

Peer Review File

Demand for low-quality offsets by major companies undermines climate integrity of the voluntary carbon market



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

This is a very difficult manuscript to review. On the one hand, it deals with a really important issue—the role of voluntary carbon credits in the global effort to address climate change. Given how many companies use these credits as part of their “net zero” strategies an empirical look at the kinds of credits used, their likely quality, etc a close look at the voluntary market is essential. On the other hand, the last five years has seen a LOT of this kind of research; moreover, there has been extensive media coverage of the same in Tech Review, Propublica, the Wall Street Journal, New York Times, FT and especially The Guardian. So the question of novelty is much more pressing for this review than it would be five years earlier.

It is important that the journal have someone who is deep in the analytics of this industry look at the data analysis and the novelty of the data collection—like Danny Cullenward or Barbara Haya.

The central finding of this paper is not new, and I am not sure that the data are new either. The big offset buyers are selecting low quality offsets—that’s what all the other studies have shown. A lot of that literature has focused on the supply side, but the attention to registries (eg, Verra) has implicitly given attention to demand as well.

I think the paper needs to do these things to be publishable.

First, the claim to novelty needs to be addressed head on in the introduction or shortly after—before presenting results. That may require SOME of the methods to be moved into the main text. (The main text is already very long for what it says.).

Second, the authors need to avoid shorthand tools for claiming quality. For example, figure 4a points to 2016 as the start date for CORSIA-eligible credits and the main text suggests that’s a reasonable proxy for quality. Why? (Similarly, the statements about additionality of renewable energy credits face similar problems—the authors focus on whether the projects come from LDCs, but why is that a good test of real quality? Some of the registries use that test, but there is no reason to believe that the registries have it right.). The statements made in the paper that most of the credits—if not all of them—are bogus is probably true, especially because of the massive volume of “avoided emissions” credits. But those statements of quality are not what’s potentially new about the paper. Rather, what’s new is the database of buyers and the attributes of what they have bought. So I’d keep it more narrowly focused on that.

Third, it could be that figure 6 is what is really new. I haven’t seen that information (about price) organized this way before, and it really complements the other analyses that suggest problems w quality. I would plot this, not just put in a table. If, indeed, this is what is new then emphasize that to a greater degree.

Reviewer #2:

Remarks to the Author:

This is an important paper as it further highlights the flaws associated with carbon offsets. While the authors discuss some real attempts to reduce carbon, by some firms, they highlight the fact that the volume of carbon purported to be saved is misleading.

The methodology used for the estimations is sound and draws on appropriate data. It is never possible to get exact details on the carbon saved or produced in schemes,

It is clear from the data presented that firms will distort systems to their advantage as much as possible. Thus, I do think that the authors are exceptionally generous in terms of their discussions of the intentions and consequences.

The Authors refer to Table 2, but there is only a Table 2S?

When discussing the age of an offset, it would be worth discussing whether these are longer term relationships. Thus, if we have an agreement for 5 years with a scheme approaching some cutoff date for inclusion, it will mean that over the agreements term, it will move from compliant to non-compliant.

While not the focus of the paper, it would be valuable to calculate the true carbon saved and the extent to which the 'reported' results are exaggerated. This could even go to a method where firms need to exceed reported reductions, thus have a net reduction, not just a net zeroing out.

I also think that the discussion of the implications needs to be more innovative. For example, governments could have approved schemes which are the only ones that count. Auditing of compliance of the schemes used by governments of companies using them could be put in place, with penalties, such as disallowing the firm using offsets for a given number of years, would make using questionable offsets truly risky.

Reviewer #3:

Remarks to the Author:

The central question of this paper – do offset credits used by the largest purchasers reflect common standards and principles of quality? – is an interesting one and worthy of investigation. However, while the authors' findings appear broadly plausible (large buyers are not meaningfully discriminating against low-quality offsets), they also appear to be overdetermined, based the application of an eclectic and redundant mix of so-called "quality" criteria.

A key deficit is that the authors advance no coherent definition of what they mean by the term "quality" when applied to carbon offset credits. The authors' assessment appears to be based on five sets of considerations:

1. Whether carbon credits are associated with emission reductions or removals (which has nothing to do with quality per se)
2. Whether carbon credits are sourced from project types that other studies have found to be problematic in terms of quality
3. Whether purchased carbon credits tend to be of older vintages (or from older projects – the
4. The additionality of the mitigation activities for which carbon credits were issued (though the authors only do this for renewable energy projects)
5. Whether purchased carbon credits tend to be low cost

The problem with applying these criteria (leaving aside #1) is that, as indicators of low carbon offset quality, they tend to be confounded with each other. Credits from project types with apparent quality issues (#2 and #4), for example, may be precisely those that remain unpurchased and unused for many years (#3), and as a result can be had for a low price (#5). The criteria are therefore mostly redundant, in essence pointing to the same underlying issue in multiple ways, rather than charting quality concerns along different dimensions. As a result, Figure 7 (p. 18), for example, provides little meaningful information beyond what is encoded in the sliding color scale for the data points.

More to point, only #2 is really indicative of relative credit quality – criteria 3-5 have more to do with strategies for mitigating quality risk, but are not quality indicators themselves. I would suggest a more

structured approach to the assessment, that more clearly defines the usefulness of looking at these variables. For example:

1. Are companies buying credits from project types that have widely noted quality concerns? (Evidently yes, but the authors need to rigorously define quality – see suggested references further below.)
2. Are they taking steps to mitigate the quality risks associated with these categories? This is where it could be useful to assess whether they are deploying strategies like being selective about the sources of renewable energy credits; purchasing more recent vintages (although see further comments on this below), or purchasing more expensive credits (though again, this may be only a weakly effective strategy – cheap does not automatically correlate with low quality, and vice versa).

More care needs to be taken not to conflate the presence or absence of these strategies with an assessment of the actual quality of the credits being used.

More detailed comments follow...

First, it is important to say a few words about the “reductions vs. removals” debate. Regardless of what the authors assert about what is “widely seen” as the “innately lower quality” of emission reductions from a climate mitigation perspective (p. 26, lines 3-5), this perception is simply incorrect. There is an abundant literature defining the key characteristics of “high quality” carbon credits. Key criteria – for any type of mitigation activity – include additionality, avoiding over-quantification, permanence, avoiding double counting, and avoiding social and environmental harms.

The first two criteria – additionality and quantification – require consideration of counterfactuals. The authors explicitly acknowledge this with respect to emission reductions (e.g., p. 5, lines 3-5). Somewhat bizarrely, however, they seem to suggest that these considerations do not apply to removals (“In contrast...”, top of p. 6; or lines 22-25 on p. 20, which imply that these issues for “avoidance” projects somehow do not apply to removals). It is crucial to recognize that, following the authors’ own formulation, all *removal* projects “pursue recognition as an emissions compensation device by claiming that a technological or nature-based intervention resulted in *enhanced levels of carbon removal* compared to a counterfactual scenario where the offsetting activity never occurred” (note, however, that the comparison should really be to “a scenario where the financial incentive provided by carbon credits was not present.”) And, it would seem, the authors *do* recognize this on some level – e.g., lines 1-4 on p. 20 – which makes their assertions about the “innately” higher quality of removal projects all the more puzzling.

In short, there is no fundamental or “innate” difference in the quality of compensation provided by either type of project. Quality depends on one’s confidence about the counterfactuals (which can actually be quite high for *some* types of avoidance projects, and low for some types of removal projects). Rather, a number of initiatives have suggested that formally – for the purposes of making “net zero” and similar claims – companies *ought* to choose removals as compensation (perhaps because this mirrors at a company level what will be required to achieve net zero globally). As the authors note, there is a “perennial” debate (p. 26, lines 3-4) about whether this is the right approach. If the authors wish to take a side in this debate, that is fine, but they cannot advance this as a carbon offset *quality* issue. The initiatives advocating use of removals themselves make this very distinction (e.g., reference #26 (the Oxford Principles), which refers at the top of page 5 to the need to ensure that all offset credits are high quality, irrespective of whether the credits are for avoidance or removals).

Second, with respect to “high-risk” project types (the assessment beginning on p.7), the authors err in claiming that projects that destroy N₂O from nitric acid production suffer from quality concerns. These projects have never been subject to additionality concerns (as implied on p. 2, line 18). Nor, according to independent analyses, have they been subject to concerns about over-crediting (unlike

N₂O destruction at *adipic* acid plants). I would encourage the authors to take a closer look at the assessment of these projects in Cames et al. (reference #38) as well as Kollmuss, A. and Lazarus, M. (2010). Industrial N₂O Projects Under the CDM: The Case of Nitric Acid Production. <http://www.sei-international.org/publications?pid=1636>. (The arguments against inclusion of nitric acid N₂O projects in the EU ETS – reference #41 – were based on policy concerns, not credit quality.)

Third, with respect to using “age” as a screen for credit quality, the issues here are a bit more nuanced than the authors suggest. Yes, projects registered under earlier methodology versions may be somewhat more likely to have lower quality (e.g., higher risk of over-crediting) (p. 29). However, this is not uniformly the case – older renewable energy projects, for example, may have *lower* additionality risk because they were implemented at a time when renewable energy was not as competitive with conventional power generation. Rules adopted under CORSIA and Article 6 of the Paris Agreement about project start dates reflect policy concerns – i.e., the need to avoid undermining future target achievement – not necessarily (or exclusively) credit quality concerns (e.g., Warnecke, C., Schneider, L., Day, T., Theuer, S. L. H. and Fearnough, H. (2019). Robust eligibility criteria essential for new global scheme to offset aviation emissions. *Nature Climate Change*, 9(3). 218. DOI:10.1038/s41558-019-0415-y). It is not clear to what extent the same logic should apply to voluntary credit purchases. Taken to the extreme, for example, one could argue that companies should never buy credits that have already been issued, because they represent reductions that have already happened, and therefore are “historical” (p. 29, line 9). But if companies were to only “forward purchase” offsets, this could severely hamper the market, even for high quality project types. If the authors want to propose using age as a screen for “good practice” corporate offsetting, then, they should take time to explain why policy-based cutoffs under CORSIA and Article 6 are relevant to voluntary purchasers. Finally, note that *vintage* restrictions do nothing to alleviate any concerns about credit quality that might arise with respect to start date. The rationale for screening credits according to vintage (in addition to or instead of start date) therefore needs more substantiation. The fact that various exchanges have adopted arbitrary vintage cutoffs (p. 29, lines 24-31) is not dispositive (they may have adopted these rules for financial reasons and/or to reflect buyer preferences – e.g., for CORSIA – which themselves may have nothing to with quality per se).

Fourth, while it is useful to understand whether buyers are being selective about renewable energy carbon credit purchases to reduce additionality risks (section starting p. 13), additionality is not an issue only for renewable energy. The authors could explain more why this is only assessed for renewable energy projects.

Finally, the paper would benefit greatly from explaining and defining what is actually meant by the term “quality,” along with providing references. Additionality is a paramount criterion, but so are avoidance of over-crediting, ensuring permanence, avoiding double counting, and avoiding social and environmental harms. Some sources to consider in refining the definition for this paper:

The IC-VMC “core carbon principles,” related to emissions impact & sustainable development:
<https://icvcm.org/the-core-carbon-principles/>

Broekhoff, D., Gillenwater, M., Colbert-Sangree, T. and Cage, P. (2019). *Securing Climate Benefit: A Guide to Using Carbon Offsets*. Stockholm Environment Institute and Greenhouse Gas Management Institute. <http://www.offsetguide.org/pdf-download/>.

Schneider, L., Michaelowa, A., Broekhoff, D., Espelage, A. and Siemons, A. (2019). *Lessons Learned from the First Round of Applications by Carbon-Offsetting Programs for Eligibility under CORSIA*. Öko-Institut / Perspectives / Stockholm Environment Institute. https://www.carbon-mechanisms.de/fileadmin/media/dokumente/Publikationen/Studie/2019_O__ko-Institut_CORSIA_Lessons.pdf.

Schneider, L., Conway, D., Kachi, A. and Hermann, B. (2018). *Crediting Forest-Related Mitigation*

under International Carbon Market Mechanisms: A Synthesis of Environmental Integrity Risks and Options to Address Them. GIZ, Berlin. <https://newclimate.org/2018/09/19/crediting-forest-related-mitigation-under-international-carbon-market-mechanisms/>.

Additional comments on specific line items...

1. Abstract: What does “underdelivering” mean? This could suggest projects are producing fewer credits than expected, rather than generating fewer real, additional reductions than credits issues. Consider a different term.

2. P. 2., line 17: “Not only are such projects unable to remove atmospheric carbon...” This is not true for forest conservation projects, which may include elements or activities that enhance removals. But more importantly, the relevance of this statement here is unclear. Why does it matter that these projects may be “unable” to remove carbon? The premise seems to be that the only valid way to “counterbalance” emissions is with removals. But that premise needs an explicit introduction and substantiation before simply asserting that the inability to remove carbon represents a quality deficit for these credit types.

3. P. 3, line 16-17: “Emission reduction approach” is a confusing term here – removals are not emission reductions. Furthermore, whether mitigation involves emission reductions or removals has no bearing on offset credit quality – the key questions (for any type of activity approach) are about additionality, quantification, permanence, and avoidance of double counting.

4. P. 4, line 25. “Though reputed...” The Gold Standard may have this reputation, but a citation is needed here. It does not appear unequivocally better based on objective assessments, e.g., Broekhoff, D. and Spalding-Fecher, R. (2021). Assessing crediting scheme standards and practices for ensuring unit quality under the Paris agreement. Carbon Management. DOI:10.1080/17583004.2021.1994016.

5. P. 4, line 32: Please take a closer look at the findings of Cames et al. (reference #38) with regard to industrial gas destruction projects. They do not support the conclusion that these projects continue to suffer quality concerns under the CDM.

6. P. 6, lines 8-9. The Methods do not appear to address the many quality concerns that have been raised with respect to the most common types of removal projects, e.g., tree planting and soil carbon enhancement projects (for example, see here: <https://carboncreditquality.org/index.html> and here: <https://www.offsetguide.org/sticking-to-lower-risk-project-types/higher-risk-project-types/>). This is a significant and concerning oversight on the part of the authors. It is simply incorrect to suggest that emission reduction or avoidance projects suffer from greater *inherent* quality concerns compared to removal projects.

Reply to reviewer comments

Reviewer #1

Comment R1-1

This is a very difficult manuscript to review. On the one hand, it deals with a really important issue—the role of voluntary carbon credits in the global effort to address climate change. Given how many companies use these credits as part of their “net zero” strategies an empirical look at the kinds of credits used, their likely quality, etc a close look at the voluntary market is essential. On the other hand, the last five years has seen a LOT of this kind of research; moreover, there has been extensive media coverage of the same in Tech Review, Propublica, the Wall Street Journal, New York Times, FT and especially The Guardian. So the question of novelty is much more pressing for this review than it would be five years earlier. It is important that the journal have someone who is deep in the analytics of this industry look at the data analysis and the novelty of the data collection—like Danny Cullenward or Barbara Haya.

The central finding of this paper is not new, and I am not sure that the data are new either. The big offset buyers are selecting low quality offsets—that’s what all the other studies have shown. A lot of that literature has focused on the supply side, but the attention to registries (eg, Verra) has implicitly given attention to demand as well.

Reply:

We are aware of the press attention that offsets have attracted and have sought to build on this as well as the academic literature, which has hitherto focused on the supply-side of carbon offsets as you point out. We understand the need to demonstrate novelty with our paper and have taken the following steps in our review to increase this:

- Added retirements from 2023 (Jan 1 – Dec 31) to our dataset. Our analysis now covers 4 years of retirement activity and is the first study to examine if the retirements by individual companies align with key quality metrics.
- Changed our analysis of the age of carbon offsets to examine both project age and vintage in parallel (**Figure 3**). This offers a novel view on the temporal character of the climate actions behind offsets and we are unaware of any similar analysis in the previous literature.

In addition, our study demonstrates novelty in the following areas, which were already present in the first version of the manuscript:

- Creates the first publicly available dataset (covering 4 years from 2020-23) to show the attributes of offsets retired by the largest buyers on the market.
- Demonstrates that the problems raised in previous studies are persistent and showing no improvement over time (e.g. the preference for avoidance credits and cheap credits, mainly from REDD+ and renewables). This is the first study to show that during 4 years and after several spouts of criticism from the media, major companies are still not aligning their offsetting behaviour with common standards and principles of quality or high climate impact.
- Demonstrates empirically that preferences for low-quality offsets can be explained by their cost. As you point out in comment R1-4, this is the first study to demonstrate this. Please note also that the addition of 2023 data now shows a much stronger preference for cheap credits (**Fig. 5**) than in our previous paper.

Comment R1-2

I think the paper needs to do these things to be publishable.

First, the claim to novelty needs to be addressed head on in the introduction or shortly after—before presenting results. That may require SOME of the methods to be moved into the main text. (The main text is already very long for what it says.)

Reply:

We appreciate this advice and have carried out the following changes:

1) Added a clarification of how we interpret quality in the introduction with the following paragraph as follows on page 2-3. Note how this outlines the limitations of existing conceptions.

“Despite considerable heterogeneity across different conceptions of quality²⁷, there is wide agreement on the need to ensure additionality and permanence (i.e., prevent emissions reversal or leakage) while avoiding over-crediting and double-counting as well as considering social and environmental harms^{7, 15, 25, 28, 29}.

Such principles, however, do not automatically guarantee real climate benefits. Consider the case of an outdated credit from a windfarm or forest conservation project that started 15 years ago. If used to offset emissions today, a company’s purchase of this credit would neither foster new decarbonisation activities³⁰ nor remove historical carbon emissions from

the atmosphere, which is crucial for achieving reaching net-zero and tackling emissions overshoot^{31, 32}. Indicators of quality should therefore consider the age of offset projects and credits^{30, 33} as well as the pressing need for carbon removal^{24, 34-37}. Quality indicators should also address price. Not only do cheap offsets typically originate from over-credited projects with low additionality¹⁵, but they also deter funds from flowing to projects with higher quality control measures, which typically cost more¹¹.

Since offset quality is strongly influenced by characteristics at the individual project level, the extant literature has focused on GHG accounting methodologies and assessing project-level climate benefits^{8, 11-13, 28, 38}. Yet there remains a need for a framework and indicators that can be applied with readily available market-level data to determine if the offsets retired by large corporate buyers correspond with key metrics of high quality and high climate benefits. However, firm-level analyses^{5, 39-41} are few, despite many companies depending heavily on offsets to pursue decarbonisation goals. Further, scholars have yet to exhaustively study the publicly available data on registries to examine the behaviour of large-scale corporate offset buyers.”

We then explain that our conception of quality is operationalized with five dimensions as follows:

(1) offsetting approach (avoidance or removal); (2) use of low/high risk project types; (3) age of projects and credits; (4) cost of credits; and (5) additionality (applied to renewable energy projects).

We then sharpened our statement of our contribution as follows in the introduction on page 3:

“Our results reinforce claims that dominating practices on the VCM are not supporting effective climate mitigation. Our contribution is to demonstrate that individual companies are a major cause of persisting quality issues because of demand for problematic and cheap offset types known to overstate emission reductions. We also provide the first publicly available dataset that compiles the attributes of offsets used by the largest corporate buyers on the VCM. This complements the extent literature’s focus on supply-side issues such as project characteristics, registries and methodologies.”

In addition, we have also moved some key explanations from our methods into the main text (the findings) to ensure that readers understand the logic the supports our application of each the five dimensions in our framework.

Comment R1-3

Second, the authors need to avoid shorthand tools for claiming quality. For example, figure 4a points to 2016 as the start date for CORSIA-eligible credits and the main text suggests that's a reasonable proxy for quality. Why? (Similarly, the statements about additionality of renewable energy credits face similar problems—the authors focus on whether the projects come from LDCs, but why is that a good test of real quality? Some of the registries use that test, but there is no reason to believe that the registries have it right.). The statements made in the paper that most of the credits—if not all of them—are bogus is probably true, especially because of the massive volume of “avoided emissions” credits. But those statements of quality are not what's potentially new about the paper. Rather, what's new is the database of buyers and the attributes of what they have bought. So I'd keep it more narrowly focused on that.

Reply:

We agree that there is a need to briefly explain to the reader at various points the justification of our use of certain indicators/standards of quality and not expect that the reader visits the Methods section, where we explain them in detail. We have thus made the following changes. Please note that both changes have simply expanded explanations of these quality indicators that were already present in the first draft, but which appear to have been insufficient.

We with respect to our use of CORSIA cut-off year (2016) as yardstick for an acceptable age of offsets, we have rewritten a paragraph in the findings (page 10) to read as follows:

“Fig. 3 indicates that three quarters (75%) of offsets retired would fail to meet eligibility rules for the Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA), managed by the United Nation's International Civil Aviation Organization (ICAO). To qualify in CORSIA, an offset credit must derive from a project that started issuing credits in 2016 or later and, similarly, have a vintage year (the year when a specific mitigation action occurred) of 2016 or later. Although this rule targets airlines and sets only a weak standard^{28, 49}, the 2016 cut-off year provides a useful indicator of what many industry stakeholders and offset trading platforms (e.g. CBL's Global Emissions Offset (GEO) and ACX) regard as an acceptable age limit for offsets³¹ (see Methods). Moreover, setting a limit on the maximum vintage and project start year addresses quality concerns in at least two ways. First, it assures that offset projects adopt newer procedures when calculating emissions avoidance/removal, since methodologies are continuously updated to correct historical

faults³¹. Second, it prevents the use of older credits, which are likely to reflect low additionality in the case that a project has continued to operate despite not having sold all its credits¹⁵.”

Regarding our use of the standards used by VCS (Verra) and GS (Gold Standard) for the country of implementation for renewable energy, this aims to ensure a minimal level of additionality, by requiring that renewable energy projects are implemented in low-income countries where the barriers to renewable energy are higher than in richer countries, and therefore where a stronger case can be made for additionality. We have expanded our explanation about this as follows on page 13: “Two registries, GS and VCS, have set explicit rules regarding the minimum conditions by which renewable energy projects seeking registration are required to demonstrate additionality. Since a comparably explicit degree of guidance lacks for other project types, including forestry, land use, industrial and commercial, we limit our additionality analysis to renewable energy projects. Two tests are applied; to be considered additional in this analysis, passing either is sufficient. The first test, using a criterion set by GS and VCS, requires that renewable electricity projects are implemented in a Least Developed Country (LDC). The second test, using criteria set by GS, requires that projects target a Low-Income Country (LIC) or Lower-Middle Income Country (LMIC) where penetration of the proposed technology is below 5% of all grid-connected generation capacity. Both tests limit implementation to developing countries where renewable energy has not yet been mainstreamed as a pre-condition for demonstrating additionality. Although these rules apply only to cases of new project registration, both provide a meaningful test of additionality from the perspective of contemporary quality standards.”

In line with your suggestion, we have also clarified our contribution and novelty as follows in the introduction, as follows:

“Our contribution is to demonstrate that individual companies are a major cause of persisting quality issues because of demand for problematic and cheap offset types known to overstate emission reductions. We also provide the first publicly available dataset that compiles the attributes of offsets used by the largest corporate buyers on the VCM.”

Comment R1-4

Third, it could be that figure 6 is what is really new. I haven't seen that information (about price) organized this way before, and it really complements the other analyses that suggest problems w quality. I would plot this, not just put in a table. If, indeed, this

is what is new, then emphasize that to a greater degree.

Reply: We appreciate this advice and have emphasize the novelty of our pricing analysis in the introduction, in the final paragraph.

Regarding your suggestion to show this data as a plot, we are unfortunately unable to do this due to the nature of our data. If we understand your comment correctly, you have suggested that we make a scatter plot to show for each company (1) the exact price paid per credit and (2) the volume of credits retired. However, to do this, we would need to know the exact cost paid by a company at a certain point in time. Because such data is not publicly available, we had to use the estimated average cost of a credit. (Regarding this last point, strictly speaking, at least one private data company [Allied Offsets] has a proprietary dataset that estimates how much a company paid for credits at a particular point in time. However, this data is not publicly available, which would decrease the replicability of our analysis. Besides, we are not sure how reliable this data would be, since it is only an estimation. Therefore, only the companies themselves know how much they actually paid per transaction).

Given this limitation, the most reliable source of price estimation data in our view are the yearly estimations from Ecosystem Marketplace. These estimations also use the same offset categories that we use, which are all based on definitions used in the Berkley Carbon Trading Project at the University of California's Berkley School of Public Policy (i.e. agriculture, chemical processes). In revising the paper, we have since removed our use of other pricing estimations (namely from S&P Platts and Trove Research), keeping only the Ecosystem Marketplace estimations. We made this decision after Ecosystem Marketplace released updates to its annual estimates, which were previously not available in our last draft. Our exclusion of S&P Platts and Trove Research pricing estimates also ensures that each yearly estimation of the average cost of a credit reflects the same set of assumptions (i.e. volume of credits, vintage) and the same definition of offset categories (i.e. agriculture, chemical processes). We hope that you can understand our inability to make the figure that you have suggested, for the reason that we do not have access to a reliable estimate of how much each company actually paid for each transaction.

Reviewer #2

This is an important paper as it further highlights the flaws associated with carbon offsets. While the authors discuss some real attempts to reduce carbon, by some firms, they highlight the fact that the volume of carbon purported to be saved is misleading.

The methodology used for the estimations is sound and draws on appropriate data. It is never possible to get exact details on the carbon saved or produced in schemes. It is clear from the data presented that firms will distort systems to their advantage as much as possible. Thus, I do think that the authors are exceptionally generous in terms of their discussions of the intentions and consequences.

Comment R2-1

The Authors refer to Table 2, but there is only a Table 2S?

Reply: Thank you for pointing this out. We have since resolved this problem in the manuscript. The correct text now reads Table 1.

Comment R2-1

When discussing the age of an offset, it would be worth discussing whether these are longer term relationships. Thus, if we have an agreement for 5 years with a scheme approaching some cutoff date for inclusion, it will mean that over the agreement's term, it will move from compliant to non-compliant.

Reply: You raise an important point and we fully agree with you. This perspective about the relative age of an offset (presumably the vintage) becomes important if considering that an offset with a vintage of say 2015 would have an age of 3 years if retired in 2018 but 9 years if retired in 2024. We therefore include a brief reference to this perspective in the main text (page 11), explaining that our finding about the preference for older vintages by some companies still holds if accounting for the age of the vintage year at the time of retirement. We also include the following table in our supplementary material to show this trend. However, since this is not the focus of our analysis, we do not provide further details in the main text.



Fig. S2. Relative share of credits (%) retired by company and by age of vintage at time of retirement

Values show the distribution of the age of vintage years for all credits sourced by a company over 2020-23. For example, 20.5% of the credits retired by Shell had a vintage aged eight years at the time of retirement. Darker colors indicator higher shares. Data excludes credits retired from CDM since this registry does not disclose vintage years. *One company (Hu-Chems) has no credits recorded because it retired all of its credits from CDM.

Comment R2-2

While not the focus of the paper, it would be valuable to calculate the true carbon saved and the extent to which the ‘reported’ results are exaggerated. This could even go to a method where firms need to exceed reported reductions, thus have a net reduction, not just a net zeroing out.

Reply: You raise an important point that we fully agree with. We regret to say however that we are unable to come up with a method that would estimate the actual carbon avoidance/removal performance of projects vs their stated estimate. This is because the 20 companies in our sample have obtained their offsets from around 500 offset projects, and making an estimation of their emission reduction performance would require an intimate understanding of each. This said, we agree with your idea that companies should purchase offsets well beyond what they actually need to arrive at a net reduction due to the high likelihood that an offset project would not achieve its stated reduction. Yet we are unsure that the strategy of simply purchasing more low-quality credits is worth highlighting as a meaningful way

to address the persisting quality issues in the VCM. Instead of promoting this idea, we instead argue as an implication in our conclusion that companies should use offsets as a “contribution” to climate mitigation and not as a “compensation” for actual emissions. We state this as follows:

“Our findings point to a need for corrective actions to address the VCM’s ongoing quality issues. We foresee two plausible options. In the first option, governments could intervene to regulate the VCM, allowing only the use of removal offsets from government-approved schemes with quality control measures within their jurisdiction. But this would be unlikely to overcome all quality issues, since many offset projects still fail to deliver or overestimate their climate benefits in government-regulated regimes, including in California^{64, 65} and Australia¹⁶. Furthermore, renewable energy offsets generated by CDM, overseen by the United Nations, also suffer from low quality and low additionality¹³. With such abundant evidence that carbon offsets do not provide a reliable means of neutralising emissions, more fundamental change is required. A second option, therefore, until verifiable and permanent removals become widely available, the VCM would be strictly limited to ‘contribution’ and no longer used for tonne-for-tonne compensation^{11, 26, 29}. In this way, companies would support climate mitigation and co-benefits for sustainable development by buying offsets, but refrain from claiming to have neutralised their emissions⁶⁶. The shift from compensation to contribution would generate a need for companies to focus their investments on reducing inhouse emissions and to implement fundamental changes to technologies and business practices. Such a decarbonisation approach would be much more expensive than offsetting. But as shown by modelling pathways, including from the International Energy Agency⁶⁷, pursuing decarbonisation by transitioning away from fossil fuels is the surest and quickest pathway to net-zero and meeting Paris Agreement targets⁶⁸.”

Comment R2-3

I also think that the discussion of the implications needs to be more innovative. For example, governments could have approved schemes which are the only ones that count. Auditing of compliance of the schemes used by governments of companies using them could be put in place, with penalties, such as disallowing the firm using offsets for a given number of years, would make using questionable offsets truly risky.

Reply: Once again, you raise an important suggestion for increasing the reliability of offsets on the VCM. We have therefore responded to your suggestion by adding the following text to the conclusion.

“Our findings point to a need for corrective actions to address the VCM’s ongoing quality issues. We foresee two plausible options. In the first option, governments could intervene to regulate the VCM, allowing only the use of removal offsets from government-approved schemes with quality control measures within their jurisdiction. But this would be unlikely to overcome all quality issues, since many offset projects still fail to deliver or overestimate their climate benefits in government-regulated regimes, including in California^{64, 65} and Australia¹⁶. Furthermore, renewable energy offsets generated by CDM, overseen by the United Nations, also suffer from low quality and low additionality¹³. With such abundant evidence that carbon offsets do not provide a reliable means of neutralising emissions, more fundamental change is required. A second option, therefore, until verifiable and permanent removals become widely available, the VCM would be strictly limited to ‘contribution’ and no longer used for tonne-for-tonne compensation^{11, 26, 29}. In this way, companies would support climate mitigation and co-benefits for sustainable development by buying offsets, but refrain from claiming to have neutralised their emissions⁶⁶. The shift from compensation to contribution would generate a need for companies to focus their investments on reducing inhouse emissions and to implement fundamental changes to technologies and business practices. Such a decarbonisation approach would be much more expensive than offsetting. But as shown by modelling pathways, including from the International Energy Agency⁶⁷, pursuing decarbonisation by transitioning away from fossil fuels is the surest and quickest pathway to net-zero and meeting Paris Agreement targets⁶⁸.”

To explain the logic behind this text, the idea that governments would only allow offset credits that have been approved has been operationalised in some jurisdictions, notable examples being California’s Emission’s Trading Scheme and Australia’s carbon offset scheme. We hesitate however to advocate this as a key approach for addressing quality issues in the VCM because academic studies have also found that credits trading under these government-regulated regimes also suffer from quality issues. See for instance a study (Badgley et al., 2022) showing that forestry-offsets used in California’s cap-and-trade program overstate their emissions reduction or this just published study (Macintosh et al., 2024) and the [supporting dataset](#) by scholars at Australian National University showing that afforestation credits in Australia’s government approved scheme also failed to deliver their stated emissions reductions. (Full references appear at end of this document). This said, we do appreciate that government oversight of the VCM is necessary. And we also suspect that the problems found even in government regulated carbon trading markets are likely pointing to the need for stricter oversight and rule setting.

Reviewer #3

Comment R3-1

The central question of this paper – do offset credits used by the largest purchasers reflect common standards and principles of quality? – is an interesting one and worthy of investigation. However, while the authors' findings appear broadly plausible (large buyers are not meaningfully discriminating against low-quality offsets), they also appear to be overdetermined, based on the application of an eclectic and redundant mix of so-called "quality" criteria.

A key deficit is that the authors advance no coherent definition of what they mean by the term "quality" when applied to carbon offset credits. The authors' assessment appears to be based on five sets of considerations:

1. Whether carbon credits are associated with emission reductions or removals (which has nothing to do with quality per se)
2. Whether carbon credits are sourced from project types that other studies have found to be problematic in terms of quality
3. Whether purchased carbon credits tend to be of older vintages (or from older projects – the
4. The additionality of the mitigation activities for which carbon credits were issued (though the authors only do this for renewable energy projects)
5. Whether purchased carbon credits tend to be low cost

The problem with applying these criteria (leaving aside #1) is that, as indicators of low carbon offset quality, they tend to be confounded with each other. Credits from project types with apparent quality issues (#2 and #4), for example, may be precisely those that remain unpurchased and unused for many years (#3), and as a result can be had for a low price (#5). The criteria are therefore mostly redundant, in essence pointing to the same underlying issue in multiple ways, rather than charting quality concerns along different dimensions. As a result, Figure 7 (p. 18), for example, provides little meaningful information beyond what is encoded in the sliding color scale for the data points.

More to point, only #2 is really indicative of relative credit quality – criteria 3-5 have more to do with strategies for mitigating quality risk, but are not quality indicators

themselves. I would suggest a more structured approach to the assessment, that more clearly defines the usefulness of looking at these variables. For example:

1. Are companies buying credits from project types that have widely noted quality concerns? (Evidently yes, but the authors need to rigorously define quality – see suggested references further below.)

2. Are they taking steps to mitigate the quality risks associated with these categories?

This is where it could be useful to assess whether they are deploying strategies like being selective about the sources of renewable energy credits; purchasing more recent vintages (although see further comments on this below), or purchasing more expensive credits (though again, this may be only a weakly effective strategy – cheap does not automatically correlate with low quality, and vice versa).

More care needs to be taken not to conflate the presence or absence of these strategies with an assessment of the actual quality of the credits being used.

More detailed comments follow...

Reply: We thank you for these comments, which prompted us to:

- (1) Explicitly link our study to existing conceptions of offset quality
- (2) Point out the limitations in these conceptions
- (3) Clearly define the quality dimensions used in our study.

To achieve the first two points, we added the following text to the introduction:

“Despite considerable heterogeneity across different conceptions of quality²⁷, there is wide agreement on the need to ensure additionality and permanence (i.e., prevent emissions reversal or leakage) while avoiding over-crediting and double-counting as well as considering social and environmental harms^{7, 15, 25, 28, 29}.

Such principles, however, do not automatically guarantee real climate benefits. Consider the case of an outdated credit from a windfarm or forest conservation project that started 15 years ago. If used to offset emissions today, a company’s purchase of this credit would neither foster new decarbonisation activities³⁰ nor remove historical carbon emissions from the atmosphere, which is crucial for achieving reaching net-zero and tackling emissions overshoot^{31, 32}. Indicators of quality should therefore consider the age of offset projects and credits^{30, 33} as well as the pressing need for carbon removal^{24, 34-37}. Quality indicators should also address price. Not only do cheap offsets typically originate from over-credited projects

with low additionality¹⁵, but they also deter funds from flowing to projects with higher quality control measures, which typically cost more¹¹.

Since offset quality is strongly influenced by characteristics at the individual project level, the extant literature has focused on GHG accounting methodologies and assessing project-level climate benefits^{8, 11-13, 28, 38}. Yet there remains a need for a framework and indicators that can be applied with readily available market-level data to determine if the offsets retired by large corporate buyers correspond with key metrics of high quality and high climate benefits.

However, firm-level analyses^{5, 39-41} are few, despite many companies depending heavily on offsets to pursue decarbonisation goals. Further, scholars have yet to exhaustively study the publicly available data on registries to examine the behaviour of large-scale corporate offset buyers.”

We have sharpened our research question, as follows, to include a focus on “climate benefits” as well as quality. This justifies our attention to removal offsets and their age.

‘To what extent could the offsets retired by these companies be considered high quality and likely to benefit the climate?’

At the end of the introduction, we then explain that our conception of quality is operationalised with five dimensions as follows:

- (1) offsetting approach (avoidance or removal)
- (2) use of low/high risk project types
- (3) age of projects and credits
- (4) cost of credits
- (5) additionality (applied to renewable energy projects).

Furthermore, in our methods we include the following table to summarise this five-dimension framework. Please note how the framework consists of a broad indicator that is then applied with a specific rule/standard. (Our choice of rules/standards is illustrative, since other rules/standards could also be used).

Table 1. Framework for assessing offset quality and climate benefits

Dimension	Indicator	Rule or standard illustratively used, coding framework adopted	Scientific basis in literature
1. Offsetting approach	Do credits come from projects that directly remove carbon from the atmosphere?	Categorisations of removal, avoidance and mixed indicated in Voluntary Registry Offsets Database by Berkeley Carbon Trading Project.	24, 31, 32, 34, 36, 37
2. Relative quality risks	Do credits come from offset project types with a lower likelihood of overstating their emissions reduction or additionality?	Categorisations of offset project types with a lower, medium and higher risk using the relative quality risks framework in the <i>Quality Offsets Guide</i> by the Stockholm Environmental Institute and GHG Management Institute ¹⁵ .	8, 10, 12, 13, 15, 16
3. Age	Is the window between the offsetting activity and the time of retirement in line with industry standards?	Rule by Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA) that excludes credits with a vintage year and project start year earlier than 2016.	15, 20, 33, 69
4. Price	Does the credit come from offset project types that typically sell for above average price?	Estimates of average price paid for offsets by project category from Ecosystem Marketplace ⁵² .	11, 15, 20
5. Additionality (applied to renewable energy projects)	Do renewable energy credits come from offset projects implemented in low-income countries with a low penetration rate of that energy source?	(1) Rule by GS and VCS that limits eligible renewable energy projects to those located in a Least Developed Country (LDC); (2) Rule set by GS that limits eligible renewable energy projects to those located in a Low-Income Country or Lower-Middle Income Country where the penetration rate of the proposed energy technology is below 5%. Country classifications and data from World Bank ⁵⁵ and IRENA ⁵⁶ .	12, 57, 70

We then explain the logic supporting each dimension of the framework at the appropriate section of the findings. The reader thus is constantly informed about what our conception of quality and high climate impact is, and how this is applied.

Please note that we have also replaced the former scatter plot (Fig 7 in the first draft) with a new figure that applies all of our quality criteria simultaneously to each offset credit in the dataset. This results in a much more comprehensive and integrated analysis and message for the paper.

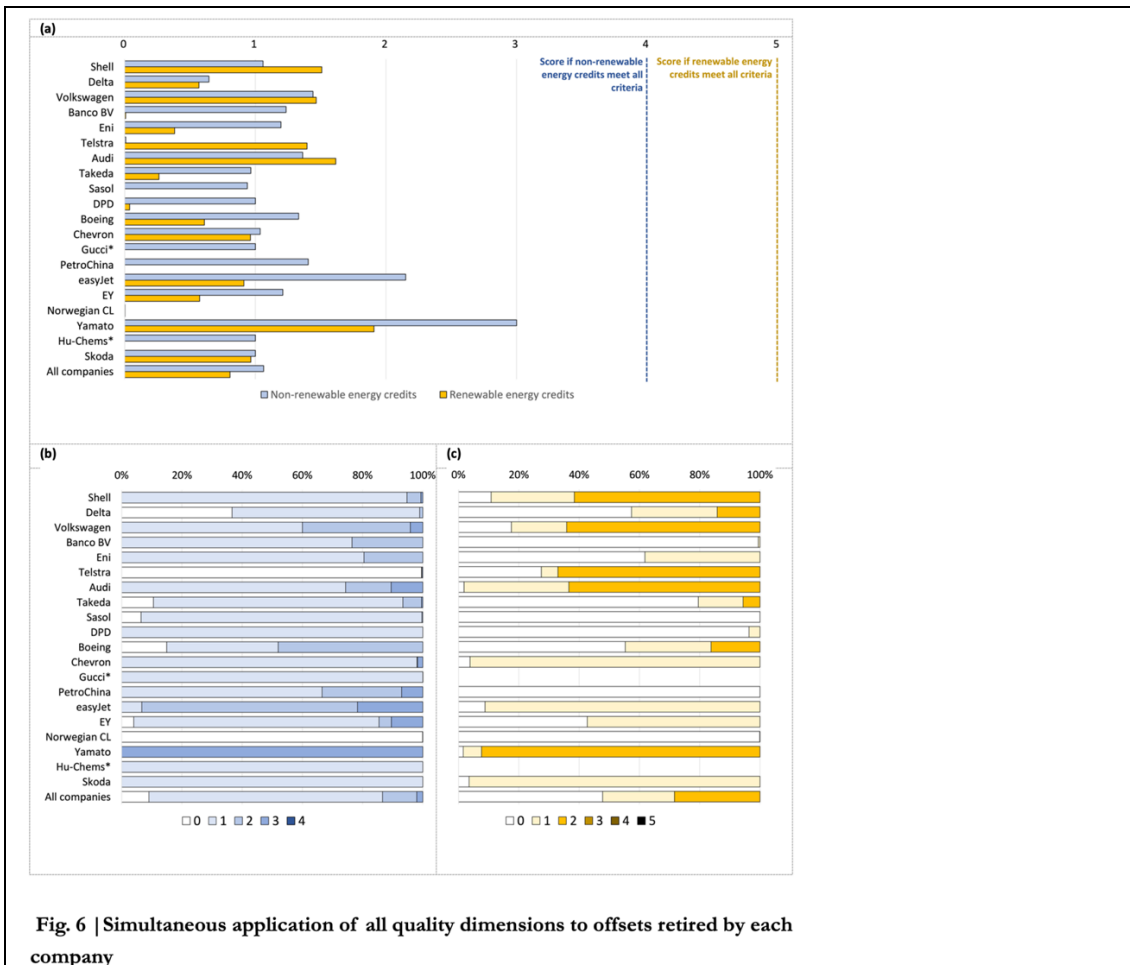


Fig. 6 | Simultaneous application of all quality dimensions to offsets retired by each company

Comment R3-2

First, it is important to say a few words about the “reductions vs. removals” debate. Regardless of what the authors assert about what is “widely seen” as the “innately lower quality” of emission reductions from a climate mitigation perspective (p. 26, lines 3-5), this perception is simply incorrect. There is an abundant literature defining the key characteristics of “high quality” carbon credits. Key criteria – for any type of mitigation activity – include additionality, avoiding over-quantification, permanence, avoiding double counting, and avoiding social and environmental harms.

The first two criteria – additionality and quantification – require consideration of counterfactuals. The authors explicitly acknowledge this with respect to emission reductions (e.g., p. 5, lines 3-5). Somewhat bizarrely, however, they seem to suggest that these considerations do not apply to removals (“In contrast...”, top of p. 6; or lines 22-25 on p. 20, which imply that these issues for “avoidance” projects somehow do not apply to removals). It is crucial to recognize that, following the authors’ own

formulation, all *removal* projects “pursue recognition as an emissions compensation device by claiming that a technological or nature-based intervention resulted in *enhanced levels of carbon removal* compared to a counterfactual scenario where the offsetting activity never occurred” (note, however, that the comparison should really be to “a scenario where the financial incentive provided by carbon credits was not present.”) And, it would seem, the authors *do* recognize this on some level – e.g., lines 1-4 on p. 20 – which makes their assertions about the “innately” higher quality of removal projects all the more puzzling.

In short, there is no fundamental or “innate” difference in the quality of compensation provided by either type of project. Quality depends on one’s confidence about the counterfactuals (which can actually be quite high for *some* types of avoidance projects, and low for some types of removal projects). Rather, a number of initiatives have suggested that formally – for the purposes of making “net zero” and similar claims – companies *ought* to choose removals as compensation (perhaps because this mirrors at a company level what will be required to achieve net zero globally). As the authors note, there is a “perennial” debate (p. 26, lines 3-4) about whether this is the right approach. If the authors wish to take a side in this debate, that is fine, but they cannot advance this as a carbon offset *quality* issue. The initiatives advocating use of removals themselves make this very distinction (e.g., reference #26 (the Oxford Principles), which refers at the top of page 5 to the need to ensure that all offset credits are high quality, irrespective of whether the credits are for avoidance or removals).

Reply: We thank you for comments and affirm that we agree with your assessment, that a removal credit is not a guarantee of quality relative to an avoidance credit. We have thus revisited our statements about this as follows, and also cited the works that you point out:

1. **Introduction (page 3):** We explain that carbon removal is essential for reaching net-zero and tackling emissions overshoot. We state here that avoidance credits are unable to “...remove historical carbon emissions from the atmosphere, which is crucial for achieving reaching net-zero and tackling emissions overshoot^{31, 32}.”
2. **Findings on page 4-6 (Section called “Offsetting approach: Few offsets remove atmospheric carbon”):** Our major change here is to remove any arguments that would suggest that removal credits are free of quality concerns. We simply point out that “VCM governance frameworks^{25, 41}, best-practice

principles^{30, 33} and researchers^{34, 42} advocate permanent removals over avoidance (see Methods)”.

3. **Discussion (page 20):** We point out the quality concerns with removal offsets, also citing a new study by researchers in Australia (Macintosh et al., 2024) supporting this concern:

“Although our findings highlight a pressing need for much larger investments in carbon removal, upscaling nature-based approaches is unlikely to solve many of the VCM’s innate problems, such as questionable additionality. Research shows that many afforestation⁴⁵ and soil-sequestration projects⁶² would have gone ahead without the extra revenue from offset sales. Furthermore, nature-based removals provide only temporary storage that is easily reversed if perturbed by wildfires, drought, disease or changed land-management practices⁶³. Nature-based removal projects therefore carry a high risk of leakage and lack permanence¹⁵. These unresolved quality risks for carbon removal projects call the utility of offsets as a decarbonisation strategy into deeper question. At the same time, these risks reinforce the urgency of pursuing emissions reductions and carbon neutrality by directly phasing out fossil fuels across society.”

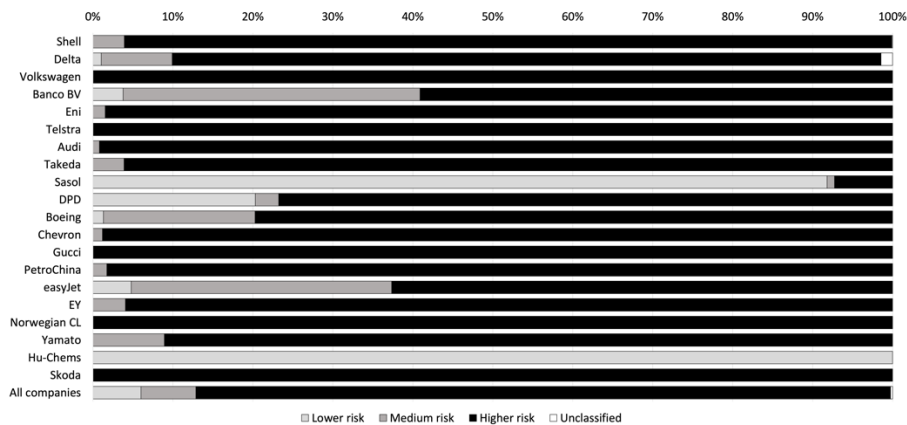
Comment R3-3

Second, with respect to “high-risk” project types (the assessment beginning on p.7), the authors err in claiming that projects that destroy N₂O from nitric acid production suffer from quality concerns. These projects have never been subject to additionality concerns (as implied on p. 2, line 18). Nor, according to independent analyses, have they been subject to concerns about over-crediting (unlike N₂O destruction at *adipic* acid plants). I would encourage the authors to take a closer look at the assessment of these projects in Cames et al. (reference #38) as well as Kollmuss, A. and Lazarus, M. (2010). Industrial N₂O Projects Under the CDM: The Case of Nitric Acid Production. <http://www.sei-international.org/publications?pid=1636>. (The arguments against inclusion of nitric acid N₂O projects in the EU ETS – reference #41 – were based on policy concerns, not credit quality.)

Reply: Once again, we appreciate your comments that have challenged our emphasis on industrial N₂O projects as a high-risk type of offset project. We have since abandoned the argument that these projects constitute by default a particularly problematic type of project.

But please note that we had previously made this decision based on studies that argued against the climate benefits of these projects as well as rules in government regulated emissions trading schemes (EU and New Zealand) that forbid use of these credits. We do of course respect your view that these decisions were due to politics and the desire to protect the EU-ETS from low-cost credits from abroad.

Please also note that we have now sharpened our conceptualization of “high risk projects” by adopting the framework provided by “A Guide to Using Carbon Credits” (Broekhoff, Gillenwater, Colbert-Sangree, & Cage, 2019) that you reference below (which we had already cited and referred to in our first draft). Concretely, we have adopted their definitions of three categories of “relative offset quality risks” (lower, medium and higher). This follows the authors’ view that some “Some types of carbon offset projects have an easier time meeting essential carbon offset criteria than others.” This decision now provides our study with a tool that is higher resolution than our previous approach, since we now are able to assign a level of potential risk to a highly varied range of project types. Although this publication from late 2019 is now around 4 years old, because the bulk of offset projects used by the 20 companies started issuing credits well before 2019, we regard this as a useful and meaningful way to start examining if companies are choosing credits reputed to be high or low quality. Our updated figures now appear on page 9 as follows:



Having stated this, we do appreciate that projects should be judged on a per country or per project basis and that this assessment of potential offset quality risk is extremely coarse and that many projects could be an exception to generalized judgements on quality. Thus, we only use this project as an indicative starting point for a multi-dimensional analysis of quality, which we attempt to carry out in our paper. We explain these caveats as follows (page 7):

“Although the framework does not assess quality based on individual project characteristics, it operationalises the view that some project types are, *ceteris paribus*, more likely to lack additionality, overestimate emissions reductions, or encounter leakage.”

Comment R3-4

Third, with respect to using “age” as a screen for credit quality, the issues here are a bit more nuanced than the authors suggest. Yes, projects registered under earlier methodology versions may be somewhat more likely to have lower quality (e.g., higher risk of over-crediting) (p. 29). However, this is not uniformly the case – older renewable energy projects, for example, may have *lower* additionality risk because they were implemented at a time when renewable energy was not as competitive with conventional power generation. Rules adopted under CORSIA and Article 6 of the Paris Agreement about project start dates reflect policy concerns – i.e., the need to avoid undermining future target achievement – not necessarily (or exclusively) credit quality concerns (e.g., Warnecke, C., Schneider, L., Day, T., Theuer, S. L. H. and Fearnough, H. (2019). Robust eligibility criteria essential for new global scheme to offset aviation emissions. *Nature Climate Change*, 9(3). 218. DOI:10.1038/s41558-019-0415-y). It is not clear to what extent the same logic should apply to voluntary credit purchases. Taken to the extreme, for example, one could argue that companies should never buy credits that have already been issued, because they represent reductions that have already happened, and therefore are “historical” (p. 29, line 9). But if companies were to only “forward purchase” offsets, this could severely hamper the market, even for high quality project types. If the authors want to propose using age as a screen for “good practice” corporate offsetting, then, they should take time to explain why policy-based cutoffs under CORSIA and Article 6 are relevant to voluntary purchasers.

Reply: We appreciate your critical views and the above reference, which is extremely helpful. To address your concern, we have carried out the following actions:

(1) We explained in the main text in more detail why vintage years and project start years are relevant from a quality and climate benefit perspective. We have added the following text:

“Moreover, setting a limit on the maximum vintage and project start year addresses quality concerns in at least two ways. First, it assures that offset projects adopt newer procedures when calculating emissions avoidance or removal, since methodologies are

continuously updated to correct historical faults³³. Second, it prevents the use of older credits, which reflect low additionality, especially in the case where a project has continued to operate despite not having sold all its credits¹⁵.”

- (2) We added a more detailed explanation as follows to our methods. This stance reflects the view from researchers (e.g. the Warnecke (2019) study you introduce above) and industry stakeholders that the age of offsets is highly relevant from a quality and climate benefit perspective.

“There are several reasons why younger offsets may offer higher climate benefits than older ones. First, newer offset projects ensure that investments are directed towards contemporary rather than historical climate actions. This is essential for supporting the formation of new or recently implemented climate mitigation initiatives and for minimising the use of offset credits from historical projects that cannot deliver further emissions reductions beyond what they are already scheduled to do³⁰. Second, old projects that still issue credits for past actions (i.e. backissuing) or that have large volumes of unsold vintages are unlikely to be additional. Because such a project has continued to operate despite not having sold all of its credits, there is a strong argument to suspect that the initiative did not need revenue from offset sales to attain bankability, and therefore is not additional³³. Third, older projects are likely to use older methodologies, which in many cases have been discredited and updated due to problems. Older projects are therefore more prone to overestimating emissions reductions and over-issuing credits^{33, 69}. Credit prices on the VCM reflects preferences for newer vintages, which generally fetch higher premiums from buyers²⁰.”

- (3) We added in the main text the nuance that CORSIA’s 2016 cut-off year *per se* does not reflect if an offset is of high quality or not.

“In applying the 2016 cut-off year to our analysis, we acknowledge that CORSIA targets airlines, that it sets only a soft standard^{30, 51}, and that meeting the cut-off year does not automatically guarantee key quality dimensions such as additionality.”

In addition, in the methods (page 27) we added the following:

Thus, there is no definitive age limit that would guarantee quality. Recognising this, we adopt as an illustrative indicator a rule from CORSIA (Carbon Offsetting and Reduction Scheme for International Aviation), which governs offsetting in the global aviation industry, which prohibits the use of offset credits with a vintage year or project start year before 2016. Although this lenient cut-off year by no means constitutes a rigorous

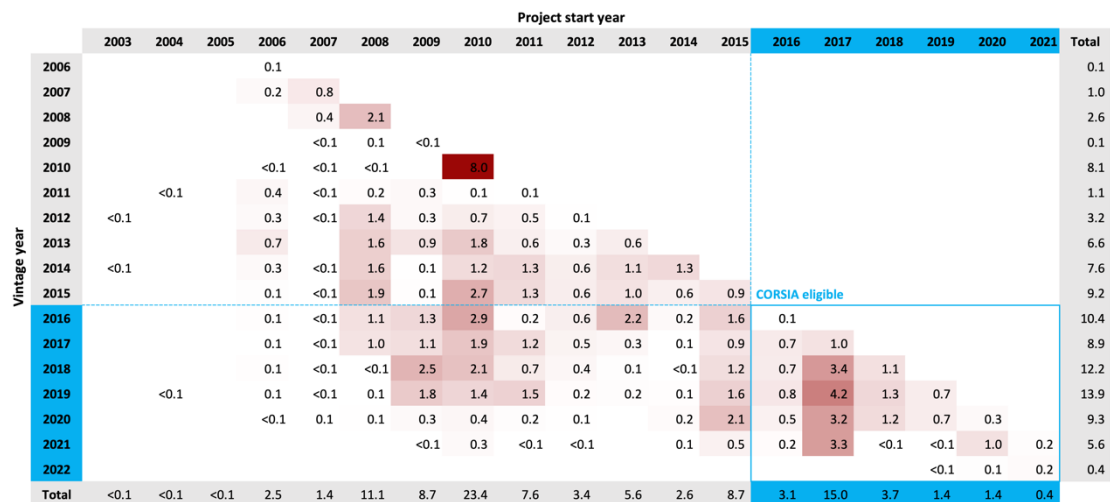
standard, it has been adopted as a defacto yardstick by some commodity trading platforms that screen for quality (e.g. ACX's Sustainable Development Goals Tonne [SDGT] or Global Nature+ Tonne [GNT+] and CBL's Global Emissions Offset [GEO]).

(4) After acknowledging this, we explained why the CORSIA cut-off year of 2016 nevertheless provides a meaningful indicator in our study, adding the following text to the findings:

“Nonetheless, the 2016 cut-off year provides a useful indicator of what many industry stakeholders and offset trading platforms (e.g. CBL and ACX) regard as an acceptable age limit for offsets³³ (see Methods).”

Thus, we have recognized that there is no universally accepted cutoff year for an acceptable degree of an offset's age, and only used the CORSIA cut-off year (2016) as an illustrative indicator, because it considered by many stakeholders and trading platforms as providing a useful minimum requirement for the age of offsetting activities.

(5) Finally, we changed our perspective of analysis on offset age to simultaneously consider both the project start year (measured by the first crediting year) and the credit vintage. We have rebuilt **Figure 3** to appear as follows:



The footnote from this figure also explains that we only use the CORSIA cut-off year as an illustrative indicator, the main reason being that it is considered by main stakeholders and trading platforms as representing an acceptable limit on the age of offsetting activities as set by an authoritative rule making body.

Incidentally, we do not state this in our manuscript, but the Taskforce on Scaling Voluntary Carbon Markets (TSVCM) attempted to use CORSIA's 2016 cut-off year as a minimum quality standard. You can see this discussion on page 65 of this [public consultation document](#) from 2021. This is further evidence that many industry stakeholders interpret the 2016 start year as a useful minimum standard in terms of what should be considered as a reasonable age range for project start years and vintages.

Comment R3-5

Finally, note that *vintage* restrictions do nothing to alleviate any concerns about credit quality that might arise with respect to start date. The rationale for screening credits according to vintage (in addition to or instead of start date) therefore needs more substantiation. The fact that various exchanges have adopted arbitrary vintage cutoffs (p. 29, lines 24-31) is not dispositive (they may have adopted these rules for financial reasons and/or to reflect buyer preferences – e.g., for CORSIA – which themselves may have nothing to do with quality per se).

Reply: We fully agree with your view and have made changes to better convey to the reader these nuances you raise. To view these changes, kindly refer to our reply to **Comment R3-4** above.

Comment R3-6

Fourth, while it is useful to understand whether buyers are being selective about renewable energy carbon credit purchases to reduce additionality risks (section starting p. 13), additionality is not an issue only for renewable energy. The authors could explain more why this is only assessed for renewable energy projects.

Reply: We fully share your view that the issue of additionality does not apply only to renewable energy projects. We focused our additionality analysis on renewable energy projects for the reason that two major registries (VCS and GS) have set explicit additionality criteria that are fairly easy to apply to our dataset. Once again, although meeting these criteria does not guarantee quality per se, once again application of these industry standards provides another means of verifying if the credits retired by companies meet minimum quality standards (in this case rules) on the VCM. Though we would also like to examine the extent to which projects from other sectors (e.g. forestry and land use) demonstrate additionality, we are

unfortunately not aware of any explicit rules set by registries that we could easily apply to our analysis. To make this justification clearer about our decision to examine additionality only for renewable energy projects, we have added the following text to the paper at page 13:

“Two registries, GS and VCS, have set explicit rules regarding the minimum conditions by which renewable energy projects seeking registration are required to demonstrate additionality. Since a comparably explicit degree of guidance lacks for other project types, including forestry, land use, industrial and commercial, we limit our additionality analysis to renewable energy projects.”

Comment R3-7

Finally, the paper would benefit greatly from explaining and defining what is actually meant by the term “quality,” along with providing references. Additionality is a paramount criterion, but so are avoidance of over-crediting, ensuring permanence, avoiding double counting, and avoiding social and environmental harms. Some sources to consider in refining the definition for this paper:

- The IC-VMC “core carbon principles,” related to emissions impact & sustainable development: <https://icvcm.org/the-core-carbon-principles/>
- Broekhoff, D., Gillenwater, M., Colbert-Sangree, T. and Cage, P. (2019). Securing Climate Benefit: A Guide to Using Carbon Offsets. Stockholm Environment Institute and Greenhouse Gas Management Institute. <http://www.offsetguide.org/pdf-download/>.
- Schneider, L., Michaelowa, A., Broekhoff, D., Espelage, A. and Siemons, A. (2019). Lessons Learned from the First Round of Applications by Carbon-Offsetting Programs for Eligibility under CORSIA.
- Öko-Institut / Perspectives / Stockholm Environment Institute. https://www.carbon-mechanisms.de/fileadmin/media/dokumente/Publikationen/Studie/2019_O_ko-Institut_CORSIA_Lessons.pdf.
- Schneider, L., Conway, D., Kachi, A. and Hermann, B. (2018). Crediting Forest-Related Mitigation under International Carbon Market Mechanisms: A Synthesis of Environmental Integrity Risks and Options to Address Them. GIZ, Berlin. <https://newclimate.org/2018/09/19/crediting-forest-related-mitigation-under-international-carbon-market-mechanisms/>.

Reply: Thank you for prompting us to better explain our conception of “quality” and to better cement this in the literature. We have done so by making the changes that we explain in our reply to **Comment R3-1** above. Please check our changes there.

In addition, please let us briefly justify these changes and our now sharpened, five-dimensional conception of quality and climate benefits as follows.

First, this conception of quality and climate benefits recognizes that there has been an ongoing debate in the literature and in the industry about what principles can assure quality, and that criteria for quality have been typically framed as (1) additional, (2) not overestimated, (3) permanent, (4) not double counted and (5) avoiding adverse social or environmental impacts (Broekhoff et al., 2019; Broekhoff & Spalding-Fecher, 2021). Though we recognize that these are widely accepted indicators of quality, they are not the *only* principles of quality out there. For example, this set of five indicators does not take into account the age of offsetting activity (Turner & Grocott, 2021) and the view that offsetting funds should flow to recent or new projects that provide additional mitigation activities rather than to historical projects that don't provide any additional emissions reduction bar those already scheduled to be implemented (Warnecke et al., 2019). Nor does this set of five criteria stress the need for removal credits to meet net-zero/carbon neutrality by mid-century and to drawdown historical emissions after that, as stressed by the IPCC. In fact, in a worst-case but very plausible scenario, a company could offset its emissions via an avoidance project such as a REDD+ or wind power in a developing country that meets all these five criteria, but started 20 years ago and does nothing to reduce contemporary emissions or drawdown atmospheric carbon. Furthermore, many studies are showing that some types of offset projects have a higher likelihood of achieving their stated emissions reductions than others (Broekhoff et al., 2019). Finally, there are several useful standards and rules in the VCM that allow us to examine if companies are following industry or stakeholder principles and standards with regard to age, the "shift to removals" and additionality.

Although you have criticised our framework that synthesises these various debates and standards of quality as "eclectic" and "sometimes redundant", our approach simply follows that view that the five quality indicators do not guarantee climate impact because they do not take into account 1) the age of offsetting activities, 2) the need for removal credits, and 3) the inherent quality risk in some offset types, and 4) the cost of credits.

Given these limitations of existing conceptions of quality, we thus stress in our paper that there is a need for a framework and set of indicators that can be applied with

readily available market-level data to a group of companies using offsets to determine if offsetting procurements correspond with quality metrics. As mentioned, we provide and explain this framework in detail in the Methods section.

In proposing this quality dimensions however, we also acknowledge in our introduction that “offset quality is strongly influenced by characteristics at the individual project level” (page 3).

Comment R3-7

Additional comments on specific line items...

1. Abstract: What does “underdelivering” mean? This could suggest projects are producing fewer credits than expected, rather than generating fewer real, additional reductions than credits issues. Consider a different term.

Reply: Thank you for noting this. We have since changed this term in the abstract to read, more simply: “fail to reduce emissions as claimed”

Comment R3-8

2. P. 2., line 17: “Not only are such projects unable to remove atmospheric carbon...” This is not true for forest conservation projects, which may include elements or activities that enhance removals. But more importantly, the relevance of this statement here is unclear. Why does it matter that these projects may be “unable” to remove carbon? The premise seems to be that the only valid way to “counterbalance” emissions is with removals. But that premise needs an explicit introduction and substantiation before simply asserting that the inability to remove carbon represents a quality deficit for these credit types.

Reply: We agree with your view that avoidance projects like REDD+ can also contribute to carbon removal through forest growth and have thus changed this statement to the following:
Criticism has particularly concerned offset REDD+ (forest conservation)⁸⁻¹¹ and renewable energy projects¹²⁻¹⁴. Such projects issue credits on the claim to have avoided emissions, but are prone to over-crediting and exaggerating their ‘additionality’ (i.e. the claim that the project would not have been implemented without revenue from selling

offsets).

We have also substantiated our criticism of avoidance and emphasis of removal credits as follows:

(1) **Introduction:** Explain that avoidance credits cannot remove historical carbon emissions from the atmosphere, which is crucial for achieving reaching net-zero and tackling emissions overshoot (Axelsson et al., 2024; IPCC, 2022)”

(2) **Findings (Offsetting approach):** Added the text:

“Removal projects directly contribute to decarbonisation and progress to net-zero by physically drawing down and sequestering historical carbon emissions²⁶. In contrast, avoidance projects do not physically capture carbon, instead generating offsets by claiming that a technological or nature-based initiative resulted in lower GHG emissions than a counterfactual scenario where the initiative never occurred.”

Comment R3-9

3. P. 3, line 16-17: “Emission reduction approach” is a confusing term here – removals are not emission reductions. Furthermore, whether mitigation involves emission reductions or removals has no bearing on offset credit quality – the key questions (for any type of activity approach) are about additionality, quantification, permanence, and avoidance of double counting.

Reply: We agree and have since changed this term to “Approach”. We reflected this change in the body text and Table 1 in the methods, which presents the five quality/climate benefit dimensions used in our analysis.

Comment R3-10

4. P. 4, line 25. “Though reputed...” The Gold Standard may have this reputation, but a citation is needed here. It does not appear unequivocally better based on objective assessments, e.g., Broekhoff, D. and Spalding-Fecher, R. (2021). Assessing crediting scheme standards and practices for ensuring unit quality under the Paris agreement. Carbon Management. DOI:10.1080/17583004.2021.1994016.

Reply: We have deleted the argument that Gold Standard is reputed to deliver higher quality offsets. Our previous draft had included this statement only to reflect the subjective views of some stakeholders in the VCM that Gold Standard attempts to provide higher quality offsets than its competitors (for example, it does not supply

REDD+ credits). However, we now fully agree that this positive view of Gold Standard is subjective and does not match the conclusions of objective studies, for example the one that you have kindly introduced us above. We however found this study to be extremely informative and have cited it in our revised manuscript.

Comment R3-11

5. P. 4, line 32: Please take a closer look at the findings of Cames et al. (reference #38) with regard to industrial gas destruction projects. They do not support the conclusion that these projects continue to suffer quality concerns under the CDM.

Reply: As we stated at our reply to **Comment R3-3**, we have since deleted the argument that industrial gas destruction projects are more likely to suffer from quality concerns than other project types. This follows our decision to introduce the different categories of “relative quality risks” for different types of offset projects following the framework developed by Swedish Environmental Institute (SEI) (Broekhoff et al., 2019).

Comment R3-12

6. P. 6, lines 8-9. The Methods do not appear to address the many quality concerns that have been raised with respect to the most common types of removal projects, e.g., tree planting and soil carbon enhancement projects (for example, see here: <https://carboncreditquality.org/index.html> and here: <https://www.offsetguide.org/sticking-to-lower-risk-project-types/higher-risk-project-types/>). This is a significant and concerning oversight on the part of the authors. It is simply incorrect to suggest that emission reduction or avoidance projects suffer from greater *inherent* quality concerns compared to removal projects.

Reply: We have deleted the previous statement that removal projects are not exposed to the same quality issues as avoidance projects due to the absence of a counterfactual emissions baseline. As we stated at our reply to **Comment R3-2**, we included the following text in our Discussion, to point out the quality concerns with removal offsets:

“Although our findings highlight a pressing need for much larger investments in carbon removal, upscaling nature-based approaches is unlikely to solve many of the VCM’s innate problems, such as questionable additionality. Research shows that many afforestation⁴⁵ and soil-sequestration projects⁶² would have gone ahead without the extra revenue from offset

sales. Furthermore, nature-based removals provide only temporary storage that is easily reversed if perturbed by wildfires, drought, disease or changed land-management practices⁶³. Nature-based removal projects therefore carry a high risk of leakage and lack permanence¹⁵. These unresolved quality risks for carbon removal projects call the utility of offsets as a decarbonisation strategy into deeper question. At the same time, these risks reinforce the urgency of pursuing emissions reductions and carbon neutrality by directly phasing out fossil fuels across society.”

References

- Axelsson, K., Wagner, A., Johnstone, I., Allen, M., Caldecott, B., Eyre, N., . . . Smith, S. M. (2024). Oxford Principles for Net Zero Aligned Carbon Offsetting (revised 2024). Retrieved from <https://www.smithschool.ox.ac.uk/sites/default/files/2024-02/Oxford-Principles-for-Net-Zero-Aligned-Carbon-Offsetting-revised-2024.pdf>
- Badgley, G., Freeman, J., Hamman, J. J., Haya, B., Trugman, A. T., Anderegg, W. R. L., & Cullenward, D. (2022). Systematic over-crediting in California's forest carbon offsets program. *Global Change Biology*, 28(4), 1433-1445. doi:10.1111/gcb.15943
- Broekhoff, D., Gillenwater, M., Colbert-Sangree, T., & Cage, P. (2019). Securing Climate Benefit: A Guide to Using Carbon Offsets. Retrieved from https://www.offsetguide.org/wp-content/uploads/2020/03/Carbon-Offset-Guide_3122020.pdf
- Broekhoff, D., & Spalding-Fecher, R. (2021). Assessing crediting scheme standards and practices for ensuring unit quality under the Paris agreement. *Carbon Management*, 12(6), 635-648. doi:10.1080/17583004.2021.1994016
- IPCC. (2022). Climate Change 2022: Mitigation of Climate Change (Summary for Policymakers). Retrieved from <https://www.ipcc.ch/report/sixth-assessment-report-working-group-3/>
- Macintosh, A., Butler, D., Larraondo, P., Evans, M. C., Ansell, D., Waschka, M., . . . Summerfield, P. (2024). Australian human-induced native forest regeneration carbon offset projects have limited impact on changes in woody vegetation cover and carbon removals. *Communications Earth & Environment*, 5(1), 149. doi:10.1038/s43247-024-01313-x
- Turner, G., & Grocott, H. (2021). The Global Voluntary Carbon Market: Dealing With the Problem of Historic Credits. Retrieved from https://trove-research.com/wp-content/uploads/2021/01/Global-Carbon-Offset-Supply_11-Jan-1.pdf
- Warnecke, C., Schneider, L., Day, T., La Hoz Theuer, S., & Fearnough, H. (2019). Robust eligibility criteria essential for new global scheme to offset aviation emissions. *Nature Climate Change*, 9(3), 218-221. doi:10.1038/s41558-019-0415-y

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

Recommendation: accept.

The authors have been highly responsive to the reviews

Reviewer #2:

Remarks to the Author:

The authors have provided additional detail of the information and undertaken additional analysis. Other reviewers have suggested that board implications of the study are not necessarily new. I tend to agree, but the detail in which this is demonstrated is important when discussing the extent to that this a fundamental issue within the domain of Carbon offsets.

The authors have addressed the comments I have raised to my satisfaction, and I acknowledge that some of my suggested extensions would be highly difficult to achieve given the diversity of types of offsets used and each programs specificity in how things are calculated. Thus, further highlighting the need for better regulation of the sector.

Reviewer #3:

Remarks to the Author:

This revised manuscript is greatly improved from the prior draft. Unfortunately, I cannot recommend it for publication, as the authors continue to imply that use of "removal" offsets is **ipso facto** an indicator of low offset quality. As I noted in my prior review, this is simply not the case. The sources the authors cite to support this contention all support use of removal offsets as a **policy preference** (e.g., to demonstrate formal achievement of "net zero" targets), not as a safeguard against low offset quality. To quote from one of the peer-reviewed sources (Cullenward et al):

"...the carbon offsets market must shift its focus from avoided emissions and temporary carbon storage to long-duration carbon removal. The implications of this change are profound and require the creation of an entire new set of industries subject to robust standards. **These new entrants must address many of the same additionality and over-crediting concerns found in the incumbent industry, as well as a host of novel challenges that include uncertain environmental impacts, measurement and monitoring complexities, and social license considerations.**"

(emphasis added - note the distinction between the need to pursue removals as a policy objective vs. the need to ensure removal offsets themselves do not suffer the same quality issues as other types of offsets, which is **not** a foregone conclusion)

Furthermore, as policy objective, preference for removal offsets is far from a settled objective among policymakers, civil society groups, and carbon market "watchdog" organizations. To cite a couple very recent examples:

From an April 2 letter to Nature: "Compared with 'business as usual', preventing emissions that would have occurred reduces atmospheric CO₂ as much as growing new trees or technological capture. If the atmosphere is a rapidly filling bathtub, reducing the tap's flow and pulling the plug out both stop it overflowing... **All types of carbon credit have potential integrity concerns.**" (Mitchard, E., Ellis, P., Cook Patton, S. and Adjei, R. F. (2024). Don't dismiss carbon credits that aim to avoid future emissions. Nature, 628(8006). 36–36. DOI:10.1038/d41586-024-00972-2.)

From an April 5 Carbon Market Watch opinion piece: "While it is true that residual emissions will need

to be matched with permanent removals eventually, we are far away from that time. Even to promote this concept now risks distracting us from the key issue at hand: the need to urgently reduce emissions. To enhance this focus, governments and the private sector must set separate targets for removals and reductions rather than combine them... This is more nuanced than a blunt "offset with removals" approach, as removal targets will initially not be required to match the volume of unabated emissions, and will eventually exceed it in order to reach what some might call "negative emissions" territory. **Ultimately, the focus for the next decade (at least) should be on how to reach reduction targets and not on how to set removal targets.**" (Cf. <https://carbonmarketwatch.org/2024/04/05/why-carbon-offsetting-should-die-out-but-healthy-carbon-markets-should-live-on/>)

These are recent examples, but the debate goes back at least several years, e.g.: <https://ghginstitute.org/2020/04/17/should-carbon-offsets-only-include-removing-co2-from-the-atmosphere/>

In short, whether companies use removal offsets cannot and should not be used as a proxy for whether the offsets they use are high or low quality. If the authors remove this criterion/indicator from their methodology, I think the paper would be fit for publication.

REPLIES TO REVIEWERS

Reviewer #1 (Remarks to the Author)

Recommendation: accept.

The authors have been highly responsive to the reviews.

Reply:

We thank this reviewer for their time taken to review the manuscript and for endorsing its publication.

Reviewer #2 (Remarks to the Author)

The authors have provided additional detail of the information and undertaken additional analysis. Other reviewers have suggested that board implications of the study are not necessarily new. I tend to agree, but the detail in which this is demonstrated is important when discussing the extent to that this a fundamental issue within the domain of Carbon offsets.

The authors have addressed the comments I have raised to my satisfaction, and I acknowledge that some of my suggested extensions would be highly difficult to achieve given the diversity of types of offsets used and each programs specificity in how things are calculated. Thus, further highlighting the need for better regulation of the sector.

Reply: We thank this reviewer for their time taken to review the manuscript and for endorsing its publication.

Reviewer #3 (Remarks to the Author):

This revised manuscript is greatly improved from the prior draft. Unfortunately, I cannot recommend it for publication, as the authors continue to imply that use of "removal" offsets is **ipso facto** an indicator of low offset quality. As I noted in my prior review, this is simply not the case. The sources the authors cite to support this contention all support use of removal offsets as a **policy preference** (e.g., to demonstrate formal achievement of "net zero" targets), not as a safeguard against low offset quality. To quote from one of the peer-reviewed sources (Cullenward et al):

"...the carbon offsets market must shift its focus from avoided emissions and

temporary carbon storage to long-duration carbon removal. The implications of this change are profound and require the creation of an entire new set of industries subject to robust standards. These new entrants must address many of the same additionality and over-crediting concerns found in the incumbent industry, as well as a host of novel challenges that include uncertain environmental impacts, measurement and monitoring complexities, and social license considerations." (emphasis added - note the distinction between the need to pursue removals as a policy objective vs. the need to ensure removal offsets themselves do not suffer the same quality issues as other types of offsets, which is *not* a foregone conclusion)

Furthermore, as policy objective, preference for removal offsets is far from a settled objective among policymakers, civil society groups, and carbon market "watchdog" organizations. To cite a couple very recent examples:

From an April 2 letter to Nature: "Compared with 'business as usual', preventing emissions that would have occurred reduces atmospheric CO₂ as much as growing new trees or technological capture. If the atmosphere is a rapidly filling bathtub, reducing the tap's flow and pulling the plug out both stop it overflowing... All types of carbon credit have potential integrity concerns." (Mitchard, E., Ellis, P., Cook Patton, S. and Adjei, R. F. (2024). Don't dismiss carbon credits that aim to avoid future emissions. *Nature*, 628(8006). 36–36. DOI:10.1038/d41586-024-00972-2.)

From an April 5 Carbon Market Watch opinion piece: "While it is true that residual emissions will need to be matched with permanent removals eventually, we are far away from that time. Even to promote this concept now risks distracting us from the key issue at hand: the need to urgently reduce emissions. To enhance this focus, governments and the private sector must set separate targets for removals and reductions rather than combine them... This is more nuanced than a blunt "offset with removals" approach, as removal targets will initially not be required to match the volume of unabated emissions, and will eventually exceed it in order to reach what some might call "negative emissions" territory. Ultimately, the focus for the next decade (at least) should be on how to reach reduction targets and not on how to set removal targets." (Cf. <https://carbonmarketwatch.org/2024/04/05/why-carbon-offsetting-should-die-out-but-healthy-carbon-markets-should-live-on/>)

These are recent examples, but the debate goes back at least several years,

e.g.: <https://ghginstitute.org/2020/04/17/should-carbon-offsets-only-include-removing-co2-from-the-atmosphere/>

In short, whether companies use removal offsets cannot and should not be used as a proxy for whether the offsets they use are high or low quality. If the authors remove this criterion/indicator from their methodology, I think the paper would be fit for publication.

Reply:

We thank this reviewer for their concerns about our previous decision to include the use of removal offsets as an indicator of quality. After careful deliberation across our team and consideration of sources such as those introduced above, **we are pleased to report that we agree with the reviewer. We have thus removed the use of removal credits from the evaluation of offset quality.** The evaluation framework now appears as follows, consisting of four dimensions and no longer five.

Table 1. Framework for assessing offset quality and climate benefits

Dimension	Indicator	Rule or standard illustratively used, coding framework and data source	Scientific basis in literature
1. Relative quality risks	Do credits come from offset project types with a lower likelihood of overstating their emissions reduction or additionality?	Categorisations of offset project types with a lower, medium and higher risk using the relative quality risks framework in the <i>Quality Offsets Guide</i> by the Stockholm Environmental Institute and GHG Management Institute ¹⁵ .	8, 10, 12, 13, 15, 16
2. Age	Is the window between the offsetting activity and the time of retirement in line with industry standards?	Rule by Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA) that excludes credits with a vintage year and project start year earlier than 2016.	15, 21, 32, 77
3. Price	Does the credit come from offset project types that typically sell for above average price?	Estimates of average price paid for offsets by project category from Ecosystem Marketplace ⁵⁷ .	11, 15, 21
4. Country of implementation (applied to renewable energy projects)	Do credits derive from projects implemented in low-income countries where the diffusion of renewable energy is low and hampered by market or policy conditions?	(1) Rule by GS and VCS that limits eligible renewable energy projects to those located in a Least Developed Country (LDC); (2) Rule set by GS that limits eligible renewable energy projects to those located in a Low-Income Country or Lower-Middle Income Country where the penetration rate of the proposed energy technology is below 5%. Country classifications and data from World Bank ⁶⁰ and IRENA ⁶¹ .	12, 13, 62

We added explanations about our decision to exclude consideration of the use of removal offsets as a quality indicator at several locations throughout the manuscript, as follows. We highlight in bold key statements designed to address your concerns:

Paragraph added to first section of Findings (page 7):

It should be emphasised that carbon removal projects are not immune to the quality issues affecting avoidance projects⁴⁶. For instance, afforestation and soil enhancement projects have been found to overestimate carbon stockage¹⁶ and to lack additionally^{47, 48}. Besides, nature-based solutions are unable to permanently store carbon on a millennial timescale that effective climate mitigation requires^{44, 49}. **These issues motivate us to separate the above analysis of mitigation strategy from the following evaluation of offset quality.**

Paragraph added to Discussion (page 21)

Our evaluation of offset quality (Table 1), however, does not link offset quality to a company's mitigation approach, in particular the choice of avoidance verses removal offsets. This methodological decision reflects three considerations. First, the registries examined in this study do not yet offer technology-based carbon removals with permanent storage, which some stakeholders^{34, 36, 38} promote as a long-term goal for quality enhancement. Second, some VCM stakeholders and researchers worry that prioritising removals may distract from the urgent need to reduce emissions at the source. **To use the 'rapidly filling bathtub' metaphor suggested by Ellis and colleagues⁶⁹, assuming that avoidance and removal offsets equally satisfy quality criteria, atmospheric emissions overflow can be stopped by both 'reducing the tap's flow and pulling the plug out'.** Third, upscaling nature-based removals is unlikely in itself to solve many of the VCM's innate quality problems^{37, 46}. Research shows that many afforestation and soil-sequestration projects do not capture carbon as stated¹⁶ or would have gone ahead without the extra revenue from offset sales^{16, 48}. Furthermore, nature-based removals are unable to store carbon on millennial time-scales⁴⁹, while stockage is easily reversed if perturbed by wildfires, drought, disease or changed land-management practices^{15, 70}. These unresolved quality risks for carbon removal projects call the utility of offsets as a decarbonisation strategy into deeper question. At the same time, these risks reinforce the urgency of pursuing emissions reductions and carbon neutrality by directly phasing out fossil fuels across society.

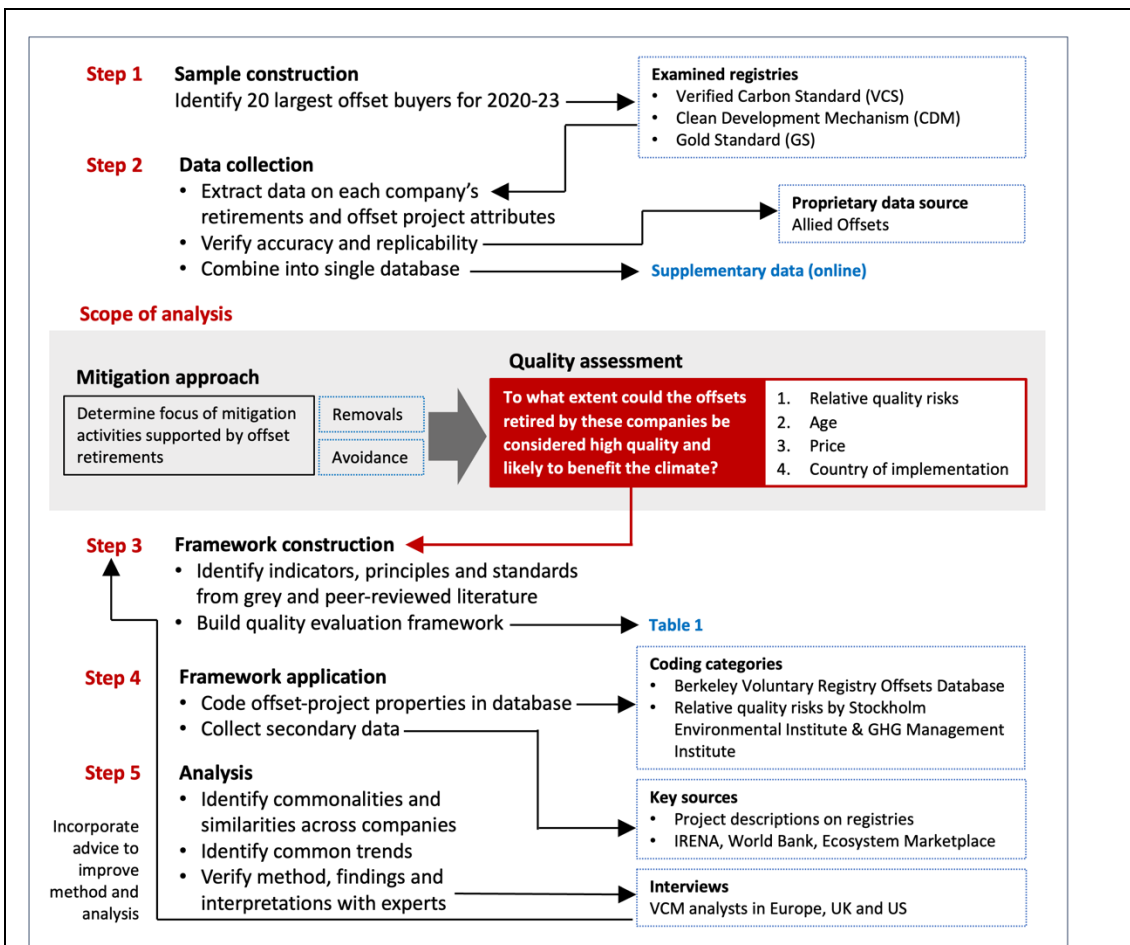
Paragraph added to Methods (page 25)

It is important to note that we do not include the mitigation approach (i.e. avoidance or removal) as an indicator of offset quality. Some industry frameworks like the Science Based Targets initiative and the International Standards Organization have stipulated that

companies should exclusively use removals when using offsets to reach net-zero targets^{25, 43, 76}. However, **as already explained, this methodological decision reflects evidence that non-durable removal techniques like afforestation and soil enhancement also suffer from quality issues, especially with regard to carbon stockage estimation^{15, 16} and additionality^{47, 48}**. Furthermore, due to technological immaturity, offsets from engineering-based removals with durable storage – promoted by some stakeholders as critical for quality enhancement^{34, 36, 44} – were not traded on the studied registries.

As a result of our decision to exclude removal offsets from the quality evaluation framework, we have accordingly omitted the use of removal credits from the summary of quality scores and the associated Figure 6. In addition, please note that we have added to the manuscript some explicit discussions about the quality issues affecting removal credits, which you can see in the above paragraphs.

Please note that after careful consideration and debate across our team, we have decided to retain the data on the share of removals/avoidance in the paper. The principal reason is that this is a topic of high relevance to many researchers and VCM stakeholders. However, **to make it clear that this is a preliminary analysis that is separate to the assessment of offset quality, we have modified our research design figure as shown on the preceding page (see block in grey: “Scope of analysis”)**.



Please note that we explain the decision to separate the assessment of removal/avoidance offsets from the quality assessment at the locations mentioned above (page 7, 21, 25). Here is are some snippets from the above paragraphs to illustrate this:

Page 7 in findings:

It should be emphasised that carbon removal projects are not immune to the quality issues affecting avoidance projects⁴⁶. For instance, (...) These issues motivate us to separate the above analysis of mitigation strategy from the following evaluation of offset quality.

Page 21 in Discussion:

Our evaluation of offset quality (Table 1), however, does not link offset quality to a company's mitigation approach, in particular the choice of avoidance verses removal offsets.

Finally, we have also removed any normative arguments that removal offsets are

superior to removal offsets from the section that discusses volumes/shares of removals (called “Mitigation approach”, starting page 5). The arguments we list there are for the purpose of illustrating the significance of the data on removals retirements are from industry frameworks and the literature and do not reflect our own subjective views.

We trust that the above changes have sufficiently addressed your concerns and that our study is now fit for publication.

Reviewers' Comments:

Reviewer #3:

None