel. Figure 3 shows the ToEs of all global pixels.

We add these explanations to the caption of Figure 3:

“Spatial effect in NBP historical simulations across five ESM large ensembles. The distribution of ToE (years after 1960) is shown for varying spatial resolutions. We aggregate the global data per pixel to different resolutions, then calculate the ToE per pixel. The line within each box indicates the median. Note that all signals are in absolute values, so the calculated ToE are all positive.”

Figure 5: there seem to be missing bars where only the light or shaded box is there? What happened to the other box?

As we stated in method section 2.4:

“Note that only the calculated signal (regression slope) with significance value  $P < 0.05$  is selected.”

We only present the ToE with signal significance  $P < 0.05$ . The missing bars have a signal that is not significant ( $P > 0.05$ ), which implies the linear slope (signal trend) has not emerged, so we did not include them.

We explained this further in the caption of Figure 6 (old Figure 5):

“In cases where both boxes are missing, the respective signal is not available (no significance of linear trend slope), or the ToEs are longer than 150 years. ~~the respective simulation was not available.~~”

A lot of content discussed is in the appendix figures – I wonder if more should be in the main text

We added Figure A.9 (Heat map of ToE, noise, and signal of NBP under future scenarios.) up as new Figure 5. Note that since we changed the noise period from 2020-2050 to 2020-2070, so the new Figure 5 with slightly different values.

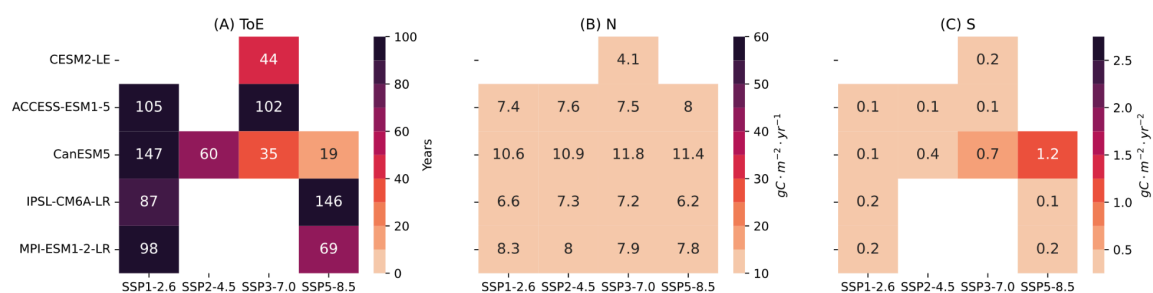


Figure 5. Heat map of ToE, noise, and signal of NBP under future scenarios.



## 6. Paper is missing a discussion section

Thanks for pointing out, we discussed mostly at the end of each section and also in conclusions. Here, we added a formal discussion:

In line 260:

“We found global scale takes shorter time to detect long-term trends induced by anthropogenic effects than at smaller scales, with ToE increasing for smaller domains as reported by Lombardozzi et al. (2014), though their study used fewer models and less than 10 simulations. ~~However, this spatial delay does not always apply, such as in regions like Russia, the interannual natural variability in three CMIP6 models is lower than in global scale.~~ For regional NBP, except large internal variability, the larger variability of the signal trend also contributes, likely reflecting different regional climate drivers (e.g., fires, decadal internal variability, land use changes (Deser et al., 2012b; Canadell et al., 2021; Eyring et al., 2021; Mercado et al., 2009). A few regions, however, show shorter ToEs than the global scale. For example, in Russia, CanESM5 and IPSL-CM6A-LR simulate relatively small noise and stronger signal trends, leading to shorter ToEs. This maybe relate to the sparsely distributed ecosystems in Russia, which are less sensitive to changes in climate drivers.

It should also be noted that the large interannual variations in NBP largely arise from variability in GPP and respiration. Future studies can focus on comparing regional GPP and respiration variabilities and attributing them to different drivers. Regional ecosystems are more influenced by precipitation than by temperature (Jung et al., 2017), implying that large variations in regional NBP maybe driven by precipitation. However, the detection of anthropogenic signals in precipitation is less robust and delayed than in temperature (Doblas-Reyes et al., 2021). Thus, detecting anthropogenic signals in precipitation and assessing how they influence regional ecosystem activity, may provide a cleaner and earlier detection of anthropogenic signals in regional land carbon sinks.”

In line 298:

“The divergence of NBP across models is much larger in future scenarios than in historical simulations. While models are inherently different, these differences in the historical period may be amplified in future scenarios due to: (a) the increasing influence of CO<sub>2</sub> fertilization and land use change (van Vuuren et al., 2011, Ciais et al., 2013, O’Neil et al., 2016, Christensen et al., 2018, Arora et al., 2020, Lee et al., 2021); and (b). Rising temperature and more frequent extreme events, which may intensify differences in climate-carbon feedbacks among models (Friedlingstein et al., 2014, Hewitt et al., 2016, Seneviratne et al., 2021).”

We also added one paragraph in conclusion (the last second paragraph), in line 368:

“The emergence approach used in this study is sensitive to the choice of the reference period, and signal and noise length. In addition, the signal trend can be biased by the large regional variability, and the noise can enhance or amplify external natural forcing in

observations (Lombardozzi et al., 2014, Bonan et al., 2021). A better understanding of regional ecosystem responses to anthropogenic signals, and improved methods may help improve the detected emergence time.”

7. Section 3.1 – the noise could be easily quantified rather than discussed qualitatively

Thanks for pointing out this very critical issue.

1) we quantified the percentage of the noise change that contributed to the ToE change in the manuscript, both changes are compared with global scale. We added the figures according to the manuscript as Figure A.5, A.7, and A.9 for historical RECCAP-2 region contribution (compared with global scale), as well as Figure for future contribution (compared with historical global scale).

2) We added one paragraph in section 2.4 (line 194) to explain the methods

3) We rewrote some parts in Section 3.2 and 3.3 (line 221-290), as well as the conclusion to include the quantified N and S contributions.

Below are details:

1) Quantify N and S contribution

Due to  $ToE = 2N * (1/S)$ , both N and S have changed in regional scale, compared with global scale.

$$ToE_{global} = 2N_{global} * (1/S_{global})$$

$$ToE_{region} = 2N_{region} * (1/S_{region})$$

We calculate the contribution of N and S by taking the Log of two sides of the equation.

$$\log(ToE_{global}) = \log(2N_{global}) + \log(1/S_{global}) \quad (1)$$

$$\log(ToE_{region}) = \log(2N_{region}) + \log(1/S_{region}) \quad (2)$$

Using equation 2 minus equation 1.

$$\begin{aligned} \log(ToE_{region}) - \log(ToE_{global}) &= \log(2N_{region}) - \log(2N_{global}) \\ &\quad + \log(1/S_{region}) - \log(1/S_{global}) \quad (3) \end{aligned}$$

Then we can calculate the percentage of log changes in the noise and in the signal that contribute to log ToE change.

Below is the calculated percentage of contribution from N and S in historical simulations, here we only show the values with N/S change as the dominant contributor.

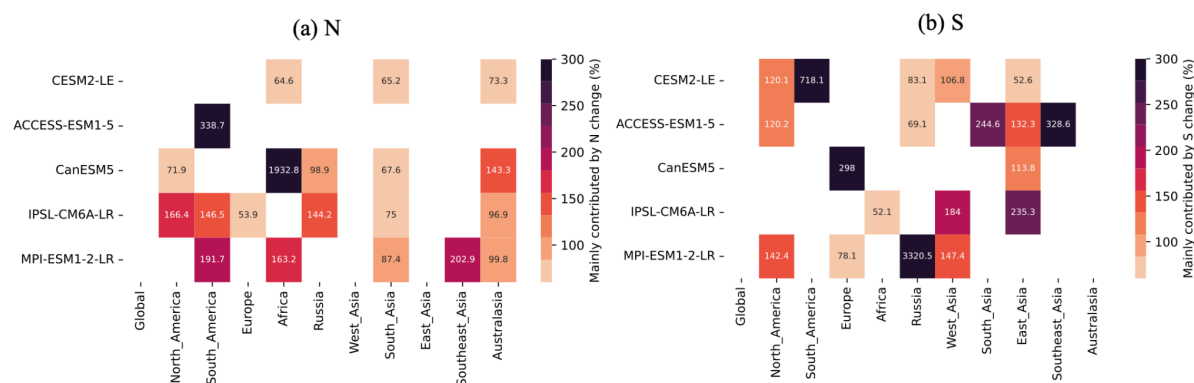


Figure A5. Contribution of N and S to ToE changes in NBP for each RECCAP-2 region, compared with global scale, in historical simulations. Note that only values where changes in N or S are the dominant contributor are shown.

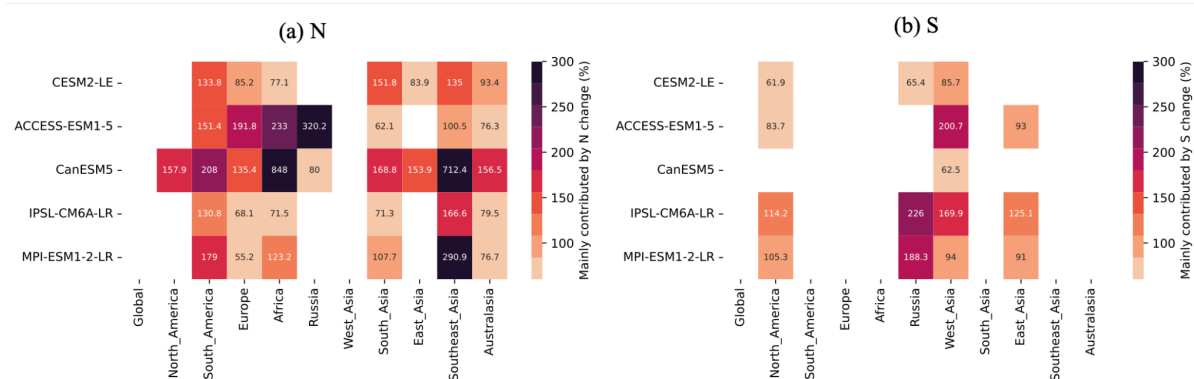


Figure A7. Contribution of N and S to ToE changes in GPP for each RECCAP-2 region, compared with global scale, in historical simulations. Note that only values where changes in N or S are the dominant contributor are shown.

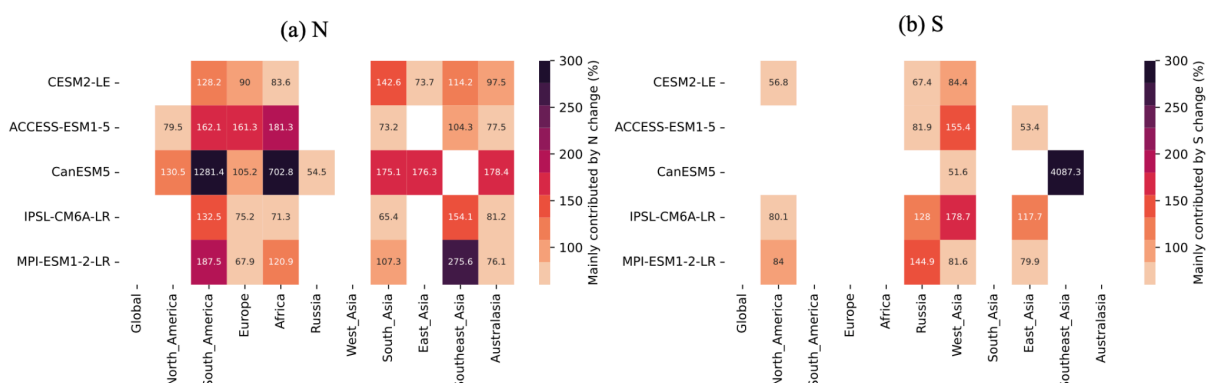


Figure A9. Contribution of N and S to ToE changes in TER for each RECCAP-2 region, compared with global scale, in historical simulations. Note that only values where changes in N or S are the dominant contributor are shown.

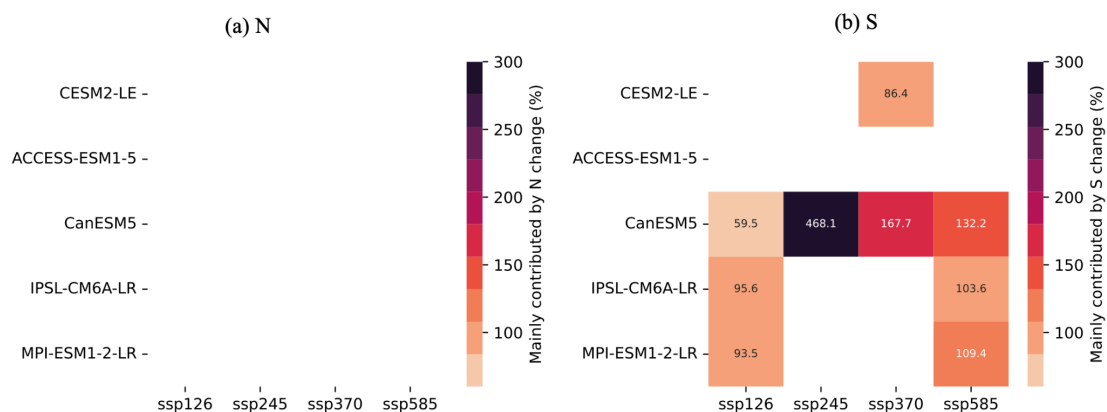


Figure A14. Contribution of N and S to ToE changes in NBP under each future scenario, compared with the historical global scale. Note that only values where changes in N or S are the dominant contributor are shown.

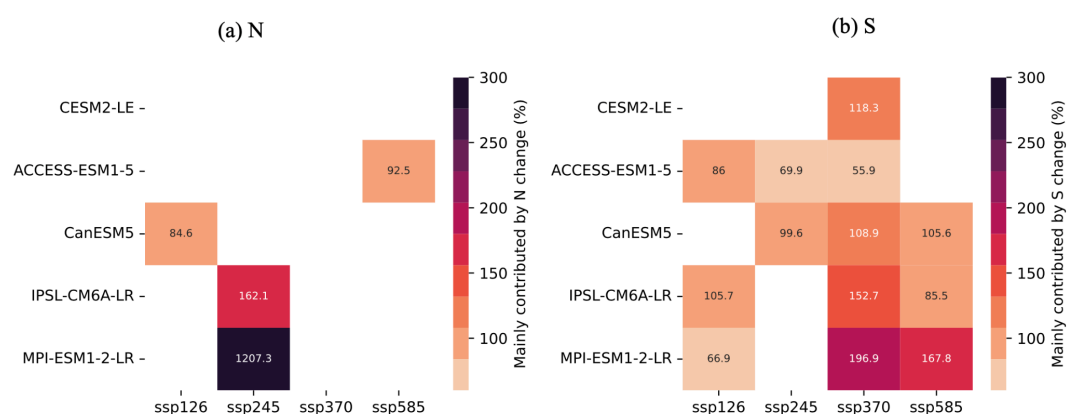


Figure A16. Contribution of N and S to ToE changes in GPP under each future scenario, compared with the historical global scale. Note that only values where changes in N or S are the dominant contributor are shown.

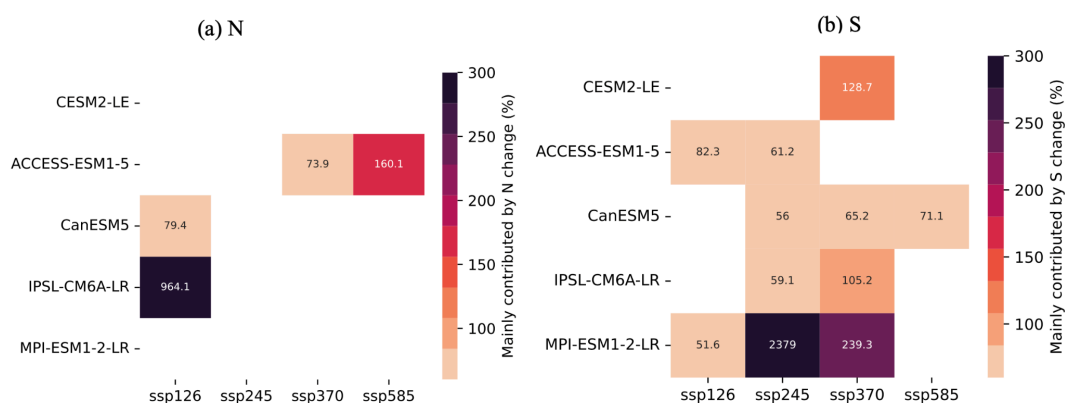


Figure A18. Contribution of N and S to ToE changes in TER under each future scenario, compared with the historical global scale. Note that only values where changes in N or S are the dominant contributor are shown.

- 2) Add a paragraph to Section 2.4, describing the method used to calculate the contribution of changes in N and S to changes in ToE. In line 194:

“We also calculated the contribution of N and S to ToE changes in each RECCAP-2 region, compared with the global scale, for NBP, GPP, and TER. The Eq. (5) is the logarithmic (in base e) form of Eq. (3). We first calculate the log change in global and regional scales. Then calculate the differences between each region to global scale (Eq. (6)).

$$\ln(\text{ToE}) = \ln(2N) + \ln(1/S) \quad (5)$$

$$\ln(\text{ToE}_{\text{region}}) - \ln(\text{ToE}_{\text{global}}) = \ln(2N_{\text{region}}) - \ln(2N_{\text{global}}) + \ln(1/S_{\text{region}}) - \ln(1/S_{\text{global}}) \quad (6)$$

The contribution of changes in N and S are:

$$N_{\text{contri}} = 100\% * (\ln(2N_{\text{region}}) - \ln(2N_{\text{global}})) / (\ln(\text{ToE}_{\text{region}}) - \ln(\text{ToE}_{\text{global}})) \quad (7)$$

$$S_{\text{contri}} = 100\% * (\ln(1/S_{\text{region}}) - \ln(1/S_{\text{global}})) / (\ln(\text{ToE}_{\text{region}}) - \ln(\text{ToE}_{\text{global}})) \quad (8)$$

Note that for future scenarios, we substitute the region's value to each future scenario's value.”

- 3) Change the results and discussion accordingly.

Section 3.2, Line 221-255:

“ We first check the ensemble mean of NBP in global scale and 10 RECCAP-2 regions (Appendix A Fig. A1). The trends of ensemble mean in NBP subjects to larger variability in regional than in global scales, particularly in Southeast Asia and Africa. Comparing noise (N) and signal (S) (Appendix A Fig. A4 and Eq. (3)), we found that the variation in detection time across models is mainly due to differences in year-to-year variability, as well as differences in signal trends.

In most of the 10 RECCAP-2 domains, ToE detection in NBP takes longer than at the global scale (Fig. 2B). The exception is ACCESS-ESM1-5, where a flat global trend after the 1960s causes a much longer detection time. We calculate the contribution of changes in N and S to regional ToE, compared with historical global scale NBP (Appendix A Fig. A5). In regions such as South Asia and Australasia, ~~The regional noise tends to be larger than at the global scale,~~ the longer regional ToE is mainly due to larger regional noise. While in other regions such as East Asia, the longer regional ToE is mainly attributable to smaller regional S. These effects ~~leading to delayed~~ delay the detection of anthropogenic signals. ~~We refer to this difference~~, a phenomenon we refer to as "spatial delay".

The spatial delay can be explained by reduced noise from internal climate variability when aggregating fluxes globally, while the signal trend is smoothed and enhanced ~~the signal remains relatively stable~~ (Appendix A Fig. A.4). This is a well-known pattern in detection and emergence studies in the climate literature (Mahlstein et al., 2011, Lehner

et al., 2017). Southeast Asia and Australasia, show the largest noise, and therefore also the longest ToE across the 10 RECCAP-2 domains. However, the spatial delay does not apply everywhere. In Russia, models like ~~ACCESS-ESM1-5~~, CanESM5, and IPSL-CM6A-LR show shorter detection time compared to the global scale. This is due to apparent smaller noise in Russia of the ~~two~~ ~~three~~ models, compared with other regions and global scale Fig. A.4).

We then evaluate the ToE for GPP and TER (Fig. 2C, D), both show similar patterns in detection time, and the relative importance of noise and signal at regional scale (Appendix A Fig. A.3, A.4}). Globally, it takes around 8 to 13 years to detect anthropogenic signal in GPP and 6 to 10 years for TER. As found for NBP, both GPP and TER show a spatial delay from global scale to regional scale. ~~For GPP and TER, longer regional ToEs are mostly due to larger regional noise rather than weaker signal trends, except in West Asia, where it is mainly driven by an apparent weaker signal trend. Australasia in GPP and TER both generally have~~ ~~generally~~ the longest detection time, due to ~~higher~~ noise compared to ~~lower~~ signal levels, indicating higher regional internal climate variability. In both GPP and TER, South America and Southeast Asia experience high levels of noise and signal, while West Asia has relatively low levels of noise and signal."

Line 254:

"To further analyse the spatial delay effect, we calculate the distribution of pixel based ToE for NBP, GPP, and TER under varying resolutions in the historical simulations. For NBP, as the resolution becomes coarser, the spread of the ToE distribution decreases substantially (Fig. 3), though the median remains similar. This ~~might be due to~~ ~~supports~~ the noise reduction by spatial aggregation through offsetting internal climate variability (Lombardozzi et al., 2014), ~~as well as smoother signal trend~~ (Fig. 3). A similar pattern is observed in GPP and TER, where aggregation reduces the spreads of ToE without substantially altering the medians (Appendix A Fig. A.10 and A.11)."

Line 302:

"In future scenarios, it takes longer to detect the anthropogenic signal in NBP, when compared to historical simulations. This delay is ~~likely~~ ~~mainly~~ due to the small anthropogenic signal caused by the compensation effect of GPP and TER, ~~and~~. In ~~addition~~, larger noise levels that driven by more frequent extreme events in a warming climate (Arias et al., 2021) ~~, which~~ amplifies year-to-year variations in the land carbon sink."

We also added this to the conclusion in line 356: "At the regional scale, ToE is longer, which might be due to larger noise from natural climate variability in most regions, ~~as well as detected weaker signal trends.~~"

Line 376: "This approach ~~can be~~ ~~has been~~ applied in observations ~~to facilitate~~ and ~~shows an~~ early detection of anthropogenic signal in global carbon cycle variability."

## Minor comments

Title should be 'the anthropogenic signal'

Thanks, we added "the" to the title.

## Intro

Line 1 'the' atmospheric CO<sub>2</sub>

Thanks, we added "the" to line 1: "..., driven by the increasing atmospheric CO<sub>2</sub>...", and also in line 24: "...mainly driven by the increasing atmospheric CO<sub>2</sub>..."

Line 2 – this is not driven by climate change, more likely it is driven by processes related to climate change please be specific here

Thanks for pointing out this, we rewrite line 2 as: "...driven by the increasing atmospheric CO<sub>2</sub> concentration and physical processes influenced by climate change."

line 6 replace by with from

We have replaced "by" with "from" in line 6: "and future scenarios (2016-2020) by from Earth system models."

Line 13 – it is odd to have a 1 sentence paragraph – this sentence is also unclear can you please rephrase

Thanks, we have rephrased in line 13 as below:

"Secondly, future scenarios show delayed signal detection compared to historical trends. ~~due to a slow-down of the increasing net land carbon sink in response to emission mitigation, compared to the higher natural variability~~ This delay is mainly due to weaker anthropogenic signal trends rather than stronger natural variability. The weaker signal reflects the slow-down of the increasing net land carbon sink in response to emission mitigation."

Line 16 – remove 'allows to' replace shorten with 'shortens'

We changed line 17: "This approach ~~allows to~~ substantially shortens the detection time "

Line 18 – replace on with 'of'

We changed in line 19: "our ability to detect long-term trends ~~on~~ of land carbon sink variability"

Line 30 – should be 'temperature, and precipitation'

We changed line 30: "such as temperature, ~~and~~ precipitation"

Line 33 – replace by with 'on'

Thanks we replaced in line 34: “The long-term trends of the global carbon cycle are superimposed ~~by~~ with substantial year-to-year”.

Line 33 – the paragraph starting on this line is hard to parse, I suggest rewriting

Thanks, we rewrote as below in line 34:

“The long-term trends ~~of in~~ the global carbon cycle are superimposed ~~by~~ with substantial year-to-year variations (Piao et al., 2020). These variations mostly ~~stem~~ originate from natural processes, ~~such as~~ including internal climate variability, ~~variability~~ –fluctuations across ~~on~~ a continuum of time scales ~~unrelated to external effects~~– as well as from ~~influences from~~ natural external forcings such as volcanic eruptions and solar radiation (Deser et al., 2012b; Canadell et al., 2021; Eyring et al., 2021; Mercado et al., 2009; Zhang et al., 2021). Internal climate variability is often ~~referred to~~ regarded as irreducible noise ~~in~~ within the signal of long-term forced climatic trends. It ~~and~~ arises from internal atmospheric dynamics and ~~interactions between the~~ from atmosphere–oceans interactions (Deser et al., 2012a, 2020; Lehner et al., 2017; Bonan et al., 2021). ~~Internal climate variability emerges as short-term weather events and long-term low climate variability patterns like~~ Such variability manifests both as short-term weather events and as low-frequency climate patterns, such as the El Niño–Southern Oscillation (ENSO), ~~which are known to influenced~~ strongly influence global land carbon sink variations through associated ~~local~~ changes in temperature and precipitation.”

Line 47 – can differ substantially from what? Please be specific

We change line 48 to: “First, internal climate variability can differ substantially in various simulations under the same external forcings,”

Line 50 – a good example of this is found in this paper:

<https://egusphere.copernicus.org/preprints/2024/egusphere-2024-3684/>

We added this citation in line 52: “...range of physically plausible internal climate variability. Moreover, internal climate variability is sensitive to the choice and length of the study period (Kumar et al., 2016; Doblas-Reyes et al., 2021; Maher et al., 2024).”

Line 53 – I would end this sentence with ‘orividing a challenge because ...’

In line 54:

~~“Short-term observational records may not capture the full dynamics of internal climate variability~~ This makes it challenging to capture the full dynamics of internal climate variability, due to the limited length observation records.”

Line 65 – this section is missing two key references:

<https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019GB006453>,

<https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011gl050087>



Thanks, we added accordingly in line 68: “Based on such large ensembles of ESM simulations, the “time of emergence (ToE)” can be determined as the time required for an external perturbed signal (anthropogenic-caused climate change) to become larger than the amplitude of natural variations (Hawkins and Sutton, 2012; Lehner et al., 2017; Schlunegger et al., 2020; Bonan et al., 2021).”

Line 73 – odd use of the word ‘need’

We removed “need” in line 75: “driven by anthropogenic perturbations ~~need~~ to be detected”

## Methods

Line 81 – in this section please add the number of ensemble members from each model

Thanks, we added below in line 85:

“The models selected include the CESM2-LE with 90 simulations (Danabasoglu et al., 2020; Rodgers et al., 2021) and four models in CMIP6 (Eyring et al., 2016; Brunner et al., 2020): ACCESS-ESM1-5 with 38 simulations (Ziehn et al., 2020), CanESM5 with 40 simulations (Swart et al., 2019), IPSL-CM6A-LR with 33 simulations (Boucher et al., 2020), and MPI-ESM1-2-LR with 41 simulations (Mauritsen et al., 2019).”

Line 82 – CESM2 is part of CMIP6

Here we use CESM2-LE dataset downloaded from NCAR directly:

“<https://rda.ucar.edu/gsearch/dataset-search/?q=CESM2-LE+>”, this dataset is a bit different from pre-processed CMIP6 by Brunner 2020, they only included 12 simulations in CMIP6.

Line 84 – this is the correct citation for MPI-ESM1-2-LR large ensemble:

<https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2023MS003790>

Thanks, but the simulations from MPI (MPI-ESM1-2-LR) and the other three CMIP6 models used in this study were pretreated by Brunner et al., 2020 (<https://zenodo.org/records/3734128>). Which means they are published before 2020. The citation you suggested is for the newest MPI (MPI-GE CMIP6) model and is published in 2023.

Line 125 – this is for CESM2 what about the other models?

We added other models in line 121: “Note that CESM2-LE only includes one future scenario (SSP3-7.0), and other models included all four future scenarios.”

Line 133 – what about the SSPs

I explained SSPs in line 175, we have moved to line 136: “In the future scenarios, we calculate the ToE for NBP, GPP, and TER, with the signal period in 2020–2070 and the noise period in 2020–2070 (with the ensemble mean removed).”

Line 134 – how do you calculate N? do you pool the data? Do you take the STD of each member and average? Please provide this detail

We added in line 132: “For the calculation of N, we first gather a data pool including all ensemble residuals from simulations in the selected period, then mix the data from all years in the selected period together and calculate the standard deviation.”

Line 137 – please add what 2N/S means in terms of significance of the emergence

We already explained in line 137: “ToE (Eq. (3)) represents the time needed for the anthropogenic perturbed signal to become larger than the amplitude of the noise (Bonan et al., 2021).”

”Line 147 – would 500mb geopotential height be a better proxy for circulation than SLP?

Yes, but sea level pressure (SLP) is physically more close to land ecosystems, and can reflect the direct influence of internal variability to land ecosystems.

Line 168 – you interchange ‘ToE’ and ‘time to detect’ this is confusing please stick to one terminology

We changed in line 177: “We analyse the ~~time to detect~~ ToE of anthropogenic perturbed signal in NBP, GPP”

## Results:

Line 213 – similar patterns to what? Each other?

Thanks, we changed to in line 210: “...~~which~~ both show similar patterns in detection time...”

Line 230 – grammar please fix ‘takes shorter’

Thanks, we changed to in line 261: “We found ~~that on a global scale, it takes shorter to detect~~ global scale takes shorter time to detect long-term trends induced by anthropogenic effects than at smaller scales,

Line 232 – please link with ‘such as’

Thanks, we changed to in line 264: “ ~~However, this spatial delay does not always apply, such as in regions like Russia, the interannual natural variability in three CMIP6 models is lower than in global scale.~~ A few regions, however, show shorter ToEs than the global scale. For example, in Russia, CanESM5 and IPSL-CM6A-LR simulate relatively small noise and stronger signal trends, leading to shorter ToEs.”

Line 233/234/236 – please reference the figures that you are discussing in this section

Thanks, we added.

Line 233 – remove ‘is apparently’

We removed in line 264: “ ~~However, this spatial delay does not always apply, such as in regions like Russia, the interannual natural variability in three CMIP6 models is lower than in global scale.~~ A few regions, however, show shorter ToEs than the global scale. For example, in Russia, CanESM5 and IPSL-CM6A-LR simulate relatively small noise and stronger signal trends, leading to shorter ToEs.”

Line 240 – is this statement true – I do not make the same conclusion from the Figure

You mean in Section 3.3 the first paragraph line 275: “We examine the time series of NBP under various future scenarios from 2016 to 2100 (Fig. 4). While NBP trends ~~are similar in the two low CO2 emission scenarios (SSP1-2.6 and SSP2-4.5), the highest emission scenario (SSP5-8.5) shows~~ show large deviations across models.”

### **Conclusions:**

Line 293 – you should be able to quantify if this is due to larger noise or not from your results not just speculate.

Thanks for the reviewer’s concrete suggestion. Here we quantified the contribution of noise and signal to the regional increased ToE, in NBP, GPP, and TER. Please check our answers above in question 7.

Line 315 – can be detected how? There are no observations in the paper

Thanks, we have added the observed results as Figure 7 (as is shown above)

## Additional changes:

1. Line 7-8: Our results show that, firstly, the anthropogenic signal in the global net land carbon sink emerges from 26 to 66 years in the period 1960–~~2009~~ ~~2019~~
2. Line 10: “Furthermore, we find that long-term trends of net land carbon sink on most regional scales take at least 20 years more to emerge, due to larger contributions from internal climate variability ~~and detected weak signals~~ at smaller scales.”
3. Line 65-66: “The externally perturbed signal (~~dominated by~~ anthropogenic signal) emerges as the ensemble mean, that is, a deterministic signal.”
4. Line 68-69: “the “time of emergence (ToE)” can be determined as the time required for an external perturbed signal (~~mostly~~ anthropogenic-caused climate change) to become...”
5. Line 75-78: “Here, we evaluate how long ~~it takes for~~ long-term trends in the global land carbon sink—~~primarily, mostly~~ driven by anthropogenic perturbations—to be detected ~~from local to global scales~~ at different spatial scales. ~~by estimating the To~~ achieve this, we estimate the ToE in ESM simulations under historical and future scenarios. ”
6. All “circulation induced” to “circulation-induced”
7. Line 155-156: “In our model, the sea level pressure (SLP) field is used as a predictor of NBP variability and as a proxy of ~~circulation-induced variability predict the~~ circulation-induced variability.”
8. Line 156-157: “As a regularized linear regression ~~method~~, ridge regression allows for including full spatiotemporal dynamics of circulation variations while overcoming multicollinearity and overfitting, ~~which typically arise normally raising~~ from a large number of predictors and relatively short study period.”
9. Line 160: “Select pixel based time series of global SLP, to be used later for predicting ~~predict~~ global carbon cycle variability.”
10. Line 162-163: “ 2) Select the time series representing global ~~land carbon~~ ~~eyele~~ variability; here, this corresponds to the global annual NBP with the ensemble mean removed. ”
11. We remove the definition of SSP, since the detailed discussion can be found in IPCC report in line 109.

“The future scenario simulations are modeled under different Shared Socioeconomic Pathways (SSPs) for the period 2015-2100, based on varying levels of human-emitted CO<sub>2</sub> and mitigation efforts (Chen et al., 2021; Lee et al., 2021; O’Neill et al., 2016):.

- ~~1. SSP1-2.6, CO<sub>2</sub> emissions decrease and reach net zero by 2050. Global surface air temperature averaged over 2081-2100 is 0.5–1.5° higher than in 1995-2014 (Chen et al., 2021; Lee et al., 2021; O'Neill et al., 2016).~~
- ~~2. SSP2-4.5, CO<sub>2</sub> emissions remain steady until 2050, with global surface air temperature averaged over 2081-2100 is 1.2–2.6° higher than in 1995-2014 (Chen et al., 2021; Lee et al., 2021; O'Neill et al., 2016).~~
- ~~3. SSP3-7.0, without additional climate policies, CO<sub>2</sub> emissions doubled by 2100, leading to the global surface air temperature averaged over 2081-2100 increase of 2.0–3.7° relative to 1995-2014 (Chen et al., 2021; Lee et al., 2021; O'Neill et al., 2016).~~
- ~~4. SSP5-8.5 Without additional climate policies, CO<sub>2</sub> emissions doubled by 2050, with global surface air temperature averaged over 2081-2100 rising by 2.4–4.8° relative to 1995-2014 (Chen et al., 2021; Lee et al., 2021; O'Neill et al., 2016)."~~

12. Other values are slightly changed due to the changes of noise period from 2020-2050 to 2020-2070, in future scenarios. Also, the ToE values are cut until 150 years.

1. Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., Krasting, J. P., Law, R. M., Lawrence, D. M., Lenton, A., Lindsay, K., Pongratz, J., Raddatz, T., Séférian, R., Tachiiri, K., Tjiputra, J. F., Wiltshire, A., Wu, T., and Ziehn, T.: Carbon-concentration and carbon-climate feedbacks in CMIP6 models and their comparison to CMIP5 models, *Biogeosciences*, 17, 4173–4222, <https://doi.org/10.5194/bg-17-4173-2020>, 2020.
2. Bonan, G., Lombardozzi, D. L., and Wieder, W. R.: The signature of internal variability in the terrestrial carbon cycle, *Environ. Res. Lett.*, 16, 034 022, <https://doi.org/10.1088/1748-9326/abd6a9>, 2021.
3. Brunner, L., Hauser, M., Lorenz, R., and Beyerle, U.: The ETH Zurich CMIP6 next generation archive: technical documentation., Zenodo, <https://doi.org/10.5281/zenodo.3734128>, 2020.
4. Canadell, J. G., Monteiro, P. M. S., Costa, M. H., Cotrim da Cunha, L., Cox, P. M., Eliseev, A. V., Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Patra, P. K., Piao, S., Rogelj, J., Syampungani, S., Zaehle, S., and Zickfeld, K.: Global Carbon and other Biogeochemical Cycles and Feedbacks. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/9781009157896.007>, chapter 5, p732, 2021.
5. Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, B. R., Piao, S., and Thornton, P.: Carbon and Other Biogeochemical Cycles. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_all\\_final.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_all_final.pdf), last access: 20 March 2022, 2013.
6. Christensen, P., Gillingham, K., and Nordhaus, W.: Uncertainty in forecasts of long-run economic growth, *Proceedings of the National Academy of Sciences*, 115, 5409–5414, <https://doi.org/10.1073/pnas.1713628115>, 2018.
7. Cleveland, W., Grosse, E., and Shyu, W. M.: *Local regression models. Statistical models in S*, Chapman Hall, ISBN 9780412830402, 309–376, 1991.
8. Deser, C., Knutti, R., Solomon, S., et al.: Communication of the role of natural variability in future North American climate, *Nat. Clim. Change*, 2, 775–779, <https://doi.org/10.1038/nclimate1562>, 2012b.
9. Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., Kay, J. E., Knutti, R., Lovenduski, N. S., Marotzke, J., McKinnon, K. A., Minobe, S., Randerson, J., Screen, J. A., Simpson, I. R., and Ting, M.: Insights from Earth system model initial-condition large ensembles and future prospects, *Nature Climate Change*, 10, 277–286, <https://doi.org/10.1038/s41558-020-0731-2>, 2020.
10. Doblas-Reyes, F. J., Sörensson, A. A., Almazroui, M., Dosio, A., Gutowski, W. J., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W. T., Lamptey, B. I., Maraun, D., Stephenson, T. S., Takayabu, I., Terray, L., Turner, A., and Zuo, Z.: Linking Global to Regional Climate Change. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/9781009157896.012>, IPCC,2021, chapter 10, p1363-1512, 2021.

11. Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P. J., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y.: Human Influence on the Climate System. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/9781009157896.005>, chapter 3, p425, 446, 2021.
12. Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks, *Journal of Climate*, 27(2), 511–526, <https://doi.org/10.1175/JCLI-D-12-00579.1>, 2014.
13. Hawkins, E. and Sutton, R.: Time of emergence of climate signals, *Geophysical Research Letters*, 39, <https://doi.org/10.1029/2011GL050087>, 2012.
14. Hewitt, H. T., Roberts, M. J., Hyder, P., Graham, T., Rae, J., Belcher, S. E., Bourdallé-Badie, R., Copsey, D., Coward, A., Guiavarch, C., Harris, C., Hill, R., Hirschi, J. J.-M., Madec, G., Mizielinski, M. S., Neiningner, E., New, A. L., Rioual, J.-C., Sinha, B., Storkey, D., Shelly, A., Thorpe, L., and Wood, R. A.: The impact of resolving the Rossby radius at mid-latitudes in the ocean: results from a high-resolution version of the Met Office GC2 coupled model, *Geoscientific Model Development*, 9, 3655–3670, <https://doi.org/10.5194/gmd-9-3655-2016>, 2016.
15. Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y.-P., Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly global land CO<sub>2</sub> sink changes to temperature, *Nature*, 541, 516–520, <https://doi.org/10.1038/nature20780>, 2017.
16. Lan, X., Tans, P., and Thoning, K. W.: Trends in globally-averaged CO<sub>2</sub> determined from NOAA Global Monitoring Laboratory measurements., <https://doi.org/10.15138/9N0H-ZH07>, version 2025-09, 2025.
17. Lee, J. Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J. P., Engelbrecht, F., Fischer, E., Fyfe, J. C., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S., and Zhou, T.: Future Global Climate: Scenario-Based Projections and Near- Term Information. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/9781009157896.006>, pp. 555, IPCC2021, Chapter 4, 2021.
18. Li, N., Sippel, S., Winkler, A. J., Mahecha, M. D., Reichstein, M., and Bastos, A.: Interannual global carbon cycle variations linked to atmospheric circulation variability, *Earth System Dynamics*, 13, 1505–1533, <https://doi.org/10.5194/esd-13-1505-2022>, 2022.
19. Lombardozzi, D., Bonan, G. B., and Nychka, D. W.: The emerging anthropogenic signal in land–atmosphere carbon-cycle coupling, *Nature Climate Change*, 4, 796–800, <https://doi.org/10.1038/nclimate2323>, 2014.
20. Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, D., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D., Raddatz, T., Stevens, B., and Marotzke, J.: The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability, *Journal of Advances in Modeling Earth Systems*, 11, 2050–2069, <https://doi.org/10.1029/2019MS001639>, 2019. Mercado, L. M., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., and Cox, P. M.: Impact of changes in diffuse radiation on the global land carbon sink, *Nature*, 458, 1014–1017, <https://doi.org/10.1038/nature07949>, 2009.

21. Milinski, S., Maher, N., and Olonscheck, D.: How large does a large ensemble need to be?, *Earth System Dynamics*, 11, 885–901, <https://doi.org/10.5194/esd-11-885-2020>, 2020.
22. O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, *Geoscientific Model Development*, 9, 3461–3482, <https://doi.org/10.5194/gmd-9-3461-2016>, 2016.
23. Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Luca, A. D., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., and Zhou, B.: *Weather and Climate Extreme Events in a Changing Climate.*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/9781009157896.013>, pp. 1513–1766, 2021.
24. Schlunegger, S., Rodgers, K. B., Sarmiento, J. L., Ilyina, T., Dunne, J. P., Takano, Y., Christian, J. R., Long, M. C., Frölicher, T. L., Slater, R., and Lehner, F.: Time of Emergence and Large Ensemble Intercomparison for Ocean Biogeochemical Trends, *Global Biogeochemical Cycles*, 34, e2019GB006453, <https://doi.org/https://doi.org/10.1029/2019GB006453>, e2019GB006453 2019GB006453, 2020.
25. van Vuuren, D. P., Edmonds, J., Kainuma, M., et al.: The representative concentration pathways: an overview., *Climatic Change*, 109, <https://doi.org/10.1007/s10584-011-0148-z>, 2011.
26. Zhang, Y., Ciais, P., Boucher, O., Maignan, F., Bastos, A., Goll, D., Lurton, T., Viovy, N., Bellouin, N., and Li, L.: Disentangling the impacts of anthropogenic aerosols on terrestrial carbon cycle during 1850–2014, *Earth's Future*, 9, e2021EF002 035, 2021.