

Forest vegetation increased across China's carbon offset projects and positively impacted neighboring areas

Corresponding Author: Professor Da Zhang

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Decision Letter:

**** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors ****

Dear Professor Zhang,

Your manuscript titled "Leakage or Benefit? Spillovers from a Forest Offset Program in China" has now been seen by 4 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they have raised quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version that fully addresses these serious concerns.

We hope you will find the reviewers' comments useful as you decide how to proceed. Should additional work allow you to • address these criticisms (that is, either to incorporate the suggestions or provide a compelling argument why the point made by the reviewer is not valid, or relevant to the editorial threshold as outlined below)

AND

• meet our editorial thresholds as outlined below,

***** Provide compelling new insights on the spillover effects of China's forest offset programs.**

***** Outline your data and methods in detail and clarify suitability of the Normalized difference vegetation index, the scale and implementation status of the projects, the reference area, and buffer zone selection.**

***** Discuss the limitation of the study design, put your findings in a broader context, and ensure that all your claims are robust and fully supported by your data and analysis.**

then we would be happy to look at a substantially revised manuscript.

If you choose to take up this option, please either highlight all changes in the manuscript text file, or provide a list of the changes to the manuscript with your responses to the reviewers.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

If the revision process takes significantly longer than three months, we will be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the reviewers' comments with a list of your changes to the manuscript text (which should be in a separate document to any cover letter), a tracked-changes

version of the manuscript (as a PDF file) and any completed checklist:

Link Redacted

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

Please do not hesitate to contact us if you have any questions or would like to discuss the required revisions further. Thank you for the opportunity to review your work.

Best regards,

Jinfeng Chang, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0003-4463-7778

Martina Grecequet, PhD
Associate Editor
Communications Earth & Environment

EDITORIAL POLICIES AND FORMAT

If you decide to resubmit your paper, please ensure that your manuscript complies with our editorial policies and complete and upload the checklist below as a Related Manuscript file type with the revised article:

Editorial Policy Policy requirements
(Download the link to your computer as a PDF.)

For Manuscripts that fall into the following fields:

- Behavioural and social science
- Ecological, evolutionary & environmental sciences
- Life sciences

An updated and completed version of our Reporting Summary must be uploaded with the revised manuscript

You can download the form here:

<https://www.nature.com/documents/nr-reporting-summary.zip>

For your information, you can find some guidance regarding format requirements summarized on the following checklist: (<https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf>) and formatting guide (<https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf>).

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

In this study (Manuscript#: COMMSENV-24-0989-T), the authors compiled unique forest offset project data from the CCER programs in China and combined remote-sensed data with causal inference techniques to estimate the spillover effects of 36 forest offset projects on the surrounding landscape's forestry activities. This is an interesting and well-designed study, with a particular focus on forest offset projects in subtropical and temperate regions, where relevant knowledge is still limited. The manuscript is well-written with a clear and compact structure. I greatly appreciate the authors' effective use of counterfactual analysis, with a suite of tests, and carefully addressing potential confounders that could introduce bias in estimation. The findings from this study have the potential to make contributions to the current literature and to the practice of forest offset projects.

However, there is still some lack of clarity in the reported results (both in the text and figures), and I have concerns regarding the data used in this study and the research design. I have the following comments and questions that need to be addressed, which I think would help improve the manuscript.

1. The authors do not appear to have reported the size of the projects in this study. However, knowing the project sizes is important for selecting the appropriate data resolution and determining the number of treated and control samples. Given that the current spatial data is at a 1 km resolution, is this resolution too coarse to capture more nuanced forest change information? Additionally, what is the original resolution of the NDVI data and how was it matched to the 1 km x 1 km grid?
2. One significant limitation of NDVI is the saturation effect, which occurs in areas with high vegetation cover and during peak seasons. This study focuses on forest offset projects, where forests can be very dense, and uses peak monthly NDVI values, which occur during peak seasons. This makes it challenging to use NDVI as an ideal outcome variable for accurately estimating the intervention effects. I suggest that the authors consider using higher-resolution forest data or other more suitable vegetation indices (such as kNDVI or LAI) to further validate the estimated effects. This would also serve as a

robust test of the study's findings.

3. From Table S12, it appears that most of the projects included in this study started after 2015. However, the estimated results indicate that these projects exhibited positive impacts, both within the project areas and in the surrounding buffer zones, as early as 2012-2015. Why might this be the case? How can these early positive impacts be genuinely attributed to the implementation of China's CCER market, rather than to other forest restoration projects or local environmental or ecological initiatives? Although the authors have taken into account the effects of other offset programs, other CCER projects, and protected areas during the matching process, China implemented many ecological restoration programs and various levels of ecological conservation projects. Could the influence of these projects potentially affect the attribution process?

4. From Fig. 1, I noticed that the authors created a fishnet covering the entire extent of the project sites. I wonder if the authors used all these pixels in the analysis (which was not reported). If so, did they consider the potential bias caused by the spatial autocorrelation of geospatial data? If not, please provide information on how the sample size was determined for each project site and how these samples in the analysis can reduce spatial autocorrelation.

5. Please add confidence intervals to Fig. 3c and Fig. 2c.

6. In the captions for Fig. 2d and Fig. 3d, I suggest indicating the criteria used to distinguish between effective and non-effective projects—specifically, whether this is based solely on the sign of the regression coefficients or if statistical significance is also considered. Although the text suggests that the authors only use the sign of the regression coefficients to determine effectiveness, there remains a question regarding Fig. 2d. For example, the ATT for Project 7 and Project 8 is negative. Why are these still classified as effective projects?

7. Lines 127-129: These negative values represent the relative difference between the treated and control samples. There is a possibility that both the project area and the corresponding reference area experienced forest gains, but the forest gain in the project area was less than that in the reference area. I suggest that the authors provide the average change trends for each project and its reference area in the supplementary materials to determine whether the negative estimates are due to a lack of forest gains.

8. How can we be certain that the positive effects within the buffer zones are entirely due to the spillover effects of the projects? Is it possible that these buffer zones are influenced by other forest restoration projects, leading to positive effects relative to the reference areas, rather than stemming from the spillover effects of the projects themselves?

9. The Project IDs in Fig. 2d do not appear to correspond directly with those in Table S2, which makes it very confusing and difficult for readers to locate and track each project's effect. This mismatch also seems to apply to Fig. 3d and Table S3. Moreover, it seems that the Project IDs in Fig. 2d and Fig. 3d do not correspond to each other either. I suggest standardizing the Project IDs across Fig. 2d, Fig. 3d, Table S2, and Table S3. Additionally, I recommend including the 95% confidence interval values and p-values in Table S2, as displayed in Fig. 2d.

10. The inconsistent labeling of Project IDs makes it difficult to track the described effects in the charts when reading site-based estimations in Lines 122-130 and Lines 147-154.

11. In Line 149, negative impacts ranging from -1.84% to -0.84% are reported, but these estimated values cannot be found in either Fig. 3d or Table S3. Are the DID estimates in Fig. 2d, Fig. 3d, Table S1, and Table S2 reported as original values or logarithm-transformed values?

12. Why is it that, regardless of the differences in project size and observations (as noted in Table S2), the number of observations for individual buffer zones is the same (as reported in Table S3)?

13. In Line 190-191, why are there only 8 years before treatment in Fig. 5a when the observation period of this study starts from 1998?

14. In Line 200, the projects were categorized based on the per capita GDP of counties. However, as noted in the caption of Fig. 1, a project may span multiple counties. How did the authors categorize projects that span multiple counties?

15. In Line 201, it appears that positive spillovers are not observed in all groups, such as privately-owned projects and projects in medium-developed counties.

16. Line 206-210, it is plausible that public agencies and state-owned enterprises potentially have better access to information, skills, and resources, which could lead to better project support. However, while this explains why public agencies and state-owned enterprises might better support their projects compared to private enterprises, it does not seem to account for why they also promote forest gain in the buffer zones.

17. Since this study primarily focuses on subtropical and temperate regions, I suggest adding a comparison of the heterogeneity between projects in these two climatic zones.

18. Lines 227-228: Regarding leakage effects, I find the site-specific estimates in Fig. 3d indicate the presence of leakage effects, as we can observe that the buffer zones of some projects have negative estimates. The authors seem to consider that only forest loss leading to an increase in cultivated land can be regarded as 'leakage effects.' However forest loss does not necessarily result in an increase in cultivated land. Also, forest loss or negative ATT values themselves indicate the presence of leakage effects. I suggest revising the conclusion to state that the majority of forest offset projects in this study show no substantial leakage.

19. The authors suggest that the occurrence of spillover effects is primarily through a learning channel. However, this seems to be based solely on inferences from prior literature. Are there any quantitative evidence or actual case studies that demonstrate the of this channel?

20. Lines 333-335: What remote sensing data was used, and from which year? A potential issue is that if some non-forest areas were removed, but these non-forest areas later transitioned to forest (forest gain), would this introduce bias into the effect estimates?

21. Lines 337-340: If the sizes of different projects vary, does this imply that the magnitude or extent of their spillover effects might also differ? In that case, is it reasonable to set buffer zones of the same width for all projects?

22. Is there a difference in the spillover effects when using buffer zones of varying widths? How was the maximum width of 5 km determined? If the actual spillover effects of a project are smaller or larger than this width, is there a possibility of overestimating or underestimating the spillover effects?

23. Line 356: Please explain what CDM and VCS are.

24. Line 347: What is the source of the protected areas data? Does "national conservation" mean that only national-level protected areas were excluded? Given that spatial data on protected areas in China is largely incomplete, such as many city-level or provincial-level protected areas, could this data gap potentially affect the estimation results of this study?

25. After matching, how many treated samples were retained, and how many were discarded? Please provide this information.

Reviewer #2 (Remarks to the Author):

The manuscript "Leakage or Benefit? Spillovers from a Forest Offset Program in China" is on an interesting topic and based on a large dataset (but see comment below). The study is based on a combination of interesting and sophisticated analytical approaches that as such are conducted properly. But I have doubts about the importance of the program itself, which doesn't seem to be widely implemented in China. I don't recognize that this article can be published without addressing the following issues.

First of all, I don't know what the ecosystem type of the buffer is, and I think that could have a big impact and error on the results. I think breaking down the ecosystem types of the buffer zone and discussing them in detail might make the study more scientifically comprehensive.

Secondly, the selection of reference areas I think has some problems, especially in the type of afforestation. Silviculture involves the size of the area, the species of trees and the ecological characteristics of the area, and it is very difficult to find a suitable control group in a number of only 100. I have doubts on this point, which may cause a big error on the accuracy of the study.

Thirdly I think the analysis of ownership is meaningless and the author himself mentions the national characteristics of China. I suggest a study to analyze the impact of facility size on spillover effects.

Last, the project area is so small and shows a fragmented distribution, but the buffer size is thirty times larger than the project area. Is the change in NDVI in the buffer zone all spillover?

Reviewer #3 (Remarks to the Author):

This paper estimates the spillover effects of forest carbon offset programs in China. The authors use a Difference-in-Difference estimation strategy with a matched sample to estimate the causal effects relative to a counterfactual situation without the introduction of the offset programs. Overall, the paper is novel, is written very clearly, and the empirical analysis is conducted carefully. The relevance of the topic and contribution of this work are motivated well in relation to the international literature on carbon offsets, and in the Chinese context in particular. The mechanisms for the potential spillover effects are also explained thoroughly.

I have some comments and questions as follows:

1) It would aid the interpretation of the results to have some additional context about (i) the implementation of the projects, for example how the sites are selected, what activities take place, how the boundaries are determined etc.; and (ii) the condition and use of the forested land in the reference area, and in the project and buffer areas prior to the CCER market. It looks like NDVI was increasing on average in all of project, buffer and reference areas since around 2000 independently of the CCER program, which is relevant to understanding the program impacts.

2) The magnitude of the estimated effects of the projects in project areas and buffer areas appear to be extremely small, especially when compared with the changes over time that were occurring prior to the program and in the reference areas throughout the period.

(i) The size of the effects should be put more explicitly in context so the reader can understand how large they are in practical terms.

(ii) It seems possible that the results are statistically significant simply because of the large sample size. As the unobserved determinants of the outcomes in pixels within the same project are likely to be very similar, at minimum the results should be shown with standard errors clustered to take account of this correlation. I also suggest estimating the DID model at the project level by aggregating the treatment and control pixels in the buffer and reference areas. This would provide additional confidence that the effects are practically meaningful.

3) Since the majority of projects included in the analysis had public comment start dates in 2015 or 2016, more explanation should be provided about whether, and if so why, 2012 is the appropriate treatment date. It seems relevant to consider using the staggered model as the main specification with the single treatment date as a robustness check. It would strengthen the analysis to use staggered start dates with new DID methods such as the Callaway and Sant'Anna (2021) estimator.

4) The Methods section describes how the pixels within each project area are matched with similar socio-economic and

geo-bio-physical pixels. However, it does not explain how the pixels within each buffer are matched with reference pixels. Please add this information.

4) On line 202 in the main text, it says that positive spillovers are observed for all groups in the heterogeneity analysis. This does not seem to fit with Fig 5b, which shows negative effects in privately owned projects and zero effect in counties with Medium economic status.

5) Since matching is carried out with replacement, it would be helpful to include information on whether there are counties or pixels that are used as controls for multiple treated observations, and if so, how many times each county/pixel is used.

Reviewer #4 (Remarks to the Author):

Please find my comments attached.

** Visit Nature Research's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Communications Earth & Environment is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on 'Modify my Springer Nature account' and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.

<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

For more information please visit <http://www.springernature.com/orcid>

If you experience problems in linking your ORCID, please contact the Platform Support Helpdesk.

Version 1:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Professor Zhang,

Your manuscript titled "Leakage or Benefit? Spillovers from a Forest Offset Program in China" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at <https://www.nature.com/documents/commsj-file-checklist.pdf>.

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit <https://www.nature.com/commsenv/open-access>

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:

Link Redacted

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Jinfeng Chang, PhD
Editorial Board Member
Communications Earth & Environment

Martina Grecequet, PhD
Senior Editor,
Communications Earth & Environment
@CommsEarth

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

I would like to thank the authors for carefully considering my comments. The authors have addressed all of my comments thoughtfully and sufficiently. There must have been a lot of effort in making thorough and substantial modifications. I really appreciate it! This revised version is now greatly improved and clearer. Overall, the revised version uses data with higher resolution and more reasonable metrics to evaluate the projects' effectiveness, along with neat robustness checks and heterogeneity analyses. I am satisfied with the updated manuscript and recommend it for publication in Communications Earth & Environment.

Reviewer #2 (Remarks to the Author):

The author has made a good modification according to the reviewer's opinion. I'd like to give an accept decision.

Reviewer #3 (Remarks to the Author):

I appreciate the careful responses to the reviewer comments.

Reviewer #4 (Remarks to the Author):

The authors have clarified remaining issues and made major updates to the paper, and I believe the manuscript would be well suited to CE&E with minor revisions to improve the presentation. Congratulations on the interesting paper!

Regarding earlier comments:

- Figure 1 is much improved. In the new version, the inlays have a font size that is too small to be read in print, and are arguably a bit overloaded. My suggestions (feel free to adopt which ones you deem sensible) would be to
- Relabel the red / blue circles as "Project with coordinates (study sample)" / "Project without coordinates" to remove clutter
- Increase the legend's font size – especially for the plots to the right

- Move the legend of plot (b) to the bottom-left
- Drop "county boundary" and "Pixel (30m)" from the legends
- Consider dropping either plot (e) or the pixel grid in plot (c) to make more space for larger text ()
- I understand that there is a figure review process, and ultimately that is down to the journal. I do appreciate the changes made.
- I'm still unsure about "synthesizing", which suggests that this occurred on its own. Perhaps "synthesized" is a better fit? That's with respect to:
"These matched pixels were eventually synthesizing into a reference area."

Minor comments:

- Line 33 could drop the "Notably, "
- Line 54 and below could be more nuanced (it's very broad and not necessarily part of the results). Maybe something like "Spillovers could play an important role in the development ..."
- Figure 2 would also benefit from larger font sizes; perhaps it would be better to focus on plots (a) and (c) that are very convincing, perhaps even dropping (b) and (d). Estimates for individual projects in (d) will be much less reliable for interpretation, and you mention the important parts in the text and supplement anyways. Same comments hold for Figure 3.
- Typo in line 179: "First, *T*o"
- Lines 185:186 are unclear to me, which other team?
- Wording in line 188 is unclear: "This is attributed to *the fact that?* ~~the~~ most of the projects consist of numerous small-scale polygons"
- Good to see that the results hold with GFW data. Since it is just a check I would not worry about time consistency, but you should be aware of <https://www.globalforestwatch.org/blog/data-and-tools/tree-cover-loss-satellite-data-trend-analysis/>
- Line 327 should probably be "Compared to *the* existing literature~~s~~"

** Visit Nature Research's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Open Access This Peer Review 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 cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review 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 <https://creativecommons.org/licenses/by/4.0/>

Leakage or Benefit? Spillovers from a Forest Offset Program in China

Referee report

May 2024

The paper investigates the off-site impacts of a forest-based carbon-offset program in China. The authors find positive spillover effects, which they attribute to learning effects. The setting is interesting and the contribution would be important, but I have major concerns that would need to be addressed before I can recommend publication.

- First, I don't see why the authors use the (maximum) NDVI as outcome, when other more precise data sources are available. If it's about forests, why not use the Hansen data or other land use data? The NDVI is a very rough proxy that may well be confounded.
- Second, why limit the sample to sites with positive local effects, and term the other ones as "invalid"? This makes the results seem **grossly misleading** to me — akin to "if we exclude interventions that don't work, then the intervention works!"
- Third, I'm not sure how the study reveals learning as "the primary driver of positive spillovers". Is this a result or a possible explanation? Why can we exclude social norms or other links? Please elaborate and clarify this in the paper.
- Fourth, the staggered design of your intervention complicates diff-in-diff estimation somewhat, but is not addressed in the paper — see Goodman-Bacon, 2021.

While these open issues are critical, I find a lot to be liked about the paper. I like that the authors are clear about possible mechanisms for spillover effects, and taking the coincidence of positive and negative spillovers seriously is

interesting. I hope that my concerns can be alleviated, and that they might be helpful in improving the manuscript. Please find other comments below.

Minor comments

- I feel that the abstract could be improved
 - First two sentences could be shortened, and maybe “less considered” could be rephrased to something like “quantifying impacts beyond borders remains an open issue”
 - I’d drop the “unique dataset”
 - The sentence on “Employing a matching [...] in multiple directions” could be clearer worded, and
 - the one afterwards is too long.
 - Last sentence could be more concrete — what do we learn?
- Figure 1 needs to be more accessible.
 - The NDVI color scale is bad for red-green blindness and printing — consider using a two-color scale (e.g. from bright or yellow to green) instead
 - The NDVI visualization adds little information to where it is needed (project sites are almost all green), maybe a log-scale helps (don’t plot the 0s)?
 - I don’t see why it is necessary to visualize any kind of EEZ, and especially not *just* part of it (in a disputed area)
- I would move Figure 4 to the Appendix
- Please reword the discussion, which states that “We find limited evidence of negative spillovers, implying that leakage is not a major issue in China at present”, which is misleading to wrong
 1. The absence of evidence is not evidence of absence
 2. Instead of qualifying the time (“at present”), it would be much more important to qualify the scope where your results have validity (selected, i.e., effective forest-offset programs)
- The small discontinuous polygons that are treated are weird to me — overall, I feel like the paper could do a better job at explaining what these forest-offset programs do and how
 - Wouldn’t we expect leakage into the immediate vicinity then?
- Line 372 has a typo (“synthesizing”)
- The 13 years before CCER seems excessive?

**Response to Reviewers' Comments on
Leakage or Benefit? Spillovers from a Forest Offset Program in China
COMMSENV-24-0989-T, resubmitted as "Communications Earth &
Environment"**

We find all the comments from the four reviewers very constructive. Below we have listed our responses to all comments. The reviewers' original comments are provided in **bold**, followed by our responses. Excerpts from the revised manuscript are provided in *italics*.

Response to Reviewer #1

In this study (Manuscript#: COMMSENV-24-0989-T), the authors compiled unique forest offset project data from the CCER programs in China and combined remote-sensed data with causal inference techniques to estimate the spillover effects of 36 forest offset projects on the surrounding landscape's forestry activities. This is an interesting and well-designed study, with a particular focus on forest offset projects in subtropical and temperate regions, where relevant knowledge is still limited. The manuscript is well-written with a clear and compact structure. I greatly appreciate the authors' effective use of counterfactual analysis, with a suite of tests, and carefully addressing potential confounders that could introduce bias in estimation. The findings from this study have the potential to make contributions to the current literature and to the practice of forest offset projects.

We sincerely thank you for your kind and encouraging comments on our study. We greatly appreciate your recognition of our research question, research design, and the potential contributions of our work. It is particularly meaningful to us to hear that you found our manuscript well-written and our counterfactual analysis appropriate. We are hopeful that our findings will indeed contribute to both the current literature and the practice of forest offset projects.

However, there is still some lack of clarity in the reported results (both in the text and figures), and I have concerns regarding the data used in this study and the research design. I have the following comments and questions that need to be addressed, which I think would help improve the manuscript.

We thank the reviewer for sharing us very constructive suggestions. We have revised our manuscript to address them as described below. Your original comments are provided in **bold**, followed by our responses. Excerpts from the revised manuscript are provided in *italics*.

- 1. The authors do not appear to have reported the size of the projects in this study. However, knowing the project sizes is important for selecting the appropriate data resolution and determining the number of treated and control samples. Given that the current spatial data is at a 1 km resolution, is this resolution too coarse to capture more nuanced forest change information? Additionally, what is the original resolution of the NDVI data and how was it matched to the 1 km x 1 km grid?

We sincerely appreciate your insightful comments regarding project sizes and data resolution. Your observations have led to significant improvements in our analysis and presentation of results.

Regarding project sizes, we apologize for any lack of clarity in our initial report. We included the area of each project in Table S11 in the last version of manuscript, measured in hectares. Our projects range from 1.77 km² to 409.32 km², with a mean area of 50.63 km². To enhance transparency, we have now additionally supplemented the Supporting Information with a table detailing project numbers, sample sizes based on pixel counts, corresponding buffer zone sizes, and control group sizes.

Your comment about data resolution was particularly valuable and led to a significant enhancement of our study. While our original analysis used 1km x 1km NDVI data, we recognized that this resolution might be too coarse to capture nuanced forest changes, especially given the scattered nature of China's forestry carbon sink projects.

In response, we made a substantial revision to our methodology:

First, we replaced the 1km x 1km data with the highest resolution data available: 30m x 30m. This change required overcoming significant computational challenges but has greatly improved our ability to accurately capture project site information.

Second, we moved the original 1km x 1km results to the Supporting Information as a robustness check. We compared the accuracy of project delineation between the two grid resolutions (as shown in Fig. S5) and analyzed the differences in estimation results.

Our findings confirm that the higher resolution data more accurately captures the small, scattered, and disconnected nature of China's forest offset projects. Notably, the revised estimates using high-resolution data are significantly larger than those from the lower resolution data. This difference is attributed to the lower resolution data capturing too many non-forested areas outside the project boundaries, leading to an underestimation of effects.

We are grateful for your constructive feedback, which has substantially improved the precision and reliability of our results.

Manuscript changes:

Page 6 paragraph 3 (line 183):

Second, we further validated our results using NDVI data estimated from different sources. Due to computational constraints, we used 1km × 1km resolution data from another team for matching and baseline regression. While the conclusions remained consistent (Fig. S4; Table S6, columns 8-9), the higher-resolution data used in our primary analysis yielded approximately 5 times larger effect. This is attributed to the most of the projects consist of numerous small-scale polygons, with nearly 45% of the 1km² grids covering project sites contained less than 10% project area, resulting in a coarser representation of green biomass (see comparison in Fig. S5).

Supporting information changes:

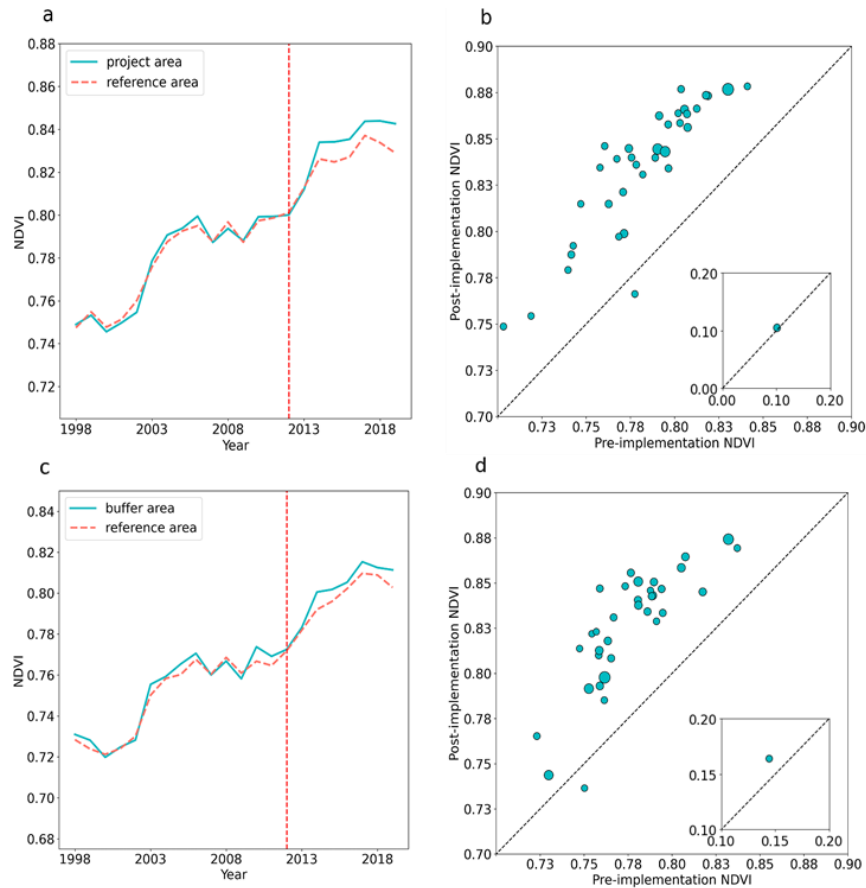


Fig.S4. Forestry activities in project sites and buffers using 1km×1km resolution data.

Note: Plot a and Plot c illustrates the annual mean trends of NDVI for project areas and buffer zones, respectively, alongside their corresponding reference areas, and the red vertical line representing the year 2012. plots b and d offer comparative analyses of NDVI values between the pre-implementation and post-implementation periods for project areas and buffer zones, respectively.

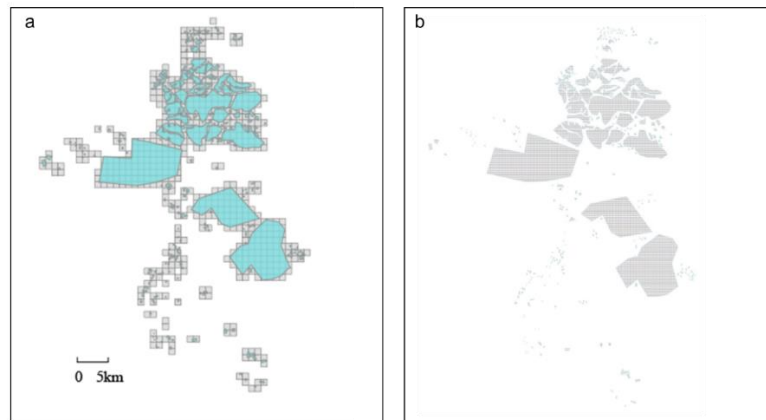


Fig.S5. Grid of different spatial resolution overlaid on an example study site.

Note: Plot a demonstrates the coverage of a 1km×1km grid over the project area. The blue region delineates the actual project boundary, while the grey grid cells represent areas that intersect with the project site. The relatively coarse granularity of this grid results in the inclusion of substantial non-project areas within grid cells that overlap the project boundary. Plot b shows the same project area overlaid with a 30m×30m grid. The finer granularity allows for a nearly seamless integration of the grid with the project area, to the extent that the distinction between the grid and the project boundary becomes visually indiscernible, effectively eliminating the blue demarcation visible in plot a.

Table S9. Project-level sample sizes.

<i>Project ID</i>	<i>Project size</i>	<i>5-km buffer size</i>	<i>Control pool</i>
<i>1</i>	<i>67,404</i>	<i>5,745,913</i>	<i>7,714,004</i>
<i>2</i>	<i>363,900</i>	<i>2,119,508</i>	<i>6,196,655</i>
<i>3</i>	<i>61,054</i>	<i>1,921,103</i>	<i>9,863,194</i>
<i>4</i>	<i>439,734</i>	<i>2,506,444</i>	<i>6,451,283</i>
<i>5</i>	<i>5,864</i>	<i>179,469</i>	<i>5,964,796</i>
<i>6</i>	<i>64,603</i>	<i>3,327,156</i>	<i>8,154,565</i>
<i>7</i>	<i>151,439</i>	<i>2,135,344</i>	<i>6,102,731</i>
<i>8</i>	<i>58,780</i>	<i>3,430,284</i>	<i>6,451,283</i>
<i>9</i>	<i>25,821</i>	<i>1,131,532</i>	<i>4,038,793</i>
<i>10</i>	<i>78,306</i>	<i>869,315</i>	<i>3,573,048</i>
<i>11</i>	<i>42,809</i>	<i>1,910,698</i>	<i>4,622,267</i>
<i>12</i>	<i>109,106</i>	<i>1,064,928</i>	<i>2,366,509</i>
<i>13</i>	<i>128,864</i>	<i>5,612,265</i>	<i>14,439,939</i>
<i>14</i>	<i>8,576</i>	<i>772,716</i>	<i>5,728,102</i>
<i>15</i>	<i>34,145</i>	<i>1,368,873</i>	<i>4,038,793</i>
<i>16</i>	<i>9,981</i>	<i>1,162,629</i>	<i>7,261,515</i>
<i>17</i>	<i>32,638</i>	<i>2,255,939</i>	<i>2,777,441</i>
<i>18</i>	<i>24,810</i>	<i>1,410,789</i>	<i>6,542,067</i>
<i>19</i>	<i>67,338</i>	<i>1,789,599</i>	<i>2,366,509</i>
<i>20</i>	<i>105,422</i>	<i>942,786</i>	<i>5,570,416</i>
<i>21</i>	<i>36,960</i>	<i>2,290,075</i>	<i>17,109,096</i>
<i>22</i>	<i>8,859</i>	<i>460,450</i>	<i>6,196,655</i>
<i>23</i>	<i>30,344</i>	<i>2,598,543</i>	<i>7,503,247</i>

24	91,021	5,202,382	10,315,414
25	17,453	477,984	3,602,251
26	4,021	179,828	5,213,675
27	4,166	329,955	7,542,131
28	8,500	1,744,232	1,956,386
29	9,817	1,038,165	30,784,970
30	68,144	6,925,761	20,940,090
31	46,142	1,674,901	21,599,203
32	1,676	134,655	4,745,923
33	62,681	3,052,663	9,838,569
34	6,545	617,846	4,608,943
35	20,373	1,089,743	1,957,359
36	95,436	654,070	76,168,318
<i>Total</i>	2,392,732	70,128,543	350,306,140

- 2. One significant limitation of NDVI is the saturation effect, which occurs in areas with high vegetation cover and during peak seasons. This study focuses on forest offset projects, where forests can be very dense, and uses peak monthly NDVI values, which occur during peak seasons. This makes it challenging to use NDVI as an ideal outcome variable for accurately estimating the intervention effects. I suggest that the authors consider using higher-resolution forest data or other more suitable vegetation indices (such as kNDVI or LAI) to further validate the estimated effects. This would also serve as a robust test of the study's findings.

We sincerely appreciate your insightful comments regarding the limitations of NDVI, particularly the saturation effect. Thank you for educating us about the more suitable vegetation indices and suggestion on higher-resolution dataset. Your feedback has led to significant improvements in our methodology and analysis.

Our initial selection of NDVI was based on its widespread use in assessing vegetation status, including its ability to capture trends in aboveground standing biomass and various aspects of forest health (MacDougall et al., 2024; Ghafoor et al., 2022; Trujillo et al., 2012; Jiang et al., 2019; Maes et al., 2023). However, we acknowledge the limitations you pointed out, particularly the saturation effect in dense forests. To address these concerns and enhance the robustness of our findings, we have

implemented several changes and additional analyses.

Firstly, we adopted higher-resolution NDVI data at a 30 m × 30 m scale, representing a significant improvement over our previous 1 km × 1 km resolution with a factor of 1100. This enhancement allowed for more accurate identification of smaller polygons and yielded results consistent with our original findings (see Figures 1-3, Tables S1-S3).

Secondly, following your suggestion, we used kNDVI proposed by Camps-Valls et al. (2021) to mitigate the potential saturation effect as a robustness check. This analysis confirmed our main findings but also revealed larger effects, with a 4.25% increase in project areas and a 1.6% increase in buffer zones, suggesting that our estimated results represent a greater degree of improvement in vegetation quality and biomass increase in practical terms (see Table S1-S3, Fig.S3).

Lastly, we incorporated the Hansen dataset, widely recognized in forest studies, to conduct an additional robustness check using forest loss data from 2000 to 2023. Our analysis of forest loss, defined as a stand-replacement disturbance or a change from forest to non-forest state, showed no additional forest loss in the buffer zones and indicated a significant reduction in forest loss trends. These findings further support the positive spillover effects observed in our main analysis and alleviate concerns about unintended negative consequences.

Manuscript changes:

Page 4 paragraph 3 (line 117):

This divergence continued to grow over time, revealing overall positive effects of the CCER program. The baseline results of overall project sites show a positive effect of 2.25% to 4.25% with p-value less than 0.01 (Table S1).

Page 5 paragraph 3 (line 151):

Over 23 years, the average forest gain spillover of the project sites to neighboring areas is 0.91% to 1.60% (NDVI and kNDVI respectively, $p < 0.01$, Table S1). This effect is increasingly positive in the post-treatment period.

Page 6 paragraph 3 (line 178):

First, To address the non-linear relationship and saturation effect of NDVI in areas with high vegetation cover and during peak seasons, we employed kNDVI⁴¹ as an alternative vegetation index. Results remained consistent, with kNDVI matching results (Fig. S3) aligning with trends in Fig. 2a and Fig. 3a. The estimated effects were slightly larger in absolute value (Tables S1, S2, S3), possibly due to NDVI limitations in dense forest areas.

Page 18 paragraph 5 (line 564):

Our primary source is 30-m resolution NDVI data derived from all available Landsat 5/7/8/9 remote sensing imagery. This dataset is processed to remove clouds and shadows, and the annual NDVI is calculated as the maximum value from all valid Landsat observations throughout the year. Additionally, we employ 1-km spatial resolution NDVI data derived from the SPOT/VEGETATION PROBA-V 1 KM PRODUCTS, which provides an annual perspective on NDVI across China from 1998 to 2019. This moderate-resolution data is generated using a maximum value synthesis method, capturing the peak monthly NDVI values for each year, with values ranging from 0 to 1. Notably, NDVI can exhibit a saturation effect in dense forest areas, where the index becomes less sensitive to vegetation changes. This could lead to underestimation of effects and increased uncertainty in our analysis. To ensure a more accurate and robust analysis, we also employed kNDVI, an index proposed by Camps-Valls et al. (2021) that corrects for saturation effects.

Page 19 paragraph 3 (line 589):

The forest loss variable is derived from the Global Forest Change 2000-2023 dataset⁵⁵. In this dataset, forest loss is defined as a stand-replacement disturbance or a change from forest to non-forest state. Using 2000 as the baseline year, the data detects loss primarily in the year 2001-2023 at a high spatial resolution of 1 arc-second per pixel (approximately 30 meters per pixel at the equator).

Page 9 paragraph 4 (line 281):

Furthermore, we examined the trends in forest loss within buffer zones during the

period 2001-2023, utilizing data from Global Forest Change dataset⁵⁵. We used the annual percentage of forest loss relative to the 2000 baseline forest cover in buffer zones as our dependent variable (mean: 0.39%). The DID estimation revealed that in the full sample, forest loss decreased by 0.1% (about 25% reduction) following policy implementation. Projects with negative spillovers showed a non-significant 0.04% decrease, while those with positive spillovers exhibited a significant 0.2% (about 25%) reduction in forest loss (Table 1, columns 4-6). These findings indicate an absence of additional forest loss or carbon leakage in the buffer zones, while supporting the positive spillover effects of forest offset projects in reducing forest loss beyond their boundaries.

Table 1. The channels of the forestry spillovers.

	Land use			Forest loss			Wild fire	
	Cultivate	Forest	Grass	Full	Negative	Positive	Frequency	Freq/km ²
DID	-0.018***	0.002	0.024***	-0.001**	-0.000	-0.002**	-0.110*	-0.002**
estimator	(-17.59)	(1.26)	(25.80)	(-2.33)	(-0.94)	(-2.14)	(-1.77)	(-1.84)
Constants	0.181***	0.570***	0.237**	-0.003***	-0.002***	-0.004**	-0.556***	-0.004***
	(156.13)	(419.11)	(221.00)	(-2.76)	(-2.72)	(-2.53)	(-7.23)	(-5.38)
Obs.	1.5e6	1.5e6	1.5e6	1,633	644	989	1672	1672
R-squared	0.887	0.317	0.155	0.419	0.678	0.381	0.720	0.427

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics in parentheses.

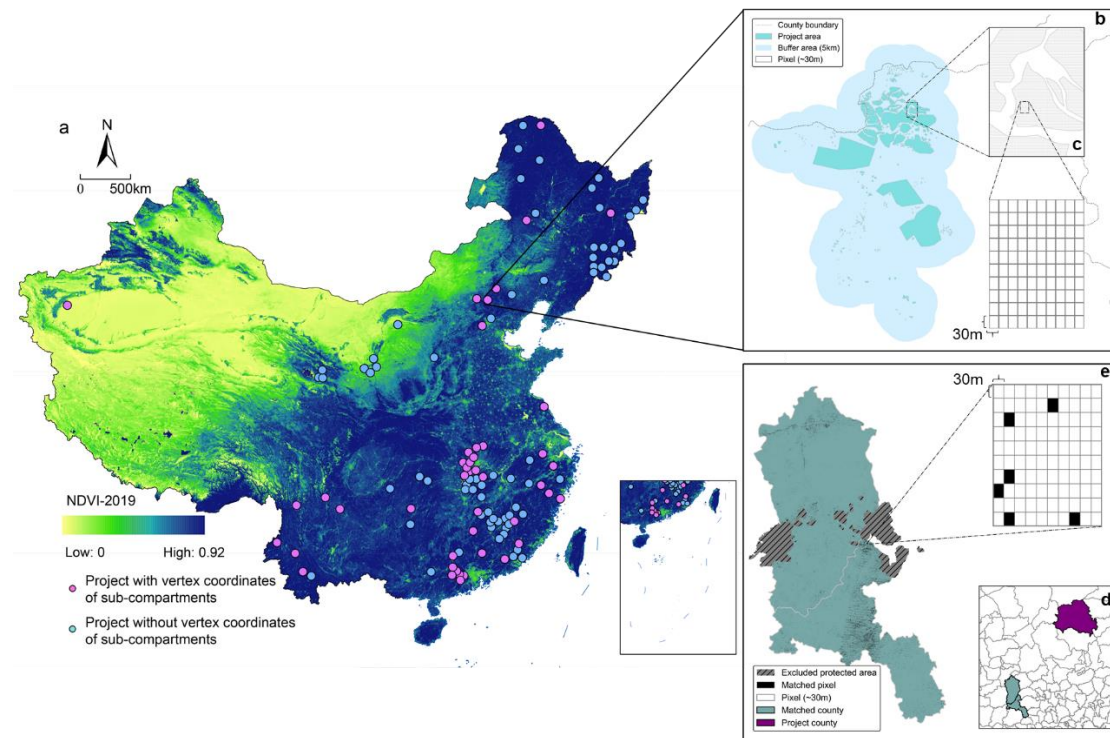


Fig.1. Study area and research design. Plot a depicts the national distribution of forest offset projects across China overlaid on the 2019 NDVI map, illustrating vegetation status. Red circles represent project locations with documented vertex coordinates (our study sample), while blue circles indicate projects without specified coordinates. Note that a single county may have multiple project applications, and a project may span multiple counties. In plot b, one specific project site (blue) and its 5 km buffer zones (light blue) are illustrated. Plot c lays out a 30m×30m grid (gray) on the project area, precisely delineating its boundary. Plot d displays the matched control counties (dark green) within the same province as the project locations (purple). Plots e demonstrate our grid-based matching results (see Methods), with control pixels sampled from the control counties, excluding those within established protected areas. The matched pixels are shown in black.

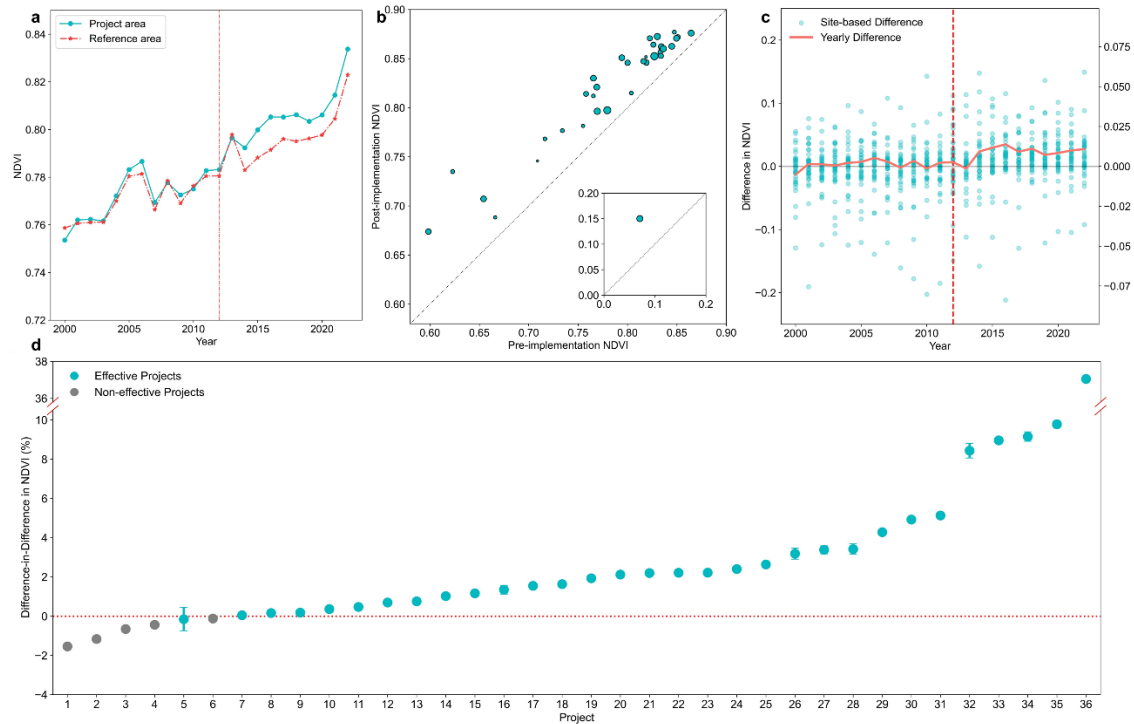


Fig.2. Forestry activities in project sites. Plot a illustrates the annual mean trends of NDVI in project areas and reference areas, with the red vertical line representing the year 2012. Plot b compares NDVI between the pre-implementation and post-implementation periods, with circle sizes indicating project area sizes. Plot c presents the annual site-based NDVI differences between each of the 36 project sites and their corresponding reference areas, as represented by blue dots for each year. The red line across the plot represents the mean of these differences across all project. Plot d presents the average treatment effect on the treated for individual projects estimated by difference-in-difference regression, where the whiskers extending above and below each point indicate the 95% confidence interval.

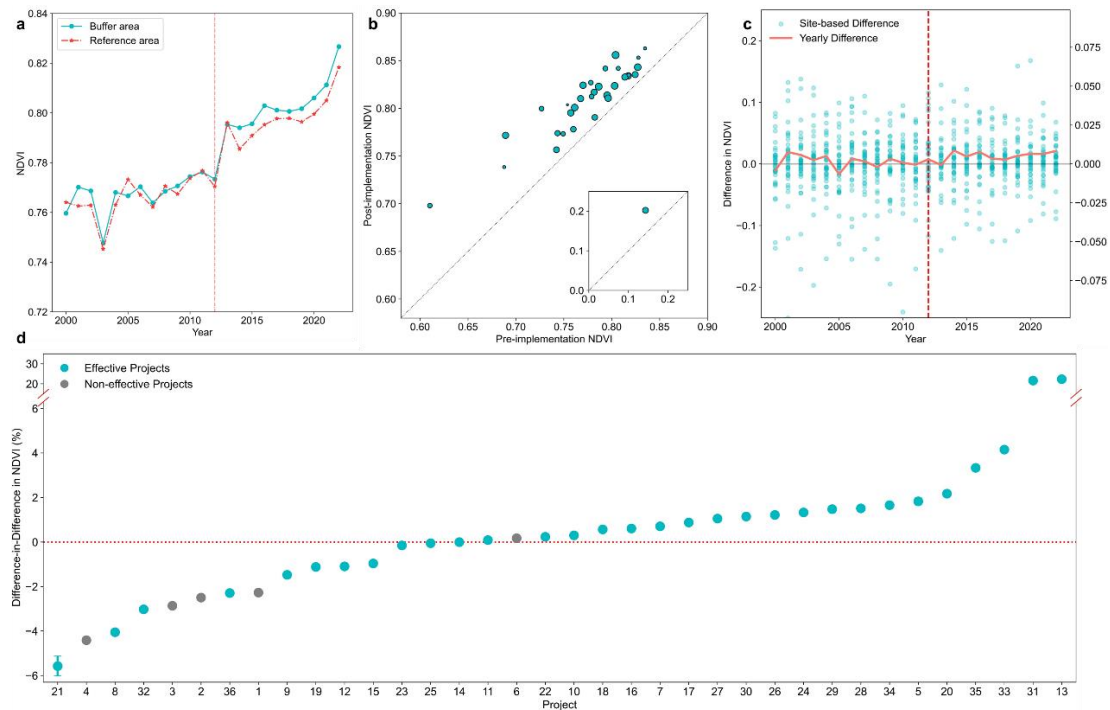


Fig.3. Forestry activities in buffers. Plot a illustrates the annual mean trends of NDVI in the buffer zones of the project and reference areas, with the red vertical line representing the year 2012. Plot b compares NDVI between the pre-implementation and post-implementation periods, with circle sizes indicating project area sizes. Plot c presents the annual site-based NDVI differences between each of the 36 project sites and their corresponding reference areas, as represented by blue dots for each year. The red line across the plot represents the mean of these differences across all projects. Plot d presents the average treatment effect on the treated for individual projects' buffer zones estimated by difference-in-difference regression, where the whiskers extending above and below each point indicate the 95% confidence interval.

Supporting information changes:

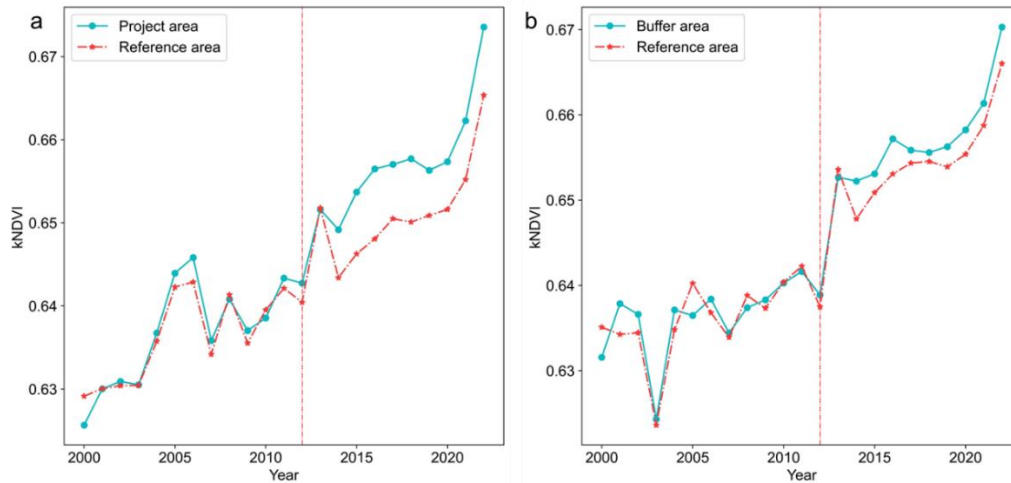


Fig.S3. Trends of kNDVI for project areas and buffer zones with the reference areas.

Note: Plot a presents the annual mean trends of kNDVI for project areas and the corresponding reference area. Plot b presents the annual mean trends of kNDVI for 5-km buffer zones and the corresponding reference area.

Table S1. Baseline results of project sites and buffer zones.

	(1)	(2)	(3)	(4)	(5)	(6)
Area	Project			Buffer		
Dependent	NDVI	NDVI	kNDVI	NDVI	NDVI	kNDVI
DID	0.0225*** (113.87)	0.0225*** (300.54)	0.0425*** (305.25)	0.0091*** (21.33)	0.0091*** (32.99)	0.0160*** (32.69)
post	0.0444*** (317.75)	0.1174*** (888.02)	0.1947*** (792.75)	0.0544*** (180.75)	0.1136*** (233.63)	0.1857*** (214.99)
treat	0.0037*** (26.73)	0.0038*** (73.12)	0.0062*** (64.209)	0.00113*** (3.83)	-0.00238*** (-12.48)	0.0084*** (24.94)
Cons.	8.869 (91771)	8.8472*** (16893)	8.402*** (8643)	18.059*** (86655)	18.094*** (34069)	8.457*** (8962)
FE	No	Yes	Yes	No	Yes	Yes
Obs.	106,566,063	106,566,063	106,566,063	16,918,564	16,918,564	16,918,564
R ²	0.003	0.86	0.87	0.005	0.58	0.60

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics in parentheses. Columns 1 and 2 display pooled regression outcomes for project sites taking NDVI as the dependent variable, with and without two-way fixed effects, respectively. Column 3 presents results for project sites using kNDVI as the dependent variable, incorporating two-way fixed effects. Columns 4 and 5 display the pooled regression results of the associated 5-km buffer zones of the project sites, using NDVI

as the dependent variable, without and with two-way fixed effects, respectively. Column 6 presents results for project sites using kNDVI as the dependent variable, incorporating two-way fixed effects. The substantial fluctuation in R^2 indicates that, captured by the fixed effect, time trends and individual characteristics have significant explanatory power and should not be omitted.

Table S2. Baseline results of individual project sites.

Project ID	Dependent Variable: ln (NDVI)				Dependent Variable: ln (kNDVI)				Obs.
	DID (%)	R^2	T-statistic	Standard error	DID (%)	R^2	T-statistic	Standard error	
1	-1.538	0.129	-51.703	0.000	-2.265	0.120	-44.91	0.001	3,100,584
2	-1.165	0.007	-43.656	0.000	-1.917	0.006	-39.071	0.000	16,733,559
3	-0.652	0.118	-34.644	0.000	-0.961	0.099	-30.892	0.000	2,808,484
4	-0.437	0.091	-53.979	0.000	-0.648	0.081	-48.839	0.000	20,227,764
5	-0.15	0.035	-0.495	0.003	0.436	0.034	0.768	0.006	269,107
6	-0.117	0.102	-7.509	0.000	-0.209	0.088	-8.4	0.000	2,876,334
7	0.051	0.059	4.398	0.000	0.096	0.047	5.017	0.000	6,966,194
8	0.159	0.055	2.24	0.001	0.114	0.052	0.877	0.001	2,703,639
9	0.182	0.161	9.672	0.000	0.289	0.142	9.75	0.000	1,187,766
10	0.365	0.061	19.814	0.000	0.652	0.05	20.585	0.000	3,602,076
11	0.474	0.027	16.374	0.000	0.607	0.018	12.025	0.001	1,969,214
12	0.698	0.098	69.753	0.000	1.278	0.085	80.448	0.000	5,018,876
13	0.76	0.019	21.491	0.000	1.287	0.015	20.075	0.001	5,923,690
14	1.025	0.086	14.632	0.001	1.69	0.071	13.843	0.001	394,496
15	1.168	0.086	47.193	0.000	1.99	0.071	47.591	0.000	1,570,670
16	1.351	0.029	12.057	0.001	2.411	0.021	11.664	0.002	459,084
17	1.547	0.129	66.724	0.000	2.503	0.110	66.129	0.000	1,404,702
18	1.637	0.125	41.08	0.000	2.721	0.112	40.655	0.001	1,110,255
19	1.933	0.053	105.922	0.000	3.452	0.047	112.973	0.000	3,097,548
20	2.121	0.139	188.258	0.000	3.22	0.123	182.622	0.000	4,849,412
21	2.203	0.089	66.626	0.000	3.601	0.073	62.931	0.001	1,758,932
22	2.214	0.087	52.09	0.000	3.549	0.080	51.581	0.001	407,514
23	2.221	0.05	85.413	0.000	3.59	0.041	82.577	0.000	1,395,824
24	2.403	0.078	70.992	0.000	3.966	0.064	71.163	0.001	1,363,575
25	2.64	0.184	113.274	0.000	4.188	0.179	116.332	0.000	802,838
26	3.189	0.173	21.801	0.001	5.126	0.172	20.441	0.003	184,966
27	3.383	0.211	32.335	0.001	5.838	0.209	32.912	0.002	140,346
28	3.421	0.073	25.245	0.001	6.364	0.065	26.082	0.002	390,264
29	4.284	0.137	88.239	0.000	6.618	0.128	83.22	0.001	451,582
30	4.924	0.097	178.549	0.000	8.386	0.080	175.594	0.000	3,134,612
31	5.135	0.182	68.307	0.001	9.696	0.166	69.937	0.001	2,122,163
32	8.438	0.238	43.908	0.002	15.32	0.233	44.307	0.003	77,096

33	8.962	0.111	333.34	0.000	13.906	0.090	297.598	0.000	2,883,302
34	9.153	0.073	73.547	0.001	15.516	0.069	69.456	0.002	301,070
35	9.785	0.224	100.959	0.001	17.903	0.220	102.448	0.002	488,474
36	37.051	0.417	449.436	0.001	72.997	0.414	450.493	0.002	4,390,051

Note: Columns 2 to 5 present the baseline regression results for 36 sample projects, with the NDVI as the dependent variable. Columns 6 to 9 replace the dependent variable with the kNDVI. "Obs." denotes the number of project observations.

Table S3. Baseline results of individual buffer zones.

Buffer ID	Dependent Variable: ln (NDVI)				Dependent Variable: ln (kNDVI)				Obs.
	DID (%)	R ²	T-statistic	Standard error	DID (%)	R ²	T-statistic	Standard error	
1	-2.273	0.077	-506.29	0.000	-3.597	0.069	-455.01	0.000	264,311,939
2	-2.497	0.013	-298.49	0.000	-4.035	0.009	-264.34	0.000	97,484,591
3	-2.866	0.193	-123.05	0.000	-4.751	0.171	-122.59	0.000	2,760,828
4	-4.414	0.218	-94.832	0.000	-8.171	0.204	-99.069	0.001	1,919,074
5	1.827	0.06	23.433	0.001	3.946	0.059	26.827	0.001	8,093,233
6	0.17	0.075	10.117	0.000	0.034	0.06	1.197	0.000	4,782,252
7	0.706	0.028	30.67	0.000	1.069	0.021	27.49	0.000	3,068,840
8	-4.053	0.128	-79.445	0.001	-7.712	0.125	-83.894	0.001	2,628,256
9	-1.475	0.023	-183.72	0.000	-2.352	0.017	-166.45	0.000	52,050,327
10	0.3	0.013	35.422	0.000	0.615	0.01	41.869	0.000	39,988,487
11	0.085	0.011	12.966	0.000	-0.137	0.008	-11.98	0.000	87,891,951
12	-1.096	0.031	-150.10	0.000	-1.637	0.024	-129.10	0.000	48,986,638
13	22.335	0.027	210.522	0.001	40.946	0.029	213.325	0.002	7,477,097
14	-0.004	0.052	-0.584	0.000	-0.009	0.043	-0.713	0.000	35,544,924
15	-0.963	0.008	-116.82	0.000	-1.484	0.006	-101.82	0.000	62,967,859
16	0.606	0.024	58.289	0.000	1.041	0.019	55.702	0.000	53,476,844
17	0.878	0.025	99.967	0.000	1.654	0.019	103.094	0.000	103,770,936
18	0.566	0.09	85.104	0.000	0.738	0.078	64.343	0.000	64,086,835
19	-1.118	0.026	-199.48	0.000	-1.784	0.02	-181.78	0.000	82,321,044
20	2.172	0.028	180.971	0.000	3.671	0.022	170.478	0.000	43,367,687
21	-5.575	0.011	-24.539	0.002	-14.207	0.011	-34.494	0.004	3,262,607
22	0.233	0.032	12.809	0.000	0.458	0.025	13.86	0.000	21,180,308
23	-0.152	0.032	-7.571	0.000	-0.176	0.025	-5.165	0.000	3,734,644
24	1.33	0.018	34.784	0.000	1.833	0.015	30.697	0.001	7,474,185
25	-0.055	0.007	-3.367	0.000	-0.471	0.006	-16.222	0.000	21,987,229
26	1.217	0.105	52.224	0.000	2.181	0.095	52.959	0.000	8,271,775
27	1.053	0.036	49.92	0.000	1.403	0.03	36.892	0.000	15,176,747
28	1.511	0.026	171.561	0.000	2.775	0.022	175.438	0.000	80,234,672
29	1.476	0.124	53.387	0.000	2.636	0.12	57.542	0.000	1,491,868
30	1.144	0.109	89.546	0.000	1.628	0.091	75.082	0.000	9,954,400

31	21.631	0.358	421.268	0.001	40.021	0.346	429.85	0.001	2,407,179
32	-3.024	0.068	-140.88	0.000	-4.83	0.06	-129.55	0.000	6,194,130
33	4.152	0.077	873.534	0.000	6.603	0.055	774.337	0.000	139,011,318
34	1.658	0.006	86.736	0.000	3.238	0.005	92.04	0.000	28,420,486
35	3.332	0.114	289.154	0.000	5.849	0.109	284.562	0.000	50,128,178
36	-2.295	0.082	-38.43	0.001	-4.546	0.081	-38.89	0.001	29,896,197

Note: Columns 2 to 5 present the baseline regression results for the buffer zones of 36 sample projects, with the NDVI as the dependent variable. Columns 6 to 9 replace the dependent variable with the kNDVI. "Obs." denotes the number of project observations.

- 3. From Table S12, it appears that most of the projects included in this study started after 2015. However, the estimated results indicate that these projects exhibited positive impacts, both within the project areas and in the surrounding buffer zones, as early as 2012-2015. Why might this be the case? How can these early positive impacts be genuinely attributed to the implementation of China's CCER market, rather than to other forest restoration projects or local environmental or ecological initiatives? Although the authors have taken into account the effects of other offset programs, other CCER projects, and protected areas during the matching process, China implemented many ecological restoration programs and various levels of ecological conservation projects. Could the influence of these projects potentially affect the attribution process?

We sincerely appreciate your insightful observation regarding the timing of project impacts and the challenge of attribution. Your comment has prompted us to provide a more detailed explanation of our methodology and findings.

Firstly, we apologize for not clearly explaining our choice of 2012 as the intervention year. This decision was based on several factors.

(i) Policy signal. 2012 marked the launch of China's voluntary emission reduction market. The policy's announcement created anticipation of potential financial support, motivating eligible entities to initiate activities even before formal project approval. This proactive approach is common in China, where work often begins before subsidy applications are submitted.

(ii) The process of project development. The initial phase of project development involves a period of project design, including defining project boundaries, demonstrating additionality, estimating emission reductions, and preparing construction contracts or project initiation documents. Consequently, project activities are likely to be in preparation or implementation stages before formal submission.

(iii) The economic incentives from emission reduction estimation. The timeframe for estimating emission reductions is not strictly tied to the project application date but rather to when activities diverge from the baseline scenario. This incentivizes project applicants to prepare afforestation or conservation activities in advance to obtain more carbon credits in the future.

(iv) Observed trends. As shown in Figures 2a and 3b, vegetation index growth rates in project sites and buffer zones significantly increased starting in 2013, the year following the policy announcement, indicating substantial improvements in vegetation quality and biomass. Therefore, 2012 is more appropriate as the start year to study the impact of the CCER program.

Secondly, regarding the attribution of forest activities to CCER implementation, we offer two possible explanations. We've incorporated additional context into our manuscript and supporting information to provide a more comprehensive understanding of CCER program implementation and its impacts in China.

(i) CCER projects must demonstrate uniqueness and additionality. Activities eligible for support from other projects are disqualified, reducing the likelihood of confounding influences within project areas. This requirement helps ensure that the observed effects can be more confidently attributed to the CCER program.

(ii) The advantages of matched difference-in-differences analysis. As you correctly noted, these initiatives are widely and uniformly distributed across the country, resulting in a general increasing trend in NDVI nationwide, which aligns with other research findings. In this context, the DID model's strength lies in its

ability to mitigate the influence of such confounding factors and common trends affecting NDVI changes in both treatment and control groups. This approach allows us to isolate the effects of the CCER program from broader environmental changes, including climate change and widespread ecological programs, thereby enabling a more accurate attribution of observed changes to the CCER implementation.

Thirdly, to further address concerns about potential confounding factors and support the possible explanations we raise above, we conducted additional analyses:

(i) Placebo tests. We performed two types of placebo tests using a bootstrapping procedure, randomly aligning treatment pixels and intervention periods. These tests simulated scenarios without forest offset projects or with interventions in different years: (a) created pseudo-treatment groups by randomly selecting samples from the control group, maintaining the original intervention period; (b) randomly selected both treatment groups and intervention periods between 2000 and 2022. Both tests were repeated 1000 times. The results (Fig. S6) demonstrate that our actual DID estimates lie at the extreme right tail of both the empirical and theoretical distributions, indicating that our observed effects are statistically significant and unlikely to result from random factors.

(ii) Robustness checks for other forestry programs. To examine whether our analysis results were influenced by other forest restoration projects or local environmental and ecological initiatives, we conducted robustness checks using representative projects. (a) We considered the National Forest City program, initiated in 2004, which has awarded this title to 219 cities to date. We categorized our study projects into two groups: those located in counties designated as National Forest Cities and those in non-designated counties. (b) We considered the National Forest Parks, established continuously since 2006, with China now hosting nearly 900 such parks. We divided our study projects into two groups: those in counties with National Forest Parks and those without. (c) We considered

the National Reserve Forest Project, which designates areas as key construction zones and moderate construction zones based on natural conditions and other factors. According to the list of counties within the project's scope, we categorized our study projects into three groups: those in counties designated as key construction areas, those in moderate implementation areas, and those beyond the project's construction scope. The results show that across all three policy-based groupings, the estimated effects remained significantly positive. This consistency indicates that our primary findings are robust to the presence of these forest restoration and environmental programs, mitigating concerns about potential confounding influences from these related initiatives to some extent.

Fourth, to account for discrepancies between the CCER launch year and the project application submission time, we substituted the intervention time with the public comment start date for each project, conducting the DID regression within a staggered framework recently proposed by Callaway and Sant'Anna (2021). The results remain consistent with baseline regression, which further confirmed the robustness of our results.

These additional analyses strengthen our confidence in attributing the observed NDVI changes primarily to spillover effects from the forest offset projects, rather than to other concurrent programs or random factors. We appreciate your critical comment, which has led to a more robust and comprehensive analysis.

Manuscript changes:

Page 17 paragraph 2 (line 505):

The post-intervention period is defined as 2012-2022, since the year 2012 marked the launch of China's voluntary emission reduction market, signaling potential financial support and incentivizing early action. Furthermore, project activities often commence before formal submission, and emission reduction estimations consider pre-application activities. Notably, Figures 2a and 3a demonstrate significant vegetation index growth starting in 2013, indicating a rapid response to the policy announcement.

These factors collectively justify 2012 as the appropriate start year for analyzing the CCER program's impact.

Page 6 paragraph 2 (line 171):

Fourth, to account for discrepancies between the CCER launch year and the project application submission time, we substituted the intervention time with the public comment start date for each project, conducting the DID regression within a staggered framework³⁹.

Page 7 paragraph 1 (line 194):

Additionally, the results are robust to the presence of other forest restoration and environmental programs.

Furthermore, we performed two types of placebo tests by randomly aligning the treatment pixels and intervention period using a bootstrapping procedure (Fig. S6). We first created pseudo-treatment groups by randomly selecting samples from the control group, maintaining the original intervention period. Then, we randomly selected both treatment groups and intervention periods between 2000 and 2022. Both tests were repeated 1000 times. These placebo tests simulated scenarios where either no forestry carbon sink projects occurred in the project areas, or other policy interventions happened in different years. The results show that our actual DID estimates lie at the extreme right tail of both the empirical distributions of the placebo estimates and the associated theoretical normal distribution. This implies a positive effect that is statistically different from zero, signifying that our observed treatment effects are unlikely to be the results of random factors and instead represent genuine policy intervention effects.

Supporting information changes:

1. Supplementing background

Chinese Certified Emission Reduction (CCER) market is a carbon credits program governed by China's government. It is open to project owners who are legally registered persons or organizations in China, and is linked to China's national carbon market

through the offset mechanism. The CCER market was initiated in 2012 when the National Development and Reform Commission issued the Interim Measures and key guidelines, followed by the launch of the National Trading Platform in October 2013. It is suspended in 2017, when the authorities had announced a total of 100 forestry projects on the official website. In 2023, it is re-launched with the Ministry of Ecology and Environment issuing new measures, guidelines, and methodologies.

Project implementation follows a process similar to many carbon credit systems, including project development and registration, validation and verification, and carbon credit issuance. Project developers must submit detailed Project Design Documents (PDDs) specifying project boundaries, planned activities, methodological compliance, baseline scenarios, and project additionality.

Site selection and boundary determination are governed by specific criteria outlined in the 2013 afforestation methodology (revised in 2023). Key conditions include: land must have been non-forested since February 16, 2005, with specific vegetation thresholds (must not simultaneously meet continuous area ≥ 0.0667 hectares, canopy cover $\geq 20\%$, tree height ≥ 2 meters after forest formation); clear land ownership rights, certified by county-level or higher government, are required; land must not be classified as wetland or organic soil; project activities must comply with soil and water conservation requirements; and the pre-existing agricultural activities should not shift from the intended project sites to other sites. It's important to note that forest property rights remain complex in China, impacting project implementation. Despite the Forest Tenure Reform initiated in the early 2000s, the confirmation and registration of forestland ownership continue to present significant challenges (Ho, 2001; Qin et al., 2011).

The 2023 revision of the interim measures for the CCER program introduced significant changes in project eligibility criteria. Notably, it stipulates that project activities cannot predate 2012. This revision also shifts the eligible project start date in the afforestation methodology from 2005 (as specified in the previous version) to 2012, ensuring consistency with the CCER program's launch. It's important to note that while

project activities could predate the generation of the PDDs, the crediting period is defined as the timeframe when the project scenario generates additional greenhouse gas emission reductions compared to the baseline scenario.

During the validation and verification period, when the project officially begins implementation, all activities within the project boundaries must be meticulously recorded and reported. This comprehensive documentation covers a wide range of aspects: afforestation activities, including planting details, tree species used, and land preparation methods; management activities such as patrols, replanting efforts, harvesting operations, pest control measures, and forest fire prevention strategies; and any occurrences of natural disasters or human disturbances.

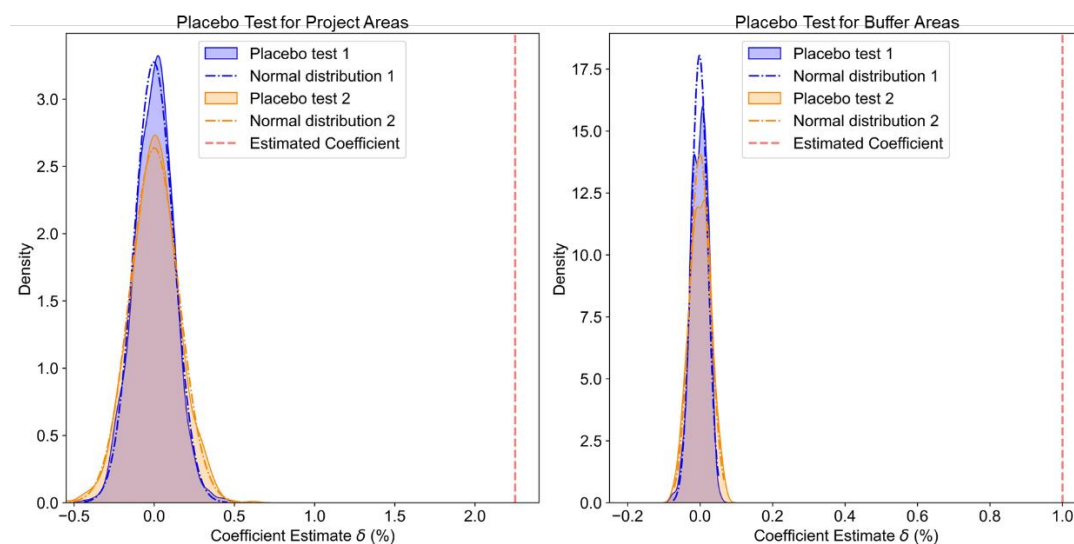


Fig.S6. Placebo test. The left plot is the placebo test for project sites and the right plot is for the buffer zones. The solid density curves illustrate the empirical distribution of the placebo estimates of the coefficient δ , the DID estimator. The superimposed dashed lines represent the theoretical normal distribution based on the central limit theorem, serving as a reference to assess how our empirical findings deviate from the expected distribution under the null hypothesis of no effect. Placebo test 1 (purple) displays the distribution of estimators for pseudo-treatment groups, while Placebo test 2 (orange) shows the distribution of estimators resulting from randomly assigning both treatment

groups and policy intervention years. The dotted vertical lines indicate our actual DID estimators obtained from the baseline regression.

Table S6. Robustness check.

	...	(11)					
	...	Staggered did					
DID		0.0032***					
		(14.39)					
Obs.		12,495,452					
...							
	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Forest_ City=1	Forest_ City=0	National_ Park=1	National_ Park=0	Reserve_ Forest=2	Reserve_ Forest=1	Reserve_ Forest=0
DID	0.0099***	0.0054***	0.0035***	0.0138***	0.0073***	0.0099***	0.0054***
	(44.12)	(11.83)	(9.11)	(25.50)	(10.49)	(44.12)	(11.83)
Obs.	6,599,342	9,833,755	9,318,591	4,773,692	2,340,814	6,599,342	9,833,755
R ²	0.20	0.60	0.28	0.78	0.23	0.20	0.60

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics are reported in parentheses (except for column 11). All columns include two-way fixed effects.Column 11 replaces the intervention year with the application submission year for each project, following the staggered difference-in-differences framework proposed by Callaway and Sant'Anna (2021). For column 11, ATT is reported with the z-values in the parentheses.

To examine whether our analysis results were influenced by other forest restoration projects or local environmental and ecological initiatives, we conducted robustness checks using representative projects. First, we considered the “National Forest City” program, initiated in 2004, which has awarded this title to 219 cities to date. We categorized our study projects into two groups: those located in counties designated as National Forest Cities and those in non-designated counties. The corresponding outcomes of baseline regression are presented in columns 10 and 11. Second, we considered the National Forest Parks, established continuously since 2006, with China now hosting nearly 900 such parks. We divided our study projects into two groups: those in counties with National Forest Parks and those without. The

corresponding outcomes of baseline regression are shown in columns 12 and 13. Lastly, we considered the National Reserve Forest Project, which designates areas as key construction zones and moderate construction zones based on natural conditions and other factors. According to the list of counties within the project's scope, we categorized our study projects into three groups: those in counties designated as key construction areas, those in moderate implementation areas, and those beyond the project's construction scope. The results are presented in columns 14-16. Notably, across all three policy-based groupings, the estimated effects remained significantly positive. This consistency indicates that our primary findings are robust to the presence of these forest restoration and environmental programs, mitigating concerns about potential confounding influences from these related initiatives to some extent.

- 4. From Fig. 1, I noticed that the authors created a fishnet covering the entire extent of the project sites. I wonder if the authors used all these pixels in the analysis (which was not reported). If so, did they consider the potential bias caused by the spatial autocorrelation of geospatial data? If not, please provide information on how the sample size was determined for each project site and how these samples in the analysis can reduce spatial autocorrelation.

We sincerely appreciate your constructive comments regarding the potential spatial autocorrelation within our geospatial data. We apologize for any confusion caused by the omission of details about our sample size in the initial manuscript.

In response to your first question: yes, we utilized all pixels covering the entire extent of the project sites, generated from the 30m $\times$ 30m grids in the revised version. We have now added this information to the appendix, providing the total sample size for each project site.

We acknowledge that we did not initially consider the potential bias caused by spatial autocorrelation. To address this concern, we have implemented spatial statistical methods and used spatially robust standard errors developed by Conley (1999, 2008). This approach is widely used to account for spatial dependence, allowing for correlation

between observations that are geographically close.

Furthermore, to enhance the robustness of our findings, we have conducted an additional analysis using a DID model at the project level. This approach involves aggregating the treatment and control pixels in both the buffer and reference areas. By aggregating data, we can reduce the spatial autocorrelation inherent in grid-based data, as the unit of analysis becomes larger and less prone to fine-scale spatial dependencies.

We believe these steps improve the robustness of our results and hope them address your concerns. We appreciate your careful review, which has led to these important improvements in our methodology.

Manuscript changes:

Page 6 paragraph 2 (line 169):`

Third, to address the concern about potential spatial autocorrelation in our geospatial data, we implemented spatial statistical methods and used spatially robust standard errors developed by Conley^{37,38}.

Page 6 paragraph 3 (line 190):

Third, to address the potential influence of the large sample size on the statistical significance of the results, we aggregated the treatment and control pixels in both project sites and the buffers to conduct baseline estimation on this smaller sample. The project-level results remain significantly positive and show only minor differences from our pixel-based estimates.

Supporting information changes:

(Previously reported changes, specifically the Table S9 (project-level sample sizes), is not repeated here)

Table S6. Robustness check.

	...	(3)	(4)	...	(10)
	...	Aggregate project	Aggregate buffer	...	Spatial SE
DID	...	0.0296*** (4.14)	0.0055*** (3.06)	...	0.0091*** (2.84)
Obs.		1,656	1,656		16,918,564

R^2	0.96	0.96	0.58
-------	------	------	------

- 5. Please add confidence intervals to Fig. 3c and Fig. 2c.

We sincerely thank you for your valuable suggestion regarding the description of Fig. 3c and Fig. 2c. We apologize for any confusion caused by the presentation of these figures. These figures illustrate the annual differences in the average and site-based NDVI between the project areas and reference areas. The x-axis represents the years, and the y-axis shows the NDVI differences between the project sites and reference areas. Each year is represented by 36 blue dots, each corresponding to the NDVI difference for one of the 36 project sites relative to its reference area, while the red line represents the mean of these differences across all projects.

Given the nature of these visual representations, there are no statistical estimates and confidence intervals depicted in these specific figures. We hope this explanation clarifies the presentation, and we have revised the manuscript to more clearly describe the results depicted in these figures.

Manuscript changes:

Page 5 paragraph 1 (line 133):

Fig.2. Forestry activities in project sites. Plot c presents the annual site-based NDVI differences between each of the 36 project sites and their corresponding reference areas, as represented by blue dots for each year. The red line across the plot represents the mean of these differences across all project.

Page 7 paragraph 3 (line 208) :

Fig.3. Forestry activities in buffers..... Plot c presents the annual site-based NDVI differences between each of the 36 project sites and their corresponding reference areas, as represented by blue dots for each year. The red line across the plot represents the mean of these differences across all project.

- 6. In the captions for Fig. 2d and Fig. 3d, I suggest indicating the criteria used to distinguish between effective and non-effective projects—specifically, whether this

is based solely on the sign of the regression coefficients or if statistical significance is also considered. Although the text suggests that the authors only use the sign of the regression coefficients to determine effectiveness, there remains a question regarding Fig. 2d. For example, the ATT for Project 7 and Project 8 is negative. Why are these still classified as effective projects?

We sincerely appreciate your astute observation regarding the classification of effective and non-effective projects in Fig. 2d and Fig. 3d. Your comment has highlighted an important area where we needed to improve both our analysis and our presentation of results.

First and foremost, we apologize for any confusion caused by unclear explanations or inconsistencies in our previous version. You are correct in pointing out the discrepancy with Projects 7 and 8 in the original Fig. 2d. We are grateful for your careful examination, which allowed us to identify and correct this error. In light of your comment and as part of our overall revision process using new, higher-resolution data, we have (i) explicitly stated in the manuscript that the classification of effective and non-effective projects is based on both the sign of the regression coefficients and their statistical significance, (ii) and updated the figures to accurately reflect our classification criteria and the results of our new analysis

Manuscript changes:

Page 4 paragraph 4 (line 126):

One site showed statistically insignificant effects and the remaining five projects (in gray) had negative DID estimators at the 1% level, most of whose nearby areas also exhibiting a negative trend (Fig 3d). The patterns observed in these negative projects suggest either the absence of forest gains or the influence of other systematic factors, including climate-related damage within a broader region. Therefore, these five significant negative projects are deemed as non-effective samples due to the lack of evidence of additional forestry activities inside the project boundaries.

- 7. Lines 127-129: These negative values represent the relative difference between the treated and control samples. There is a possibility that both the project area and the corresponding reference area experienced forest gains, but the forest gain in the project area was less than that in the reference area. I suggest that the authors provide the average change trends for each project and its reference area in the supplementary materials to determine whether the negative estimates are due to a lack of forest gains.

We sincerely appreciate your insightful observation regarding the interpretation of negative values in our results. You are absolutely correct in pointing out that negative values could indeed represent a scenario where both the project area and the corresponding reference area experienced forest gains, but the gain in the project area was less than that in the reference area. This is plausible given the overall upward trend in NDVI observed across project sites, buffer zones, control areas, and even nationwide.

We acknowledge that our initial terminology may have been imprecise. To clarify, what we are measuring is the additional forest gain attributable to the project activities, set against the context of a national increasing trend in forest cover (Piao et al., 2015; Wei et al., 2022). In cases where the project site's forest gain does not significantly exceed that of the reference area (which we consider as the counterfactual scenario representing the absence of project-related activities), we cannot conclude that there has been an additional forest gain due to the project. This interpretation applies both in cases where the relative difference is negative and where it is positive but not statistically significant.

To address this important point and provide greater transparency, we have revised our terminology as “additional forest gain” rather than simply “forest gain” when discussing project impacts. Following your suggestion, we have added supplementary materials showing the average change trends for each project and its reference area. This additional information allows readers to discern whether negative estimates are due to a lack of forest gains overall or a lack of additional gains relative to the reference area.

We are grateful for your careful review and constructive feedback, which has allowed us to present our findings with greater precision and nuance.

Manuscript changes:

Page 4 paragraph 4 (line 130):

Therefore, these five significant negative projects are deemed as non-effective samples due to the lack of evidence of additional forestry activities inside the project boundaries.

Supporting information changes:

Table S4. T test of individual project sites.

Project ID	Before	After	Diff1	T-Test	Treat	Control	Diff2	T-Test
1	0.769	0.821	0.052	330.98	0.794	0.802	-0.008	-72.29
2	0.780	0.797	0.018	188.09	0.788	0.788	0.000	0.62
3	0.850	0.871	0.021	239.40	0.860	0.830	0.030	417.88
4	0.827	0.852	0.025	610.49	0.839	0.828	0.012	374.00
5	0.666	0.688	0.022	21.67	0.677	0.660	0.017	23.01
6	0.834	0.862	0.028	312.68	0.848	0.847	0.001	14.41
7	0.835	0.860	0.025	409.83	0.847	0.845	0.001	32.52
8	0.654	0.707	0.053	189.04	0.680	0.633	0.046	228.58
9	0.851	0.872	0.021	230.23	0.861	0.834	0.027	350.36
10	0.864	0.876	0.012	239.41	0.870	0.847	0.023	358.27
11	0.819	0.846	0.027	182.64	0.832	0.827	0.005	48.71
12	0.836	0.860	0.024	529.84	0.848	0.841	0.006	159.32
13	0.769	0.796	0.026	183.79	0.782	0.777	0.005	48.19
14	0.765	0.812	0.047	133.54	0.788	0.783	0.005	20.33
15	0.816	0.847	0.031	298.50	0.831	0.815	0.016	173.56
16	0.734	0.776	0.043	102.33	0.754	0.766	-0.012	-39.33
17	0.826	0.864	0.038	345.19	0.844	0.824	0.021	225.73
18	0.758	0.814	0.056	250.56	0.785	0.786	-0.001	-9.49
19	0.845	0.863	0.018	240.53	0.853	0.839	0.014	217.32
20	0.830	0.872	0.042	635.50	0.850	0.852	-0.002	-38.75
21	0.800	0.846	0.046	287.49	0.822	0.835	-0.013	-108.96
22	0.804	0.815	0.011	52.17	0.809	0.804	0.005	29.22
23	0.834	0.853	0.019	160.78	0.843	0.825	0.018	191.89
24	0.823	0.871	0.048	317.90	0.846	0.860	-0.014	-110.57
25	0.833	0.857	0.023	219.40	0.844	0.823	0.021	212.78

26	0.819	0.852	0.033	92.05	0.835	0.742	0.092	181.51
27	0.709	0.746	0.037	67.69	0.727	0.754	-0.028	-67.71
28	0.716	0.768	0.052	83.21	0.741	0.799	-0.058	-141.13
29	0.847	0.877	0.030	126.74	0.862	0.860	0.001	5.53
30	0.766	0.830	0.065	556.74	0.796	0.799	-0.003	-31.51
31	0.598	0.674	0.075	254.51	0.635	0.769	-0.134	-665.01
32	0.709	0.736	0.027	38.25	0.722	0.687	0.035	51.98
33	0.794	0.851	0.057	486.18	0.821	0.836	-0.014	-152.61
34	0.755	0.781	0.026	56.03	0.768	0.731	0.036	93.35
35	0.623	0.735	0.112	254.92	0.677	0.679	-0.003	-7.66
36	0.070	0.150	0.080	746.53	0.109	0.076	0.032	474.70

Note: “Before” refers to the average NDVI values in the pre-implementation phase. “After” denotes the average NDVI values for the years from 2012 to 2022, indicating the post-implementation phase. “Mean” signifies the overall average NDVI values across all projects. “Control” represents the average NDVI values for the control group.

Table S5. T test of individual buffer zones.

Buffer ID	Before	After	Diff1	T-Test	Treat	Control	Diff2	T-Test
1	0.757	0.806	0.049	2305.88	0.781	0.772	0.008	552.23
2	0.795	0.814	0.018	559.98	0.804	0.797	0.007	312.10
3	0.835	0.863	0.028	269.60	0.848	0.792	0.056	634.08
4	0.754	0.804	0.050	249.06	0.778	0.697	0.081	506.22
5	0.541	0.581	0.040	157.74	0.560	0.677	-0.117	-716.50
6	0.807	0.842	0.035	389.08	0.824	0.818	0.006	95.71
7	0.828	0.853	0.025	220.22	0.840	0.836	0.004	50.97
8	0.688	0.738	0.051	241.89	0.712	0.653	0.059	362.05
9	0.817	0.834	0.017	474.87	0.825	0.802	0.024	927.18
10	0.825	0.835	0.011	265.05	0.830	0.829	0.000	9.43
11	0.803	0.824	0.021	687.07	0.813	0.808	0.005	234.73
12	0.814	0.833	0.019	587.60	0.823	0.803	0.020	826.33
13	0.545	0.576	0.031	98.22	0.560	0.579	-0.019	-94.86
14	0.782	0.817	0.035	941.42	0.799	0.796	0.003	108.48
15	0.796	0.811	0.014	392.85	0.803	0.794	0.010	368.96
16	0.757	0.795	0.038	853.75	0.775	0.784	-0.009	-275.08
17	0.786	0.823	0.036	1038.67	0.804	0.769	0.035	1424.40
18	0.770	0.824	0.054	1666.31	0.796	0.774	0.022	948.68
19	0.827	0.843	0.016	645.31	0.835	0.814	0.021	1133.36
20	0.768	0.810	0.043	807.47	0.788	0.811	-0.023	-646.13
21	0.433	0.450	0.016	37.35	0.441	0.497	-0.056	-189.85
22	0.760	0.778	0.018	233.77	0.769	0.811	-0.042	-812.42

23	0.818	0.835	0.017	166.44	0.826	0.836	-0.010	-140.30
24	0.778	0.827	0.049	349.20	0.802	0.849	-0.048	-246.12
25	0.782	0.790	0.008	108.17	0.786	0.789	-0.003	-52.72
26	0.779	0.812	0.033	392.23	0.795	0.732	0.063	836.36
27	0.744	0.774	0.030	353.03	0.758	0.783	-0.025	-418.94
28	0.761	0.801	0.039	1057.73	0.780	0.783	-0.003	-105.50
29	0.876	0.886	0.010	78.47	0.881	0.884	-0.003	-30.00
30	0.793	0.842	0.048	745.75	0.817	0.811	0.006	122.74
31	0.573	0.743	0.171	746.97	0.654	0.760	-0.106	-585.45
32	0.749	0.773	0.024	222.74	0.761	0.780	-0.019	-263.55
33	0.804	0.856	0.052	2717.74	0.829	0.868	-0.040	-2835.2
34	0.742	0.757	0.014	193.84	0.749	0.748	0.001	13.90
35	0.689	0.772	0.082	1513.24	0.729	0.709	0.020	523.21
36	0.141	0.200	0.058	649.93	0.169	0.165	0.004	62.174

Note: “Before” refers to the average NDVI values in the pre-implementation phase. “After” denotes the average NDVI values for the years from 2012 to 2022, indicating the post-implementation phase. “Mean” signifies the overall average NDVI values across all projects. “Control” represents the average NDVI values for the control group.

- 8. How can we be certain that the positive effects within the buffer zones are entirely due to the spillover effects of the projects? Is it possible that these buffer zones are influenced by other forest restoration projects, leading to positive effects relative to the reference areas, rather than stemming from the spillover effects of the projects themselves?

We sincerely appreciate your insightful question regarding the attribution of positive effects in the buffer zones to project spillovers, and we are grateful for the opportunity to address this concern.

Our attribution process of spillovers can be summarized as follows. We use matching techniques and difference-in-differences models to estimate NDVI changes, constructing counterfactual scenarios for both project sites and buffer zones. As illustrated in Fig. 2a and Fig. 3a, our constructed reference areas (control groups) show nearly identical pre-intervention trends with the treatment groups, diverging only after policy implementation. This DID analysis effectively eliminates the influence of confounding factors and common trends affecting NDVI changes in both treatment and

control groups, such as climate change.

However, we acknowledge your valid concern about potential confounding factors, such as other policies affecting vegetation changes in the region. To address this, we have undertaken the following additional analyses.

First, we performed placebo tests using a bootstrapping procedure, randomly aligning treatment pixels and intervention periods and repeating the process for 1000 times. The results (Fig. S6) demonstrate that our actual DID estimates lie at the extreme right tail of both the empirical and theoretical distributions, indicating that our observed effects are statistically significant and unlikely to result from random factors.

Second, we conducted robustness checks for other forestry programs, including the National Forest City program, National Forest Parks, and the National Reserve Forest Project. Our estimated effects remained significantly positive across all these policy-based groupings, suggesting that our findings are robust to the presence of these other forest restoration and environmental programs.

Third, we tested the time frame sensitivity by conducting analyses with shorter time frame (2007-2017 and 2009-2015) to mitigate potential confounding effects from long-term trends. The results of shorter framework generate larger spillover estimates.

Additionally, we compared our results with other studies using NDVI. Our study shows annual growth rates of 0.0035 in project areas and 0.0029 in buffer zones from 2000-2023, significantly higher than the range of 0.0006 to 0.0027 reported for China as a whole in previous studies. Furthermore, we compared our findings with studies on similar conservation programs. For instance, Chen et al. (2017) found NDVI increases of 0.0362 and 0.0148 for two programs in Inner Mongolia, while Zheng et al. (2024) reported growth rates ranging from 0.0005 to 0.0019 across eight afforestation initiatives in China. Our effects are comparable to these studies, providing additional confidence in our results.

While these measures significantly strengthen our confidence in attributing the observed positive effects to project spillovers, we acknowledge that we cannot be absolutely certain that these effects are entirely due to spillovers from our studied

projects. We have added a discussion of these limitations to our manuscript, emphasizing the need for further research to disentangle the complex interactions between various forestry initiatives and their impacts on surrounding areas. We appreciate your critical comment, which has led us to provide a more nuanced interpretation of our results.

Manuscript changes:

(Previously reported changes, including the placebo tests and analysis of other forestry programs, are not repeated here)

Page 6 paragraph 2 (line 173):

Lastly, to address potential temporal confounding during our long study period, particularly in light of China's intensified forestation efforts post-2000 and the initiation of several large-scale tree plantation programs in the early 2000s⁴⁰, we conducted additional estimations using shorter intervals 2007-2017 and 2009-2015.

Page 11 paragraph 3 (line 326):

Compared to existing literatures our findings also show a considerable magnitude. Our study reveals an annual NDVI growth rate of 0.0035 in project areas and 0.0029 in buffer zones from 2000-2023, significantly higher than the range of 0.0006 to 0.0027 observed for China as a whole during different periods in previous studies⁶¹⁻⁶⁴. Using a difference-in-differences approach, we identified additional average NDVI growth of 2.25% in project areas and 0.91% in buffer zones from 2012-2022, after controlling for common trends. These results are not only comparable to but in some instances surpass the effects found in other conservation programs and afforestation initiatives⁶⁴⁻⁶⁶, indicating the significant impact of CCER forest offset projects on vegetation quality.

Supporting information changes:

Table S6. Robustness check.

.....	(6)	(7)
.....	2007-2017	2009-2015
DID	0.0119***	0.0144***

	(30.35)	(29.10)
Obs.	8,088,606	5,148,618
R^2	0.58	0.58

*Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics are reported in parentheses..... Column 6 and 7 present the results of using study period of 2007-2017 and 2009-2015, respectively.*

- 9. The Project IDs in Fig. 2d do not appear to correspond directly with those in Table S2, which makes it very confusing and difficult for readers to locate and track each project's effect. This mismatch also seems to apply to Fig. 3d and Table S3. Moreover, it seems that the Project IDs in Fig. 2d and Fig. 3d do not correspond to each other either. I suggest standardizing the Project IDs across Fig. 2d, Fig. 3d, Table S2, and Table S3. Additionally, I recommend including the 95% confidence interval values and p-values in Table S2, as displayed in Fig. 2d.

We sincerely appreciate your meticulous observation regarding the inconsistencies in Project IDs across related figures and tables. We apologize for any confusion this may have caused and are grateful for the opportunity to address this issue.

Following your valuable suggestion, we have made significant revisions to ensure consistency across all relevant figures and tables. Specifically, we have standardized the Project IDs in Fig. 2d, Fig. 3d, Table S2, and Table S3, allowing readers to easily track and compare each project's effects across different representations of our data. This standardization should greatly enhance the clarity and usability of our results.

Regarding your recommendation to include 95% confidence interval values and p-values in Table S2, we appreciate the suggestion and have given it careful consideration. However, we decided not to include these additional statistics in the table for several reasons. Firstly, our tables S2 and S3 are already quite informative, especially after incorporating the standard errors and the new kNDVI reexamination results following your insightful suggestions. Secondly, our analysis using the new data revealed extremely narrow confidence intervals, with standard errors approaching zero (0.000). This is visually represented in the updated Fig. 2d and Fig. 3d, where the confidence

interval bars are barely visible. Given these factors, and in the interest of maintaining the table's clarity and visual appeal, we chose not to include the confidence interval values explicitly. Additionally, it's worth noting that the T-values, which are included in the tables, provide a good indication of statistical significance, serving a similar purpose to p-values in this context.

We hope you understand our decision to prioritize the most informative and essential statistics in our tables. We are deeply grateful for your attentive review and constructive suggestions, which have significantly improved the clarity and consistency of our data presentation.

Manuscript changes:

Page 5 paragraph 1 (line 133):

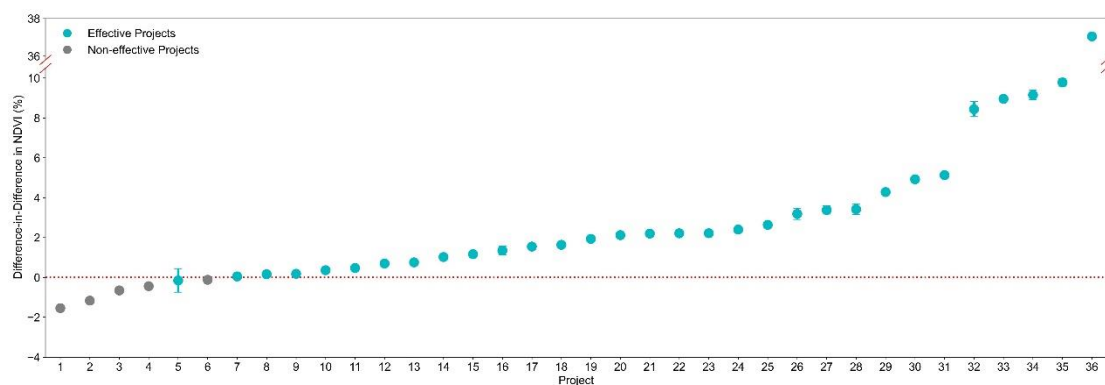


Fig.2. Forestry activities in project sites.....Plot d presents the average treatment effect on the treated for individual projects estimated by difference-in-difference regression, where the whiskers extending above and below each point indicate the 95% confidence interval.

Page 7 paragraph 3 (line 208):

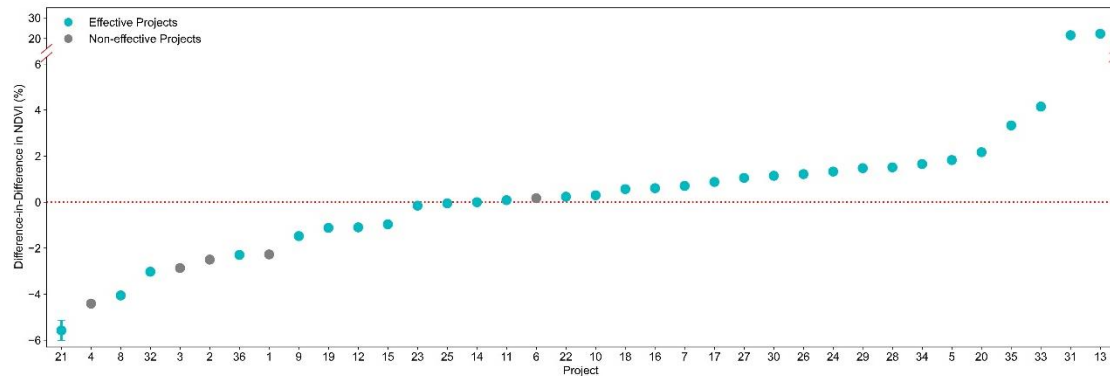


Fig.3. Forestry activities in buffers..... Plot d presents the average treatment effect on the treated for individual projects' buffer zones estimated by difference-in-difference regression, where the whiskers extending above and below each point indicate the 95% confidence interval.

Supporting information changes:

(Previously reported changes, including the Table S2 and S3 are not repeated here)

- **10. The inconsistent labeling of Project IDs makes it difficult to track the described effects in the charts when reading site-based estimations in Lines 122-130 and Lines 147-154.**

We sincerely appreciate your observation regarding the inconsistent labeling of Project IDs. We apologize for this oversight and any confusion it may have led.

To address this issue, we have implemented a comprehensive revision of our manuscript. We have standardized all Project IDs across figures, tables, and the main text. The descriptions in these lines have been updated to accurately reference the revised figures. Additionally, we've included an appendix with a comprehensive list of all projects, their standardized IDs, and brief descriptions of key characteristics to serve as a quick reference guide.

We hope these changes will make it easier for readers to track and understand the site-based estimations and their corresponding visual representations. We are grateful for your careful reading and feedback, which has been instrumental in enhancing the clarity and accessibility of our research. If you have any further questions or

suggestions for improvement, please don't hesitate to let us know.

Manuscript changes:

Page 4 paragraph 4 (line 122) :

Furthermore, we conducted a site-based examination to analyze individual projects (Fig. 2d; Table S2). Among all the available projects, 29 out of 36 project sites demonstrated statistically significant positive DID estimators, indicating positive program impacts in terms of forest gain. The majority of the estimated effects ranged between 1% and 5%, with the highest effect reaching 37%. One site showed statistically insignificant effects and the remaining five projects (in gray) had negative DID estimators at the 1% level, most of whose nearby areas also exhibiting a negative trend (Fig. 3d). The patterns observed in these negative projects suggest either the absence of forest gains or the influence of other systematic factors, including climate-related damage within a broader region. Therefore, these five significant negative projects are deemed as non-effective samples due to the lack of evidence of additional forestry activities inside the project boundaries.

Page 5 paragraph 3 (line 150):

The full sample trends indicate that the pixels in buffer zones and their control groups exhibited consistent patterns prior to 2012, with notable changes emerging after 2012 (Fig. 3a). Over 23 years, the average forest gain spillover from the project sites to neighboring areas ranges from 0.91% to 1.60% (for NDVI and kNDVI respectively, $p < 0.01$, Table S1). This effect becomes increasingly positive in the post-treatment period, reflecting strengthened forestry efforts and cumulative impacts as the CCER policy was implemented.

However, our project-specific analyses revealed significant variations in spillover effects. Individual site-based estimations (Fig. 3d; Table S3) showed that, out of 36 projects, regardless the five non-effective projects, 10 projects induced negative impacts to the buffer zones, ranging from -10.27% to -0.06% with $p < 0.01$; conversely, 20 projects significantly induced positive impacts, ranging from 0.09% to 22.34%; while

one project showed no significant effect. This highlights the complex nature of spillovers, which can vary across a wide range of settings, including the geographical location, the ecological setting, and the project features. Therefore, the existence, magnitudes and even directions of spillovers should be assessed within specific settings.

- 11. In Line 149, negative impacts ranging from -1.84% to -0.84% are reported, but these estimated values cannot be found in either Fig. 3d or Table S3. Are the DID estimates in Fig. 2d, Fig. 3d, Table S1, and Table S2 reported as original values or logarithm-transformed values?

We sincerely appreciate your careful examination of our results. We apologize for any confusion caused by the discrepancy of the manuscript descriptions and the appendix. Yes, the DID estimates in both Figures and Tables are logarithm-transformed values, which represent the percentage change of NDVI.

We acknowledge that we should have made this transformation more explicit and prominent in our presentation. To address this issue and improve clarity, we have made the following changes:

- (1) We have emphasized the logarithmic transformation in the Methods section to ensure it is more noticeable to readers;
- (2) In our updated figures (Fig.2d and Fig.3d), we now express the estimated effects as percentages. The y-axis labels have been changed to “Difference-in-Difference in NDVI (%)” for clarity.
- (3) In Tables S2 and S3, we have explicitly stated in the column headers that the dependent variables are $\ln(\text{NDVI})$ and $\ln(\text{kNDVI})$.

We believe these changes will provide a clearer and more consistent presentation of our results. We are grateful for your attentive review, which has helped us improve the accuracy and transparency of our work.

Manuscript changes:

(Previously reported changes, including the Fig.2d and Fig. 3d are not repeated

here)

Page 17 paragraph 1 (line 498):

$$NDVI_{it} = \alpha + \beta \cdot Treat_i + \gamma \cdot Post_t + \delta \cdot (Treat_i \times Post_t) + \mu_i + \theta_t + \varepsilon_{it} \quad (1)$$

where $NDVI_{it}$ represents vegetation status for project i in year t and is logarithm-transformed.

Supporting information changes:

Table S2. Baseline results of individual project sites.

Project ID	Dependent Variable: ln (NDVI)				Dependent Variable: ln (kNDVI)				Obs.
	DID (%)	R ²	T-statistic	Standard error	DID (%)	R ²	T-statistic	Standard error	
...	...	...	...	...	...	...	...	...	...

Table S3. Baseline results of individual buffer zones.

Project ID	Dependent Variable: ln (NDVI)				Dependent Variable: ln (kNDVI)				Obs.
	DID (%)	R ²	T-statistic	Standard error	DID (%)	R ²	T-statistic	Standard error	
...	...	...	...	...	...	...	...	...	...

- 12. Why is it that, regardless of the differences in project size and observations (as noted in Table S2), the number of observations for individual buffer zones is the same (as reported in Table S3)?

We sincerely appreciate your keen observation regarding the observation numbers for individual buffer zones in Table S3. We apologize for any confusion this may have caused and acknowledge that we should have provided a more detailed explanation in our previous manuscript.

In our earlier version, to manage computational costs, we merged buffer zones at different distances for each project into a single sample for calculation. This approach resulted in identical sample observation numbers for all buffer zones.

However, in response to your comment and other valuable feedbacks, we have significantly revised our methodology in the new version. We have overcome the computational challenges and now use pixel-level datasets for regression analysis in buffer zones, which is the same approach as the project sites. As a result, we have

updated Table S3 to accurately report the sample size for each project's buffer zone in the rightmost column, reflecting the true variation in observation numbers across different projects.

Furthermore, to enhance the robustness of our findings, we have also conducted an additional analysis using a DID model at the project level. This involves aggregating the treatment and control pixels in both the buffer and reference areas.

We hope these changes and additional analyses significantly improve the transparency and reliability of our results. We are grateful for your attentive review, which has prompted us to refine our methodology and presentation.

Manuscript changes:

Page 6 paragraph 3 (line 190):

Third, to address the potential influence of the large sample size on the statistical significance of the results, we aggregated the treatment and control pixels in both project sites and the buffers to conduct baseline estimation on this smaller sample. The project-level results remain significantly positive and show only minor differences from our pixel-based estimates. These findings underscore both the robustness of our conclusions and the importance of data resolution.

Supporting information changes:

(Previously reported changes, including the Table S2 and S3 are not repeated here)

Table S6. Robustness check.

	(3)	(4)
	Aggregate project	Aggregate buffer
DID	0.0296***	0.0055***
	(4.14)	(3.06)
Obs.	1,656	1,656
R ²	0.96	0.96

- 13. In Line 190-191, why are there only 8 years before treatment in Fig. 5a when

the observation period of this study starts from 1998?

We sincerely appreciate your astute observation regarding the discrepancy between the stated observation period and the pre-treatment years shown in the previous Fig. 5a. We want to clarify that while Fig. 5a displayed a truncated pre-treatment period of 8 years for visual symmetry, we did include the complete 13-year pre-treatment analysis in the appendix. However, we acknowledge that this presentation decision may have led to confusion and potentially misrepresented the full extent of our data. We apologize for any misunderstanding this may have caused.

In light of your comment and suggestions from other reviewers, we have made a significant change to this part of our analysis. We have decided to remove this figure and replace it with a new heterogeneity analysis that focuses on the effects across different buffer zone distances, which is expected to provide more informative and valuable insights relevant to our research questions.

We are grateful for your careful review, which has contributed to improving the clarity and relevance of our analysis.

- 14. In Line 200, the projects were categorized based on the per capita GDP of counties. However, as noted in the caption of Fig. 1, a project may span multiple counties. How did the authors categorize projects that span multiple counties?

We thank the reviewer for this insightful question. Among the 36 projects in our study, four span multiple counties. To address this complexity, we employed a weighted approach to categorize the per capita GDP of these projects. Specifically, for each of these four projects, we calculated a weighted average of the per capita GDP based on the proportion of the project area in each county. This weighted GDP was then used to rank and categorize these projects.

To ensure the robustness of our approach, we also conducted an alternative analysis where we ranked and divided all counties with offset projects into terciles based on their individual GDP values. Notably, this alternative method yielded almost identical results to our weighted approach.

- 15. In Line 201, it appears that positive spillovers are not observed in all groups, such as privately-owned projects and projects in medium-developed counties.

We sincerely appreciate your attentive observation regarding the discrepancy between our statement and the figure, which has helped us identify and correct this inaccuracy in our manuscript. We apologize for this oversight and the lack of precision in our original statement.

In light of your feedback, we have (i) deleted the inaccurate statements in Line 201 as you rightly pointed out; (ii) revised our discussion to accurately reflect the heterogeneity in spillover effects across different groups; (iii) thoroughly reviewed the entire manuscript to ensure consistency between our statements and the presented results.

Manuscript changes:

(Previously reported changes to statements regarding robustness checks are not repeated here)

- 16. Line 206-210, it is plausible that public agencies and state-owned enterprises potentially have better access to information, skills, and resources, which could lead to better project support. However, while this explains why public agencies and state-owned enterprises might better support their projects compared to private enterprises, it does not seem to account for why they also promote forest gain in the buffer zones.

We sincerely appreciate your insightful comment regarding the explanation of spillover effects in relation to project ownership. We acknowledge that this interpretation does not adequately explain the observed forest gains in buffer zones, as you rightly pointed out.

In light of your valuable feedback and suggestions from other reviewers, we have decided to make a substantial revision to this part of our analysis. We have removed the

discussion on ownership entirely, as it did not provide a robust explanation for the observed patterns of spillover effects.

Instead, we have incorporated a new discussion focusing on ecosystem types and their influence on spillover effects (see the response to your next comment). This new analysis provides a more ecologically grounded explanation for the variations in forest gains both within project areas and in buffer zones.

We are grateful for your critical evaluation, which has led us to reassess and improve this aspect of our study.

- 17. Since this study primarily focuses on subtropical and temperate regions, I suggest adding a comparison of the heterogeneity between projects in these two climatic zones.

We sincerely appreciate your insightful suggestion regarding the analysis of climate regions on spillover effects. We agree that this is a valuable heterogeneity analysis given China's diverse climatic conditions. In light of your feedback and suggestions from other reviewers, we have conducted several additional analyses.

First, we analyzed the influence of climate regions on outcomes, comparing 9 projects in temperate regions with 27 in subtropical regions. As shown in Table R1.1, both groups yield positive spillovers at a 1% statistical significance level. Projects in temperate regions generate larger effects (1.97%) compared to those in subtropical regions (0.48%).

Second, we divided the 36 projects into three equal groups based on project area (small, moderate, and large) and performed baseline regressions for each group. We found no significant differences in the estimated results across these groups, though larger projects exhibit slightly larger estimated effects.

Third, following other reviewers' suggestion, we also categorized the buffer zones of all 36 projects into four main ecosystem types: forest (26 projects), grassland (3 projects), farmland (6 projects), and desert (1 project). Key findings include an overall decrease in cultivated land and an increase in grassland and forest cover within the 5-

km buffer zones, suggesting positive spillovers through land-use conversion. Interestingly, we found a significant decrease in vegetation quality in the desert ecosystem buffer zone (one project in Xinjiang), possibly due to increased water consumption exacerbating local water scarcity. Further examination revealed a notable decrease in water surface area from 3.5% in 2010 to 1.9% in 2020, suggesting that water resource shortages may have intensified drought conditions, possibly leading to decreased vegetation quality in surrounding areas. These findings underscore the complex interactions between forestry projects and their surrounding ecosystems, emphasizing the importance of considering local ecological conditions in project planning and implementation.

Given the significance of these findings and space constraints, we have incorporated the ecosystem type analysis into the main manuscript. The climate region and project size analyses are included in the Supporting Information. We are grateful for your valuable suggestion, which has significantly enhanced the comprehensiveness and quality of our research.

Table R1.1. Additional heterogeneity test.

indicator	group	coef_did	s.e.	t-value	p> t	95% conf. interval	
climate	subtropical	0.0048	0.0003	15.2826	0.000	0.0042	0.0054
region	temperate	0.0197	0.0006	34.8353	0.000	0.0186	0.0208
	small	0.0084	0.0003	24.3956	0.000	0.0077	0.0090
project size	moderate	0.0079	0.0005	15.0025	0.000	0.0068	0.0089
	large	0.0112	0.0005	20.6950	0.000	0.0101	0.0122
	forest	0.0002	0.0003	-0.6755	0.499	-0.0008	0.0004
ecosystem	cultivated	0.0643	0.0007	88.0498	0.000	0.0629	0.0657
type	grass	0.0209	0.0007	30.7533	0.000	0.0195	0.0222
	desert	-0.0199	0.0044	-4.4823	0.000	-0.0285	-0.0112

Manuscript changes:

Page 8 paragraph 2 (line 227):

We further investigated how project features impact outcomes by categorizing projects based on county economic development level (grouped by per capita GDP), project type (forest management=6, afforestation=25, and vegetation restoration=5), and primary ecosystem type (forest=26, grassland=3, cultivated land=6, desert =1) (Fig. 4b). The positive spillovers are larger in less-developed regions, vegetation restoration projects, and farmland-predominant ecosystems.

First, regarding the economic status, the potential economic benefits and new technologies from the projects are appealing to economically disadvantaged communities^{42,43}, leading to larger positive spillovers in less-developed regions.

Second, we discuss the heterogeneity with respect to project type. The ecosystem services provided by afforestation and vegetation restoration programs have been extensively studied^{44–46}. These projects, including cropland-to-forest conversion, mine rehabilitation, and reduced deforestation, have been shown to influence local climate patterns⁴⁷, mitigate soil erosion and land degradation^{48,49}, and enhance ecosystem resilience. Such effects likely contribute to more substantial spillovers in surrounding areas.

Third, we distinguish the primary ecosystem type within the buffer zones. The substantial spillovers observed in farmland ecosystems align with the mechanism tests on land-use changes (see Table 1), where cultivated land experienced an overall decrease while grassland and forest cover increased within the 5-km buffers. This suggests that the spillovers may be partly attributed to the promotion of farmland-to-forest conversion, potentially driven by farmers' re-evaluation of revenue from forestry activities and offset projects when these are perceived as more profitable^{50,51}.

Conversely, a significant decrease in vegetation quality was found in the buffer zone primarily composed of desert ecosystem. Satellite imagery reveals predominant desert conditions both within and beyond the project sites (see the subplots in Fig. 2b and 3b). While local water resources are scarce, the PDD states that the project site is irrigated by rivers formed from mountain glacier melt and spring water. A plausible explanation is that increased water consumption by the afforestation project may have

exacerbated water scarcity in downstream and surrounding areas. This is supported by further examination revealing a notable decrease in water surface area from 3.5% in 2010 to 1.9% in 2020, suggesting intensified drought conditions and consequently decreased vegetation quality in surrounding areas⁵². These findings highlight the need to consider the complex interactions between forestry activities and local ecological contexts when planning and implementing related projects.

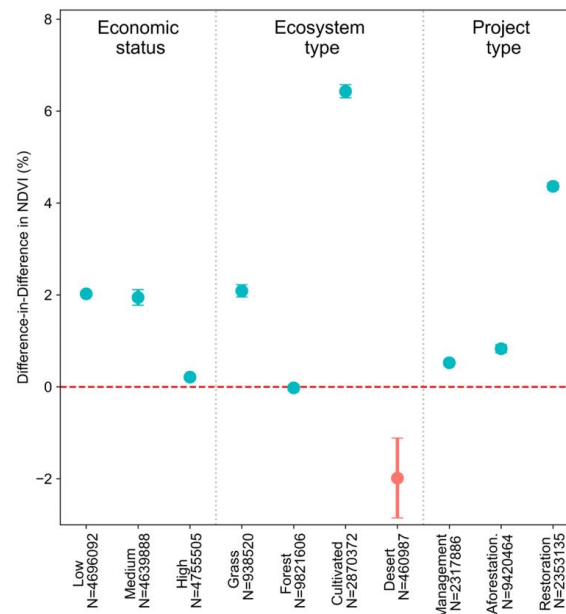


Fig.4. Heterogeneity test. ... The right plot presents a heterogeneity analysis based on three key features: economic status, primary ecosystem type, and project type; the whiskers extending above and below each point indicate the 95% confidence intervals for the respective estimates.

Supporting information changes:

Table S8. Additional heterogeneity test

	(1)	(2)	(3)	(4)	(5)
	Subtropical	Temperate	Small	Moderate	Large
<i>DID</i>	0.005*** (15.28)	0.020*** (34.83)	0.008*** (24.40)	0.008*** (15.00)	0.011*** (20.70)
<i>post</i>	0.080*** (143.04)	0.212*** (212.44)	0.082*** (135.92)	0.093*** (100.88)	0.166*** (174.06)
<i>treat</i>	0.001***	0.001***	-0.021***	0.005***	0.009***

	(4.53)	(3.11)	(-86.57)	(13.17)	(23.41)
Cons.	18.12***	18.06***	0.468***	18.17***	18.05***
	(32,394)	(22,186)	(530.5)	(23,092)	(22,527)
Obs.	12,620,372	4,291,444	5,656,596	5,645,265	5,616,703
R ²	0.267	0.796	0.208	0.320	0.731

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics in parentheses. The results of 9 projects in temperate regions and 27 in subtropical regions are presented in columns 1 and 2, respectively. The 36 projects are divided into three equal groups based on project area, namely the small, moderate, and large in columns 3-5.

- 18. Lines 227-228: Regarding leakage effects, I find the site-specific estimates in Fig. 3d indicate the presence of leakage effects, as we can observe that the buffer zones of some projects have negative estimates. The authors seem to consider that only forest loss leading to an increase in cultivated land can be regarded as 'leakage effects.' However forest loss does not necessarily result in an increase in cultivated land. Also, forest loss or negative ATT values themselves indicate the presence of leakage effects. I suggest revising the conclusion to state that the majority of forest offset projects in this study show no substantial leakage.

We sincerely appreciate your insightful comment regarding the interpretation of leakage effects in our study. Your observation is astute and has prompted us to reassess and clarify our findings.

You are correct in pointing out that forest loss or negative ATT values themselves can indicate the presence of leakage effects, and that forest loss does not necessarily result in an increase in cultivated land. We apologize for our imprecise wording and thank you for bringing this to our attention.

In light of your feedback, we have conducted additional analyses and revised our discussion of leakage effects.

First, we examined changes in forest land proportion in buffer zones from 2000 to 2020 (at five-year intervals). Contrary to leakage concerns, we found no decrease in forest land proportion. In fact, pooled regression analysis showed an increase in forest

land across all projects.

Second, we incorporated the Hansen dataset to conduct an additional robustness check using forest loss data from 2000 to 2023. Our analysis of forest loss, defined as a stand-replacement disturbance or a change from forest to non-forest state, showed no additional forest loss in the buffer zones and indicated a significant reduction in forest loss trends. These findings support the positive spillover effects observed in our main analysis and mitigate concerns about unintended negative consequences.

Third, we have revised our discussion on leakage to more accurately reflect why substantial leakage is unlikely in our study context. The CCER program's strict eligibility criteria often result in scattered, eligible plots within a broader landscape. While overall project sites are large, they could be very small in a single village or county, making it unlikely to yield large fluctuations in local timber markets or cause market-driven carbon leakage.

Fourth, following your suggestion, we have revised our conclusion to state that the majority of forest offset projects in this study show "no substantial leakage". We acknowledge that some projects show negative effects in buffer zones, which could potentially indicate localized leakage, but these effects are not widespread or substantial enough to negate the overall positive impacts of the projects.

We are grateful for your critical evaluation, which has led us to provide a more nuanced and accurate interpretation of our results.

Manuscript changes:

Page 9 paragraph 3 (line 267):

Overall, our analysis does not provide evidence of substantial leakage from forest offset projects in China. To investigate potential leakage, we examined land use dynamics in buffer zones of projects exhibiting negative spillovers, including non-effective projects. The leakage hypothesis would predict an increase in cultivated land use in buffer zones, reflecting a displacement of forest harvest activities from project sites to surrounding areas. However, our results contradict this expectation: a mean

decrease of 1.8% is observed in cultivated land use in 2015 and 2020 compared to the 2005-2010 average, accompanied by increases in grassland and forest cover of 2.4% and 0.2%, respectively (Table 1, columns 1-3). These land use changes align closely with those observed in the full sample (Table S7).

Page 9 paragraph 4 (line 281):

Furthermore, we examined the trends in forest loss within buffer zones during the period 2001-2023, utilizing data from Global Forest Change dataset⁵⁵. We used the annual percentage of forest loss relative to the 2000 baseline forest cover in buffer zones as our dependent variable (mean: 0.39%). The DID estimation revealed that in the full sample, forest loss decreased by 0.1% (about 25% reduction) following policy implementation. Projects with negative spillovers showed a non-significant 0.04% decrease, while those with positive spillovers exhibited a significant 0.2% (about 25%) reduction in forest loss (Table 1, columns 4-6). These findings indicate an absence of additional forest loss or carbon leakage in the buffer zones, while supporting the positive spillover effects of forest offset projects in reducing forest loss beyond their boundaries.

Page 11 paragraph 1 (line 311):

Using a quasi-experimental approach with a 23-year panel dataset from 2000 to 2022, we find that forest offset projects induce overall positive forestry spillovers to their buffer zones, with learning potentially being one of the contributing channels. Within the scope of our analysis, substantial leakage is not detected.

Page 12 paragraph 3 (line 371):

Additionally, considering the re-launch of CCER program in 2023, further research across a wider range of forestry projects in China is necessary to draw more comprehensive insights.

Page 11 paragraph 2 (line 318):

Second, the scattered and limited-scale nature of the offset projects makes it challenging for any single project to exert a substantial influence on the supply and demand dynamics of global timber or crop markets^{18,59}. Forest offset projects in China

are typically characterized by small, scattered, and disconnected polygons, mostly due to the land use features and strict site selection criteria. These include requirements for non-forested land since 2005 and have certified ownership rights, with the latter complicated by China's complex forest property system (see SI for details). These factors make it difficult to yield large fluctuations in the local timber market, as relatively small polygons are being regulated. Thus, leakage through shifts in market demand is less likely to occur.

- 19. The authors suggest that the occurrence of spillover effects is primarily through a learning channel. However, this seems to be based solely on inferences from prior literature. Are there any quantitative evidence or actual case studies that demonstrate the of this channel?

We sincerely appreciate your insightful question regarding the learning channel. Your comment has prompted us to clarify our findings and acknowledge the limitations of our study.

Our suggestion that learning could be one of the drivers of positive spillovers is based on both quantitative evidence from our study and inferences from prior literature. Our hypothesis posits that learning effects occur through knowledge dissemination and behavioral adaptation, leading to improved forest management practices in areas surrounding the project sites. We expected this learning effect to manifest in two ways. First, reduced forest-related risks: if learning occurs, we would expect to see a reduction in forest-related challenges in buffer zones as best practices spread. Second, increased project applications: if neighboring communities learn about the benefits of these projects, we would expect to see patterns of successive project applications in nearby areas.

Our quantitative evidence supports these expectations.

(i) Our analysis of wildfire frequency, using MODIS data from NASA spanning 2001 to 2022, demonstrates a notable reduction in wildfire occurrences. Specifically, we observed an average decrease in wildfire detection frequency of 11% in the entire

buffer zones and 0.2% per square kilometer (Table 1, columns 7-8). This reduction is consistent with improved forest management practices, potentially resulting from knowledge dissemination.

(ii) We observed that 60 out of 100 CCER forest offset projects were applied in the same cities during adjacent years. Over half of the studied projects, and 50% of the projects with evidence of positive spillovers, were located in cities with successive rounds of project applications. This pattern suggests information flow and learning among neighboring communities.

These findings, combined with insights from prior literature on knowledge dissemination and behavioral learning (references 15 and 27), support our hypothesis about the learning channel. However, we acknowledge that our study has limitations. We do not have actual case studies or community and individual-level data to directly demonstrate this channel, and we cannot entirely exclude other channels, which constitute limitations of our analysis.

In light of your comment and to ensure more rigorous presentation of our findings, we have taken the following steps:

(i) revised the conclusion to highlight the limitation of lacking actual cases at the resident level and presented the opportunity for future research.

(ii) revised our discussion of the learning channel, which is now presented as one possible drivers of positive spillovers rather than the primary driver.

(iii) thoroughly reviewed our manuscript and made appropriate revisions to ensure that our language accurately reflects the strength of our evidence throughout the paper.

Manuscript changes:

Page 1 paragraph 1 (Abstract, line 15):

Further analysis finds limited evidence of leakage, supported by land use change and the pattern of forest loss, possibly due to China's land policies and project features; instead, the positive spillovers are found to be facilitated through the learning channel of knowledge diffusion and information flow, supported by the reduction of wildfire

activities and the patterns of project applications.

Page 10 paragraph 2 (line 290):

Our study suggests that learning could be one of the drivers of positive spillovers in forest offset projects. This learning effect operates through knowledge dissemination and behavioral learning^{15,27}.

Page 11 paragraph 1 (line 311):

Using a quasi-experimental approach with a 23-year panel dataset from 2000 to 2022, we find that forest offset projects induce overall positive forestry spillovers to their buffer zones, with learning potentially being one of the contributing channels.

Page 12 paragraph 3 (line 363):

We acknowledge several limitations to this study. First, while we have some evidence regarding the impact of learning effects, a more comprehensive understanding of why and how forest managers respond to policies necessitates more granular data. This could include individual behavioral data obtained through field investigations and interviews, which would also facilitate exploration of other potential mechanisms, such as social norms.

- 20. Lines 333-335: What remote sensing data was used, and from which year? A potential issue is that if some non-forest areas were removed, but these non-forest areas later transitioned to forest (forest gain), would this introduce bias into the effect estimates?

We sincerely appreciate your thoughtful question. We apologize for any confusion caused by our lack of clarity on this point.

To address your first question, we used satellite imagery from the year each project's application was submitted to delineate the project boundaries.

For the issue of non-forest areas, we want to clarify that our process was more nuanced than simply removing non-forest areas. The project area geographic information we received was not in shapefile format (which is not publicly accessible), but rather as coordinates of the four corners of each polygon listed in the Project Design

Documents. We plotted these coordinates on a map to form irregular quadrilaterals. We then refined these boundaries using additional information from the PDDs, including project site boundaries images, satellite images, and project area map images.

To further improve accuracy, we used real-time satellite imagery from sources like Google Maps to refine and delineate each polygon, removing features such as rivers and residential areas. Our goal in this process was to create a more detailed and accurate representation of the project boundaries, aiming to replicate as closely as possible the actual project area delineation.

We acknowledge that our original wording of “removed some non-forest areas” was unclear and potentially misleading. A more accurate description would be that we removed some “non-project areas”. Our intention was to refine the project site boundaries to more accurately and comprehensively reflect the conditions within the project areas, rather than specifically excluding areas that were non-forested prior to project implementation.

We appreciate your careful consideration of these methodological details. Your question has prompted us to provide a more precise explanation of our approach in the manuscript. Thank you for helping us improve the clarity and rigor of our study.

Manuscript changes:

Page 13 paragraph 3 (line 401):

We utilized the coordinate data from the PDDs to delineate the boundaries of forestry sub-compartments for the 36 projects. These boundaries were further refined using additional information from the PDDs, including project area maps and satellite images. Remote sensing imagery was also used to create a more accurate representation of the project boundaries, removing non-project areas such as rivers and residential lands, ultimately generating the final project polygons (Fig.1).

- 21. Lines 337-340: If the sizes of different projects vary, does this imply that the magnitude or extent of their spillover effects might also differ? In that case, is it reasonable to set buffer zones of the same width for all projects?

We sincerely appreciate your insightful comments, which are pertinent and helpful, highlighting an important methodological consideration in our study. You are correct in suggesting that it may be more reasonable to set buffer zones of different widths for projects of varying sizes.

In our previous version, we had already addressed this concern by adjusting the width of buffers for each project based on their size as a robustness check. Specifically, we adjusted the buffer zone size for each project according to the quintiles of project area, using 0-1km, 0-2km, ..., 0-5km buffers. The pooled regression results of this adjusted sample remained robust, showing minimal difference from the main results.

However, your comment has prompted us to conduct more comprehensive analyses to further ensure the reliability and persuasiveness of our results. In our revised version, we have implemented the following additional analyses:

First, we categorized projects into three groups based on their size and conducted baseline regression tests using 0-5km buffer zones for each group. This analysis revealed no significant differences in spillover effects across project sizes.

Second, for each project, we examined effects in 1km intervals (0-1km, 1-2km, 2-3km, 3-4km, 4-5km). We found that most projects' effects were contained within the 5km range, with the strongest effects typically observed within the first kilometer (See details in the response to the next comment).

The results from these grouped tests have been incorporated into our manuscript as part of our robustness checks and heterogeneity analyses.

We appreciate your critical evaluation, which has led us to provide a more nuanced and comprehensive analysis of spillover effects in relation to project size and buffer zone width.

Manuscript and supporting information changes:

(Previously reported changes to statements regarding robustness checks, specifically the Table S9, are not repeated here)

- 22. Is there a difference in the spillover effects when using buffer zones of varying widths? How was the maximum width of 5 km determined? If the actual spillover effects of a project are smaller or larger than this width, is there a possibility of overestimating or underestimating the spillover effects?

We sincerely appreciate the insightful comment which touches on a crucial aspect of our study, and we are grateful for the opportunity to clarify our methodological approach and findings. We examined the spillover effects across varying buffer distances from 1 km to 5 km for all 36 projects in our sample, and the results of this analysis have been incorporated into the manuscript as a new section on spatial heterogeneity (see Fig. 5a).

This investigation yielded two key findings. First, regarding the spatial gradient of spillover effects, the density distributions reveal a notable pattern - as the buffer distance increases, the distribution of estimators becomes more concentrated around zero. This trend suggests a gradual decline in spillover effects as we move further from the project sites. Second, the average effects decline as the distance increases. We observed a substantial drop in average effects between the 1 km and 2 km buffers, from 0.85% to 0.15%, followed by a gradual reduction to approximately zero at greater distances. Notably, at the 5 km boundary, the effect is nearly negligible. These observations indicate a spatial gradient of forest project impacts on vegetation quality beyond project boundaries, with the primary impact occurring within the 1 km buffer zone. This pattern supports our choice of a 5 km buffer zone as an appropriate analytical boundary for capturing spillover effects.

Furthermore, our choice of the 5 km maximum width is also influenced by the unique characteristics of forest offset projects in China. These projects are often characterized by small, scattered, and disconnected polygons, unlike the large, contiguous areas typical in tropical rainforest regions. This fragmented nature results in

disproportionately large buffer zones. Our calculations show that, on average, the 5 km buffer area is more than 30 times the size of the project area itself, suggesting that this range is sufficiently comprehensive for our analysis.

In summary, our selection of the 5 km buffer zone represents a balanced approach, considering the observed spatial patterns of spillover effects, the specific characteristics of China's forest offset projects, and computational resource constraints. We hope this explanation addresses your concerns and provides a clearer justification for our methodological choices. Thank you again for your valuable questions.

Manuscript changes:

Page 8 paragraph 1 (line 218):

The spatial distribution of estimated spillovers for the 36 sample projects is presented in Fig. 4a, facilitating a comparison of spillover effects across different spatial scales. The density patterns reveal that as buffer distance increases, the distribution of estimators becomes more concentrated around zero. The average effects diminish, with a notable drop between the 1 km and 2 km buffers from 0.85% to 0.15%, followed by a gradual reduction to approximately zero at greater distances. This observed trend suggests a spatial gradient of forest project impacts on vegetation quality beyond project boundaries, with the primary impact occurring within the 1 km buffer zone. The pattern also indicates that the 5-km buffer zone serves as an appropriate analytical boundary for capturing spillover effects.

Page 9 paragraph 2 (line 258):

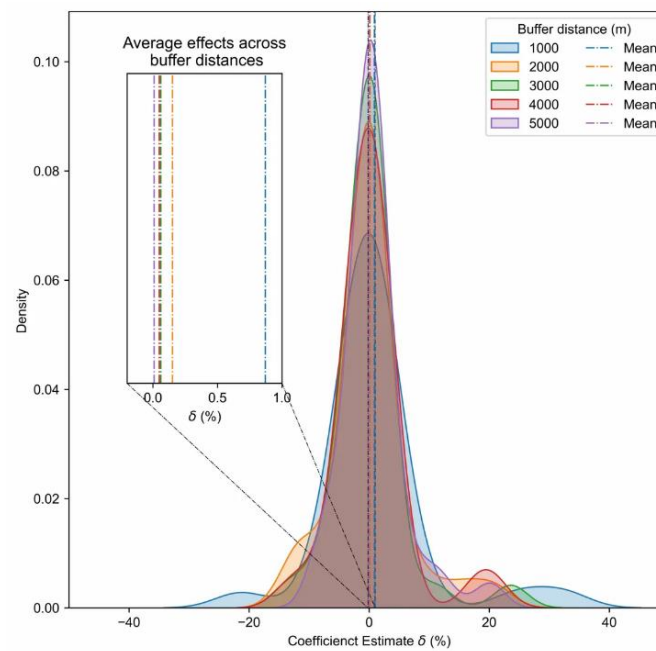


Fig.4. Heterogeneity test. The left plot illustrates the spatial pattern of estimated spillovers in the buffer zones. The main plot depicts the density distribution of DID estimators for buffer zones at distances of 1 km, 1-2 km, 2-3 km, 3-4 km, and 4-5 km from the project boundaries. The inset in the upper left corner displays the average effects across these buffer distances, with each bar representing the mean estimator value for a specific buffer interval.

Page 14 paragraph 2 (line 411):

As a result, the 5-km buffer zones around each project are, on average, thirty times larger than the project area. This characteristic is illustrated in the lower part of the example project shown in Fig. 1b. Our high-resolution data accurately captures both the project boundaries and the scattered small polygons.

- 23. Line 356: Please explain what CDM and VCS are.

We appreciate the reviewer's insightful comment regarding the use of acronyms. We recognize that these two abbreviations may not be immediately clear to all readers, especially those who are not familiar with the specific domain. To address this, we have revised the manuscript to include the full names and brief descriptions of these terms upon their first occurrence. We apologize for any confusion caused and hope the amendment enhances the clarity and accessibility of our work.

Manuscript changes:

Page 14 paragraph 3 (line 420):

First, counties with forest projects supported by other offset programs were excluded, including those under the Clean Development Mechanism of the Kyoto Protocol and projects certified by the Verified Carbon Standard, a widely used greenhouse gas crediting program.

- 24. Line 347: What is the source of the protected areas data? Does “national conservation” mean that only national-level protected areas were excluded? Given that spatial data on protected areas in China is largely incomplete, such as many city-level or provincial-level protected areas, could this data gap potentially affect the estimation results of this study?

Thank you for the thoughtful question regarding the protected area data used in our study. It sourced from the Geographic Remote Sensing Ecological Network Platform (gisrs.cn). This comprehensive dataset is compiled based on information provided by relevant authorities up to the end of 2020, including the Ministry of Ecology and Environment, National Forestry and Grassland Administration, and Ministry of Natural Resources.

We apologize for any ambiguity in our original wording. The term “national conservation” was not meant to imply only national-level protected areas, but rather nationwide protected areas across various administrative levels. To clarify, our dataset encompasses protected areas at four administrative levels: national, provincial, municipal, and county.

To improve clarity, we have added the data sources and revised the manuscript to provide a more accurate description. We appreciate your attention to this important aspect of our study, as it has allowed us to improve the clarity and completeness of our work.

Manuscript changes:

Page 14 paragraph 3 (line 424):

Second, established protected areas, namely the area involving conservation, were excluded.

Page 19 paragraph 2 (line 583):

The nationwide established protected area and ecosystem dataset was sourced from the Geographic Remote Sensing Ecological Network Platform (gisrs.cn), which is compiled based on information provided by relevant authorities.

- 25. After matching, how many treated samples were retained, and how many were discarded? Please provide this information.

We thank the reviewer for this insightful comment. The suggestion has indeed enhanced the clarity and transparency of our matching approaches and results.

In response, we have included additional information regarding the frequency of county and pixel usage as controls in our study. Out of approximately 2,840 counties in China, 45 counties host our sample projects. Following our filtering and matching processes, as detailed in the methodology section, 86 counties were selected for the matched control group, with 73 being unique. The repeated matched counties primarily occurred in Guangdong and Hubei provinces, where several cities were excluded pre-matching due to existing forest offset projects. On average, each matched county appears 1.17 times in our control group.

After converting the matched counties into grid cells, we obtained about 350 million pixels to serve as control pixels. The project sites, comprising approximately 2.39 million points, were matched 1:1, resulting in 0.6 million unique pixel points. Consequently, a control pixel is used an average of 4 times in the control group.

To provide a more comprehensive view, we have included detailed information on the pixel size of each project and its corresponding control area in Table S8. We believe this additional data offers a clearer understanding for both reviewers and readers. We are grateful for the opportunity to clarify and thank you for your valuable suggestion,

which has undoubtedly improved the quality of our study.

Manuscript changes:

Page 14 paragraph 3 (line 433):

For those implemented in more than one county, 1:1 matching was performed (Fig. 6a). We evaluated the quality of the matching, adjusted the relevant covariates, and iteratively optimized the control group counties (Fig. 6b). In the first few iterations, we found that some counties had been funded by other forestry carbon sequestration projects recently, indicating that our matching results were well representative. A total of 86 counties were selected into the matched control group after a series of filtration and matching, with 73 being unique. The repeated matched county mainly occurred in projects in Guangdong and Hubei provinces, where a number of cities were excluded before matching for the implementation of forest offset projects. After converting the unique matched counties into grid cells, we obtained about 350 million pixels to serve as control pixels.

.....

Thus, we performed 1:1 nearest-neighbor matching with replacement on each pixel plot from the pool of control pixels group, based on NDVI, land use, landform, precipitation, solar radiation, ecosystem, as well as land use type surrounding the project pixels (Fig. 6c, 6d; Table S10). Similarly, a Mahalanobis distance metric was used to measure the similarity between treated and control pixels. This process yielding 0.6 million unique control pixels for project sites, indicating that a matched control pixel is used 4 times on average in the control group. An identical matching approach was applied to the buffer zones, with more than 70 million pixels within the 5-km buffer zones. These matched pixels were eventually synthesizing into a reference area. As a result, the matching process has to be carried out individually for each of the 36 projects and their 5-km buffer zones, about 73 million pixels, which makes the process computationally demanding.

Supporting information changes:*Table S9. Project-level sample sizes.*

<i>Project ID</i>	<i>Project size</i>	<i>5-km buffer size</i>	<i>Control pool</i>
1	67,404	5,745,913	7,714,004
2	363,900	2,119,508	6,196,655
3	61,054	1,921,103	9,863,194
4	439,734	2,506,444	6,451,283
5	5,864	179,469	5,964,796
6	64,603	3,327,156	8,154,565
7	151,439	2,135,344	6,102,731
8	58,780	3,430,284	6,451,283
9	25,821	1,131,532	4,038,793
10	78,306	869,315	3,573,048
11	42,809	1,910,698	4,622,267
12	109,106	1,064,928	2,366,509
13	128,864	5,612,265	14,439,939
14	8,576	772,716	5,728,102
15	34,145	1,368,873	4,038,793
16	9,981	1,162,629	7,261,515
17	32,638	2,255,939	2,777,441
18	24,810	1,410,789	6,542,067
19	67,338	1,789,599	2,366,509
20	105,422	942,786	5,570,416
21	36,960	2,290,075	17,109,096
22	8,859	460,450	6,196,655
23	30,344	2,598,543	7,503,247
24	91,021	5,202,382	10,315,414
25	17,453	477,984	3,602,251
26	4,021	179,828	5,213,675
27	4,166	329,955	7,542,131
28	8,500	1,744,232	1,956,386
29	9,817	1,038,165	30,784,970
30	68,144	6,925,761	20,940,090
31	46,142	1,674,901	21,599,203
32	1,676	134,655	4,745,923
33	62,681	3,052,663	9,838,569
34	6,545	617,846	4,608,943
35	20,373	1,089,743	1,957,359
36	95,436	654,070	76,168,318
<i>Total</i>	2,392,732	70,128,543	350,306,140

References Cited in Response to Reviewer:

-
- [1] MacDougall, A. S., Esch, E., Chen, Q., Carroll, O., Bonner, C., Ohlert, T., ... & Seabloom, E. W. (2024). Widening global variability in grassland biomass since the 1980s. *Nature ecology & evolution*, 1-12.
- [2] Ghafoor, G. Z., Sharif, F., Shahid, M. G., Shahzad, L., Rasheed, R., & Khan, A. U. H. (2022). Assessing the impact of land use land cover change on regulatory ecosystem services of subtropical scrub forest, Soan Valley Pakistan. *Scientific Reports*, 12(1), 10052.
- [3] Trujillo, E., Molotch, N. P., Goulden, M. L., Kelly, A. E., & Bales, R. C. (2012). Elevation-dependent influence of snow accumulation on forest greening. *Nature Geoscience*, 5(10), 705-709.
- [4] Jiang, P., Liu, H., Piao, S., Ciais, P., Wu, X., Yin, Y., & Wang, H. (2019). Enhanced growth after extreme wetness compensates for post-drought carbon loss in dry forests. *Nature Communications*, 10(1), 195.
- [5] Maes, J., Bruzón, A. G., Barredo, J. I., Vallecillo, S., Vogt, P., Rivero, I. M., & Santos-Martín, F. (2023). Accounting for forest condition in Europe based on an international statistical standard. *Nature Communications*, 14(1), 3723.
- [6] Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., ... & Wang, Y. (2015). Detection and attribution of vegetation greening trend in China over the last 30 years. *Global change biology*, 21(4), 1601-1609.
- [7] Wei, Y., Sun, S., Liang, D., & Jia, Z. (2022). Spatial-temporal variations of NDVI and its response to climate in China from 2001 to 2020. *International Journal of Digital Earth*, 15(1), 1463-1484.
- [8] Chen, H., Shao, L., Zhao, M., Zhang, X., & Zhang, D. (2017). Grassland conservation programs, vegetation rehabilitation and spatial dependency in Inner Mongolia, China. *Land Use Policy*, 64, 429-439.
- [9] Zheng, L., Lu, J., Liu, H., Chen, X., & Yesou, H. (2024). Evidence of vegetation greening benefitting from the afforestation initiatives in China. *Geo-Spatial Information Science*, 27(3), 683-702.
- [10] Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S.,

Duveiller, G., Cescatti, A., ... & Running, S. W. (2021). A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, 7(9), eabc7447.

[11]Conley T G. GMM estimation with cross sectional dependence[J]. *Journal of econometrics*, 1999, 92(1): 1-45.

— (2008). *Spatial Econometrics*. In: *Microeconometrics*. Ed. by S. N. Durlauf and L. E. Blume. London: Palgrave Macmillan, 303–313.

We would like to express our deepest gratitude to the reviewer again. We feel incredibly fortunate to have received such thoughtful feedback from you and to have the opportunity to improve our analysis. We have implemented substantial modifications based on your valuable feedback. Your meticulous review helps us significantly enhance the quality and accuracy of our research. Thank you again for your valuable time and professional guidance.

Response to Reviewer #2

We thank the reviewer for the valuable feedback on our original manuscript. Following your suggestions, we have thoroughly revised the paper to enhance the clarity and readability. We also rewritten the discussion and conclusion section to highlight our contributions. We hope the revised manuscript has been significantly improved. Please do not hesitate to let us know if you have any further concerns or comments.

Below we have addressed all your comments. Your original comments are provided in **bold**, followed by our responses. Excerpts from the revised manuscript are provided in *italics*.

The manuscript “Leakage or Benefit? Spillovers from a Forest Offset Program in China” is on an interesting topic and based on a large dataset (but see comment below). The study is based on a combination of interesting and sophisticated analytical approaches that as such are conducted properly. But I have doubts about the importance of the program itself, which doesn't seem to be widely implemented in China. I don't recognize that this article can be published without addressing the following issues.

-1. First of all, I don't know what the ecosystem type of the buffer is, and I think that could have a big impact and error on the results. I think breaking down the ecosystem types of the buffer zone and discussing them in detail might make the study more scientifically comprehensive.

We sincerely appreciate your insightful comment regarding the ecosystem types in buffer zones. Your suggestion is indeed crucial and has led us to conduct a more comprehensive analysis of this aspect.

Given the wide geographical distribution of the projects, the ecosystem composition of their buffer zones is indeed complex and varied, comprising seven main

types: forest, farmland, grassland, desert, freshwater, settlement, and other ecosystems. We performed additional heterogeneity tests to explore the spillover effects across different ecosystems, categorizing the buffer zones of all 36 projects into four main types based on the predominant ecosystem: forest (26 projects), grassland (3 projects), farmland (6 projects), and desert (1 project). As you correctly anticipated, the results reveal significant variations across different ecosystems.

The largest spillover effects were observed in farmland ecosystems, followed by grassland ecosystems. This aligns with our mechanism tests on land-use changes, which found an overall decrease in the proportion of cultivated land and an increase in grassland and forest cover within the 5-km buffer zones. These findings suggest that positive spillovers may be partly attributed to the promotion of farmland-to-forest or farmland-to-grassland conversion. It could be possibly through the dissemination of technology or social norms, or driving by farmers' re-evaluation of potential revenues from forestry activities and offset projects (Latta et al., 2016). The economic incentive can influence land use choices and agricultural practices, particularly when they perceive the forestry activities more profitable, promoting participation in programs like the Returning Farmland to Forest Program (Song et al., 2014; Zhang et al., 2017).

Interestingly, we found a significant decrease in vegetation quality in the buffer zone primarily composed of desert ecosystem. This pertains to only one project in our sample located in Xinjiang. Satellite imagery shows that this project was predominantly desert before the project implementation (see Fig. 2b NDVI mean, where the pre-implementation NDVI mean was approximately 0.07). Post-implementation, the NDVI mean more than doubled. The project site is irrigated by rivers formed from high mountain glacier melt and spring water. The increased water consumption by the afforestation project may have exacerbated water scarcity in downstream and surrounding areas. Further examination revealed a notable decrease in water surface area from 3.5% in 2010 to 1.9% in 2020, suggesting that water resource shortages may have intensified drought conditions, possibly leading to decreased vegetation quality in surrounding areas. This finding aligns with previous research in similar regions (Lu et

al., 2018), highlighting the potential for unintended consequences in water-scarce environments.

These findings underscore the complex interactions between forestry projects and their surrounding ecosystems, emphasizing the importance of considering local ecological conditions in project planning and implementation. We have incorporated these additional analyses and discussions into our manuscript to provide a more nuanced and comprehensive understanding of the spillover effects across different ecosystem types. We are grateful for your valuable suggestion, which has undoubtedly improved the quality and comprehensiveness of our research.

Manuscript changes:

Page 8 paragraph 2 (line 227):

We further investigated how project features impact outcomes by categorizing projects based on county economic development level (grouped by per capita GDP), project type (forest management=6, afforestation=25, and vegetation restoration=5), and primary ecosystem type (forest=26, grassland=3, cultivated land=6, desert =1) (Fig. 4b). The positive spillovers are larger in less-developed regions, vegetation restoration projects, and farmland-predominant ecosystems.

.....

Third, we distinguish the primary ecosystem type within the buffer zones. The substantial spillovers observed in farmland ecosystems align with the mechanism tests on land-use changes (see Table 1), where cultivated land experienced an overall decrease while grassland and forest cover increased within the 5-km buffers. This suggests that the spillovers may be partly attributed to the promotion of farmland-to-forest conversion, potentially driven by farmers' re-evaluation of revenue from forestry activities and offset projects when these are perceived as more profitable^{48,49}.

Conversely, a significant decrease in vegetation quality was found in the buffer zone primarily composed of desert ecosystem. Satellite imagery reveals predominant desert conditions both within and beyond the project sites (see the subplots in Fig. 2b

and 3b). While local water resources are scarce, the PDD states that the project site is irrigated by rivers formed from mountain glacier melt and spring water. A plausible explanation is that increased water consumption by the afforestation project may have exacerbated water scarcity in downstream and surrounding areas. This is supported by further examination revealing a notable decrease in water surface area from 3.5% in 2010 to 1.9% in 2020, suggesting intensified drought conditions and consequently decreased vegetation quality in surrounding areas⁵⁰. These findings highlight the need to consider the complex interactions between forestry activities and local ecological contexts when planning and implementing related projects.

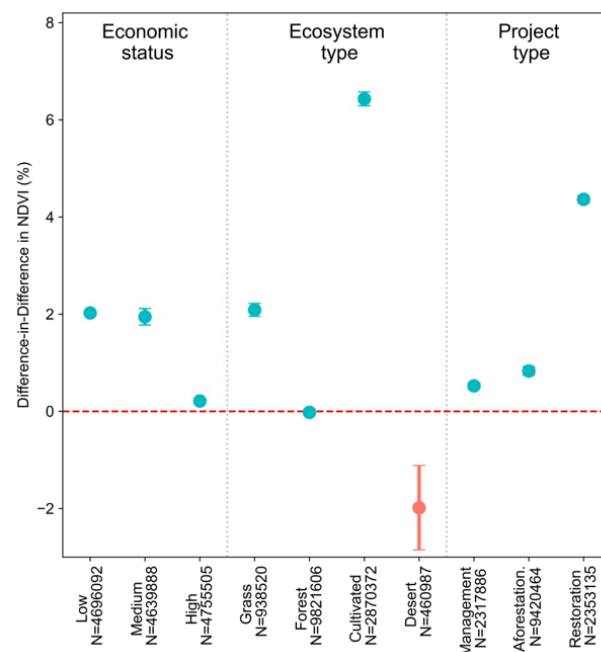


Fig.4. Heterogeneity test.The right plot presents a heterogeneity analysis based on three key features: economic status, primary ecosystem type, and project type; the whiskers extending above and below each point indicate the 95% confidence intervals for the respective estimates.

-2. Secondly, the selection of reference areas I think has some problems, especially in the type of afforestation. Silviculture involves the size of the area, the species of trees and the ecological characteristics of the area, and it is very difficult to find a

suitable control group in a number of only 100. I have doubts on this point, which may cause a big error on the accuracy of the study.

We sincerely appreciate your insightful comment regarding the selection of reference areas, which addresses a crucial aspect of our methodology.

We apologize if our initial description led to any misunderstanding. To clarify, our approach does not rely on finding a suitable control group from a limited sample of 100. Instead, we employ a grid-based sampling approach and two-step statistical matching among over 350 million potential control pixels:

1. Grid-based sampling: We utilize a high-resolution grid-based sampling approach at 30m×30m, allowing for a much finer-grained analysis.

2. Two-step statistical matching: This approach ensures that each pixel within our sample projects is matched with pixels from areas within the same province that have not undergone forestry carbon sink projects or been designated as conservation areas. The matching process considers both socioeconomic conditions and geographic-biological factors (including vegetation quality, land use, landform, precipitation, solar radiation, ecosystem, as well as land use surrounding the project sites), addressing the ecological characteristics you mentioned.

3. Extensive control pool: After converting the matched areas into grid cells, we obtained approximately 350 million pixels to serve as potential control pixels, compared to 2.3 million pixels within the project boundaries. This extensive pool of control pixels allows for more accurate and reliable matching.

To demonstrate the effectiveness of our matching process, we compared the standardized mean difference of covariates before and after matching, a crucial step in ensuring the validity of our causal analysis. The Fig. 6b illustrates the standardized mean differences of covariates between treated and control counties before and after matching. The Fig. 6d presents a similar comparison but at the pixel level. Notably, the plot demonstrates a substantial reduction in the standardized mean differences across all covariates post-matching, which is evident from the clustering of filled circles closer to the zero-line compared to the open circles. These visualizations collectively confirm

the successful alignment of covariates between treated and control areas, thereby strengthening the robustness of our methodological approach in constructing appropriate control groups.

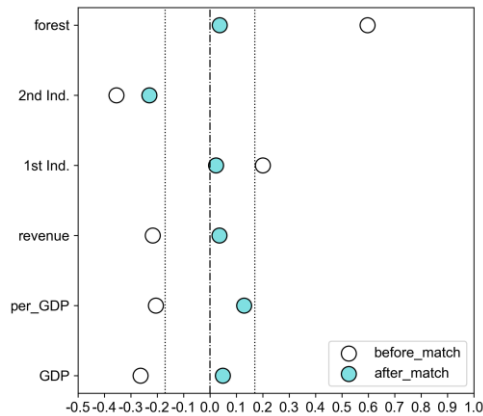


Fig. 6b

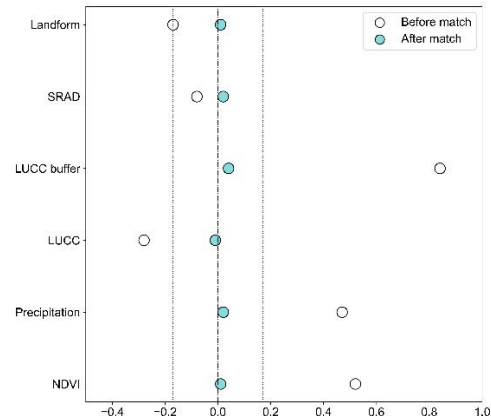


Fig. 6d

We have revised our methods to provide a clearer explanation of our control group selection process. Additionally, we have updated our results to reflect the application of refined methodology. To further address potential concerns about the adequacy of our control group, we have included an appendix that details the number of pixels for each project and the corresponding number of available matching control pixels.

We believe the revised methodological details and the extensive nature of our control pixel pool should alleviate concerns about the accuracy of our study with respect to reference area selection. The large number of potential control pixels allows us to account for the various factors you mentioned to a much greater extent than might be possible with a smaller sample. We appreciate your valuable feedback, which has prompted us to provide a more comprehensive explanation of our methodology.

Manuscript changes:

Page 13 paragraph 1 (line378):

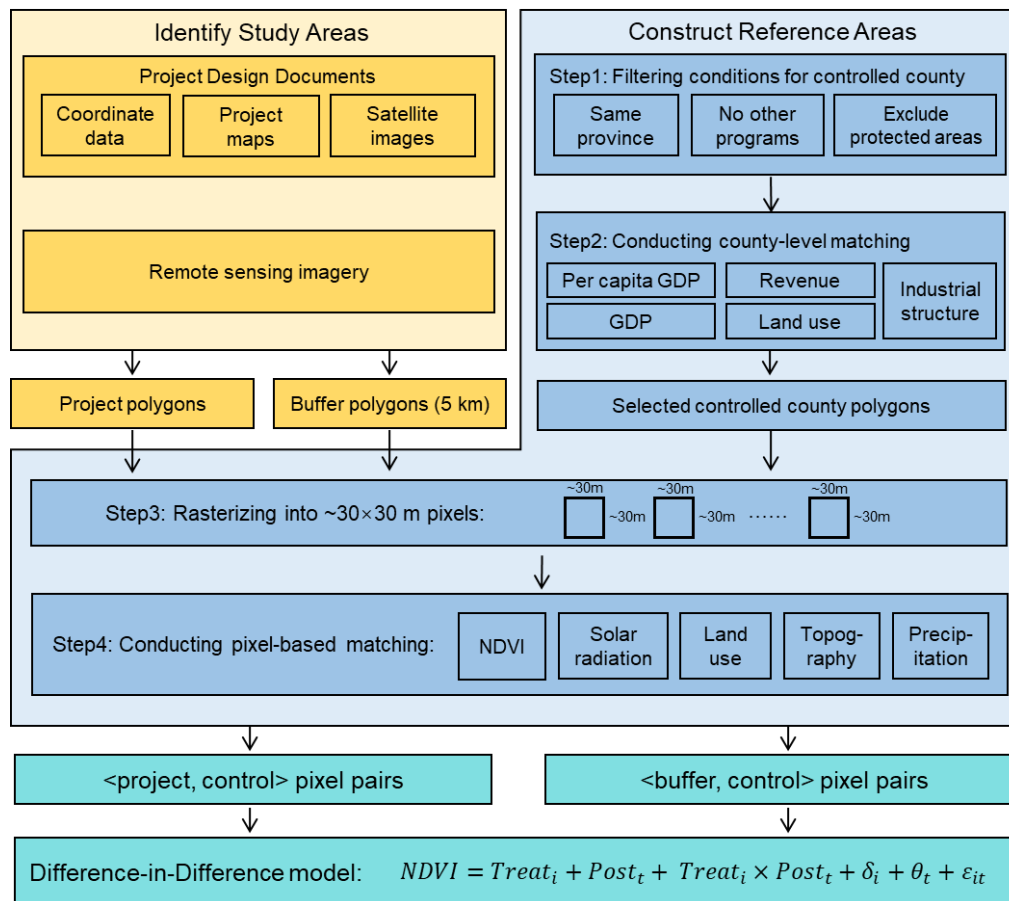


Fig.5. Flowchart of methods. First, the project areas were identified and calibrated using remote sensing imagery. Thus, 5-km buffer zones were created for each project. Both the project and buffer polygons were rasterized into about 30m × 30m pixels. Second, the reference areas were constructed following data exclusion and a two-step statistical matching process: the control groups were restricted to the same provinces without other forest offset projects, and the protected areas were excluded; county matching was conducted using county-level socioeconomic covariates; and pixel matching was conducted using pixel-level geo-biophysical covariates. These steps constructed references areas for both project sites and buffers, generating the matching pixels. Third, difference-in-differences regressions were run for the entire offset portfolio using pooled data, as well as for each individual project, to estimate the impact of CCER programs on additional forestry activities of forest offset projects.

Page 14 paragraph 2 (line 413):

Our high-resolution data accurately captures both the project boundaries and the

scattered small polygons. In total, the 36 projects yielded approximately 2.39 million pixels

.....

Before matching, we conducted data exclusions and filtration. First, counties with forest projects supported by other offset programs were excluded, including those under the Clean Development Mechanism of the Kyoto Protocol and projects certified by the Verified Carbon Standard, a widely used greenhouse gas crediting program. Additionally, counties with other CCER projects were not considered in our analysis. Second, established protected areas, namely the area involving conservation, were excluded. Third, control counties were restricted to the same province as project sites.

.....

A total of 86 counties were selected into the matched control group after a series of filtration and matching, with 73 being unique. The repeated matched county mainly occurred in projects in Guangdong and Hubei provinces, where a number of cities were excluded before matching for the implementation of forest offset projects. After converting the unique matched counties into grid cells, we obtained about 350 million pixels to serve as control pixels.

Secondly, we constructed a reference area, namely a counterfactual control area for each project and their corresponding buffer zones, based on biophysical and land-use variables, using a grid-based sampling strategy and statistical matching. This ensured equal coverage and similar features of project areas, which had multiple project types and initial land use types (non-forested, sparse-forested, cultivated, grasslands, unused, or degraded land). A nationwide 30m×30m resolution grid (consistent with the NDVI base data grid) was overlaid on the project area and the matched county, using the Create Fishnet tool in ArcMap v.10.5. Thus, we performed 1:1 nearest-neighbor matching with replacement on each pixel plot from the pool of control pixels group, based on NDVI, land use, landform, precipitation, solar radiation, ecosystem, as well as land use type surrounding the project pixels (Fig. 6c, 6d; Table S10). Similarly, a Mahalanobis distance metric was used to measure the similarity

between treated and control pixels. This process yielding 0.6 million unique control pixels for project sites, indicating that a matched control pixel is used 4 times on average in the control group. An identical matching approach was applied to the buffer zones, with more than 70 million pixels within the 5-km buffer zones. These matched pixels were eventually synthesizing into a reference area. As a result, the matching process has to be carried out individually for each of the 36 projects and their 5-km buffer zones, about 73 million pixels, which makes the process computationally demanding.

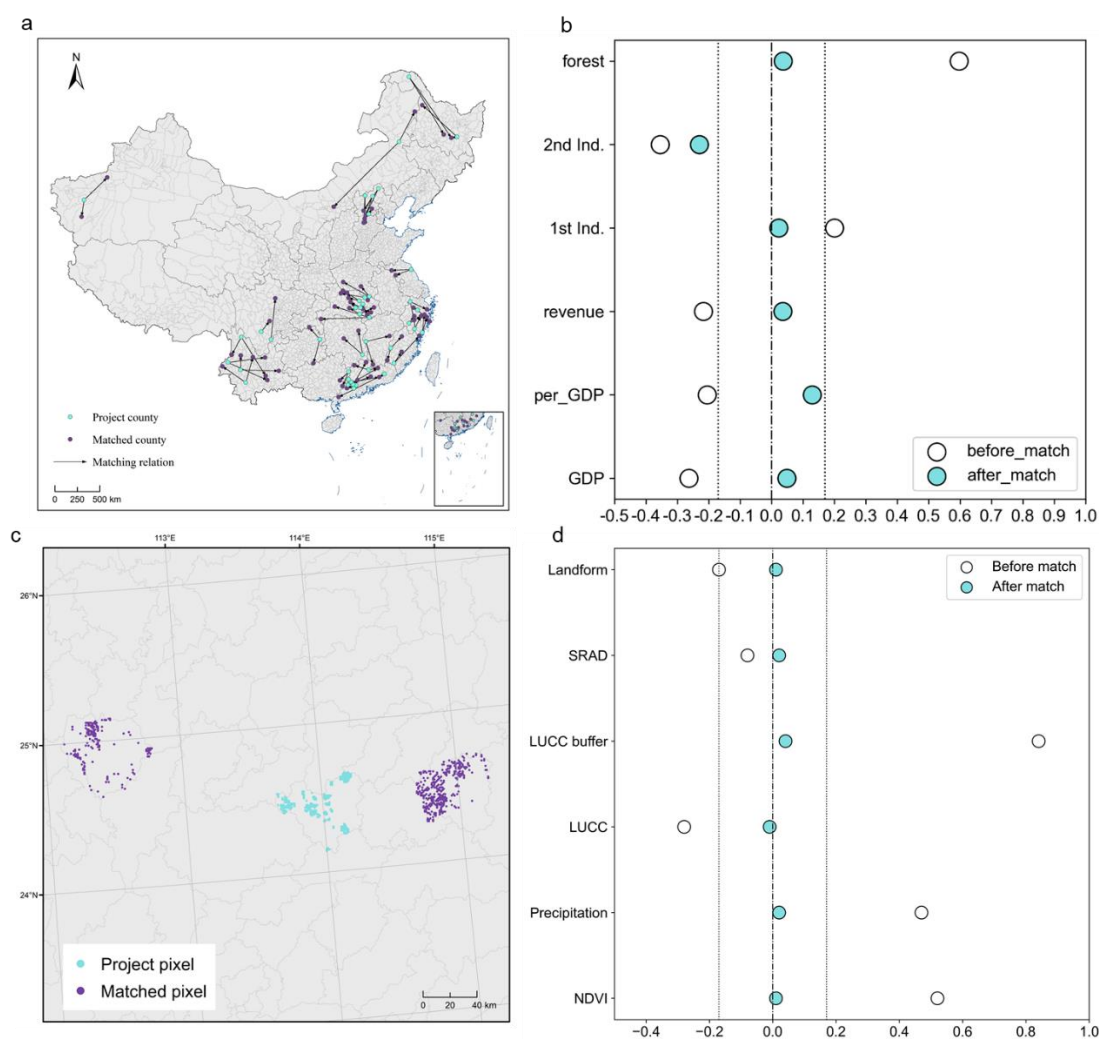


Fig.6. Spatial matching and balance assessment. Plot (a) delineates the distribution of project sites and their matched counties, connected by lines to emphasize the geographic pairing. Plot (b) compares the standardized mean difference of covariates between treated and control counties before and after matching. The open circles

represent the standardized mean differences before matching, while filled circles (blue) indicate these differences after matching. The vertical dashed lines demarcate standardized mean differences of -0.1, 0, and 0.1, providing reference points for assessing the quality of covariate balance. The covariates include forest coverage (forest), industrial structure (1st Ind, 2nd Ind), and economic indicators (revenue, per_GDP, GDP). Plot (c) provides an example of pinpointing individual project and matched pixels at 1km $\times$ 1km resolution (30m $\times$ 30m resolution is exhibited in Fig. 1c) where the blue pixels denote the project sites and the purple pixels denote the reference area. Plot (d) compares the standardized mean difference of covariates between treated and control pixels in buffer zones before and after matching. The covariates include land use within and beyond the project sties (LUCC, LUCC Buffer) and landform, where the proportion of each category is considered, as well as NDVI, precipitation, and solar radiation (SRAD).

Supporting information changes:

Table S9. Project-level sample sizes.

Project ID	Project size	5-km buffer size	Control pool
1	67,404	5,745,913	7,714,004
2	363,900	2,119,508	6,196,655
3	61,054	1,921,103	9,863,194
4	439,734	2,506,444	6,451,283
5	5,864	179,469	5,964,796
6	64,603	3,327,156	8,154,565
7	151,439	2,135,344	6,102,731
8	58,780	3,430,284	6,451,283
9	25,821	1,131,532	4,038,793
10	78,306	869,315	3,573,048
11	42,809	1,910,698	4,622,267
12	109,106	1,064,928	2,366,509
13	128,864	5,612,265	14,439,939
14	8,576	772,716	5,728,102
15	34,145	1,368,873	4,038,793
16	9,981	1,162,629	7,261,515

17	32,638	2,255,939	2,777,441
18	24,810	1,410,789	6,542,067
19	67,338	1,789,599	2,366,509
20	105,422	942,786	5,570,416
21	36,960	2,290,075	17,109,096
22	8,859	460,450	6,196,655
23	30,344	2,598,543	7,503,247
24	91,021	5,202,382	10,315,414
25	17,453	477,984	3,602,251
26	4,021	179,828	5,213,675
27	4,166	329,955	7,542,131
28	8,500	1,744,232	1,956,386
29	9,817	1,038,165	30,784,970
30	68,144	6,925,761	20,940,090
31	46,142	1,674,901	21,599,203
32	1,676	134,655	4,745,923
33	62,681	3,052,663	9,838,569
34	6,545	617,846	4,608,943
35	20,373	1,089,743	1,957,359
36	95,436	654,070	76,168,318
<i>Total</i>	2,392,732	70,128,543	350,306,140

-3. Thirdly I think the analysis of ownership is meaningless and the author himself mentions the national characteristics of China. I suggest a study to analyze the impact of facility size on spillover effects.

We sincerely appreciate your insightful suggestion regarding the analysis of facility size on spillover effects. We agree that this is a more meaningful heterogeneity analysis given the national characteristics of China. In light of your feedback, we have made the following adjustments to our analysis.

First, we have removed the previously included analysis on ownership, as you correctly pointed out its limited significance.

Second, we have conducted an analysis of how project size influences outcomes following your recommendation. We divided the 36 projects into three equal groups based on project area (small, moderate, and large size) and performed baseline

regressions for each group. Consequently, we found no significant differences in the estimated results across these three groups of varying project sizes, though the large projects exhibit slightly larger estimated effect, as seen in Table R2.1.

Table R2.1. Heterogeneity test of the project size.

ln_NDVI	coef_did	standard_error	t-value	p> t	95% conf. interval	
small	0.0084	0.0003	24.3956	0.0000	0.0077	0.0090
moderate	0.0079	0.0005	15.0025	0.0000	0.0068	0.0089
large	0.0112	0.0005	20.6950	0.0000	0.0101	0.0122

Third, we have incorporated a comprehensive analysis of how buffer zone size affects spillover effects. This new section on spatial heterogeneity examines spillover effects across varying buffer distances from 1 km to 5 km for all 36 projects in our sample (see Fig. 5a in the revised manuscript).

This analysis yielded two key findings. First, regarding the spatial gradient of spillover effects, the density distributions reveal a notable pattern - as the buffer distance increases, the distribution of estimators becomes more concentrated around zero. This trend suggests a gradual decline in spillover effects as we move further from the project sites. Second, the average effects decline as the distance increases. We observed a substantial drop in average effects between the 1 km and 2 km buffers, from 0.85% to 0.15%, followed by a gradual reduction to approximately zero at greater distances. Notably, at the 5 km boundary, the effect is nearly negligible. These observations indicate a spatial gradient of forest project impacts on vegetation quality beyond project boundaries, with the primary impact occurring within the 1 km buffer zone.

Given the significance of these findings and space constraints, we have incorporated the ecosystem type analysis into the main manuscript. The analysis of project size and climate region (suggested by other reviewers) are included in the Supporting Information. We believe these additional analyses provide a more nuanced and meaningful understanding of how project characteristics, particularly size and

spatial extent, influence spillover effects. We are grateful for your suggestion, which has significantly enhanced the depth and relevance of our study.

Manuscript changes:

Page 8 paragraph 1 (line 218):

The spatial distribution of estimated spillovers for the 36 sample projects is presented in Fig. 4a, facilitating a comparison of spillover effects across different spatial scales. The density patterns reveal that as buffer distance increases, the distribution of estimators becomes more concentrated around zero. The average effects diminish, with a notable drop between the 1 km and 2 km buffers from 0.85% to 0.15%, followed by a gradual reduction to approximately zero at greater distances. This observed trend suggests a spatial gradient of forest project impacts on vegetation quality beyond project boundaries, with the primary impact occurring within the 1 km buffer zone. The pattern also indicates that the 5-km buffer zone serves as an appropriate analytical boundary for capturing spillover effects.

Page 9 paragraph 2 (line 258):

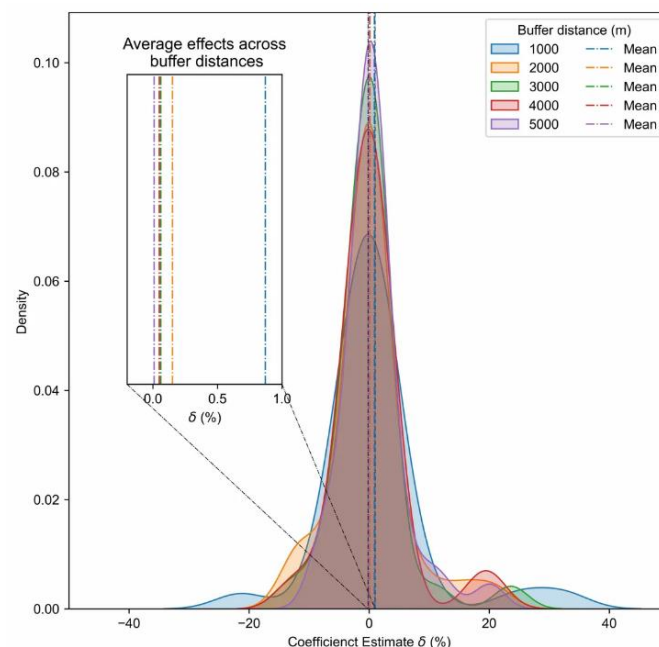


Fig.4. Heterogeneity test. The left plot illustrates the spatial pattern of estimated spillovers in the buffer zones. The main plot depicts the density distribution of DID estimators for buffer zones at distances of 1 km, 1-2 km, 2-3 km, 3-4 km, and 4-5 km

from the project boundaries. The inset in the upper left corner displays the average effects across these buffer distances, with each bar representing the mean estimator value for a specific buffer interval.

Supporting information changes:

Table S8. Additional heterogeneity test

	(1)	(2)	(3)	(4)	(5)
	<i>Subtropical</i>	<i>Temperate</i>	<i>Small</i>	<i>Moderate</i>	<i>Large</i>
<i>DID</i>	0.005*** (15.28)	0.020*** (34.83)	0.008*** (24.40)	0.008*** (15.00)	0.011*** (20.70)
<i>post</i>	0.080*** (143.04)	0.212*** (212.44)	0.082*** (135.92)	0.093*** (100.88)	0.166*** (174.06)
<i>treat</i>	0.001*** (4.53)	0.001*** (3.11)	-0.021*** (-86.57)	0.005*** (13.17)	0.009*** (23.41)
<i>Cons.</i>	18.12*** (32,394)	18.06*** (22,186)	0.468*** (530.5)	18.17*** (23,092)	18.05*** (22,527)
<i>Obs.</i>	12,620,372	4,291,444	5,656,596	5,645,265	5,616,703
<i>R</i> ²	0.267	0.796	0.208	0.320	0.731

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics in parentheses. The results of 9 projects in temperate regions and 27 in subtropical regions are presented in columns 1 and 2, respectively. The 36 projects are divided into three equal groups based on project area, namely the small, moderate, and large in columns 3-5.

-4. Last, the project area is so small and shows a fragmented distribution, but the buffer size is thirty times larger than the project area. Is the change in NDVI in the buffer zone all spillover?

We sincerely thank the reviewers for raising this important issue regarding the attribution of NDVI changes. Your comment has prompted us to provide a more comprehensive explanation of our methodology and findings.

Our attribution process can be summarized as follows. First, the choice of analytical boundaries. Our spatial heterogeneity analysis of the 5 km buffer zone reveals a clear gradient in the spillover effects. We observe that the effect diminishes

significantly with increasing distance from the project sites. The most substantial impacts occur within the first 1 km, while effects in the 4-5 km range approach zero (as mentioned in the response to comment 3). This pattern supports our choice of a 5 km buffer zone as an appropriate analytical boundary for capturing the full extent of spillover effects.

Second, our methodological approach employs matching techniques and difference-in-differences models to estimate NDVI changes. This approach constructs counterfactual scenarios for both project sites and buffer zones, adhering to the principle of estimating what would have occurred in the absence of policy intervention. As illustrated in Fig. 2a and Fig. 3a, our constructed reference areas (control groups) show nearly identical pre-intervention trends with the treatment groups, diverging only after policy implementation.

We conducted difference-in-differences analysis for causal estimation, which effectively eliminates the influence of confounding factors and common trends affecting NDVI changes in both treatment and control groups, such as climate change. This ability to control for common trends is a key advantage of the DID model. Our analysis was further supplemented by a series of robustness checks to ensure reliable causal inference and to attribute NDVI changes to project implementation with greater confidence.

However, we acknowledge your valid concern about potential confounding factors, such as other policies affecting vegetation changes in the region. To address this, we have undertaken the following additional analyses.

First, Placebo tests. We performed two types of placebo tests using a bootstrapping procedure, randomly aligning treatment pixels and intervention periods. These tests simulated scenarios without forest offset projects or with interventions in different years: (a) created pseudo-treatment groups by randomly selecting samples from the control group, maintaining the original intervention period; (b) randomly selected both treatment groups and intervention periods between 2000 and 2022. Both tests were repeated 1000 times. The results (Fig. S6) demonstrate that our actual DID estimates

lie at the extreme right tail of both the empirical and theoretical distributions, indicating that our observed effects are statistically significant and unlikely to result from random factors.

Second, robustness checks for other forestry programs. To examine whether our analysis results were influenced by other forest restoration projects or local environmental and ecological initiatives, we conducted robustness checks using representative projects. (a) We considered the National Forest City program, initiated in 2004, which has awarded this title to 219 cities to date. We categorized our study projects into two groups: those located in counties designated as National Forest Cities and those in non-designated counties. (b) We considered the National Forest Parks, established continuously since 2006, with China now hosting nearly 900 such parks. We divided our study projects into two groups: those in counties with National Forest Parks and those without. (c) We considered the National Reserve Forest Project, which designates areas as key construction zones and moderate construction zones based on natural conditions and other factors. According to the list of counties within the project's scope, we categorized our study projects into three groups: those in counties designated as key construction areas, those in moderate implementation areas, and those beyond the project's construction scope. The results show that across all three policy-based groupings, the estimated effects remained significantly positive. This consistency indicates that our primary findings are robust to the presence of these forest restoration and environmental programs, mitigating concerns about potential confounding influences from these related initiatives to some extent.

These additional analyses strengthen our confidence in attributing the observed NDVI changes primarily to spillover effects from the forest offset projects, rather than to other concurrent programs or random factors. We appreciate your critical comment, which has led to a more robust and comprehensive analysis.

Manuscript changes:

Page 7 paragraph 1 (line 194):

Additionally, the results are robust to the presence of other forest restoration and environmental programs.

Furthermore, we performed two types of placebo tests by randomly aligning the treatment pixels and intervention period using a bootstrapping procedure (Fig. S6). We first created pseudo-treatment groups by randomly selecting samples from the control group, maintaining the original intervention period. Then, we randomly selected both treatment groups and intervention periods between 2000 and 2022. Both tests were repeated 1000 times. These placebo tests simulated scenarios where either no forestry carbon sink projects occurred in the project areas, or other policy interventions happened in different years. The results show that our actual DID estimates lie at the extreme right tail of both the empirical distributions of the placebo estimates and the associated theoretical normal distribution. This implies a positive effect that is statistically different from zero, signifying that our observed treatment effects are unlikely to be the results of random factors and instead represent genuine policy intervention effects.

Supporting information changes:

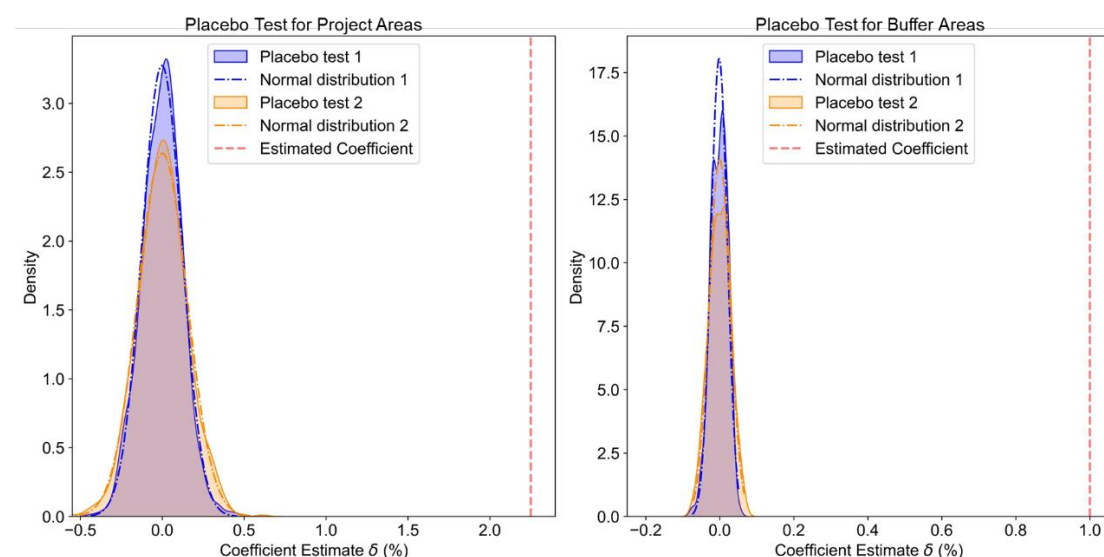


Fig.S6. Placebo test. The left plot is the placebo test for project sites and the right plot is for the buffer zones. The solid density curves illustrate the empirical distribution of the placebo estimates of the coefficient δ , the DID estimator. The superimposed dashed

lines represent the theoretical normal distribution based on the central limit theorem, serving as a reference to assess how our empirical findings deviate from the expected distribution under the null hypothesis of no effect. Placebo test 1 (purple) displays the distribution of estimators for pseudo-treatment groups, while Placebo test 2 (orange) shows the distribution of estimators resulting from randomly assigning both treatment groups and policy intervention years. The dotted vertical lines indicate our actual DID estimators obtained from the baseline regression.

To examine whether our analysis results were influenced by other forest restoration projects or local environmental and ecological initiatives, we conducted robustness checks using representative projects. First, we considered the “National Forest City” program², initiated in 2004, which has awarded this title to 219 cities to date. We categorized our study projects into two groups: those located in counties designated as National Forest Cities and those in non-designated counties. The corresponding outcomes of baseline regression are presented in columns 12 and 13. Second, we considered the National Forest Parks³, established continuously since 2006, with China now hosting nearly 900 such parks. We divided our study projects into two groups: those in counties with National Forest Parks and those without. The corresponding outcomes of baseline regression are shown in columns 14 and 15. Lastly, we considered the National Reserve Forest Project⁴, which designates areas as key construction zones and moderate construction zones based on natural conditions and other factors. According to the list of counties within the project’s scope, we categorized our study projects into three groups: those in counties designated as key construction areas, those in moderate implementation areas, and those beyond the project’s construction scope. The results are presented in columns 16-18. Notably, across all three policy-based groupings, the estimated effects remained significantly positive. This consistency indicates that our primary findings are robust to the presence of these forest restoration and environmental programs, mitigating concerns about potential confounding influences from these related initiatives to some extent.

Table S6. Robustness check.

...

	(12) <i>Forest_City=</i> <i>ty=1</i>	(13) <i>Forest_City=</i> <i>0</i>	(14) <i>National_P</i> <i>ark=1</i>	(15) <i>National_P</i> <i>ark=0</i>	(16) <i>Reserve_F</i> <i>orest=2</i>	(17) <i>Reserve_F</i> <i>orest=1</i>	(18) <i>Reserve_F</i> <i>orest=0</i>
<i>DID</i>	0.0099*** (44.12)	0.0054*** (11.83)	0.0035*** (9.11)	0.0138*** (25.50)	0.0073*** (10.49)	0.0099*** (44.12)	0.0054*** (11.83)
<i>Obs.</i>	6,599,342	9,833,755	9,318,591	4,773,692	2,340,814	6,599,342	9,833,755
<i>R²</i>	0.20	0.60	0.28	0.78	0.23	0.20	0.60

Thank you once again for your time, effort, and valuable insights. We are deeply committed to incorporating your feedback and are grateful for the opportunity to revise and strengthen our work based on your recommendations. Your dedication to the peer review process and to advancing scientific knowledge is truly commendable. Your contribution to the improvement of our manuscript cannot be overstated, and we are sincerely appreciative of your role in enhancing the quality and impact of our research.

Response to Reviewer #3

This paper estimates the spillover effects of forest carbon offset programs in China. The authors use a Difference-in-Difference estimation strategy with a matched sample to estimate the causal effects relative to a counterfactual situation without the introduction of the offset programs. Overall, the paper is novel, is written very clearly, and the empirical analysis is conducted carefully. The relevance of the topic and contribution of this work are motivated well in relation to the international literature on carbon offsets, and in the Chinese context in particular. The mechanisms for the potential spillover effects are also explained thoroughly.

We are pleased to read that the reviewer recognizes our contribution to both scientific literature and ongoing policy discussions. We are grateful for your valuable comments, which helps us further strengthen our analysis. Following your suggestions, we have thoroughly revised our manuscript. We have also provided additional information to improve the clarity of our study.

Below we have addressed all your comments. Your original comments are provided in **bold**, followed by our responses. Excerpts from the revised manuscript are provided in *italics*.

- 1. It would aid the interpretation of the results to have some additional context about (i) the implementation of the projects, for example how the sites are selected, what activities take place, how the boundaries are determined etc.; and (ii) the condition and use of the forested land in the reference area, and in the project and buffer areas prior to the CCER market. It looks like NDVI was increasing on average in all of project, buffer and reference areas since around 2000 independently of the CCER program, which is relevant to understanding the program impacts.

We sincerely appreciate your insightful comments regarding the need for additional

context about the implementation of the projects, and we are grateful for the opportunity to provide more clarity on these points.

We've incorporated additional context into our manuscript and supporting information to provide a more comprehensive understanding of CCER program implementation and its impacts in China.

Regarding the pre-CCER condition of the project, buffer, and reference areas, we acknowledge your observation about the increasing NDVI trends since around 2000. This trend is indeed relevant to understanding the program's impacts and aligns with broader reforestation efforts in China during this period. To address this, our difference-in-differences approach aims to isolate the CCER program's effects from these broader trends. Additionally, we've conducted robustness checks with shorter time frames (2007-2017 and 2009-2015) to mitigate potential confounding effects from long-term trends. Also, alternative model specifications and different vegetation indices are used, and grouped tests have been conducted to examine influences of other forest restoration projects or environmental initiatives.

We appreciate your valuable feedback, which has significantly enhanced our study's depth and clarity.

Manuscript changes:

Page 2 paragraph 4 (line 67):

As the world's largest emitter, China faces substantial pressure for carbon offsets to achieve carbon neutrality. Its abundant forestry carbon sequestration resources are expected to play a pivotal role³³⁻³⁶. In 2012, China launched its voluntary emission reduction market, the Chinese Certified Emissions Reduction (CCER) program, inviting project applications upon announcement. Project developers were required to submit detailed PDDs and materials specifying eligible project boundaries, planned or ongoing activities, and demonstrating project additionality. By 2017, when authorities suspended new CCER project applications, the official website had listed 100 forestry projects. In 2023, the CCER program was re-launched with updated measures,

guidelines, and methodologies. Although the program was temporarily halted, our results suggest that forest managers continued to maintain and improve their forests, anticipating future offset credits.

Page 11 paragraph 2 (line 318):

Second, the scattered and limited-scale nature of the offset projects makes it challenging for any single project to exert a substantial influence on the supply and demand dynamics of global timber or crop markets^{18,59}. Forest offset projects in China are typically characterized by small, scattered, and disconnected polygons, mostly due to the land use features and strict site selection criteria. These include requirements for non-forested land since 2005 and have certified ownership rights, with the latter complicated by China's complex forest property system (see SI for details). These factors make it difficult to yield large fluctuations in the local timber market, as relatively small polygons are being regulated. Thus, leakage through shifts in market demand is less likely to occur.

Page 11 paragraph 4 (line 341):

Moreover, although not the primary focus of our study, our findings of forestry gains observed from 2012 both within and beyond the project sites lend support to the 2023 revision of the administrative measures for the CCER market. This revision shifted the project eligibility criterion, changing the initiation date for eligible project activities from 2005 to 2012, aligning it with the CCER program's implementation date⁶⁰. This adjustment reflects a more targeted approach to capturing the effects of the CCER initiative.

Supporting information changes:

1. Supplementing background

Chinese Certified Emission Reduction (CCER) market is a carbon credits program governed by China's government. It is open to project owners who are legally registered persons or organizations in China, and is linked to China's national carbon market through the offset mechanism. The CCER market was initiated in 2012 when the

National Development and Reform Commission issued the Interim Measures and key guidelines, followed by the launch of the National Trading Platform in October 2013. It is suspended in 2017, when the authorities had announced a total of 100 forestry projects on the official website. In 2023, it is re-launched with the Ministry of Ecology and Environment issuing new measures, guidelines, and methodologies.

Project implementation follows a process similar to many carbon credit systems, including project development and registration, validation and verification, and carbon credit issuance. Project developers must submit detailed Project Design Documents (PDDs) specifying project boundaries, planned activities, methodological compliance, baseline scenarios, and project additionality.

Site selection and boundary determination are governed by specific criteria outlined in the 2013 afforestation methodology (revised in 2023). Key conditions include: land must have been non-forested since February 16, 2005, with specific vegetation thresholds (must not simultaneously meet continuous area ≥ 0.0667 hectares, canopy cover $\geq 20\%$, tree height ≥ 2 meters after forest formation); clear land ownership rights, certified by county-level or higher government, are required; land must not be classified as wetland or organic soil; project activities must comply with soil and water conservation requirements; and the pre-existing agricultural activities should not shift from the intended project sites to other sites. It's important to note that forest property rights remain complex in China, impacting project implementation. Despite the Forest Tenure Reform initiated in the early 2000s, the confirmation and registration of forestland ownership continue to present significant challenges (Ho, 2001; Qin et al., 2011).

The 2023 revision of the interim measures for the CCER program introduced significant changes in project eligibility criteria. Notably, it stipulates that project activities cannot predate 2012. This revision also shifts the eligible project start date in the afforestation methodology from 2005 (as specified in the previous version) to 2012, ensuring consistency with the CCER program's launch. It's important to note that while project activities could predate the generation of the PDDs, the crediting period is

defined as the timeframe when the project scenario generates additional greenhouse gas emission reductions compared to the baseline scenario.

During the validation and verification period, when the project officially begins implementation, all activities within the project boundaries must be meticulously recorded and reported. This comprehensive documentation covers a wide range of aspects: afforestation activities, including planting details, tree species used, and land preparation methods; management activities such as patrols, replanting efforts, harvesting operations, pest control measures, and forest fire prevention strategies; and any occurrences of natural disasters or human disturbances.

- 2. The magnitude of the estimated effects of the projects in project areas and buffer areas appear to be extremely small, especially when compared with the changes over time that were occurring prior to the program and in the reference areas throughout the period.

(i) The size of the effects should be put more explicitly in context so the reader can understand how large they are in practical terms.

We sincerely thank the reviewer for pointing out this important issue regarding the magnitude of our estimated effects.

After adopting a more refined spatial resolution of 30m x 30m NDVI data, we more accurately captured the characteristics of the project sites. This resulted in estimates nearly 5 times larger than our previous results (as detailed in the newly revised sections of our manuscript). We now find that overall project sites show a positive forest gain of 2.25%, along with positive spillover to neighboring areas of 1.0% using NDVI. This increase is due to project sites typically being composed of scattered, smaller polygons (as mentioned in our response to comment 1, with relevant background now added to the manuscript). The previous 1km x 1km data captured many non-forested areas around project boundaries, leading to underestimation of results.

We have made efforts to provide a more detailed explanations for the practical significance in the following three aspects.

First, we explored the actual implications of NDVI. This index is measured by the difference between the near-infrared (NIR) and the visible red light reflectance (RED) from satellite imagery. It can reflect the background influence of plant canopy, such as soil, wet ground, dead leaves, coarse ultra, etc., which is related to vegetation coverage. Negative values indicate ground cover like clouds, water, snow, etc., which are highly reflective of visible light; 0 indicates rocks or bare soil, where NIR and RED are approximately equal; positive values indicate vegetation cover, increasing with cover density. NDVI can judge the state of plant growth. Plant chlorophyll absorbs red light due to photosynthesis, so greater plant growth leads to more red light absorption and NIR light reflection. Therefore, NDVI can reflect the amount of plant biomass. Specifically, larger NDVI indicates better plant growth. Studies have shown that NDVI can capture overall trends in aboveground standing biomass (MacDougall et al., 2024). It is widely used for expressing forest status including density and health (Ghafoor et al., 2022), forest greening (Trujillo et al., 2012), forest growth (Jiang et al., 2019), and forest productivity (Maes et al., 2023).

However, NDVI growth is not linear with vegetation changes when environmental conditions and biomass reach high levels, leading to a saturation effect. For example, Van Der Meer et al. (2001) found NDVI values become saturated when broadleaf forest biomass values are large (> 100 tons/hectare). We attempted to retest using another vegetation index, kNDVI (Camps-Valls et al., 2021), which corrects for saturation effects. Our results show a 4.25% increase in project areas and a 1.6% increase in buffer zones. This implies our estimated results represent a greater degree of improvement in vegetation quality and biomass increase in practical terms.

Second, we compared our results with other studies using NDVI. Existing research (Table R3.1) shows China's NDVI annual growth rate ranges from 0.0006 to 0.0027 over different time periods. Our study shows an annual growth rate of 0.0035 in project areas and 0.0029 in buffer zones from 2000-2023, significantly higher than the range of China as a whole. Using a difference-in-differences approach, we found additional average growth of 2.25% and 0.91% in project areas and buffer zones respectively from

2012-2022, after eliminating common NDVI change trends. This comparison also reflects the effects of our projects.

We also compared our results with studies on similar conservation programs’ effects on NDVI. Chen et al. (2017) found that during 2000-2013, the Beijing–Tianjin Wind/Sand Source Control Program and the Grazing Withdrawal Program significantly increased NDVI values in Inner Mongolia by 0.0362 and 0.0148, while Wu et al. (2013) found an annual NDVI increase of 0.0121 in the same program. Zheng et al. (2024), evaluating eight afforestation initiatives in China, found increasing trends in annual average NDVI values, with growth rates ranging from 0.0005 to 0.0019, among which four programs had lower rates than China as a whole (0.0008) during 1982-2020. Our effects are also comparable to these studies.

Table R3.1: Annual growth rates of vegetation indices in China from existing literatures

Studies	Scale	Indices	Period	Annual growth rate
Song et al., 2022	National	NDVI	1999-2015	0.0017
Piao et al., 2015	National	LAI	1982-2009	0.0070
Wei et al., 2022	National	NDVI	2001-2021	0.0021
Zheng et al., 2024	National	NDVI	1982-2020	0.0008
Zhang et al., 2016	National	NDVI	1982-2000	0.0006
			2000-2013	0.0027

Lastly, to support our main results from another perspective, we used forest loss from the Hansen dataset as another indicator to assist in understanding NDVI changes, where forest loss is defined as a stand-replacement disturbance or a change from forest to non-forest state.

We hope this additional context clarifies the practical significance of our findings. Thanks for your valuable inputs again.

Manuscript changes:

Page 18 paragraph 4 (line 553):

NDVI (Normalized Difference Vegetation Index) is the key dependent variable of this study, which is widely used for expressing forest status including density and health⁷¹, forest greening⁷², forest growth⁷³, and forest productivity⁷⁴. It is measured by the difference between the near-infrared (NIR) and the visible red light reflectance (RED) from satellite imagery. Mathematically, NDVI is calculated using the formula:

$$NDVI = (NIR - RED)/(NIR + RED) \quad (6)$$

NDVI values range from -1 to +1. Negative values typically indicate non-vegetative cover such as water, snow, or clouds. Values near zero suggest bare soil or rock, while positive values indicate vegetation cover, with values closer to 1 representing denser and healthier vegetation. As plant chlorophyll absorbs red light for photosynthesis, greater plant growth leads to increased red light absorption and NIR reflection. Consequently, higher NDVI values generally indicate better plant growth and higher biomass⁶⁸. In this study, we utilize two sources of NDVI data. Our primary source is 30-m resolution NDVI data derived from all available Landsat 5/7/8/9 remote sensing imagery. This dataset is processed to remove clouds and shadows, and the annual NDVI is calculated as the maximum value from all valid Landsat observations throughout the year. Additionally, we employ 1-km spatial resolution NDVI data derived from the SPOT/VEGETATION PROBA-V 1 KM PRODUCTS, which provides an annual perspective on NDVI across China from 1998 to 2019. This moderate-resolution data is generated using a maximum value synthesis method, capturing the peak monthly NDVI values for each year, with values ranging from 0 to 1. Notably, NDVI can exhibit a saturation effect in dense forest areas, where the index becomes less sensitive to vegetation changes. This could lead to underestimation of effects and increased uncertainty in our analysis. To ensure a more accurate and robust analysis, we also employed kNDVI, an index proposed by Camps-Valls et al. (2021) that corrects for saturation effects.

Page 11 paragraph 3 (line 326):

Compared to existing literatures our findings also show a considerable magnitude.

Our study reveals an annual NDVI growth rate of 0.0035 in project areas and 0.0029 in buffer zones from 2000-2023, significantly higher than the range of 0.0006 to 0.0027 observed for China as a whole during different periods in previous studies^{60–63}. Using a difference-in-differences approach, we identified additional average NDVI growth of 2.25% in project areas and 0.91% in buffer zones from 2012-2022, after controlling for common trends. These results are not only comparable to but in some instances surpass the effects found in other conservation programs and afforestation initiatives^{63–65}, indicating the significant impact of CCER forest offset projects on vegetation quality.

Page 9 paragraph 4 (line 281):

Furthermore, we examined the trends in forest loss within buffer zones during the period 2001-2023, utilizing data from Global Forest Change dataset⁵⁵. We used the annual percentage of forest loss relative to the 2000 baseline forest cover in buffer zones as our dependent variable (mean: 0.39%). The DID estimation revealed that in the full sample, forest loss decreased by 0.1% (about 25% reduction) following policy implementation. Projects with negative spillovers showed a non-significant 0.04% decrease, while those with positive spillovers exhibited a significant 0.2% (about 25%) reduction in forest loss (Table 1, columns 4-6). These findings indicate an absence of additional forest loss or carbon leakage in the buffer zones, while supporting the positive spillover effects of forest offset projects in reducing forest loss beyond their boundaries.

Table 1. The channels of the forestry spillovers.

		Forest loss		
		Cultivate	Full	Negative
DID estimator		-0.018*** (-17.59)	-0.001** (-2.33)	-0.000 (-0.94)
Constants		0.181*** (156.13)	-0.003*** (-2.76)	-0.002*** (-2.72)
Obs.		1.5e6	1,633	644
R-squared		0.887	0.419	0.678

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics in parentheses.

Page 19 paragraph 3 (line 589):

The forest loss variable is derived from the Global Forest Change 2000-2023 dataset⁵⁵. In this dataset, forest loss is defined as a stand-replacement disturbance or a change from forest to non-forest state. Using 2000 as the baseline year, the data detects loss primarily in the year 2001-2023 at a high spatial resolution of 1 arc-second per pixel (approximately 30 meters per pixel at the equator).

(ii) It seems possible that the results are statistically significant simply because of the large sample size. As the unobserved determinants of the outcomes in pixels within the same project are likely to be very similar, at minimum the results should be shown with standard errors clustered to take account of this correlation. I also suggest estimating the DID model at the project level by aggregating the treatment and control pixels in the buffer and reference areas. This would provide additional confidence that the effects are practically meaningful.

We sincerely appreciate your insightful comments regarding the statistical significance of our results and the potential influence of large sample size.

We acknowledge that our large sample size could potentially lead to statistically significant results even for small effect sizes. To address this concern, we have added a note in Table S1 explaining that the standard errors in our baseline regression are indeed very small (<0.0000) due to the large number of observations. Following your recommendation, we have also revisited our analysis to account for potential correlation within projects. We have now included standard error information for all individual site regressions in Tables S2 and S3, providing a more comprehensive view of the statistical reliability of our results at the project level.

We greatly appreciate your suggestion to estimate the DID model at the project level. As per your recommendation, we aggregated the treatment and control pixels in both project sites and the buffers to conduct a new baseline estimation on this smaller sample. The results of this analysis are presented in Table R3.2 and included in the Supporting Information. Importantly, these project-level results remain significantly positive and show only minor differences from our pixel-based estimates. The standard

errors in this analysis are less than 0.01, which further confirms the robustness of our findings. This project-level analysis provides additional confidence that the effects we observe are not only statistically significant but also practically meaningful.

We are grateful for your constructive feedback, which has allowed us to strengthen the statistical foundation of our study and provide a more nuanced interpretation of our results.

Table R3.2 Baseline regression results of the aggregated project-level analysis

Ln_NDVI	coef_did	standard_error	t-value	p> t	[95% conf. interval]	
project_site	0.0296	0.0071	4.1387	0.0000	0.0155	0.0436
buffer_zone	0.0055	0.0018	3.0400	0.0060	0.0018	0.0093

Manuscript changes:

Page 6 paragraph 3 (line 190):

Third, to address the potential influence of the large sample size on the statistical significance of the results, we aggregated the treatment and control pixels in both project sites and the buffers to conduct baseline estimation on this smaller sample. The project-level results remain significantly positive and show only minor differences from our pixel-based estimates. These findings underscore both the robustness of our conclusions and the importance of data resolution. Additionally, the results are robust to the presence of other forest restoration and environmental programs.

Supporting information changes:

Table S2. Baseline results of individual project sites.

Project ID	Dependent Variable: ln (NDVI)				Dependent Variable: ln (kNDVI)				Obs.
	DID (%)	R ²	T-statistic	Standard error	DID (%)	R ²	T-statistic	Standard error	
1	-1.538	0.129	-51.703	0.000	-2.265	0.120	-44.91	0.001	3,100,584
2	-1.165	0.007	-43.656	0.000	-1.917	0.006	-39.071	0.000	16,733,559
3	-0.652	0.118	-34.644	0.000	-0.961	0.099	-30.892	0.000	2,808,484
4	-0.437	0.091	-53.979	0.000	-0.648	0.081	-48.839	0.000	20,227,764
5	-0.15	0.035	-0.495	0.003	0.436	0.034	0.768	0.006	269,107
6	-0.117	0.102	-7.509	0.000	-0.209	0.088	-8.4	0.000	2,876,334

7	0.051	0.059	4.398	0.000	0.096	0.047	5.017	0.000	6,966,194
8	0.159	0.055	2.24	0.001	0.114	0.052	0.877	0.001	2,703,639
9	0.182	0.161	9.672	0.000	0.289	0.142	9.75	0.000	1,187,766
10	0.365	0.061	19.814	0.000	0.652	0.05	20.585	0.000	3,602,076
11	0.474	0.027	16.374	0.000	0.607	0.018	12.025	0.001	1,969,214
12	0.698	0.098	69.753	0.000	1.278	0.085	80.448	0.000	5,018,876
13	0.76	0.019	21.491	0.000	1.287	0.015	20.075	0.001	5,923,690
14	1.025	0.086	14.632	0.001	1.69	0.071	13.843	0.001	394,496
15	1.168	0.086	47.193	0.000	1.99	0.071	47.591	0.000	1,570,670
16	1.351	0.029	12.057	0.001	2.411	0.021	11.664	0.002	459,084
17	1.547	0.129	66.724	0.000	2.503	0.110	66.129	0.000	1,404,702
18	1.637	0.125	41.08	0.000	2.721	0.112	40.655	0.001	1,110,255
19	1.933	0.053	105.922	0.000	3.452	0.047	112.973	0.000	3,097,548
20	2.121	0.139	188.258	0.000	3.22	0.123	182.622	0.000	4,849,412
21	2.203	0.089	66.626	0.000	3.601	0.073	62.931	0.001	1,758,932
22	2.214	0.087	52.09	0.000	3.549	0.080	51.581	0.001	407,514
23	2.221	0.05	85.413	0.000	3.59	0.041	82.577	0.000	1,395,824
24	2.403	0.078	70.992	0.000	3.966	0.064	71.163	0.001	1,363,575
25	2.64	0.184	113.274	0.000	4.188	0.179	116.332	0.000	802,838
26	3.189	0.173	21.801	0.001	5.126	0.172	20.441	0.003	184,966
27	3.383	0.211	32.335	0.001	5.838	0.209	32.912	0.002	140,346
28	3.421	0.073	25.245	0.001	6.364	0.065	26.082	0.002	390,264
29	4.284	0.137	88.239	0.000	6.618	0.128	83.22	0.001	451,582
30	4.924	0.097	178.549	0.000	8.386	0.080	175.594	0.000	3,134,612
31	5.135	0.182	68.307	0.001	9.696	0.166	69.937	0.001	2,122,163
32	8.438	0.238	43.908	0.002	15.32	0.233	44.307	0.003	77,096
33	8.962	0.111	333.34	0.000	13.906	0.090	297.598	0.000	2,883,302
34	9.153	0.073	73.547	0.001	15.516	0.069	69.456	0.002	301,070
35	9.785	0.224	100.959	0.001	17.903	0.220	102.448	0.002	488,474
36	37.051	0.417	449.436	0.001	72.997	0.414	450.493	0.002	4,390,051

Note: Columns 2 to 5 present the baseline regression results for 36 sample projects, with the NDVI as the dependent variable. Columns 6 to 9 replace the dependent variable with the kNDVI. "Obs." denotes the number of project observations.

Table S3. Baseline results of individual buffer zones.

Buffer ID	Dependent Variable: ln (NDVI)				Dependent Variable: ln (kNDVI)				Obs.
	DID (%)	R ²	T-statistic	Standard error	DID (%)	R ²	T-statistic	Standard error	
1	-2.273	0.077	-506.29	0.000	-3.597	0.069	-455.01	0.000	264,311,939
2	-2.497	0.013	-298.49	0.000	-4.035	0.009	-264.34	0.000	97,484,591
3	-2.866	0.193	-123.05	0.000	-4.751	0.171	-122.59	0.000	2,760,828
4	-4.414	0.218	-94.832	0.000	-8.171	0.204	-99.069	0.001	1,919,074
5	1.827	0.06	23.433	0.001	3.946	0.059	26.827	0.001	8,093,233

6	0.17	0.075	10.117	0.000	0.034	0.06	1.197	0.000	4,782,252
7	0.706	0.028	30.67	0.000	1.069	0.021	27.49	0.000	3,068,840
8	-4.053	0.128	-79.445	0.001	-7.712	0.125	-83.894	0.001	2,628,256
9	-1.475	0.023	-183.72	0.000	-2.352	0.017	-166.45	0.000	52,050,327
10	0.3	0.013	35.422	0.000	0.615	0.01	41.869	0.000	39,988,487
11	0.085	0.011	12.966	0.000	-0.137	0.008	-11.98	0.000	87,891,951
12	-1.096	0.031	-150.10	0.000	-1.637	0.024	-129.10	0.000	48,986,638
13	22.335	0.027	210.522	0.001	40.946	0.029	213.325	0.002	7,477,097
14	-0.004	0.052	-0.584	0.000	-0.009	0.043	-0.713	0.000	35,544,924
15	-0.963	0.008	-116.82	0.000	-1.484	0.006	-101.82	0.000	62,967,859
16	0.606	0.024	58.289	0.000	1.041	0.019	55.702	0.000	53,476,844
17	0.878	0.025	99.967	0.000	1.654	0.019	103.094	0.000	103,770,936
18	0.566	0.09	85.104	0.000	0.738	0.078	64.343	0.000	64,086,835
19	-1.118	0.026	-199.48	0.000	-1.784	0.02	-181.78	0.000	82,321,044
20	2.172	0.028	180.971	0.000	3.671	0.022	170.478	0.000	43,367,687
21	-5.575	0.011	-24.539	0.002	-14.207	0.011	-34.494	0.004	3,262,607
22	0.233	0.032	12.809	0.000	0.458	0.025	13.86	0.000	21,180,308
23	-0.152	0.032	-7.571	0.000	-0.176	0.025	-5.165	0.000	3,734,644
24	1.33	0.018	34.784	0.000	1.833	0.015	30.697	0.001	7,474,185
25	-0.055	0.007	-3.367	0.000	-0.471	0.006	-16.222	0.000	21,987,229
26	1.217	0.105	52.224	0.000	2.181	0.095	52.959	0.000	8,271,775
27	1.053	0.036	49.92	0.000	1.403	0.03	36.892	0.000	15,176,747
28	1.511	0.026	171.561	0.000	2.775	0.022	175.438	0.000	80,234,672
29	1.476	0.124	53.387	0.000	2.636	0.12	57.542	0.000	1,491,868
30	1.144	0.109	89.546	0.000	1.628	0.091	75.082	0.000	9,954,400
31	21.631	0.358	421.268	0.001	40.021	0.346	429.85	0.001	2,407,179
32	-3.024	0.068	-140.88	0.000	-4.83	0.06	-129.55	0.000	6,194,130
33	4.152	0.077	873.534	0.000	6.603	0.055	774.337	0.000	139,011,318
34	1.658	0.006	86.736	0.000	3.238	0.005	92.04	0.000	28,420,486
35	3.332	0.114	289.154	0.000	5.849	0.109	284.562	0.000	50,128,178
36	-2.295	0.082	-38.43	0.001	-4.546	0.081	-38.89	0.001	29,896,197

Note: Columns 2 to 5 present the baseline regression results for the buffer zones of 36 sample projects, with the NDVI as the dependent variable. Columns 6 to 9 replace the dependent variable with the kNDVI. "Obs." denotes the number of project observations.

Table S6. Robustness check.

	(3)	(4)
	Aggregate project	Aggregate buffer
<i>DID</i>	0.0296*** (4.14)	0.0055*** (3.06)
<i>Obs.</i>	1,656	1,656
<i>R</i> ²	0.96	0.96

- 5. Since the majority of projects included in the analysis had public comment start dates in 2015 or 2016, more explanation should be provided about whether, and if so why, 2012 is the appropriate treatment date. It seems relevant to consider using the staggered model as the main specification with the single treatment date as a robustness check. It would strengthen the analysis to use staggered start dates with new DID methods such as the Callaway and Sant’Anna (2021) estimator.

We sincerely appreciate your insightful observation regarding the timing of project impacts. Your comment has prompted us to provide a more detailed explanation of our methodology and findings.

We apologize for not clearly explaining our choice of 2012 as the intervention year. This decision was based on several factors.

(i) Policy signal. 2012 marked the launch of China’s voluntary emission reduction market. The policy’s announcement created anticipation of potential financial support, motivating eligible entities to initiate activities even before formal project approval. This proactive approach is common in China, where work often begins before subsidy applications are submitted.

(ii) The process of project development. The initial phase of project development involves a period of project design, including defining project boundaries, demonstrating additionality, estimating emission reductions, and preparing construction contracts or project initiation documents. Consequently, project activities are likely to be in preparation or implementation stages before formal

submission.

(iii) The economic incentives from emission reduction estimation. The timeframe for estimating emission reductions is not strictly tied to the project application date but rather to when activities diverge from the baseline scenario. This incentivizes project applicants to prepare afforestation or conservation activities in advance to obtain more carbon credits in the future.

(iv) Observed trends. As shown in Figures 2a and 3b, vegetation index growth rates in project sites and buffer zones significantly increased starting in 2013, the year following the policy announcement, indicating substantial improvements in vegetation quality and biomass. Therefore, 2012 is more appropriate as the start year to study the impact of the CCER program.

As highlighted in your first comment, a more detailed background is crucial for understanding our research design and outcomes. In light of this, we have incorporated additional context into both our manuscript and supporting information.

To further address concerns about the choice of policy year, we conducted additional robustness analysis.

(i) Placebo tests. We performed two types of placebo tests using a bootstrapping procedure, randomly aligning treatment pixels and intervention periods. These tests simulated scenarios without forest offset projects or with interventions in different years: (a) created pseudo-treatment groups by randomly selecting samples from the control group, maintaining the original intervention period; (b) randomly selected both treatment groups and intervention periods between 2000 and 2022. Both tests were repeated 1000 times. The results (Fig. S6) demonstrate that our actual DID estimates lie at the extreme right tail of both the empirical and theoretical distributions, indicating that our observed effects are statistically significant and unlikely to result from random factors.

(ii) Staggered DID. Following your recommendation, to account for

discrepancies between the CCER launch year and the project application submission time, we substituted the intervention time with the public comment start date for each project, conducting the DID regression within a staggered framework recently proposed by Callaway and Sant'Anna (2021). The results remain consistent with baseline regression, which further confirmed the robustness of our results.

We have added our explanations into the discussion part and the additional robustness tests into the results part. We appreciate your critical comment, which has led to a more robust and comprehensive analysis.

Manuscript changes:

Page 17 paragraph 3 (line 506):

The post-intervention period is defined as 2012-2022, since the year 2012 marked the launch of China's voluntary emission reduction market, signaling potential financial support and incentivizing early action. Furthermore, project activities often commence before formal submission, and emission reduction estimations consider pre-application activities. Notably, Figures 2a and 3a demonstrate significant vegetation index growth starting in 2013, indicating a rapid response to the policy announcement. These factors collectively justify 2012 as the appropriate start year for analyzing the CCER program's impact.

Page 6 paragraph 2 (line 171):

Fourth, to account for discrepancies between the CCER launch year and the project application submission time, we substituted the intervention time with the public comment start date for each project, conducting the DID regression within a staggered framework³⁹.

Page 7 paragraph 1 (line 194):

Additionally, the results are robust to the presence of other forest restoration and environmental programs.

Furthermore, we performed two types of placebo tests by randomly aligning the

treatment pixels and intervention period using a bootstrapping procedure (Fig. S6). We first created pseudo-treatment groups by randomly selecting samples from the control group, maintaining the original intervention period. Then, we randomly selected both treatment groups and intervention periods between 2000 and 2022. Both tests were repeated 1000 times. These placebo tests simulated scenarios where either no forestry carbon sink projects occurred in the project areas, or other policy interventions happened in different years. The results show that our actual DID estimates lie at the extreme right tail of both the empirical distributions of the placebo estimates and the associated theoretical normal distribution. This implies a positive effect that is statistically different from zero, signifying that our observed treatment effects are unlikely to be the results of random factors and instead represent genuine policy intervention effects.

Page 7 paragraph 3 (line 207):

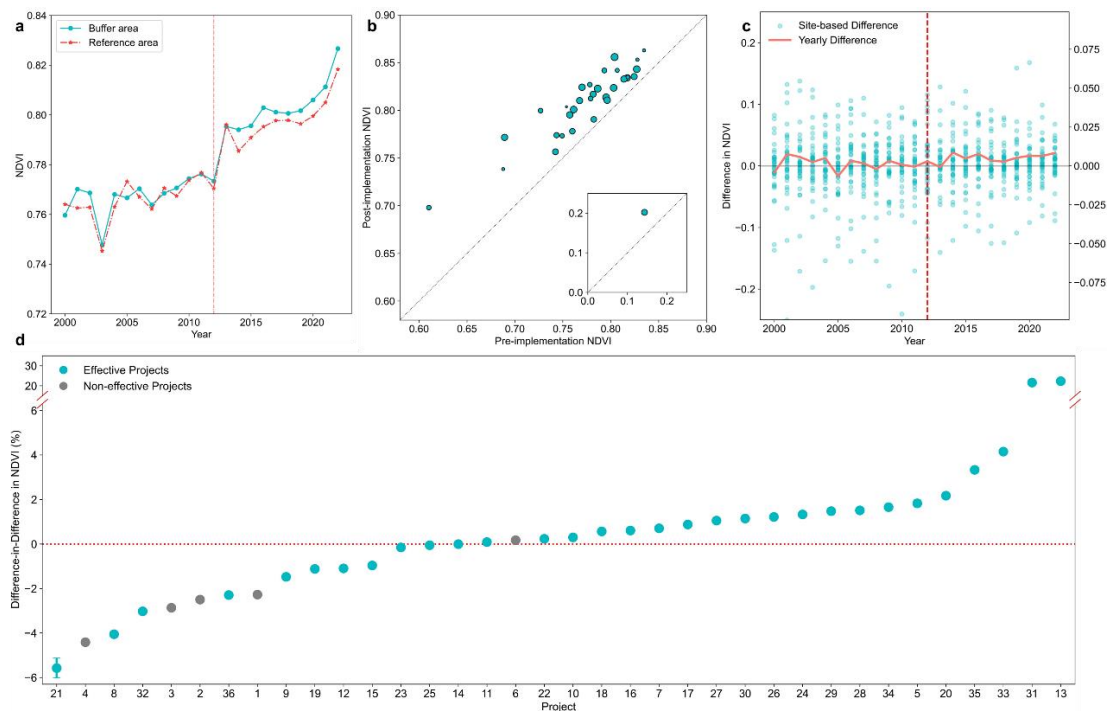


Fig.3. Forestry activities in buffers. Plot a illustrates the annual mean trends of NDVI in the buffer zones of the project and reference areas, with the red vertical line representing the year 2012. Plot b compares NDVI between the pre-implementation and post-implementation periods, with circle sizes indicating project area sizes. Plot c

presents the annual site-based NDVI differences between each of the 36 project sites and their corresponding reference areas, as represented by blue dots for each year. The red line across the plot represents the mean of these differences across all projects. Plot d presents the average treatment effect on the treated for individual projects' buffer zones estimated by difference-in-difference regression, where the whiskers extending above and below each point indicate the 95% confidence interval.

Supporting information changes:

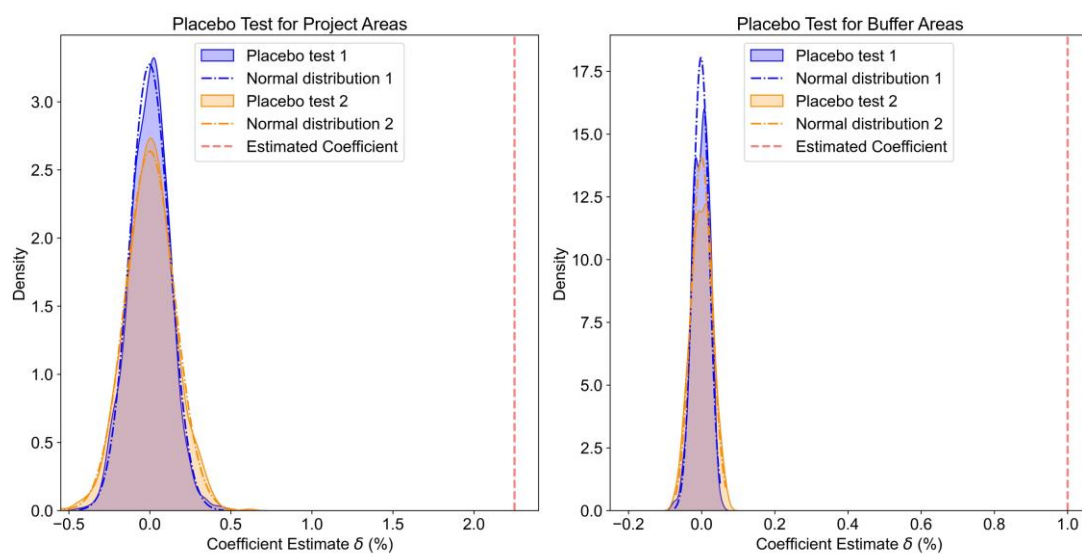


Fig.S6. Placebo test. The left plot is the placebo test for project sites and the right plot is for the buffer zones. The solid density curves illustrate the empirical distribution of the placebo estimates of the coefficient δ , the DID estimator. The superimposed dashed lines represent the theoretical normal distribution based on the central limit theorem, serving as a reference to assess how our empirical findings deviate from the expected distribution under the null hypothesis of no effect. Placebo test 1 (purple) displays the distribution of estimators for pseudo-treatment groups, while Placebo test 2 (orange) shows the distribution of estimators resulting from randomly assigning both treatment groups and policy intervention years. The dotted vertical lines indicate our actual DID estimators obtained from the baseline regression.

Table S6. Robustness check.

...	(11)
...	<i>Staggered did</i>
<i>DID</i>	0.0032***
	(14.39)
<i>Obs.</i>	12,495,452

*Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics are reported in parentheses (except for column 11). All columns include two-way fixed effects.Column 11 replaces the intervention year with the application submission year for each project, following the staggered difference-in-differences framework proposed by Callaway and Sant'Anna (2021). For column 11, ATT is reported with the z-values in the parentheses.*

- 6. The Methods section describes how the pixels within each project area are matched with similar socio-economic and geo-bio-physical pixels. However, it does not explain how the pixels within each buffer are matched with reference pixels. Please add this information.

We sincerely appreciate your astute observation regarding the matching process for pixels within the buffer zones. Your comment has led to a significant improvement in our methodology and overall analysis.

In response to your insightful feedback, we have revised our approach to include a matching process for the buffer zones, similar to that used for the project areas. This change has resulted in updates to all the results presented in the paper. We agree that it is crucial to clearly explain the buffer matching process and present its results.

We have revised the flowchart of methods and added a detailed explanation of the buffer matching process in the Methods section. To summarize, the matching process for buffer zone pixels follows the same procedure and uses the same variables as the project area matching. This ensures consistency in our methodology across all areas of analysis.

To visually represent this change, we have replaced the original Fig. 6d with a new figure showing the matching results for the buffer zone pixels. The matching results for the project areas have been moved to the appendix for completeness.

We are grateful for your suggestion, as it has significantly enhanced the scientific rigor and comprehensiveness of our study. The inclusion of matched reference pixels for buffer zones strengthens our analysis and provides a more robust foundation for our findings on spillover effects.

Manuscript changes:

Page 13 paragraph 1 (line378):

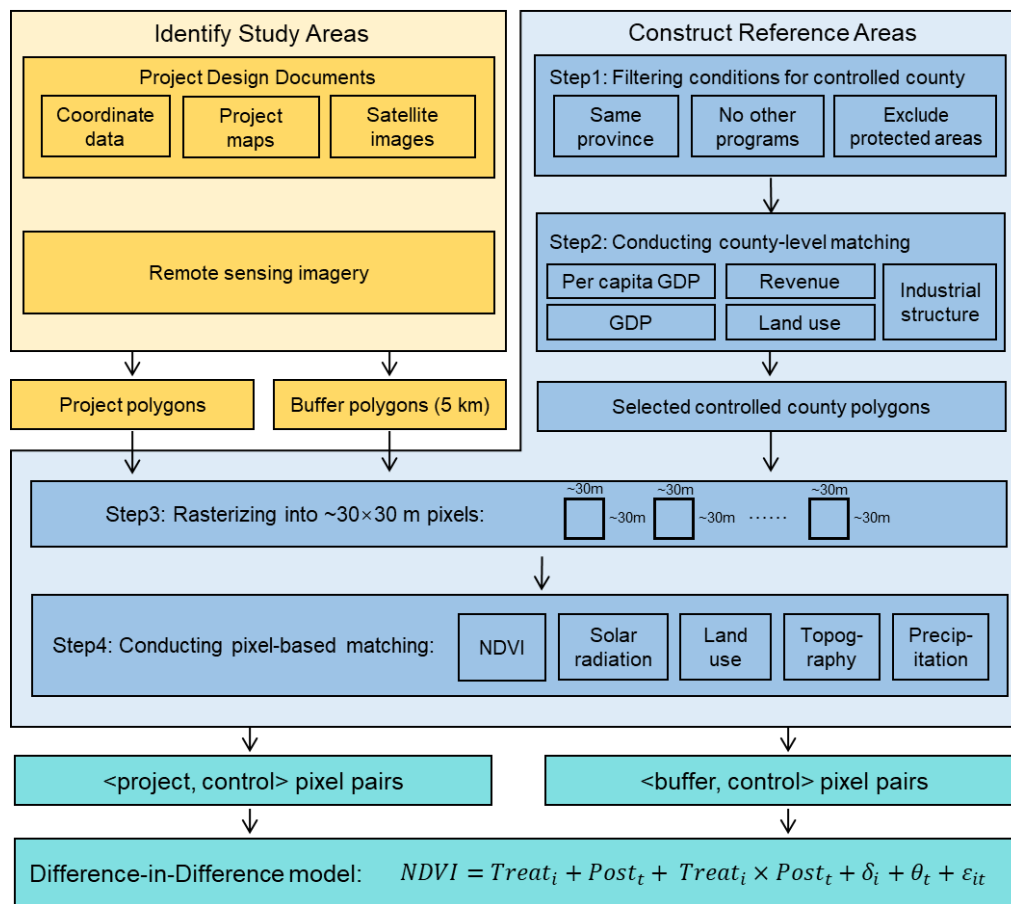


Fig.5. Flowchart of methods. First, the project areas were identified and calibrated using remote sensing imagery. Thus, 5-km buffer zones were created for each project. Both the project and buffer polygons were rasterized into about 30m × 30m pixels. Second, the reference areas were constructed following data exclusion and a two-step statistical matching process: the control groups were restricted to the same provinces without other forest offset projects, and the protected areas were excluded; county matching was conducted using county-level socioeconomic covariates; and pixel matching was conducted using pixel-level geo-biophysical covariates. These steps

constructed references areas for both project sites and buffers, generating the matching pixels. Third, difference-in-differences regressions were run for the entire offset portfolio using pooled data, as well as for each individual project, to estimate the impact of CCER programs on additional forestry activities of forest offset projects.

Page 15 paragraph 2 (line 451):

Thus, we performed 1:1 nearest-neighbor matching with replacement on each pixel plot from the pool of control pixels group, based on NDVI, land use, landform, precipitation, solar radiation, ecosystem, as well as land use type surrounding the project pixels (Fig. 6c, 6d; Table S10). Similarly, a Mahalanobis distance metric was used to measure the similarity between treated and control pixels. An identical matching approach was applied to the buffer zones, with more than 70 million pixels within the 5-km buffer zones. These matched pixels were eventually synthesizing into a reference area. As a result, the matching process has to be carried out individually for each of the 36 projects and their 5-km buffer zones, about 73 million pixels, which makes the process computationally demanding.

To demonstrate the effectiveness of our matching process, we compared the standardized mean difference of covariates before and after matching, a crucial step in ensuring the validity of our subsequent causal analysis. Fig. 6b and 6d demonstrate a substantial reduction in the standardized mean differences across all covariates after matching, as evidenced by the clustering of filled blue circles closer to the zero line compared to the open circles. These confirm the successful alignment of covariates between treated and control areas, thereby strengthening the robustness of our methodological approach in constructing appropriate control groups.

For robustness checks, we expanded the matching process for project sites to include 1:3 and 1:5 nearest neighbors. Given the computational costs and the need for rigorous validation stemming from the abundance of observations in buffer zones, we applied this replacement approach to all 36 project sites. Our findings indicate that this replacement yielded nearly identical estimates.

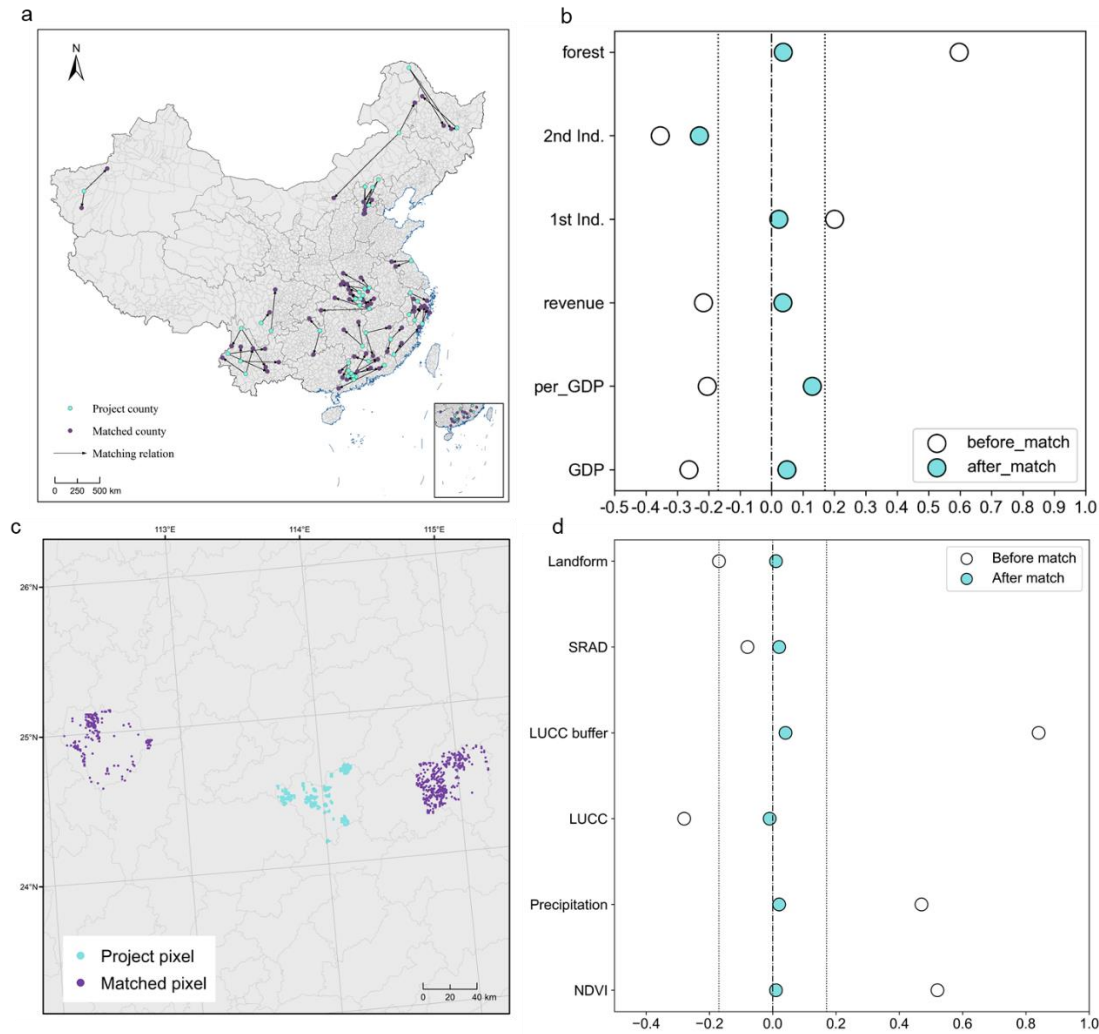


Fig.6. Spatial matching and balance assessment. Plot (a) delineates the distribution of project sites and their matched counties, connected by lines to emphasize the geographic pairing. Plot (b) compares the standardized mean difference of covariates between treated and control counties before and after matching. The open circles represent the standardized mean differences before matching, while filled circles (blue) indicate these differences after matching. The vertical dashed lines demarcate standardized mean differences of -0.1, 0, and 0.1, providing reference points for assessing the quality of covariate balance. The covariates include forest coverage (forest), industrial structure (1st Ind, 2nd Ind), and economic indicators (revenue, per_GDP, GDP). Plot (c) provides an example of pinpointing individual project and matched pixels at 1km x 1km resolution (30m x 30m resolution is exhibited in Fig. 1c) where the blue pixels denote the project sites and the purple pixels denote the reference

area. Plot (d) compares the standardized mean difference of covariates between treated and control pixels in buffer zones before and after matching. The covariates include land use of the pixel and the surrounding pixels (LUCC, LUCC Buffer), landform, where the proportion of each category is considered, as well as NDVI, precipitation, and solar radiation (SRAD).

Supporting information changes:

Table S10. The outcomes of matching.

	Group	Treat Mean	Control Mean	%bias	reduction %bias	T-statistic	P-value
Panel A: County-level matching							
GDP	Unmatched	13.253	13.367	-0.855	177.409	-0.847	0.401
	Matched	13.253	13.166	0.662		0.513	0.609
Per_GDP	Unmatched	9.570	9.692	-1.259	173.022	-1.147	0.257
	Matched	9.570	9.483	0.919		0.641	0.523
revenue	Unmatched	10.014	10.121	-1.056	140.704	-0.615	0.541
	Matched	10.014	9.971	0.430		0.194	0.846
1st Ind.	Unmatched	0.281	0.235	19.712	90.891	2.627	0.011
	Matched	0.281	0.276	1.796		0.232	0.817
2nd Ind.	Unmatched	0.349	0.427	-18.292	62.475	-4.686	0.000
	Matched	0.349	0.375	-6.864		-1.282	0.203
Forest	Unmatched	0.524	0.390	34.522	109.703	3.168	0.003
	Matched	0.524	0.542	-3.350		-0.361	0.719
Panel B: Pixel-level matching in project areas							
Solar radiation	Unmatched	4974.945	5209.49	-4.502	158.819	-1.696	0.090
	Matched	4974.945	4846.597	2.648	158.819	0.625	0.534
Precipitation	Unmatched	1083.239	774.148	39.927	94.842	3.805	0.000
	Matched	1083.239	1061.381	2.059	94.842	0.119	0.905
NDVI	Unmatched	0.7871	0.6174	27.502	97.774	2.995	0.003
	Matched	0.7871	0.7824	0.612	97.774	0.160	0.873
Forest land	Unmatched	0.714	0.381	87.402	93.389	1.682	0.102
	Matched	0.714	0.675	5.778	93.389	0.199	0.844
Grassland	Unmatched	0.103	0.148	-30.405	100	-1.742	0.090
	Matched	0.103	0.103	0.000	100	-0.006	0.995
Residential area	Unmatched	0.024	0.039	-38.462	100	-0.919	0.365
	Matched	0.024	0.024	0.000	100	0.077	0.939
Unused land	Unmatched	0.008	0.021	-61.905	67.692	-2.556	0.015
	Matched	0.008	0.010	-20	67.692	-0.342	0.735
Water area	Unmatched	0.042	0.161	-73.913	110.407	-3.429	0.002
	Matched	0.042	0.039	7.692	110.407	0.093	0.927

Plain	Unmatched	0.056	0.147	-61.905	59.077	-2.139	0.039
	Matched	0.056	0.075	-25.333	59.077	-0.442	0.661
Hill	Unmatched	0.147	0.203	-27.586	213.281	-1.030	0.310
	Matched	0.147	0.112	31.25	213.281	0.656	0.516
Gently undulating mountain	Unmatched	0.226	0.112	101.786	20.618	1.492	0.145
	Matched	0.226	0.125	80.8	20.618	1.322	0.195
Moderately undulating mountain	Unmatched	0.336	0.156	115.385	79.608	2.054	0.048
	Matched	0.336	0.272	23.529	79.608	0.733	0.468

Panel C: Pixel-level matching in buffer zones

Solar radiation	Unmatched	4904.686	5209.49	-5.851	124.352	-1.684	0.092
	Matched	4904.686	4835.786	1.425	124.352	0.558	0.578
Precipitation	Unmatched	1054.009	774.148	36.151	97.014	3.785	0
	Matched	1054.009	1042.754	1.079	97.014	0.059	0.953
NDVI	Unmatched	7818.671	6173.657	26.646	98.547	2.513	0.012
	Matched	7818.671	7788.516	0.387	98.547	-0.064	0.95
Forest land	Unmatched	0.661	0.381	73.491	103.216	1.968	0.057
	Matched	0.661	0.677	-2.363	103.216	-0.116	0.908
Grassland	Unmatched	0.143	0.148	-3.378	298.806	-0.104	0.917
	Matched	0.143	0.134	6.716	298.806	0.188	0.852
Water area	Unmatched	0.014	0.039	-64.103	112	-5.391	0
	Matched	0.014	0.013	7.692	112	0.655	0.517
Residential area	Unmatched	0.014	0.021	-33.333	100	-2.165	0.037
	Matched	0.014	0.014	0	100	-0.013	0.99
Unused land	Unmatched	0.012	0.161	-92.547	109.823	-14.15	0
	Matched	0.012	0.011	9.091	109.823	0.078	0.938
Plain	Unmatched	0.053	0.147	-63.946	123.797	-6.997	0
	Matched	0.053	0.046	15.217	123.797	0.521	0.606
Hill	Unmatched	0.121	0.203	-40.394	137.724	-2.277	0.029
	Matched	0.121	0.105	15.238	137.724	0.458	0.65
Gently undulating mountain	Unmatched	0.212	0.112	89.286	61.722	1.504	0.142
	Matched	0.212	0.158	34.177	61.722	0.812	0.422
Moderately undulating mountain	Unmatched	0.277	0.156	77.564	96.657	1.534	0.134
	Matched	0.277	0.27	2.593	96.657	0.09	0.929

Note: The 'Reduction |%Bias|' column in our statistical table plays a critical role in assessing the effectiveness of our matching methodology. This column quantifies the change in percentage bias before and after matching, offering a clear measure of improvement in the balance between the treatment and control groups.

- 7. On line 202 in the main text, it says that positive spillovers are observed for all groups in the heterogeneity analysis. This does not seem to fit with Fig 5b, which shows negative effects in privately owned projects and zero effect in counties with Medium economic status.

We sincerely appreciate your attentive observation regarding the discrepancy between our statement and the Fig 5b, which has helped us identify and correct this inaccuracy in our manuscript. We apologize for this oversight and the lack of precision in our original statement.

In light of your comment and suggestions from other reviewers, we have undertaken a substantial revision of our heterogeneity analysis section. These revisions include (i) correcting the inaccurate statement on line 202 to accurately reflect the results shown in Fig. 4; (ii) removing the discussion on ownership as per suggestions from other reviewers; and (iii) incorporating a new discussion on ecosystem types and their influence on spillover effects. We also thoroughly reviewing the entire manuscript to ensure consistency between our statements and the presented results.

We believe these changes have significantly improved the accuracy and depth of our heterogeneity analysis, providing a more nuanced and accurate representation of our findings. We are truly grateful for your meticulous review, which has not only helped us correct this specific issue but also prompted a more comprehensive improvement of our manuscript.

Manuscript changes:

Page 8 paragraph 2 (line 227):

We further investigated how project features impact outcomes by categorizing projects based on county economic development level (grouped by per capita GDP), project type (forest management=6, afforestation=25, and vegetation restoration=5), and primary ecosystem type (forest=26, grassland=3, cultivated land=6, desert =1) (Fig. 4b). The positive spillovers are larger in less-developed regions, vegetation restoration projects, and farmland-predominant ecosystems.

First, regarding the economic status, the potential economic benefits and new technologies from the projects are appealing to economically disadvantaged communities^{42,43}, leading to larger positive spillovers in less-developed regions.

Second, we discuss the heterogeneity with respect to project type. The ecosystem services provided by afforestation and vegetation restoration programs have been extensively studied^{44–46}. These projects, including cropland-to-forest conversion, mine rehabilitation, and reduced deforestation, have been shown to influence local climate patterns⁴⁷, mitigate soil erosion and land degradation^{48,49}, and enhance ecosystem resilience. Such effects likely contribute to more substantial spillovers in surrounding areas.

Third, we distinguish the primary ecosystem type within the buffer zones. The substantial spillovers observed in farmland ecosystems align with the mechanism tests on land-use changes (see Table 1), where cultivated land experienced an overall decrease while grassland and forest cover increased within the 5-km buffers. This suggests that the spillovers may be partly attributed to the promotion of farmland-to-forest conversion, potentially driven by farmers' re-evaluation of revenue from forestry activities and offset projects when these are perceived as more profitable^{50,51}.

Conversely, a significant decrease in vegetation quality was found in the buffer zone primarily composed of desert ecosystem. Satellite imagery reveals predominant desert conditions both within and beyond the project sites (see the subplots in Fig. 2b and 3b). While local water resources are scarce, the PDD states that the project site is irrigated by rivers formed from mountain glacier melt and spring water. A plausible explanation is that increased water consumption by the afforestation project may have exacerbated water scarcity in downstream and surrounding areas. This is supported by further examination revealing a notable decrease in water surface area from 3.5% in 2010 to 1.9% in 2020, suggesting intensified drought conditions and consequently decreased vegetation quality in surrounding areas⁵². These findings highlight the need to consider the complex interactions between forestry activities and local ecological contexts when planning and implementing related projects.

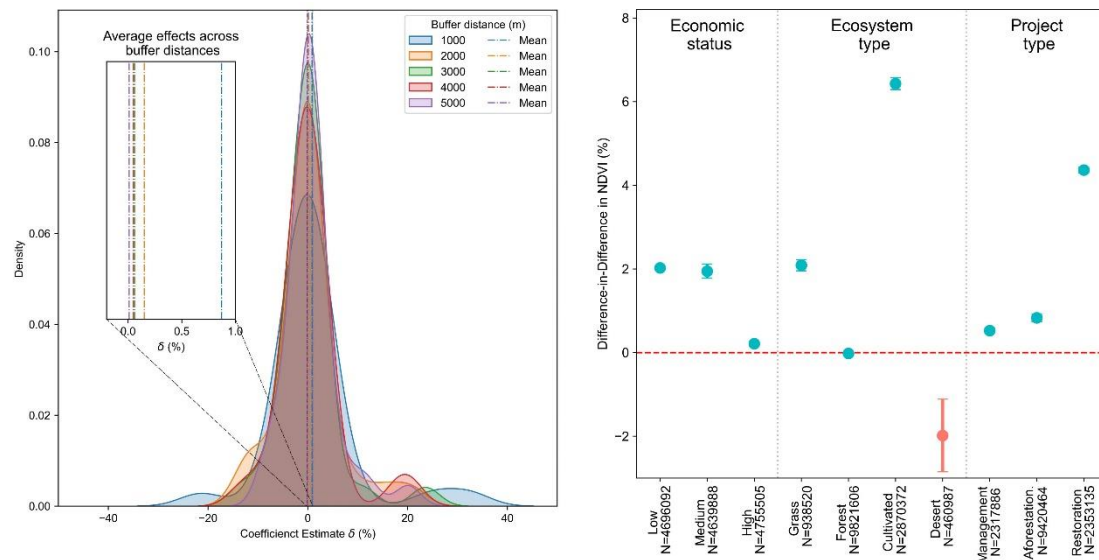


Fig.4. Heterogeneity test. The left plot illustrates the spatial pattern of estimated spillovers in the buffer zones. The main plot depicts the density distribution of DID estimators for buffer zones at distances of 1 km, 1-2 km, 2-3 km, 3-4 km, and 4-5 km from the project boundaries. The inset in the upper left corner displays the average effects across these buffer distances, with each bar representing the mean estimator value for a specific buffer interval. The right plot presents a heterogeneity analysis based on three key features: economic status, primary ecosystem type, and project type; the whiskers extending above and below each point indicate the 95% confidence intervals for the respective estimates.

- 8. Since matching is carried out with replacement, it would be helpful to include information on whether there are counties or pixels that are used as controls for multiple treated observations, and if so, how many times each county/pixel is used.

We thank the reviewer for this insightful comment. The suggestion has indeed enhanced the clarity and transparency of our matching approaches and results.

In response, we have included additional information regarding the frequency of county and pixel usage as controls in our study. Out of approximately 2,840 counties in China, 45 counties host our sample projects. Following our filtering and matching processes, as detailed in the methodology section, 86 counties were selected for the matched control group, with 73 being unique. The repeated matched counties primarily

occurred in Guangdong and Hubei provinces, where several cities were excluded pre-matching due to existing forest offset projects. On average, each matched county appears 1.17 times in our control group.

After converting the matched counties into grid cells, we obtained about 350 million pixels to serve as control pixels. The project sites, comprising approximately 2.39 million points, were matched 1:1, resulting in 0.6 million unique pixel points. Consequently, a control pixel is used an average of 4 times in the control group.

To provide a more comprehensive view, we have included detailed information on the pixel size of each project and its corresponding control area in Table S9. We believe this additional data offers a clearer understanding for both reviewers and readers. We are grateful for the opportunity to clarify and thank you for your valuable suggestion, which has undoubtedly improved the quality of our study.

Manuscript changes:

Page 14 paragraph 3 (line 433):

For those implemented in more than one county, 1:1 matching was performed (Fig. 6a). We evaluated the quality of the matching, adjusted the relevant covariates, and iteratively optimized the control group counties (Fig. 6b). In the first few iterations, we found that some counties had been funded by other forestry carbon sequestration projects recently, indicating that our matching results were well representative. A total of 86 counties were selected into the matched control group after a series of filtration and matching, with 73 being unique. The repeated matched county mainly occurred in projects in Guangdong and Hubei provinces, where a number of cities were excluded before matching for the implementation of forest offset projects. After converting the unique matched counties into grid cells, we obtained about 350 million pixels to serve as control pixels.

.....

Thus, we performed 1:1 nearest-neighbor matching with replacement on each pixel plot from the pool of control pixels group, based on NDVI, land use, landform,

precipitation, solar radiation, ecosystem, as well as land use type surrounding the project pixels (Fig. 6c, 6d; Table S10). Similarly, a Mahalanobis distance metric was used to measure the similarity between treated and control pixels. This process yielding 0.6 million unique control pixels for project sites, indicating that a matched control pixel is used 4 times on average in the control group. An identical matching approach was applied to the buffer zones, with more than 70 million pixels within the 5-km buffer zones. These matched pixels were eventually synthesizing into a reference area. As a result, the matching process has to be carried out individually for each of the 36 projects and their 5-km buffer zones, about 73 million pixels, which makes the process computationally demanding.

Supporting information changes:

Table S9. Project-level sample sizes.

Project ID	Project size	5-km buffer size	Control pool
1	67,404	5,745,913	7,714,004
2	363,900	2,119,508	6,196,655
3	61,054	1,921,103	9,863,194
4	439,734	2,506,444	6,451,283
5	5,864	179,469	5,964,796
6	64,603	3,327,156	8,154,565
7	151,439	2,135,344	6,102,731
8	58,780	3,430,284	6,451,283
9	25,821	1,131,532	4,038,793
10	78,306	869,315	3,573,048
11	42,809	1,910,698	4,622,267
12	109,106	1,064,928	2,366,509
13	128,864	5,612,265	14,439,939
14	8,576	772,716	5,728,102
15	34,145	1,368,873	4,038,793
16	9,981	1,162,629	7,261,515
17	32,638	2,255,939	2,777,441
18	24,810	1,410,789	6,542,067
19	67,338	1,789,599	2,366,509
20	105,422	942,786	5,570,416
21	36,960	2,290,075	17,109,096
22	8,859	460,450	6,196,655
23	30,344	2,598,543	7,503,247
24	91,021	5,202,382	10,315,414

25	17,453	477,984	3,602,251
26	4,021	179,828	5,213,675
27	4,166	329,955	7,542,131
28	8,500	1,744,232	1,956,386
29	9,817	1,038,165	30,784,970
30	68,144	6,925,761	20,940,090
31	46,142	1,674,901	21,599,203
32	1,676	134,655	4,745,923
33	62,681	3,052,663	9,838,569
34	6,545	617,846	4,608,943
35	20,373	1,089,743	1,957,359
36	95,436	654,070	76,168,318
<i>Total</i>	2,392,732	70,128,543	350,306,140

References Cited in Response to Reviewer:

- [1] Ho, P. (2001). Who owns China's land? Policies, property rights and deliberate institutional ambiguity. *The China Quarterly*, 166, 394-421.
- [2] Qin, P., Carlsson, F., & Xu, J. (2011). Forest tenure reform in China: A choice experiment on farmers' property rights preferences. *Land Economics*, 87(3), 473-487.
- [3] MacDougall, A. S., Esch, E., Chen, Q., Carroll, O., Bonner, C., Ohlert, T., ... & Seabloom, E. W. (2024). Widening global variability in grassland biomass since the 1980s. *Nature ecology & evolution*, 1-12.
- [4] Ghafoor, G. Z., Sharif, F., Shahid, M. G., Shahzad, L., Rasheed, R., & Khan, A. U. H. (2022). Assessing the impact of land use land cover change on regulatory ecosystem services of subtropical scrub forest, Soan Valley Pakistan. *Scientific Reports*, 12(1), 10052.
- [5] Trujillo, E., Molotch, N. P., Goulden, M. L., Kelly, A. E., & Bales, R. C. (2012). Elevation-dependent influence of snow accumulation on forest greening. *Nature Geoscience*, 5(10), 705-709.
- [6] Jiang, P., Liu, H., Piao, S., Ciais, P., Wu, X., Yin, Y., & Wang, H. (2019). Enhanced growth after extreme wetness compensates for post-drought carbon loss in dry forests. *Nature Communications*, 10(1), 195.
- [7] Maes, J., Bruzón, A. G., Barredo, J. I., Vallecillo, S., Vogt, P., Rivero, I. M., &

- Santos-Martín, F. (2023). Accounting for forest condition in Europe based on an international statistical standard. *Nature Communications*, 14(1), 3723.
- [8] Van Der Meer, F., Bakker, W., Scholte, K., Skidmore, A., De Jong, S., Clevers, J. G. P. W., ... & Epema, G. (2001). Spatial scale variations in vegetation indices and above-ground biomass estimates: implications for MERIS. *International Journal of Remote Sensing*, 22(17), 3381-3396.
- [9] Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., ... & Running, S. W. (2021). A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, 7(9), eabc7447.
- [10] Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., ... & Wang, Y. (2015). Detection and attribution of vegetation greening trend in China over the last 30 years. *Global change biology*, 21(4), 1601-1609.
- [11] Zhang, Y., Peng, C., Li, W., Tian, L., Zhu, Q., Chen, H., ... & Xiao, X. (2016). Multiple afforestation programs accelerate the greenness in the ‘Three North’ region of China from 1982 to 2013. *Ecological Indicators*, 61, 404-412.
- [12] Chen, H., Shao, L., Zhao, M., Zhang, X., & Zhang, D. (2017). Grassland conservation programs, vegetation rehabilitation and spatial dependency in Inner Mongolia, China. *Land Use Policy*, 64, 429-439.
- [13] Song, W., Feng, Y., & Wang, Z. (2022). Ecological restoration programs dominate vegetation greening in China. *Science of the Total Environment*, 848, 157729.
- [14] Wu, Z., Wu, J., Liu, J., He, B., Lei, T., & Wang, Q. (2013). Increasing terrestrial vegetation activity of ecological restoration program in the Beijing–Tianjin Sand Source Region of China. *Ecological Engineering*, 52, 37-50.
- [15] Wei, Y., Sun, S., Liang, D., & Jia, Z. (2022). Spatial–temporal variations of NDVI and its response to climate in China from 2001 to 2020. *International Journal of Digital Earth*, 15(1), 1463-1484.
- [16] Zheng, L., Lu, J., Liu, H., Chen, X., & Yesou, H. (2024). Evidence of vegetation greening benefitting from the afforestation initiatives in China. *Geo-Spatial Information Science*, 27(3), 683-702.

- [17] Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2), 200-230.
- [18] Song, C. et al. Sustainability of Forests Created by China's Sloping Land Conversion Program: A comparison among three sites in Anhui, Hubei and Shanxi. *For. Policy Econ.* 38, 161–167 (2014).
- [19] Latta, G. S., Adams, D. M., Bell, K. P. & Kline, J. D. Evaluating land-use and private forest management responses to a potential forest carbon offset sales program in western Oregon (USA). *For. Policy Econ.* 65, 1–8 (2016).
- [20] Lu, C., Zhao, T., Shi, X. & Cao, S. Ecological restoration by afforestation may increase groundwater depth and create potentially large ecological and water opportunity costs in arid and semiarid China. *J. Clean. Prod.* 176, 1213–1222 (2018).

In conclusion, we would like to express our sincere and heartfelt gratitude again for your time, expertise, and invaluable feedback. Your thoughtful and detailed comments have been instrumental in significantly enhancing the depth, rigor, and overall quality of our study.

Response to Reviewer #4

The paper investigates the off-site impacts of a forest-based carbon-offset program in China. The authors find positive spillover effects, which they attribute to learning effects. The setting is interesting and the contribution would be important, but I have major concerns that would need to be addressed before I can recommend publication.

We are pleased to read that the reviewer recognizes the quality and contributions of our study. We are grateful for your valuable comments. We have significantly revised our manuscript to strengthen our analysis and streamline the storyline. We have also provided additional information to improve the clarity of our study.

Below we have addressed all your comments. Your original comments are provided in **bold**, followed by our responses. Excerpts from the revised manuscript are provided in *italics*.

- 1. First, I don't see why the authors use the (maximum) NDVI as outcome, when other more precise data sources are available. If it's about forests, why not use the Hansen data or other land use data? The NDVI is a very rough proxy that may well be confounded.

We sincerely appreciate your insightful comment regarding our choice of NDVI as the primary outcome measure. Your feedback has prompted us to conduct additional analyses and provide a more comprehensive justification for our methodological choices.

Our initial selection of NDVI was based on its widespread use in assessing vegetation status. Studies have shown that NDVI effectively captures trends in aboveground standing biomass (MacDougall et al., 2024) and is widely used to express various aspects of forest status, including density, health, greening, growth, and

productivity (Ghafoor et al., 2022; Trujillo et al., 2012; Jiang et al., 2019; Maes et al., 2023). However, we acknowledge the limitations you pointed out, particularly the saturation effect in dense forests.

Based on your valuable feedback, we have taken several steps to address this issue.

First, following your suggestion, we carefully studied and incorporated the Hansen dataset. We found that only the forest loss data (2000-2023) suited our research needs for long-term time series analysis. Other data like forest gain were either cross-sectional or only available up to 2012. The forest loss is defined as a stand-replacement disturbance or a change from forest to non-forest state, which is somewhat strict for our study's focus on improving forest quality rather than forest cover increase. It doesn't capture the concept of greening or biomass increase in existing forested areas.

Therefore, we used it for robustness checks and channel tests rather than as our primary research outcome. Using this forest loss data, we examined spillover effects in buffer zones. Our analysis showed no additional forest loss in the buffer zones and indicated a significant reduction in forest loss trends, alleviating concerns about unintended negative consequences and supporting the positive spillover effects observed.

Second, we attempted to retest using another vegetation index, kNDVI (Camps-Valls et al., 2021), which corrects for saturation effects. Our results show a 4.25% increase in project areas and a 1.6% increase in buffer zones. This implies our estimated results represent a greater degree of improvement in vegetation quality and biomass increase in practical terms. Our analysis confirmed that the main findings remained consistent and even improved compared to the previous results. Detailed results are provided in the Supporting Information.

Lastly, we have adopted higher-resolution NDVI data at a $30\text{ m} \times 30\text{ m}$ scale, which is the highest precision data currently available. This represents a significant improvement over our previous data, enhancing the resolution by a factor of 1100. We reconducted all analyses using this higher-resolution data, successfully overcoming significant computational challenges. The results were consistent with our original

findings. The improved precision allowed for more accurate identification of smaller polygons, further validating the robustness of our results. Additionally, we re-examined land use changes as part of our mechanism tests, providing a more comprehensive understanding of the project impacts.

Manuscript changes:

All figures in the manuscript have been updated to reflect the outcomes using the 30 m × 30 m resolution dataset.

Page 4 paragraph 3 (line 117):

This divergence continued to grow over time, revealing overall positive effects of the CCER program. The baseline results of overall project sites show a positive effect of 2.25% to 4.25% with p-value less than 0.01 (Table S1).

Page 5 paragraph 3 (line 151):

Over 23 years, the average forest gain spillover of the project sites to neighboring areas is 0.91% to 1.60% (NDVI and kNDVI respectively, $p < 0.01$, Table S1). This effect is increasingly positive in the post-treatment period.

Page 6 paragraph 3 (line 178):

First, To address the non-linear relationship and saturation effect of NDVI in areas with high vegetation cover and during peak seasons, we employed kNDVI⁴¹ as an alternative vegetation index. Results remained consistent, with kNDVI matching results (Fig. S3) aligning with trends in Fig. 2a and Fig. 3a. The estimated effects were slightly larger in absolute value (Tables S1, S2, S3), possibly due to NDVI limitations in dense forest areas.

Page 9 paragraph 4 (line 281):

Furthermore, we examined the trends in forest loss within buffer zones during the period 2001-2023, utilizing data from Global Forest Change dataset⁵⁵. We used the annual percentage of forest loss relative to the 2000 baseline forest cover in buffer zones as our dependent variable (mean: 0.39%). The DID estimation revealed that in the full sample, forest loss decreased by 0.1% (about 25% reduction) following policy

implementation. Projects with negative spillovers showed a non-significant 0.04% decrease, while those with positive spillovers exhibited a significant 0.2% (about 25%) reduction in forest loss (Table 1, columns 4-6). These findings indicate an absence of additional forest loss or carbon leakage in the buffer zones, while supporting the positive spillover effects of forest offset projects in reducing forest loss beyond their boundaries.

Table 1. The channels of the forestry spillovers.

	Land use			Forest loss			Wild fire	
	Cultivate	Forest	Grass	Full	Negative	Positive	Frequency	Freq/km ²
DID	-0.018***	0.002	0.024***	-0.001**	-0.000	-0.002**	-0.110*	-0.002**
estimator	(-17.59)	(1.26)	(25.80)	(-2.33)	(-0.94)	(-2.14)	(-1.77)	(-1.84)
Constants	0.181***	0.570***	0.237**	-0.003***	-0.002***	-0.004**	-0.556***	-0.004***
	(156.13)	(419.11)	(221.00)	(-2.76)	(-2.72)	(-2.53)	(-7.23)	(-5.38)
Two-way fixed effect	YES	YES	YES	YES	YES	YES	YES	YES
Obs.	1.5e6	1.5e6	1.5e6	1,633	644	989	1672	1672
R-squared	0.887	0.317	0.155	0.419	0.678	0.381	0.720	0.427

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics in parentheses.

Page 18 paragraph 5 (line 559):

NDVI values range from -1 to +1. Negative values typically indicate non-vegetative cover such as water, snow, or clouds. Values near zero suggest bare soil or rock, while positive values indicate vegetation cover, with values closer to 1 representing denser and healthier vegetation. As plant chlorophyll absorbs red light for photosynthesis, greater plant growth leads to increased red light absorption and NIR reflection. Consequently, higher NDVI values generally indicate better plant growth and higher biomass⁶⁸. In this study, we utilize two sources of NDVI data. Our primary source is 30-m resolution NDVI data derived from all available Landsat 5/7/8/9 remote sensing imagery. This dataset is processed to remove clouds and shadows, and the annual NDVI is calculated as the maximum value from all valid Landsat observations throughout the year. Additionally, we employ 1-km spatial resolution NDVI data derived from the SPOT/VEGETATION PROBA-V 1 KM PRODUCTS, which provides an annual

perspective on NDVI across China from 1998 to 2019. This moderate-resolution data is generated using a maximum value synthesis method, capturing the peak monthly NDVI values for each year, with values ranging from 0 to 1. Notably, NDVI can exhibit a saturation effect in dense forest areas, where the index becomes less sensitive to vegetation changes. This could lead to underestimation of effects and increased uncertainty in our analysis. To ensure a more accurate and robust analysis, we also employed *k*NDVI, an index proposed by Camps-Valls et al. (2021) that corrects for saturation effects.

.....

The forest loss variable is derived from the Global Forest Change 2000-2023 dataset⁵⁵. In this dataset, forest loss is defined as a stand-replacement disturbance or a change from forest to non-forest state. Using 2000 as the baseline year, the data detects loss primarily in the year 2001-2023 at a high spatial resolution of 1 arc-second per pixel (approximately 30 meters per pixel at the equator).

Supporting information changes:

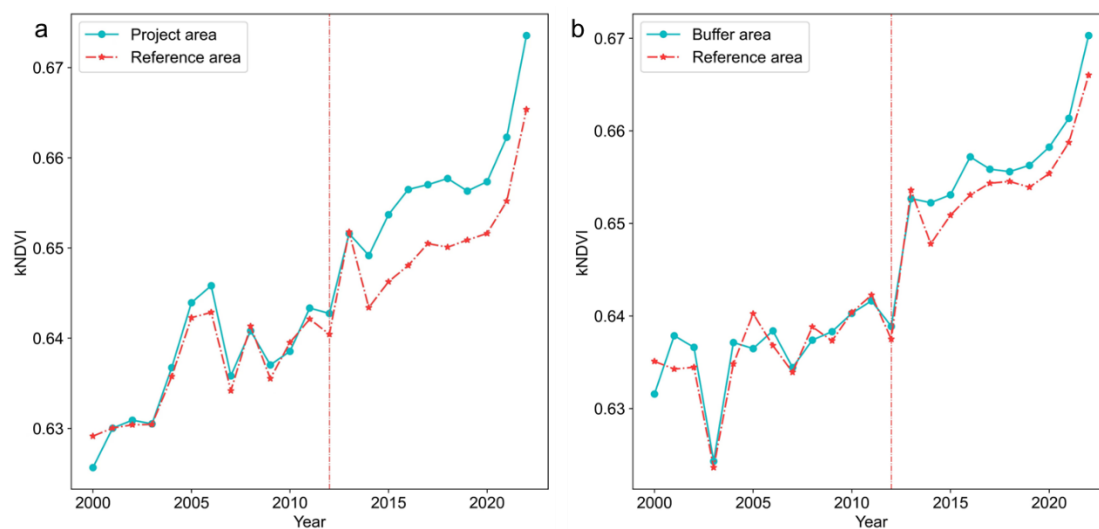


Fig.S3. Trends of *k*NDVI for project areas and buffer zones with the reference areas.

Note: Plot a presents the annual mean trends of *k*NDVI for project areas and the corresponding reference area. Plot b presents the annual mean trends of *k*NDVI for 5-km buffer zones and the corresponding reference area.



Fig.S5. Grid of different spatial resolution overlaid on an example study site.

Note: Plot a demonstrates the coverage of a 1km×1km grid over the project area. The blue region delineates the actual project boundary, while the grey grid cells represent areas that intersect with the project site. The relatively coarse granularity of this grid results in the inclusion of substantial non-project areas within grid cells that overlap the project boundary. Plot b shows the same project area overlaid with a 30m×30m grid. The finer granularity allows for a nearly seamless integration of the grid with the project area, to the extent that the distinction between the grid and the project boundary becomes visually indiscernible, effectively eliminating the blue demarcation visible in plot a.

Table S1. Baseline results of project sites and buffer zones.

	(1)	(2)	(3)	(4)	(5)	(6)
Area	Project			Buffer		
Dependent	NDVI	NDVI	kNDVI	NDVI	NDVI	kNDVI
DID	0.0225*** (113.87)	0.0225*** (300.54)	0.0425*** (305.25)	0.0091*** (21.33)	0.0091*** (32.99)	0.0160*** (32.69)
post	0.0444*** (317.75)	0.1174*** (888.02)	0.1947*** (792.75)	0.0544*** (180.75)	0.1136*** (233.63)	0.1857*** (214.99)
treat	0.0037*** (26.73)	0.0038*** (73.12)	0.0062*** (64.209)	0.00113*** (3.83)	-0.00238*** (-12.48)	0.0084*** (24.94)
Cons.	8.869 (91771)	8.8472*** (16893)	8.402*** (8643)	18.059*** (86655)	18.094*** (34069)	8.457*** (8962)
FE	No	Yes	Yes	No	Yes	Yes

Obs.	106,566,063	106,566,063	106,566,063	16,918,564	16,918,564	16,918,564
R ²	0.003	0.86	0.87	0.005	0.58	0.60

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics in parentheses.

Table S2. Baseline results of individual project sites.

Project ID	Dependent Variable: ln (NDVI)				Dependent Variable: ln (kNDVI)				Obs.
	DID (%)	R ²	T-statistic	Standard error	DID (%)	R ²	T-statistic	Standard error	
1	-1.538	0.129	-51.703	0.000	-2.265	0.120	-44.91	0.001	3,100,584
2	-1.165	0.007	-43.656	0.000	-1.917	0.006	-39.071	0.000	16,733,559
3	-0.652	0.118	-34.644	0.000	-0.961	0.099	-30.892	0.000	2,808,484
4	-0.437	0.091	-53.979	0.000	-0.648	0.081	-48.839	0.000	20,227,764
5	-0.15	0.035	-0.495	0.003	0.436	0.034	0.768	0.006	269,107
6	-0.117	0.102	-7.509	0.000	-0.209	0.088	-8.4	0.000	2,876,334
7	0.051	0.059	4.398	0.000	0.096	0.047	5.017	0.000	6,966,194
8	0.159	0.055	2.24	0.001	0.114	0.052	0.877	0.001	2,703,639
9	0.182	0.161	9.672	0.000	0.289	0.142	9.75	0.000	1,187,766
10	0.365	0.061	19.814	0.000	0.652	0.05	20.585	0.000	3,602,076
11	0.474	0.027	16.374	0.000	0.607	0.018	12.025	0.001	1,969,214
12	0.698	0.098	69.753	0.000	1.278	0.085	80.448	0.000	5,018,876
13	0.76	0.019	21.491	0.000	1.287	0.015	20.075	0.001	5,923,690
14	1.025	0.086	14.632	0.001	1.69	0.071	13.843	0.001	394,496
15	1.168	0.086	47.193	0.000	1.99	0.071	47.591	0.000	1,570,670
16	1.351	0.029	12.057	0.001	2.411	0.021	11.664	0.002	459,084
17	1.547	0.129	66.724	0.000	2.503	0.110	66.129	0.000	1,404,702
18	1.637	0.125	41.08	0.000	2.721	0.112	40.655	0.001	1,110,255
19	1.933	0.053	105.922	0.000	3.452	0.047	112.973	0.000	3,097,548
20	2.121	0.139	188.258	0.000	3.22	0.123	182.622	0.000	4,849,412
21	2.203	0.089	66.626	0.000	3.601	0.073	62.931	0.001	1,758,932
22	2.214	0.087	52.09	0.000	3.549	0.080	51.581	0.001	407,514
23	2.221	0.05	85.413	0.000	3.59	0.041	82.577	0.000	1,395,824
24	2.403	0.078	70.992	0.000	3.966	0.064	71.163	0.001	1,363,575
25	2.64	0.184	113.274	0.000	4.188	0.179	116.332	0.000	802,838
26	3.189	0.173	21.801	0.001	5.126	0.172	20.441	0.003	184,966
27	3.383	0.211	32.335	0.001	5.838	0.209	32.912	0.002	140,346
28	3.421	0.073	25.245	0.001	6.364	0.065	26.082	0.002	390,264
29	4.284	0.137	88.239	0.000	6.618	0.128	83.22	0.001	451,582
30	4.924	0.097	178.549	0.000	8.386	0.080	175.594	0.000	3,134,612
31	5.135	0.182	68.307	0.001	9.696	0.166	69.937	0.001	2,122,163
32	8.438	0.238	43.908	0.002	15.32	0.233	44.307	0.003	77,096
33	8.962	0.111	333.34	0.000	13.906	0.090	297.598	0.000	2,883,302
34	9.153	0.073	73.547	0.001	15.516	0.069	69.456	0.002	301,070
35	9.785	0.224	100.959	0.001	17.903	0.220	102.448	0.002	488,474

36 37.051 0.417 449.436 0.001 72.997 0.414 450.493 0.002 4,390,051

Note: Columns 2 to 5 present the baseline regression results for 36 sample projects, with the NDVI as the dependent variable. Columns 6 to 9 replace the dependent variable with the kNDVI. "Obs." denotes the number of project observations.

Table S3. Baseline results of individual buffer zones.

Buffer ID	Dependent Variable: ln (NDVI)				Dependent Variable: ln (kNDVI)				Obs.
	DID (%)	R ²	T-statistic	Standard error	DID (%)	R ²	T-statistic	Standard error	
1	-2.273	0.077	-506.29	0.000	-3.597	0.069	-455.01	0.000	264,311,939
2	-2.497	0.013	-298.49	0.000	-4.035	0.009	-264.34	0.000	97,484,591
3	-2.866	0.193	-123.05	0.000	-4.751	0.171	-122.59	0.000	2,760,828
4	-4.414	0.218	-94.832	0.000	-8.171	0.204	-99.069	0.001	1,919,074
5	1.827	0.06	23.433	0.001	3.946	0.059	26.827	0.001	8,093,233
6	0.17	0.075	10.117	0.000	0.034	0.06	1.197	0.000	4,782,252
7	0.706	0.028	30.67	0.000	1.069	0.021	27.49	0.000	3,068,840
8	-4.053	0.128	-79.445	0.001	-7.712	0.125	-83.894	0.001	2,628,256
9	-1.475	0.023	-183.72	0.000	-2.352	0.017	-166.45	0.000	52,050,327
10	0.3	0.013	35.422	0.000	0.615	0.01	41.869	0.000	39,988,487
11	0.085	0.011	12.966	0.000	-0.137	0.008	-11.98	0.000	87,891,951
12	-1.096	0.031	-150.10	0.000	-1.637	0.024	-129.10	0.000	48,986,638
13	22.335	0.027	210.522	0.001	40.946	0.029	213.325	0.002	7,477,097
14	-0.004	0.052	-0.584	0.000	-0.009	0.043	-0.713	0.000	35,544,924
15	-0.963	0.008	-116.82	0.000	-1.484	0.006	-101.82	0.000	62,967,859
16	0.606	0.024	58.289	0.000	1.041	0.019	55.702	0.000	53,476,844
17	0.878	0.025	99.967	0.000	1.654	0.019	103.094	0.000	103,770,936
18	0.566	0.09	85.104	0.000	0.738	0.078	64.343	0.000	64,086,835
19	-1.118	0.026	-199.48	0.000	-1.784	0.02	-181.78	0.000	82,321,044
20	2.172	0.028	180.971	0.000	3.671	0.022	170.478	0.000	43,367,687
21	-5.575	0.011	-24.539	0.002	-14.207	0.011	-34.494	0.004	3,262,607
22	0.233	0.032	12.809	0.000	0.458	0.025	13.86	0.000	21,180,308
23	-0.152	0.032	-7.571	0.000	-0.176	0.025	-5.165	0.000	3,734,644
24	1.33	0.018	34.784	0.000	1.833	0.015	30.697	0.001	7,474,185
25	-0.055	0.007	-3.367	0.000	-0.471	0.006	-16.222	0.000	21,987,229
26	1.217	0.105	52.224	0.000	2.181	0.095	52.959	0.000	8,271,775
27	1.053	0.036	49.92	0.000	1.403	0.03	36.892	0.000	15,176,747
28	1.511	0.026	171.561	0.000	2.775	0.022	175.438	0.000	80,234,672
29	1.476	0.124	53.387	0.000	2.636	0.12	57.542	0.000	1,491,868
30	1.144	0.109	89.546	0.000	1.628	0.091	75.082	0.000	9,954,400
31	21.631	0.358	421.268	0.001	40.021	0.346	429.85	0.001	2,407,179
32	-3.024	0.068	-140.88	0.000	-4.83	0.06	-129.55	0.000	6,194,130
33	4.152	0.077	873.534	0.000	6.603	0.055	774.337	0.000	139,011,318
34	1.658	0.006	86.736	0.000	3.238	0.005	92.04	0.000	28,420,486

35	3.332	0.114	289.154	0.000	5.849	0.109	284.562	0.000	50,128,178
36	-2.295	0.082	-38.43	0.001	-4.546	0.081	-38.89	0.001	29,896,197

Note: Columns 2 to 5 present the baseline regression results for the buffer zones of 36 sample projects, with the NDVI as the dependent variable. Columns 6 to 9 replace the dependent variable with the kNDVI. “Obs.” denotes the number of project observations.

- 2. Second, why limit the sample to sites with positive local effects, and term the other ones as “invalid”? This makes the results seem grossly misleading to me — akin to “if we exclude interventions that don’t work, then the intervention works!”

We sincerely appreciate your critical observation regarding the sample selection. Your comment has highlighted an important aspect of our study that requires clarification, and we are grateful for the opportunity to address this potential misunderstanding.

We want to emphasize that we did not limit our sample to sites with positive local effects. Our analysis was conducted using the full sample of all projects, including the baseline regression, heterogeneity analysis, and mechanism tests. We apologize for using the misleading term “invalid”, which we have now revised in our manuscript to “non-effective”. The distinction between “effective” and “non-effective” projects was only made when examining the effects within the project sites themselves, not for the overall analysis or spillover effects. Our intention in making this distinction was to explore why some projects showed negative effects, although this was not the primary focus of our study. We observed that most projects showing significant negative effects within their boundaries also demonstrated negative effects in their buffer zones. This pattern suggests that these projects could be influenced by broader systematic factors, such as region-wide climate-related impacts.

We fully agree with your implicit point that empirical research should not selectively focus only on projects with positive outcomes. Our intention is to present a comprehensive analysis of all projects, including those with both positive and negative outcomes, to provide a balanced and accurate representation of the overall impact of forest offset projects.

We are grateful for your reminder and question, which has prompted us to reconsider our presentation of results and improve the clarity of our findings.

Manuscript changes:

Page 4 paragraph 4 (line 126):

One site showed statistically insignificant effects and the remaining five projects (in gray) had negative DID estimators at the 1% level, most of whose nearby areas also exhibiting a negative trend (Fig 3d). The patterns observed in these negative projects suggest either the absence of forest gains or the influence of other systematic factors, including climate-related damage within a broader region. Therefore, these five significant negative projects are deemed as non-effective samples due to the lack of evidence of forestry activities inside the project boundaries.

Page 5 paragraph 1 (line 133):

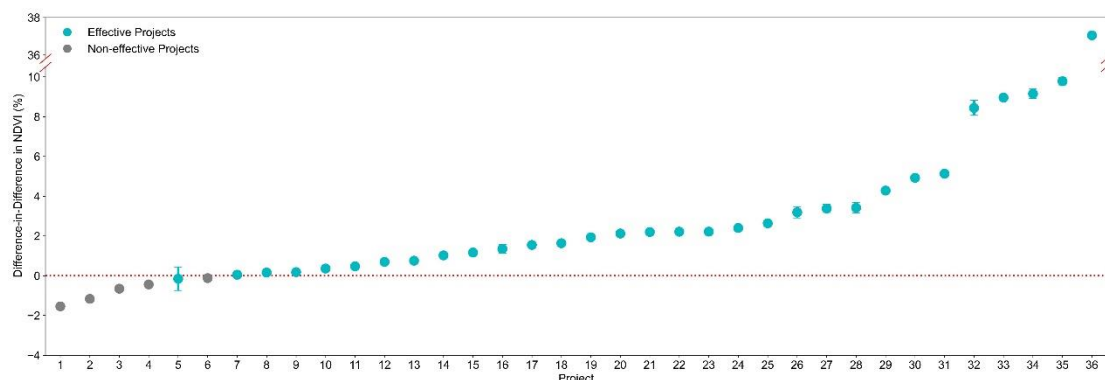


Fig.2. Forestry activities in project sites.....Plot d presents the average treatment effect on the treated for individual projects estimated by difference-in-difference regression, where the whiskers extending above and below each point indicate the 95% confidence interval.

Page 5 paragraph 2 (line 142):

Generally, the forest status in all buffer zones follows a similar trend to that of the project sites (Fig. S2), indicating regional ecological and physical links.

Page 5 paragraph 3 (line 150):

The full sample trends indicate that the pixels in buffer zones and their control groups exhibited consistent patterns prior to 2012, with notable changes emerging after 2012 (Fig. 3a). Over 23 years, the average forest gain spillover from the project sites to neighboring areas ranges from 0.91% to 1.60% (for NDVI and kNDVI respectively, $p < 0.01$, Table S1). This effect becomes increasingly positive in the post-treatment period, reflecting strengthened forestry efforts and cumulative impacts as the CCER policy was implemented.

Page 7 paragraph 3 (line 208):

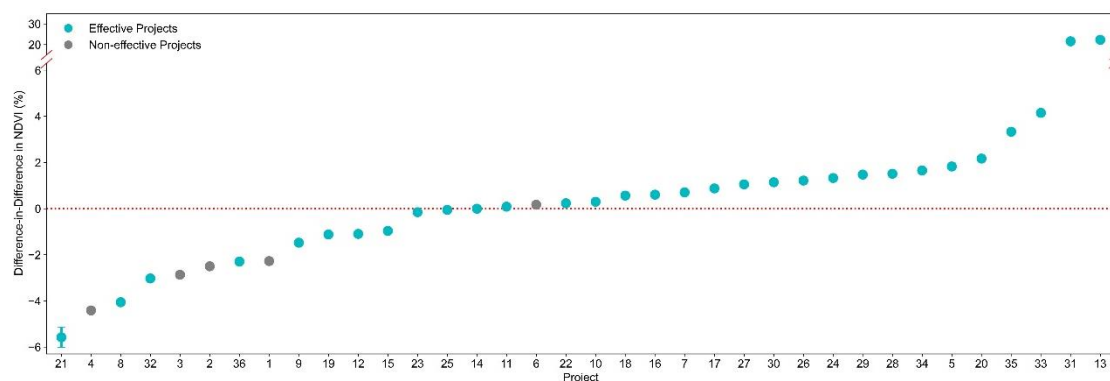


Fig.3. Forestry activities in buffers..... Plot d presents the average treatment effect on the treated for individual projects' buffer zones estimated by difference-in-difference regression, where the whiskers extending above and below each point indicate the 95% confidence interval.

- 3. Third, I'm not sure how the study reveals learning as "the primary driver of positive spillovers". Is this a result or a possible explanation? Why can we exclude social norms or other links? Please elaborate and clarify this in the paper.

We sincerely appreciate your insightful comment regarding our statement about the potential channels.

We apologize for any confusion caused by our imprecise wording. We acknowledge that we cannot conclusively exclude other potential mechanisms, such as social norms and other links, as you rightly mentioned. You are correct in pointing out that this is more accurately described as a possible explanation rather than a definitive

result.

In light of your feedback, we have reviewed our manuscript for similar instances where we may have overstated our findings, and we have made appropriate revisions to ensure our language accurately reflects the strength of our evidence throughout the paper. As for social norms, it would require more detailed data at the resident level. This limitation in our current study presents an opportunity for future research, which we now highlight in our discussion of the study's limitations and areas for further investigation.

Manuscript changes:

Page 1 paragraph 1 (Abstract, line 15):

Further analysis finds limited evidence of leakage, supported by land use change and the pattern of forest loss, possibly due to China's land policies and project features; instead, the positive spillovers are found to be facilitated through the learning channel of knowledge diffusion and information flow, supported by the reduction of wildfire activities and the patterns of project applications.

Page 10 paragraph 2 (line 290):

Our study suggests that learning could be one of the drivers of positive spillovers in forest offset projects. This learning effect operates through knowledge dissemination and behavioral learning^{15,27}.

Page 11 paragraph 1 (line 311):

Using a quasi-experimental approach with a 23-year panel dataset from 2000 to 2022, we find that forest offset projects induce overall positive forestry spillovers to their buffer zones, with learning potentially being one of the contributing channels.

Page 12 paragraph 3 (line 363):

We acknowledge several limitations to this study. First, while we have some evidence regarding the impact of learning effects, a more comprehensive understanding of why and how forest managers respond to policies necessitates more granular data. This could include individual behavioral data obtained through field investigations and interviews, which would also facilitate exploration of other potential mechanisms, such as social norms.

- 4. Fourth, the staggered design of your intervention complicates diff-in- diff estimation somewhat, but is not addressed in the paper — see Goodman-Bacon, 2021.

We sincerely appreciate your insightful comment and for educating us about the complications of the staggered design.

In response to your comment and in recognition of the issues highlighted by Goodman-Bacon (2021), we have conducted an additional DID regression following the methodology recently proposed by Callaway and Sant’Anna (2021). We are pleased to report that the results from this staggered DID analysis remain consistent with our baseline regression.

We have revised the corresponding statement in both manuscript and supporting information. We are grateful for your careful review and your reference to the relevant literature, which has led to this important improvement in our analysis.

Manuscript changes:

Page 6 paragraph 2 (line 171):

Fourth, to account for discrepancies between the CCER launch year and the project application submission time, we substituted the intervention time with the public comment start date for each project, conducting the DID regression within a staggered framework³⁹.

Supporting information changes:

Table S6. Robustness check.

	...	(11)
	...	Staggered did
DID		0.0032***
		(14.39)
Obs.		12,495,452

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; T-statistics are reported in parentheses (except for column 11). All columns include two-way fixed effects.Column 11 replaces the intervention year with the application submission year for each project, following the staggered difference-in-differences framework proposed by Callaway and Sant'Anna (2021). For column 11, ATT is reported with the z-values in the parentheses.

- 5. I feel that the abstract could be improved. First two sentences could be shortened, and maybe “less considered” could be rephrased to something like “quantifying impacts beyond borders remains an open issue”. I’d drop the “unique dataset”. The sentence on “Employing a matching [. . .] in multiple directions” could be clearer worded, and the one afterwards is too long. Last sentence could be more concrete — what do we learn?

We sincerely appreciate your thoughtful feedback on our abstract. Your suggestions for improvement are valuable and have helped us refine our presentation of the study. We have carefully considered all your comments and have revised the abstract accordingly to enhance its clarity and impact.

Manuscript changes:

Abstract

As forest-based carbon offset programs gain increasing attention, quantifying their impacts beyond project boundaries remains an open issue, particularly in subtropical and temperate regions. This study focuses on the local spillover effects of 36 forest offset projects in China’s voluntary carbon market. Using a matching approach and a difference-in-difference model, we compare the forest status of the

project areas and buffer zones to their reference areas from 2000 to 2022. Results show overall positive forest gains of 2.25% to 4.25% in project sites, with positive spillovers of 0.91% to 1.60% in neighboring areas, while the individual projects exhibit various spillover patterns. Further analysis finds limited evidence of leakage, supported by land use change and the pattern of forest loss, possibly due to China's land policies and project features; instead, the positive spillovers are found to be facilitated through the learning channel of knowledge diffusion and information flow, supported by the reduction of wildfire activities and the patterns of project applications. This study demonstrates that well-designed forest offset programs can yield benefits beyond their boundaries, providing insights for offset policy design and project implementation in China and similar contexts.

- 6. Figure 1 needs to be more accessible. The NDVI color scale is bad for red-green blindness and printing consider using a two-color scale (e.g. from bright or yellow to green) instead The NDVI visualization adds little information to where it is needed (project sites are almost all green), maybe a log-scale helps don't plot the 0s)? I don't see why it is necessary to visualize any kind of EEZ, and especially not just part of it (in a disputed area).

We sincerely appreciate your thoughtful and detailed feedback on Figure 1, and thanks for educating us the more inclusive color setting.

Following your advice, we have revised the NDVI color scale to use a gradient from bright yellow to purple. This new color scheme not only provides a clearer contrast but also ensures better accessibility for all readers, including those with red-green color blindness. We have also reviewed and adjusted the color schemes in all other figures throughout the manuscript for consistency and improved accessibility.

Concerning the NDVI visualization, we agree that the original 2019 NDVI visualization provided limited information. Therefore, we have replaced the 2019 NDVI data with 2005 NDVI data, which allows for a more informative representation of early vegetation status of project sites. Furthermore, due to the scarcity of zero values

in the dataset, the log-scale did not yield significant improvement as we expected, and instead, we have adopted a classified symbology approach.

Regarding the EEZ visualization, we understand your concern about the inclusion of the potential disputed area and we apologize for any confusion this may have caused. The inclusion of this element is a necessary requirement of our institution's figure review process. However, we have significantly reduced its prominence and moved it to a less conspicuous part of the figure to maintain focus on the main content. We appreciate your understanding in this matter.

Your feedback has been invaluable in enhancing the quality and inclusivity of our visualizations, making our research more accessible to a broader audience. Once again, we thank you for your insightful comments and the opportunity to improve our work.

Manuscript changes:

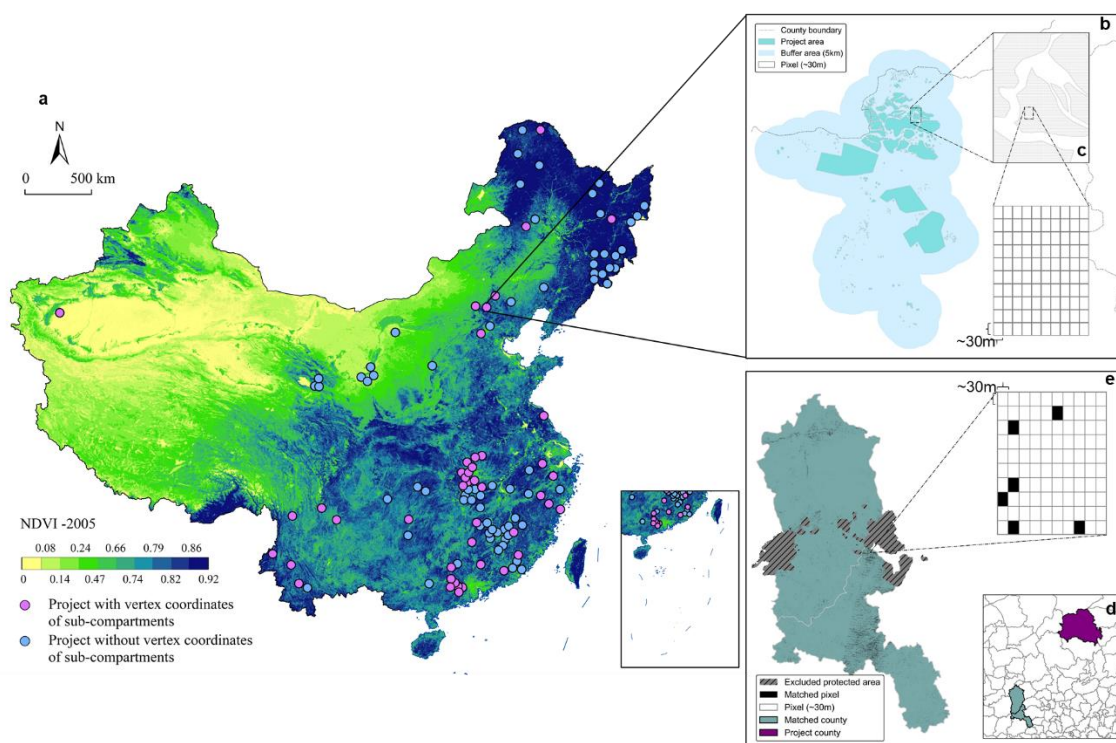


Fig.1. Study area and research design. Plot a depicts the national distribution of forest offset projects across China overlaid on the 2005 NDVI map, illustrating vegetation status. Red circles represent project locations with documented vertex coordinates (our study sample), while blue circles indicate projects without specified coordinates. Note

that a single county may have multiple project applications, and a project may span multiple counties. In plot b, one specific project site (blue) and its 5 km buffer zones (light blue) are illustrated. Plot c lays out a 30m×30m grid (gray) on the project area, precisely delineating its boundary. Plot d displays the matched control counties (dark green) within the same province as the project locations (purple). Plots e demonstrate our grid-based matching results (see Methods), with control pixels sampled from the control counties, excluding those within established protected areas. The matched pixels are shown in black.

- 7. I would move Figure 4 to the Appendix.

We sincerely appreciate your suggestion regarding the placement of Figure 4. As per your suggestion, we have moved the original Figure 4 to the Appendix. We agree that this relocation helps streamline the main text while still keeping the information accessible to readers who wish to delve deeper into our analysis.

In addition to relocating the figure, we have also taken this opportunity to enhance and enrich our placebo tests. We now randomly align both treatment pixels and intervention periods, repeating this process 1,000 times for both project sites and buffer zones. As a result, the content of placebo test Figure (original Figure 4) has undergone significant changes with the aim of further strengthening the credibility of our conclusions.

Manuscript changes:

Page 7 paragraph 2 (line 196):

Furthermore, we performed two types of placebo tests by randomly aligning the treatment pixels and intervention period using a bootstrapping procedure. We first created pseudo-treatment groups by randomly selecting samples from the control group, maintaining the original intervention period. Then, we randomly selected both treatment groups and intervention periods between 2000 and 2022. Both tests were repeated 1000 times. These placebo tests simulated scenarios where either no forestry

carbon sink projects occurred in the project areas, or other policy interventions happened in different years. The results (Fig. S6) show that our actual DID estimates lie at the extreme right tail of both the empirical distributions of the placebo estimates and the associated theoretical normal distribution. This implies a positive effect that is statistically different from zero, signifying that our observed treatment effects are unlikely to be the results of random factors and instead represent genuine policy intervention effects.

Supporting information changes:

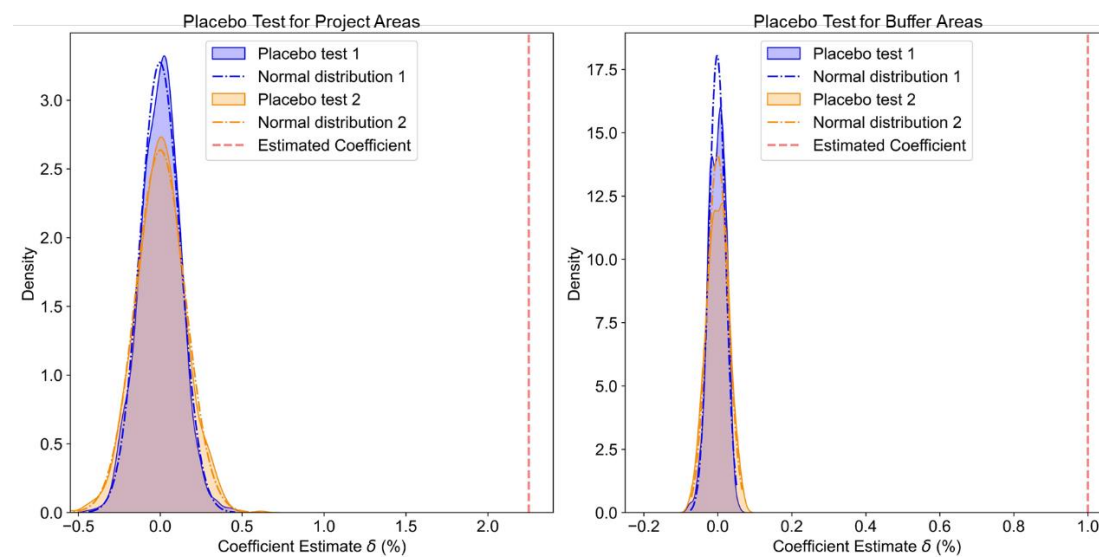


Fig.S6. Placebo test. The left plot is the placebo test for project sites and the right plot is for the buffer zones. The solid density curves illustrate the empirical distribution of the placebo estimates of the coefficient δ , the DID estimator. The superimposed dashed lines represent the theoretical normal distribution based on the central limit theorem, serving as a reference to assess how our empirical findings deviate from the expected distribution under the null hypothesis of no effect. Placebo test 1 (purple) displays the distribution of estimators for pseudo-treatment groups, while Placebo test 2 (orange) shows the distribution of estimators resulting from randomly assigning both treatment groups and policy intervention years. The dotted vertical lines indicate our actual DID estimators obtained from the baseline regression.

- 8. Please reword the discussion, which states that “We find limited evidence of negative spillovers, implying that leakage is not a major issue in China at present”, which is misleading to wrong. 1/The absence of evidence is not evidence of absence. 2/Instead of qualifying the time (“at present”), it would be much more important to qualify the scope where your results have validity (selected, i.e., effective forest-offset programs)

We sincerely appreciate your critical and insightful comment. Your observation has been truly eye-opening, making us realize that our previous statement was imprecise and potentially misleading. We apologize for our inappropriate wording and are grateful for your guidance on the importance of properly quantifying the scope of our findings.

Your feedback has prompted us to reassess and improve the accuracy of our statements throughout the manuscript. This valuable insight not only enhances the quality of our current work but also serves as an important lesson for our future academic writing.

Manuscript changes:

Page 11 paragraph 1 (line 311):

Using a quasi-experimental approach with a 23-year panel dataset from 2000 to 2022, we find that forest offset projects induce overall positive forestry spillovers to their buffer zones, with learning potentially being one of the contributing channels. Within the scope of our analysis, substantial leakage is not detected.

Page 12 paragraph 3 (line 371):

Additionally, considering the re-launch of CCER program in 2023, further research across a wider range of forestry projects in China is necessary to draw more comprehensive insights.

- 9. The small discontinuous polygons that are treated are weird to me — overall, I feel like the paper could do a better job at explaining what these forest-offset programs do and how – Wouldn't we expect leakage into the immediate vicinity then?

We sincerely appreciate your insightful comment. We acknowledge that we should have provided more comprehensive background information which could help to explaining the results. We thank you for the opportunity to clarify these important aspects of our study.

This fragmented nature of the forest-offset projects in our study are a unique feature of China's forestry projects, primarily due to China's land tenure system and the specific requirements of the CCER program.

Specifically, in China, forestland is often collectively owned, with use rights allocated to individual households. This results in a mosaic of small, fragmented forest plots. Despite the Forest Tenure Reform initiated in the early 2000s, the confirmation and registration of forestland ownership continue to present significant challenges. As a result, forest property rights remain complex in China, impacting the development of forestry projects.

As for the eligibility of CCER program, it has strict requirements for project sites. For instance, land must have been non-forested since a specific date (originally 2005, later revised to 2012), and there are limitations on single polygon size; moreover, clear land ownership rights must also be demonstrated. These criteria often lead to the selection of scattered, eligible plots within a broader landscape. Though the overall project sites are large, they could be very small in a single village or even a county. Consequently, these features are unlikely to yield large fluctuations in their local timber market, that is, makes it more difficult for market-driven carbon leakage to occur.

We have added more detailed background on these aspects to our manuscript to provide readers with a clearer understanding of the context and nature of these forest-offset programs.

Manuscript changes:

Page 2 paragraph 4 (line 67):

As the world's largest emitter, China faces substantial pressure for carbon offsets to achieve carbon neutrality. Its abundant forestry carbon sequestration resources are expected to play a pivotal role^{33–36}. In 2012, China launched its voluntary emission reduction market, the Chinese Certified Emissions Reduction (CCER) program, inviting project applications upon announcement. Project developers were required to submit detailed PDDs and materials specifying eligible project boundaries, planned or ongoing activities, and demonstrating project additionality. By 2017, when authorities suspended new CCER project applications, the official website had listed 100 forestry projects. In 2023, the CCER program was re-launched with updated measures, guidelines, and methodologies. Although the program was temporarily halted, our results suggest that forest managers continued to maintain and improve their forests, anticipating future offset credits.

Page 11 paragraph 2 (line 318):

Second, the scattered and limited-scale nature of the offset projects makes it challenging for any single project to exert a substantial influence on the supply and demand dynamics of global timber or crop markets^{18,59}. Forest offset projects in China are typically characterized by small, scattered, and disconnected polygons, mostly due to the land use features and strict site selection criteria. These include requirements for non-forested land since 2005 and have certified ownership rights, with the latter complicated by China's complex forest property system (see SI for details). These factors make it difficult to yield large fluctuations in the local timber market, as relatively small polygons are being regulated. Thus, leakage through shifts in market demand is less likely to occur.

Page 11 paragraph 4 (line 341):

Moreover, although not the primary focus of our study, our findings of forestry gains observed from 2012 both within and beyond the project sites lend support to the 2023 revision of the administrative measures for the CCER market. This revision shifted

the project eligibility criterion, changing the initiation date for eligible project activities from 2005 to 2012, aligning it with the CCER program's implementation date⁶⁰. This adjustment reflects a more targeted approach to capturing the effects of the CCER initiative.

Supporting information changes:

1. Supplementing background

Chinese Certified Emission Reduction (CCER) market is a carbon credits program governed by China's government. It is open to project owners who are legally registered persons or organizations in China, and is linked to China's national carbon market through the offset mechanism. The CCER market was initiated in 2012 when the National Development and Reform Commission issued the Interim Measures and key guidelines, followed by the launch of the National Trading Platform in October 2013. It is suspended in 2017, when the authorities had announced a total of 100 forestry projects on the official website. In 2023, it is re-launched with the Ministry of Ecology and Environment issuing new measures, guidelines, and methodologies.

Project implementation follows a process similar to many carbon credit systems, including project development and registration, validation and verification, and carbon credit issuance. Project developers must submit detailed Project Design Documents (PDDs) specifying project boundaries, planned activities, methodological compliance, baseline scenarios, and project additionality.

Site selection and boundary determination are governed by specific criteria outlined in the 2013 afforestation methodology (revised in 2023). Key conditions include: land must have been non-forested since February 16, 2005, with specific vegetation thresholds (must not simultaneously meet continuous area ≥ 0.0667 hectares, canopy cover $\geq 20\%$, tree height ≥ 2 meters after forest formation); clear land ownership rights, certified by county-level or higher government, are required; land must not be classified as wetland or organic soil; project activities must comply with soil and water conservation requirements; and the pre-existing agricultural

activities should not shift from the intended project sites to other sites. It's important to note that forest property rights remain complex in China, impacting project implementation. Despite the Forest Tenure Reform initiated in the early 2000s, the confirmation and registration of forestland ownership continue to present significant challenges (Ho, 2001; Qin et al., 2011).

The 2023 revision of the interim measures for the CCER program introduced significant changes in project eligibility criteria. Notably, it stipulates that project activities cannot predate 2012. This revision also shifts the eligible project start date in the afforestation methodology from 2005 (as specified in the previous version) to 2012, ensuring consistency with the CCER program's launch. It's important to note that while project activities could predate the generation of the PDDs, the crediting period is defined as the timeframe when the project scenario generates additional greenhouse gas emission reductions compared to the baseline scenario.

During the validation and verification period, when the project officially begins implementation, all activities within the project boundaries must be meticulously recorded and reported. This comprehensive documentation covers a wide range of aspects: afforestation activities, including planting details, tree species used, and land preparation methods; management activities such as patrols, replanting efforts, harvesting operations, pest control measures, and forest fire prevention strategies; and any occurrences of natural disasters or human disturbances.

- 10. Line 372 has a typo (“synthesizing”)

Thank you for your attentive reading of our manuscript and for bringing this potential issue to our attention. We greatly appreciate your commitment to ensuring the accuracy of our work.

Upon careful review of line 372, we found that the word “synthesizing” is used in our manuscript. We understand that you may have been considering the alternative spelling “synthesizing”, which is also correct. After consulting dictionaries, we’ve

confirmed that both “synthesizing” and “synthesizing” are accepted and commonly used spellings, with the former one being the standard American English spelling and another more common in British English.

In our manuscript, we’ve consistently used American English spelling conventions, which is why “synthesizing” appears. However, we sincerely appreciate your attention to detail, as it prompted us to double-check our spelling choices and ensure consistency throughout the manuscript.

- 11. The 13 years before CCER seems excessive?

We sincerely appreciate your insightful observation regarding the length of the pre-CCER period in our study. Your comment has prompted us to conduct additional analyses to address this concern.

We acknowledge that a 13-year pre-intervention period (now has changed into 11-year after we change our dataset) could potentially introduce uncertainties into our estimates. To address this concern, we have conducted additional analyses with shorter time frames: (i) we performed an analysis for the period 2007-2017, which yielded an estimation of 1.19%; (ii) we also analyzed the period 2009-2015, which resulted in an estimation of 1.44%. These results are consistent with our primary findings (0.91%) and, interestingly, exhibit larger effects with shorter intervals. This pattern not only validates your astute observation but also further reinforces the robustness of our conclusions.

We have included these additional analyses in the robustness check section of our manuscript. We are grateful for your critical evaluation, which has led to a more comprehensive analysis and strengthened the validity of our study.

Manuscript changes:

Page 6 paragraph 2 (line 173):

Lastly, to address potential temporal confounding during our long study period, particularly in light of China's intensified forestation efforts post-2000 and the initiation of several large-scale tree plantation programs in the early 2000s⁴⁰, we conducted additional estimations using shorter intervals 2007-2017 and 2009-2015.

Supporting information changes:

Table S6. Robustness check.

(6)	(7)
2007-2017	2009-2015
0.0119***	0.0144***
(30.35)	(29.10)
0.0599***	0.036***
(120.98)	(69.38)
-0.0047***	-0.0052***
(-16.53)	(-13.74)
18.09***	18.11***
(26899)	(22269)
YES	YES
8,088,606	5,148,618
0.58	0.58

References Cited in Response to Reviewer:

- [1] MacDougall, A. S., Esch, E., Chen, Q., Carroll, O., Bonner, C., Ohlert, T., ... & Seabloom, E. W. (2024). Widening global variability in grassland biomass since the 1980s. *Nature ecology & evolution*, 1-12.
- [2] Ghafoor, G. Z., Sharif, F., Shahid, M. G., Shahzad, L., Rasheed, R., & Khan, A. U. H. (2022). Assessing the impact of land use land cover change on regulatory ecosystem services of subtropical scrub forest, Soan Valley Pakistan. *Scientific Reports*, 12(1), 10052.
- [3] Trujillo, E., Molotch, N. P., Goulden, M. L., Kelly, A. E., & Bales, R. C. (2012). Elevation-dependent influence of snow accumulation on forest greening. *Nature Geoscience*, 5(10), 705-709.
- [4] Jiang, P., Liu, H., Piao, S., Ciais, P., Wu, X., Yin, Y., & Wang, H. (2019). Enhanced

growth after extreme wetness compensates for post-drought carbon loss in dry forests. *Nature Communications*, 10(1), 195.

- [5] Maes, J., Bruzón, A. G., Barredo, J. I., Vallecillo, S., Vogt, P., Rivero, I. M., & Santos-Martín, F. (2023). Accounting for forest condition in Europe based on an international statistical standard. *Nature Communications*, 14(1), 3723.
- [6] Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2), 200-230.
- [7] Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., ... & Running, S. W. (2021). A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, 7(9), eabc7447.

In conclusion, we would like to express our deepest gratitude to the reviewer. We feel incredibly fortunate to have received such thoughtful and comprehensive feedback from you. Your comments have been truly constructive, feasible, specific, and invaluable, addressing not only the content and logic of our study but also improving our figures, wording, and overall writing. We have embraced all of your suggestions and implemented modifications based on your feedback. Your meticulous review has not only improved this particular study but has also provided us with valuable insights that will undoubtedly influence our future research endeavors.

**Response to Reviewers' Comments on
Forest vegetation increased across China's carbon offset projects and positively
impacted neighboring areas (COMMSENV-24-0989-A)**

The reviewers' original comments are provided in **bold**, followed by our responses. Excerpts from the revised manuscript are provided in *italics*.

[Response to Reviewer #1](#)

I would like to thank the authors for carefully considering my comments. The authors have addressed all of my comments thoughtfully and sufficiently. There must have been a lot of effort in making thorough and substantial modifications. I really appreciate it! This revised version is now greatly improved and clearer. Overall, the revised version uses data with higher resolution and more reasonable metrics to evaluate the projects' effectiveness, along with neat robustness checks and heterogeneity analyses. I am satisfied with the updated manuscript and recommend it for publication in Communications Earth & Environment.

We sincerely thank you for your thoughtful acknowledgment and kind words. We are glad to know that you recognize the dedication involved and are satisfied with the revised manuscript. Your support means a lot to us, and we are grateful for all your constructive feedback.

[Response to Reviewer #2](#)

The author has made a good modification according to the reviewer's opinion. I'd like to give an accept decision.

Thank you very much for your positive feedback and for acknowledging the modifications we have made based on your suggestions. We are pleased to hear that you have decided to accept the manuscript.

Your support and constructive comments have been invaluable in improving the quality of our work.

Response to Reviewer #3

I appreciate the careful responses to the reviewer comments.

Thank you for your acknowledgment and all the constructive comments, which have helped to improve the quality of the work.

Response to Reviewer #4

The authors have clarified remaining issues and made major updates to the paper, and I believe the manuscript would be well suited to CE&E with minor revisions to improve the presentation. Congratulations on the interesting paper!

We sincerely thank you for acknowledging the modifications we have made! We also greatly appreciate all your insightful feedback, which has significantly improved our manuscript. We have taken all your comments into consideration and made corresponding modifications. We believe they have indeed helped to enhance the presentation of our work. Thank you once again for your time and support, which mean a lot to us.

Regarding earlier comments:

- Figure 1 is much improved. In the new version, the inlays have a font size that is too small to be read in print, and are arguably a bit overloaded. My suggestions (feel free to adopt which ones you deem sensible) would be to

- Relabel the red / blue circles as “Project with coordinates (study sample)” / “Project without coordinates” to remove clutter

- Increase the legend’s font size – especially for the plots to the right

- Move the legend of plot (b) to the bottom-left

- Drop “county boundary” and “Pixel (30m)” from the legends

- Consider dropping either plot (e) or the pixel grid in plot (c) to make more space for larger text ()

- I understand that there is a figure review process, and ultimately that is down to the journal. I do appreciate the changes made.

- I'm still unsure about “synthesizing”, which suggests that this occurred on its own. Perhaps “synthesized” is a better fit? That's with respect to:

"These matched pixels were eventually synthesizing into a reference area. "

Thank you very much for your meticulous and valuable comments regarding Figure 1. We have taken all your suggestions into consideration and have made the modifications to the figure accordingly.

Regarding the use of the word “synthesizing”, we fully agree with your perspective and appreciate your precise observation regarding the language formulation.

Manuscript changes:

Figure1:

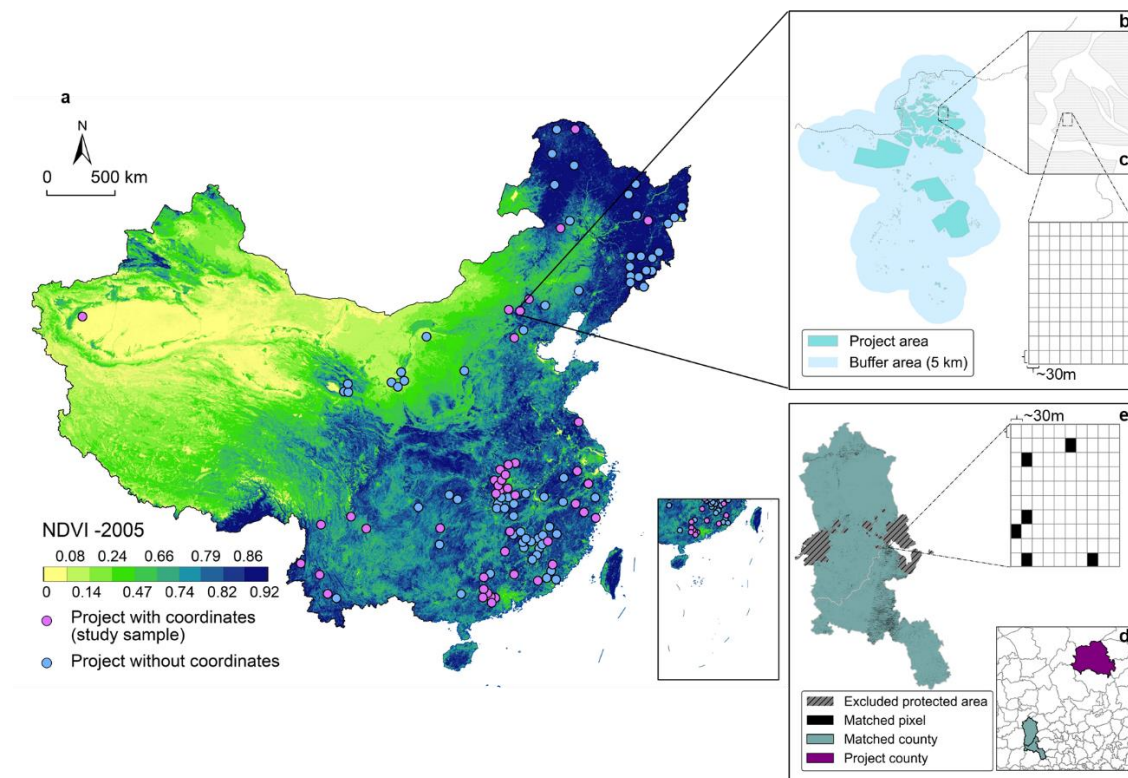


Fig.1. Study area and research design.

Line 418:

These matched pixels were eventually synthesized into a reference area.

Minor comments:**- Line 33 could drop the “Notably, ”**Manuscript changes:

Line 33 (where “Notably” has been dropped):

Leakage (negative spillover), referring to shifts in forest loss to areas that are unprotected, might undermine the value of offsets.

- Line 54 and below could be more nuanced (it’s very broad and not necessarily part of the results). Maybe something like “Spillovers could play an important role in the development ...”

We thank the reviewer for helping us improve the readability and logistics of this sentence, and we have made modification accordingly.

Manuscript changes:

Line 55:

Spillovers could play an important role in developing methodologies for forest offset programs, as they may impact the effectiveness and sustainability of these programs. A spillover aligned with the program’s objectives could enhance the overall project benefits, while a negative spillover undermining these goals can cause over-crediting, necessitating explicit mitigation measures in the methodologies.

- Figure 2 would also benefit from larger font sizes; perhaps it would be better to focus on plots (a) and (c) that are very convincing, perhaps even dropping (b) and (d). Estimates for individual projects in (d) will be much less reliable for interpretation, and you mention the important parts in the text and supplement anyways. Same comments hold for Figure 3.

Thanks for your valuable suggestions regarding Figures 2 and 3. We have increased the font sizes and, following your advice, moved the original plot (b) to the supplementary information to highlight the more valuable and convincing main results. Your feedback has greatly enhanced the visualization and clarity of our figures.

Manuscript changes:

Figure 2:

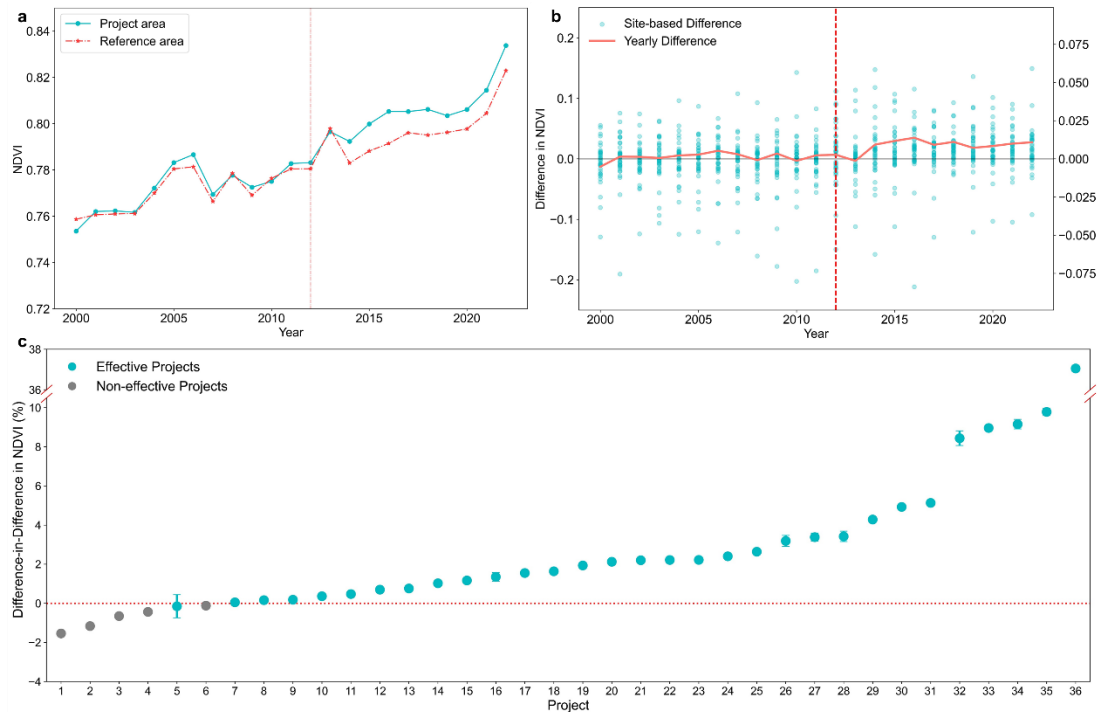


Fig.2. Forestry activities in project sites. Plot a illustrates the annual mean trends of NDVI in project areas and reference areas, with the red vertical line representing the year 2012. Plot b presents the annual site-based NDVI differences between each of the 36 project sites and their corresponding reference areas, as represented by blue dots for each year. The red line across the plot represents the mean of these differences across all project. Plot c presents the average treatment effect on the treated for individual projects estimated by difference-in-difference regression, where the whiskers extending above and below each point indicate the 95% confidence interval.

Manuscript changes:

Figure 3:

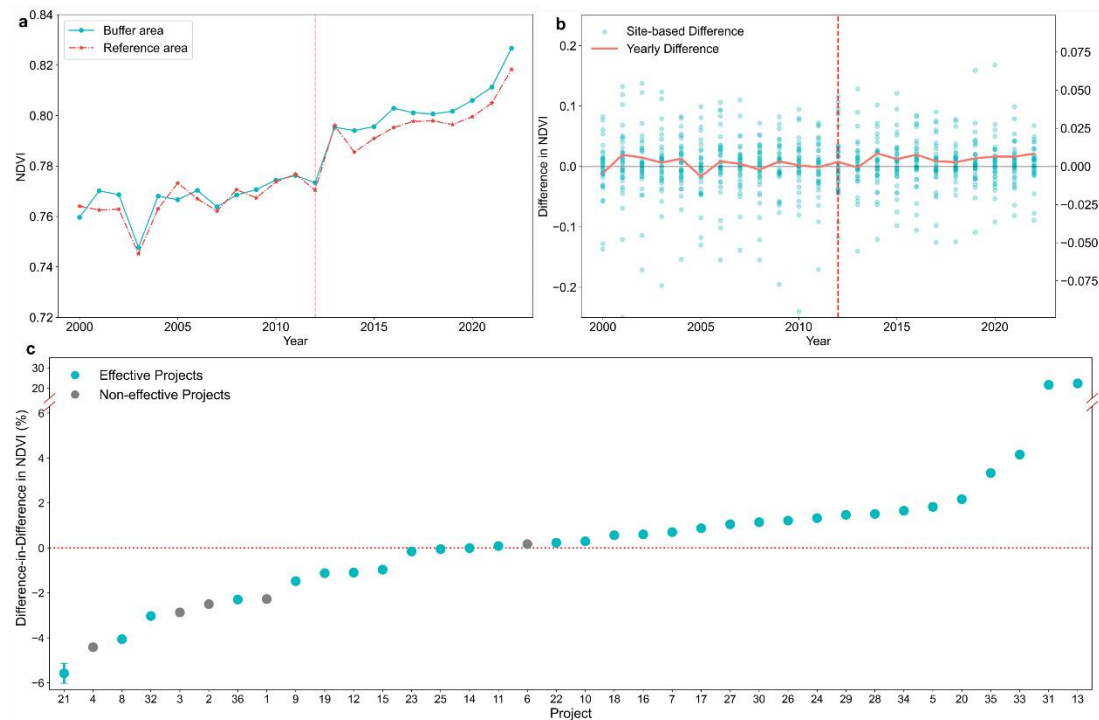


Fig.3. Forestry activities in buffers. Plot a illustrates the annual mean trends of NDVI in the buffer zones of the project and reference areas, with the red vertical line representing the year 2012. Plot b presents the annual site-based NDVI differences between each of the 36 project sites and their corresponding reference areas, as represented by blue dots for each year. The red line across the plot represents the mean of these differences across all projects. Plot c presents the average treatment effect on the treated for individual projects' buffer zones estimated by difference-in-difference regression, where the whiskers extending above and below each point indicate the 95% confidence interval.

- Typo in line 179: "First, *T*o"

- Lines 185:186 are unclear to me, which other team?

- Wording in line 188 is unclear: "This is attributed to *the fact that?* ~the~ most of the projects consist of numerous small-scale polygons"

- Line 327 should probably be "Compared to *the* existing literature~s~"

Thank you so much for your meticulous review and invaluable suggestions. We deeply apologize for any ambiguities and incorrections in our language expression. We

have addressed the typos and corrected the inaccurate phrasing by changing “another team” to “another source”. Furthermore, we have incorporated your other recommended changes, which we found very helpful. We have also reviewed the entire article to ensure that all corrections and improvements have been made consistently. Thank you once again for your contributions!

Manuscript changes:

Line 165:

First, to address...

Line 172:

...we used 1km×1km resolution data from another source for matching and baseline regression.

Line 175:

This is attributed to the fact that most of the projects consist of numerous small-scale polygons...

Line 297:

Compared to the existing literature, our findings also show a considerable magnitude.

- Good to see that the results hold with GFW data. Since it is just a check I would not worry about time consistency, but you should be aware of <https://www.globalforestwatch.org/blog/data-and-tools/tree-cover-loss-satellite-data-trend-analysis/>

Thank you for your positive feedback on the consistency of our results with GFW data, inspired by your valuable comments. We greatly appreciate your understanding of our process and usage of GFW data. We have taken careful note of the blog post you mentioned on the GFW website and have reviewed the information provided to ensure that our work takes it into account. Thank you once again for bringing this to our attention and for your helpful guidance.