

Stronger institutional performance correlates with ecological effectiveness

Corresponding Author: Professor Tobias Böhmelt

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.

This manuscript has been previously reviewed at another Nature Portfolio journal. This document only contains reviewer comments and rebuttal letters for versions considered at Communications Earth & Environment.

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 Böhmelt,

Your manuscript titled "Are Institutionally Effective Environmental Organizations Also Ecologically Effective?" has now been seen by 3 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. We would be interested in considering a revised version that fully addresses these serious concerns.

For publication in Communications Sustainability to be appropriate, your study must

* provide compelling novel insights to support causal estimates on the question of institutional effectiveness;

* transparently report methods and limitations.

We hope you will find the reviewers' comments useful as you decide how to proceed. Should additional work allow you to address these criticisms, 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.

When resubmitting, please provide a point-by-point response to the reviewers' comments. Please submit your responses as a separate file, distinct from your cover letter where you can add responses to the Editors' comments that you do not want to be made available to the reviewers. Word files are preferred. We recommend that any figures, tables or graphs that are included in the response to reviewers are also included in the main article or Supplementary Information.

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 Sustainability 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,

Yann Benetreau, PhD
Deputy Editor, Communications Sustainability
Consulting Editor, Communications Earth & Environment
Nature Portfolio
ORCID: 0000-0002-1897-0887
New York Office

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:

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.pdf>

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):

Summary

This paper examines two forms of effectiveness of international environmental agreements and measures the effect of organizational effectiveness on ecological (environmental) effectiveness. Using three sets of quantification strategies, the authors find a negative effect of institutions on PM2.5 emissions.

This paper is interesting and taps into long-standing debates regarding the role of international agreements in addressing critical environmental issues. The paper is well-written and clearly organized. Nevertheless, I have several concerns. The plausibility of the estimates having a causal interpretation, in particular, is unclear to me.

Comments

(a) Contribution and current debates

I would normally not comment on this, but there are several instances where the authors contextualize their study in a literature that is dated. To give a few examples:

- I40: I'm not sure a study by Zürn from nearly thirty years ago is indicative of current trends in the study of international relations.

- I54: is our understanding still incomplete in the same manner as it was described in a paper from 15 years ago?

- I56: is a study from 10 years ago still offering an overview of the state of the art today?

It may well be that these are the latest references in this literature, but it would then be worth clarifying why so little progress has been made (which, in my own view, touches upon the concerns raised by the authors: getting causal quantitative estimates is very hard in this topic).

(b) Concepts and theory

The paper articulates a theory that contrasts "institutional effectiveness" and "ecological effectiveness". This is interesting, but I am not clear as to where the boundaries of the two concepts are. The authors for instance cite: "Jackson and Bührs (2015: 65) not only emphasize the need to distinguish between institutional effectiveness—focused on implementation and

compliance ..." But the way institutional effectiveness was earlier defined ("the extent to which they [organizations] achieve their stated objectives") does not match. The latter's output is also environmental (or ecological) effectiveness ("stated objectives").

Note that this doesn't undermine the authors' main theoretical idea (implementation vs biophysical impact), but it makes the fit of this study in existing debates awkward.

On a side note, there is a large literature in organizational theory and management that, I think, connects to the authors' argument, and I suspect readers from that audience would find this paper interesting. It might make sense to cite this literature as well to put this paper on these readers' map.

(c) Empirical analysis

I have several questions here, some stylistic in nature and others about the analysis itself.

First, it would be helpful to the reader if the authors briefly discuss the six international organizations. As I arrive at Table 1, I have no idea what exactly I am looking at. I don't even know why PM2.5 is the relevant outcome variable. The SM is useful and addresses this in general, but I don't think it would hurt to add a bit more details in the core text.

Second, it would also be good to describe institutional effectiveness. Sentences such as "the estimated coefficient of Institutional Effectiveness in Models 1-4 is -0.175, which translates into a reduction of PM2.5 by 0.175 units for every increase in an organization's institutional performance" are meaningless without knowing the units of the treatment.

Third -- and this is my main concern -- the authors' main "selling" point is the notion that this analysis would generate causal estimates (though they are careful to qualify this statement on l.101). The argument is that the use of three models (fixed effects, ECM, and between-within) addresses problems of endogeneity and self-selection. These are the core issues that have plagued this literature: states self-select into environmental treaties that themselves are set at a bar that is achievable. Here, we would therefore have two worries: states self-select into one of the 6 agreements, and these agreements themselves are calibrated to achieve feasible goals.

The effectiveness of a strategy like two-way fixed effect to address such problems is only guaranteed under very strict conditions that are not discussed in this paper. For instance, the FE model will capture causal effects in an unbiased manner if confounders are time-invariant and there is no heterogeneity in treatment effects. The plausibility of this statement is not discussed. And in fact, the model specification (including a third polynomial of GDP per capita -- far from a conventional difference-in-difference model) does suggest that this analysis is model-based rather than drawing on exogenous source of variation in effectiveness. Nor do the FE address in an of themselves self-selection.

As such, I don't see how the causal interpretation (however qualified by the author) is obtained via these three measurement strategies. The underlying problem remains. Contrast this for instance with pieces such as Lupu (2013, AJPS) which is not free of problem but tries to tackle self-selection.

Fourth, it was not clear to me why countries-years are the appropriate unit of analysis. If we are interested in the effectiveness of an institution, wouldn't the appropriate approach be to aggregate PM2.5 at the level of all member organizations? Aren't there really 6 observations? The theory appeared to be rooted in the international organization's structure. Alternatively, this is a story about domestic action, in which indeed, states are the unit of analysis. But then this raises all the questions asked previously about endogeneity as well as add questions such as domestic implementation of international rules, domestic state capacity, etc. and whether these interact (or mediate) international rules.

Fifth, the authors write "It is also interesting that out of the 1,000 simulations, only 22 of them (2.1 percent) are positively signed when using the dyadic data structure" Eyeballing model 2, the t statistic is about 6. A distribution centered at 6 will have virtually no negative value. Is the simulation then really surprising? Doesn't it mechanically relate to the regression estimates?

Reviewer #2 (Remarks to the Author):

Referee Report

Summary

This paper investigates whether higher institutional effectiveness of international environmental organizations, defined as the successful performance of their core functions, leads to better ecological outcomes in member countries. The author also examines how organizational design influences this relationship and addresses concerns about self-selection. The question is timely and important, given the role international organizations play in global environmental governance. Despite this relevance, empirical evidence on the topic remains scarce. And it is a gap the author intends to fill.

To shed some light on this issue, the author uses a dataset containing measures of institutional effectiveness for environmental organizations and ecological outcomes, specifically PM2.5 in member countries. Institutional effectiveness is measured using the performance indicators compiled by Lall. The empirical strategy includes two-way fixed effects OLS regressions, a within-between estimator, and a general error correction model.

The main finding is that institutional effectiveness is robustly associated with improved ecological outcomes. Across all models implemented, higher institutional effectiveness leads to statistically significant reductions in PM2.5 pollution. The

general error correction model shows both an immediate effect and a larger long-term effect. The within-between estimator indicates economically large within-country and between-country effects, even though the latter is not statistically significant.

Comments

This paper contributes to the literature by providing empirical evidence on a previously understudied question. It is a well-written and well-executed paper. Although it does not present a perfect setting for causal inference, I generally find the author's conclusions convincing. It offers useful guidance to policymakers, particularly in terms of considerations around joining international organizations. By shedding light on the relationship between organizational effectiveness and ecological outcomes, it informs both institutional design and strategic membership decisions.

That said, I have several comments:

1. The most common concern in studies examining the causal effects of membership in international organizations is the issue of self-selection. Countries that join environmental institutions, especially those with higher effectiveness scores, might be countries becoming more environmentally committed. The author addresses it in part through a Heckman selection model presented in the Supporting Information, which is much appreciated. I still wonder if it is possible to conduct placebo tests that examine the impact of institutional effectiveness on contributor countries. If self-selection is not a major concern, the estimated effects on contributors, who, according to the author, are not expected to benefit directly from the institutions' activities, should be close to zero.
2. Also, the paper would benefit from more detailed summary statistics on the institutional effectiveness variable, particularly for the six organizations analyzed and their membership structures, in the Supporting Information. It would be helpful to understand whether these organizations target countries with similar geographic or developmental profiles. This information would give readers a clearer view of the sample and the sources of variation in the independent variable.
3. A further concern for me is the inclusion of certain organizations in the sample. Specifically, the Adaptation Fund, the Climate Investment Funds, and the Multilateral Fund for the Implementation of the Montreal Protocol are not primarily designed to reduce PM2.5 pollution. While it is plausible that some of their projects, such as the retirement of coal as a part of the decarbonization scheme, generate air quality co-benefits, they most likely focus on unrelated objectives, like adaptation infrastructure. So, my guess is that the institutional effectiveness of these international organizations should not have a large impact on PM2.5 of their members. It would be helpful to report results excluding these three organizations from the sample to assess whether the main findings hold when focusing only on other institutions with a more direct connection to air quality outcomes.
4. For the reproducibility of the paper, the author should also be more explicit about how target and contributor countries are defined and assigned for each of the six international organizations. While some description is provided, it would help readers and future researchers if the rules used to classify countries into these categories were clearly stated, ideally with a short summary table or reference to an appendix with coding decisions.

Reviewer #3 (Remarks to the Author):

Title: Are Institutionally Effective Environmental Organizations Also Ecologically Effective?

This manuscript explores the relationship between institutional effectiveness and ecological effectiveness in international environmental organizations. Using panel data on six organizations from 2008 to 2018 and employing three econometric models, the study identifies a robust, likely causal relationship between the two dimensions, as measured by reductions in PM2.5 pollution. While the topic is relevant and the methodological rigor is commendable, there are significant concerns about the clarity of the concepts, the precision of the empirical claims, and the originality of the findings. The paper would benefit from a more focused theoretical framework, more precise terminology, and a clearer articulation of its contribution to the literature.

I recommend major revisions before the manuscript can be considered for publication.

Major Concerns and Suggestions

1. Conceptual Clarity: Ambiguity in Core Terms

The manuscript relies heavily on the distinction between institutional effectiveness and ecological effectiveness, yet these terms are not sufficiently defined early on.

Suggestion: Provide clear, operational definitions of these concepts in the introduction, with concrete examples of what is included or excluded in each. If ecological effectiveness is proxied solely by PM2.5 improvement, consider using more precise terminology such as "air quality improvement" or "pollution reduction."

2. Overuse of Generalized Terms

The term "ecological effectiveness" suggests broad environmental success, but the analysis focuses exclusively on PM2.5 reduction.

Suggestion: Either expand the dependent variable to include multiple ecological indicators or revise the terms to avoid overstating the ecological scope.

3. Research Motivation and Contribution

While the manuscript offers a rigorous empirical analysis linking institutional and ecological effectiveness, the core conclusion—that better-performing international environmental organizations are associated with improved environmental outcomes—is arguably intuitive and may align with widely held assumptions among scholars and policymakers. Many readers are likely to expect that reputable and well-established institutions would contribute positively to ecological outcomes, particularly in measurable domains like air quality.

In this context, the current framing of the research question lacks a strong sense of novelty. The paper would benefit from

clarifying: (1) what specific gap in the literature it seeks to address, and (2) what non-obvious insight is produced by the analysis. For example, if the novelty lies in identifying the causal magnitude of institutional effectiveness using new data, or in isolating the role of organizational design features, this should be made explicit early on—especially in the abstract and introduction.

Moreover, the motivation for the study remains somewhat abstract. It is not clear what problem or empirical ambiguity this research is intended to resolve. If existing studies have assumed, but not empirically demonstrated, that institutional effectiveness drives air quality improvements, this assumption should be problematized more clearly. If the study contributes by confirming this link through more robust econometric modeling and high-quality data, then this contribution should be framed as a specific and novel advance rather than a validation of conventional wisdom.

Suggestion: Strengthen the motivation and contribution sections by clearly articulating (1) what prior studies have missed or inadequately addressed, (2) how this study fills that gap, and (3) what new, specific knowledge is gained beyond the affirmation of a presumed positive relationship between institutional and ecological effectiveness.

4. Data validity and Policy Timing Bias in Measuring Ecological Effectiveness

While the manuscript aims to provide a robust empirical link between institutional and ecological effectiveness, it relies exclusively on modeled PM_{2.5} concentration data derived from satellite observations and chemical transport models (van Donkelaar et al., 2021). Although such data offer broad spatial coverage and address the issue of uneven in-situ monitoring, they carry notable limitations in accuracy. The authors do not discuss the uncertainty inherent in these modeled estimates, nor how such uncertainties may affect the measurement of ecological effectiveness or bias the inferred institutional impact. A discussion of error margins, data validation (e.g., cross-referencing with ground-based data), and sensitivity analyses using alternative pollution metrics would significantly enhance the credibility of the findings.

Furthermore, the analysis does not adequately control for the heterogeneous timing and intensity of national air quality policies, which may be independently responsible for PM_{2.5} reductions. For instance, many east Asian countries implemented aggressive air pollution control measures at different time points during the study period. These external policy shocks may confound the observed relationship between institutional effectiveness and pollution levels, especially if countries with high institutional performance are also those undertaking strong domestic reforms. Relying solely on country and year fixed effects may not be sufficient to disentangle these effects.

Suggestion: The authors should either (1) incorporate additional controls to account for national environmental policy timelines and implementation intensity, or (2) clearly acknowledge this as a limitation that could undermine the causal interpretation of the results. Additionally, a detailed discussion on the validity and uncertainty of the PM_{2.5} data used as the dependent variable is essential for evaluating the robustness of the ecological effectiveness measure.

Minor Comments

[Line 1–3] The term “institutionally effective” in the title and abstract may be unclear to readers unfamiliar with the literature. Consider rephrasing or briefly defining it early in the abstract.

[Line 7–8] The phrase “organizations’ performance” is overly general. It would be helpful to briefly specify what aspects of performance are being evaluated—e.g., goal achievement, transparency, cost-effectiveness—as this would enhance the credibility and informativeness of the abstract.

[Line 9] The term “ecological effectiveness” appears to refer exclusively to reductions in PM_{2.5} concentrations. If that is the only environmental indicator used in the study, the expression may be too broad and risk misleading readers. Consider specifying that the ecological effectiveness assessed here refers to air quality improvements (PM_{2.5} reduction).

[Line 41] I don't understand why their effectiveness is the dominant force in international relations analysis. Please add an explanation to supplement this argument.

[Line 48–49] What kind of problem structure and organizational design? It would be helpful if you could give a clear example of one.

[Line 59–70] While the conceptual distinction between institutional and ecological effectiveness is theoretically grounded, its necessity and clarity could be better justified in the introduction. For readers unfamiliar with the framework, the rationale for separating these dimensions should be more explicitly developed, perhaps with real-world examples to illustrate their divergence.

[Line 81] It is necessary to explain what factors are needed to evaluate institutional and ecological effectiveness. Even in the aforementioned previous studies, it is unclear what data was used to produce the results. However, when they suddenly say “high-quality data,” I don't understand what they mean.

[Line 171–173] The robustness of the results is notable; however, the consistently strong associations raise questions about whether potential confounding factors or model overfitting have been fully ruled out. Further discussion on model limitations or sensitivity tests in appendix would strengthen confidence in the findings.

[Line 190–191] I am not sure if this factor is important in considering the effects within a country. How are income, population, democracy, and globalization related to institutional effectiveness?

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

Communications Sustainability 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:

Dear Professor Böhmelt,

Please first accept our sincere apologies for the delay in reaching a decision on your submission as a review was much delayed.

Your manuscript titled "Are Institutionally Effective Environmental Organizations Also Ecologically Effective?" 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 Sustainability.

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 maximize the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

* Throughout the text, please avoid self-laudatory language (e.g., claiming that you provide robust evidence).

* Please thoroughly address the remaining reviewer concerns, avoid overstating causal findings, and transparently report limitations.

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 Sustainability 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 Portfolio, please visit <https://www.nature.com/commssustain/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.**

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

Best regards,

Yann Benetreau, PhD
Deputy Editor, Communications Sustainability
Consulting Editor, Communications Earth & Environment
Nature Portfolio
ORCID: 0000-0002-1897-0887
New York Office

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

Dear Author,

Thank you for engaging with my review. I found the answers and edits by and large convincing. This being said, I remain skeptical regarding the ability to make causal claims from these estimates. First, the fixed effects aren't helping self-selection bias based on differential treatment effects.

Second, the Heckman selection models (Table 8, SI) too, raise questions. They rely on very strong assumptions regarding the exclusion restriction on the instrument used in the selection stage, which are here: democracy and population (and a cubic time trend). Given that you articulate the notion that the story in this paper is a domestic politics one (in response to my question about the unit of analysis), and given that we know that there exists a large literature on the effects of variables such as democratic institutions on environmental outcomes, it seems difficult to believe that these variables used in Table 8, SI (democracy, population) shouldn't be included in the main equation either.

I continue to find this paper interesting and well-written, but I also continue to have serious doubts regarding this promise made in the paper: "The main contribution of this article is thus to provide a deeper understanding – also with a view towards identifying a causal effect – of the links between institutional and ecological effectiveness" and "Second, establishing causality between an institution's actions and actual environmental improvements remains a significant hurdle. This study addresses both challenges, making a significant contribution to our understanding of environmental organizations' overall performance".

Reviewer #2 (Remarks to the Author):

My concerns have been addressed, and I hereby recommend the article for publication as I believe it meets the expected standard. The insights provided by this paper are useful, and the author is transparent about the methodology.

Reviewer #3 (Remarks to the Author):

I have carefully reviewed the revised manuscript and find that all of my previous concerns have been adequately addressed. The revisions have significantly improved the clarity and quality of the work. I believe the manuscript now meets the criteria for publication, and I recommend it be accepted in its current form.

** Visit Nature Portfolio's author and reviewers' website at <http://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/>

REVIEWER 1

- *This paper is interesting and taps into long-standing debates regarding the role of international agreements in addressing critical environmental issues. The paper is well-written and clearly organized. Nevertheless, I have several concerns. The plausibility of the estimates having a causal interpretation, in particular, is unclear to me.*

Reply: I thank R1 for their positive assessment of my manuscript and the constructive feedback provided. Please see my detailed responses below, particularly where I outline that I strengthened the causality component of my work.

- *I would normally not comment on this, but there are several instances where the authors contextualize their study in a literature that is dated. To give a few examples, l40: I'm not sure a study by Zürn from nearly thirty years ago is indicative of current trends in the study of international relations. l54: is our understanding still incomplete in the same manner as it was described in a paper from 15 years ago? l56: is a study from 10 years ago still offering an overview of the state of the art today? It may well be that these are the latest references in this literature, but it would then be worth clarifying why so little progress has been made (which, in my own view, touches upon the concerns raised by the authors: getting causal quantitative estimates is very hard in this topic).*

Reply: In response, I have made several revisions in the paper. First, I dropped the reference to Zürn (1998). Second, while the literature has advanced since Gutner and Thompson (2010), and I refer to some studies of the corresponding literature in the sentence before, the fact remains that the complexities of *overall institutional performance* remain less well understood in current work. To avoid the impression that this assessment is based on a study from 15 years ago, however, I also cite Mitchell et al. (2020) or Goyes (2023) next to Gutner and Thompson (2010) or Jackson and Bührs (2015) as they arrive at the same conclusion. Third, the reference to the “state of the art” was not entirely accurate (in what was l56 of the previous version of the paper). I dropped this expression and now cite more thoroughly Mitchell et al. (2020), Jackson and Bührs (2015), Goyes (2023), and Lall (2017, 2021, 2023) when discussing the current literature. In sum, I hope that these revisions address R1’s valid concern about “little progress in the field” – as I show clearly now, the literature advanced over the last 10-15 years. Yet, the relationship between institutional and ecological effectiveness, both of which together constitute an organization’s overall effectiveness, has not been comprehensively explored. As I argue in the introduction, this is driven by the limitations I have identified in existing work – which I seek to address in this article.

- *The paper articulates a theory that contrasts “institutional effectiveness” and “ecological effectiveness.” This is interesting, but I am not clear as to where the boundaries of the two concepts are. The authors for instance cite: “Jackson and Bührs (2015: 65) not only emphasize the need to distinguish between institutional effectiveness—focused on implementation and compliance.” But the way institutional effectiveness was earlier defined (“the extent to which they [organizations] achieve their stated objectives”) does not match. The latter’s output is also environmental (or ecological) effectiveness (“stated objectives”). Note that this doesn’t undermine the authors’ main theoretical idea (implementation vs biophysical impact), but it makes the fit of this study in existing debates awkward.*

Reply: I realize that the definitions of the core concepts have not been clearly spelled out in the previous version. I thus revised the text by elaborating on the definitions used in Jackson and Bührs (2015) more effectively. In the end, I hope that conceptual clarity has now been achieved. While institutional effectiveness is indeed about achieving organizations’ stated objectives and, thus,

addresses implementation and compliance (I return to this as well in the Online Methods section), ecological effectiveness solely focuses on improvements in environmental quality (at the outcome level). The former does not necessarily translate into the latter as, for example, states could simply agree on targets that are easy to implement and comply with, but have little to no impact on actually improving environmental quality. Note that this lies at the heart of the selection problem in the literature, which I also discuss in more detail in the revised paper.

- On a side note, there is a large literature in organizational theory and management that, I think, connects to the authors' argument, and I suspect readers from that audience would find this paper interesting. It might make sense to cite this literature as well to put this paper on these readers' map.

Reply: I have added three studies on organizational theory and management (Lynn and Waters Robichau 2013; Krause et al. 2014; Tompkins 2023), which link my work to this literature.

- First, it would be helpful to the reader if the authors briefly discuss the six international organizations. As I arrive at Table 1, I have no idea what exactly I am looking at. I don't even know why PM2.5 is the relevant outcome variable. The SM is useful and addresses this in general, but I don't think it would hurt to add a bit more details in the core text.

Reply: The word limit prevents me from discussing these organizations in detail in the main text. However, I have added a reference to the SI where I introduce the six organizations, and I also list the sources I used for coding countries as donors/contributors and target/program states. In addition, in the Online Methods section, I elaborate in more detail that PM2.5 is used as a general proxy for ecological effectiveness (that is defined more clearly in the introduction now, thus also establishing the link to PM2.5 more thoroughly) and, hence, environmental quality on the ground.

- Second, it would also be good to describe institutional effectiveness. Sentences such as "the estimated coefficient of Institutional Effectiveness in Models 1-4 is -0.175, which translates into a reduction of PM2.5 by 0.175 units for every increase in an organization's institutional performance" are meaningless without knowing the units of the treatment.

Reply: Next to defining institutional effectiveness more clearly in the introduction, I have also extended the discussion of this variable in the Online Methods section. The most relevant sections read as follows (I have marked the clarification of the variable's units in **bold**):

"In more detail, based on the general definition of institutional effectiveness outlined in the main text, Lall (2017, 2021, 2023) develops a measure that incorporates the views of a variety of groups, supplements these with "objective" data, and is multi-dimensional in nature. First, the different views are incorporated as the DFID review considers the Multilateral Organization Performance Assessment Network survey, the Survey on Monitoring the Paris Declaration, the Heavily Indebted Poor Countries Capacity Building Project evaluations, as well as stakeholder consultations, workshops, interviews, and written submissions. Reports from the other stakeholders supplement this. Second, "objective" information is given by quantitative data from multiple sources, including the Quality of Official Development Assistance Assessment, the Publish What You Fund Aid Transparency Index, the Heavily Indebted Poor Countries Capacity Building Project, and the Common Performance Assessment System. Third, multi-dimensionality is taken into account as Lall (2017, 2021, 2023) focuses on institutions' achievements of stated objectives, their responsiveness to stakeholders,

and their cost-effectiveness. Against this background, Lall (2017, 2021, 2023) suggests coding six performance indicators: (1) delivery of results, (2) contribution to meeting the international community's objectives, (3) cost and value consciousness, (4) financial resources management, (5) accountability and transparency, and (6) strategic/performance management. The first two variables pertain to the first dimension (goal attainment), the third and fourth items stand for the second dimension (cost-effectiveness), and the last two variables capture the third dimension (responsiveness to diverse stakeholders).

While the codings in Lall (2017) are based on earlier reports like the 2011 DFID report, the DFID and other stakeholders issued updates of their reports, which Lall (2021, 2023) used to update his original data. Eventually, I use Lall's (2021, 2023) most recent data and merge the six performance indicators (Australian, Danish, MOPAN, Dutch, Swedish, and British) into a general institutional effectiveness variable using principal component analysis, as it is done in Lall (2017). **The final item is represented by scores on the first principal component in the analysis, and this is the institutional effectiveness variable used in the main text. It is continuously scaled and originally ranges in [-5.95; 3.65], with higher values standing for more institutional effectiveness.**"

- Third – and this is my main concern – the authors' main "selling" point is the notion that this analysis would generate causal estimates (though they are careful to qualify this statement on l.101). The argument is that the use of three models (fixed effects, ECM, and between-within) addresses problems of endogeneity and self-selection. These are the core issues that have plagued this literature: states self-select into environmental treaties that themselves are set at a bar that is achievable. Here, we would therefore have two worries: states self-select into one of the 6 agreements, and these agreements themselves are calibrated to achieve feasible goals. The effectiveness of a strategy like two-way fixed effect to address such problems is only guaranteed under very strict conditions that are not discussed in this paper. For instance, the FE model will capture causal effects in an unbiased manner if confounders are time-invariant and there is no heterogeneity in treatment effects. The plausibility of this statement is not discussed. And in fact, the model specification (including a third polynomial of GDP per capita – far from a conventional difference-in-difference model) does suggest that this analysis is model-based rather than drawing on exogenous source of variation in effectiveness. Nor do the FE address in and of themselves self-selection. As such, I don't see how the causal interpretation (however qualified by the author) is obtained via these three measurement strategies. The underlying problem remains. Contrast this for instance with pieces such as Lupu (2013, AJPS) which is not free of problem but tries to tackle self-selection.

Reply: I understand R1's concerns in this regard and, thus, have implemented the following changes. First, I now clarify in the Online Methods section how each estimation procedure contributes to identifying a causal effect. Note that the conclusion as stated in the main text still applies, however: "While each method has certain limitations and cannot fully resolve concerns over causal inference [...]." To this end, second, I now point to a range of additional estimation procedures in the SI, which address the problem of self-selection and the strategic design/use of international organizations more directly: the two-part model (2PM) as discussed in Vance and Ritter (2014) and a Heckman selection model. Third, I considered Lupu (2013, now cited in the manuscript as well). Eventually, I decided not to implement his matching approach (as it is based on observables only, which does not necessarily add to the empirical leverage over what I use), but to use one of the latest advancements in causal inference work: the fixed effects counterfactual estimator by Liu et al. (2024). The corresponding results are discussed in the SI. Fourth, though

not related to the issue of causal inference, I now spell out more clearly in the Online Methods section why there is a polynomial GDP per capita specification: this is driven by the literature on the Environmental Kuznets Curve, which argues for a curvilinear effect of income on environmental quality. By adding income to the power of 3, I allow for more flexibility in the functional form and the variable's impact. As shown by several models in the paper, however, leaving out this item (or any of the other controls) does not qualitatively affect the findings.

- Fourth, it was not clear to me why countries-years are the appropriate unit of analysis. If we are interested in the effectiveness of an institution, wouldn't the appropriate approach be to aggregate PM2.5 at the level of all member organizations? Aren't there really 6 observations? The theory appeared to be rooted in the international organization's structure. Alternatively, this is a story about domestic action, in which indeed, states are the unit of analysis. But then this raises all the questions asked previously about endogeneity as well as add questions such as domestic implementation of international rules, domestic state capacity, etc. and whether these interact (or mediate) international rules.

Reply: I understand R1's concern about the unit of analysis as there might be the risk of artificially increasing sample size if the underlying dynamics are exclusively situated at the institutional level. However, the outcome variable, ecological effectiveness, is not located there, nor are the mechanisms. That is, first, the ecological performance of institutions materializes at the country level as it pertains to improvements in environmental quality on the ground. Second, and derived from this, though the argument centers on international organizations' structures and their performance, whether and to what extent this has an impact only emerges not at the international level, but domestically. I fully agree with R1, though, that parts of this give rise to endogeneity issues as well as questions about domestic-level factors, since these may equally shape country-level environmental quality. As a result, I control for a broad array of domestic-level factors, which are all derived from the established environmental-politics literature. I have also followed R3's suggestion and now explore and control for the influence of domestic-level policymaking in the SI. What is more, I employ a large range of different estimation procedures, including causal estimation approaches and a placebo test (as suggested by R2) in the revised version of the manuscript, to rule out endogeneity problems, to control for self-selection dynamics, and to obtain a causal effect with greater confidence.

- Fifth, the authors write "It is also interesting that out of the 1,000 simulations, only 22 of them (2.1 percent) are positively signed when using the dyadic data structure." Eyeballing model 2, the t-statistic is about 6. A distribution centered at 6 will have virtually no negative value. Is the simulation then really surprising? Doesn't it mechanically relate to the regression estimates?"

Reply: The simulation exercise based on King et al. (2000; see also Rittman et al. 2025) is of course based on the regression estimates, but treats sample variability and uncertainty more seriously than coefficients alone. That said, given the *t*-statistics of the models, the outcome of only a small share of cases being positively signed is indeed not surprising. I thus revised the discussion of the simulation results and formulate the interpretation more carefully.

REVIEWER 2

- This paper contributes to the literature by providing empirical evidence on a previously understudied question. It is a well-written and well-executed paper. Although it does not present a perfect setting for causal inference, I generally find the author's conclusions convincing. It offers useful guidance to policymakers, particularly in terms of considerations around joining

international organizations. By shedding light on the relationship between organizational effectiveness and ecological outcomes, it informs both institutional design and strategic membership decisions.

Reply: I am also very grateful to R2 for their detailed comments and suggestions. In response, I have revised the article thoroughly in both the main text and the SI. R2, like the other referees and the editor, are concerned about the causality of my estimates and, thus, I spent a great amount of time addressing this valid point.

- The most common concern in studies examining the causal effects of membership in international organizations is the issue of self-selection. Countries that join environmental institutions, especially those with higher effectiveness scores, might be countries becoming more environmentally committed. The author addresses it in part through a Heckman selection model presented in the Supporting Information, which is much appreciated. I still wonder if it is possible to conduct placebo tests that examine the impact of institutional effectiveness on contributor countries. If self-selection is not a major concern, the estimated effects on contributors, who, according to the author, are not expected to benefit directly from the institutions' activities, should be close to zero.

Reply: This is an excellent suggestion, and I believe implementing it has considerably strengthened the causal component of my analysis. I have followed R2's instructions and the placebo test in the SI looks at the ecological performance in organizations' *contributing states*. As expected, when arguing for a causal effect on ecological effectiveness in target or program states, the estimates for contributing countries are indeed all indistinguishable from 0. I point to this placebo test in the main text, too.

- Also, the paper would benefit from more detailed summary statistics on the institutional effectiveness variable, particularly for the six organizations analyzed and their membership structures, in the Supporting Information. It would be helpful to understand whether these organizations target countries with similar geographic or developmental profiles. This information would give readers a clearer view of the sample and the sources of variation in the independent variable.

Reply: This mirrors the other referees' comments. In the main text and, in more detail in the SI, I thus discuss the six organizations and present summary statistics for their institutional effectiveness. I also list the sources I used for the membership (donors and target/program countries) of each organization in the SI. As shown there, donors or contributing members are predominantly higher income countries, while target/program states come from the Global South and are lower income.

- A further concern for me is the inclusion of certain organizations in the sample. Specifically, the Adaptation Fund, the Climate Investment Funds, and the Multilateral Fund for the Implementation of the Montreal Protocol are not primarily designed to reduce PM2.5 pollution. While it is plausible that some of their projects, such as the retirement of coal as a part of the decarbonization scheme, generate air quality co-benefits, they most likely focus on unrelated objectives, like adaptation infrastructure. So, my guess is that the institutional effectiveness of these international organizations should not have a large impact on PM2.5 of their members. It would be helpful to report results excluding these three organizations from the sample to assess whether the main findings hold when focusing only on other institutions with a more direct connection to air quality outcomes.

Reply: I use PM2.5 as a general proxy for ecological effectiveness, which is based on the rationale that all organizations should have – direct and indirect – effects on such a variable if their institutional effectiveness is indeed strongly pronounced. In the SI, I discuss three other measures, however. Particularly the consumption-based carbon footprint measure and the greenhouse gas emissions item are both general in nature and perhaps more suitable for the organizations I analyze. Having said that, I do see R2’s point and, thus, now present four models in the SI, which omit one of the three organizations at a time or all three jointly. As discussed in the SI, the results are qualitatively unchanged when excluding these three organizations from the sample.

- *For the reproducibility of the paper, the author should also be more explicit about how target and contributor countries are defined and assigned for each of the six international organizations. While some description is provided, it would help readers and future researchers if the rules used to classify countries into these categories were clearly stated, ideally with a short summary table or reference to an appendix with coding decisions.*

Reply: In the SI, I now introduce the six organizations I focus on in my sample, and I list the source links I use for compiling the target and contributing countries of each organization for the analysis.

REVIEWER 3

- *While the topic is relevant and the methodological rigor is commendable, there are significant concerns about the clarity of the concepts, the precision of the empirical claims, and the originality of the findings. The paper would benefit from a more focused theoretical framework, more precise terminology, and a clearer articulation of its contribution to the literature. I recommend major revisions before the manuscript can be considered for publication.*

Reply: The comments raised by R3 are extremely helpful and, by incorporating them, have significantly improved the manuscript in terms of a more focused theoretical framework, a more precise terminology, and a clearer articulation of the contribution to existing studies. I thank R3 for their time and efforts, please see my detailed responses to the specific issues below.

- *Conceptual Clarity: Ambiguity in Core Terms: The manuscript relies heavily on the distinction between institutional effectiveness and ecological effectiveness, yet these terms are not sufficiently defined early on. Suggestion: Provide clear, operational definitions of these concepts in the introduction, with concrete examples of what is included or excluded in each. If ecological effectiveness is proxied solely by PM2.5 improvement, consider using more precise terminology such as “air quality improvement” or “pollution reduction.”*

Reply: R1 raises a similar point. In response, I now define institutional effectiveness and ecological effectiveness early on and more clearly than it was the case in the previous version of the manuscript. Second, due to the fact that existing literature uses the term “ecological effectiveness,” I would prefer keeping it for my work to avoid any potential confusion. Having said that, I now clarify that PM2.5 is a proxy for ecological effectiveness, but (a) argue more convincingly why this general air pollution measure is suitable and (b) stress more thoroughly that the SI presents models with three alternative variables for ecological effectiveness – including measures on the consumption-based carbon footprint and greenhouse gas emissions, which are among the most general environmental quality indicators available.

- *Overuse of Generalized Terms: The term “ecological effectiveness” suggests broad environmental success, but the analysis focuses exclusively on PM2.5 reduction. Suggestion:*

Either expand the dependent variable to include multiple ecological indicators or revise the terms to avoid overstating the ecological scope.

Reply: Please see the reply above and the SI. In the latter, I summarize models using alternative dependent variables including the consumption-based carbon footprint and greenhouse gas emissions – environmental quality indicators that are also general in nature. In light of this, and also due to the fact that the existing literature works with the term “ecological effectiveness,” I prefer keeping it so to avoid any potential conceptual confusion.

- Research Motivation and Contribution: While the manuscript offers a rigorous empirical analysis linking institutional and ecological effectiveness, the core conclusion—that better-performing international environmental organizations are associated with improved environmental outcomes—is arguably intuitive and may align with widely held assumptions among scholars and policymakers. Many readers are likely to expect that reputable and well-established institutions would contribute positively to ecological outcomes, particularly in measurable domains like air quality. In this context, the current framing of the research question lacks a strong sense of novelty. The paper would benefit from clarifying: (1) what specific gap in the literature it seeks to address, and (2) what non-obvious insight is produced by the analysis. For example, if the novelty lies in identifying the causal magnitude of institutional effectiveness using new data, or in isolating the role of organizational design features, this should be made explicit early on—especially in the abstract and introduction.

Reply: This comment mirrors the next one. In response, I have revised the introduction to stress (1) what prior studies have not addressed, (2) how I seek to address this, and (3) what new, specific knowledge is gained. Especially with regards to the latter, although perhaps intuitive, I do not think that the core conclusion, i.e., better-performing international environmental organizations are associated with improved environmental outcomes, necessarily follows or aligns with widely held assumptions. Existing literature argues for a self-selection mechanism underlying membership in international institutions – states tend to select themselves into institutions if the benefits for doing so are high and the costs of participation are low. That is, hardly any obligations follow, laws and regulations can be circumvented, and commitments of organizations are soft, not monitored, and not enforced. I spell out this contribution more clearly in the revised introduction, next to fact that I add to previous studies by identifying the causal magnitude of institutional effectiveness using recently compiled data (and, although minor in nature, by discussing the role of organizational design features).

- Moreover, the motivation for the study remains somewhat abstract. It is not clear what problem or empirical ambiguity this research is intended to resolve. If existing studies have assumed, but not empirically demonstrated, that institutional effectiveness drives air quality improvements, this assumption should be problematized more clearly. If the study contributes by confirming this link through more robust econometric modeling and high-quality data, then this contribution should be framed as a specific and novel advance rather than a validation of conventional wisdom. Suggestion: Strengthen the motivation and contribution sections by clearly articulating (1) what prior studies have missed or inadequately addressed, (2) how this study fills that gap, and (3) what new, specific knowledge is gained beyond the affirmation of a presumed positive relationship between institutional and ecological effectiveness.

Reply: The contribution is indeed what R3 describes. On one hand, existing studies have assumed, but not empirically demonstrated, that institutional effectiveness is related to ecological effectiveness, i.e., improvements in environmental outcome measures such as air quality. On the other hand, derived from the first point, I seek to contribute to the literature by confirming this link

through more robust econometric modeling and recently compiled data (in particular, the data by Lall). I understand that this was not well presented in the previous version of the manuscript and, thus, I revised the introduction to stress (1) what prior studies have not addressed (no systematic empirical analysis), (2) how I seek to address this gap (recently compiled data is analyzed in a rigorous empirical approach that helps addressing causality and sample-selection issues), and (3) what new, specific knowledge is gained (institutional effectiveness facilitates ecological effectiveness, and this is particularly important as the relationship between institutional and ecological effectiveness may well be negative or only weakly pronounced due to the underlying selection problem (see above)).

- Data validity and Policy Timing Bias in Measuring Ecological Effectiveness: While the manuscript aims to provide a robust empirical link between institutional and ecological effectiveness, it relies exclusively on modeled PM2.5 concentration data derived from satellite observations and chemical transport models (van Donkelaar et al., 2021). Although such data offer broad spatial coverage and address the issue of uneven in-situ monitoring, they carry notable limitations in accuracy. The authors do not discuss the uncertainty inherent in these modeled estimates, nor how such uncertainties may affect the measurement of ecological effectiveness or bias the inferred institutional impact. A discussion of error margins, data validation (e.g., cross-referencing with ground-based data), and sensitivity analyses using alternative pollution metrics would significantly enhance the credibility of the findings.

Reply: I agree with R3 on this assessment, but would like to stress that the data set from van Donkelaar et al. (2021) is one of the general sources on environmental quality available. That said, I do discuss the limitations of this data set in more detail in the revised Online Methods section and I explain that PM2.5 is used as a general proxy for ecological effectiveness (that is defined more clearly in the introduction now, thus establishing the link to PM2.5 more thoroughly). After discussing the weaknesses of the PM2.5 data, I conclude that there is “evidence that global-monitored locations tend to be in cleaner regions [...], with large measurement gaps in the Global South. Uncertainty estimates exhibit regional consistency with observed differences between ground-based and satellite derived PM2.5. The evaluation of uncertainty for agglomerated values indicates that hybrid PM2.5 estimates provide precise regional scale representation, with residual uncertainty inversely proportional to the sample size.” I also stress more the use of alternative pollution metrics in the SI – most notably the environmental-footprint indicator and the greenhouse gas variable.

- Furthermore, the analysis does not adequately control for the heterogeneous timing and intensity of national air quality policies, which may be independently responsible for PM2.5 reductions. For instance, many east Asian countries implemented aggressive air pollution control measures at different time points during the study period. These external policy shocks may confound the observed relationship between institutional effectiveness and pollution levels, especially if countries with high institutional performance are also those undertaking strong domestic reforms. Relying solely on country and year fixed effects may not be sufficient to disentangle these effects. Suggestion: The authors should either (1) incorporate additional controls to account for national environmental policy timelines and implementation intensity, or (2) clearly acknowledge this as a limitation that could undermine the causal interpretation of the results. Additionally, a detailed discussion on the validity and uncertainty of the PM2.5 data used as the dependent variable is essential for evaluating the robustness of the ecological effectiveness measure.

Reply: These are excellent suggestions. In response, I have implemented two major changes. First, I fully agree with R3 that domestic-level forces, including national air quality policies, may be of importance and, thus far, have not been considered adequately in my analysis. In the SI, I now

discuss two models that incorporate domestic policy measures from the Climate Policy Database. The results are robust. Second, I have added a discussion on the validity and uncertainty of the PM2.5 data to the Online Methods section, while stressing that the SI also summarizes robustness checks using other environmental quality items, including the general consumption-based carbon footprint and greenhouse gas emissions.

- [Line 1–3] *The term “institutionally effective” in the title and abstract may be unclear to readers unfamiliar with the literature. Consider rephrasing or briefly defining it early in the abstract.*

Reply: “Institutional effectiveness” is indeed the term used in the literature, e.g., see Jackson and Bührs (2015) and I am thus reluctant to rephrase that. However, I have revised the abstract and the introduction so that I define institutional and ecological effectiveness clearly there.

- [Line 7–8] *The phrase “organizations’ performance” is overly general. It would be helpful to briefly specify what aspects of performance are being evaluated—e.g., goal achievement, transparency, cost-effectiveness—as this would enhance the credibility and informativeness of the abstract.*

Reply: In the previous version of the abstract, “organizational performance” in lines 7-8 referred to both institutional effectiveness and ecological effectiveness. I realize that this was potentially confusing and, thus, revised the abstract and dropped the term. In addition, I have revised the introduction to clearly define institutional and ecological effectiveness, also by highlighting what aspects of performance are being evaluated. Please see the Online Methods section in addition to that, where I elaborate on the variables’ operationalizations more comprehensively.

- [Line 9] *The term “ecological effectiveness” appears to refer exclusively to reductions in PM2.5 concentrations. If that is the only environmental indicator used in the study, the expression may be too broad and risk misleading readers. Consider specifying that the ecological effectiveness assessed here refers to air quality improvements (PM2.5 reduction).*

Reply: Due to the fact that existing literature works with the term “ecological effectiveness,” I would prefer keeping it for my work to avoid any potential confusion. Having said that, I now argue more convincingly why PM2.5 as a general air pollution measure is suitable as a proxy for ecological effectiveness and, in addition, stress more thoroughly that the SI presents models with three alternative variables for ecological effectiveness – including a measure on the consumption-based carbon footprint and a variable on greenhouse gas emissions, which are among the most general environmental quality indicators available.

- [Line 41] *I don’t understand why their effectiveness is the dominant force in international relations analysis. Please add an explanation to supplement this argument.*

Reply: As R1 noted as well, this formulation and the corresponding reference were unnecessary and, thus, I have removed both from the revised version of the paper.

- [Line 48–49] *What kind of problem structure and organizational design? It would be helpful if you could give a clear example of one.*

Reply: I have added examples to illustrate what problem structure and organizational design are. For additional information on these concepts, please see the literature (e.g., Hovi et al. 2003; Andresen 2013; Breitmeier et al. 2011; Grundig and Ward 2015; Böhmelt and Pilster 2011; Mitchell et al. 2020; Kim et al. 2017, 2020).

- [Line 59–70] *While the conceptual distinction between institutional and ecological effectiveness is theoretically grounded, its necessity and clarity could be better justified in the introduction. For readers unfamiliar with the framework, the rationale for separating these dimensions should be more explicitly developed, perhaps with real-world examples to illustrate their divergence.*

Reply: R1 raises a similar point and this comment mirrors one of R3’s suggestions above. In response, I now define institutional effectiveness and ecological effectiveness early on and more clearly than it was the case in the previous version of the manuscript. I also illustrate the two concepts using examples.

- [Line 81] *It is necessary to explain what factors are needed to evaluate institutional and ecological effectiveness. Even in the aforementioned previous studies, it is unclear what data was used to produce the results. However, when they suddenly say: “high-quality data,” I don't understand what they mean.*

Reply: On one hand, also in light of the other referees’ comments, I now define institutional and ecological effectiveness precisely in the introduction and provide examples. This hopefully clarifies what elements are needed to empirically capture institutional and ecological effectiveness. On the other hand, I now explain in detail in the introduction that previous research simply does not offer a systematic empirical analysis of institutional and ecological effectiveness – existing accounts are rather of a theoretical nature, which is likely driven by a lack of data thus far. I have decided to drop the reference to “high-quality” as I understand this was potentially confusing.

- [Line 171–173] *The robustness of the results is notable; however, the consistently strong associations raise questions about whether potential confounding factors or model overfitting have been fully ruled out. Further discussion on model limitations or sensitivity tests in appendix would strengthen confidence in the findings.*

Reply: In response, I have strengthened the discussion of the additional analysis in the SI: what analyses are discussed and what underlying rationale they have. Please see the last paragraph before the discussion section in the main text on that (“In the SI, I present numerous additional analyses, which, among others, discuss potential confounding factors, provide sensitivity tests, employ alternative estimation procedures, and rule out overfitting of the models. I also explicitly control for countries’ self-selection into organizations (see Lupu 2013; Spilker and Böhmelt 2013; Mitchell et al. 2020) using a two-part model and a Heckman selection model. Furthermore, I implement a newly developed counterfactual model and a placebo test. The results in the SI continue to support the main findings presented above, and they lend further support to the causal interpretation of the findings”). Second, in the Online Methods section, I now discuss the limitations of the analysis and, especially, the main estimation procedures used in the main text.

- [Line 190–191] *I am not sure if this factor is important in considering the effects within a country. How are income, population, democracy, and globalization related to institutional effectiveness?*

Reply: As discussed in the main text, including the items for income, population, democracy, and globalization does not affect the core finding, as demonstrated in Table 1. What is more, they are also unlikely drivers of institutional effectiveness. According to the literature (e.g., Miles et al. 2002; Gutner and Thompson 2010; Young 2011; see also Helm and Sprinz 2000; Böhmelt and Pilster 2010), institutional effectiveness is primarily shaped by an organization’s problem structure (e.g., whether an environmental issue is characterized as a public good or a common pool resource) and organizational design (e.g., whether organizations’ terms are precise and legally binding, and

whether mechanisms are in place that allow to monitor countries' behavior and enforce agreement terms). The main reason for considering the controls is, therefore, not institutional effectiveness, but ecological effectiveness: the environmental quality of a country is affected by both domestic and country-level characteristics (i.e., precisely the controls I consider) as well as international forces – institutional effectiveness.

REVIEWER 1

- Thank you for engaging with my review. I found the answers and edits by and large convincing. This being said, I remain skeptical regarding the ability to make causal claims from these estimates. First, the fixed effects aren't helping self-selection bias based on differential treatment effects.

Reply: I agree that fixed effects alone cannot address self-selection if heterogeneous treatment effects are correlated with treatment assignment. The estimates should therefore be interpreted as causal only under the assumption that, conditional on fixed effects and observables, selection is uncorrelated with potential outcomes. As Angrist and Pischke (2009) note, fixed effects can still deliver consistent estimates under this assumption, though I acknowledge the limitation now more thoroughly in the Methods section, and I complement the analysis with robustness checks in the main text and the SI, including the placebo test (Table 9 in the SI) and the counterfactual estimator for causal inference (Table A10 in the SI).

- Second, the Heckman selection models (Table 8, SI) too, raise questions. They rely on very strong assumptions regarding the exclusion restriction on the instrument used in the selection stage, which are here: democracy and population (and a cubic time trend). Given that you articulate the notion that the story in this paper is a domestic politics one (in response to my question about the unit of analysis), and given that we know that there exists a large literature on the effects of variables such as democratic institutions on environmental outcomes, it seems difficult to believe that these variables used in Table 8, SI (democracy, population) shouldn't be included in the main equation either.

Reply: R1 is correct in that the Heckman model has an exclusion restriction. I now spell this out more clearly on p. 10 in the SI. In addition, the justification for excluding democracy and population in the outcome stage is now more comprehensive. In essence, I elaborate that these items seemed to perform best at the selection stage (empirically) and, at least in terms of democracy, are theoretically more likely to shape selection into organizations than ecological effectiveness. That is, both variables appeared largely unrelated to or are inconsistently associated with the dependent variable in the outcome stage in various model specifications I explored. Moreover, “theoretically,” the existing literature offers mixed evidence on the overall effect of democracy on environmental quality outcomes (i.e., ecological effectiveness), but findings are more consistent with regard to the role of regime type in shaping environmental policy outputs at the domestic or at the international level (and this relates to organizational selection processes in the form of participation in or ratification of international environmental agreements). I also include references to support the theoretical claims.

- I continue to find this paper interesting and well-written, but I also continue to have serious doubts regarding this promise made in the paper: “The main contribution of this article is thus to provide a deeper understanding – also with a view towards identifying a causal effect – of the links between institutional and ecological effectiveness” and “Second, establishing causality between an institution’s actions and actual environmental improvements remains a significant hurdle. This study addresses both challenges, making a significant contribution to our understanding of environmental organizations’ overall performance.”

Reply: Following the editor’s advice, I now avoid overstating causal findings. To this end, for example, I now refer to “likely causal effects” (and have re-written the second sentence R1 refers to as: “This study seeks to address both challenges by providing an empirical analysis that produces likely causal estimates, thus aiming at facilitating our understanding of environmental

organizations' overall performance and the conditions under which they can drive meaningful environmental change”).

REVIEWER 2

- My concerns have been addressed, and I hereby recommend the article for publication as I believe it meets the expected standard. The insights provided by this paper are useful, and the author is transparent about the methodology.

Reply: I am very grateful to R2 for the comments and suggestions they have provided, which prompted me to improve the paper.

REVIEWER 3

- I have carefully reviewed the revised manuscript and find that all of my previous concerns have been adequately addressed. The revisions have significantly improved the clarity and quality of the work. I believe the manuscript now meets the criteria for publication, and I recommend it be accepted in its current form.

Reply: I also thank R3 for their feedback and positive assessment of the revised manuscript.

Title: *Are Institutionally Effective Environmental Organizations Also Ecologically Effective?*

This manuscript explores the relationship between institutional effectiveness and ecological effectiveness in international environmental organizations. Using panel data on six organizations from 2008 to 2018 and employing three econometric models, the study identifies a robust, likely causal relationship between the two dimensions, as measured by reductions in PM_{2.5} pollution. While the topic is relevant and the methodological rigor is commendable, there are significant concerns about the clarity of the concepts, the precision of the empirical claims, and the originality of the findings. The paper would benefit from a more focused theoretical framework, more precise terminology, and a clearer articulation of its contribution to the literature.

I recommend **major revisions** before the manuscript can be considered for publication.

Major Concerns and Suggestions

1. Conceptual Clarity: Ambiguity in Core Terms

The manuscript relies heavily on the distinction between *institutional effectiveness* and *ecological effectiveness*, yet these terms are not sufficiently defined early on.

Suggestion: Provide clear, operational definitions of these concepts in the introduction, with concrete examples of what is included or excluded in each. If ecological effectiveness is proxied solely by PM_{2.5} improvement, consider using more precise terminology such as “air quality improvement” or “pollution reduction.”

2. Overuse of Generalized Terms

The term “ecological effectiveness” suggests broad environmental success, but the analysis focuses exclusively on PM_{2.5} reduction.

Suggestion: Either expand the dependent variable to include multiple ecological indicators or revise the terms to avoid overstating the ecological scope.

3. Research Motivation and Contribution

While the manuscript offers a rigorous empirical analysis linking institutional and ecological effectiveness, the core conclusion—that better-performing international environmental organizations are associated with improved environmental outcomes—is arguably intuitive and may align with widely held assumptions among scholars and policymakers. Many readers are likely to expect that reputable and well-established institutions would contribute positively to ecological outcomes, particularly in measurable domains like air quality.

In this context, the current framing of the research question lacks a strong sense of novelty. The paper would benefit from clarifying: (1) what specific gap in the literature it seeks to address, and (2) what non-obvious insight is produced by the analysis. For example, if the novelty lies in identifying the causal magnitude of institutional effectiveness using new data, or in isolating the role of organizational design features, this should be made explicit early on—especially in the abstract and introduction.

Moreover, the motivation for the study remains somewhat abstract. It is not clear what problem or empirical ambiguity this research is intended to resolve. If existing studies have assumed, but not empirically demonstrated, that institutional effectiveness drives air quality improvements, this assumption should be problematized more clearly. If the study contributes by confirming this link through more robust econometric modeling and high-quality data, then this contribution should be framed as a specific and novel advance rather than a validation of conventional wisdom.

Suggestion: Strengthen the motivation and contribution sections by clearly articulating (1) what prior studies have missed or inadequately addressed, (2) how this study fills that gap, and (3) what new, specific knowledge is gained beyond the affirmation of a presumed positive relationship between institutional and ecological effectiveness.

4. Data validity and Policy Timing Bias in Measuring Ecological Effectiveness

While the manuscript aims to provide a robust empirical link between institutional and ecological effectiveness, it relies exclusively on modeled PM_{2.5} concentration data derived from satellite observations and chemical transport models (van Donkelaar et al., 2021). Although such data offer broad spatial coverage and address the issue of uneven in-situ monitoring, they carry notable limitations in accuracy. The authors do not discuss the uncertainty inherent in these modeled estimates, nor how such uncertainties may affect the measurement of ecological effectiveness or bias the inferred institutional impact. A discussion of error margins, data validation (e.g., cross-referencing with ground-based data), and sensitivity analyses using alternative pollution metrics would significantly enhance the credibility of the findings.

Furthermore, the analysis does not adequately control for the heterogeneous timing and intensity of national air quality policies, which may be independently responsible for PM_{2.5} reductions. For instance, many east Asian countries implemented aggressive air pollution control measures at different time points during the study period. These external policy shocks may confound the observed relationship between institutional effectiveness and pollution levels, especially if countries with high institutional performance are also those undertaking strong domestic reforms. Relying solely on country and year fixed effects may not be sufficient to disentangle these effects.

Suggestion: The authors should either (1) incorporate additional controls to account for national environmental policy timelines and implementation intensity, or (2) clearly acknowledge this as a limitation that could undermine the causal interpretation of the results. Additionally, a detailed discussion on the validity and uncertainty of the PM_{2.5} data used as the dependent variable is essential for evaluating the robustness of the ecological effectiveness measure.

Minor Comments

[Line 1–3] The term “institutionally effective” in the title and abstract may be unclear to readers unfamiliar with the literature. Consider rephrasing or briefly defining it early in the abstract.

[Line 7–8] The phrase “organizations’ performance” is overly general. It would be helpful to briefly specify what aspects of performance are being evaluated—e.g., goal achievement, transparency, cost-effectiveness—as this would enhance the credibility and informativeness of the abstract.

[Line 9] The term “ecological effectiveness” appears to refer exclusively to reductions in PM_{2.5} concentrations. If that is the only environmental indicator used in the study, the expression may be too broad and risk misleading readers. Consider specifying that the ecological effectiveness assessed here refers to air quality improvements (PM_{2.5} reduction).

[Line 41] I don't understand why their effectiveness is the dominant force in international relations analysis. Please add an explanation to supplement this argument.

[Line 48–49] What kind of problem structure and organizational design? It would be helpful if you could give a clear example of one.

[Line 59–70] While the conceptual distinction between institutional and ecological effectiveness is theoretically grounded, its necessity and clarity could be better justified in the introduction. For readers unfamiliar with the framework, the rationale for separating these dimensions should be more explicitly developed, perhaps with real-world examples to illustrate their divergence.

[Line 81] It is necessary to explain what factors are needed to evaluate institutional and ecological effectiveness. Even in the aforementioned previous studies, it is unclear what data was used to produce the results. However, when they suddenly say “high-quality data,” I don't understand what they mean.

[Line 171–173] The robustness of the results is notable; however, the consistently strong associations raise questions about whether potential confounding factors or model overfitting have been fully ruled out. Further discussion on model limitations or sensitivity tests in appendix would strengthen confidence in the findings.

[Line 190–191] I am not sure if this factor is important in considering the effects within a country. How are income, population, democracy, and globalization related to institutional effectiveness?