Remark: During the revision process (but unrelated to any of the reviewers' comments), we detected a small error in some of the scenario simulations that are used as an illustration of our simulation framework. We have re-run the affected simulations and updated our manuscript accordingly (updated Figs. 3 and 4 and adjusted some numbers in Sec. 5). This error does not affect any of our conclusions, and it leads to only small adjustments of some of the numbers presented in Sec. 5. For example, the net atmospheric CO_2 reduction (NAR) is now 20.3 (was 19.5).

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."

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

We thank Ben Sanderson for reviewing our manuscript, and for his 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.

Schwinger et al. present a conceptual framework for assessing Earth system responses to carbon dioxide removal (CDR). The central contribution is the operational distinction between net atmospheric removal (NAR) and process carbon removal (PCR), through construction of a no-CDR baseline in the IAM-ESM modelling chain.

This is a timely paper, a complete emissions-driven framework necessitates activity-driven CDR representations, but the community lacks a shared vocabulary for what coupled CDR assessment is actually measuring. The no-CDR baseline formalism is well thought through, and the ocean alkalinity enhancement (OAE) illustration gives a concrete number for the carbon-cycle offset that can be used for benchmarking and comparison.

I recommend minor revisions. The framework is sound but four points warrant development. The treatment of internal variability in Section 5.1 identifies a real problem but doesn't follow through on the implications for simulation design, and in particular stops short of recommending offline experiments that would sidestep the noise problem for PCR estimation. The multi-model intercomparison gap flagged in the conclusions deserves a more concrete protocol suggestion than it currently receives. The freezing approach to A/R in the ex-post baseline construction conflates policy-driven afforestation with passive forest regrowth from land abandonment, which likely biases the attributable A/R signal. And the counterfactual baseline framing in Section 3.2 understates how different ex-post and counterfactual approaches actually are: they answer different questions and the paper should distinguish them more carefully rather than presenting counterfactual as a cleaner-in-principle alternative.

Thank you for the positive evaluation of our manuscript. See below for a detailed response to all points raised.

Major Comments

1. The internal variability discussion suggests a complementary offline experiment tier

Section 5.1 is one of the most valuable parts of the paper. The authors show that the CDR scenario and Baseline diverge in ENSO phase within a few years of OAE deployment beginning, and that noise in the annual PCR estimate remains comparable to the signal even for decadal averages with a three-member ensemble. Larger initial-condition ensembles are clearly needed for emission-driven NAR estimation. For the PCR specifically, though, the authors could consider a complementary approach worth recommending alongside the large-ensemble strategy.

The PCR is the CO2 drawdown attributable to the CDR activity before carbon-cycle feedbacks. For OAE this depends on ocean carbonate chemistry, circulation, and mixing, all explicitly represented in ocean biogeochemistry models that do not require an interactively coupled atmosphere. Running the ocean in offline mode, forced by prescribed circulation from a scenario climatology, controls the modes of variability driving the noise in Figure 4 rather than having them emerge interactively. A single offline simulation pair in this mode could be statistically informative for PCR estimation with minimal additional computational cost. The same logic holds for A/R and BECCS land carbon; offline land surface models (TRENDY-like) give cleaner process-level estimates and can span multiple scenario backgrounds cheaply.

The paper hints at this in the conclusions but doesn't develop it as a design recommendation. I'd suggest a short paragraph distinguishing what offline runs can and cannot do. For NAR estimation, coupling is essential, since the outgassing response of the ocean and the terrestrial sink weakening under declining CO2 require the full coupled system. For PCR, coupling adds noise without adding information. This is a useful distinction to make explicit for groups planning CDRMIP contributions; as currently framed, readers could reasonably conclude that very expensive coupled ensembles are the only way to do activity-driven CDR assessment - and if there are alternatives, it would be good to spell them out.

Thank you for pointing this out. We agree that off-line experiments can be useful, but there is one important limitation: Our simulation framework includes biogeophysical effects in the PCR, while the proposed offline protocol would not. For OAE this is most likely not relevant (unless changes in ocean surface albedo due to plankton community changes would significantly influence surface climate and/or ocean circulation). So, for the results presented in Section 5.1, the offline protocol would indeed be a cheaper alternative to simulating many ensemble members. However, more generally, and for A/R in particular, the omission of biogeophysical effects in the offline simulations would not be desirable. A/R changes surface temperature and humidity, as well as precipitation and circulation patterns, which all might feed back on the PCR of A/R. In an offline simulation $B_{S\text{-offline}}$, where all forcing fields (not only atmospheric CO_2) are taken from the scenario simulation S, all these changes would not be present and would not enter the estimation of the PCR. For a full assessment of the PCR of A/R these effects should be included, so it is difficult to avoid the need for relatively large ensembles in this case. Nevertheless, we agree that it is useful to add a short discussion on the benefits and limits of offline-simulations. We have therefore added the following paragraph to the revised version of our manuscript at the end of Section 5.1:

“To avoid this problem, one could think of setting up the non-CDR baseline B_S as an offline simulation forced by the complete atmospheric state of S (instead of the atmospheric CO_2 concentration only). Thereby, $B_{S\text{-offline}}$ would be in the same state of internal variability as S, so the PCR could be determined cleanly from just one realization. However, our definition of the PCR intentionally includes biogeophysical effects (Eq. 9 and 10), and such effects would be suppressed by an offline simulation approach. This is most relevant for A/R, which changes surface temperature and humidity, as well as precipitation and circulation patterns, all of which might feed back on the PCR. Generally, for CDR methods that entail significant biogeophysical effects, the offline approach is not suitable, and the full assessment of the PCR requires a relatively large ensemble."

In the conclusions-section we have added the following text and recommendation: “The PCR, specifically, could also be estimated through offline simulations of the land or ocean component that are forced by the atmospheric state of the fully coupled scenario simulation. However, this is only viable if the biogeophysical effects of CDR are negligible, which will likely not be the case for methods with a large land-use footprint. We nevertheless recommend that ESM modelling groups develop and/or maintain the capability to force the land and ocean components of their ESM by the outputs of a fully coupled scenario simulation. ”

2. The multi-model dimension deserves a more concrete recommendation

The authors note that NorESM2-LM results are model-dependent and that the land feedback in particular may vary greatly across ESMs. Given the authors' central involvement in CDRMIP planning, this caveat should be translated into a concrete protocol suggestion rather than left as a general observation.

A coordinated offline PCR intercomparison, with models forced by a common prescribed atmospheric state and alkalinity input, is tractable where a multi-model emission-driven ensemble is not. It would also give the community something that doesn't currently exist: a separation of CDR process uncertainty from feedback uncertainty, without which the inter-model spread in NAR is fundamentally ambiguous (analagous to the TRENDY protocol's assessment of land use impacts on carbon budgets). An analogous ocean protocol under prescribed alkalinity forcing, run across a small set of background CO2 trajectories, could run alongside the coupled simulation. I'm not suggesting this replaces the coupled runs, but a development of supporting offline experiments would be valuable.

We agree that offline experiments for PCR intercomparison of OAE (and possibly other CDR methods that have no or little biogeophysical effects; see our response above) will be extremely useful in a multi-model context. For OAE, a MIP is currently planned, OAEMIP led by the Carbon to Sea Initiative (https://www.carbontosea.org/2025/03/20/oae-model-intercomparisonproject/). A protocol paper is in preparation, but not yet published. In OAEMIP, ocean-only hindcast models of the type used in the Global Carbon Budget will perform pulse deployments of OAE to determine the inter-model spread in the time-dependent PCR curves for OAE. Such hindcast-based protocols using pulse of OAE deployment are superior when it comes to a precise determination of the modelled PCR, since the problem of internal variability is circumvented, but this protocol is limited to the historical period.

When it comes to assessing the PCR in future scenarios, an offline protocol as discussed under comment #1 above could be useful for OAE, where biogeophysical effects are most likely small. However, from a technical viewpoint this protocol is more complicated than prescribing atmospheric CO_2. In contrast to land, it needs to be ensured that changes in ocean circulation in the offline simulation are constrained to follow the fully coupled scenario. This can probably be achieved by applying a restoring of surface salinity towards the scenario-SSS, but it is currently unclear how many ESM groups would be able to run such a protocol. For instance, the model used as an example in our study (NorESM2), would not be able to run such experiments out of the box for future scenarios (since time-varying SSS-relaxation is not implemented for a future scenario case in NorESM2).

In summary, we will leave the description of an offline modelling protocol to the upcoming OAEMIP exercise. Regarding future scenarios, which is the focus of this manuscript, we add the following recommendation in Section 6 (lines 621-626): “The PCR, specifically, could also be estimated through offline simulations employing the land or ocean component of an ESM, forced by the atmospheric state of the fully coupled scenario simulation. However, this is only viable if the biogeophysical effects of CDR are negligible, which will likely not be the case for methods with a large land-use footprint. We nevertheless recommend that ESM modelling groups develop the capability to force their land and ocean component by the outputs of a fully coupled scenario simulation.”

3. The scenario dimension could be acknowledged more explicitly

The 35% feedback offset is derived from a single REMIND-MAgPIE scenario. The authors have demonstrated in past papers that OAE efficiency depends on the background CO2 trajectory, so this number is scenario-specific. A high-emission world and a deep mitigation world produce different carbon-cycle states and therefore different feedback fractions; the scenario dependence of the PCR is itself a key quantity for carbon budget applications. The paper doesn't need to fully resolve this dependency, but scenario dependence should be more highlighted in the conclusions and abstract.

The connection to remaining carbon budget calculations - the zero emissions commitment (ZEC), for example, is also not mentioned. Extending the emission-driven coupled runs past the net-zero point would let the framework diagnose how active CDR deployment modifies the ZEC. ZEC under ongoing CDR is potentially different to that in idealised experiments e.g. it's not immediately obvious if the ocean carbon ZEC dynamics would differ under OAE and a short discussion on this would be useful.

Thank you for pointing this out. We have highlighted the scenario dependence better in the conclusions and the abstract. We have added the following sentence at the end of the abstract: "Both, the efficiency of CDR as well as related carbon cycle feedbacks and biogeophysical effects are expected to be scenario- and model-dependent. Our simulation design of concentration- and emission-driven no-CDR baselines, together with an improved representation of CDR in the IAM - ESM modelling chain, opens an avenue towards estimating CDR efficiencies and their uncertainties under various future scenarios in upcoming model intercomparison activities."

In the conclusions-section: We have highlighted the scenario dependence better: "Both, carbon-cycle-CDR feedbacks and the simulated PCR are model- and scenario-dependent, and we therefore recommend applying our simulation protocol in future model intercomparison exercises that deal with activity-driven CDR, for example the next phase of CRDMIP in CMIP7"

Regarding the connection to remaining carbon budgets and ZEC, this is an important question, which is planned to be addressed in CDRMIP with stylized experiments based on flat10-zec. However, for mitigation scenarios created by IAMs, which is the focus of our paper, the question of ZEC and how it is modified by CDR implementations is much more difficult. Neither CMIP6 nor CMIP7 scenarios have a prolonged period of net-zero emissions where residual positive emissions are balanced by CDR (in the extension periods of the scenarios, both negative and positive emissions are ramped down to zero). To be able to analyse ZEC in ESMs in the presence of internal variability one would need at least 30 years of such a net-zero setting. Also, the construction of a corresponding no-CDR net-zero baseline would be difficult. We believe that this question is therefore best addressed in stylized ESM simulations (as planned for CDRMIP). For our paper, with a focus on scenarios, we feel that considering this issue would add complexity to an already complex paper about an issue that cannot really be resolved by our simulation framework, and we would therefore prefer not to discuss this.

4. The A/R freezing approach conflates policy-driven afforestation with passive regrowth

Section 3.1.2 proposes to freeze all land-use transitions into forest in grid cells where forest area expands, treating these as attributable to A/R. The authors acknowledge on lines 343-346 that A/R can also happen on abandoned land and choose not to distinguish between unintentional and policy-driven A/R, citing inventory reporting conventions. This works for inventory purposes but leaves ambiguity in the NAR/PCR framework, which is specifically trying to decompose CDR effects on the Earth system rather than align with national reporting.

Deep mitigation scenarios, including the REMIND-MAgPIE scenario used here, typically combine demographic transition (population peak and decline), continued agricultural yield improvement, and shifts in dietary patterns. All of these reduce cropland demand and produce secondary forest expansion through passive regrowth, independently of any climate mitigation policy. In scenarios where abandonment-driven forest expansion is a meaningful fraction of total secondary forest gain, freezing all forest transitions in the no-CDR baseline attributes passive regrowth to A/R as a CDR method. This overstates the PCR, and by extension the feedback offset. The scenario and baseline thus differ by more than just CDR deployment,.

For OAE this is not an issue, since the land-use differences between S and B are small. The framework is meant to apply to A/R as well, though, and for A/R-heavy scenarios the attribution bias could be meaningful. A refined freezing approach using IAM-internal flags for policy-driven versus economically-driven land-use transitions could give a cleaner A/R signal. An alternative worth discussing is a partial counterfactual in which the no-CDR baseline uses the land-use pattern from a no-A/R-policy IAM variant if one is available, limiting the counterfactual adjustment to land use rather than re-running the full scenario. The current treatment likely sets an upper bound on attributable A/R PCR rather than a best estimate, and the paper should be explicit about this

We agree that the treatment of policy-driven versus unintentional A/R deserves more explanation and discussion. It was not our primary intention to align with reporting conventions. Our decision to treat intentional (policy-driven) and unintentional A/R equally, was primarily motivated by practical considerations. While REMIND-MAgPIE can make this distinction, there is currently no way to communicate this information to ESMs through LUH data. In a ScenarioMIP/CDRMIP context, it is also unclear if other IAMs are able to distinguish policy-driven from unintentional A/R in a reasonably coherent way. Since unintentional A/R will, in strong mitigation scenarios with large pressure on the land-use system, in most cases only make a small part of total A/R, we still believe that treating all forest expansion as A/R is a justifiable first-order assumption. However, in the revised version of our manuscript, we have better described the problem, avoiding that our "solution" to the problem is perceived as a "recommendation". We acknowledge that our approach might overestimate the carbon removals achieved by A/R policies. We also added a paragraph in Section 6 recommending that the IAM-community works towards providing more fine-grained information about A/R that distinguishes passive regrowth from policy- and carbon-price-driven A/R in the future.

Specifically, we rewrote the paragraph in question, which now reads (lines 361-370): "A/R can also happen unintentionally on abandoned land, for example in scenarios that assume a population peak and a decline towards the end of the century. These cases would not represent policy-driven A/R, and such unintentional regrowth should ideally not be conflated with policy-driven A/R by "freezing" the corresponding transitions in our no-CDR baseline. However, the LUH data for CMIP6 and CMIP7 does not offer a way to distinguish between the two cases. Since it seems reasonable to assume, as a first order approximation, that in scenarios with strong land-based mitigation, unintentional regrowth will contribute only little to total CDR, we will not distinguish between unintentional and policy-driven A/R. By doing so, we will overestimate the carbon removals achieved by policy-driven A/R, while the degree of overestimation will depend on the scenario (i.e., the amount of passive regrowth). As a side note, we mention that current reporting practices and recent assessments (e.g. Smith et al. 2024), do not make a distinction between passive regrowth and policy-driven A/R either."

We have also added a recommendation in Section 6 (Conclusions and recommendations) as follows (lines 646-649): "Likewise, LUH currently offers no way to distinguish passive forest regrowth from policy-driven A/R, which may lead to an overestimation of the effect of policy-driven A/R in our simulation framework. We therefore recommend extending future versions of the LUH management data by an indicator that allows flagging passive regrowth of forests."

5. The counterfactual baseline framing needs a more careful treatment of IAM behaviour

Section 3.2 presents the counterfactual IAM baseline as a cleaner alternative to ex-post adjustment, limited mainly by practical tractability. However - this conceptual framing understates how different the two approaches actually are, and the paper would benefit from engaging more directly with what a counterfactual baseline would and wouldn't tell you.

The first issue is that CDR options in IAMs are not independent. They substitute for each other through the cost-optimization framework and interact through shared resource constraints. Disabling BECCS in REMIND-MAgPIE doesn't just remove BECCS; it frees CCS capacity for fossil applications, shifts DACCS onto the efficient frontier at lower carbon prices, releases land that was growing bioenergy crops for other uses (including potentially more A/R), and alters the marginal abatement cost curve across the whole mitigation portfolio. The counterfactual baseline therefore isn't a "no-BECCS-everything-else-equal" world but a world with substantially more DACCS, different land use, different fossil CCS deployment, and different energy mix. For attribution this raises significant complexity. The atmospheric CO2 difference between scenario and counterfactual baseline reflects not just the absence of BECCS but also all the compensating adjustments the IAM makes in response.

A/R is the hardest case, and the paper doesn't properly acknowledge this. A/R is not a discrete technology with a well-defined deployment rate that can be switched off. In MAgPIE it emerges from the interaction of land-use optimization, carbon pricing applied to land-use emissions, agricultural productivity projections, and forest management assumptions. Disabling A/R requires deciding what that operation actually means, and the options have quite different implications. Removing the land-use carbon price eliminates the A/R incentive but also removes the incentive to avoid deforestation, producing more forest loss in the baseline than in the scenario. Capping forest area at present-day levels prevents A/R but creates a physically unrealistic land system where cropland is abandoned but not allowed to revert to forest, which connects directly to the passive regrowth problem raised in Comment 4. Requiring the IAM to solve for net-zero without A/R lets the model re-optimize but effectively reframes the whole land-use problem rather than isolating A/R. None of these produces a clean "no A/R" baseline.

As such, ex-post and counterfactual baselines answer different questions. Ex-post holds the socio-economic world fixed except for CDR deployment, which is closer to an attribution question. Counterfactual lets the IAM re-optimize, which is closer to a systems-response question. Both are useful, but they give different information and shouldn't be treated as better or worse versions of the same approach. The paper's current framing suggests counterfactual is preferable in principle and limited only by cost; in practice the choice depends on what you want to measure. For quantifying carbon-cycle feedbacks attributable to a specific CDR method, ex-post is probably the right tool despite its socio-economic inconsistency. The counterfactual approach is better suited to asking how the mitigation portfolio as a whole responds when one option is removed, which is a different and separately interesting question. Making this distinction explicit would be useful.

It was not our intention to present the counterfactual no-CDR baseline as a “cleaner” alternative to the ex-post adjustments. We completely agree that the two approaches are very different with specific limitations and strengths (as we already point out in the beginning of Section 3: “The two options offer different insights and have their specific limitations.”).

We agree, the no-CDR baseline would not be a “no-CDR, everything else equal” scenario, and this is the intention. We propose the counterfactual no-CDR baseline in order to compare it to the ex-post approach, particularly the differences in land-use dynamics and the energy system. For example, such comparison would allow us to assess the problem of indirect land-use changes through A/R and BECCS (which might be misattributed in the ex-post baseline, but not in the counterfactual baseline). We believe that these kinds of scenarios are a gap in the current scenario literature. On a side-note, the effects of switching off a distinct CDR option would probably be less extreme than suggested by the reviewer, because the climate target is relaxed at the same time when a CDR method is deactivated (e.g., BECCS is switched off, the CO_2 budget would be increased by the corresponding amount, the CO_2 price might even fall and the incentive for DACCS would be lower; i.e. we do not expect the scenario to show way more DACCS, despite the fact that one competitor for CCS is gone). Also, it is not given that less land would be used for bioenergy, given that the counterfactual no-CDR baseline still needs to decarbonise the energy system. But exactly such questions would be interesting to analyse and to compare to the ex-post no-CDR baseline approach.

Regarding the difficulty to switch off AR in a counterfactual baseline, we only partially agree. In the MAgPIE model it is possible to deactivate only the incentive for afforestation/reforestation, while keeping existing natural land types (e.g. forests) protected with a CO_2 price. However, as discussed above, the “natural regrowth” on abandoned land cannot be deactivated and in a scenario with less targeted A/R, we might see more natural regrowth, so the net difference in overall forest area could be smaller than in the “ex-post no-CDR baseline” approach.

In order to avoid the impression that we see the counterfactual no-CDR baseline as the better alternative, rather than a very different view on the problem, we revised the introduction to section 3 and the description of the counterfactual no-CDR baseline scenario in section 3.2 (line 438-442): “However, land use-patterns and the energy-system would be completely different between the scenario and the corresponding counterfactual no-CDR baseline. Regional biogeophysical effects could therefore not be assessed using this approach. Generally, the two approaches for generating no-CDR baselines deliver very different but complementary information. While the ex-post adjustments are targeted towards attribution, the counterfactual is suited to evaluate the socio-economic system response to a no-CDR assumption.”

We also have added the following sentence to the conclusions (Section 6, lines 600-604): “The two approaches are very different and deliver complementary information. The ex-post adjustments allow attributing effects of a specific CDR method down to a regional level at the cost of disregarding socio-economic consistency. In contrast, the counterfactual no-CDR baseline describes the system response to a no-CDR assumption in a socio-economic consistent way, but a regional attribution to specific CDR methods will not be possible. ”

Minor Comments

- The spatially homogeneous BECCS storage ratio (kgCO2/tDM uniform across all bioenergy cropland) is a reasonable pragmatic choice but the assumption potentially introduces a regional bias in PCR attribution which is worth a caveat.

Agreed, we have added the following sentence to section 3.1.3 (lines 394-396): “The use of Eq. 5 attributes BECCS-induced land-use changes proportionally and homogeneously to all bioenergy related land-use transitions and may thereby introduce regional biases in the estimated PCR of BECCS.”

- The IAM output recommendations in Section 6 read as a wish list. Tables 2 and 3 do some of the needed work but don't cover the gross positive emissions disaggregation needed specifically for no-CDR baseline construction. A clear table specifying required outputs, temporal and spatial resolution for each CDR method would be useful reference for IAM groups.

Agreed, we have added a new table (Table 5) that summarizes the recommendations for additional IAM outputs as follows:

Table 5: Summary of IAM output data needed as input for activity-driven simulation of CDR in ESMs and the construction of no-CDR baselines.

	Data available in CMIP6 / CMIP7	**New data required for activity-driven representation of CDR in ESMs**
General (applies for all CDR methods)		· Separate data for gross removals and gross positive emissions related to each CDR method. Gross positive emissions need to be included in the emission forcing for the ESM in scenario simulation S, but need to be omitted for the no-CDR baselines B. Gross removals are needed to compare the PCR between IAM and ESM. · Separate data for all flavours of a specific CDR method considered in the IAM, e.g. BECCS sourced from bioenergy crops, forest plantations, or forestry residues. Enables stepwise implementation of activity-driven CDR in ESMs.
A/R	· spatially explicit land-cover change · forest management information (wood harvest)	· Spatially explicit information on policy-driven A/R versus passive regrowth added to a future version of LUH
BECCS	· spatially explicit distribution of 2nd generation bioenergy crops	· Either: global average BECCS ratio expressed as CO_2 stored per tonne of dry biomass harvested (kg CO_2/t(DM)) · or, preferably: spatially explicit fraction of bioenergy crops used in combination with BECCS, together with spatially explicit BECCS efficiency (kg CO_2 stored per tonne of dry biomass harvested) added to a future version of LUH
OAE		· Spatially explicit flux of alkalinity addition to the surface ocean, together with fluxes of micronutrients and contaminants if applicable

Summary

The NAR/PCR framework is a strong contribution to understanding and attributing the effectiveness of CDR in ESM simulations. I suggest minor revisions to address the issues raised above.

Thanks again for reviewing our manuscript and for the positive evaluation.