Response to a review by Irina Melnikowa of "Assessing Earth system responses in deep mitigation scenarios with activity-driven simulation of carbon dioxide removal"

We thank Irina Melnikowa for reviewing our manuscript, and for her helpful and constructive comments. We have addressed all the points raised as detailed in our point-by-point response below. The review comments are in blue fonts, our responses in black, and changes to the manuscript text are highlighted in orange fonts. Line numbers refer to the revised version of our manuscript.

In this manuscript, Schwinger et al. discuss the challenges of representing carbon dioxide removal (CDR) in Earth system models (ESMs) and the resulting inconsistencies with integrated assessment model (IAM) simulations. The authors identify key issues, propose a framework to address them, and provide an example using NorESM2-LM (ESM) and REMIND-MAgPIE (IAM).

This is a timely contribution in the context of the upcoming CMIP7 phase, particularly with respect to CDRMIP. The manuscript is clearly written and has the potential to make a significant contribution to both the IAM and ESM communities.

Below, I provide several suggestions for improving the manuscript along with questions that arose during my reading.

Thank you for the positive evaluation of our manuscript.

- The title specifies "deep-mitigation scenarios". While I understand that inconsistencies between IAMs and ESMs may be most pronounced in such scenarios, it is not entirely clear why this specification is necessary. Would the proposed framework not also be applicable to moderate or low mitigation scenarios? If its applicability is indeed limited, this could be clarified more explicitly

Thank you for pointing this out. It is correct that our framework is generally applicable if CDR is present in a scenario. The magnitude of the deployment is not critical (other than that the signal of the CDR becomes of course smaller and might be more difficult to isolate against the background of internal variability as discussed in our manuscript). To avoid confusion, we therefore changed "deep mitigation scenarios" to "mitigation scenarios" in the title and also where applicable in the main text of the revised version of our manuscript.

- The proposed framework is beneficial in many respects. However, it may also have some limitations (or specific use-case constraints) that could be more clearly articulated. For example, for assessing biogeophysical effects, the traditional approach of comparing CDR vs no-CDR simulations in a concentration-driven setup may already be sufficient. Similarly, disentangling carbon–concentration and carbon–climate feedbacks can be achieved via standard carbon-cycle switch experiments. It is not entirely clear how such analyses would be incorporated within the proposed framework.

In addition, depending on the research question, there may be a need to estimate (e.g., regional) Earth system feedbacks independently of associated socioeconomic or land-use changes, i.e., isolating "pure" Earth system responses to CDR.

While the manuscript does a very good job of explaining the framework itself, I suggest adding a short section or paragraph that explicitly discusses its limitations or intended scope of application and clarifies which sources of uncertainty are included or excluded (especially, taking into account the broad ESM-IAM community, for whom the paper is addressed).

Thank you for pointing this out. We agree that it is a good idea to add a short paragraph that explains the scope and limitations of our simulation framework. We have done so at the beginning of Section 2 (Motivation and definitions) as follows: "Our work focuses on the IAM – ESM modelling chain that has been established for the assessment of future scenarios in ScenarioMIP over the past decades (O'Neill et al., 2016; Taylor et al., 2012; Van Vuuren et al., 2026). We aim at assessing the efficiency of CDR together with carbon-cycle feedbacks and biogeophysical effects caused by the deployment of CDR in mitigation scenarios. Understanding the Earth system response to CDR is multifaceted, and we do not intend to propose a framework that is suited or preferable for all research questions. We acknowledge that stylized experiments with ESMs or with land and ocean offline models are important for understanding CDR independently of socio-economic constraints. We also focus solely on the feedbacks caused by CDR deployment, which is distinct from determining carbon-cycle feedbacks in response to rising atmospheric CO_2 and to climate change in ESMs (e.g., Arora et al., 2020a)."

Regarding climate- and CO_2-concentration-driven feedbacks, it is correct that our framework does not attempt to separate those. In theory, this would be possible by running biogeochemically coupled variants of all required simulations. Probably, to make the analysis easier, one would additionally consider running these BGC simulations with prescribed atmospheric CO_2 of the corresponding emission-driven fully coupled simulations. However, we feel that covering these aspects would be beyond the scope of this manuscript, and we propose to refrain from adding more details beyond the general statement added to Section 2 above.

- The manuscript highlights the importance of large ensembles due to interannual variability when analysing CDR impacts. While I understand the motivation, I was not fully convinced that this issue is specific to CDR.

  E.g., I speculate that Figure 4a would look about the same if both "CDR" and "no-CDR" were compared but just two perturbed ensemble members of the same "CDR" scenario. This raises the question of whether the illustrated variability is uniquely tied to CDR effects or simply reflects general internal variability.

In addition, the framework involves comparisons between emission-driven and concentration-driven simulations (e.g., for estimating biogeophysical effects), which are also subject to internal variability. I suggest clarifying why interannual variability is particularly critical in the context of CDR. A slight refinement of the text may be sufficient to make this point clearer.

We agree, the issue of internal variability is in no way specific to CDR, and it was not our intention to claim this. As the reviewer rightly points out, the same (or very similar) differences would be obtained by a small perturbation of the same scenario without any differences in CDR deployment. The point that we want to highlight is that CDR deployment, even on small scales during early years of deployment, acts as a small perturbation, which means the no-CDR baselines (both the emission-driven and concentration-driven) will be in a different state with respect to internal variability than the CDR scenario. We have clarified this by adding the following sentences to Section 5.1 (lines 593-596): "CDR deployment in a scenario, even if it is at very small scales initially, acts as a perturbation relative to the no-CDR baselines. This has the same effect as a small perturbation of initial conditions, which is a technique utilized to create initial condition ensembles of ESM simulations."

Some other comments/questions:

L 93 "which is always activity-driven in ESMs": maybe add smth like "see Section …" (reference to a section that provides explanation why A/R is always activity-driven

We have added "…, see section 2.2)", where more details about activity-driven CDR is provided.

L126 would the reduction of CO2 be always less than the gross amount removed? Even during early mitigation phases under increasing emissions?

Even when emissions are still positive (early mitigation) or under increasing emissions, this would be the case, as long as the airborne fraction f remains smaller than 1. Let's assume f = const < 1. In this case, CDR of R PgC would reduce the amount of positive emissions E by E-R PgC. However, the atmospheric reduction of CO_2 would be (E-R)*f, so the amount effectively removed would be R*f < R. Of course, f is not constant, but a more accurate calculation would still yield $\int r(t)*f(t)\, dt$ < R as long as $f(t)$ < 1 over the time interval in question. In theory, one could think about a case where f > 1 (if the total carbon cycle feedback would be positive, i.e. amplifying an emission into the atmosphere). Then indeed, the atmospheric CO_2 reduction could be larger than the removal. However, even under quite strong forcing in the CMIP 1%-CO2 experiments none of the ESMs in Aurora et al. 2013 and 2020 showed a positive total feedback, so this case is quite hypothetical. To make this clearer, we have modified the sentence in question (line 137-138), which now reads: "Hence, due to the buffering nature of carbon cycle feedbacks, the net atmospheric reduction of CO_2 in CDR scenarios will always be less than the gross amount of CO_2 that has been removed."

L134 NAR abbreviation. CDR and PCR have "carbon" in abbreviations. Is there any particular reason for not defining Net Atmospheric Carbon Removal or Net Carbon Removal?

There is no particular reason, but we have discussed (and modified) these acronyms extensively during the writing of our manuscript. "Net Carbon Removal" is ambiguous because it could refer to removals from the active carbon cycle as a whole, so we wanted to have "atmospheric" in it. Also, we would prefer to have not more than three letters. In the context of CDR, we believe there is no danger that "NAR" could be misinterpreted, so we would tend to keep this abbreviation as it is.

Table 1. The absence of symbol for biogeophysical effects made me think whether it should be explained in radiative forcing (or temperature) space? Would including it improve clarity or overly complicate the framework?

Thank you for this suggestion, we have added a symbol for biogeophysical effects. The biogeophysical effects can be calculated for any direct or derived (e.g. frequency of extremes) ESM output. We therefore keep the symbol flexible by defining ΔX^{bgp} for biogeophysical effects in Table 1. We have extended Section 4 accordingly by adding the following text (lines 488-490): "Hence, the biogeophysical effect of CDR deployment for any direct or derived ESM output *X* can be calculated as

$$\Delta X^{bgc} = X^{S} - X^{BS}. \qquad (13)$$"

Section 3.1.2 A/R was a bit difficult to follow (together with Figure 2, which I still do not understand completely). Maybe first explaining the difference between states and transitions could help?

We have added more explanation to Section 3.1.2, which reads (lines 336-339): "Two consecutive land-use states are connected through land-use transitions that specify the rates of conversion from one land-use type to another in each grid cell. Freezing a specific land-use state can be achieved by setting all corresponding transitions to zero. For the case of A/R, where we are interested in eliminating forest expansion (but not deforestation), we can also disable transitions to forest only"

Equations 9-12. Is the “no-more-active” carbon from BECCS (stored in geological reservoirs) included in delta_Cland?

The calculations in Equations 9-12 are based on air-sea and air-land carbon-fluxes. The air-land flux of the simulations with activity-driven BECCS includes the carbon uptake of bioenergy crops, which is later harvested and used for BECCS. So, yes the Delta C_L includes the carbon removed by BECCS and stored geologically. We have clarified this by adding (lines 484-485): “We note that the carbon stored geologically through BECCS is included in ΔC_L^{PCR}, since it was part of $F_L^S(t)$.”

Figure 3b: what does the difference between yellow and grey lines (B and BS) indicate?

We assume, the reviewer means “the difference between the yellow and purple lines (B and BS)”. This difference is the carbon cycle feedback contribution according to Equation 12, i.e. the amount of carbon that is lost from the ocean due to the additional carbon cycle feedbacks in the CDR scenario. We have clarified this by adding the following sentence to the figure caption of Fig. 3: “The difference between the yellow and purple lines in panels b and c is used to derive the feedback contributions according to Eqs. 11 and 12, which is shown by the yellow lines in panel d.”

L 514 “reasonable agreement”: is it given for comparing green and purple lines (with about 5 GtC difference)? Does the difference arise because “it takes up to 10-15 years until the full efficiency of an alkalinity addition has been reached” (L516)?

We see that we have been a bit sloppy here. The wording “reasonable agreement” is of course little accurate and prone to misinterpretation. We have replaced this sentence by a more objective description of the deviations between IAM and ESM, which now read as follows (lines 551-557): “In the case of our example scenario, the IAM accumulated PCR (green line in Fig. 3d) is larger by about 5 PgC (roughly 15%) in 2100 than the ESM PCR. Note that the IAM assumes instantaneous CO_2 sequestration upon addition of alkalinity, whereas in simulations with an ocean biogeochemistry model it takes up to 10-15 years until the full efficiency of an alkalinity addition has been reached (Zhou et al., 2025). Parts of the discrepancy between IAM and ESM will be caused by this simplification in the IAM, but our simulations provide no means of testing the exact amount of delayed OAE uptake in the ESM.”