Profiling Scientific Influences on Climate Policy: A Case Study on IPCC and Country Submissions to UNFCCC

Anonymous ACL submission

Abstract

Scientists produce a lot of information on climate change, and policymakers craft plenty of documents on measures to mitigate and adapt to climate change. A key question is whether or not, and how much of the former influences the latter. This work takes a first step towards profiling scientific influences on climate policy. We present a case study that extracts the mentions and quotes of IPCC, the intergovernmental scientific body, among 1.1M paragraphs of policy text from 198 countries to the United Nations. We use three different methods: counting mentions, modeling topics, and finding quotes. We observe that 80% of the documents and 2% paragraphs mention IPCC. Such mentions can be categorized into six broad topical groups, five of which concerns the measurement and sector-specific Greenhouse Gas (GHG) Emissions, and one on Climate Change Scenarios and Impacts. Upon further examining the phrases that mention IPCC, we found that mentions disproportionally focus on the IPCC GHG guidelines. We hope that this study serves as a first step towards profiling the intersection between science and policy and generate valuable discussions in the NLP and Climate research community.

1 Introduction

Addressing climate change requires coordinated efforts across diverse stakeholders, including scientists, policymakers, activists, and the public. In particular, it is important for the scientific and policy-making community to be aligned to ensure that scientific findings have a direct and strong influence on policy actions.

Organisations like the Intergovernmental Panel on Climate Change (IPCC) was specifically created for this purpose - "to provide policymakers with regular scientific assessments on climate change, its implications and potential future risks, as well as put forward adaptation and mitigation options" (IPCC). The IPCC's work includes comprehensive Assessment Reports, which synthesise the available scientific basis on climate change, and technical guidelines to assist policymakers.

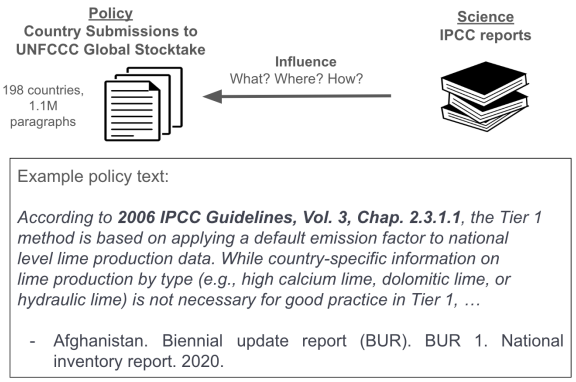


Figure 1: We profile the mentions of science (IPCC reports, on the right) on country-level policy (UNFCCC submissions, on the left) by counting mentions, modeling topics and finding quotes. An example mention is shown at the bottom of this figure.

The UNFCCC is the UN entity responsible for supporting the global response to the threat of climate change. The UNFCCC coordinates the Global Stocktake (GST), which evaluates the world's progress on climate action and support, and identify gaps and ultimately informs the next round of climate action plans under the Paris Agreement. The first ever stocktake concluded in late 2023 and the aim is for the GST to occur once every 5 years. As part of the GST, each country takes stock of their climate policies and commitments, and assess progress towards the Paris Agreement goals. The stocktake[1] also covers gathering the relevant scientific and technical output from key sources such as the IPCC.

Natural language processing and machine

[1] https://unfccc.int/topics/global-stocktake

learning methods have recently been used to analyze datasets of climate science (Callaghan et al., 2021), policy (Sietsma et al., 2023) and to create chatbot applications (Vaghefi et al., 2023). However, quantifying the influence of science on policy remains an open question, e.g., to what extent does science influence policy, and in what context are key scientific outputs being discussed in a policy domain?

In this study, we take the first step towards computationally quantifying the influence of IPCC on policy using natural language processing. To measure country-level policy actions and considerations, we use their submissions to the UNFCCC Global Stocktake (GST) as discussed above.

We use three types of techniques. First, we counted the number of 'IPCC' mentions in the policy documents and observe that it is more prevalent in documents that discuss Greenhouse Gas (GHG) inventories. We then apply topic modelling to understand the context in which the IPCC is being discussed. We find that the topics can be broadly clustered to GHG Emissions Measurement and Methodology, Climate Change Scenarios and Impacts, GHG Emissions by key sectors. Further analysis via quote-finding confirms our observation that GHG inventories are the main context for IPCC mentions, even though GHG inventories only consists of one component of the IPCC's output.

This work represents a first step towards generating a data-driven proxy measure for how IPCC's output influences policy-making. We believe these methods and results paves the way for future work on methodology and the broader science-policy discussion among different stakeholders.

2 Related work

On the science side, the IPCC reports have been analysed and leveraged in a range of natural language processing research. One area of research focuses on extracting insights from the reports - through LLM-based summarisation (Vaghefi et al., 2023), scientific evidence (Lacombe et al., 2023) or sentiment analysis (Barkemeyer et al., 2016). For example, chatClimate is a climate-focused LLM that was created by integrating the IPCC's Sixth Assessment Report into GPT-4(Vaghefi et al., 2023). Another area of work leverages the text from the IPCC reports to support downstream NLP analysis - such as on climate change scientific literature (Callaghan et al., 2020) (Callaghan et al., 2021) or climate-related development goals (Zhou et al., 2022).

On the policy side, there has been increased interest in recent years to compile and annotate datasets on policy text related to climate change. The annotations include tagging the policies by sector and mitigation area (Nascimento et al., 2022) or rating country's progress towards climate targets (Climate Action Tracker). One tool in particular facilitates the full text search of global climate change-related policies and laws (Climate Policy Radar).

Natural language processing techniques can be applied to climate policy texts to provide valuable insights into the themes, trends, and priorities within the documents. The types of climate change policy texts that have been analysed include country-level policies and country's reporting to the UNFCCC. For country-specific policies, Żółkowski et al. (2022) creates a topic modelling and clustering pipeline to automatically summarise and compare climate policies across the EU. In regards to the country's reporting to the UNFCCC, reports such as country's National Communications and Nationally Determined Contributions (NDCs) have been analysed using topic modelling (Biesbroek et al., 2022; Hsu et al., 2021; Wright et al., 2023).

A limited number of studies have emerged that analyse the totality of the country's submissions to the UNFCCC Global Stocktake in 2023. This includes a broad range of different document types, in addition to National Communications and NDCs as mentioned above. One notable study that does this applies BERTtopic to surface the themes covered in the country submissions, and compares them to the themes covered in the Synthesis Report, which is a summary of the country submissions that the UNFCCC is tasked to compile (Sietsma et al., 2023).

This paper aims to further build on the analysis of the country submissions to the UNFCCC Global Stocktake. The main contribution of this paper is to analyse its intersection with the IPCC reports, to profile the influence of science on country-level policy.

3 Data

We leverage the Global Stocktake dataset that was created by Sietsma et al. (2023). Their dataset builds on a dataset of climate policy links and metadata by Climate Policy Radar (Climate Policy Radar). Using the links provided in Climate Policy Radar, Sietsma et al.'s work performs the text extraction and parsing of the documents into paragraphs. The dataset contains 1,241,872 text rows, covering 1,529 documents from 371 entities (based on first author), of which 198 are countries. Additional details on the dataset including its completeness and known issues are included in Appendix A.

We then pre-process the columns (or features) of the dataset. For documents that have more than one author or document type, we extract the first field and use it for analysis.[2] We also classify the authors by type, and where the author is a country, we map it to the relevant country groups (e.g. G20) . Additional details on this column-level processing can be found in Appendix B.

Next, we specify the key datasets by author type for analysis. The first key dataset we create is the *Country Policy Submissions* (CPS) dataset, by filtering where the author type is 'Country'. The second key dataset the *IPCC reports* dataset, by filtering the author type to 'IPCC'. By delineating these two datasets, we can more effectively analyse the influence of IPCC on country-level policies.

Country Policy Submissions (CPS) The CPS dataset contains 1,102,177 rows and 1,017 documents. This distribution varies widely in terms of words per row, rows per document, documents per author (Table 2). The dataset includes 14 different types of documents are submitted by countries (Figure 4).

To facilitate further analysis, we then filter the CPS dataset for when the term 'IPCC' occurs in the 'text' field to create the *Country Policy Submissions IPCC mentions* (CPS-IPCC) dataset. This truncates the dataset to 25,948 rows and 794 unique documents.

Next, we clean the dataset so it only includes text from the main 'body' of the documents. We remove references, author lists, figure or table-related text. In addition, we specify the type of the field as 'text' and specify number of words (in each 'text' field) to be greater than 10. This yields a dataset of 18,848 rows from 750 unique documents. More details on cleaning can be found in Appendix B.

We also identified potential limitations of our dataset to take into consideration during our analysis. In terms of coverage, some types documents, such as *National Communications* and *Nationally Determined Contributions*, are more consistently present in the dataset compared to other documents. Future work could address the known gaps in the dataset. In terms of parsing, our analysis suggests that some documents are parsed alot more finely than others. The cleaning that we perform above to limit our analysis to body text aims to deal with this. These points are discussed in more detail in Appendix C.

IPCC reports The *IPCC reports* dataset consists of 37,745 rows, and includes 4 assessment reports related to the Sixth Assessment cycle (AR6) and 3 special reports. These reports were published from 2018 onwards. It does not include procedural documents such as those related to Greenhouse Gas (GHG) inventory reporting and guidelines. For the *IPCC reports*, we perform similar cleaning so it only includes text from the main body of the document. This yields a dataset of 16,666 rows.

A summary of the key datasets created for downstream analysis can be found in Table 1.

4 Profiling IPCC mentions in Country Submissions

The overall goal of profiling the mentions is to better understand the *extent* and *context* in which IPCC is being discussed. We use mention counts to measure the extent to which it is being discussed. To understand the context in which it is discussed, we use topic modelling as well as quote-finding.

4.1 Counting mentions

The most straight-forward method for identifying IPCC influence on country-level submissions to the UNFCCC can be to identify mentions of this organisation and its key outputs. Key outputs of the IPCC include the Fifth Assessment Report (AR5) and Sixth Assessment Report (AR6). We counted the number of mentions of the acronym 'IPCC', 'AR5' and 'AR6' in the CPS dataset. We

[2] This is justified as 95% of the documents only have 1 author.

	Rows	Docs	Analysis
Original (Unprocessed)	1,241,872	1,529	
Country Policy Submissions (CPS)	1,102,177	1,017	*Counting mentions*
Country Policy Submissions IPCC mentions (CPS-IPCC)	25,948	794	
- Body	18,848	750	*Topic modelling, Quote-finding*
IPCC Reports	37,745	52	
- Body	16,666	49	*Topic modelling, Quote-finding*

Table 1: Summary of Created Datasets

	Mean	St. Dev.	Min.	Max.
Words/Row	28	36	1	755
Rows/Document	1083	2960	1	85409
Documents/Author	5	3	1	20

Table 2: Distribution of *Country Policy Submissions* Dataset (n = 1,102,177)

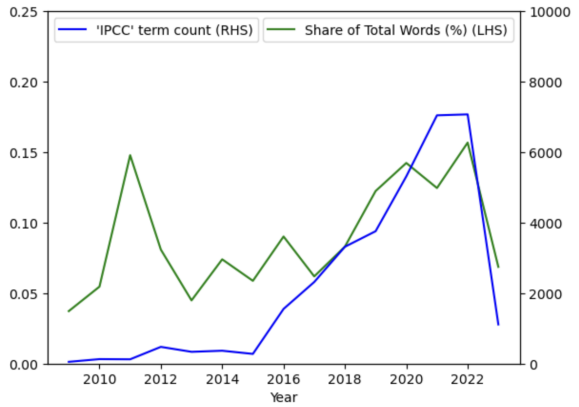


Figure 2: 'IPCC' mention counts and share of total words

take into account both the lower and upper cases of the term. We don't count the full text mentions to avoid double-counting instances where the full text mentions are accompanied by the acronymn. Our current method is high precision but could potentially miss certain instances, say when the full text is mentioned by itself without the acronymn, expanding on this is left for future work. Around 2% of rows and 80% of unique documents contain a mention of 'IPCC'. In recent years, it appears that there was a steady upward trend of the term 'IPCC', though this increase is in line with the increase in the number of words, resulting in the 'IPCC' term count as a share of total words remaining roughly stable (Figure 2). The increase in the 'IPCC' mention counts was driven by the mention counts of '2006 IPCC' and '2006 IPCC Guidelines' (Figure 3). The '2006 IPCC Guidelines' terms are likely references to the *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, a guidance that was produced by the IPCC to

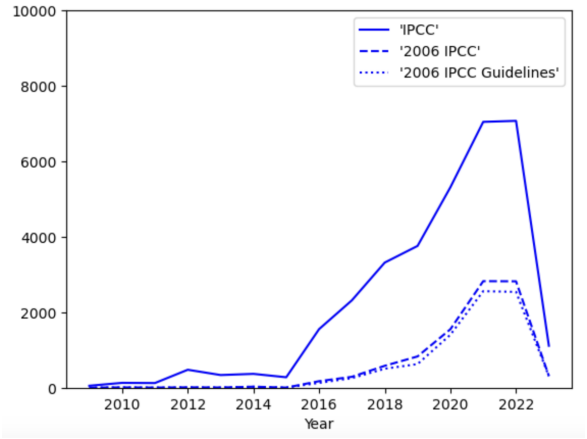
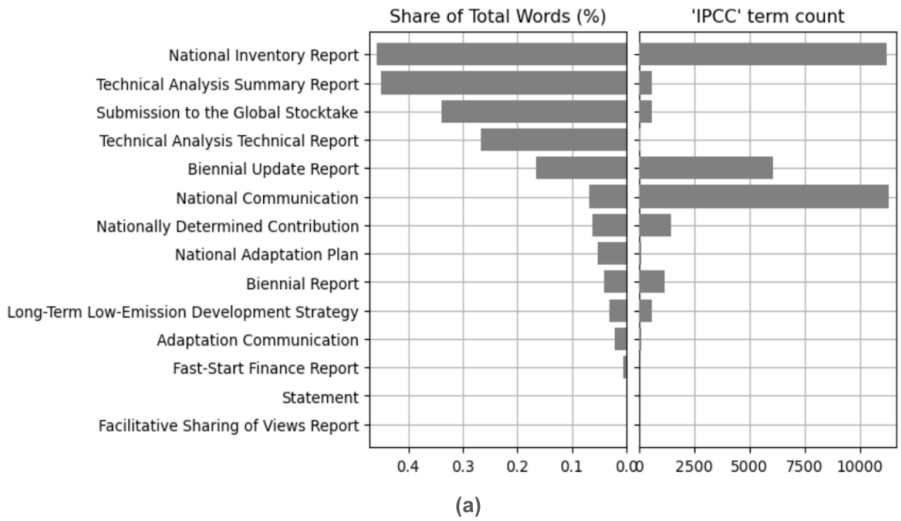


Figure 3: 'IPCC' and related phrases mention counts

assist countries in fulfilling their reporting commitments on greenhouse gas emissions to the UNFCCC (Eggleston et al., 2006). The counts for 'AR5' and 'AR6', that were published in 2014 and between 2021-2023 respectively, are negligible compared to the 'IPCC' counts.

The type of document with the highest share of 'IPCC' mentions was National Inventory Reports (NIR) (Figure 4). The NIRs are part of industrialised country's national greenhouse gas inventory reporting obligations under the UNFCCC (UNFCCC). The reports are compiled according to guidelines provided by the IPCC, such as the *2006 IPCC Guidelines for National Greenhouse Gas Inventories* as discussed above. The 'Technical Analysis Summary Reports' and 'Submission to the Global Stocktake' documents also have a fairly high share of 'IPCC' mentions, however their term count, and relative representation in the dataset is low.

By country, Argentina's reports has the most mentions of 'IPCC', where around 1 in 300 words was 'IPCC'. This is followed by South Africa (1 in 400), then Mexico (1 in 500) (Figure 5). This is largely driven by the availability of their National Inventory Report in our dataset, indeed, they are the only 3 of the G20 countries have have their report in the dataset.



Rows	Documents
105947	41
3733	12
8584	29
469	2
130495	105
596110	251
73712	379
9907	6
96596	58
57764	63
15647	34
3181	36
8	1
24	1

Figure 4: By Document Type (a) IPCC mentions (b) CPS

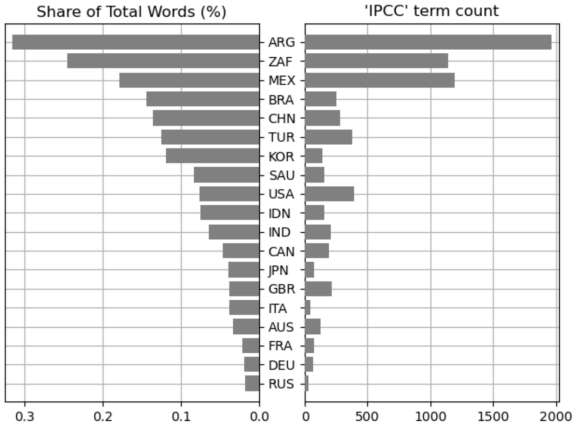


Figure 5: IPCC mentions by G20 Country

4.2 Topic Modelling

To better understand how IPCC is being discussed in the policy texts, we apply topic modelling to the portions of the text where 'IPCC' is mentioned. Topic modeling is a type of statistical modeling used to surface abstract topics within a set of documents. It automatically identifies themes or patterns in text data, classifying the information into different topics based on the co-occurrence of words. It is the method of choice here as we aim to apply a scalable and efficient way to understand the key themes in this large textual dataset.

Method We use BERTopic (Grootendorst, 2022) to discover topics in the CPS-IPCC dataset. BERTopic is a recent technique that employs sentence transformers to generate sentence embeddings, UMAP to reduce the dimensionality of the sentence vectors, and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to create dense clusters, and c-TF-IDF to generate representative words for each topic. This approach is observed to produce readable topics by leveraging recent neural language models. The results are often comparable or better than widely-used topic modeling techniques based on Latent Dirichlet Allocation (LDA). Benchmarking BERTopic and LDA is left as future work. We removed standard stopwords along with a list of domain-specific stop words.[3] The number of topics is determined automatically by BERTopic – yielding 29 topics with a handful of outliers documents. For each outlier document, we use the soft-clustering provided by HDBSCAN to find the best matching topic based on the topic with the highest probability.

Findings We named each of the 29 topics to make sense of the results. For each topic, we provide ChatGPT with the most frequent terms and the most 3 representative documents in order to obtain a suggested topic name in return.[4] We take the suggested topic name, along with the inputs we provided chatGPT, to manually assign a label to each topic. Overall the output from ChatGPT was quite representative. Some of the topic names from ChatGPT overfitted to the top 3 most represented documents, so in most cases,

[3]The list of custom stop words are: 'et', 'al', 'institute', 'university','climate', 'change', 'box', 'figure', 'table', 'ipcc', 'emission', 'emissions', 'guidelines'

[4]The prompt used and an example interaction with chatGPT is found in the appendix. Cross-checking this output against other methods like TopicGPT is left to future work.

our manual label was slightly more general than the ChatGPT-generated label. Some topics that were very similar were combined, leading to a final topic count of 25.

To make the topics more intepretable and useful, we grouped topics to a set of broader categories. We first classified the topics that referred to greenhouse gas (GHG) emissions from a particular source into broader sector-level categories that are provided by the IPCC Guidelines. These sector-level categories are Energy, Agriculture and Land Use, Land-Use Change and Forestry (LULUCF), Industrial Processes and Product Use (IPPU) and Waste (Eggleston et al., 2006). We then labelled the topics related to measurement and methodologies in the general category as 'GHG Emissions Measurement and Methods (General)'. Given that the 'Climate Change Scenarios and Impacts' category was distinct from the other two types of categories, it was classified as it as its own group.

Observations. The breakdown by topic category and topic can be found in Figure 6. The largest topic category was 'GHG Measurement and Methodologies (General)', then 'GHG Emissions - Agriculture and LULUCF' then 'Climate Change Scenarios and Impacts'. The shares were 49.0, 21.1 and 13.3 per cent respectively. The largest individual topics were 'Climate Change Scenario and Impacts' (part of the 'Climate Change Scenario and Impacts'), 'National Greenhouse Gas Inventories' (part of the 'GHG Emissions Measurement and Methods (General)') and 'Emissions from Forest Management' (part of the 'GHG Emissions - Agriculture and LULUCF'), with the respective shares of 13.3 per cent, 12.5 per cent and 10.9 per cent.

Figure 7 plots topic breakdown results by document type. National Communications and National Inventory Reports have the highest representation in the dataset, accounting for 35 and 33 per cent of rows respectively. Focusing on comparing these two reports, we find that National Communications reports contains more discussions on 'Climate Change Scenarios and Impacts' (blue topic group) than National Inventory Reports. Figure 8 breaks down IPCC mentions in National Communications by G20 countries. We observe that France (FRA), South Africa (ZAF), Germany (DEU) and Great Britain (GBR) discuss climate impacts to a greater extent than other G20 countries, whereas this topic group is underrepresented in China and Korea *National Communications*.

We also attempted to benchmark some of these findings by comparing them to the topics found in the *IPCC Reports* Dataset. The aim of this was to analyse if the topic distribution of how IPCC was discussed in the country policy submissions (as discussed above) was comparable to the topics actually discussed in the key outputs by the IPCC. However, the *IPCC Reports* dataset does not include procedural documents that discuss GHG measurements and methods, such as the *2006 IPCC Guidelines on Greenhouse Gas Inventories*, which is a key area of focus of the country mentions, so it is challenging to make a balanced comparison. Indeed, preliminary results from topic modelling on *IPCC Reports* do suggest that it covers topics related to climate change impacts and response actions. The results for this can be found in Appendix D.

We compare this work to Sietsma et al. (2023), which laid the groundwork for performing topic models in UNFCCC documents, and provided the dataset we are using. In terms of dataset, Sietsma et al. (2023) performs topic modelling on the entirety of the Global Stocktake including country submissions, whereas our topics are focused only on parts of the Country Policy Submissions that mention IPCC. In terms of topic-naming, Sietsma et al groups and aligns their topics based on major themes in the UNFCCC Synthesis report, while our topics are grouped by categories found in the IPCC Guidelines. In terms of the topics discovered, both our topics and Sietsma's have topic groups related to emissions by sector. A broader range of topics are reflected in Sietsma's work, such as Mitigation, Adaptation, and Financing. This is expected given that our dataset is a subset of theirs. One possible interpretation of our initial observations can be that IPCC mentions do not discuss mitigation or adaptation in detail , but this needs to be verified with further work.

4.3 Quote-finding

We perform quote-finding, aiming to identify text fragments (from phrases to sentences) from IPCC reports that have been adopted by the country submissions.

To do this, we further breakdown the paragraph-level CPS-IPCC dataset into sentences using the NLTK package, and then filter out

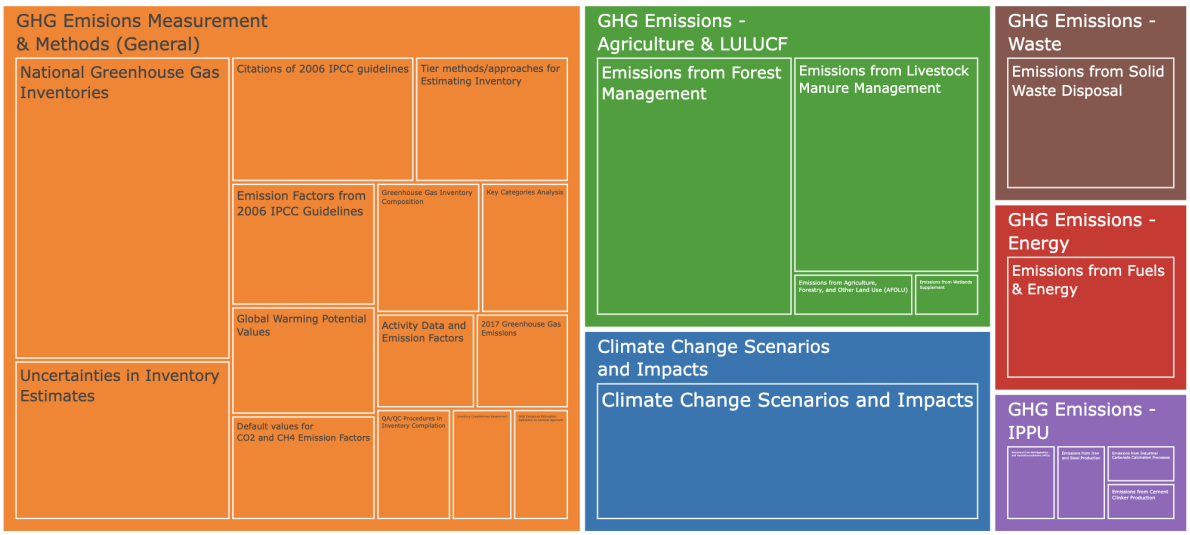


Figure 6: A treemap of topics in CPS-IPCC dataset (n = 18,848). Colors indicate six broader topic groups, block sizes reflect the frequency of the 25 topics. Acronyms: GHG - Greenhouse Gases. LULUCF - Land Use, Land-Use Change and Forestry. IPPU - Industrial Processes and Product Use

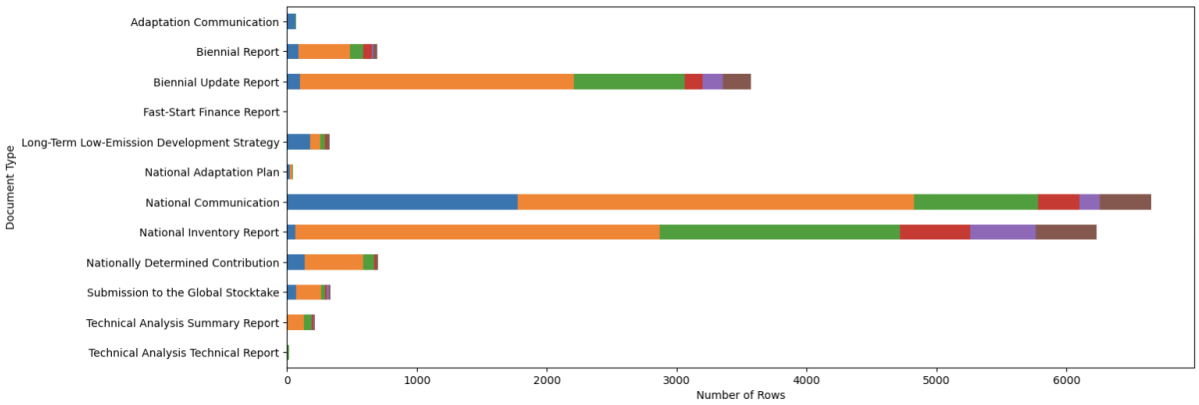


Figure 7: Topic Breakdown by document type in CPS-IPCC. Colors correspond to six topic groups in Figure 6.

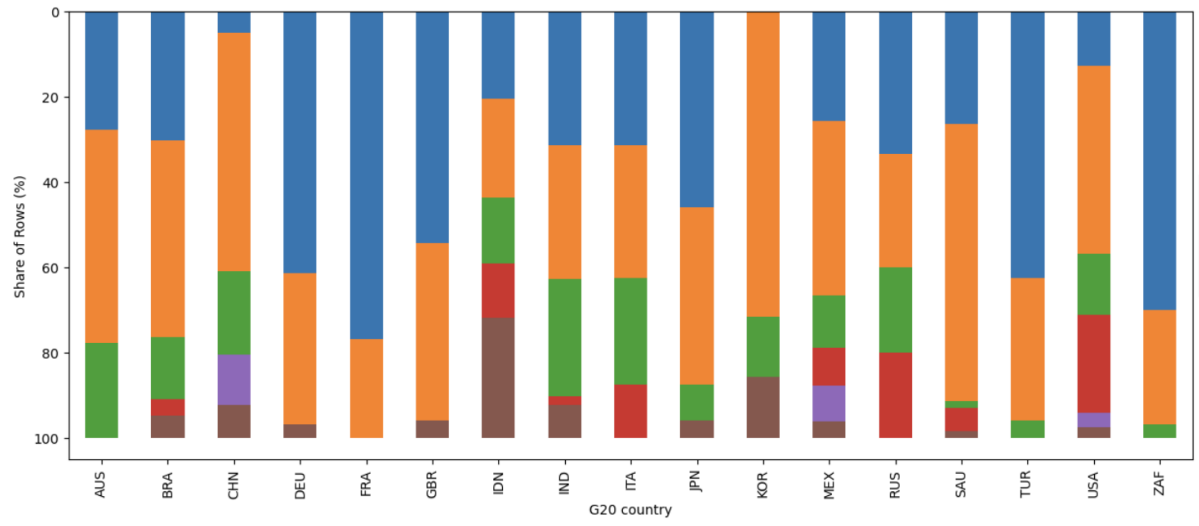


Figure 8: Topic Distribution in G20 Countrys' *National Communications.* Colors correspond to six topic groups in Figure 6.

sentences that are less than 5 words. This results in a set of 56,077 sentences in *IPCC reports* denoted as $\mathscr{S}_I$ and 49,541 sentence in CPS-IPCC denoted as $\mathscr{S}_C$.

We then compute the Levenshtein's edit distance between strings[5] for sentence pairs in $\mathscr{S}_I$ and $\mathscr{S}_C$, keeping pairs with a match score over 0.6. The edit distance calculates the minimum number of insertions and deletions required to change one sequence into the other, and the matching score is the normalised edit distance.[6] We have tried a number of thresholds, and 0.6 seems to yield a sufficient number of meaningful matches. We examine the matching text, and perform post-processing to further remove matched sentences that appear to be titles, references, or referential sentences to tables and figures.

The sentences from the IPCC reports that had the highest number of matches in the CPS-IPCC dataset mainly refer to the IPCC Guidelines on National Greenhouse Gas Inventories (see Table 8 in Appendix E). The matching text was slight variations of document reference itself: "*IPCC Guidelines for National Greenhouse Gas Inventories*". This corroborates findings from the mention counts and topic modelling, that the dominant focus of the policy texts in relation to the IPCC is the inventory guidelines.

We also analysed the matches with a score above 0.8, that is, pairs that have a larger portion of the string matched. For this, the main sentence from the IPCC reports that is matched is the UNFCCC's definition of climate change: *Note that the United Nations Framework Convention on Climate Change (UNFCCC), in its Article 1, defines climate change as: 'a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural dimate variability observed over comparable time periods'.* This is quoted 8 times in the policy texts by 5 different countries.

Additional details of the results can be found in Appendix E.

5 Conclusions

In complex and evolving issues like climate change, it is crucial to ensure that scientific findings have a direct and significant influence on policy actions. Our study marks an initial step in computationally quantifying the influence of scientific outputs from the IPCC on policy-making. By employing natural language processing techniques to analyze policy documents submitted to the UNFCCC's Global Stocktake, our findings suggest that the IPCC's scientific assessments are indeed being leveraged in policy-making, particularly in areas concerning GHG inventories measurement and methodologies. The most commonly cited IPCC output appears to be the *IPCC Guidelines on National Greenhouse Gas Inventories* (Eggleston et al., 2006).

This work provides a data-driven proxy measure of the IPCC's impact on policy and paves the way for future research to refine these methodologies and deepen the understanding of the science-policy intersection.

Discussions Several limitations of the current work warrants discussions and scoping for future work.

In terms of the dataset, we have identified several data quality issues. In terms of coverage, poor coverage of certain document types, such as the *National Inventory Reports* may skew the distribution of the topic modelling results. In addition, there are some very long paragraphs and reshuffled and broken sentences (see Appendix C). Future work should focus on enhancing the completeness and robustness of this dataset.

Our computational methods for identifying and profiling IPCC mentions rely on string matching of a small set of acronyms, while the matches seems of high precision, whether this hurts the overall recall needs further consideration. The computation is also oblivious of the political process underlying the different policy document types, which could be incorporated in the future (Stede and Patz, 2021). Our concurrent work on information extraction and linking of IPCC reports (Anonymous, 2024) aims to provide tools for better content linking.

IPCC and UNFCCC are one possible science-policy data pair, it is likely that other pairs will yield more informative results. Another important question is *what should be in the intersection between science and policy*? We expect input from stakeholders and policy experts would be informative for shaping further work on this topic.

[5] Using the python package Levenshtein.

[6] This is calculated as 1 - (distance / (length string 1 + length string 2)).

References

Authors Anonymous. 2024. Profiling and analyzing climate change statements in IPCC reports. In *Submitted*.

Ralf Barkemeyer, Suraje Dessai, Beatriz Monge-Sanz, Barbara Gabriella Renzi, and Giulio Napolitano. 2016. Linguistic analysis of IPCC summaries for policymakers and associated coverage. *Nature Climate Change*, 6(3):311–316.

Robbert Biesbroek, Sarah Judith Wright, Stefanie Korswagen Eguren, Anita Bonotto, and Ioannis N Athanasiadis. 2022. Policy attention to climate change impacts, adaptation and vulnerability: a global assessment of National Communications (1994–2019). *Climate Policy*, 22(1):97–111.

Max Callaghan, Carl-Friedrich Schleussner, Shruti Nath, Quentin Lejeune, Thomas R Knutson, Markus Reichstein, Gerrit Hansen, Emily Theokritoff, Marina Andrijevic, Robert J Brecha, et al. 2021. Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nature climate change*, 11(11):966–972.

Max W Callaghan, Jan C Minx, and Piers M Forster. 2020. A topography of climate change research. *Nature Climate Change*, 10(2):118–123.

Climate Action Tracker. Climate Action Tracker rating methodology — climateactiontracker.org. https://climateactiontracker.org/methodology/cat-rating-methodology/. [Accessed 08-05-2024].

Climate Policy Radar. Climate Policy Radar | AI for climate law and policy research — climatepolicyradar.org. https://climatepolicyradar.org/. [Accessed 08-05-2024].

HS Eggleston, Leandro Buendia, Kyoko Miwa, Todd Ngara, and Kiyoto Tanabe. 2006. 2006 IPCC guidelines for National Greenhouse Gas Inventories.

Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Angel Hsu, John Brandt, Oscar Widerberg, Sander Chan, and Amy Weinfurter. 2021. Exploring links between national climate strategies and non-state and subnational climate action in nationally determined contributions (NDCs). In *Making Climate Action More Effective*, pages 39–53. Routledge.

IPCC. IPCC; Intergovernmental Panel on Climate Change — ipcc.ch. https://www.ipcc.ch/. [Accessed 08-05-2024].

Romain Lacombe, Kerrie Wu, and Eddie Dilworth. 2023. ClimateX: Do LLMs accurately assess human expert confidence in climate statements? *arXiv preprint arXiv:2311.17107*.

Leonardo Nascimento, Takeshi Kuramochi, Gabriela Iacobuta, Michel den Elzen, Hanna Fekete, Marie Weishaupt, Heleen Laura van Soest, Mark Roelfsema, Gustavo De Vivero-Serrano, Swithin Lui, et al. 2022. Twenty years of climate policy: G20 coverage and gaps. *Climate policy*, 22(2):158–174.

Anne J Sietsma, Rick W Groenendijk, and Robbert Biesbroek. 2023. Progress on climate action: a multilingual machine learning analysis of the global stocktake. *Climatic Change*, 176(12):173.

Manfred Stede and Ronny Patz. 2021. The climate change debate and natural language processing. In *Proceedings of the 1st Workshop on NLP for Positive Impact*, pages 8–18.

UNFCCC. UNFCCC; National Inventory Submissions 2023. https://unfccc.int/ghg-inventories-annex-i-parties/2023. [Accessed 08-05-2024].

Saeid Ashraf Vaghefi, Qian Wang, Veruska Muccione, Jingwei Ni, Mathias Kraus, Julia Bingler, Tobias Schimanski, Chiara Colesanti-Senni, Nicolas Webersinke, Christrian Huggel, et al. 2023. chatIPCC: Grounding conversational AI in climate science. *arXiv preprint arXiv:2304.05510*.

Sarah Judith Wright, Anne Sietsma, Stefanie Korswagen, Ioannis N Athanasiadis, and Robbert Biesbroek. 2023. How do countries frame climate change? A global comparison of adaptation and mitigation in UNFCCC National Communications. *Regional Environmental Change*, 23(4):129.

Xin Zhou, Kshitij Jain, Mustafa Moinuddin, and Patrick McSharry. 2022. Using natural language processing for automating the identification of climate action interlinkages within the sustainable development goals.

Artur Żółkowski, Mateusz Krzyziński, Piotr Wilczyński, Stanisław Giziński, Emilia Wiśnios, Bartosz Pieliński, Julian Sienkiewicz, and Przemysław Biecek. 2022. Climate policy tracker: Pipeline for automated analysis of public climate policies. *arXiv preprint arXiv:2211.05852*.

Appendix

A Information on Original Dataset

This appendix provides information noted in the dataset's HuggingFace repository.[7] In terms of data completeness, the data was last updated on 18 November 2023. 148 of the 1726 documents have yet to be processed due to formatting issues.

There are also two known issues with the dataset: (1) Some of the 'author' fields are corrupted and (2) 'Language' field are sometimes missing or marked as "nan".

[7] https://huggingface.co/datasets/ClimatePolicyRadar/global-stocktake-documents

B Data Pre-processing

This appendix contains details of the pre-processing that was completed on this dataset.

- *Document Type*: 6 per cent of documents have two 'types' associated with it (e.g. Biannual Report and National Communication). Our analysis in this paper is based on the first type that is listed. There are also some spelling errors in this column which have been adjusted.

- *Author*: Around 3 per cent of documents have more than one author associated with it. Our analysis in this paper is based on the first type that is listed. We then categorise the authors into broad groups: Country, IPCC, UNFCCC and Other. Where the author is a country, we map it to its corresponding continent.

- *Geography*: The geography column, which contains the ISO-3 code for respective countries, has field noted as 'XAB'. 'XAB' is not a country code and it seems to contain documents where the authors are groups of nations - e.g. Independent Alliance of Latin America and the Caribbean, European Commission, European Union. However, it seems also includes two countries: India and Bhutan. For India and Bhutan, we map them back to their respective country code. Overall, we include the 'XAB' field when conducting analysis of country policy submissions.

- *Text type*: There is a column that aims to specify the type of text in the row - for example, whether it is text, table, figure or footnote. There is no available documentation outlining how this column was produced. From manual review, it seems like the 'text' category still contains rows that could be associated with table or figure titles. To take this into account, we perform our own filtering by text length and certain terms for subsequent analysis.

C Data Checks

We manually checked the dataset and made adjustments where necessary.

We checked the coverage of the dataset by reviewing the types of documents submitted by G20 countries. Figure 9 shows the number of each type of document submitted by those countries. The gaps in the graph could either be due to the country not submitting that document, or the country submitting the document but it not being included in the database due to processing issues (as outlined in Appendix A). The documents that are most consistently submitted by the G20 countries are: Nationally Determined Contribution(17/19), National Communication(17/19), and Long-Term Low-Emission Development Strategy (14/19). More broadly, the global coverage of the dataset can be seen in Figure 10.

We also checked the parsing of the document text into rows by reviewing some summary statistics. Around 25 per cent of the rows have fewer than 5 words, on manual checking these are mainly titles, headings and subheadings. The maximum number of words per row is 1000 words, this row contains a lengthy list of references.

There appears to be variation in how finely parsed different documents are on average - see Figure 11. The mean and the standard deviation of the words per row, per document, is 34 and 16 respectively. A list of the top 5 most finely parsed documents can be found in Table 3.

For a given page in a document, the paragraphs are not scraped in a sequential order. This was based on manual comparison with the source documents. This prevents us from being able to re-join the paragraphs to form a larger piece of text. Given our analysis mainly focuses on mention counts and topic modelling, this shown have a limited impact on our analysis.

Around 12 per cent of rows in the dataset start with a lowercase letter, suggesting that the parsing method has broken up many sentences midway. This could be due to sentences being cut-off at the end of the page. This could limit the effectiveness of the subsequent analysis such as topic modelling and sentence similarity.

D Topic Modelling Methods

Cleaning To clean the dataset, we filter out references and add filters to identify body text. For references, we identify them by common citation formats (for example, strings that start with 'IPCC.', 'IPCC (' or 'IPCC,') or name strings of certain report references. To identify body texts, we

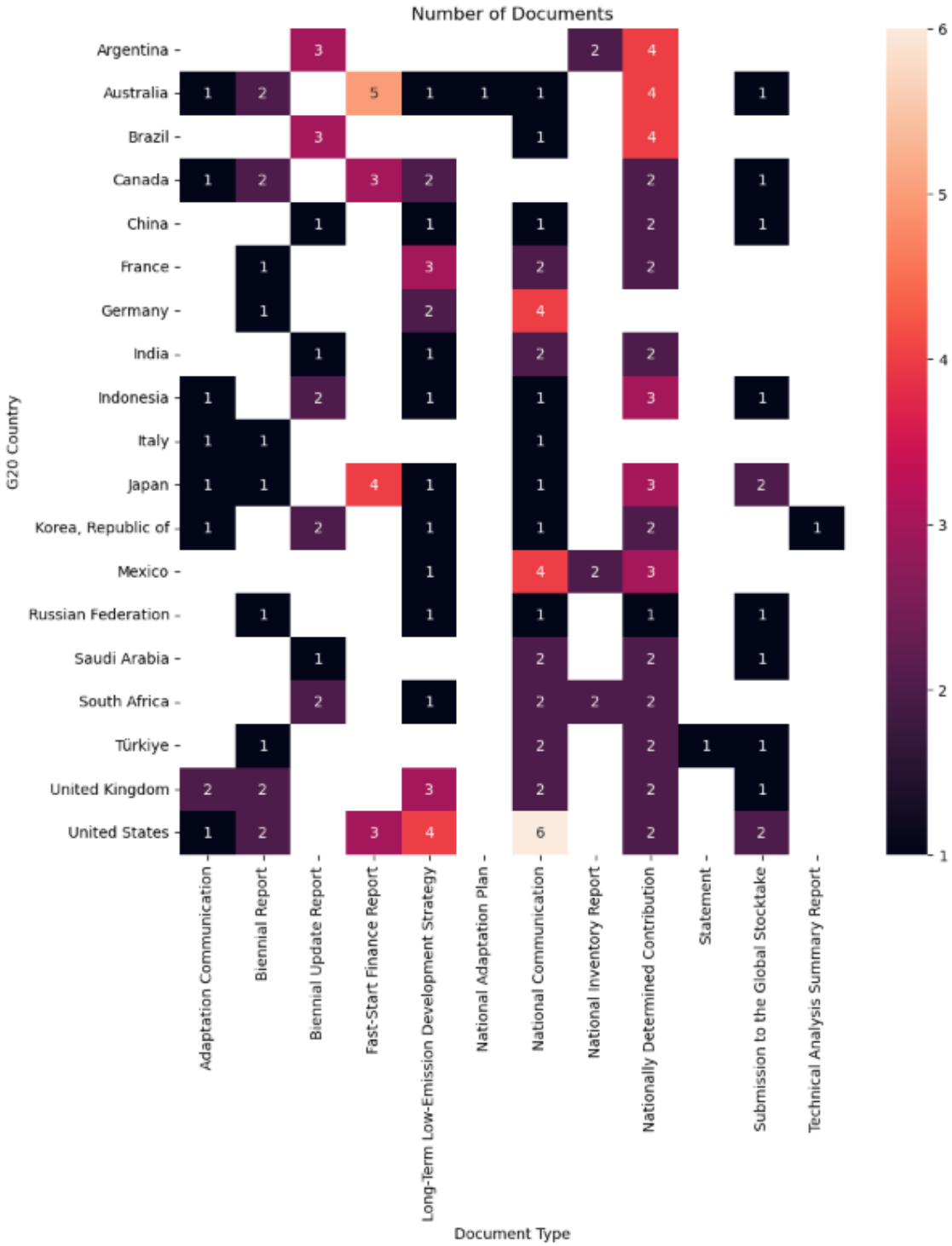


Figure 9: Number of Documents - By G20 country and Dcoument Type

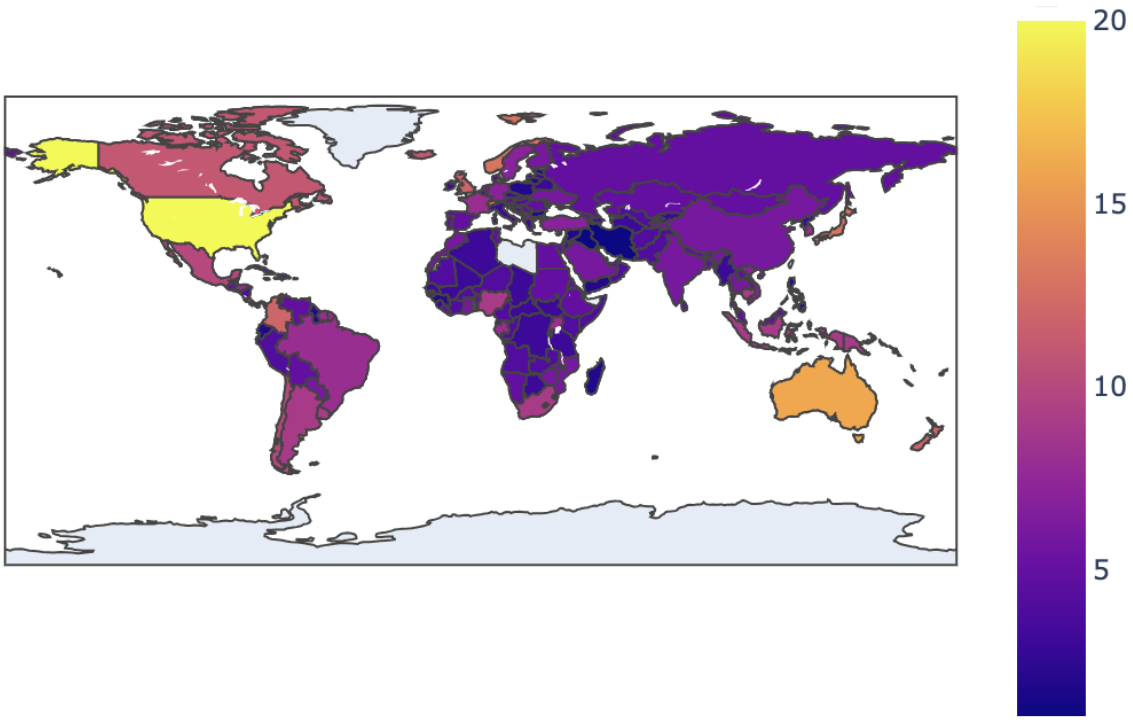


Figure 10: Document Count by Country

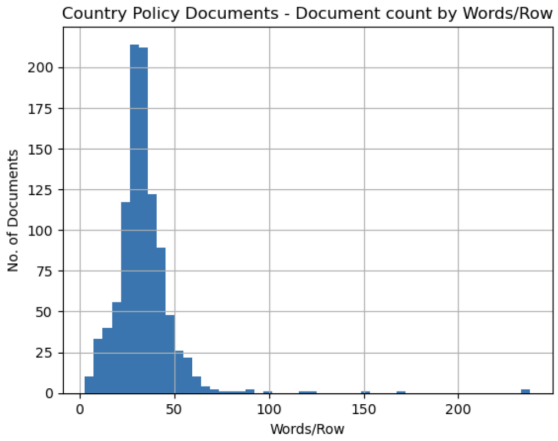


Figure 11: Distribution - Document Count by Words per Row

Document	**Words**	**Rows**	**Rows (Unique Text)**	**Words/ Row**
Individual Actions Supported by EU Fast Start Financing II	94	32	28	2.94
Cameroon First NDC (Updated submission)	4	1	1	4.00
Individual Actions Supported by EU Fast Start Financing I	1418	349	5	4.06
Denmark. National Communication (NC). NC 8. Biennial Reports (BR). BR 5.	369224	85409	21235	4.32
Figures Translation LTS Slovenia	477	91	75	5.24

Table 3: Top 5 Most Finely Parsed Documents (by Words per Row)

remove strings that start with 'Source:', 'Legend:' or '*'. We also only include fields where the type of 'text', as provided by the dataset, is 'Text' and where the number of words is greater than 10.

Model parameters For this, we specified the HDBSCAN model to have minimum size for the final cluster to be 100, and the minimum number of neighbours to a core point to be 5.

Processing outliers The model initially found 25 per cent of the dataset to be outliers. For each outlier, we use the soft-clustering provided by HDBSCAN to find the best matching topic based on the topic with the highest probability. The topic distribution before and after we process the outliers can be found the table below (Figure 5).

Manual labelling 29 topics were manually labelled, with assistance from Chat GPT. The prompt that was provided to ChatGPT can be found in Figure 12, and an example interaction can be found in Figure 13. Following the labelling, topics that were similar were combined, reducing the total number of topics to 25. For example, 3 topics referring to 'National Greenhouse Gas Inventories' were combined into 1. 2 topics referring to 'Key Categories Analysis' were combined into 1. 2 topics related to tier-based methods for estimating inventory were combined into 'Tier methods/approaches for Estimating Inventory' (Table 6).

Topic modelling results figures The numerical figures related to Figure 6 can be found in the Table 4. The top words and most representative paragraph for each topic can be Table 6.

Topic modelling bench-marking We aimed to benchmark our findings with topics found in the IPCC dataset. The topics found are shown in Table 7.

Find a topic label (a few words) for the following:
Top words: [Insert top words]
Top document 1: [Insert most relevant document]
Top document 2: [Insert 2nd most relevant document]
Top document 3 : [Insert 3rd most relevant document]

Figure 12: ChatGPT prompt to generate label

Find a topic label (a few words) for the following:
Top words: steel, production, iron, coke

Top document 1:A Tier I approach (country-specific emission factors) was applied to calculate the emissions from iron and steel production for the different process types in 2017. Default IPCC emission factors were used for the calculation of GHG emissions from basic oxygen furnace, electric furnace and pig iron production, direct reduced iron production and sinter. The separation of energy and process emissions emanating from the use of coke was not done due to a lack of disaggregated information on coke consumption. Hence, energy-related emissions from iron and steel production have been accounted for through the application of default IPCC emission factors.

Top document 2:Iron and steel production results in the emission of CO_2, CH, and N,O. According to the 2006 IPCC

Top document 3: Metal production includes the production of metals such as iron and steel. The quantity of iron and steel produced in 2000 was 2.7 and 7.6 kt respectively. Emission factors were the default values recommended by the IPCC 2006. For iron production, default CO2 emission factor from pig iron production was used and as for steel production, CO_2 emission was calculated by electric arc furnace steel making method.

"GHG Emissions from Iron and Steel Production"

Figure 13: Example ChatGPT prompt and Output

Topic Group	Topic	Share (%)
GHG Emisions Measurement & Methods (General)		**49.0**
	2017 Greenhouse Gas Emissions	1.7
	Activity Data and Emission Factors	1.8
	Citations of 2006 IPCC guidelines	4.4
	Default values for CO2 and CH4 Emission Factors	2.9
	Emission Factors from 2006 IPCC Guidelines	3.4
	GHG Emissions Estimation: Reference vs Sectoral Approach	1.2
	Global Warming Potential Values	3.0
	Greenhouse Gas Inventory Composition	2.6
	Inventory Completeness Assessment	1.3
	Key Categories Analysis	2.2
	National Greenhouse Gas Inventories	12.5
	QA/QC Procedures in Inventory Compilation	1.6
	Tier methods/approaches for Estimating Inventory	3.7
	Uncertainties in Inventory Estimates	6.6
GHG Emissions - Agriculture & LULUCF		**21.1**
	Emissions from Agriculture, Forestry, and Other Land Use (AFOLU)	1.1
	Emissions from Forest Management	10.9
	Emissions from Livestock Manure Management	8.6
	Emissions from Wetlands Supplement	0.6
Climate Change Scenarios and Impacts		**13.3**
	Climate Change Scenarios and Impacts	13.3
GHG Emissions - Waste		**6.2**
	Emissions from Solid Waste Disposal	6.2
GHG Emissions - Energy		**5.9**
	Emissions from Fuels & Energy	5.9
GHG Emissions - Industrial Process and Product Use (IPPU)		**4.5**
	Emissions from Cement Clinker Production	0.9
	Emissions from Industrial Carbonate Calcination Processes	0.9
	Emissions from Iron and Steel Production	1.3
	Emissions from Refridgeration and Hydroflourocarbons (HFCs)	1.3

Table 4: Topic Distribution of *Country Policy Submissions IPCC mentions* dataset (n = 18,848)

Topic	Count (Before)	Count (After)	Count (Difference)
NaN	4571	0	-4571
Climate Change Scenarios and Impacts	2461	2502	41
Emissions from Forest Management	1884	2049	165
Emissions from Livestock Manure Management	1578	1618	40
Emissions from Solid Waste Disposal	1140	1175	35
Uncertainties in Inventory Estimates	1093	1253	160
Citations of 2006 IPCC guidelines	740	832	92
Emissions from Fuels & Energy	736	1118	382
Global Warming Potential Values	450	564	114
Emission Factors from 2006 IPCC Guidelines	358	638	280
National Greenhouse Gas Inventories	345	693	348
National Greenhouse Gas Inventories	303	329	26
National Greenhouse Gas Inventories	286	1329	1043
Default values for CO2 and CH4 Emission Factors	277	545	268
2017 Greenhouse Gas Emissions	259	312	53
Tier methods/approaches for Estimating Inventory	257	457	200
QA/QC Procedures in Inventory Compilation	233	304	71
Greenhouse Gas Inventory Composition	210	482	272
Emissions from Refridgeration and Hydroflourocarbons (HFCs)	204	242	38
Key Categories Analysis	194	236	42
Emissions from Agriculture, Forestry, and Other Land Use (AFOLU)	161	199	38
Emissions from Industrial Carbonate Calcination Processes	153	165	12
Emissions from Cement Clinker Production	127	179	52
Activity Data and Emission Factors	126	346	220
Tier methods/approaches for Estimating Inventory	124	245	121
Emissions from Iron and Steel Production	121	254	133
GHG Emissions Estimation: Reference vs Sectoral Approach	120	235	115
Key Categories Analysis	119	185	66
Inventory Completeness Assessment	115	245	130
Emissions from Wetlands Supplement	103	117	14

Table 5: Topic Distribution of *Country Policy Submissions IPCC mentions* dataset (n = 18,848) - Before and After Outlier Processing

Topic GPT	Topic Manual	Topic Manual - Final	Topic Manual - Final - Group	Top words	Most Representative Document
Climate Change Impacts	Climate Change Scenarios and Impacts	Climate Change Scenarios and Impacts	Climate Change Scenarios and Impacts	global, scenarios, report, temperature	Under its sixth assessment cycle, the IPCC released a series of special reports on global warming and its impacts in 2018 and 2019. According to these reports, human activities have been responsible for approximately 1.0°C of global warming since pre-industrial times, and temperature rise is likely to breach 1.5°C between 2030 and 2052 at current rates of warming. More intense and frequent climate and weather extremes have been observed. Warming from historical anthropogenic emissions since the pre-industrial period will continue to drive long-term shifts in the climate system such as sea level rise. According to the Special Report on the Ocean and Cryosphere in a Changing Climate, under a "business-as-usual" (RCP8.5) scenario, global mean sea level rise could increase to 1.1m by 2100, which is 10cm above the most likely estimate from the Fifth Assessment Report released in 2014. This is due to a larger projected ice mass loss from the Antarctica ice sheet.
Carbon Accounting in Forest Management	Emissions from Forest Management	Emissions from Forest Management	GHG Emissions - Agriculture & LULUCF	land, forest, biomass, carbon	Methodologies applied in this national inventory were based on the Revised 1996 IPCC Guideline for National Greenhouse Gas Inventories (IPCC, 1997) and the uncertainty analysis of the activity data and emission factor was undertaken according to the Good Practice Guidance for Land Use, Land-Use Change and Forestry (IPCC, 2003). Forest carbon pools were classified into 6 categories including living aboveground biomasses, living below-ground biomasses, dead wood, litter, soil organic carbon and harvested wood products (IPCC, 1997). Only the fate and amount of aboveground biomass, however, were taken into account. Appropriate methodology tiers, either with or without change of forest land to other land use were chosen according to the Good Practice Guidance for Land Use, Land-Use Change and Forestry (IPCC, 2003). Similar to previous reports, Thailand adopted multiple tiers in the national GHG inventory for LULUCF sector in this report. Activity data were obtained from the interpretation of satellite imageries and statistical reports from relevant agencies. Tier 2 emission factors were adopted in most activities using country-specific data from local publications and IPCC defaults were applied as Tier 1 emission factors when appropriate.
Livestock Manure Management	Emissions from Livestock Manure Management	Emissions from Livestock Manure Management	GHG Emissions - Agriculture & LULUCF	manure, management, livestock, cattle	The methodology used to estimate the emission factors corresponded to level 2 of the IPCC. Country-specific methane emission factors, Enteric Fermentation and Manure Management, and nitrogen excretion factors were estimated for the calculation of nitrous oxide by Manure Management (included in
Waste Management and Treatment	Emissions from Solid Waste Disposal	Emissions from Solid Waste Disposal	GHG Emissions - Waste	waste, wastewater, solid, treatment	Based on the country context and the data availability, 4.A - Solid Waste Disposal, 4.B - Biological Treatment of Solid Waste, 4.C - Incineration and Open Burning of Waste, 4.D - Wastewater and Treatment and Discharge categories are reported under the waste sector. The GHG emissions were estimated using either available data from various literature or by using default values provided in the 2006 IPCC guidelines. The current results show a slight growth of emissions in all subcategories mainly due to the increasing population in the same period and changes in the main mode of waste management.

Topic GPT	Topic Manual	Topic Manual - Final	Topic Manual - Final - Group	Top words	Most Representative Document
Uncertainty Assessment in Emissions Inventory	Uncertainties in Inventory Estimates	Uncertainties in Inventory Estimates	GHG Emisions Measurement & Methods (General)	uncertainty, uncertainties, data, activity	The uncertainty of the LULUCF sector activity data comes mainly from the estimation error of the geospatial data. While the uncertainties of the emission factors used come from the default values of the 2006 IPCC Guidelines. The uncertainty of the activity data has been included in the uncertainty calculation process and are reported together as uncertainties of emission factors.
Citation of IPCC Guidelines Equations	Citations of 2006 IPCC guidelines	Citations of 2006 IPCC guidelines	GHG Emisions Measurement & Methods (General)	2006, chapter, volume, page	Equation 3.2 (chapter 3 of the 2006 IPCC guidelines Volume 1):
Energy Fuel Consumption	Emissions from Fuels & Energy	Emissions from Fuels & Energy	GHG Emissions - Energy	fuels, energy, fuel, consumption	According to the guidelines of the 2006 IPCC Guidelines, the emissions associated with the consumption of fossil fuels in the fuel tanks of international air and maritime transport, and the CO2 emissions from the burning of fossil fuels, are estimated and reported separately from the national total. biomass for energy purposes.
Global Warming Potential (GWP) Reporting	Global Warming Potential Values	Global Warming Potential Values	GHG Emisions Measurement & Methods (General)	gwp, warming, report, global	The estimated CH4, N_2O, HFCs and SF emissions were converted to CO_2 equivalent (CO2 eq) using Global Warming Potentials (GWPS) values provided by the IPCC in its Second Assessment Report ("1995 IPCC GWP Values") based on the effects of GHGS over a 100-year time horizon (Table 2.1).
Default Emission Factor Calculation (2006 IPCC Guidelines)	Emission Factors from 2006 IPCC Guidelines	Emission Factors from 2006 IPCC Guidelines	GHG Emisions Measurement & Methods (General)	factors, default, 2006, calculation	Emission factors are used by default from the 2006 IPCC Guidelines.
National Greenhouse Gas Inventories Guidelines	National Greenhouse Gas Inventories 1	National Greenhouse Gas Inventories	GHG Emisions Measurement & Methods (General)	greenhouse, gas, inventories, national	as well as 1996 IPCC Guidelines for National Greenhouse Gas Inventories.
National GHG Inventory Compilation	National Greenhouse Gas Inventories 2	National Greenhouse Gas Inventories	GHG Emisions Measurement & Methods (General)	inventory, ghg, national, inventories	This Third National Communication (TNC) details a comprehensive national GHG emission inventory by sources and removals by sinks for the base years 2006 and 2012 using comparable methodologies of IPCC 2006. The scope of improvement with reference to the inventories presented in TNC include: (i) adherence to the TACCC (Transparency, Accuracy, Consistency, Comparability and Completeness) principle for developing a National GHG Inventory; (ii) development of different inventory life cycle documents such as institutional arrangement (IA) and the template methods and data documentation (MDD), detailed QA/QC procedures to ensure the highest quality of GHG Inventory; (iii) identification of data gaps on different sectors and addressing those with proper explanation in light of IPCC GL; (iv) the development for the first time, of an archiving system in DoE to preserve and regularly update the necessary inventory works and data-base; (v) a national inventory improvement plan with a focus on addressing the data gaps and sustainable inventory development capacities within and outside of the nodal agency, i.e. DoE; (vi) inclusion of additional GHG pools identified in IPCC 2006 guidelines (IPCC 2006) for the preparation of national greenhouse gas emission inventories that were not included in INC and SNC; (vii) a strong emphasis on QA/QC procedures as identified in IPCC Good Practices Guidance 2000 and 2003 (GPG 2000, 2003).

Topic GPT	Topic Manual	Topic Manual - Final	Topic Manual - Final - Group	Top words	Most Representative Document
National GHG Inventory Development	National Greenhouse Gas Inventories 3	National Greenhouse Gas Inventories	GHG Emisions Measurement & Methods (General)	ghg, inventory, national, inventories	This GHG inventory is prepared using methodology developed in the revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories. The UNFCCC software "Non Annex 1 National Greenhouse Gas Inventory Software" has been used for the estimation of GHG.
Default CH4 Emission Factors	Default values for CO2 and CH4 Emission Factors	Default values for CO2 and CH4 Emission Factors	GHG Emisions Measurement & Methods (General)	ch4, factors, default, efs	For all other emission sources, default values from the 2006 IPCC Guidelines are used for CO_2 and CH4 emission factors.
Greenhouse Gas Emissions by Sector (2017)	2017 Greenhouse Gas Emissions	2017 Greenhouse Gas Emissions	GHG Emisions Measurement & Methods (General)	gg, sector, 2017, total	In the period 1990 to 2017 GHG emissions from the Agriculture Sector increased by 73% from 11,623.10 Gg CO_2 eq in 1990 to 20,073.90 Gg CO_2 eq in 2017. Emissions from the Agriculture sector increased by 38% from 11,623.10 Gg CO_2 equivalents in 1990 to 16,036.89 Gg CO_2 equivalents in 2005. In the period 2005 to 2017 GHG emissions from the Agriculture sector increased by 25% from 16,036.89 Gg CO_2 equivalents in 2005 to 20,073.90 Gg CO_2 equivalents in 2017. The increase of emissions is mainly caused by increasing emissions from Enteric Fermentation and Manure Management (IPCC subcategory 3.A and 3.B) and Agricultural Soils (IPCC subcategory 3.D).
Tiered Method for Emission Factor Calculation	Tier methods for Estimating Inventory	Tier methods/approaches for Estimating Inventory	GHG Emisions Measurement & Methods (General)	tier, factors, method, default	For the estimation of GHG emissions for the category, Tier 1 and Tier 2 methods were applied for the subcategories, using disaggregated country-specific activity data and default emission factors in accordance with the 2006 IPCC Guidelines.
QA/QC Procedures in Inventory Compilation	QA/QC Procedures in Inventory Compilation	QA/QC Procedures in Inventory Compilation	GHG Emisions Measurement & Methods (General)	qc, quality, qaqc, procedures	The general and category-specific QC procedures are performed by the experts during inventory calculation and compilation according to the QA/QC and verification plan. The QC procedures used in Finlandś GHG inventory comply with the 2006 IPCC Guidelines. General inventory QC checks (2006 IPCC Guidelines, Vol 1, Chapter 6, Table 6.1) include routine checks of the integrity, correctness and completeness of the data, identification of errors and deficiencies, and documentation and archiving of the inventory data and quality control actions. Category-specific QC checks including reviews of the activity data, emission factors and methods are applied on a case-by-case basis focusing on key categories and on categories where significant methodological changes or data revisions have taken place.
Greenhouse Gas Inventory Composition	Greenhouse Gas Inventory Composition	Greenhouse Gas Inventory Composition	GHG Emisions Measurement & Methods (General)	dioxide, gases, inventory, nox	This NIRS covers the full territory of the country for the recommended time series 1990 to 2016 and the results are presented at the national level. The inventory addressed all the IPCC sectors Energy, Industrial Processes and Product Use (IPPU), Agriculture, Forestry, and Other Land Use (AFOLU) and Waste and categories subject to Activity Data (AD) availability. The gases covered in this inventory are the direct gases carbon dioxide (CO_2), methane (CH4), nitrous oxide (N_2O) and hydro-fluorocarbons (HFCs) as well as the indirect gases nitrogen oxides (NO_2), carbon monoxide (CO), non-methane organic volatile compounds (NMVOCs) and sulphur dioxide (SO_2).

Topic GPT	Topic Manual	Topic Manual - Final	Topic Manual - Final - Group	Top words	Most Representative Document
Refrigeration and Air Conditioning HFC Usage	Emissions from Refridgeration and Hydroflouro-carbons (HFCs)	Emissions from Refridgeration and Hydroflouro-carbons (HFCs)	GHG Emissions - IPPU	refrigeration, hfc, conditioning, air	Hydrofluorocarbons (HFCs) and, to a very limited extent, perfluoro-carbons (PFCs), have high global warming potentials and are being used as alternatives to different classes of ozone-depleting substances (ODS) that are being phased out under the Montreal Protocol. According to the IPCC 2006 guidelines, HFCs and PFCs are being used in a variety of applications that includes refrigeration and air conditioning, fire suppression and explosion protection, aerosols, solvent cleaning, foam blowing, and other applications such as equipmentś sterilization.
Key Categories Analysis	Key Categories Analysis	Key Categories Analysis	GHG Emisions Measurement & Methods (General)	key, categories, trend, level	According to the IPCC definition, a key category is one that is prioritized within the national inventory system because its estimate has a significant influence on a countryś total inventory of direct GHGs in terms of the absolute level of emissions, the trend in emissions, or both. Total emissions from the key categories amount to 95% of the total emissions included in the inventory. Key categories for Lebanon are determined with Approach 1 described in Volume 1, Chapter 4 of the 2006 IPCC Guidelines. 4 categories have been identified as key in the analysis, with CO_2 being the main gas and the energy sector being the main key category. Analyses of key categories is performed collectively for all sectors and a list of all key categories is presented in Table 10.
Agriculture, Forestry, and Other Land Use (AFOLU) Sector Analysis	Emissions from Agriculture, Forestry, and Other Land Use (AFOLU)	Emissions from Agriculture, Forestry, and Other Land Use (AFOLU)	GHG Emissions - Agriculture & LULUCF	afolu, sector, categories, 3a	The main categories of the inventory were obtained following Method 1 by evaluation of level, tendency and uncertainty. For this, the emissions and removals in absolute value were taken into account based on the participation that each source has in the inventory until 95% of the total was covered. This analysis was carried out following the guidelines established in Table 4.1 of the 2006 IPCC Guidelines (Volume 1, Chapter 4) to determine the level of aggregation of the categories of the Energy, PIUP and Waste sectors. In the case of the AGSOUT sector, given that it contains several significant subcategories, the recommendation was made to carry out the analysis of main categories with a higher level of disaggregation. In particular, category 3A was subdivided into 3A1ai Dairy cattle; 3A1aii Beef cattle and 3A1ab-j Other livestock. Regarding subcategories 3C4 and 3C5, they are broken down by type of animal production following the same criteria as in category 3A, by harvest residues, by synthetic fertilizers and by direct N_2O emissions due to the loss of organic matter from the soil. . On the other hand, since Argentina has a consistent representation of Method 1 lands, it is not possible to assign the changes in soil carbon to land use categories 3B1 to 3B6. For this reason, a new subcategory *3B7-Variation of soil organic matter (carbon)" was generated to report the change in total soil carbon corresponding to the area of the country included in the coherent representation of lands, and it is this subcategory 3B7 that included in the main category analysis.

Topic GPT	Topic Manual	Topic Manual - Final	Topic Manual - Final - Group	Top words	Most Representative Document
Industrial Carbonate Calcination Processes	Emissions from Industrial Carbonate Calcination Processes	Emissions from Industrial Carbonate Calcination Processes	GHG Emissions - IPPU	lime, glass, production, soda	According to 2006 IPCC Guidelines, Vol. 3, Chap. 2.3.1.1, the Tier 1 method is based on applying a default emission factor to national level lime production data. While country-specific information on lime production by type (e.g., high calcium lime, dolomitic lime, or hydraulic lime) is not necessary for good practice in Tier 1, where data are available to identify the specific types of lime produced in the country, this may be used. It is not necessary for good practice to account for LKD in Tier 1.
Cement Clinker Production Emissions	Emissions from Cement Clinker Production	Emissions from Cement Clinker Production	GHG Emissions - IPPU	cement, clinker, production, cao	GHG emissions from cement production were estimated using a Tier 2 methodology (2006 IPCC Guidelines), based on activity data on clinker production. Data obtained directly from the producer on CaO and MgO ratio in clinker and cement were used for calculating the CO_2 emission factors.
Quality Control for Activity Data and Emission Factors	Activity Data and Emission Factors	Activity Data and Emission Factors	GHG Emisions Measurement & Methods (General)	activity, data, sources, factors	As a quality control, the recommendations indicated in Chapter 6 - Volume 1 of the 2006 IPCC Guidelines were followed. Additionally, the hypotheses used for the selection of the activity data, the appropriate calculation methodologies, the selection of emission factors and calculations. Likewise, the transcription of the activity data from the original sources to the spreadsheets was reviewed, mainly contrasting the totals of the original reports, with the sums made in the spreadsheets.
Estimating GHG Emissions using IPCC Tier 1 Approach	Tier approaches for Estimating Inventory	Tier methods/approaches for Estimating Inventory	GHG Emisions Measurement & Methods (General)	tier, ch4, approach, method	For estimating the GHG emissions (CO_2, CH4, N_2O) the 2006 IPCC Guidelines Tier 1 approach has been applied:
GHG Emissions from Iron and Steel Production	Emissions from Iron and Steel Production	Emissions from Iron and Steel Production	GHG Emissions - IPPU	steel, production, iron, coke	A Tier I approach (country-specific emission factors) was applied to calculate the emissions from iron and steel production for the different process types in 2017. Default IPCC emission factors were used for the calculation of GHG emissions from basic oxygen furnace, electric furnace and pig iron production, direct reduced iron production and sinter. The separation of energy and process emissions emanating from the use of coke was not done due to a lack of disaggregated information on coke consumption. Hence, energy-related emissions from iron and steel production have been accounted for through the application of default IPCC emission factors.
GHG Emissions Estimation: Reference vs Sectoral Approach	GHG Emissions Estimation: Reference vs Sectoral Approach	GHG Emissions Estimation: Reference vs Sectoral Approach	GHG Emisions Measurement & Methods (General)	reference, approach, sectoral, fuel	According to the 2006 IPCC, the emissions level is estimated using reference and sectoral approaches. The reference approach is a top-down approach, using a countryś energy supply data to calculate the emissions of CO_2 from combustions of fossil fuels, while the sectoral approach is a bottom-up approach, using fuel data of stationary as well as mobile combustions in several sectors of energy category to calculate CO2, CH2, N_2O emissions. It is good practice to apply both sectoral and reference approaches to estimate a countryś CO_2 emissions from fuel combustion and to compare the results of these two estimates. Figure 2-3 presents GHG emissions level estimated under reference approach by type of fuels and under sectoral approach by type of sector category. In 2015, GHG emissions by sectoral approach reached 4,437 Gg CO2e, while by reference approach were 4,441 Gg CO2e.

Topic GPT	Topic Manual	Topic Manual - Final	Topic Manual - Final - Group	Top words	Most Representative Document
GHG Key Category Analysis: IPCC Tier 2 Level	Key Category Level Analysis	Key Categories Analysis	GHG Emisions Measurement & Methods (General)	key, analysis, categories, level	Table 5.8 presents the results of the IPCC Tier 2 key category level analysis for the year 2019. There is a total of 21 key categories based on the Tier 2 current year level analysis, with LULUCF.
Inventory Completeness Assessment and Software Utilization	Inventory Completeness Assessment	Inventory Completeness Assessment	GHG Emisions Measurement & Methods (General)	inventory, software, completeness, source	An assessment of the completeness of the inventory was made for individual activity areas within each source category and the results are presented within the sections covering the individual sectors. The methodology adopted was according to the IPCC 2006 Guidelines (IPCC 2007) with the following notation keys used:
Wetlands Emission Estimation and IPCC Supplement Utilization	Emissions from Wetlands Supplement	Emissions from Wetlands Supplement	GHG Emissions - Agriculture & LULUCF	wetlands, supplement, reservoirs, peatlands	(IPCC 2006). In wetlands category emissions are estimated only for managed wetlands due to human

Table 6: Topics from *Country Policy Submissions IPCC mentions* dataset -Labelling output, Top words and Representative Documents.

E Sentence Similarity Results

The sentences from the IPCC reports that had the highest number of matches in the CPS-IPCC dataset mainly refer to the IPCC Guidelines on National Greenhouse Gas Inventories (Table 8).

For the top matching sentence when the threshold is set to 0.6, Table 9 shows the highest scoring matches and Table 10 shows the lowest scoring matches. The matching text was generally the slight variations of document reference itself: "IPCC Guidelines for National Greenhouse Gas Inventories".

The top matching sentence when the threshold is set to 0.8 is about the UNFCCC climate change definition. The list of 8 occurrences across 5 different countries can be found in Table 11.

Topic	Top words	Topic Count)	Share of Total
0	confidence, warming, ice, changes	8320	50.0
1	energy, development, pathways, mitigation	3567	21.0
2	food, production, security, land	1164	7.0
3	desertification, degradation, land, erosion	557	3.0
4	bioenergy, cdr, beccs, land	399	2.0
5	options, response, management, land	300	2.0
6	srm, aerosol, radiation, radiative	269	2.0
7	coastal, slr, protection, sea	268	2.0
8	scenarios, scenario, socioeconomic, ssps	244	1.0
9	united, states, america, sciences	238	1.0
10	report, findings, likelihood, evidence	200	1.0
11	water, irrigation, drought, groundwater	171	1.0
12	carbon, costs, discount, price	171	1.0
13	carbon, ocean, atmospheric, sink	137	1.0
14	soil, carbon, organic, soils	121	1.0
15	disaster, insurance, risk, instruments	119	1.0
16	feasibility, mitigation, option, options	110	1.0
17	tropical, cyclones, cyclone, tc	105	1.0
18	ecosystem, services, es, human	106	1.0
19	cryosphere, ocean, srocc, changes	100	1.0

Table 7: Topic Distribution of *IPCC Reports* dataset (n = 16,666)

IPCC Report Sentence	**Number of matches in *Country Policy Reports IPCC mentions*** dataset
It links to the IPCC Guidelines on National Greenhouse Gas Inventories in the land sector.	231
In the context of the IPCC Guidelines for National Greenhouse Gas Inventories, a tier represents a level of methodological complexity.	99
Tier In the context of the IPCC Guidelines for National Greenhouse Gas Inventories, a tier represents a level of methodological complexity.	83
[Note: More details can be found in 2006 IPCC Guidelines for National GHG Inventories, Volume 4, Chapter 1.]	78
13387-[Note: For a discussion of the term 'forest' in the context of National GHG inventories, see the 2006 IPCC Guidelines for National GHG Inventories.]	55
d y [Note: For a discussion of the term 'forest in the context of National GHG inventories, see the 2006 IPCC Guidelines for National GHG Inventories (PCC 2006).)	40
[Note: More details can be found in IPCC 2006 Guidelines for National GHG Inventories, Volume 4, Chapter 1.)	37
The most important trace gases contributing to the greenhouse effect are carbon dioxide (CO_2), ozone (O_2), methane (CH_2), nitrous oxide (NO), perfluorocarbons (PFCs), chlorofluorocarbons (CFCs), hydrofluorocarbons (HFCs), sulphur hexafluoride (SF) and water vapour (H_2O).	32
Emission sources follows the definitions by the IPCC Task Force on National Greenhouse Gas Inventories (TFI) (IPCC 2019).	26
The ocean warming trend documented in the IPCC Fifth Assessment Report (ARS) has continued.	22

Table 8: IPCC Reports Sentences (n=56,077) - Top 10 matches by Count

Sentence from *Country Policy Submissions IPCC mentions*	**Distance**	**Ratio**	**Matching Text**
9812-In 2006 IPCC Guideline for National Greenhouse Gas Inventories.	34	0.75	[' IPCC Guideline', ' National Greenhouse Gas Inventories']
9809-In 2006 IPCC Guideline for National Greenhouse Gas Inventories.	34	0.75	[' IPCC Guideline', ' National Greenhouse Gas Inventories']
9808-In 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Vol.	31	0.75	[' IPCC Guidelines ', ' National Greenhouse Gas Inventories']
9807-In 2006 IPCC Guideline for National Greenhouse Gas Inventories.	34	0.75	[' IPCC Guideline', ' National Greenhouse Gas Inventories']
9811-In 2006 IPCC Guideline for National Greenhouse Gas Inventories.	34	0.75	[' IPCC Guideline', ' National Greenhouse Gas Inventories']

Table 9: Top-scoring Matches of the Sentence,*It links to the IPCC Guidelines on National Greenhouse Gas Inventories in the land sector.*, from the IPCC report. In the sentence column, the number at the start refers to the indexing of the sentence, to show that it is a unique sentence in the dataset.

Sentence from *Country Policy Submissions IPCC mentions*	**Distance**	**Ratio**	**Matching Text**
12720-In line with the IPCC guidelines, we used national emission factors wherever possible (in some activities in the sectors of	66	0.6	[' the IPCC ', 'uidelines', 'ational ', 'es in the ', 'sector']
4564-The analysis was performed using Approach 1 recommended in the 2006 IPCC Guidelines for National Greenhouse Gas Inventories Vol.4 Ch.4.	75	0.6	['IPCC Guidelines ', ' National Greenhouse Gas Inventories ']
14781-Intergovernmental Panel on Climate Change (IPCC), 1997: Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories.	80	0.6	['Guidelines ', ' National Greenhouse Gas Inventories']
4498-Methods are based on 2006 Intergovernmental Panel on Climate Change (IPCC) Guidelines for National Greenhouse Gas Inventories.	80	0.6	[' Guidelines ', ' National Greenhouse Gas Inventories']
15177-The application of methods has been proof-read and follows the 2006 IPCC Guidelines for National Greenhouse Gas Inventories (IPCC 2006).	75	0.6	['IPCC Guidelines ', ' National Greenhouse Gas Inventories ']

Table 10: Lowest-scoring Matches of the Sentence,*It links to the IPCC Guidelines on National Greenhouse Gas Inventories in the land sector.*, from the IPCC report. In the sentence column, the number at the start refers to the indexing of the sentence, to show that it is a unique sentence in the dataset.

Sentence	Geography	Document Name	Date	Text Index	Ratio	Matching Text
Note that the United Nations Framework Convention on Climate Change (UNFCCC), in its Article 1, defines climate change as: "a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods."	TTO	Trinidad and Tobago. National Communication (NC). NC 3.	2021-12-29	p21b2	0.99	[Note that the United Nations Framework Convention on Climate Change (UNFCCC), in its Article 1, defines climate change as: , a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural , imate variability observed over comparable time periods]
EXECUTIVE SUMMAR T he United Nations Framework Convention on Climate Change (UNFCCC) defines climate change as: "change of climate that is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and that is in addition to natural climate variability observed over comparable period of time".	CIV	Côte dIvoire. National Communication (NC). NC 3.	2017-12-31	p24b7	0.88	[he United Nations Framework Convention on Climate Change (UNFCCC), defines climate change as: , change of climate , is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and , is in addition to natural , imate variability observed over comparable]
This usage differs from that in the United Nations Framework Convention on Climate Change (UNFCCC), where climate change refers to a change of climate that is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and that is addition to natural climate variability observed over comparable time periods.	NZL	New Zealands Adaptation communication	2017-12-21	p438b3	0.88	[the United Nations Framework Convention on Climate Change (UNFCCC), , climate change , a change of climate , is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and , addition to natural , imate variability observed over comparable time periods]
This usage differs from that in the United Nations Framework Convention on Climate Change (UNFCCC), where climate change refers to a change of climate that is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and that is addition to natural climate variability observed over comparable time periods.	NZL	New Zealands Adaptation communication	2017-12-21	p471b4	0.88	[the United Nations Framework Convention on Climate Change (UNFCCC), , climate change , a change of climate , is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and , addition to natural , imate variability observed over comparable time periods]
This usage differs from that in the United Nations Framework Convention on Climate Change (UNFCCC), where climate change refers to a change of climate that is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and that is addition to natural climate variability observed over comparable time periods.	NZL	New Zealand. National Communication (NC). NC 7.	2017-12-21	p438b3	0.88	[the United Nations Framework Convention on Climate Change (UNFCCC), , climate change , a change of climate , is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and , addition to natural , imate variability observed over comparable time periods]

Sentence	Geography	Document Name	Date	Text Index	Ratio	Matching Text
This usage differs from that in the United Nations Framework Convention on Climate Change (UNFCCC), where climate change refers to a change of climate that is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and that is addition to natural climate variability observed over comparable time periods.	NZL	New Zealand. National Communication (NC). NC 7.	2017-12-21	p471b4	0.88	[the United Nations Framework Convention on Climate Change (UNFCCC), , climate change , a change of climate , is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and , addition to natural , imate variability observed over comparable time periods]
Under the UNFCCC, climate change is defined as "a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods".	GNB	Guinea-Bissau. National Communication (NC). NC 3.	2018-03-09	p87b2	0.84	[the U, limate , hange , define, a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural , imate variability observed over comparable time periods]
he United Nations Framework Convention on Climate Change (UNFCCC) defines climate change in the following terms: A change of climate directly or indirectly attributable to human activity, which alters the composition of the global atmosphere and adds to the natural climate variability observed over comparable time periods.	CIV	Côte dIvoire. National Communication (NC). NC 3.	2017-12-31	p14b8	0.82	[e United Nations Framework Convention on Climate Change (UNFCCC), change of climate , irectly , to human activity, alters the composition of the global atmosphere and , natural , imate variability observed over comparable time periods]
The revised act directly addresses the climate change related challenges and also provides a definition of the term, reflecting common IPCC terminology (p. 3): "a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods".	VUT	Vanuatu. National communication (NC). NC 3.	2021-03-22	p130b8	0.76	[a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural , imate variability observed over comparable time periods]

Table 11: Matches of the Sentence, *Note that the United Nations Framework Convention on Climate Change (UNFCCC), in its Article 1, defines climate change as: 'a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural dimate variability observed over comparable time periods'.*, from the IPCC report.