**Revision of # SDATA-24-02534 (A dataset of structural breaks in greenhouse gas emissions for climate policy evaluation)**

We would like to thank the editor and reviewers for their thoughtful comments, which have contributed to significantly improving the paper and making it accessible to a broader audience. Below you can find a description of how we addressed the comments of the editor and the reviewers in the revised version of the manuscript.

We hope you are satisfied with the revised version of the paper and look forward to hearing from you again

**Editor comments**

|  |  |
| --- | --- |
| **Editor Comments** | **Responses** |
| Because at least some of your data looks to originate from third parties, please check the following and confirm/state what has been done in your ‘response to reviewers’ document: All sources are **clearly described** in the main text. For general mentions of services, databases, or other platforms, please name the resource in the main text and provide a **URL** in ()s. Instructions should be sufficient for the **exact input data** to be retrieved, so please provide specific links, accession numbers, or the exact search term/query when discussing any data from general databases. For specific datasets downloaded from repositories, or other items with formal metadata, please use a full citation in the reference list containing the relevant **DOI**. Note this should be the dataset, rather than any **relevant publication**, however these should be cited as well if required, or if the data was newly extracted from a document. | Additional detail has been added to the description of the three underlying datasets and URL links to all datasets have been included in the text. Indicator codes have also been added to ensure the exact indicators used are clear to the reader. URLs have also been incorporated to the references for ease of access. |
| 2) Confirm all the sources are **openly available** - i.e. your re-use or re-distribution is compliant with the terms and conditions/licenses for data sharing. If you cannot see a fully open licence at the source please check with the data provider. Please note it is your legal responsibility to ensure all data have been used or re-distributed appropriately and we cannot support publication of descriptions of data obtained or re-shared without this check | All input sources have been verified as being openly accessible to the public. All underlying data have also been provided in the Zenodo repository accompanying the database. |
| Please add a **data citation** for the dataset to the reference list using these instructions (<https://www.nature.com/sdata/publish/submission-guidelines>#data\_citations - note that a DOI URL should be used). Please add the **reference number** to wherever the dataset is mentioned in the text - the main position should be the **first part of the Data Record in a sentence describing where the data has been deposited**. Also **cite this in your Code Availability** section. | A data citation has been added to the reference list according to the citation instructions, including a DOI URL. The citation is referenced throughout the report including in the Data Record and Code Availability sections. |

**Reviewer #2 (Remarks to the Authors (required)):**

|  |  |
| --- | --- |
| **Reviewer Comments** | **Responses** |
| This paper contributes a new dataset to identifying structural breaks in greenhouse gas emissions. However, this data might be a bit more confusing for scholars and researchers in other disciplines.  1. The content of **'structural break**' is the key to this work and in data. The authors need **more explanations**, background descriptions, and instructions to help readers understand the data. | In the revised version of the paper, we have significantly expanded the explanation of structural breaks in the context of the model specification used in the statistical analysis. We have expanded the verbal explanation of structural breaks in the *Study Overview* section, including an example of a break as a country-year combination (Austria-2010) for a given gas and sector. In addition, we are also now more precise on the definition of the structural break when presenting the model that is estimated in section *Identifying Structural Breaks in Emissions Data*. |
| 2. **Data methods need a more detailed and comprehensive description**. At present, I do not understand the data process and algorithm for generating the data. | In addition to presenting the method in more detail, in the revised version of the manuscript we include a paragraph that summarizes the nature of the data input and output at the end of the section *Identifying Structural Breaks in Emissions Data*. We also added more detail on the choice of sample groups, and a comparison of the approach to standard difference-in-differences estimation in the “Study Overview” sub-section. |
| 3. **Need more deep validation**. **Statistical significance testing** is essential for data validation. Whereas, significance itself is convinced for data technical validation. Authors should provide comprehensive data validation. | The method used to obtain the estimates of the structural breaks assesses both model uncertainty and the statistical significance of individual structural breaks, and as such ensures a high degree of robustness in the estimates. We have included more details on the way in which the method guarantees the empirical robustness of the structural break estimates in the data validation section. |
| 4. The last and most important, please provide a **data usage instruction** in this paper. | A section titled “Usage notes” has been added with greater detail on how the structural break dataset can be used as the basis for reverse causal analysis of climate policies across all sectors and all major GHGs. Examples have been provided of other studies that have used this approach in the past. These provide a basis for similar analysis with a different coverage of countries, sectors or gases. Further, suggestions are provided on areas of the literature that have not yet been explored, such as identifying role model countries, best policies or the analysis of positive structural breaks, all of which the dataset can serve as a foundation. |

**Reviewer #3 (Remarks to the Authors (required)):**

|  |  |
| --- | --- |
| **Reviewer Comments** | **Responses** |
| In terms of general comments, one of the disadvantages of the structural break approach is that it overlooks more gradual changes that accumulate over many years. One can see this for example in Figure 1, where peak power sector emissions for the UK was already reached in the mid-2000s – but a structural break was only detected approximately one decade later. Therefore it might be important to clarify early on in the text what kind of potential policy effects are discovered with this approach, and what kinds are not. | The reviewer is right to highlight this limitation of the method. We have added a paragraph at the end of the *“Study Overview”* sub-section which comments on this issue. Given the conservative setting we employ in the GETS method, our results will tend to err on the side of not finding small breaks or slow, gradual shifts in emissions. Given the potential uses of the dataset, minimizing such false-positives appears to be the most reasonable setting for the modelling exercise. |
| Line 55-56: I suppose the authors are now aware that a new study does exactly this (Stechemesser et al; <https://doi.org/10.1126/science.adl6547>), which should probably be cited. I think it would be important to also compare the methodologies used here and in Stechemesser et al., as their paper also includes a dataset of structural breaks using EDGAR! | We have included the reference to Stechemesser et al., and note that our dataset has a much higher degree of granularity at the sectoral level, which is expected to allow for more detailed analysis in the future. In addition, the scope of GHG coverage is much broader in our study. |
| Line 67-68: The important information here is the time horizon for calculating CO2e. I assume it is “global warming potentials with a 100-year time horizon (i.e. GWP100) from the IPCC 5th Assessment Report”. If so please write that here. | This has been clarified in the text and is indeed GWP100 from the AR5 report. |
| Line 68-69: Which World Bank indicator for GDP is used? | This is GDP in constant 2015 US dollars (code: NY.GDP.MKTP.KD), which has now been specified in the text. |
| Line 71-75: If only (1) GHG emissions (2) population and (3) GDP is needed, one could in principle calculate structural breaks for almost every country – and not just the OECD. Is there a particular reason not to? And should that be stated here? | The EU15 and OECD sample groups were chosen to maintain homogeneity within the sample groups, noting the fact that the impact of GDP, population and policy changes on emissions may differ between developed and developing countries. Within the EU15 sample, EU-wide regulations mean that the countries are subject to similar regulatory environments, while the OECD sample of developed countries reflects the fact that these countries have a greater access to decarbonization technologies than developing ones. A new paragraph has been added to the text to explain this choice of sample groups in more detail. |
| Line 91: Looking here and ahead, it seems that the analysis covers the highest level of sector detail in EDGAR. I wonder if this makes sense. Many of these sector categories are highly specific (e.g. Urea application) or are essentially residual categories included in the EDGAR inventory for completeness (Indirect N2O, fossil fuel fires). The question is whether users of this data might benefit from a pre-selection or aggregation of sectors to some meaningful minimum. For example, why not aggregate the overall waste sector? Also, without including the EDGAR/IPCC codes, users basically don’t know the aggregate sector in which “Non-specified” sits (it’s the energy sector). Nor would it be obvious to many where the energy sector stops and the process emissions sector starts. | The reviewer’s comment is correct. The sectors in the analysis are at the highest level of detail as provided in the EDGAR database, which was the level chosen to provide users with the richest data possible on which to conduct further analyses. We have also added the IPCC code based on the 2006 IPCC Guidelines for National Greenhouse Gas Inventories to the dataset so users can more easily identify in which aggregated sector a given category falls. We have added a further column aggregating the categories to a more intuitive level of the IPCC categorization, using the following five groups: Energy; Industrial processes and product use (IPPU); Agriculture, forestry, and other land use; Waste; and Other. |
| Figure 1: is the interpretation from figure 1 that each structural break counterfactual is associated with a “start” and an “end”? If so, shouldn’t the end year also be listed in the database? | Counterfactuals are not associated with a start and end date. Rather, a structural break represents a persistent step-change in emissions, not accounted for by the main determinants of emissions (GDP and population). As such, there is no end date. The red counterfactual lines are an illustration of emissions in the absence of a break for a period of 5 years after the identified break date. This has been clarified in the Figure 1 description and is clarified by the revised explanation of structural breaks in the main text. |
| Table 4: the units here differ across gases, contrary to the earlier statement that emissions are in CO2e. Which is it? | The units of this table are number of breaks per gas, not any units of emissions. The chemical symbol for the different gases are displayed in brackets after each gas, and this has been described in the column header. |
| Database: please also provide the **country ISO codes** as a column. This makes things a lot easier for users! | 2-digit ISO codes have been added for all countries and has been described in the text. |