Mapping forward-looking mitigation studies at country level

Claire Lepault and Franck Lecocq
CIRED, AgroParisTech, Université Paris Saclay, CNRS, ENPC, CIRAD, EHESS, 45 bis, avenue de la Belle Gabrielle, 94736 Nogent-sur-Marne Cedex, France
E-mail: claire.lepault@enpc.fr and lecocq@centre-cired.fr

Keywords: mitigation pathway, forward-looking, national, topic modeling, natural language processing

Supplementary material for this article is available online

Abstract
We provide the first survey of the rapidly expanding literature on country-level mitigation pathways using systematic mapping techniques. We build a database of 4691 relevant papers from the Web of Science and Scopus. We analyze their abstracts and metadata using text mining and natural language processing techniques. To discover common topics within the abstracts, we use an innovative and fully reproducible topic modeling approach based on two machine learning models. We find that the number of papers per country is well correlated with current levels of greenhouse gas (GHG) emissions, with few papers for (current) low emitters, notably in Africa. Time horizons of 2030 and 2050 each account for one-third of the papers, with the former actually more frequent in recent years, spurred by interest in the (Intended) Nationally Determined Contributions. Topic modeling analysis of the data set reveals that forward-looking mitigation papers encompass all dimensions of mitigation, save for financial issues, that are lacking. However, energy and to a lesser degree land use, land use change and forestry are very dominant relative to other sectors. Topics are unevenly addressed across countries, reflecting national circumstances and priorities, but also pointing to gaps in the literature. The limited number of forward-looking papers in (currently) low-emitting countries raises questions about the lack of research capacity in support of the construction of domestic climate policies.

1. Introduction

The Paris Agreement, signed in 2015, emphasizes nationally determined contributions (NDCs) as the building block of global action against climate change, today and over time as countries are expected to ramp up their ambition over time in subsequent NDCs. An increasing number of countries has also communicated long-term low greenhouse gas emission development strategies under Article 4 of the Agreement and/or adopted mid-century mitigation goals.

Although principles exist for mitigation that are general enough to apply everywhere (e.g. decarbonizing electricity, electrifying end-uses, promoting energy efficiency, enhancing carbon sinks), building effective mitigation strategies at country level requires one to take into account local economic, social, technological, institutional and cultural circumstances. This is even more so when mitigation objectives are ambitious. To inform such a process, country-specific analysis is required (Fragkos et al. 2021).

While the country-level assessment of ambitious climate objectives have been conducted for many countries, via both individual exercises and multi-country projects, such as the Deep Decarbonization Pathways project (Waisman et al. 2019), CD-LINKS (2019), COMMIT (2019) or ENGAGE (www.engage-climate.org/), to our knowledge no comprehensive survey of this literature exists despite its relevance for policy-making. This may be explained by the large number of countries, large number of research teams with diverse backgrounds (energy, macroeconomics, environment, etc) working on national mitigation pathways, and by the lack of institutions that would bring them together. In contrast, the literature on mitigation pathways at global level originates from a limited number of research teams worldwide and benefits from well-developed institutions, such as global mitigation scenario databases hosted by the International Institute for Applied Science Analysis or the Integrated Assessment Modeling Consortium. It has been extensively surveyed, in particular in the IPCC 4th and 5th Assessment Reports.
In this paper, we bridge this gap by providing a comprehensive overview of the literature on mitigation pathways at country level. Specifically, we ask how comprehensive is the geographical coverage of this literature? Up to what time horizons does it consider mitigation strategies? What models do these analyses use? What are the main aspects of mitigation it addresses?

To do so, we harvest forward-looking mitigation papers at country level from the Web of Science (WoS) and Scopus databases, resulting in a data set of 4691 abstracts and other paper metadata. We use language processing techniques to extract additional information from the abstracts, such as country, time horizon of the analysis or name of the model used. Finally, we use topic modeling techniques to identify the main issues discussed in each of the papers in the database. We improve the method proposed by Lamb et al. (2019) by reducing subjectivity bias and by optimizing the parameters of the method to maximize the explanatory power of the topics.

Overall, our paper contributes to a growing literature that mobilizes big data and machine learning techniques to analyze the academic literature on climate change (Belter and Seidel 2013, Wang et al. 2014, Li and Zhao 2015, Haunschild et al. 2016, Aleixandre-Benavent et al. 2017, Lamb et al. 2019, Callaghan et al. 2020).

In addition to our main findings presented below, the data set of papers and related topics produced in this research is of interest on its own as it allows researchers, policymakers and stakeholders to ‘zoom in’ on particular topics and/or countries of interest to inform policy processes and/or identify research gaps. We strive here to provide the method and results in a clear, transparent and fully reproducible way, with a view to making the results easier to communicate (Minx et al. 2017, Donnelly et al. 2018, Haddaway and Macura 2018).

The rest of this article is structured as follows: section 2 details our methodology, section 3 presents and discusses our results, and section 4 provides the conclusion.

2. Methodology

2.1. Database construction

To find mitigation pathway(s) at national level, we search the academic databases WoS and Scopus for papers that meet the following three conditions: (i) include the name of a country in the title, abstract or keywords, and (ii) include ‘mitigation’ or a synonym in the title, abstract or keywords, and (iii) include a year in the period [2025–2100] in the title, abstract or keywords.

The two selections are then merged into one database, and duplicates are eliminated. Due to differences in the coverage of peer-reviewed journals, the results of the searches from WoS and Scopus differ significantly, with 944 references that appear only in Scopus and 574 only in WoS. The search expressions and the resulting database of 4691 papers obtained on 14 November 2020, can be found in the supplementary material (available online at stacks.iop.org/ERL/16/083001/mmedia).

Limiting the search for country names to the title of the reference is based on the observation that papers focusing on national mitigation pathways typically have the name of the country in the title. Conversely, attempts to use search equations with country names in the abstract led to harvesting too many irrelevant papers. Finally, adding a year is critical to restricting the search to papers devoted to future pathways. Without this condition, the vast literature on current mitigation policies, for instance, would also be embarked upon in the search.

While the search string is precise and encompasses all countries, our identification strategy has two shortcomings. First, we focus on papers in English only, although there are relevant papers on forward-looking mitigation at country level in other languages. Second, we search two major databases (WoS and Scopus), while other relevant papers may be indexed elsewhere. Overall, however, we believe that our approach provides an extensive view of the available literature.

2.2. Additional treatments

The database is post-treated using the Pandas library (Wes McKinney 2010) of Python software. We search country names, demonyms and acronyms in the title to associate each of the papers to a country. When the title of a paper contains more than one country name (which occurs for 153 papers), one entry is created for each. The resulting extended database contains 4884 rows. We use it to analyze the geographical coverage of our data set.

Two additional parameters are added. First, we search the title, keywords and abstracts of each of the papers for a horizon year in the [2025, 2100] ranges. If more than one number is found, we retain the largest one. Second, we search for model names, again duplicating entries if several models are identified.

We use a combined list of models from the comparative review of scenario modeling tools for national pathways to the Sustainable Development
Goals (SDGs) by Allen et al (2016), the list of models documented by the IAMC, as well as the generic expressions ‘computable general equilibrium (CGE)’ and ‘integrated assessment model’. The database with one entry for each combination of paper, country and model has 4996 rows.

2.3. Topic modeling

2.3.1. Overview

Topic modeling is a machine learning method aimed at discovering common topics within a corpus of documents; here the abstracts of the papers selected above. Specifically, we use the non-negative matrix factorization (NMF) classification method (Lee and Seung 1999). The starting point is to build the so-called term frequency-inverse document frequency (TF-IDF) matrix, in which each row corresponds to a paper and each column to a word, and in which coefficients measure the frequency of a given word in the abstract of a given paper, weighted by the frequency of that particular word in the whole corpus.

The next step is to decompose the TF-IDF matrix, i.e. to search for a combination of matrices W and H so that their product $W \times H$ best approximates TF-IDF. The columns of matrix $W$, as well as the rows of matrix $H$, can then be interpreted as topics. Matrix $H$ indicates the weight of each word in each topic, while matrix $W$ indicates the weight of each topic in each abstract.

The outcome of the method is sensitive to the number of topics (the number of columns of matrix $W$ and the number of rows of matrix $H$) as well as to other parameters of the optimization process. Previous studies using the NMF classification method have explored several topics and selected the value based on expert judgment (Lamb et al 2019, Callaghan et al 2020). Here, we reduce the risk of arbitrariness and subjectivity bias by selecting exogenous parameters from a topic coherence measure (O’Callaghan et al 2015), based on the Word2vec word embedding algorithm (Mikolov et al 2013a, 2013b). The following details each step.

2.3.2. Corpus identification

To identify the corpus, abstracts are pre-treated. All characters are put in lower case, punctuation signs, connectors and commonly used words are deleted, and words are grouped according to common radicals. Since the country scope and time horizon of each of the papers are already identified through the search equation, country names and time horizons are deleted. Terms related to mitigation listed in the search equation are also deleted, since by construction of the database each abstract contains at least one of them. Finally, we exclude terms that are either too rare (i.e. that appear in less than 1% of the abstracts) or too frequent (i.e. that appear in more than 95% of the abstracts). The final corpus contains 1300 terms.

2.3.3. TF-IDF matrix construction

We measure the weight of each term using the TF-IDF index, defined for each abstract $a$ and each term $t$ as follows (Salton and Buckley 1988):

$$TFIDF_{at} = \frac{tfidf(a, t)}{\sum_{i=1}^{l} tfidf(a, i)^2},$$

where

$$tfidf(a, t) = tf(a, t) \times \left( \log \left( \frac{A}{df(t)} \right) + 1 \right),$$

with $tf(a, t)$ the number of occurrences of term $t$ in abstract $a$ and $df(t)$ the number of abstracts containing term $t$. The TF-IDF index thus weights a particular term in a particular abstract if it appears frequently in that abstract but not frequently in the rest of the corpus.

2.3.4. Topic identification

We use the NMF method to identify relevant clusters of words (hereafter topics). The algorithm searches for a set of $K$ topics so that the product of the non-negative matrices abstract-topic $W_{a \times K}$ and topic-terms $H_{K \times T}$ best approximates $TFIDF_{a \times T}$. The $W$ matrix can be interpreted as the weight of each topic in each abstract, while the $H$ matrix represents the weight of each term in each topic.

Since the selected number of topics is small relative to the total number of abstracts (typically less than 5%), there is no algorithm of polynomial complexity that converges to a unique solution\(^3\). However, one can iteratively converge to local solutions by solving the optimization problem (3), in which $||\cdot||_F$ and $||\cdot||_1$ are the Frobenius and $L_1$ norms, respectively, and where $\alpha \geq 0$ and $0 \leq l_1 \leq 1$ are coefficients.

$$\min_{W,H \geq 0} \frac{1}{2}||X - WH||_F^2 + \alpha [l_1 (||W||_1 + ||H||_1) + (1 - l_1)(||W||_F^2 + ||H||_F^2)].$$

The first term in equation (3) ensures the convergence of the $WH$ product towards $TF - IDF$, while the second imposes additional constraints on the structure of $W$ and $H$. The $L_1$ regularization (second term) favors the presence of null coefficients in the matrices, thereby limiting the number of topics each abstract is related to, and limiting the number of terms each topic contains. The minimization of the $L_2$ regularization (third term), on the other hand, tends to limit differences across coefficients in the matrices.

We initialize the NMF algorithm using the Non-Negative Double Singular Value Decomposition method (Boutsidis and Gallopoulos 2008, Belford...}
et al 2018). To ensure that our results are reproducible, we set the random seed of the algorithm to 1511.

2.3.5. Parameter optimization in the NMF method
The set of topics identified with the NMF method is contingent on the choice of \( K \) (number of topics), \( \alpha \) (intensity of regularization relative to the optimization criteria) and \( l_1 \) (regularization parameter). We thus build a performance measure for each set of topics, and then select the triplet \((K,\alpha,l_1)\) that produces the highest ranking set. To our knowledge, this is the first time that the algorithm below has been used to select the best triplet (it has previously been used to select \( K \) only (O’Callaghan et al 2015)).

The index we use is called ‘coherence’. It measures how similar the semantic environments of each of the terms that compose a given topic are. The higher the index, the more consistent are the words that compose the topic. The coherence of a set of topics is the average of the coherence of each individual topic.

Following O’Callaghan et al 2015, we produce a vectorial representation of the semantic environment of each term within the corpus of abstracts using the Word2vec word-embedding algorithm (Mikolov et al 2013a, 2013b). Word2vec is a two-layer neuronal network that maps words into vectors that account for the semantic environment of the word. Words that share similar contexts are characterized by similar multi-dimensional vectors.

We use the Skip-gram method to train the neural network. This approach seeks to predict the semantic context of a term. The error is computed based on the corrected prediction of the words surrounding this term. The coherence of each topic coherence is then the mean of the pairwise cosine similarities between terms that characterize the topic. Precisely, for a topic \( k \), the coherence index \( TCW2V_k \) is computed as follows:

\[
TCW2V_k = \frac{1}{\binom{N}{2}} \sum_{j=2}^{N} \sum_{i=1}^{j-1} \text{similarity}(wv_{gi},wv_{gj}),
\]

where \( N \) is the number of terms that we choose to characterize each topic \( c^i \), \( wv_{gj} \) is the vector associated to term \( j \) characterizing topic \( k \), \( wv_{gi} \) is the vector associated to term \( i \) characterizing topic \( k \) and \( \text{similarity}(A,B) \) is the cosine similarity of vectors \( A \) and \( B \), defined as:

\[
\text{similarity}(A,B) = \frac{\sum_{i=1}^{D} A_i B_i}{\sqrt{\sum_{i=1}^{D} A_i^2} \sqrt{\sum_{i=1}^{D} B_i^2}}.
\]

4 For each topic, we retain the five terms with the highest ranking score in the topic—terms matrix \( H \).

We assign the set of topics resulting from each triplet \((K,\alpha,l_1)\) with the mean of the scores obtained for each individual topic.

We systematically explore the set of triplets \((K,\alpha,l_1)\) in the range \([2, 40] \times [0.00, 0.31] \times [0.0, 1.0] \) (with increments of 1, \((0.01:0.1)\) and \(0.1\) respectively). For each combination, we perform the NMF algorithm and compute the coherence index of the resulting set of topics. The highest score is obtained for the triplet \((39, 0.1, 0.9)\), which we retain for the remainder of the study.

2.4. Relationship across topics
In order to visualize how topics relate to each other (figure A6), we use LDAvis (Sievert and Shirley 2014), a system initially developed to explore topic-term relationships in a fitted latent Dirichlet allocation model. The intertopic distance is based on the Jensen–Shannon divergence calculated from the \( H \) matrix coefficients characterizing the topic-terms relationships. Principal component analysis then projects the set of intertopic distances onto two dimensions. In the online supplementary material, interactive visualization is available and represents the individual terms that are most useful for interpreting each topic. In particular, it enables one to look at the corpus-wide frequency of a given term as well as the topic-specific frequency of the term.

2.5. Relationship between abstracts and topics
The \( W \) matrix links topics to abstracts. However, it has too many non-zero coefficients, and a threshold is required is to ascribe a topic to an abstract. This, in turn, requires that the weights of each topic in each abstract be comparable across abstracts.

We normalize the \( W \) matrix so that the sum of the coefficients of each row is equal to one. In this way, each line of the matrix can be interpreted as a share of each topic in a given abstract. To do so, we transform each coefficient in the \( W \) matrix as:

\[
W_{ak}^{\ast} = \frac{W_{ak} \times \sum_{i=1}^{T} H_{ki}}{\sum_{i=1}^{T} WH_{ai}}.
\]

We then ascribe topic \( k \) to abstract \( a \) if \( W_{ak}^{\ast} > 0.02 \), as per Lamb et al (2019).

To check how relevant the resulting mapping is, we build another mapping based on titles. Specifically, we ascribe topic \( k \) to abstract \( a \) if \( W_{ak}^{\ast} > 0.02 \), and if at least one of the five terms best characterizing topic \( k \) appears in the title of the paper. Figure A5 presents the number of papers per topic in each mapping. As the figure illustrates, these distributions are similar. The one on the bottom is scaled down from the one on top. Since the presence of a word characterizing a topic in the title of a paper is a strong indication that the paper is indeed related to that particular
Figure 1. Geographical distribution of papers in the database. Papers related to the EU as a whole are not represented. (Source: Authors).

3. Results

3.1. Papers are distributed in proportion to countries’ greenhouse gas (GHG) emissions

Overall, 136 countries (plus the European Union) appear in the database (figure 1). However, the geographical distribution of papers is particularly skewed. China accounts for 24.3% of all papers. A distant second is the US (9.0%), followed by the UK (6.0%), the EU as a region (5.3%) (this figure excludes papers related to individual EU Member States), and India (4.9%). Region-wise, almost half of the papers (46.5%) focus on Asia, a little more than a quarter on Europe (28.4%) and a sixth on the Americas (17.3%, of which 7.0% on Latin America and 10.3% on North America). Each of the other regions accounts for less than 5% of the papers. Africa, in particular, is very poorly represented (4.4%), with all but seven countries in the region with less than ten papers, and almost half with no paper at all.

The representation of each country in the database appears to be well correlated with its GHG emissions (figure 2). This is not surprising since the larger the problem, the more likely it is to attract the attention of the (domestic and foreign) research community, either suo motu or at the request of governments or other interested parties. A prominent exception is the UK, which features much more frequently in the database than its GHG emissions would suggest. This may translate to the strength of the UK research community on mitigation, and/or the fact that with the adoption of the Climate Change Act in 2008, the UK has a longer history of national climate policies than most high-emission countries.

Another exception is Russia, with significantly fewer papers than its share of emissions would suggest. This might reflect the fact that our search is confined to papers written in English. It might also point to a research community that has invested more in other priorities than mitigation. Among countries with medium- or low-emissions, OECD countries, particularly in Europe (e.g. Finland, Switzerland, Sweden) have more papers in their database than their share of GHG emissions would suggest. Developing countries, on the other hand, tend to be closer to the line or below.

The tail of the distribution is also relevant for policy making. Of a total of 197 Parties to the UN Framework Convention on Climate Change, 143 have less than ten papers in the database, 127 fewer than five and 65 have do not have any. In their survey of urban climate mitigation case studies, Lamb et al. (2019) similarly find a very uneven distribution of papers by country. While it is difficult to determine a threshold below which the number of forward-looking publications on mitigation would be ‘insufficient’ to inform policies, ten papers or fewer (to be compared with 39 major topics in the database, see below) leaves little chance that even the different sectoral aspects of mitigation are adequately covered. Policymakers and stakeholders in Africa, in particular, have for the most part scant scientific literature to rely on, despite rapidly increasing emissions. Informing strategies to limit growth in GHG emissions (and ultimately start reducing them) while continuing with other development goals requires a major shift in the focus of research towards this continent.

3.2. Paris Agreement has spurred increased attention to the 2030 time horizon

The distribution of the time horizons of papers (figure 3(b)) presents two very clear peaks in 2030...
and 2050, respectively, each accounting for 34% of all papers. Only 14% of all papers have a time horizon beyond 2050, a major difference with the literature on mitigation at the global level, in which 2100 is the norm. This translates to a difference in research questions. Forward-looking mitigation studies at global level are typically conducted to assess mitigation scenarios against long-term temperature goals, whereas forward-looking studies at national level typically have the objective of assessing more detailed policy packages. For this purpose, 2050 is already a long time horizon.

Figure 3(a) presents papers by publication year and time horizon. The results show a rapid expansion of the forward-looking mitigation literature at country level over the past two decades, with a clear inflexion point around 2014. The annual increment of publications is markedly higher in the 2014–2020 period than the 2007–2013 period. This inflexion does not appear when looking at the climate change literature as a whole. When doing so, in contrast, Callaghan et al (2020) found annual increments more or less constant over the whole 2007–2020 period. The timing of the inflexion point (2014) suggests that the prior negotiation and adoption of the Paris Agreement, with its strong emphasis on national-level mitigation, may have spurred increased interest in national-level mitigation.

This hypothesis is further supported by the fact that 2014 also marks an inflexion in the distribution of time horizons across papers. As can be seen in figure 3(a), the share of papers with a time horizon

---

6 The Paris Agreement was signed in 2015. However, Parties to the UNFCCC were asked to submit Intended NDCs at COP19 in Warsaw (Poland) in November 2013.
up to 2030 was decreasing before 2014. But from 2014 onwards, this trend has reversed. Since 2030 is the time horizon of nearly all the intended NDCs communicated in 2015 (and of most of the NDCs to date), this finding suggests again that the negotiation and signature of the Paris Agreement have spurred increased interest in mitigation at national level. The adoption of the SDGs in 2015, also with 2030 as a time horizon, may also have played a role, though mitigation is only one of the 17 SDGs\(^7\).

Finally, figure 3(b) shows that the distribution of time horizons differs by region. Europe and North America represent more than 62% of the literature with a 2050 time horizon, against 25% of the literature with a 2030 time horizon. Conversely, Asia represents 68% of the literature with a 2030 time horizon, against 40% up to 2050. This suggests that research in Europe and North America is already focused on mid-century time horizons, consistent with the mid-century mitigation strategies that several European countries and the EU have adopted. However the focus in Asia would be more on the conditions under which NDCs can be achieved by 2030. If this explanation is correct, then we should soon see an increase in the share of papers with regard to 2050 and 2060 time horizons in Asia following the recent announcement of the long-term mitigation objective by China.

3.3. Studies offer a comprehensive but uneven coverage of major mitigation issues
Table 1 presents the outcome of the topic modeling analysis described in Methods. The topics are ranked by numbers of papers attached (column T0.02), and characterized by their five most relevant words (column Terms)\(^8\). The ‘title’ attached to each topic (column Topic) is our work.

Almost all of the papers in the database (4687 out of 4691) are related to topic No. 1, characterized by the words ‘policy—develop—econom—countri—use’ . This is not surprising since papers on mitigation scenarios at country level typically discuss policy implications, including in the abstract. More interesting is the fact that the corpus is then split nearly in half between papers related to topic No. 2 (Climate Change) and those related to topic No. 3 (Energy Efficiency). The two ensembles are largely disjointed, as can be seen from the mapping of the strength of the pairwise combinations of topics (figure A7). Papers associated with topic No. 2 (Climate Change) tend to be also associated with topics, such as Drought, Flood, Water, Crop Yield, Forest, Land Use, Agriculture or Air Pollution. However papers associated with topic No. 3 (Energy Efficiency) tend to be associated with topics, such as Hydrogen, Steel/Iron, Nuclear, Peak, Oil, CCS, Wind/Solar or Buildings. The other topics can be organized into five groups: (i) methods (Scenarios and Systems), (ii) policies (e.g. Costs or Targets/INDC),

\(^7\) A breakdown of papers by publication year and region (figure A1) shows a rapid increase in the share of papers in Asian countries since 2016, mostly driven by China. This is consistent with the hypothesis that the Paris Agreement spurred increased interest in national level mitigation, given the prominent attention given to the Chinese NDC.

\(^8\) Topics are characterized by word stems rather than full words. For example, the word stem corresponding to ≪country≫ or ≪countries≫ is ‘countri’. We use the stemming algorithm from the library stemming.porter2 (https://pypi.org/project/stemming/1.0/).
Table 1. Description of the 39 topics, including short name (column Topic), five most relevant words (column Terms), type of topic (column Category), sector when relevant, related IPCC WG3 AR6 chapter for classification purpose when relevant, and number of related papers with threshold of 0.02 (column T0.02) and with threshold of 0.02 plus keywords in title (column T0.02 Title) (see Methods and figure A5).

| Topic               | Terms                                      | Category  | Sector | Relevant IPCC WG3 chapter | T0.02 | T0.02 Title |
|---------------------|--------------------------------------------|-----------|--------|---------------------------|-------|-------------|
| Policy-Dvlpt-Eco    | polici_develop_econom_countri_use          | Context   |        |                           | 4687  | 1324        |
| Climate change      | climat_chang_temperatur_futur_project      | Context   |        |                           | 2000  | 1038        |
| Energy efficiency   | energi_effici_consumpt_save_demand         | Context   |        |                           | 1732  | 1018        |
| Scenario            | scenario_model_bau_refer_three             | Method    |        |                           | 1428  | 553         |
| Consumption         | intens_structure_consumpt_factor_growth    | Other     |        |                           | 936   | 221         |
| Electricity         | electr_generat_demand_grid_suppli          | Sectoral  | Energy |                           | 904   | 471         |
| Power               | powerGenerat_plant_sector_capac            | Sectoral  | Energy |                           | 797   | 416         |
| Costs               | cost_abat_option_benefit_margin            | Policy    |        |                           | 12    | 761         |
| Target/INDC         | target_achiev_indc_meet_ndc                | Policy    |        |                           | 13    | 706         |
| Air pollution       | air_pollut_pm2_qualiti_health              | Other     |        |                           | 17    | 674         |
| Fuel                | fuel_fossil_diesel_altern_biofuel           | Sectoral  | Energy |                           | 586   | 157         |
| Transport           | transport_sector_road_passeng_freight      | Sectoral  | Transport|                          | 567   | 298         |
| System              | system_model_transit_integr_pathway        | Method    |        |                           | 543   | 339         |
| Renewable energies  | renew_energi_sourc_share_res                | Sectoral  | Energy |                           | 540   | 359         |
| Cement              | industri_cement_sector_product_process     | Sectoral  | Industry|                          | 497   | 301         |
| Technology          | technolog_advanc_deploy_low_clean          | Technology|        |                           | 421   | 168         |
| Land use            | land_use_soil_area_chang                   | Sectoral  | Land use|                          | 420   | 253         |
| Permit market       | price_et_market_trade_polic                | Policy    |        |                           | 419   | 241         |
| Vehicle             | vehicl_fleet_car Hybrid_passeng             | Sectoral  | Transport|                         | 387   | 227         |
| Coal                | coal_fire_plant_natur_phase                | Sectoral  | Energy |                           | 360   | 126         |
| Buildings           | build_residenti_stock_sector_construct     | Sectoral  | Buildings|                         | 356   | 180         |
| Agriculture         | agricultur_food_product_livestock_farm      | Sectoral  | Land use|                          | 351   | 188         |
| Bioenergy           | biomass_bioenergi_biofuel_residu_wood       | Sectoral  | Energy |                           | 345   | 160         |
| Urban               | urban_citi_area_popul_develop              | Policy    |        |                           | 331   | 197         |
| Wind/Solar          | wind_solar_capac_instal_photovolta          | Sectoral  |        |                           | 324   | 107         |
| Forest              | forest_sequestr_water_sink_stock           | Sectoral  | Land use|                          | 314   | 188         |
| Crop yield          | crop_yield_soil_wheat_fertil               | Sectoral  | Land use|                          | 302   | 113         |
| Water               | water_resource_basin_river_irrig           | Impacts   |        |                           | 291   | 133         |
| Flood risk          | flood_risk_coastal_sea_disast              | Impacts   |        |                           | 284   | 127         |
| Heat pump           | heat_pump_district_cool_boiler             | Sectoral  | Buildings|                         | 272   | 94          |
| CCS                 | ccs_storage_captur_geolog_plant            | Technology|        |                           | 251   | 127         |
| Waste               | wast_landfill_solid_municip_treatment       | Sectoral  | Waste  |                           | 222   | 73          |
| Tax                 | tax_revenu_polic_equilibrium_model         | Policy    |        |                           | 213   | 117         |
| Oil                 | oil_product_crude_natur_export             | Sectoral  | Energy |                           | 200   | 72          |
| Peak                | peak_around_reach_earlier_non              | Policy    |        |                           | 173   | 65          |
| Nuclear             | nuclear_plant_power_mix_new                | Sectoral  | Energy |                           | 160   | 87          |
| Steel/Iron          | steel_iron_product_materi_save             | Sectoral  | Industry|                          | 105   | 70          |
| Hydrogen            | hydrogen_cell_chain_produc_product         | Sectoral  | Energy |                           | 94    | 60          |
| Drought             | drought_precipit_sever_frequenc_index      | Impacts   |        |                           | 89    | 39          |

Note: pm2 is the word stem resulting from ≪PM2.5≫ after text preprocessing.

(iii) sectors; (iv) air pollution; and (v) climate change impacts (Drought, Flood and Crop Yield). The latter are not all primarily about mitigation, as the search equation also picks forward-looking impact assessment or adaptation study at national level that have in the abstract the word ‘mitigation’ or a demonym.

Using the outline of the IPCC Working Group III 6th Assessment Report as a rough mapping of the topics associated with mitigation (table 1, column 5), one can see that the forward-looking mitigation papers at national level cover all IPCC WGIII AR6 sectoral chapters (6–11) as well as issues related to demand (5), policies (13) and innovation (16). The absence of international policies (Chapter 14) is understandable since the search equation focuses on mitigation at national level. The absence of a topic related to finance (Chapter 15), on the other hand, confirms anecdotal evidence that few forward-looking national mitigation pathways have been analyzed along that line so far. Finally, the lack of a standalone topic dedicated to SDGs (Chapter 17) may be related to the fact that individual SDGs are discussed in the abstracts, it may be in a diffuse way that does not get picked up in a topic (except for Air Pollution). Among the sectors that are represented, there is a considerable imbalance: energy is the one with the largest number of related papers with threshold of 0.02 (column T0.02) and with threshold of 0.02 plus keywords in title (column T0.02 Title).
of related papers (27%) followed by land use, land-use change and forestry (LULUCF) (9%), while the other sectors are far less represented. Although the attribution of topics to particular sectors may be debatable in some cases (for example, bioenergy could also be related to LULUCF), and though sector-specific information may also be present in the body of the paper, the observation of an imbalance between sectors appears robust.

### 3.4. Topics reflect country circumstances

The distribution of countries for individual topics mostly reflects the overall distribution of countries in the database (see figures A3 and A4). At one end of the spectrum, China has the largest number of papers for all topics except Heat Pump (preceded by the UK), Nuclear (preceded by Japan, the UK and South Korea) and Hydrogen (preceded by Germany, Japan and the UK). At the other end of the scale, African countries appear only once in the top five for a topic (Ethiopia for Drought). There are, however, differences across topics. Forward-looking mitigation studies of industrial sectors (Cement or Steel-Iron) have been conducted predominantly for China, while the distribution of papers across countries is much more balanced for topics, such as Renewable Energy or Buildings. The imbalance in research across countries in the Urban topic is particularly surprising since urban development issues are not confined to China. It is, however, consistent with the finding of Lamb et al (2019) that urban case studies in China overwhelmingly dominate the literature.

Figure 4 maps the distribution of topics for the 55 most represented countries in the database. The

| Country          | Topic Representation by country (in %) |
|------------------|---------------------------------------|
| China            |                                        |
| United Kingdom   |                                        |
| European Union   |                                        |
| India            |                                        |
| Japan            |                                        |
| Germany          |                                        |
| Brazil           |                                        |
| Indonesia        |                                        |
| South Korea      |                                        |
| Malaysia         |                                        |
| Thailand         |                                        |
| Canada           |                                        |
| Australia        |                                        |
| Italy            |                                        |
| Finland          |                                        |
| Poland           |                                        |
| France           |                                        |
| Russia           |                                        |
| Spain            |                                        |
| Sweden           |                                        |
| Switzerland      |                                        |
| Norway           |                                        |
| UK               |                                        |
| Netherlands      |                                        |
| South Africa     |                                        |
| Portugal         |                                        |
| New Zealand      |                                        |
| Turkey           |                                        |
| Ireland          |                                        |
| Nigeria          |                                        |
| Greece           |                                        |
| Norway           |                                        |
| Viet Nam         |                                        |
| Bangladesh       |                                        |
| Pakistan         |                                        |
| Palestinian      |                                        |
| Austria          |                                        |
| Cyprus           |                                        |
| Colombia         |                                        |
| Colombia         |                                        |
| Ethiopia         |                                        |
| Nepal            |                                        |
| Denmark          |                                        |
| Romania          |                                        |
| South Africa     |                                        |
| Ecuador          |                                        |
| Kenya            |                                        |
| Ghana            |                                        |
| Philippines      |                                        |
| Ireland          |                                        |
| Latvia           |                                        |
| Georgia          |                                        |
| Croatia          |                                        |
| Kocaelen         |                                        |
| Belgium          |                                        |
| Kuwait           |                                        |

Figure 4. Topic distribution in the 55 most represented countries in the database. Topics, from left to right, and countries, from top to bottom, are presented in descending order of their representation in the database. Topic representation per country (in %) is the number of country studies associated to the topic divided by the total number of country studies. Since individual papers can be associated to several topics, the sum of the topic representations is not equal to 100. The Policy-Devlpt-Eco topic is not included, since it is present in almost every paper. Scenario and System topics are also not included since they refer to the method of the paper rather than to the issues that the paper addresses.
number in each cell is the number of papers devoted to country x and related to topic k. The shade of the cell indicates the share of the topic in the total number of papers devoted to the country. Reading the figure vertically provides a view of the relative importance of a given topic across countries.

Patterns emerge. Some topics appear in 10% or more of the papers in nearly all 55 countries, such as Energy Efficiency, Electricity or Power. Others stand out only in a limited set of countries, such as Oil in Saudi Arabia, Kuwait, Canada, Malaysia, Indonesia, Ecuador, Mexico and Austria, Coal in China, India, Australia, Malaysia, Poland, South Africa, Vietnam, Chile and Czechia or Nuclear in Japan, Korea, France and Romania. It is not surprising that topics that are related to country-specific circumstances (e.g. fossil fuel endowments or share of nuclear in electricity mix) appear in a smaller set of countries than topics that relate to broadly shared elements of mitigation (e.g. the electricity grid or the power sector). However, the list of countries where ‘specialized’ topics appear, suggests gaps in the literature. For example, the Oil topic does not stand out in major oil exporting countries such as Nigeria, Russia or Norway. Regarding the ‘policy’ topics, Costs and Target/INDC appear evenly distributed across countries. Permit market is often present, notably in papers devoted to the EU. On the other hand, Tax is poorly represented (five countries present this topic in more than 10% of related publications), reflecting at least a higher degree of attention in the academic literature to the former relative to the latter.

Reading the figure horizontally provides a country-by-country snapshot of the issues identified by the academic literature as most important in the context of future mitigation. Some countries have balanced literature that covers nearly every topic, while others have much more ‘specialized’ literature. For instance, literature on Indonesia is balanced, with three major topics (Energy Efficiency, Target/INDC and Land Use) plus eight other topics including power generation, forest and oil. The literature on Poland, on the other hand, is more focused on Power, Costs, Renewable Energies and Coal. Although the former are mostly countries with a large number of papers attached (China, the U.S., the EU) and the latter are by construction mostly countries with a smaller number of papers, the relationship does not necessarily hold everywhere. Portugal or Ireland, for example, have less yet more balanced literature than Japan (with a higher than average number of papers devoted to Nuclear and Hydrogen) or Brazil (Forest, Land Use, Agriculture).

Patterns of countries also emerge, based on natural resource endowments (e.g. Brazil, Canada, Indonesia, Finland, New Zealand, Norway and Ghana all having higher-than-average papers on Forest and LULUCF-related topics), technology (e.g. Japan, Germany, Italy, France, Denmark and Norway on Hydrogen) or specific policies (e.g. Tax in South Africa and Switzerland). Forward-looking studies related to the impacts of climate change are particularly frequent (in relative terms) in some countries, mostly in the global south (e.g. Ethiopia, Pakistan, Bangladesh and Nigeria for Drought). Finally, for each country, it is also interesting to examine topics that are not addressed. Some may just be less relevant in that particular context (e.g. Coal in France). Others may point to gaps in the literature. For instance, one may argue that given their importance for the French 2050 net zero target, Forest and Bioenergy are currently under-represented in the literature on France.

3.5. Models are mainly identified in studies devoted to Asia

Finally, we attempt to analyze the methods used in the papers to study mitigation at country level. This is not easy given the limited amount of information present in the metadata. We focus on models, checking metadata against a database of 80 scenario modeling tools for national pathways to the SDGs (Allen et al 2016) and the list of 48 models documented by the IAMC. We identify the model names in only 16% of the abstracts (734). Compared with the country coverage of the overall data set (figure 1), Asia is even more represented (60.3%) in this subset. For example, Thailand is three times more represented than in the general database (6.2% against 2.2%) (see figure F8). At the other end of the spectrum, Africa is scarcely present (4.7%) with only 13 countries represented. Model-wise, the CGE model is the most common category of models in the corpus. The three individual models that dominate, LEAP, TIMES and MARKAL are all bottom-up. They are highly used for Asia and Europe (figure A9). Unsurprisingly, these three models are mainly used for energy-related questions (top three for topics Energy Efficiency, Electricity, Power, Fuel, Transport, Vehicle, Renewable Energies) (figure A10). Finally, it is interesting to note that the Japanese AIM model is present in 62 publications (of which 59 are in Asia, given its many regional spin-offs AIM Korea, AIM Vietnam, etc), illustrating the importance of regional clusters.

However, these findings are limited to an arguably small sample of papers that name their model (or model type) in the abstract and whose models are state-of-the-art tools in our reference list. The term ‘model’ is actually present in 52% (2446) abstracts and characterizes mainly, as well as the terms ‘scenario’, ‘bau’, ‘refer’ and ‘three’, the topic Scenario. This topic is represented in 30% (1428) publications from the database and associated to the ‘Method’ category as it characterizes papers detailing the methodology in the abstract. To better identify the modeling tools used in these studies, a deeper analysis of the publications based on the full text is needed. Although limited to a small sample of papers, these findings
nonetheless emphasize again the inequalities between countries.

4. Conclusion

In this paper, we provide the first mapping of the forward-looking mitigation literature at country level, using systematic mapping techniques. We find the number of papers per country is well correlated with current levels of GHG emissions, with few papers for (current) low emitters. Time horizons of 2030 and 2050 each account for one-third of the papers, with the former more frequent in recent years, spurred by interest in the (1)NDCs. Topic modeling analysis of the data set reveals that forward-looking mitigation papers encompass all dimensions of mitigation, save for financial issues, which are lacking. However, energy and to a lesser degree LULUCF are very dominant relative to other sectors. These topics are unevenly addressed across countries, reflecting national circumstances and priorities, but also pointing to gaps in the literature.

From a methodological point of view, this paper builds upon and improves on Lamb et al (2019) by providing a systematic way to maximize the accuracy of topic modeling. It also illustrates how topic modeling can complement traditional methods of evidence synthesis. Precisely, most systematic reviews are based on a search query that yields thousands of publications. These are then screened to set irrelevant papers aside and scale down the number of papers to a manageable level. Here, topic modeling is used to aid the screening process and provide an overview of the publications identified along all the steps of the database construction.

This paper has three main limitations. First, the term mitigation (or its demonyms) that we use in the search equation harvests too broad a set of papers, since papers about impacts and adaptation to climate change may still refer to mitigation in the abstract. Deep interactions between mitigation and adaptation make this limitation difficult to overcome. Next, despite instructions by journals, abstracts remain written in very different ways across papers. For the purpose of textual analysis, abstracts that are as close as possible to the method and key findings of the paper are preferable, though that may come at the expense of readability. General sentences providing context about climate change may be easily recognizable as such in a full paper (of which they would only represent a tiny fraction), whereas in an abstract they may be confused with a substantive result of the paper. The ubiquitousness of the Climate Change topic is a demonstration of that risk. Third, as all analysis is based on metadata (abstract, title and keywords), we may miss relevant material that does not make it to the abstract. For example, we cannot rule out that non-energy sectors (e.g. industry or transport) are discussed more frequently in the body of the paper than the abstract suggests.

Our attempt to survey the methods used for conducting forward-looking mitigation studies is limited by the fact that detailed methods, let alone model names, are not systematically presented in the abstract. An overview of national mitigation models, similar to the overview of global mitigation models supported by the Integrated Assessment Modeling Consortium, would be helpful to complement this first attempt.

Finally, our paper has policy implications, as forward-looking mitigation studies typically aim at informing decisions, notably in the context of the Paris Agreement. Where we find such papers scarce, policymakers and stakeholders do not benefit from this source of insight. This is all the more regrettable as these countries, typically with low emissions at present, may still have options to avoid getting into high emission paths. Possible explanations for our findings include a lack of domestic research capacity, lack of data or lack of interest or incentive for foreign research teams to work on other national contexts. In any case, our paper adds quantitative evidence to existing qualitative analysis of the capacity to prepare forward-looking climate policies (e.g. UNFCCC (2019)). In addition, it also provides a basis to further explore the reasons for discrepancies across countries.

Data availability statement

The data that support the findings of this study are openly available at https://github.com/ClaireLepault/Mapping-country-mitig-pathways, with the exception of the abstracts from Scopus and Web of Science, which can not be published for copyright reasons. However, the code and explanation are fully provided to reproduce the analysis and obtain complete databases. In particular, a tutorial on the systematic search on Scopus and Web of Science at https://github.com/ClaireLepault/systematic-search-wos-scopus shows how to download the initial metadata databases from WoS and Scopus.

Acknowledgments

We warmly thank two anonymous reviewers who contributed to improve the quality of the manuscript. Claire Lepault acknowledges funding from the French environment and Energy Management Agency (Ademe) (research Contract 18MAR000099186715 ‘State of the Art of prospective modeling’, 2018–2019).

Code Availability

The complete code used to produce this paper is available in Jupyter notebook format. The entire process, from the search query to the figures is
explained and clearly reproducible using Python and R at https://github.com/ClaireLepault/Mapping-country-mitig-pathways.

ORCID iDs

Claire Lepault @ https://orcid.org/0000-0002-4976-2319
Franck Lecomq @ https://orcid.org/0000-0003-4875-1156

References

Alexandre-Benavent R, Aleixandre-Tudó J L, Castelló-Cogollos L and Aleixandre J L 2017 Trends in scientific research on climate change in agriculture and forestry subject areas (2005–2014) J. Cleaner Prod. 147 406–18
Allen C, Metternicht G and Wiedmann T 2016 National pathways to the sustainable development goals (SDGs): a comparative review of scenario modelling tools Environ. Sci. Policy 66 199–207
Belford M, Mac Namee B and Greene D 2018 Stability of topic modeling via matrix factorization Expert Syst. Appl. 91 159–69
Belter C W and Seidel D J 2013 A bibliometric analysis of climate engineering research Wiley Interdiscip. Rev.: Clim. Change 4 417–27
Boutsidis C and Gallopoulos E 2008 SVD based initialization: a head start for nonnegative matrix factorization Pattern Recognit. 41 1350–62
Callaghan M W, Minx J C and Forster P M 2020 A topography of climate change research Nat. Clim. Change 10 118–23
CD-LINKS 2019 Linking climate and sustainable development: policy insights from national and global pathways Int. Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria (available at: https://pure.iiasa.ac.at/id/eprint/16235/ (CD-Links-for-web_final_November_2019.pdf))
COMMIT 2019 COMMIT deliverable D2.2: long-term, low-emission pathways in Australia, Brazil, Canada, China, EU, India, Indonesia, Japan, Republic of Korea, Russia, and United States The Hague: PBL. Netherlands Environmental Assessment Agency (available at: https://themesites.pbl.nl/commit/wp-content/uploads/COMMIT-Long-term-Low-emission-pathways-in-Australia-Brazil-Canada-China-EU-India-Indonesia-Japan-Republic-of-Korea-Russia-USA-2.pdf)
Donnelly C A, Boyd I, Campbell P, Craig C, Vallance P, Walport M, Whitty C J, Woods E and Wormald C 2018 Four principles to make evidence synthesis more useful for policy Nature 558 361–4

Fragkos P et al 2021 Energy system transitions and low-carbon pathways in Australia, Brazil, Canada, China, EU-28, India, Indonesia, Japan, Republic of Korea, Russia and the United States Energy 216 119385
Haddaway N R and Macura B 2018 The role of reporting standards in producing robust literature reviews Nat. Clim. Change 8 446–7
Haunschild R, Bornmann L and Marx W 2016 Climate change research in view of bibliometrics PLoS One 11 7
Lamb W F, Creutzig F, Callaghan M W and Minx J C 2019 Learning about urban climate solutions from case studies Nat. Clim. Change 9 279–87
Lee D D and Seung H S 1999 Learning the parts of objects by non-negative matrix factorization Nature 401 788–91
Li W and Zhao Y 2015 Bibliometric analysis of global environmental assessment research in a 20-year period Environ. Impact Asses. Rev. 50 158–66
McKinney W 2010 Data structures for statistical computing in Python Proc. 9th Python in Conf. ed van der W Stéfan and J Millman pp 56–61
Mikolov T, Chen K, Corrado G and Dean J 2013a Efficient estimation of word representations in vector space (arXiv:1301.3781)
Mikolov T, Sutskever I, Chen K, Corrado G S and Dean J 2013b Distributed representations of words and phrases and their compositionality Advances in Neural Information Processing Systems pp 3111–19
Minx J C, Callaghan M, Lamb W F, Garard J and Edenhofer O 2017 Learning about climate change solutions in the IPCC and beyond Environ. Sci. Policy 77 252–9
O’Callaghan D, Greene D, Carthy J and Cunningham P 2015 An analysis of the coherence of descriptors in topic modeling Expert Syst. Appl. 42 5645–57
Salton G and Buckley C 1988 Term-weighting approaches in automatic text retrieval Inf. Process. Manage. 24 513–23
Sievert C and Shirley K 2014 LDAvis: a method for visualizing and interpreting topics Proc. Workshop on Interactive Language Learning, Visualization, and Interfaces 63–70
UNFCCC 2019 National-level pilot exercise on capacity gaps and needs related to the implementation of nationally determined contributions Technical Report, United Nations Framework Convention on Climate Change (UNFCCC), Paris Committee on Capacity-building (PCCB) (available at: https://unfccc.int/sites/default/files/resource/PCCB_TP_capacity%20gaps%20and%20needs_NDCs_final.pdf)
Waissman H et al 2019 A pathway design framework for national low greenhouse gas emission development strategies Nat. Clim. Change 9 261–8
Wang B, Pan S-Y, Ke R-X, Wang K and Wei Y-M 2014 An overview of climate change vulnerability: a bibliometric analysis based on web of science database Nat. Hazards 74 1649–66