Article:
Lesnikowski, A, Belfer, E, Rodman, E et al. (5 more authors) (2019) Frontiers in data analytics for adaptation research: Topic modeling. Wiley Interdisciplinary Reviews: Climate Change, 10 (3). e576. ISSN 1757-7780

https://doi.org/10.1002/wcc.576

Frontiers in data analytics for adaptation research: Topic modeling

Alexandra Lesnikowski, Ella Belfer, Emma Rodman, Julie Smith, Robbert Biesbroek, John D. Wilkerson, James D. Ford, Lea Berrang-Ford

Cite as: Alexandra Lesnikowski, Ella Belfer, Emma Rodman, Julie Smith, Robbert Biesbroek, John D. Wilkerson, James D. Ford & Lea Berrang-Ford (2019). Frontiers in data analytics for adaptation research: Topic modeling. WIREs Climate Change, 10(3): e576.

Abstract

Rapid growth over the past two decades in digitized textual information represents untapped potential for methodological innovations in the adaptation governance literature that draw on machine learning approaches already being applied in other areas of computational social sciences. This article examines the potential for text mining techniques, specifically topic modeling, to leverage this data for large-scale analysis of the content of adaptation policy documents. We provide an overview of the assumptions and procedures that underlie the use of topic modeling and discuss key areas in the adaptation governance literature where topic modeling could provide valuable insights. We demonstrate the diversity of potential applications for topic modeling with two examples that examine: (a) how adaptation is being talked about by political leaders in United Nations Framework Convention on Climate Change; and (b) how adaptation is being discussed by decision-makers and public administrators in Canadian municipalities using documents collected from 25 city council archives.

6.1 Introduction

Text-based research methods have been a cornerstone of qualitative social science methods since the 1950s (Lasswell 1952). These approaches see documents as meaningful artifacts that can be analyzed for their thematic and semantic content (Krippendorff 2013), and they form a core component of the climate change adaptation governance literature. In lieu of directly observable and measurable indicators such as greenhouse gas emissions, adaptation governance research relies on written records, surveys, and interviews as its primary information sources about how different actors are responding to climate change impacts. Content analysis methods are commonly applied to sources such as strategic planning documents, government reports, peer-reviewed and grey literature, and media stories (Lesnikowski et al. 2016; Araos et al. 2016; Ford et al. 2015; Labbé et al. 2017; Belfer, Ford, and Maillet 2017; Biesbroek et al. 2018). These studies indicate a growing number of adaptation policies, programs, and interventions being adopted in the public sector to address current and projected risks.

The reliance on hand-coding textual data sources, however, has two major limitations. First, its use in large comparative analyses is constrained by the limited volume of documentation that can reasonably be analyzed using manual techniques. This challenge is becoming increasingly relevant with the proliferation of ‘big data’ sources such as social media or digitized legislative records (Beelen et al. 2017). The adaptation governance literature is certainly not alone in this challenge; computational tools for extracting data from large volumes of text are increasingly being used across the humanities and social sciences, where most data available to researchers are in the form of text (Benoit, Laver, and Mikhaylov 2009; DiMaggio, Nag, and Blei 2013; Shim, Park, and Wilding 2015; Laver and Benoit 2003).

Second, the design of research protocols for manual content analysis often relies on the *a priori* determination of conceptual categories, which is challenging given the mutable and contested nature of key concepts in adaptation governance (Levin et al. 2012; Pollitt 2015; Head 2014), the fuzziness of adaptation as a distinct problem from issues like risk management (Dabrowski 2017; Hetz 2016; Viguié and Hallegatte 2012; Uittenbroek, Janssen-Jansen, and Runhaar 2013; Bauer and Steurer 2014; Wamsler and Pauleit 2016), and differences in the understanding and use of these concepts across places and sectors (Keenan, King, and Willis 2016; Dupuis and Knoepfel 2013). While identification and classification of adaptation in stand-alone climate policies is relatively straightforward, identifying adaptation-relevant policies from related domains such as water management or sustainable development is a key limitation in current content analysis approaches (Dupuis and Knoepfel 2013).

These limitations have significant implications for what gets ‘counted’ as adaptation, and have generated debate about the extent to which existing datasets are representative of the approaches that different actors are taking to address adaptation (Craft and Fisher 2018). Issues of reporting bias in document retrieval and analysis pose challenges for the validity of results from manual content analysis. A larger empirical investigation of how policy-makers talk about adaptation and position it relative to intersecting policy issues would nuance our interpretations of textual data and improve future research designs that use code-based analysis. Balancing feasibility, representativeness, and conceptual validity in methodological approaches is thus a major challenge for adaptation governance research (Ford et al., 2015), but the rapid increase of information available through government websites, legislative databases, academic databases, and internet search engines provides an opportunity to integrate text mining research techniques into adaptation governance research that can help make sense of this complexity (Ford et al., 2016).

We argue here that the ability to efficiently analyze large volumes of text could contribute important insights on adaptation governance practices across contexts, revealing relationships between ideas and issues or even uncovering new ways of thinking about adaptation. This could shed light on how key concepts or themes are understood in policy documents or grey literature, and how consistent the conceptual categories and definitions used in adaptation governance research are with their use by practitioners and decision-makers.

The absence of text mining approaches in adaptation governance research suggests a lack of awareness around computational text techniques. The integration of methods from other disciplines into adaptation research is observable in the case of systematic review protocols, which were developed in the health sciences and are increasingly popular for synthesizing emerging evidence around adaptation policies and practices (Berrang-Ford, Pearce, and Ford 2015). Here we demonstrate the untapped potential of computational text methods to address the limitations of manual analysis.

We focus on one text mining technique in particular: topic modelling. Topic models are statistical models that use unsupervised machine learning algorithms to discover the existence and distribution of 'topics' across a body of documents based on word frequencies and co-occurrences. This technique can be understood as a form of automated content analysis, which can be helpful for interpreting the content of documents given questions such as:

- How do politicians, policy-makers, or private sector actors talk about adaptation, and how has this changed over time?
- In what context(s) is adaptation talked about?
- How is interest in, and discourse around, adaptation evolving?

- How can we conceptualize adaptation as a relational construct that is sensitive to place, scale, and time?

A number of recent papers discuss applications – and potential perils – of topic modelling in social science and environmental science research (Hillard, Purpura, and Wilkerson 2008; Grubert and Algee-Hewitt 2017; Wiedemann 2013; Quinn et al. 2010; Vilares and He 2017; Wilkerson and Casas 2017; Grimmer and Stewart 2013). Nonetheless, topic modelling has barely permeated the climate change literature, with the majority of existing examples limited to studies that use social media data to analyze coverage of climate change issues (Jang and Hart 2015; Kirilenko and Stepchenkova 2014; Cody et al. 2015; Williams et al. 2015), including skepticism and belief about climate change (Boussalis and Coan 2016; Elgesem, Steskal, and Diakopoulos 2015; Farrell 2016), and social representations of adaptation (Lynam 2016; Lynam and Walker 2016). Applications of topic modelling for adaptation research are thus largely unexplored, despite the potential to expand text-based analysis to much larger scales than is currently possible. This has the potential to make significant contributions to the study of adaptation governance, both with regards to exploratory research and hypothesis generation, and for adaptation tracking.

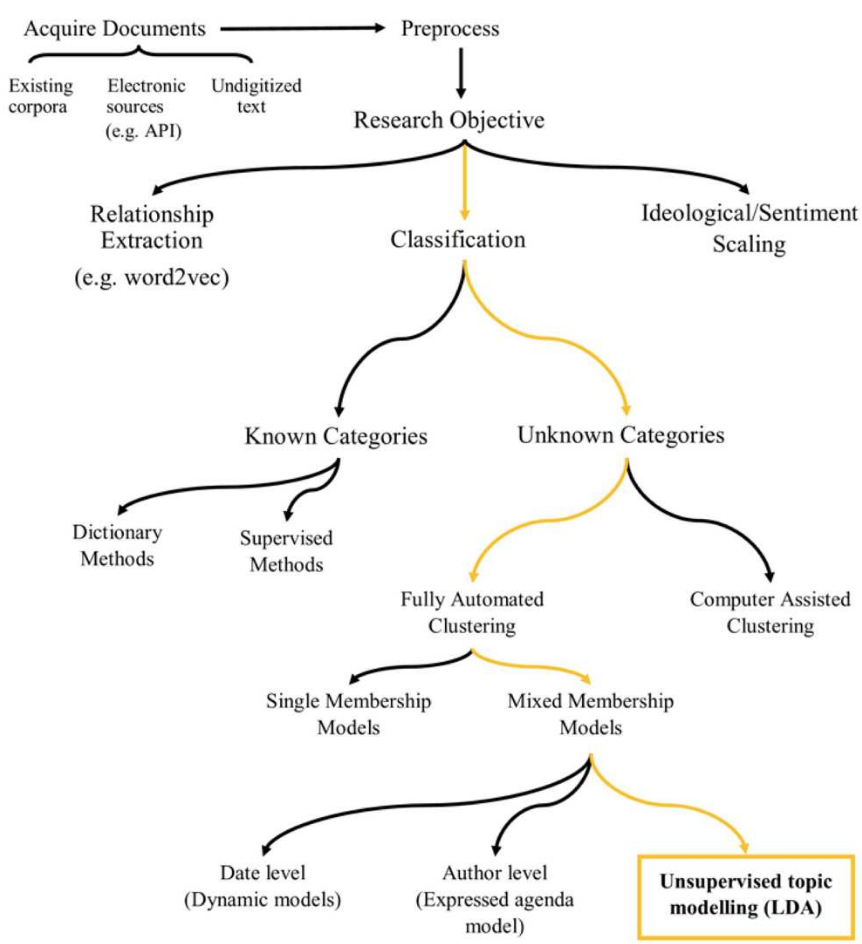
The following section elaborates on the key ideas and assumptions underlying topic modelling. We then demonstrate the topic modelling process using two examples. The first example analyzes speeches given by country representatives to the United Nations Framework Convention on Climate Change (UNFCCC) at the beginning of the annual Conference of the Parties (2010-2016), providing insight into how the issue of adaptation is discussed by politicians within the UNFCCC negotiations. The second example uses city council meeting minutes and staff reports for the 25 largest cities in Canada to analyze how adaptation policy is being approached by Canadian local governments. These two examples demonstrate: i) that topic modelling can be applied to different scales of analysis; ii) diverse types of text can be analyzed using this method; and iii) there are multiple approaches to implementing topic models and assessing model robustness when selecting and validating models. We conclude with a discussion on areas in the adaptation governance field where this approach could be applicable.

6.2 An introduction to topic modelling

Over the past two decades, text mining approaches have proliferated in social science research (Grimmer and Stewart 2013; Hopkins and King 2010). A primary benefit of text mining is the ability to scale up text analysis to sort and categorize large volumes of data that would otherwise require resource-intensive hand-coding (Jelodar et al. 2018; Quinn et al. 2010). Accordingly, it is particularly valuable in exploratory research, where little is known about a dataset, and researchers are interested in discovering unknown patterns or trends in the data or are seeking external validation of inductively determined categories. Recent advances in topic modelling also mean that this approach can also be used for research of a more deductive nature, supporting development of hypothesis-based models that use information such as document author, scale, location, or relationships between documents to understand topic results (Blei & Lafferty, 2006a; Chang & Blei, 2009; Mcauliffe & Blei, 2008; Rosen-Zvi, Griffiths, Steyvers, & Smyth, 2004; Yin, Cao, Han, Zhai, & Huang, 2011). Nevertheless, it remains essential that researchers externally validate the results of such models, including bringing subject matter expertise to bear on the substantive interpretation of model results (Grimmer and Stewart 2013).

At its most fundamental level, text mining extracts information about structures and patterns from large volumes of text, such as word frequency or correlation between words. This approach can be used for various applications (Figure 6.1). For example, common uses for text mining in social science research are the classification, clustering, and analysis of word patterns in texts (Bickel 2017), and the extraction of semantic meaning from text, for example with regards to the identification of sentiment or emotion (Onyimadu et al. 2013; Ravi and Ravi 2015; Cambria et al. 2013), the positions held by political parties or individuals on a given issue (Will et al. 2011; Laver and Benoit 2003), or the evolution of document content over time (Allee, Elsig, and Lugg 2017; Wilkerson, Smith, and Stramp 2015) (see Grubert and Siders 2016 for a more extended review of text mining approaches in the environmental sciences).

Figure Error! No text of specified style in document.**.1** Topic modelling as a text mining technique



Adapted from Grimmer and Stewart (2013)

Topic modelling deals with the problem of document classification using themes (i.e. topics) contained in each document (Figure 1). It produces a generative probabilistic model that relies on three analytical layers: i) a collection of documents for analysis, referred to as a corpus; ii) the individual documents within the corpus; iii) and the individual words within each document. Essentially, the model assumes that a particular corpus contains some pre-existing set of topics, and that each document within the corpus contains some mix of these topics. Each topic has a set of words most strongly associated with that topic, which are identified based on the probability of co-occurrence between words.

The topic model will thus generate three observations: i) lists of words that are most important to a particular topic; ii) the topics that are most important to any particular document within a corpus; and iii) a set of topics that characterize an entire corpus. Topic models can be single-membership, where each document can belong to a single topic (Grimmer 2010; Quinn et al. 2010), or mixed-membership, where each document is assumed to be composed of multiple topics (Blei, Ng, and Jordan 2003). After the model identifies a set of topics in a corpus, researchers interpret and label these topics. For example, a collection of parliamentary speeches might contain words such as ‘hospital,’ ‘doctor,’ and ‘medicine,’ which a researcher might interpret as broadly related to health. Similarly, terms such as ‘emissions,’ ‘resources,’ and ‘green’ could be interpreted as concerning the environment. The topic model examines the frequency of co-occurrence between these words; the algorithm will then predict if a particular speech that discusses the public health implications of climate change has a high prevalence of both the health and environment topics, relative to words associated with other topics such as 'economy', or 'military.'

Several important assumptions underlie the most common types of topic models (e.g. latent Dirichlet analysis). First is the 'bag-of-words' assumption (BoW), which states simply that the order of words in a document is irrelevant, and language particularities such as syntax and grammar can be ignored. Essentially, this means that the model does not ascribe inherent meaning to words; rather, meaning is derived from the frequency of word appearance in documents, and relative to other words within a single document. In processing a topic model, a simplified representation of a corpus is produced in the form of a word-document matrix, which specifies the frequency of each word over each document (Liu et al. 2016). In some cases, however, word order can be central to topic identification and interpretation; hierarchical topic modelling techniques have been developed to overcome the BoW assumptions, which assume that words within a topic are conditional on the previous word and use bigrams rather than unigrams (Wallach 2006). The extent to which the BoW assumption is appropriate to the topic modelling task in question is for researchers to consider when selecting a topic modelling algorithm (Blei 2012).

Second, all topic models assume that the number of topics (denoted by the letter *k*) is fixed, and derives this information based on instructions from the researcher about the number of topics to search for. Selecting *k* is a critical step in topic modelling and implies that while topic models are considered an unstructured form of machine learning, they still require input and interpretation from the researcher. In short, there is never any entirely automated topic model. Various techniques are available to assist in the selection of *k*. Strictly mathematical approaches to *k* selection calculate the log-likelihood of held-out training and testing documents and identify how well the model predicts topics in the test set. This approach is based on maximizing model fit, however, and has been shown to not necessarily correlate well with human judgment (Chang et al. 2009). Selecting the number of topics to run in a topic model therefore requires some level of researcher judgement and iteration. As guiding principles for model selection, Roberts et al. suggest that *k* identification should be guided by the *cohesiveness* of the topics (meaning that high-probability words co-occur within documents), and the *exclusivity* of the topics (meaning the likelihood that top words for each topic also appear in other topics) (Roberts et al. 2014).

There are a number of topic model algorithms available, and they make additional assumptions of which researchers need to be aware (Alghamdi 2015). In the examples described here, we apply a latent Dirichlet algorithm (LDA), which is one of the most commonly used topic models in the social sciences and available to new topic modelling users through various R packages, and an LDA variation called a robust latent Dirichlet algorithm (Jelodar et al. 2018; Liu et al. 2016; Grubert and Algee-Hewitt 2017; Goldstone and Underwood 2012; Mimno and David 2012; Wilkerson and Casas 2017). Our first example (COP speeches) uses the *Topic Models* R package (Grün and Hornik 2011), a LDA model explained by Blei et al. (Blei, Ng, and Jordan 2003). Our second example uses a robust latent Dirichlet allocation model (*rlda* package in Python), which builds on the LDA model by using a spectral clustering algorithm to identify K. The explanation for this approach can be found in Wilkerson and Casas's study of United States Congressional floor speeches (Wilkerson and Casas 2017).

Similar to the BoW assumption, LDA makes an assumption that the order of documents in a corpus is irrelevant and all documents are independent from one another and non-hierarchical (Blei, Ng, and Jordan 2003). For simpler research questions this assumption may be appropriate, but in other cases it may not hold, for instance in longitudinal research where we would like to know how topic prevalence changes over time (Grubert 2018). For these cases, LDA has been adapted into various other algorithms that can perform different functions, such as taking into account sequences of distributions over topics. Dynamic topic models, for example, allow the researcher to identify documents by increments of time (e.g. years) and look longitudinally at how topics change over time (Blei and Lafferty 2006b). Correlated topic models examine the relationship between topics to show where the existence of one topic is correlated with the existence of another (Blei and Lafferty 2006a; Roberts et al. 2014). The appropriateness of these models will vary depending on research questions of interest and document characteristics.

In preparing a corpus for analysis, the researcher must also deal with the various idiosyncrasies of document sets. Dissection of documents into document-term-matrices requires simplification of text, such as translation into the same language, removal of numbers, punctuation, and symbols, elimination of very common words (stopwords) with little substantive meaning (e.g. 'it', 'and', 'or', 'he', 'she') or very rare words, and stemming of similar words (e.g. stemming 'adaptation', 'adaptive', 'adapting' to 'adapt'). These pre-processing steps aim to balance simplification of the complexity inherent to textual data with interpretability, and have implications for the results generated from a topic model (de Vries, Schoonvelde, and Schumacher 2018; Denny and Spirling 2018). As such it is critical that the researcher be aware of how the pre-processing stage can affect their results. The following section details the pre-processing steps taken in the two examples presented here.

6.3 Implementing an LDA model

Language is highly complex and requires simplification for algorithmic analysis. Generating an output from a topic model requires several steps, including i) data collection, ii) document pre-processing, iii) corpus processing, and iv) interpretation (see Table 6.1 for a summary of steps). Here we provide an overview of these steps (see Appendix C for additional details).

Table Error! No text of specified style in document.**.1** Summary of approach

Stage	**Steps**	**Case 1: COP speeches**	**Case 2: Canadian local government records**
1. Model selection	*1. Specify research question*	How do country leaders talk about adaptation within the UNFCCC process?	How is adaptation being approached among local governments in Canada?
	2. Select algorithm	LDA	Robust LDA
2. Data collection	*3. Identify data source*	UNFCCC website	City Council online archives
	4. Document type	Speech	Council minutes, staff reports, strategic planning documents, by-laws
	5. Format	Machine-readable PDFs, scanned text	Machine-readable PDFs, scanned text
	6. Language	English, Spanish, French, Arabic, Russian	English, French
3. Pre-processing	*7. Translation (to English)*	Tesseract engine and Google Translate	Google Translate
	8. Text extraction	Selection of thirty words surrounding any mention of ‘adapt*’	200-word window surrounding terms inductively identified as relevant to adaptation
	9. Stemming	Yes	No
	10. Stopwords removal	SMART stopwords, plus additional corpus-specific words identified by reviewing top features	SMART stopwords, plus additional corpus-specific words identified by reviewing top features

	11. Additional character removal	Punctuation, separators, numbers and symbols were removed	Punctuation, separators, numbers and symbols were removed
4. Processing	*11. Method of selecting K*	Perplexity used to guide selection of categories with most sematic coherence; K= 25	Spectral Clustering; K=20-40
	12. Meta-topic aggregation	Not applicable	Based on Wilkerson and Casas 2017
5. Interpretation	*13. Topic labels*	Based on discussion by research team	Based on discussion by research team

6.3.1 Data collection

A topic model requires a large corpus of documentation to produce robust results, often on the order of thousands or even millions of texts; where documents are very short or very few in number then LDA will often not perform well (Tang et al. 2014). Where there are many very small documents (e.g. tweets), documents can be grouped by author (Hong and Davison 2010) or conversation (Alvarez-Melis and Saveski 2016) to generate larger documents. A variety of tools, such as application programming interfaces (APIs) or pre-existing databases like digitized parliamentary records, can support researchers in identifying and downloading large volumes of data. Web-scraping tools can also be implemented to construct unique databases of texts. With adaptation policy now widely being adopted into climate change policy agendas, there has been a rapid growth in text available through online archives that may be appropriate for thematic analysis via topic modelling.

In this Focus Article, two types of data are used: i) speeches made by country representatives to the Conference of Parties (COP) of the United Nations Framework Convention on Climate Change (UNFCCC) covering the period from 2010 (COP16 in Cancun) to 2016 (COP22 in Marrakech) (document number=1,315); and ii) city council meeting records with containing references to climate change in Canada's largest 25 cities for the period from January 2010 to May 2017 (document number=1,814). Once these documents were manually collected from online archives, they were streamlined into identical formatting that can be read by a computer (text file format).

6.3.2 Data pre-processing

The texts used in both the examples here include multiple languages and both machine-readable and not machine-readable documents (i.e. non-searchable PDFs). Here, we processed the documents into a readable format using R.

Data pre-processing: COP speeches

In the case of the COP speeches example, translation of non-English texts was completed at this stage using built-in translation capabilities for French and Spanish in the *Tesseract* package in R, and manual translation using Google's Neural Machine Translation for other languages (e.g. Arabic, Russian). COP speeches include both mitigation and adaptation content, so to isolate adaptation content for the topic modelling analysis only the 30 words surrounding each reference to 'adapt*' were extracted from the speeches to create the COP speech corpus.

Data pre-processing: Canadian local government records

The Canadian local government documents contained two added layers of pre-processing complexity. First, it became apparent that the in-text language was more varied than in the COP speeches. Second, in addition to climate change, these documents contained references to a whole range of issues and policies being considered by local governments, resulting in sometimes enormous documents (e.g. pages≥200). We therefore had to isolate adaptation-relevant text from a highly diverse range of content. To address these issues, two of the authors manually identified a list of all adaptation-relevant keywords from within the texts and selected the 400 words surrounding each of these terms to generate the corpus (keywords: adapt*, risk*, protect*, vulnerab*, emergenc*, security, resilien*, recover*, prevent*, hazard*, prepar*, disaster*, impact*, mitigate).

Data pre-processing: Final corpus preparation (both datasets)

The final step was cleaning both corpuses of stopwords. This involves removing words and punctuation symbols with no substantive information (e.g. 'the', 'and', and 'or') to improve topic coherence and reduce computational time (Hoffmann, Bach, and Blei 2010; Boyd-Graber and Blei 2009). The most frequently occurring features of the remaining corpuses were then inspected, and additional stopwords specific to that corpus were identified and removed (e.g. formalities such as 'madame', 'gentlemen', place names, boilerplate terms, procedural terms) (Benoit, Muhr, and Watanabe 2017; Lewis et al. 2004). We observed fewer cases of multiple tenses in the local government corpus as compared to the COP corpus, and so opted not to stem the vocabulary in this model. It is worth noting that there is an ongoing debate regarding the impacts of stemming on model results, with some studies suggesting that stemming can negatively impact topic coherence (Schofield and Mimno 2016). The final size of each corpus was 3,069 unique words for the COP speeches, and 21,243 words for the local government documents.

6.3.3 Processing

After pre-processing the texts but before running the models, the researcher must still provide instructions to the algorithm with regards to one key feature: the number of topics (referred to as '*k*') to be generated. To some extent the choice of how to determine *k* reflects the aim of the research question itself, whether it is to classify documents into known categories or to conduct exploratory research. A purely inductive approach to selecting *k* relies on statistical estimates (perplexity) of topic stability to tell the researcher which model output is most stable. Recall, however, that LDA does not associate semantic meanings with words, so the number of topics chosen by purely quantitative methods may not always generate the most coherent output from perspective of the researcher (Chang et al. 2009). Social scientists therefore tend to follow a 'middle-ground' approach to *k*-selection that combines statistical estimates of topic stability with expert judgement about the interpretability of results with regards to the cohesiveness and distinctiveness of topics (Blei, 2012; Roberts et al., 2014). In the case of adaptation, where debate about the relationships between different concepts like resilience, adaptive capacity, and vulnerability is ongoing, this middle-ground approach also seems likely to provide the greatest likelihood of generating meaningful results. *K*-selection has important implications for establishing the conceptual validity of topic model outputs, an issue we return to later in the discussion.

K-selection: COP speeches

In the case of the COP speeches example, perplexity was measured at a range of *k*-values between $k = \{5, 100\}$ to determine an initial range of suitable *k* -values. The final selection of model parameters followed an inductive analysis of the coherence of the outputs generated from each *k* value; $k = 25$ was identified as having the most coherent model output. This approach reflects the exploratory nature of this example, wherein the model is intended to provide an overview of major themes that emerge in COP speeches. Subsequently, the research team calculated the most commonly occurring topics by country and by year using posterior probabilities for each topic in a document.

K-selection: Canadian local government records

For the Canadian local government example, the robust LDA model was used (rlda) (Wilkerson and Casas 2017). Using the Python package *rlda*, a set of topic models was generated for $k = \{20, 21, ..., 40\}$, for a total of 21 models containing 630 topics. Model stability was then approximated using pairwise cosine similarity, which uses a clustering algorithm to group the 630 topics generated across all models by similarity. This process identified a stable model output of approximately 30 topics.

6.3.4 Interpretation

Even exploratory analyses require the researcher to examine model output and interpret meaning from the word clusters identified. Robust interpretation of topic model results therefore requires familiarity with the subject matter, and a strong understanding of texts used to create the corpus. Here, two researchers independently examined the model outputs from each example and assigned topic titles based on expert interpretation of the word clusters; together their interpretations were compared and discussed to resolve any differences.

6.4 Applying LDA topic models to climate change adaptation

6.4.1 Case 1: COP speeches (2010-2016)

The United Nations Framework Convention on Climate Change (UNFCCC) is a key site for the debate, establishment, and harmonization of global and national climate change policy (Gupta 2010). At the start of each annual UNFCCC Conference of Parties (COP), heads of state and government gather to make brief statements regarding their positions before negotiations begin. With almost all countries submitting a statement each year, these brief speeches give insight into national priorities and overarching discursive trends around climate change (Bagozzi, 2015; Ford & Maillet, 2016). This example looks at Party statements concerning adaptation from COP16 in 2010 to COP22 in 2016, with an interest in identifying trends by country and over time. We apply an LDA model to the corpus and analyze the overall results, probabilities of topic occurrence by year, and differences in topic occurrence between high-income countries (Annex I Parties) and medium- and low-income countries (non-Annex I Parties). It is worth noting that this approach differs from that taken by correlated topic models (e.g. structural topic models), which uses regression models to estimate the relationship between topic prevalence and specified co-variates (Roberts et al. 2014).

Twenty-five topics were generated by the model that represent five broad themes (see Table 6.2). The first theme is an emphasis on the governance architecture for adaptation (topics 1-9), including efforts under the UNFCCC process and national planning processes. Second is the urgent need to take action given the negative consequences of climate change (topics 10-12). The third theme consists of intersections between adaptation and other policy goals, including sustainable development and mitigation (topics 13-18). Two additional themes are detected around implementation procedures, including support for capacity-building and project implementation (topics 19-22), and climate financing, including financing for African countries, payment into the Green Climate Fund, and addressing the issue of loss and damage (topics 22-25).

Table Error! No text of specified style in document.**.2** Topics in COP speeches

Number	Topic	Terms
1	Paris Agreement	agreement, must, new, balac, element, pari, comprehens, legal, essenti, natur
2	Cooperation	climat, chang, impact, strengthen, import, cooper, ensur, activ, inform, becom
3	Adaptation framework	framework, committee, mechan, establish, cancun, institute, convent, work, made, durban
4	Global governance	chang, climat, govern, world, assist, ambit, just, promot, current, holistic

5	Leadership	prioriti, presid, remain, like, equal, given, import, resourc, alreadi, impact
6	Party commitments	mitig, commit, financ, pari, order, period, presid, protocol, continu, activ
7	Enhanced action	action, includ, enhanc, implement, mean, program, provis, appropri, nation, assist
8	National planning	nation, plan, strategi, program, process, prepar, polici, adopt, integr, communic
9	Least developed countries	Countri, developedcountri, developingcountri, least, small, ldcs, island, african, continu, especi
10	Negative climate change impacts	climat, chang, effect, impact, advers, limit, negat, approach, resourc, convent
11	Need to act	need, mitig, urgent, countri, strong, alreadi, financi, cooper, futur, appropri
12	Risk and vulnerability	vulner, particular, challeng, level, increas, risk, requir, high, extrem, take
13	Sustainable development	develop, sustain, low, econom, achiev, economi, goal, carbon, includ, object
14	Mitigation action	mitig, action, climate, key, intern, achiev, balanc, unfccc, govern, carbon
15	Mitigation effort	global, effort, mitig, contribut, necessari, activ, part, implement, climat, relat
16	Emissions reduction	Emiss, climat, measur, reduc, reduct, effort, greenhous, help, includ, aim
17	Community resilience	mitig, respons, increas, resili, ensur, address, communiti, common, need, capabl
18	Food-water-energy	sector, agricultur, measur, energi, water, initi, secur, food, manag, strengthen
19	Technical and financial support	support, financi, resourc, adequ, mitig, access, call, technic, direct, area
20	Technical capacity	technolog, capac, build, financ, transfer, transpar, enabl, share, forward, join
21	Project implementation	implement, project, import, term, long, mean, ensur, programm, mitig, includ
22	Developing country support	countri, support, developingcountri, provid, enabl, project, clean, first, major, requir
23	Climate finance for Africa	financ, year, africa, addit, billion, cost, toward, million, alloc, start
24	Loss and damage	loss, address, damag, issu, intern, work, time, target, mani, critic
25	Green Climate Fund	fund, green, mechan, contribut, decis, howev, predict, one, special, must

Mean topic probabilities were analyzed by year and by country development status. The yearly results provide intuitive validation of the coherence of the categories (Table 6.3). Overall, we detect a shift between 2010 and 2016 from an emphasis on technical and financial support for least developed countries, to an emphasis on the governance of adaptation at global and national levels. Indeed, COP16-18 were important for the elaboration of the Cancun Adaptation Framework, including enhanced action and cooperation on adaptation and the set-up of the Green Climate Fund, and the establishment of a process for supporting national adaptation planning in least developed country Parties (Schipper 2006; Hall and Persson 2018). In the run-up to the adoption of the Paris Agreement at COP21 we see a move towards emphasizing governance aspects of the UNFCCC process, including intersections with other issue areas like mitigation and sustainable development. A focus on technical capacity is still apparent but is no longer a dominant topic emerging from the model.

Table Error! No text of specified style in document.**.3** Probability of topic occurrence by COP event

	COP16 (2010)	**COP17 (2011)**	**COP18 (2012)**	**COP19 (2013)**	**COP20 (2014)**	**COP21 (2015)**	**COP22 (2016)**
1	Adaptation framework (.044)	Adaptation framework (.045)	National planning (.043)	National planning (.043)	National planning (.046)	Paris Agreement (.043)	National planning (.046)
2	Technical capacity (.043)	National planning (.043)	Negative climate change impacts (.043)	Global governance (.042)	Technical capacity (.043)	Negative climate change impacts (.043)	Food-water-energy (.044)
3	Technical and financial support (.041)	Technical capacity (.042)	Developing country support (.042)	Cooperation (.041)	Paris Agreemen t (.043)	Mitigation (.042)	Sustainable developmen t (.042)
4	Developing country support (0.041)	Developing country support (.042)	Technical capacity (.042)	Climate finance for Africa (.041)	Mitigation (.041)	Sustainable developme nt (.042)	Global governance (.042)

5	Enhanced action (0.04)	Least developed countries (.041)	Least developed countries (.041)	Adaptation framework (0.41)	Global governance (.041)	Least developed countries (.041)	Technical capacity (.042)

Separate examination of the most commonly occurring topic per country for the middle and low-income country block (non-Annex I Parties, n = 155) and the high-income country block (Annex I Parties, n = 42) reveal further insights into these patterns that broadly echo themes found in hand-coded analyses of UNFCCC decision texts (Figure 6.2) (Ford et al. 2016). While non-Annex I Parties tend to focus on national adaptation planning and technical capacity in COP speeches, Annex I Parties are emphasizing climate financing and intersections with mitigation efforts. This is consistent with the polluter pays principle underlying the UNFCCC's approach to adaptation, with developing countries prioritizing national adaptation planning and Annex I Parties (who carry greater mitigation responsibilities) providing the technical and financial support for those efforts.

Figure Error! No text of specified style in document.**.2** Topics by country development status

Figure 6.2a Most likely topic (Non-Annex I Parties)

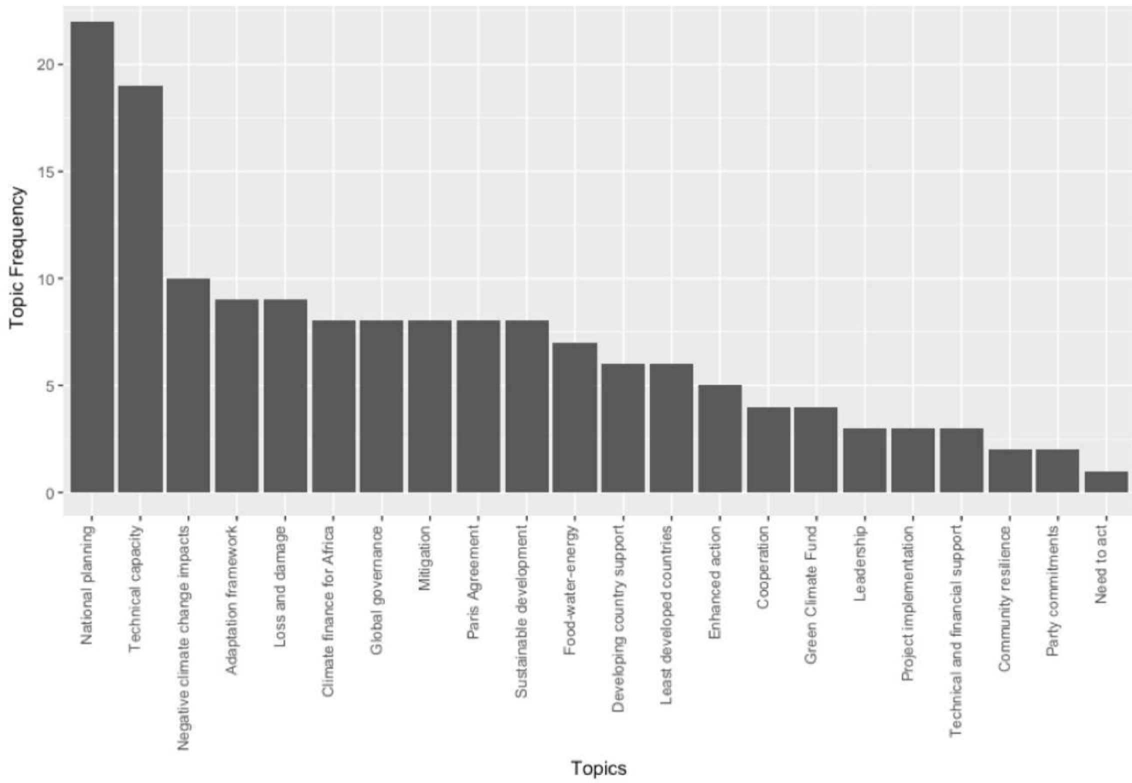
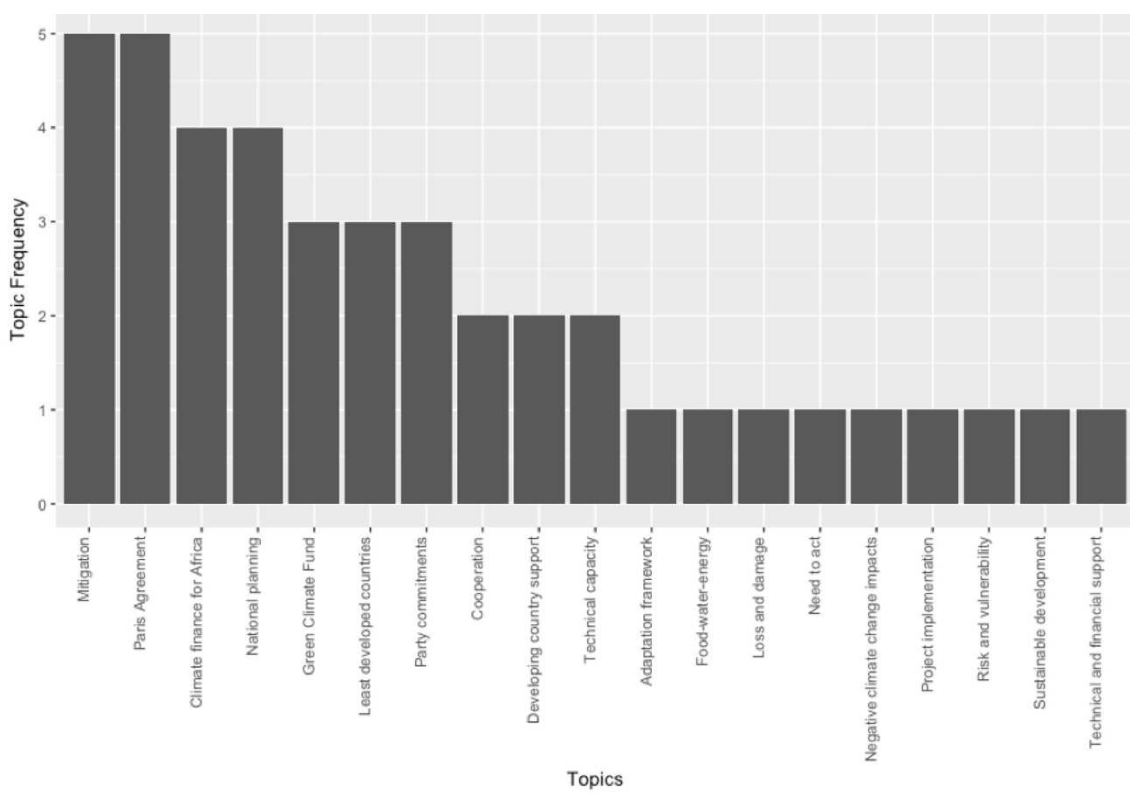


Figure 6.2b Most likely topic (Annex I Parties)



6.4.2 Case 2: Adaptation policy in 25 Canadian cities (2010-2017)

Local governments are considered key sites for adaptation policy development and implementation (Nalau, Preston, and Maloney 2015). A growing body of research is focusing on emerging patterns of policy adoption among local governments with the goal of understanding how decision-makers are integrating adaptation considerations in local operations, plans, and services (Castán Broto & Bulkeley, 2013; Hughes, 2015; Mees, 2017; Shi, Chu, & Debats, 2015; Swart et al., 2014). This case examines topics pertaining to adaptation in 25 Canadian local governments using records from city council meetings between 2010 and May 2017. It demonstrates how topic modelling can be used to get a sense of key adaptation issues facing governments, and broadly how local governments are approaching adaptation as policy issue. We apply a robust LDA model to the corpus to identify a suitable K.

We interpret five overarching themes from the topics generated by the model, which indicate that adaptation in Canadian cities is largely being considered from the perspective of the built environment (see Table 6.4). The largest discernible theme in topics is around land use management (topics 1-8), which concerns zoning, area planning, and project development, strategic planning around key sectors, and neighbourhood conservation. Several topics are also concerned with public works, including freshwater and wastewater management, waste management, and grey infrastructure (topics 13-16). While about half of the topics identified by the model center around hard infrastructure, several other topics are related to urban greening, including ecological areas, environmental assessment, and the urban forest (topics 23-25). The remaining topics capture a series of substantive issues for local governments that intersect with adaptation, including local resources, transportation, flood protection, mitigation, and local food systems (topics 17-20, 22).

Figure Error! No text of specified style in document.**.3** Topics by Canadian local government records

Number	Topic	Terms
1	Subdivision	plan, owner, subdivision, satisfaction, draft, engineer, road, lands, development, design, construction, sanitary, prior, lots, required
2	Site development	residential, site, development, street, density, building, zoning, plan, area, zone, planning, lands, commercial, planner, design
3	Project planning	project, planning, development, street, building, company, district, plan, area, amount, services, construction, integration, prepared, site
4	Land use planning	plan, area, lands, development, land, uses, planning, industrial, site, areas, official, growth, planner, natural, commercial
5	Re-zoning	development, site, rezoning, building, district, housing, plan, community, application, zoning, street, residential, planning, centre, engineering
6	Urban growth planning	plan, community, development, strategy, management, growth, environmental, transportation, land, infrastructure, planning, economic, sustainability, sustainable, services
7	Strategic planning	energy, water, food, river, waste, climate, community, flood, services, downtown, transit, plan, risk, health, street
8	Heritage protection	heritage, conservation, district, building, plan, street, property, guidelines, original, cultural, village, south, old, wortley, buildings
9	Legal and records services	law, services, street, community, information, road, planning, development, file, plan, avenue, solicitor, part, act, property
10	Community services	services, corporate, community, management, environmental, law, service, 'business, risk, fire, safety, back, protective, parks, park
11	Financial resources	budget, capital, million, services, funding, service, year, management, cost, operating, financial, asset, fund, water, infrastructure
12	Health and safety	health, services, prevention, unit, planning, community, care, safety, fire, team, housing, middlesex, healthy, ace, lake
13	Freshwater management	water, drinking, system, stormwater, wastewater, sewer, management, quality, treatment, lake, systems, act, infrastructure, environment, response

14	Waste management	waste, landfill, resource, recovery, diversion, recycling, environmental, solid, gas, management, collection, garbage, disposal, environment, materials
15	Wastewater management	stormwater, water, sewer, management, storm, system, treatment, wastewater, infrastructure, flooding, sanitary, green, control, engineering, property
16	Grey infrastructure	dike, area, road, protection, management, island, phase, river, existing, ecological, land, strategy, lands, infrastructure, park
17	Local resources	municipalities, infrastructure, funding, communities, housing, national, standing, development, provincial, forum, local, safety, provided, update, issues
18	Transportation	transit, downtown, transportation, street, cycling, design, parking, pedestrian, road, rapid, project, plan, service, traffic, bridge
19	Flood protection	river, flood, thames, dike, mitigation, dam, assessment, protection, area, lake, flooding, measures, property, level, project
20	Mitigation	energy, emissions, climate, community, gas, carbon, ghg, greenhouse, plan, corporate, change, reduction, local, green, sustainability
21	Impacts and adaptation	climate, change, adaptation, risk, weather, impacts, flood, heat, extreme, dike, events, mitigation, strategy, health, increased
22	Local food systems	food, local, system, community, agriculture, agricultural, urban, production, health, security, farm, strategy, land, flood, governments
23	Ecological areas	natural, areas, ecological, river, species, eis, dike, area, habitat, study, environmental, management, heritage, features, flood
24	Environmental assessment	environmental, study, project, river, engineering, stormwater, thames, creek, works, assessment, plan, process, water, flood, design
25	Urban forest	trees, tree, urban, forest, species, strategy, planting, 'canopy', 'invasive', 'cover', 'forests', 'management', 'green', 'ace', 'forestry']

The topics reflect the high visibility of flood risk management in local Canadian adaptation planning (Thistlethwaite and Henstra 2017; Henstra et al. 2019); ‘flood’ appears in topics 7, 15, 19, and 21-24. Topics 13 (‘freshwater management’) and 16 (‘grey infrastructure’) can also be interpreted as related to flood risk management. Topic 21 (‘impacts and adaptation’) suggests that municipalities are concerned about heat risk in a changing climate, but this seems disconnected from the ‘health and safety’ topic that is composed of words relating to community health services and emergency services.

We draw four observations from these topic interpretations. First, climate change adaptation approaches among local governments seem to be embedded in local regulatory tools related to land use decision-making and public works projects. Second, Canadian municipalities seem to be primarily concerned about risks from extreme events, particularly flooding but also extreme heat. Third, the relative balance of topics indicate that adaptation is more often linked with ‘hard’ aspects of the built environment like infrastructure, buildings, and public works (topics 1-5, 7-8, 11, 13-19, 21, 24), with only two topics composed of terms related to green infrastructure (topics 23 and 25). Finally, these topics suggest that local adaptation in Canada is being framed as an issue of vulnerability to climate change risks, and a planning issue connected to activities like land use management, services provision, and environmental assessment (Juhola, Keskitalo, and Westerhoff 2011). It is worth noting that the presence of mitigation and transportation categories suggests that the decision to take a larger selection of words around the adaptation keywords that were used to generate the corpus (see section 3.2.2 for detailed description) also captured mitigation content; further narrowing of the text might have generate somewhat different topic outputs.

6.5 What does topic modelling offer adaptation governance research?

The aim of this Focus Article is to provide an overview of topic modelling and its uses, and discuss potential applications for the study of adaptation governance. The two cases illustrated here are intended to be interpreted only in an exploratory light, and demonstrate the range of document sources that can be used in topic models and how different types of insights can be drawn from these various sources. The examples demonstrate two approaches to dealing with a key methodological debate in topic modelling, namely how to optimize model performance by selecting an appropriate number of topics around which the algorithm builds its output: a partial inductive approach typical of LDA applications in the social sciences (COP speeches), and a spectral clustering technique for grouping topics of a similar nature used in the robust LDA model (Canadian local government documents).

There are several important takeaways for adaptation governance researchers considering the use of topic models in their research. First, topic models are never an entirely automated affair. Model outputs require interpretation by researchers, and validity of results must be assessed based on clear criteria. Chuang et al., for example, offer several suggestions as a general guideline for establishing model validity, including use of multiple models to determine model consistency and measuring topic similarity (Chuang et al. 2015). Several existing topic modelling packages include features for estimating model robustness, such as the *stm* package in R for structural topic modelling, which helps to simplify this interpretive process (Roberts et al. 2014).

Second, decisions made in pre-processing are critical to the interpretability of model results (Denny and Spirling 2018). Determining whether removal of stopwords, stemming, and language translation will impact the validity of results are important steps in the process of implementing topic modelling. Here we provide only a limited introduction to pre-processing considerations, but there is a growing empirical literature testing the implications of various pre-processing decisions for model robustness.

Third, topic modelling can be used alone as an exploratory or hypothesis-testing technique, but it can also be used to strengthen the validity of manual coding protocols, and to inform the identification of future research questions (Potter and Levine-Donnerstein 1999). For example, the model results discussed here offer several interesting directions for qualitative research projects: 1) How are issues around technical capacity and financial support for non-Annex I States being treated under the emerging global governance framework emerging from the Paris Agreement? Are we seeing a shift in how States are addressing these gaps in light of this emphasis on global climate change governance? 2) To what extent is there coherence between national adaptation planning efforts in non-Annex I Parties and emerging climate finance plans from Annex I Parties? 3) How do regulatory powers around land use and development affect the scope of adaptation responses to key vulnerabilities in Canadian municipalities? 4) To what extent are local governments in Canada adopting 'soft' approaches to flood risk management, or do they continue to rely on more traditional grey infrastructure approaches?

We suggest four key ways that topic modelling might inform adaptation governance research in the future. First, topic modelling can be used to analyze framing and issue salience. Frames are key components of decision-making processes because problem detection and definition shape how actors think about adaptation and what kind of responses they propose (Dewulf 2013). These frames are often implicit, however, and not easy to identify. Topic modelling can be used for inductively detecting frames embedded within the latent structure of policy documents, with the added advantage of reducing potential bias from the application of *a priori* frame definitions that may not translate easily across contexts. This type of frame analysis can also be triangulated with more fine-grained studies of policy adoption to advance understandings of how framing is related to motivations behind policy and financing decisions. Incorporation of a longitudinal perspective using dynamic topic models can also shed light on how the framing of adaptation is changing over time.

Second, expanding our ability to parse latent adaptation content across larger volumes of text also offers a new approach to the study of adaptation policy integration (Candel and Biesbroek 2016; Massey et al. 2015; Schmidt and Fleig 2018). Identifying keyword similarities in policy documents across jurisdictions, administrative units, or organizations can be used to examine the climate change concerns of politicians and decision-makers and shed light on coherence of ideas, issues, and approaches across sectors and scales. Similarly, it can also inform our understanding of how adaptation is distinct from related policy areas (Runhaar et al. 2017; Roeck, Orbie, and Delputte 2018).

Third, policy document analyses can be used for evaluative research by connecting thematic patterns generated by topic models with global climate model projections or climate impact assessments that identify key vulnerable sectors or regions. This type of analysis can inform us about the extent to which there is alignment between the projected environmental risk and the focus or concerns of decision-makers. These evaluative questions are highly pertinent in more applied areas of adaptation governance research, which aim to determine whether current adaptation efforts are aligned with priorities for vulnerability reduction.

Finally, here we presented exploratory examples of the LDA model, but application of correlated topic models that look for covariance between topics can be used for hypothesis testing studies. In the absence of large data-sets on adaptation policies and processes, descriptive and causal research has been largely limited to case studies or small-n comparisons. Topic modelling would enable larger hypothesis testing studies that use document identifiers determined by the researcher to test relationships between the content of texts and variables like institutional structure, development status, political culture, or environmental exposure.

6.6 Conclusion

The efficiency gains that come with topic modelling represent an opportunity for adaptation governance research to engage with large-n comparative research. With rapid technical progress being made in the social sciences around the application of topic models, this approach will be an important tool for making sense of the growing volume of qualitative information available for research and policy purposes. Harnessing opportunities to use quantitative text approaches like topic modelling for adaptation research will require competency-building among researchers in the adaptation community, and deeper collaboration with quantitative social scientists already applying these techniques in their research. We argue that the chance to scale-up text-based analysis is well-worth the effort and will open new methodological horizons for adaptation research that have been previously underexplored.

References

Alghamdi, Rubayyi. 2015. “A Survey of Topic Modeling in Text Mining.” *International Journal of Advanced Computer Science and Applications* 6 (1): 147–53.

Allee, Todd, Manfred Elsig, and Andrew Lugg. 2017. “Is the European Union Trade Deal with Canada New or Recycled? A Text-as-Data Approach.” *Global Policy* 8 (2): 246–52. doi:10.1111/1758-5899.12420.

Alvarez-Melis, David, and Martin Saveski. 2016. “Topic Modeling in Twitter: Aggregating Tweets by Conversations.” In *Proceedings of the Tenth International AAAI Conference on Web and Social Media*, 519–22.

Araos, Malcolm, Lea Berrang-Ford, James Ford, Stephanie Austin, Robbert Biesbroek, and Alexandra Lesnikowski. 2016. “Climate Change Adaptation Planning in Large Cities: A Systematic Global Assessment.” *Environmental Science and Policy* 66: 375–82.

Bagozzi, Benjamin E. 2015. “The Multifaceted Nature of Global Climate Change Negotiations.” *Review of International Organizations* 10: 439–64. doi:10.1007/s11558-014-9211-7.

Bauer, Anja, and Reinhard Steurer. 2014. “National Adaptation Strategies, What Else? Comparing Adaptation Mainstreaming in German and Dutch Water Management.” *Regional Environmental Change* 15 (2): 341–52.

Beelen, Kaspar, Timothy Alberdingk Thijm, Christopher Cochrane, Kees Halvemaan, Graeme Hirst, Michael Kimmins, Sander Lijbrink, et al. 2017. “Digitization of the Canadian Parliamentary Debates.” *Canadian Journal of Political Science* 50 (3): 849–64. doi:10.1017/S0008423916001165.

Belfer, Ella, James Ford, and Michelle Maillet. 2017. “Representation of Indigenous Peoples in Climate Change Reporting.” *Climatic Change* 145 (1–2): 57–70. doi:10.1007/s10584-017-2076-z.

Benoit, Kenneth, Michael Laver, and Slava Mikhaylov. 2009. “Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions.” *American Journal of Political Science* 53 (2): 495–513. doi:10.1111/j.1540-5907.2009.00383.x.

Benoit, Kenneth, David Muhr, and Kohei Watanabe. 2017. “Package ‘Stopwords.’”

Berrang-Ford, Lea, Tristan Pearce, and James Ford. 2015. “Systematic Review Approaches for Climate Change Adaptation Research.” *Regional Environmental Change* 15 (5): 755–69.

Bickel, Manuel W. 2017. “A New Approach to Semantic Sustainability Assessment: Text Mining via Network Analysis Revealing Transition Patterns in German Municipal Climate

Action Plans." *Energy, Sustainability and Society* 7 (1): 1–25. doi:10.1186/s13705-017-0125-0.

Biesbroek, Robbert, Lea Berrang-Ford, Alexandra Lesnikowski, Stephanie Austin, and James Ford. 2018. "Data, Concepts and Methods for Large-n Comparative Climate Change Adaptation Policy Research: A Systematic Literature Review." *Wiley Interdisciplinary Reviews: Climate Change* 9 (6): 1–15.

Blei, David. 2012. "Introduction to Probabilistic Topic Modeling." *Communications of the ACM* 55 (4): 77–84. doi:10.1145/2133806.2133826.

Blei, David, and John D. Lafferty. 2006a. "Correlated Topic Models." *Advances in Neural Information Processing Systems 18*. MIT Press. doi:10.1145/1143844.1143859.

Blei, David, and John D Lafferty. 2006b. "Dynamic Topic Models." In *Proceedings of the 23rd International Conference on Machine Learning*, 113–20. ICML '06. New York, NY, USA: ACM. doi:10.1145/1143844.1143859.

Blei, David, Andrew Y Ng, and Michael I Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022. doi:10.1162/jmlr.2003.3.4-5.993.

Boussalis, Constantine, and Travis G. Coan. 2016. "Text-Mining the Signals of Climate Change Doubt." *Global Environmental Change* 36: 89–100. doi:10.1016/j.gloenvcha.2015.12.001.

Boyd-Graber, Jordan L, and David Blei. 2009. "Syntactic Topic Models." In *Advances in Neural Information Processing Systems*, 1:1–26.

Cambria, Erik, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. 2013. "New Avenues in Opinion Mining and Sentiment Analysis." *IEEE Intelligent Systems* 28 (2): 15–21. doi:10.1109/MIS.2013.30.

Candel, Jeroen J. L., and Robbert Biesbroek. 2016. "Toward a Processual Understanding of

Policy Integration." *Policy Sciences* 49 (3): 211–31.

Castán Broto, Vanesa, and Harriet Bulkeley. 2013. "A Survey of Urban Climate Change Experiments in 100 Cities." *Global Environmental Change* 23 (1): 92–102.

Chang, J, and David Blei. 2009. "Relational Topic Models for Document Network." In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, 81–88. Clearwater Beach, FL, USA: AISTATS. doi:10.1016/j.nimb.2008.03.150.

Chang, Jonathan, Sean Gerrish, Chong Wang, and David Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." *Advances in Neural Information Processing Systems 22*, 288--296. doi:10.1.1.100.1089.

Chuang, Jason, Margaret E Roberts, Brandon M Stewart, Rebecca Weiss, Dustin Tingley, Justin Grimmer, and Jeffrey Heer. 2015. "TopicCheck: Interactive Alignment for Assessing Topic Model Stability." In *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, 175–84. doi:10.1016/j.ijhcs.2017.03.007.

Cody, Emily M., Andrew J. Reagan, Lewis Mitchell, Peter Sheridan Dodds, and Christopher M. Danforth. 2015. "Climate Change Sentiment on Twitter: An Unsolicited Public Opinion Poll." *PLoS ONE* 10 (8): e0136092. doi:10.1371/journal.pone.0136092.

Craft, Brianna, and Susannah Fisher. 2018. "Measuring the Adaptation Goal in the Global Stocktake of the Paris Agreement Agreement." *Climate Policy* 18 (9): 1203–9. doi:10.1080/14693062.2018.1485546.

Dąbrowski, Marcin. 2018. "Boundary Spanning for Governance of Climate Change Adaptation in Cities: Insights from a Dutch Urban Region." *Environment and Planning C: Politics and Space* 36 (5): 837–55. doi:10.1177/2399654417725077.

de Vries, Erik, Martijn Schoonvelde, and Gijs Schumacher. 2018. "No Longer Lost in

Translation: Evidence That Google Translate Works for Comparative Bag-of-Words Text Applications." *Political Analysis* 26 (4): 417–30. doi:10.1017/pan.2018.26.

Denny, Matthew James, and Arthur Spirling. 2018. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* 26: 168–89. doi:10.2139/ssrn.2849145.

Dewulf, Art. 2013. "Contrasting Frames in Policy Debates on Climate Change Adaptation." *Wiley Interdisciplinary Reviews: Climate Change* 4 (4): 321–30. doi:10.1002/wcc.227.

DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41 (6): 570–606. doi:10.1016/j.poetic.2013.08.004.

Dupuis, Johann, and Peter Knoepfel. 2013. "The Adaptation Policy Paradox: The Implementation Deficit of Policies Framed as Climate Change Adaptation." *Ecology and Society* 18 (4): 31–47.

Elgesem, Dag, Lubos Steskal, and Nicholas Diakopoulos. 2015. "Structure and Content of the Discourse on Climate Change in the Blogosphere: The Big Picture." *Environmental Communication* 9 (2): 169–88. doi:10.1080/17524032.2014.983536.

Farrell, Justin. 2016. "Corporate Funding and Ideological Polarization about Climate Change." *Proceedings of the National Academy of Sciences* 113 (1): 92–97. doi:10.1073/pnas.1509433112.

Ford, James, Lea Berrang-Ford, Robbert Biesbroek, Malcolm Araos, Stephanie Austin, and Alexandra Lesnikowski. 2015. "Adaptation Tracking for a Post-2015 Climate Agreement." *Nature Climate Change* 5 (11): 967–69.

Ford, James, Lea Berrang-Ford, Anna Bunce, Courtney McKay, Maya Irwin, and Tristan Pearce. 2015. "The Status of Climate Change Adaptation in Africa and Asia." *Regional Environmental Change* 15 (5): 801–14. doi:DOI 10.1007/s11027-014-9627-7.

Ford, James, Michelle Maillet, Vincent Pouliot, Thomas Meredith, Alicia Cavanaugh, and IHACC Research Team. 2016. "Adaptation and Indigenous Peoples in the United Nations Framework Convention on Climate Change." *Climatic Change* 139 (3–4): 429–43.

Ford, James, Simon Tilleard, Lea Berrang-Ford, Malcolm Araos, Robbert Biesbroek, Alexandra Lesnikowski, Graham K. MacDonald, Angel Hsu, Chen Chen, and Livia Bizikova. 2016. "Opinion: Big Data Has Big Potential for Applications to Climate Change Adaptation." *Proceedings of the National Academy of Sciences* 113 (39): 10729–32.

Goldstone, Andrew, and Ted Underwood. 2012. "What Can Topic Models of PMLA Teach Us about the History of Literary Scholarship?" *Journal of Digital Humanities* 2 (1): 39–48.

Grimmer, Justin. 2010. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis* 18 (1): 1–35. doi:10.1093/pan/mpp034.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97. doi:10.1093/pan/mps028.

Grubert, Emily. 2018. "Rigor in Social Life Cycle Assessment: Improving the Scientific Grounding of SLCA." *International Journal of Life Cycle Assessment* 23 (3): 481–91. doi:10.1007/s11367-016-1117-6.

Grubert, Emily, and Mark Algee-Hewitt. 2017. "Villainous or Valiant? Depictions of Oil and Coal in American Fiction and Nonfiction Narratives." *Energy Research and Social Science*

31: 100–110. doi:10.1016/j.erss.2017.05.030.

Grubert, Emily, and Anne R. Siders. 2016. "Benefits and Applications of Interdisciplinary Digital Tools for Environmental Meta-Reviews and Analyses." *Environmental Research Letters* 11 (9): 93001.

Grün, Bettina, and Kurt Hornik. 2011. "Topicmodels : An R Package for Fitting Topic Models." *Journal of Statistical Software* 40 (13): 1–30. doi:10.18637/jss.v040.i13.

Gupta, Joyeeta. 2010. "A History of International Climate Change Policy." *Wiley Interdisciplinary Reviews: Climate Change* 1 (5): 636–53. doi:10.1002/wcc.67.

Hall, Nina, and Åsa Persson. 2018. "Global Climate Adaptation Governance: Why Is It Not Legally Binding?" *European Journal of International Relations* 24 (3): 540–66. doi:10.1177/1354066117725157.

Head, Brian W. 2014. "Evidence, Uncertainty, and Wicked Problems in Climate Change Decision Making in Australia." *Environment and Planning C: Government and Policy* 32 (4): 663–79. doi:10.1068/c1240.

Henstra, Daniel, Jason Thistlethwaite, Craig Brown, and Daniel Scott. 2019. "Flood Risk Management and Shared Responsibility: Exploring Canadian Public Attitudes and Expectations." *Journal of Flood Risk Management* 12 (1): e12346. doi:10.1111/jfr3.12346.

Hetz, Karen. 2016. "Contesting Adaptation Synergies: Political Realities in Reconciling Climate Change Adaptation with Urban Development in Johannesburg, South Africa." *Regional Environmental Change* 16 (4): 1171–82. doi:10.1007/s10113-015-0840-z.

Hillard, Dustin, Stephen Purpura, and John Wilkerson. 2008. "Computer-Assisted Topic Classification for Mixed-Methods Social Science Research." *Journal of Information Technology and Politics* 4 (4): 31–46. doi:10.1080/19331680801975367.

Hoffmann, Matthew, Francis Bach, and David Blei. 2010. "Online Learning for Latent Dirichlet Allocation." In *Advances in Neural Information Processing Systems*, 1–9. Denver, CO, USA: Neural Information Processing Systems Foundation. doi:10.1.1.187.1883.

Hong, Liangjie, and Brian D. Davison. 2010. "Empirical Study of Topic Modeling in Twitter." In *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*, 80–88. Washington, DC: ACM. doi:10.1145/1964858.1964870.

Hopkins, Daniel J, and Gary King. 2010. "Analysis for Social Science." *American Journal of Political Science* 54 (1): 229–47. doi:10.1111/j.1540-5907.2009.00428.x.

Hughes, Sara. 2015. "A Meta-Analysis of Urban Climate Change Adaptation Planning in the U.S." *Urban Climate* 14: 17–29.

Jang, S. Mo, and P. Sol Hart. 2015. "Polarized Frames on 'Climate Change' and 'Global Warming' across Countries and States: Evidence from Twitter Big Data." *Global Environmental Change* 32: 11–17. doi:10.1016/j.gloenvcha.2015.02.010.

Jelodar, Hamed, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Li Yanchao, and Liang Zhao. 2018. "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey." *Multimedia Tools and Applications*, 1–43.

Juhola, Sirkku, E. Carina H. Keskitalo, and Lisa Westerhoff. 2011. "Understanding the Framings of Climate Change Adaptation across Multiple Scales of Governance in Europe." *Environmental Politics* 20 (4): 445–63.

Keenan, Jesse M., David A. King, and Derek Willis. 2016. "Understanding Conceptual Climate Change Meanings and Preferences of Multi-Actor Professional Leadership in New York." *Journal of Environmental Policy and Planning* 18 (3): 261–85. doi:10.1080/1523908X.2015.1104628.

Kirilenko, Andrei P., and Svetlana O. Stepchenkova. 2014. "Public Microblogging on Climate Change: One Year of Twitter Worldwide." *Global Environmental Change* 26 (1): 171–82. doi:10.1016/j.gloenvcha.2014.02.008.

Krippendorff, Klaus. 2013. *Content Analysis: An Introduction to Its Methodology*. 3rd ed. Sage Publications Inc.

Labbé, Jolène, James Ford, Malcolm Araos, and Melanie Flynn. 2017. "The Government-Led Climate Change Adaptation Landscape in Nunavut, Canada." *Environmental Reviews* 25 (1): 12–25. doi:10.1139/er-2016-0032.

Lasswell, Harold. 1952. *The Comparative Study of Symbols: An Introduction*. Palo Alto, California: Stanford University Press.

Laver, Michael, and Kenneth Benoit. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97 (2): 311–31. doi:10.1017/S0003055403000698.

Lesnikowski, Alexandra, James Ford, Robbert Biesbroek, Lea Berrang-Ford, and Jody Heymann. 2016. "National-Level Progress on Adaptation." *Nature Climate Change* 6 (3): 261–64.

Levin, Kelly, Benjamin Cashore, Steven Bernstein, and Graeme Auld. 2012. "Overcoming the Tragedy of Super Wicked Problems: Constraining Our Future Selves to Ameliorate Global Climate Change." *Policy Sciences* 45 (2): 123–52. doi:10.1007/s11077-012-9151-0.

Lewis, David D., Yiming Yang, Tony G. Rose, and Fan Li. 2004. "RCV1: A New Benchmark Collection for Text Categorization Research." *Journal of Machine Learning Research* 5: 361–97. doi:10.1145/122860.122861.

Liu, Lin, Lin Tang, Wen Dong, Shaowen Yao, and Wei Zhou. 2016. "An Overview of Topic

Modeling and Its Current Applications in Bioinformatics." *SpringerPlus* 5 (1): 1608. doi:10.1186/s40064-016-3252-8.

Lynam, Timothy. 2016. "Exploring Social Representations of Adapting to Climate Change Using Topic Modeling and Bayesian Networks" 21 (4). doi:10.5751/ES-08778-210416.

Lynam, Timothy, and Iain Walker. 2016. "Making Sense of Climate Change: Orientations to Adaptation." *Ecology and Society* 21 (4): 17. doi:10.5751/ES-08886-210417.

Massey, Eric, Dave Huitema, Heiko Garrelts, Kevin Grecksch, Heleen Mees, Tim Rayner, Sofie Storbjork, Catrien J.A.M. Termeer, and Maik Winges. 2015. "Handling Adaptation Policy Choices in Sweden, Germany, the UK and the Netherlands." *Journal of Water and Climate Change* 6 (1): 9–24.

Mcauliffe, Jon D, and David Blei. 2008. "Supervised Topic Models." In *Advances in Neural Information Processing Systems 20*, edited by J C Platt, D Koller, Y Singer, and S T Roweis, 121–28. Curran Associates, Inc.

Mees, Heleen. 2017. "Local Governments in the Driving Seat? A Comparative Analysis of Public and Private Responsibilities for Adaptation to Climate Change in European and North-American Cities." *Journal of Environmental Policy and Planning* 19 (4): 374–90. doi:10.1080/1523908X.2016.1223540.

Mimno, David, and David. 2012. "Computational Historiography: Data Mining in a Century of Classics Journals." *Journal on Computing and Cultural Heritage* 5 (1): 1–19. doi:10.1145/2160165.2160168.

Nalau, Johanna, Benjamin L. Preston, and Megan C. Maloney. 2015. "Is Adaptation a Local Responsibility?" *Environmental Science & Policy* 48: 89–98.

Onyimadu, Obinna, Keiichi Nakata, Tony Wilson, David Macken, and Kecheng Liu. 2013.

"Towards Sentiment Analysis on Parliamentary Debates in Hansard." In *Joint International Semantic Technology Conference*, 48–50. Seoul, South Korea: Springer. doi:10.1007/978-3-319-06826-8_4.

Pollitt, Christopher. 2015. "Wickedness Will Not Wait: Climate Change and Public Management Research." *Public Money & Management* 35: 181–86. doi:10.1080/09540962.2015.1027490.

Potter, W. James, and Deborah Levine-Donnerstein. 1999. "Rethinking Validity and Reliability in Content Analysis." *Journal of Applied Communication Research* 27 (3): 258–84. doi:10.1080/00909889909365539.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54 (1): 209–28. doi:10.1111/j.1540-5907.2009.00427.x.

Ravi, Kumar, and Vadlamani Ravi. 2015. "A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications." *Knowledge-Based Systems* 89: 14–46. doi:10.1016/j.knosys.2015.06.015.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–82. doi:10.1111/ajps.12103.

Roeck, Frederik De, Jan Orbie, and Sarah Delputte. 2018. "Mainstreaming Climate Change Adaptation into the European Union's Development Assistance." *Environmental Science and Policy* 81: 36–45. doi:10.1016/j.envsci.2017.12.005.

Rosen-Zvi, Michal, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. 2004. "The Author-Topic Model for Authors and Documents." In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 487–94. UAI '04. Arlington, Virginia, United States: AUAI Press.

Runhaar, Hens, Bettina Wilk, Åsa Persson, Caroline Uittenbroek, and Christine Wamsler. 2017. "Mainstreaming Climate Adaptation: Taking Stock about 'What Works' from Empirical Research Worldwide." *Regional Environmental Change*, 1–10. doi:10.1007/s10113-017-1259-5.

Schipper, Lisa. 2006. "Conceptual History of Adaptation in the UNFCCC Process." *RECIEL* 15 (1): 82–92.

Schmidt, Nicole M, and Andreas Fleig. 2018. "Global Patterns of National Climate Policies : Analyzing 171 Country Portfolios on Climate Policy Integration." *Environmental Science and Policy* 84: 177–85. doi:10.1016/j.envsci.2018.03.003.

Schofield, Alexandra, and David Mimno. 2016. "Comparing Apples to Apple: The Effects of Stemmers on Topic Models." *Transactions of the Association for Computational Linguistics* 4: 287–300. https://transacl.org/ojs/index.php/tacl/article/view/868.

Shi, Linda, Eric K. Chu, and Jessica Debats. 2015. "Explaining Progress in Climate Adaptation Planning Across 156 U.S. Municipalities." *Journal of the American Planning Association* 81 (3): 191–202.

Shim, Junseop, Chisung Park, and Mark Wilding. 2015. "Identifying Policy Frames through Semantic Network Analysis: An Examination of Nuclear Energy Policy across Six Countries." *Policy Sciences* 48: 51–83. doi:10.1007/s11077-015-9211-3.

Swart, R., A.G.J. Sedee, F. de Pater, H Goosen, M. Pijnappels, and P. Vellinga. 2014. "Climate-

Proofing Spatial Planning and Water Management Projects: An Analysis of 100 Local and Regional Projects in the Netherlands." *Journal of Environmental Policy & Planning* 16 (1): 55–74.

Tang, Jian, Zhaoshi Meng, XuanLong Nguyen, Qiaozhu Mei, and Ming Zhang. 2014. "Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis." In *Proceedings of the 31st International Conference on Machine Learning*. Vol. 32. International Machine Learning Society.

Thistlethwaite, Jason, and Daniel Henstra. 2017. "Municipal Flood Risk Sharing in Canada: A Policy Instrument Analysis." *Canadian Water Resources Journal* 42 (4): 349–63. doi:10.1080/07011784.2017.1364144.

Uittenbroek, Caroline J., Leonie B. Janssen-Jansen, and Hens A.C. Runhaar. 2013. "Mainstreaming Climate Adaptation into Urban Planning: Overcoming Barriers, Seizing Opportunities and Evaluating the Results in Two Dutch Case Studies." *Regional Environmental Change* 13 (2): 399–411.

Viguié, Vincent, and Stéphane Hallegatte. 2012. "Trade-Offs and Synergies in Urban Climate Policies." *Nature Climate Change* 2 (5): 334–37.

Vilares, David, and Yulan He. 2017. "Detecting Perspectives in Political Debates." In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1573–82. Association for Computational Linguistics.

Wallach, Hanna M. 2006. "Topic Modeling: Beyond Bag-of-Words." In *Proceedings of the 23rd International Conference on Machine Learning*, 977–84. Pittsburgh, PA, USA: International Machine Learning Society. doi:10.1145/1143844.1143967.

Wamsler, Christine, and Stephan Pauleit. 2016. "Making Headway in Climate Policy

Mainstreaming and Ecosystem-Based Adaptation: Two Pioneering Countries, Different Pathways, One Goal." *Climatic Change* 137 (1–2): 71–87. doi:10.1007/s10584-016-1660-y.

Wiedemann, Gregor. 2013. "Opening up to Big Data: Computer-Assisted Analysis of Textual Data in Social Sciences." *Qualitative Social Research* 14 (2): 332–57. doi:10.1016/j.ijhcs.2008.09.006.

Wilkerson, John, and Andreu Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20 (1): 529–44. doi:10.1146/annurev-polisci-052615-025542.

Wilkerson, John, David Smith, and Nicholas Stramp. 2015. "Tracing the Flow of Policy Ideas in Legislatures: A Text Reuse Approach." *American Journal of Political Science* 59 (4): 943–56.

Will, Lowe, Kenneth Benoit, Mikhaylov Slava, and Michael Laver. 2011. "Scaling Policy Preferences from Coded Political Texts." *Legislative Studies Quarterly* 36 (1): 123–55. doi:10.1111/j.1939-9162.2010.00006.x.

Williams, Hywel T.P., James R. McMurray, Tim Kurz, and F. Hugo Lambert. 2015. "Network Analysis Reveals Open Forums and Echo Chambers in Social Media Discussions of Climate Change." *Global Environmental Change* 32: 126–38. doi:10.1016/j.gloenvcha.2015.03.006.

Yin, Zhijun, Liangliang Cao, Jiawei Han, Chengxiang Zhai, and Thomas Huang. 2011. "Geographical Topic Discovery and Comparison." In *Proceedings of the 20th International Conference on World Wide Web*, 247–56. WWW '11. New York, NY, USA: ACM. doi:10.1145/1963405.1963443.