University of Vermont

UVM ScholarWorks

3-1-2017

Transitions in climate and energy discourse between Hurricanes Katrina and Sandy

Emily M. Cody
University of Vermont

Jennie C. Stephens
University of Vermont

James P. Bagrow
University of Vermont

Peter Sheridan Dodds
University of Vermont

Christopher M. Danforth
University of Vermont

CrossMark

Transitions in climate and energy discourse between Hurricanes Katrina and Sandy

Emily M. Cody[1,2,3] · Jennie C. Stephens[4] · James P. Bagrow[2,3] ·
Peter Sheridan Dodds[1,2,3] · Christopher M. Danforth[1,2,3]

Abstract Although climate change and energy are intricately linked, their explicit connection is not always prominent in public discourse and the media. Disruptive extreme weather events, including hurricanes, focus public attention in new and different ways offering a unique window of opportunity to analyze how a focusing event influences public discourse. Media coverage of extreme weather events simultaneously shapes and reflects public discourse on climate issues. Here, we analyze climate and energy newspaper coverage of Hurricanes Katrina (2005) and Sandy (2012) using topic models, mathematical techniques used to discover abstract topics within a set of documents. Our results demonstrate that post-Katrina media coverage does not contain a climate change topic, and the energy topic is limited to discussion of energy prices, markets, and the economy with almost no explicit linkages made between energy and climate change. In contrast, post-Sandy media coverage does contain a prominent climate change topic, a distinct energy topic, as well as integrated representation of climate change and energy, indicating a shift in climate and energy reporting between Hurricane Katrina and Hurricane Sandy.

Keywords Climate change · Energy · Linguistics · Topic modeling · Journalism · Public discourse · Media analysis

✉ Emily M. Cody
emcody@uvm.edu

1 Computational Story Lab, University of Vermont, 16 Colchester Ave, Burlington, VT 05405, USA

2 Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405, USA

3 Department of Mathematics and Statistics, University of Vermont, Burlington, VT 05405, USA

4 Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT 05405, USA

Introduction

Climate change is one of the most challenging issues of our time. Anticipated climate disruptions, including a 4° C increase in the Earth's average temperature by the end of the twenty-first century (IPCC 2014) and more frequent and intense extreme weather events, result from increased atmospheric concentrations of greenhouse gases attributed primarily to fossil fuel burning for energy.

Given probable links between the increasing ocean temperature and the severity and frequency of hurricanes and tropical storms (Mann and Emanuel 2006; Field 2012; Huber 2011), extreme weather events have potential to raise awareness and increase public concern about climate change. The disruptions caused by hurricanes and other storms can also raise awareness and focus attention on energy system vulnerability. These extreme events can serve as a teachable experience for those not previously engaged with these issues (Myers et al. 2013). Indeed, previous research has shown that after experiencing a large hurricane, citizens are more likely to adopt a pro-environmental belief system and support politicians who are climate change activists (Rudman et al. 2013). Populations living as far as 800 km from the path of a hurricane report having experienced it in some way (Howe et al. 2014). Extensive news coverage of extreme weather events has also been found to

increase public awareness of climate change by highlighting tangible and specific risks (Bell 1994; Wilson 2000). It has also been shown that individuals affected by a natural disaster are more likely to strengthen interactions on social media (Phan and Airoldi 2015). As climate change news are prominent on social media (Cody et al. 2015), these interactions provide another mechanism for raising climate change awareness following a natural disaster.

This research recognizes the complex relationship between the news media and public discourse on science and policy. The news media both shapes public perceptions and public discourse and reflects and represents public perceptions and public discourse (Graber 2009; Gamson and Modigliani 1989). The media shapes public opinion of science by avoiding complex scientific language and displaying information for the layperson (Murray et al. 2001; Peterson and Thompson 2009; Priest 2009). People are more likely to learn about environmental and other science-related risks through the media than through any other source (Corbett and Durfee 2004; Peterson and Thompson 2009). Research indicates that news media establish the context within which future information will be interpreted (Peterson and Thompson 2009). In this research, we analyze media coverage to characterize differences in the public discourse about climate change and energy after Hurricane Katrina and Hurricane Sandy.

Links between climate change and energy are often focused on climate mitigation, e.g., reducing greenhouse gas emissions from energy systems by shifting low-carbon energy systems. However, climate change and energy are also linked in terms of increased energy system vulnerability in a changing climate (Stephens et al. 2013). Hurricanes and other extreme weather events often cause disruptions to energy systems including infrastructure damage, fuel supply shortages, and increases in energy prices. Flooding and high wind speeds reveal multiple energy system vulnerabilities including evacuations of oil rigs and power outages at refineries, which can contribute to energy supply shortages and price increases.

Despite the multiple linkages between climate change and energy systems, the issues of climate and energy are still often discussed in the media separately (Stephens et al. 2009; Wilson et al. 2009). Greater integration of the public discourse on climate change and energy could facilitate more sophisticated consideration of the opportunities for changing energy systems to prepare for climate change (IPCC 2014; Metz 2009).

A 2005 study on climate change in the media revealed that articles often frame climate change as a debate, controversy, or uncertainty, which is inconsistent with how the phenomenon is framed within the scientific community (Antilla 2005). A recent 2015 linguistic study determined that the IPCC summaries, intended for non-scientific audiences, are becoming increasingly more complex and more difficult for people to understand (Barkemeyer et al. 2015), which highlights the critical interpretive role of the media in public discourse.

Here, we quantitatively compare media coverage of climate change, energy, and the links between climate and energy after Hurricanes Katrina and Sandy, two of the most disruptive and costly hurricanes to ever hit the USA (Knabb et al. 2006; Blake et al. 2013). Since energy system disruption represents a tangible consequence of climate change, the linking of these two topics in post-hurricane newspaper coverage provides readers with a portal for climate change education and awareness. Newspaper media were selected for analysis rather than social media because in the rapidly changing media landscape the circulation patterns of these well-established newspapers have been relatively stable during the study period. Also, a 2014 study by the American Press Institute determined that 61 % of Americans follow the news through print newspapers and magazines alone. Sixty-nine percent of Americans use laptops and computers which include online newspapers. Eighty-eight percent of Americans find their news directly from a news organization, as opposed to roughly 45 % from social media and 30 % from electronic news ads (Media Insight Project 2014). With this high percentage of Americans getting news from the media, analysis of climate change reporting provides insights on shifts in public discourse and awareness.

We apply two topic modeling techniques stemming from different areas of mathematics to a corpus (collection of text) of newspaper articles about each hurricane. A topic model uses word frequencies within a corpus to assign one or more topics to each text. For our present analysis, we employ latent semantic analysis (LSA), which uses singular value decomposition to reduce a term-document matrix to latent semantic space, and Latent Dirichlet Allocation (LDA), a probabilistic bayesian modeling technique, which defines each hidden topic as a probability distribution over all of the words in the corpus (we provide more details in the methods section, Section Methods).

We apply a topic modeling approach as a way to assess the integration of climate change, energy, and the links between climate and energy within post-hurricane media coverage. Topic modeling is a valuable tool for the kind of research we perform as it does not require manual coders to read thousands of articles. Instead, a specified number of topics are determined through analysis of the frequency of each word in each article in the corpus. The resulting model explains the corpus in detail by categorizing the articles and terms into topics.

We focus on the two most disruptive and costly hurricanes in U.S. history. In August 2005, Hurricane Katrina struck Louisiana as a category 3 storm, affecting the Gulf Coast from central Florida to Texas, causing over 100

billion dollars in damage and roughly 1800 deaths. Katrina destroyed or severely damaged much of New Orleans and other heavily populated areas of the northern Gulf Coast, resulting in catastrophic infrastructure damage and thousands of job losses (Knabb et al. 2006). Hurricane Sandy hit the northeastern USA in October 2012. It was the largest hurricane of the 2012 Atlantic hurricane season, caused 233 reported deaths, and over 68 billion dollars in damage to residential and commercial facilities as well as transportation and other infrastructure (Blake et al. 2013). Many businesses faced short-term economic losses, while the travel and tourism industry experienced long-term economic difficulties. In the time shortly after Sandy hit, repairs and reconstructions were estimated to take 4 years (Henry et al. 2013).

We use this quantitative approach to assess the degree to which climate change or energy-related topics are included in newspaper coverage following Hurricanes Sandy and Katrina. The individual words that define each topic reveal how climate change and energy were represented in post-event reporting, which in turn shapes public discourse.

We first describe the dataset and methods of analysis in Section Methods. We then describe the results of each topic modeling technique for each hurricane and make comparisons between the two corpora in Section Results. We explore the significance of these results in Sections Discussion and Conclusion.

Methods

Data collection

We collected newspaper articles published in major U.S. newspapers in the year following each of the hurricanes. We chose the timespan of 1 year to capture the duration of media coverage following each hurricane and also to ensure we had enough articles from each hurricane to conduct a proper mathematical analysis. We identified newspaper articles through a search that included the name of the hurricane and either the word "hurricane" or "storm" in either the title or leading paragraphs of the article. To account for regional variation in post-hurricane reporting, we chose four newspapers spanning major regions of the USA: northeast, New England, midwest, and west. We chose the following four newspapers due to their high Sunday circulation, and because they are high-profile, established newspapers with high readership: The New York Times, The Boston Globe, The Los Angeles Times, and The Chicago Tribune are influential and well-respected nationally as well as locally. These four newspapers are consistently in the top 25 U.S. Sunday newspapers and were available for article collection through online databases. We collected articles appearing

onwards from the first of the month the hurricane occurred in throughout the subsequent year using the ProQuest, LexisNexis, and Westlaw Campus Research online databases. The total number of articles collected and included in the corpora for analysis are 3100 for Hurricane Katrina and 1039 for Hurricane Sandy. We transform each corpus into a term-document matrix for the analysis.

Latent semantic analysis

LSA is a method of uncovering hidden relationships in document data (Deerwester et al. 1990). LSA uses the matrix factorization technique singular value decomposition (SVD) to reduce the rank of the term-document matrix and merge the dimensions that share similar meanings. SVD creates the following matrices:

$$M = USV^T,$$

where the matrix M is the original $t \times d$ matrix (number of terms by number of documents), the columns of the matrix U are the eigenvectors of MM^T, the entires in the diagonal of the matrix S are the square roots of the eigenvalues of MM^T, and the rows of the matrix V^T are the eigenvectors of $M^T M$. Retaining the k largest singular values and setting all others to 0 gives the best rank k approximation of M. This rank reduction creates a $t \times k$ term matrix, $U_k S_k$, consisting of term vectors in latent semantic space as its columns, and a $k \times d$ document matrix, $S_k V_k^T$, consisting of document vectors as its rows. The documents and terms are then compared in latent semantic space using cosine similarity as the distance metric (Berry and Browne 2005). If two term vectors have cosine distances close to 1, then these terms are interpreted to be related to each other in meaning. We explain this process further in Fig. 1.

We load the documents into a term-document matrix and remove common and irrelevant terms. The terms we removed included terms common to the articles like "hurricane", "storm", "sandy", and "katrina", along with names of authors and editors of the articles. We then convert each frequency in the matrix to term frequency-inverse document frequency (tf-idf) via the following transformation (Baeza-Yates et al. 1999):

$$w_{i,j} = \begin{cases} (1 + \log_2 f_{i,j}) \times \log_2 \frac{N}{n_i} & f_{i,j} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where the variable $w_{i,j}$ is the new weight in the matrix at location (i, j), $f_{i,j}$ is the current frequency in position (i, j), N is the number of documents in the corpus, and n_i is the number of documents containing word i. This weighting scheme places higher weights on rarer terms because they are more selective and provide more information about the corpus, while placing lower weights on common words such as "the" and "and".

a

$$M = {}_{t}\left[\begin{array}{c} d \\ \text{freqs} \end{array}\right]$$

b

$$M = {}_{t}\left[\begin{array}{c} t \\ U \end{array}\right]{}_{t}\left[\begin{array}{c} d \\ S \end{array}\right]{}_{d}\left[\begin{array}{c} d \\ V^{\mathsf{T}} \end{array}\right]$$

c

$$M_k = {}_{t}\left[\begin{array}{c} k \\ U_k \end{array}\right]{}_{k}\left[\begin{array}{c} k \\ S_k \end{array}\right]{}_{k}\left[\begin{array}{c} d \\ V_k^{\mathsf{T}} \end{array}\right]$$

d

$$\text{terms} = {}_{t}\left[\begin{array}{c} k \\ U_k S_k \end{array}\right] \qquad \text{docs} = {}_{k}\left[\begin{array}{c} d \\ S_k V_k^{\mathsf{T}} \end{array}\right]$$

e

$$\text{cosine distance} = \frac{term_1 \bullet term_2}{\|term_1\| \bullet \|term_2\|}$$

Fig. 1 **a** M is a $t \times d$ matrix where t and d are the number of terms and documents in the corpus. An entry in this matrix represents the number of times a specific term appears in a specific document. **b** Singular value decomposition factors the matrix M into three matrices. The matrix S has singular values on its diagonal and zeros everywhere else. **c** The best rank k approximation of M is calculated by retaining the k highest singular values. k represents the number of topics in the corpus. **d** Each term and each document is represented as a vector in latent semantic space. These vectors make up the rows of the term matrix and the columns of the document matrix. **e** Terms and documents are compared to each other using cosine similarity, which is determined by calculating the cosine of the angle between two vectors

We run LSA on the tf-idf term-document matrix for each hurricane. We then compare the documents and terms in the corpus to a given query of terms in latent semantic space. We transform the words that the query is composed of into term vectors and calculate their centroid to give the vector representation of the query. If the query is only one word in length, then the vector representation of the query equals the vector representation of the word. We analyze three queries using LSA: "climate", "energy", and "climate, energy". LSA gives the terms most related to this query vector, which we then use to determine how climate change and energy are discussed both separately and together in the media after Hurricanes Katrina and Sandy.

Latent dirichlet allocation

LDA, a probabilistic topic model (Blei et al. 2003; Blei 2012), defines each hidden topic as a probability distribution over all of the words in the corpus, and each document's content is then represented as a probability distribution over all of the topics. Figure 2 gives illustrations of distributions for a potential LDA model.

LDA assumes that the documents were created via the following generative process. For each document:

1. Randomly choose a distribution of topics from a dirichlet distribution. This distribution of topics contains a nonzero probability of selecting each word in the corpus.
2. For each word in the current document:
   a) Randomly select a topic from the topic distribution in part 1.
   b) Randomly choose a word from the topic just selected and insert it into the document.
3. Repeat until document is complete.

The distinguishing characteristic of LDA is that all of the documents in the corpus share the same set of k topics,

Fig. 2 **a** Examples of two topic distributions that may arise from an LDA model. In this example, each topic is made up of 10 words and each word contributes to the meaning of the topic in a different proportion. **b** Examples of two document distributions that may arise from an LDA model. Document 1 is made up of four major topics, while document 2 is made up of three major topics
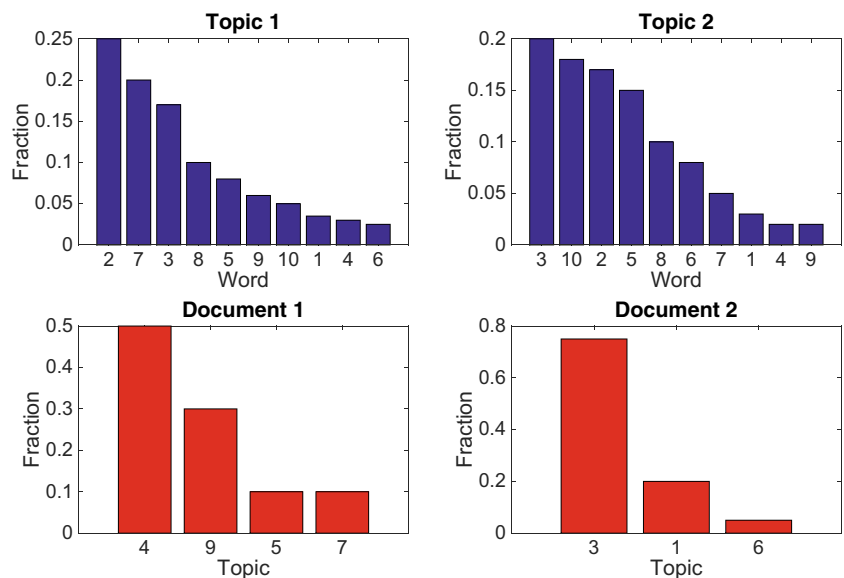
however, each document contains each topic in a different proportion. The goal of the model is to learn the topic distributions. The generative process for LDA corresponds to the following joint distribution:

$$P(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})$$
$$= \prod_{i=1}^{K} P(\beta_i) \prod_{d=1}^{D} P(\theta_d) \left(\prod_{n=1}^{N} P(z_{d,n}|\theta_d) P(w_{d,n}|\beta_{1:K}, z_{d,n}) \right),$$

where β_k is the distribution over the words, $\theta_{d,k}$ is the topic proportion for topic k in document d, $z_{d,n}$ is the topic assignment for the nth word in document d, and $w_{d,n}$ is the nth word in document d. This joint distribution defines certain dependencies. The topic selection $z_{d,n}$ is dependent on the topic proportions each the article, θ_d. The current word $w_{d,n}$ is dependent on both the topic selection, $z_{d,n}$ and topic distribution $\beta_{1:k}$. The main computational problem is computing the posterior. The posterior is the conditional distribution of the topic structure given the observed documents

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}|w_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})}.$$

The denominator of the posterior represents the probability of seeing the observed corpus under any topic model. It is computed using the sampling-based algorithm, Gibbs Sampling.

We generate topic models for the Hurricane Sandy and Katrina articles using LDA-C, developed by Blei in (Blei et al. 2003). We remove a list of common stop words from the corpus, along with common words specific to this corpus such as "Sandy", "Katrina", "hurricane", and "storm". After filtering through the words, we use a Porter word stemmer to stem the remaining words, so each word is represented in one form, while it may appear in the articles in many different tenses (Porter 1980).

Determining the number of topics

The number of topics within a particular corpus depends on the size and scope of the corpus. In our corpora, the scope is already quite narrow as we only focus on newspaper articles about a particular hurricane. Thus, we do not expect the number of topics to be large and to choose the number of topics for the analysis, we implement several techniques.

First, to determine k, the rank of the approximated term-document matrix used in LSA, we look at the singular values determined via SVD. The 100 largest singular values are plotted in Fig. 3 for Hurricanes Sandy and Katrina. The singular value decay rate slows considerably between singular values 20 and 30 for both matrices. We find that topics become repetitive above $k = 20$, and thus we choose $k = 20$ as the rank of the approximated term-document matrix in LSA.

To determine the number of topics for LDA to learn, we use the perplexity, a measure employed in (Blei et al. 2003) to determine how accurately the topic model predicts a sample of unseen documents. We compute the perplexity of a held out test set of documents for each hurricane and vary the number of learned topics on the training data. Perplexity will decrease with the number of topics and should eventually level out when increasing the number of topics no longer increases the accuracy of the model. The perplexity may begin to increase when adding topics causes the model to overfit the data. Perplexity is defined in (Blei et al. 2003) as

$$\text{perplexity}(D_{\text{test}}) = \exp\left\{ -\frac{\sum_{d=1}^{M} \log p(\mathbf{w}_d)}{\sum_{d=1}^{M} N_d} \right\},$$

where the numerator represents the log-likelihood of unseen documents $\mathbf{w}_d$, and the denominator represents the total number of words in the testing set. We separate the data into 10 equal testing and training sets for 10-fold cross validation on each hurricane. We run LDA on each of the 10 different

Fig. 3 The 100 largest singular values in the **a** Hurricane Sandy and **b** Hurricane Katrina tf-idf matrices. The elbow around 20 topics (see *dashed line*) determines the value of k for SVD in LSA
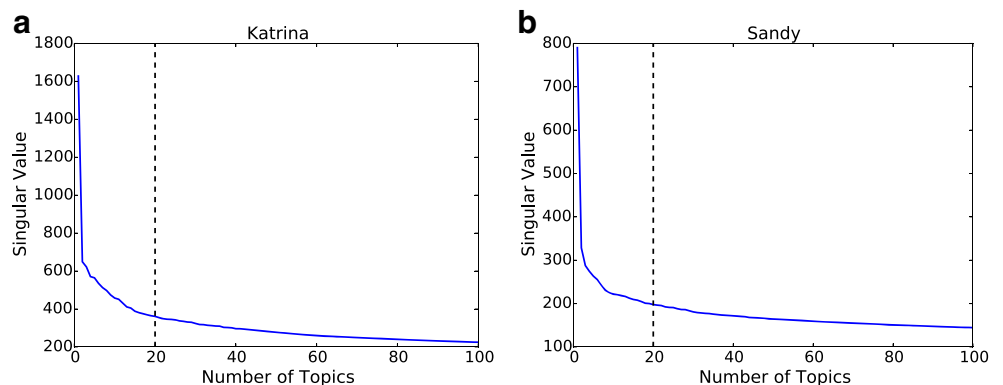
training sets consisting of 90 % of the articles in each hurricane corpus. We then calculate the perplexity for a range of topic numbers on the testing sets, each consisting of 10 % of the articles. We average the perplexity at each topic number over the testing sets and plot the result in Fig. 4a, b.

Figure 4 indicates that the optimal number of topics in the Hurricane Sandy corpus is roughly 20 distinct topics, while the optimal number in the Hurricane Katrina corpus is between 280 and 300 distinct topics. Compared to the Sandy corpus, the Hurricane Katrina corpus contains three times as many articles and about double the number of unique words (17,898 vs. 9521). On average, an article in the Hurricane Sandy corpus contains 270 words, while an article in the Hurricane Katrina corpus contains 376 words. The difference in these statistics may account for the difference in optimal topic numbers in Fig. 4. To test this hypothesis, we take 100 random samples of size 1039 (the size of the Sandy corpus) from the Katrina corpus and calculate the average perplexity over these samples. For each of the 100 random samples, we use 10 testing and training sets for 10-fold

cross validation, as was done in the previous calculations of perplexity. We calculate the average perplexity over the 10 testing sets for each topic number, and then average over the 100 samples for each topic number, showing the result in Fig. 4c. We find that on average, the optimal number of topics for a smaller Katrina corpus is around 30.

Based on the above analysis, we opt to use a 20-topic model for Hurricane Sandy and a 30-topic model for Hurricane Katrina in our LDA analysis of the post-event media coverage.

Results

Latent semantic analysis

We compute a topic model for each corpus using LSA as described in the preceding methods section. We provide 40 words most related to the three queries of interest in Tables 1 and 2. We list the 100 most related words to each

Fig. 4 Average perplexity (over 10 testing sets) vs. number of topics for the full **a** Sandy and **b** Katrina corpora. Perplexity measures how well the model can predict a sample of unseen documents. A lower perplexity indicates a better model. *Dashed lines* show the optimal number of topics. **c** The average perplexity over 100 random samples of 1039 (the size of the Sandy corpus) documents from the Katrina corpus. Each topic number is averaged first over 10 testing sets and then over 100 random samples from the full Katrina corpus. Topic numbers increase by 2. Error bars indicate the 95 % confidence intervals
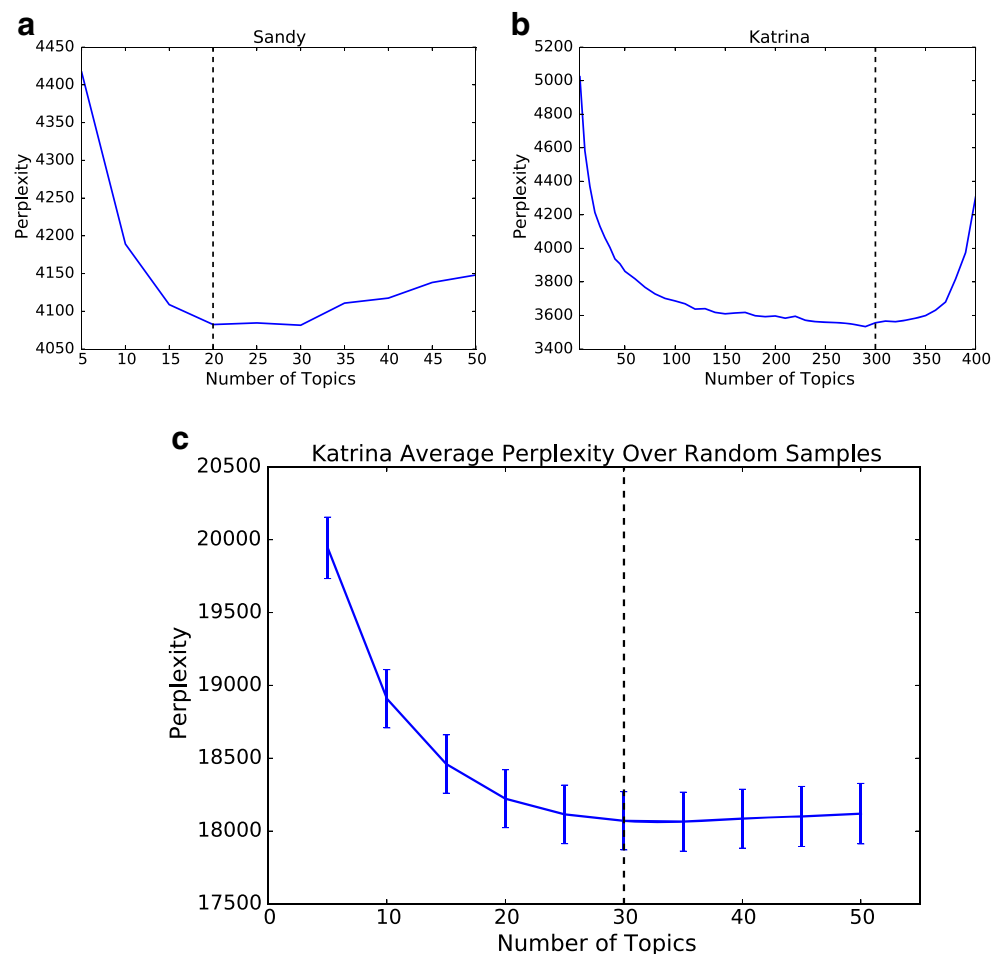
Table 1 Results of LSA for Hurricane Katrina for three different queries. Words are ordered based on their cosine similarity with the query vector

Hurricane Katrina

"climate"	Similarity	"energy"	Similarity	"climate,energy"	Similarity
climate	1.000	energy	1.000	energy	0.979
larger	0.866	prices	0.986	prices	0.952
destroy	0.861	exchange	0.968	deutsche	0.945
formally	0.848	consumers	0.966	price	0.943
theory	0.844	weinberg	0.966	underinvestment	0.943
sound	0.837	argus	0.964	signaling	0.941
gale	0.826	reidy	0.962	discounting	0.940
reinforced	0.817	splurge	0.960	java	0.940
journal	0.815	hummer	0.960	argus	0.939
sensitive	0.814	markets	0.959	hummer	0.938
unlikely	0.812	downers	0.958	oil	0.937
belief	0.809	highs	0.958	consumers	0.937
phenomenon	0.809	underinvestment	0.957	shocks	0.934
rail	0.800	exporting	0.954	weinberg	0.934
studying	0.796	price	0.954	markets	0.934
wealthy	0.795	reserves	0.954	profits	0.931
brings	0.792	signaling	0.953	reserves	0.931
barge	0.792	dampening	0.950	exchange	0.931
ancient	0.791	oil	0.950	peaks	0.931
masters	0.786	java	0.949	highs	0.929
politicians	0.785	cents	0.948	splurge	0.927
professor	0.783	deutsche	0.948	exporting	0.927
recommendations	0.782	gasoline	0.947	gasoline	0.923
thick	0.782	traders	0.946	dampening	0.923
marked	0.780	nariman	0.946	pinch	0.922
alter	0.779	discounting	0.945	oils	0.922
sounds	0.776	behravesh	0.944	soaring	0.922
hole	0.776	retailers	0.943	exported	0.920
peril	0.775	barrel	0.942	reidy	0.919
extremely	0.771	heating	0.942	output	0.919
avoided	0.770	oils	0.942	exporter	0.917
loose	0.770	shocks	0.941	easing	0.917
multi	0.769	idled	0.941	putins	0.917
appear	0.767	jolted	0.941	record	0.916
devastating	0.766	output	0.940	tumbling	0.916
draft	0.764	peaks	0.937	demand	0.915
possibility	0.764	profits	0.936	downers	0.915
roiled	0.759	soared	0.936	automaker	0.913
retracted	0.758	exported	0.936	heating	0.913
mismanagement	0.758	premcor	0.935	disruptions	0.913

query in the Supplementary Materials (see Tables 5 and 6). While it is not possible to objectively explain why each word ranks where it does in the following lists, we search for a common theme within the words to determine how climate and energy were discussed in the media following these hurricanes.

Table 2 Results of LSA for Hurricane Sandy for three different queries. Words are ordered based on their cosine similarity with the query vector

Hurricane Sandy

"climate"	Similarity	"energy"	Similarity	"climate,energy"	Similarity
climate	1.000	energy	1.000	climate	0.979
change	0.963	technologies	0.949	warmer	0.961
reduce	0.957	fuels	0.946	georgetown	0.956
warming	0.957	fossil	0.943	warming	0.955
reducing	0.956	hydroelectric	0.936	reduce	0.955
pressures	0.952	renewable	0.932	energy	0.952
georgetown	0.947	rogue	0.932	reducing	0.951
lowering	0.943	employing	0.921	pressures	0.948
talks	0.942	warmer	0.920	fossil	0.947
devise	0.938	supplying	0.918	fuels	0.946
expands	0.938	firing	0.913	change	0.946
outweigh	0.937	efficiency	0.911	technologies	0.945
warmer	0.937	streamlined	0.911	coal	0.943
plants	0.934	generating	0.908	global	0.942
drought	0.933	altering	0.906	hydroelectric	0.941
manipulation	0.929	coal	0.906	emissions	0.940
emissions	0.929	consumption	0.900	firing	0.937
global	0.929	adapt	0.898	outweigh	0.936
imperative	0.927	sparked	0.895	generating	0.933
arizona	0.924	dimming	0.894	carbon	0.930
attribute	0.923	georgetown	0.892	arizona	0.930
scientists	0.923	carbon	0.889	editorials	0.929
planet	0.920	masonry	0.888	plants	0.927
pollution	0.919	global	0.886	humanitys	0.926
curbing	0.918	erratic	0.885	altering	0.926
coal	0.917	searchable	0.884	manipulation	0.924
editorials	0.915	faster	0.882	pollution	0.923
targets	0.914	emissions	0.881	employing	0.923
oceans	0.912	skeptics	0.880	drought	0.922
vigil	0.912	proportion	0.877	extracted	0.921
scenarios	0.911	trillions	0.876	foretaste	0.920
extracted	0.911	foretaste	0.876	skeptics	0.919
humanitys	0.911	warming	0.875	lowering	0.919
distraction	0.910	reduce	0.875	dioxide	0.918
pentagon	0.910	editorials	0.875	efficiency	0.918
contiguous	0.909	humanitys	0.875	planet	0.917
controlling	0.908	eco	0.875	curbing	0.917
carbon	0.907	ton	0.874	consumption	0.915
dioxide	0.906	efficient	0.872	expands	0.914
extremes	0.905	cities	0.872	subtler	0.913

Hurricane Katrina

Within the Hurricane Katrina news media coverage, explicit reference to climate change was infrequent. The set of words most related to "climate" includes words such as "theory", "unlikely", "belief", and "possibility", indicating that linkages with climate change after Hurricane Katrina were tentative. The uncertain link between hurricanes and climate change is often present in political discussions, thus the appearance of the word "politician" in the "climate" list

is not surprising. A direct quote from the article most related to the "climate" query reads:

> "When two hurricanes as powerful as Katrina and Rita pummel the Gulf Coast so close together, many Americans are understandably wondering if something in the air has changed. Scientists are wondering the same thing. The field's leading researchers say it is too early to reach unequivocal conclusions. But some of them see evidence that global warming may be increasing the share of hurricanes that reach the monster magnitude of Katrina, and Rita" (Brownstein 2005).

Words such as "studying", "professor", and "masters" also indicate that reporting on climate change focused on research and academics. The "climate" list does not contain words relating to energy or energy systems and does not focus on the science or consequences of climate change.

Within the 40 words most related to the "energy" query, the majority pertain to energy prices and the stock market. Within the "climate" and "energy" lists, there is no overlap in the 40 most related words to these queries.

The "climate" and "energy" vectors are averaged to create the "climate, energy" query vector. The list of words most similar to this query is far more comparable to the "energy" list than the "climate" list. Of the 100 most related words to each query, there are 84 shared words between the "energy" and "climate, energy" lists. This list again focuses on energy prices and not at all on climate change or infrastructure vulnerability, indicating that discussions about climate change, energy, and power outages were independent of one another within media reporting following Hurricane Katrina.

Hurricane Sandy

In the Hurricane Sandy corpus, we find the word "climate" is most related to words describing climate change and global warming. We also see words related to energy such as "emissions", "coal", "carbon", and "dioxide". Including the top 100 words most related to "climate", we see more energy related words including "fossil", "hydroelectric", "technologies", and "energy" itself. This list differs substantially from that of the Hurricane Katrina analysis.

The word "energy" in the Hurricane Sandy corpus is most related to words describing climate change, such as the contributions of fossil fuels and the potential of renewable ("hydroelectric", "renewable") energy resources. This list of words focuses largely on how energy consumption is contributing to climate change, and, unlike the Katrina corpus, considerably overlaps with the list of "climate" words.

Of the 100 words most related to "energy", 58 of them are also listed in the 100 words most related to "climate". Of the 20 documents most related to the word "energy", 15

of them are also listed in the 20 documents most related to "climate". Many of these articles discuss harmful emissions, renewable energy, and fossil fuels.

In the Hurricane Sandy corpus, the "climate, energy" query is again most related to the climate change and global warming-related terms. There are 87 shared terms in the "climate" and "climate, energy" lists and 66 shared terms in the "energy" and "climate, energy" related lists. This result illustrates that when climate change was discussed in the media following Hurricane Sandy, energy-related themes were often present.

Latent dirichlet allocation

We generate LDA models for both the Sandy and Katrina corpora using 20 topics and 30 topics for Sandy and Katrina respectively (see Methods). The 20 most probable words in 10 selected topic distributions are given in Tables 3 and 4. The full models are given in the Supplementary Materials (see Tables 7 and 8). In addition to creating a distribution of topics over words, LDA also creates a distribution of documents over topics. Each topic is present in each document with some nonzero probability. We counted the number of times each topic appeared as one of the top two ranked topics in an article and divided this number by the number of articles in the corpus. Figure 5 summarizes the overall results of LDA for Katrina (a) and Sandy (b) by giving the proportion of articles that each topic appears in with high probability. We determined the topic names by manually analyzing the probability distribution of words in each topic. We go into more detail on the topics of importance in the following sections.

Hurricane Katrina

In Table 3, we give 10 of the 30 topics in the LDA model for Hurricane Katrina. In the Hurricane Katrina model, we see topics relating to deaths, relief, insurance, flooding, and energy. We also see location-specific topics such as sporting events, Mardi Gras, and music. A major topic that is absent from this model is climate change. Similar to the results we saw for the Katrina LSA model, the energy topic (topic 8) in the Katrina LDA model contains words relating to energy prices, the market, and the economy. In addition to a missing climate change topic, there is no mention of the climate within topic 8 either, indicating that Hurricane Katrina did not only lack in climate change reporting but it also did not highlight the link between climate change and energy.

Hurricane Sandy

In the Hurricane Sandy LDA model, we see topics related to medics, insurance, fundraisers, government, damage, power

Table 3 The 20 most probable words within 10 of the 30 topic distributions given by LDA for Hurricane Katrina. The words are stemmed according to a Porter stemmer (Porter 1980), where for example *flooded*, *flooding*, and *floods* all become *flood*

Hurricane Katrina

7: deaths	8: energy	12: relief	13: family	14: mardi gras	16: shows	19: travel	21: insurance	28: evacuation	29: response
bodi	price	red	famili	gras	music	ship	insur	hous	fema
death	oil	cross	home	mardi	jazz	airlin	flood	evacue	respons
offici	percent	donat	children	french	band	show	damag	fema	feder
state	energi	relief	day	restaur	musician	news	billion	peopl	agenc
home	gas	organ	live	parad	art	time	state	offici	brown
die	gasolin	volunt	back	street	cultur	northrop	compani	home	disast
victim	rate	victim	school	back	museum	network	loss	houston	govern
peopl	market	fund	mother	peopl	perform	travel	mississippi	feder	emerg
famili	week	peopl	friend	quarter	play	air	home	agenc	secur
parish	product	million	peopl	time	festiv	nbc	homeown	hotel	offici
st	month	chariti	im	home	artist	million	pay	trailer	homeland
louisiana	consum	disast	call	day	song	broadcast	claim	famili	hous
identifi	report	american	hous	citi	work	report	cost	state	depart
morgu	economi	money	stay	make	show	abc	allstat	shelter	report
relat	compani	group	time	club	time	cruis	area	emerg	manag
coron	increas	rais	dont	louisiana	concert	program	properti	live	chertoff
dr	gulf	effort	work	cook	includ	film	louisiana	month	white
dead	fuel	food	life	krew	orchestra	channel	industri	apart	bush
found	expect	org	son	hotel	event	televis	feder	govern	plan
remain	gallon	shelter	left	celebr	record	navi	polici	assist	investig

Table 4 The 20 most probable words within 10 of the 20 topic distributions given by LDA for Hurricane Sandy. The words are stemmed according to a Porter stemmer (Porter 1980)

Hurricane Sandy

0: utility	1: election	2: climate	3: community	7: transportation	8: medical	9: insurance	12: impact	13: media	15: fundraising
power	obama	climat	hous	train	hospit	insur	wind	show	concert
util	romney	flood	home	author	home	compani	power	time	perform
servic	presid	chang	water	station	patient	percent	day	stewart	ticket
compani	campaign	protect	beach	line	health	sale	close	peopl	music
author	elect	build	car	servic	medic	month	weather	make	show
electr	state	rise	live	tunnel	nurs	market	coast	photo	million
island	republican	sea	flood	jersey	evacu	busi	expect	live	money
custom	vote	water	peopl	gas	emerg	increas	servic	twitter	benefit
state	polit	risk	point	transport	center	million	travel	call	hall
system	governor	level	fire	power	dr	loss	area	work	rais
grid	voter	energi	street	damag	peopl	industri	offici	news	song
long	day	natur	rockaway	subway	citi	home	peopl	stori	peopl
verizon	poll	power	back	street	offici	report	state	includ	night
nation	democrat	weather	day	manhattan	resid	expect	damag	inform	work
work	peopl	develop	insur	offici	island	billion	flood	magazin	relief
phone	debat	make	damag	transit	day	rate	nation	photograph	refund
commiss	candid	cost	resid	long	care	week	massachusett	design	springsteen
network	presidenti	state	work	system	bird	retail	center	post	jersey
con	time	plan	famili	day	mayor	consum	report	print	sale
edison	nation	surg	neighborhood	island	mold	claim	hour	page	band

a

Fraction of Katrina Articles About Each Topic

b

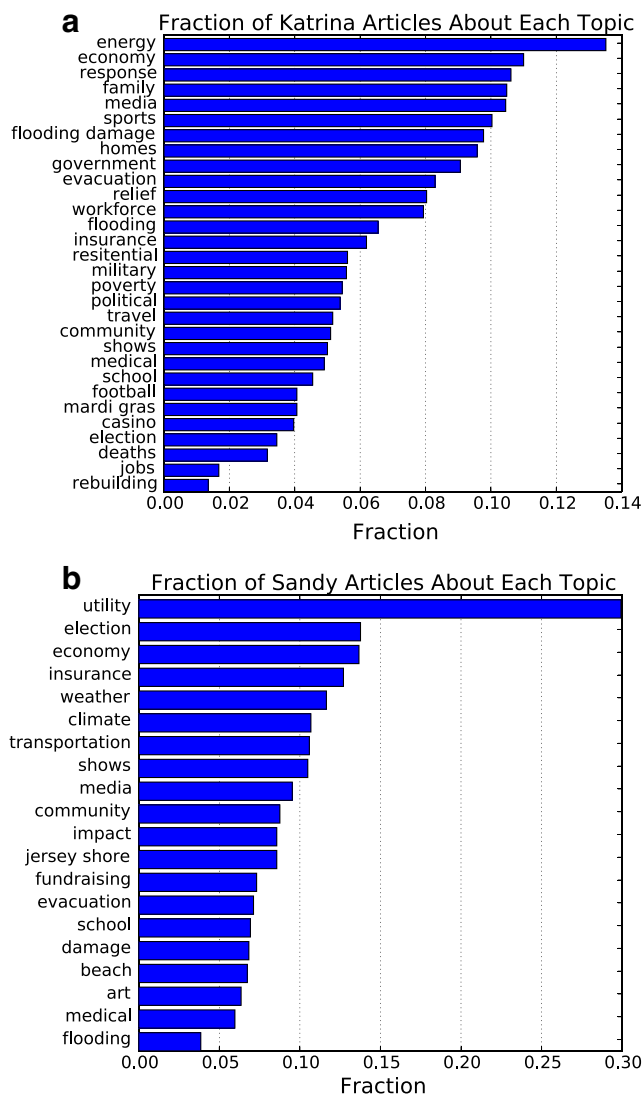### Fraction of Sandy Articles About Each Topic

Fig. 5 The proportion of articles ranking each topic as the first or second most probable topic, i.e., the proportion of articles that each topic appears in with high probability in the **a** Hurricane Katrina and **b** Hurricane Sandy corpora. The topics order is by decreasing proportions

outages, and climate change. Unlike the Katrina model, we find that topic 2 clearly represents climate change. Words such as "flood", "weather", and "natural" indicate that the reporting on climate change within articles about Hurricane Sandy discussed how climate change is contributing to weather extremes and natural disasters. There was also considerable reporting on the rising sea levels, which are expected to contribute to the intensity of hurricanes and tropical storms (Michener et al. 1997).

Dispersed throughout the weather-related words in topic 2, we see the words "energy", "power", and "develop", indicating that power outages and energy system development were often discussed within articles that mentioned climate change, highlighting a link between climate change and

the energy disruption caused by Hurricane Sandy. Extending the number of words in topic 2, we find more energy related words including "infrastructure" (23), "carbon" (28), "resilience" (35), and "emissions" (37). A list of the 100 most probable words in topic 2 is given in the Supplementary Information. While "carbon" and "emissions" are clearly linked to climate change, words like "infrastructure" and "resilience" indicate a link between climate change discussion and energy system vulnerability.

Topic 0 also contains words pertaining to energy systems. This topic, however, does not contain any words pertaining to climate change. Topic 0 is about electricity ("company", "electricity", "system"), power outages ("power","utility", "service"), and communication ("verizon", "phone", "network"). One benefit of LDA is that the model not only creates distributions of words over topics but also creates distributions of topics over documents. Of the 162 articles that are made up of more than 1 % topic 2, 24 of them also contain topic 0, demonstrating that these two topics were sporadically reported on in the same article. For example, an article in *The New York Times* entitled "Experts Advise Cuomo on Disaster Measures" discusses how New York City can better prepare for drastic outages caused by extreme weather and directly quotes Governor Cuomo's concerns about climate change:

> " 'Climate change is dramatically increasing the frequency and the severity of these situations,' Mr. Cuomo said. 'And as time goes on, we're more and more realizing that these crises are more frequent and worse than anyone had predicted.' " (Kaplan 2013)

Although the models for each hurricane generate some similar topics, there are some topics in one model that do not appear in the other. Both models give topics on politics, community, government aid, fundraisers, insurance, family, travel, medics, flooding, damage, evacuations, and energy. The Hurricane Katrina model also gives topics relating to sporting events, Mardi Gras, music, military, and the death toll, while the Sandy model gives topics relating to museums, beaches, weather, broadway, and climate change. Many of the topics only appearing in one of the models appear there due to the hurricane's location. The climate change topic, however, appears only in the Hurricane Sandy corpus and its absence in the Hurricane Katrina corpus cannot be simply be a consequence of the different locations of the hurricanes.

Discussion

Through this analysis using topic models, we discover that climate change and energy were often discussed together within coverage of Hurricane Sandy, whereas the climate

change topic is largely absent in post Hurricane Katrina reporting. This difference can be attributed in part to changing public perceptions about climate change over time. As early as 2001, the scientific consensus that climate change is occurring and resulting from human activity was legitimized by the IPCC assessment reports (Griggs and Noguer 2002). A 2003 national study on climate change risk perceptions, however, revealed that while most Americans demonstrate awareness of climate change, 68 % considered it only a moderate risk issue more likely to impact areas far from the USA (Leiserowitz 2005). In Fall 2008 (years after Hurricane Katrina), 51 % of Americans were either alarmed or concerned about global warming (Maibach et al. 2011), and in March 2012 (months before Hurricane Sandy), this number decreased to 39 % (Leiserowitz et al. 2012b). In April 2013, 38 % of Americans believed that people around the world are currently affected or harmed by the consequences of climate change (Leiserowitz et al. 2013). Those in the "alarmed" and "concerned" categories are also far more likely to report that they experienced a natural disaster within the last year (Leiserowitz et al. 2012b), implying a potential relationship between personal experience of consequences and the perception of climate change risks (Myers et al. 2013). Participants in the Yale School of Forestry & Environmental Studies "Americans and Climate Change" conference in 2005 determined that since science is the main source of climate change information, there is room for misinterpretation and disconnects in society's understanding of the issue (Abbasi 2006).

The 2004 and 2005 Atlantic hurricane seasons were among the costliest in United States history (Beven et al. 2008). In 2004, scientists began to propose that the intensity of the latest hurricane season may be linked to global warming. However, the state of climate science at the time could not support such a hypothesis, and linkages between global warming and the impacts of hurricanes were deemed premature (Pielke et al. 2005). Media coverage of climate change often presents the scientific consensus and has influenced public opinion and risk perceptions on climate change (Antilla 2008). Complexity and uncertainty within the scientific community regarding the link between climate change and hurricanes may be why climate change does not appear as a prominent topic in the 2005 news media analysis of Hurricane Katrina.

Conversely, media reporting following Hurricane Sandy did connect explicitly with climate change. By the time Hurricane Sandy occurred in 2012, climate science research had progressed and begun exploring the link between hurricanes and global warming (Mann and Emanuel 2006; Field 2012; Huber 2011). The Yale Project on Climate Change and Communications poll in March 2012 showed that a large majority of Americans believed at that time that certain weather extremes and natural disasters are caused by global warming (Leiserowitz et al. 2012a). This evolution of climate change research and public awareness is reflected in the different coverage of climate change after Hurricane Sandy.

Also unique to Hurricane Sandy coverage was the presence of climate and energy topics together. While Hurricane Katrina reporting focused on the increase in energy prices following the storm, this increase in price was not explicitly linked to the consequences of climate change within media reporting. Hurricane Katrina caused massive disruptions in oil and gas production in the Gulf of Mexico, which caused large spikes in the cost of oil and natural gas. During Katrina, 2.6 million customers lost power in Louisiana, Mississippi, Alabama, Florida, and Georgia (Energy UD 2005). The destruction caused by Katrina (followed shortly after by Hurricane Rita) encouraged drilling companies to upgrade their infrastructure to better withstand the forceful waves and wind from a large hurricane (Heidrick 2013). During Hurricane Sandy, 8.66 million customers lost power from North Carolina to Maine, and it took 10 days for the utilities to restore power to 95 % of these affected customers. Reporting on these outages is reflected in the LDA climate change topic. Flooding and power outages at refineries, pipelines, and petroleum terminals in the New York Harbor area lead to gasoline shortages and prices increases (Energy UD 2013). These impacts illustrated some of the consequences of climate change and an increase in severity of natural disasters. Hurricane Sandy news reporting not only highlighted the consequences of climate change but also the relationship between climate change, energy, and energy system vulnerability.

Conclusion

Given that the media both shapes and reflects public discourse, this analysis characterizing stark differences in media coverage between Hurricane Katrina and Hurricane Sandy demonstrates a shift in public discourse on climate change and energy systems. Although energy systems were disrupted in both storms, the connections between energy and climate change were made much more explicitly in the post-Hurricane Sandy news coverage as compared to the post-Hurricane Katrina coverage. This shift is likely to represent multiple changes including: (1) increased public awareness and concern about climate change, (2) improved scientific understanding of the link between hurricane intensity and climate change, and (3) greater understanding of the energy system risks associated with climate change. The ways that climate and energy are connected in the media coverage also reflects a larger shift toward increasing attention toward climate change adaptation in addition to climate mitigation (Hess 2013).

Our investigation presents a mathematical approach to assessing public discourse of climate and energy, one that could be applied to assessing news media of other key areas in environmental studies. This analysis focuses on Hurricanes Katrina and Sandy due to their disruption and societal impact as focusing events. Future research could expand to investigate how energy and climate are presented in other climate- and energy-related media coverage over time.

References

Abbasi DR (2006) Americans and climate change: closing the Gap between Science and Action. Yale school of forestry & environmental studies publication series

Antilla L (2005) Climate of scepticism: Us newspaper coverage of the science of climate change. Glob Environ Chang 15(4):338–352

Antilla L (2008) Self-censorship and science: a geographical review of media coverage of climate tipping points. Public Understanding of Science

Baeza-Yates R, Ribeiro-Neto B et al (1999) Modern information retrieval, vol 463. ACM press, New York

Barkemeyer R, Dessai S, Monge-Sanz B, Renzi BG, Napolitano G (2015) Linguistic analysis of IPCC summaries for policymakers and associated coverage. Nature Climate Change. doi:10.1038/nclimate2824

Bell A (1994) Media (mis) communication on the science of climate change. Public Underst Sci 3(3):259–275

Berry MW, Browne M (2005) Understanding search engines: mathematical modeling and text retrieval, vol 17. Siam

Beven JL, Avila LA, Blake ES, Brown DP, Franklin JL, Knabb RD, Pasch RJ, Rhome JR, Stewart SR (2008) Atlantic hurricane season of 2005. Mon Weather Rev 136(3):1109–1173

Blake ES, Kimberlian TB, Berg RJ, Cangialosi JP, Beven JL (2013) Tropical cyclone report, hurricane sandy. National Hurricane Center

Blei DM (2012) Probabilistic topic models. Commun ACM 55(4):77–84

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

Brownstein R (2005) Hard choices blow in the winds of katrina, and now rita. The Los Angeles Times, http://search.proquest.com/docview/422011345?accountid=14679

Cody EM, Reagan AJ, Mitchell L, Dodds PS, Danforth CM (2015) Climate change sentiment on twitter: an unsolicited public opinion poll. PLoS ONE 10(8)

Corbett JB, Durfee JL (2004) Testing public (un) certainty of science media representations of global warming. Sci Commun 26(2):129–151

Deerwester SC, Dumais ST, Landauer TK, Furnas GW, Harshman RA (1990) Indexing by latent semantic analysis. JASIS 41(6):391–407

Energy UD (2005) Hurricane Katrina situation report #11. Office of Electricity Delivery and Energy Reliability (OE)

Energy UD (2013) Comparing the impacts of northeast hurricanes on energy infrastructure. Office of Electricity Delivery and Energy Reliability (OE)

Field CB (2012) Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change

Gamson WA, Modigliani A (1989) Media discourse and public opinion on nuclear power: a constructionist approach. Am J Sociol:1–37

Graber DA (2009) Mass media and American politics. SAGE

Griggs DJ, Noguer M (2002) Climate change 2001: the scientific basis. contribution of working group i to the third assessment report of the intergovernmental panel on climate change. Weather 57(8):267–269

Heidrick R (2013) Hurricane season could bring higher energy prices. Texas Enterprise

Henry DK, Cooke-Hull S, Savukinas J, Yu F, Elo N, Vac Arnum B (2013) Economic impact of Hurriane Sandy: potential economic activity lost and gained in New Jersey and New York. Tech. rep., U.S. Department of Commerce

Hess DJ (2013) Transitions in energy systems: the mitigation–adaptation relationship. Sci Cult 22(2):197–203

Howe PD, Boudet H, Leiserowitz A, Maibach EW (2014) Mapping the shadow of experience of extreme weather events. Clim Chang 127(2):381–389

Huber DG (2011) Gulledge J, Extreme weather and climate change: understanding the link, managing the risk. Pew Center on Global Climate Change Arlington

IPCC (2014) Climate change 2014 mitigation of climate change, intergovernmental panel on climate change, https://www.ipcc.ch/report/ar5/wg3/

Kaplan T (2013) Experts advise Cuomo on disaster measures. The New York Times

Knabb R, Rhome J, Brown D (2006) Tropical cyclone report?hurricane katrina. National hurricane center, Miami

Leiserowitz A, Maibach E, Roser-Renouf C, Hmielowski J (2012a) Extreme weather, climate & preparedness in the american mind. Yale University and George Mason University New Haven, CT)(Report)

Leiserowitz A, Maibach E, Roser-Renouf C, Smith N (2012b) Global warmings six americas, march 2012 and november 2011. Yale University and George Mason University

Leiserowitz A, Maibach E, Roser-Renouf C, Feinberg G, Howe P (2013) Climate change in the american mind: Americans' global warming beliefs and attitudes in April. Yale University and George Mason University New Haven, CT: Yale Project on Climate Change Communication

Leiserowitz AA (2005) American risk perceptions: is climate change dangerous? Risk Anal 25(6):1433–1442

Maibach EW, Leiserowitz A, Roser-Renouf C, Mertz C (2011) Identifying like-minded audiences for global warming public engagement campaigns: an audience segmentation analysis and tool development. PloS one e17(3):571

Mann ME, Emanuel KA (2006) Atlantic hurricane trends linked to climate change. Eos, Trans Am Geophys Union 87(24):233–241

Media Insight Project (2014) How Americans get their news. The Personal News Cycle

Metz B (2009) Controlling climate change. Cambridge University Press

Michener WK, Blood ER, Bildstein KL (1997) Climate change, hurricanes and tropical storms, and rising sea level in coastal wetlands, vol 7, Gardner LR, pp 770–801

Murray D, Schwartz JB, Lichter SR (2001) It ain't necessarily so: how media make and unmake the scientific picture of reality. Rowman & Littlefield

Myers TA, Maibach EW, Roser-Renouf C, Akerlof K, Leiserowitz AA (2013) The relationship between personal experience and belief in the reality of global warming. Nat Clim Chang 3(4):343–347

Peterson T, Thompson J (2009) Environmental risk communication: responding to challenges of complexity and uncertainty. Handbook of risk and crisis communication. Routledge, New York, p 591Á606

Phan TQ, Airoldi EM (2015) A natural experiment of social network formation and dynamics. Proc Natl Acad Sci 112(21):6595–6600

Pielke RA, Landsea C, Mayfield M, Laver J, Pasch R (2005) Hurricanes and global warming. Bull Am Meteorol Soc 86(11):1571–1575

Porter MF (1980) An algorithm for suffix stripping. Program 14(3):130–137

Priest SH (2009) Doing media research: an introduction. Sage

Rudman LA, McLean MC, Bunzl M (2013) When truth is personally inconvenient, attitudes change the impact of extreme weather on implicit support for green politicians and explicit climate-change beliefs

Stephens JC, Rand GM, Melnick LL (2009) Wind energy in us media: a comparative state-level analysis of a critical climate change mitigation technology. Environ Commun 3(2):168–190

Stephens JC, Wilson EJ, Peterson TR, Meadowcroft J (2013) Getting smart? climate change and the electric grid. Challenges 4(2):201–216

Wilson EJ, Stephens JC, Peterson TR, Fischlein M (2009) Carbon capture and storage in context: the importance of state policy and discourse in deploying emerging energy technologies. Energy Procedia 1(1):4519–4526

Wilson KM (2000) Drought, debate, and uncertainty: measuring reporters' knowledge and ignorance about climate change. Public Underst Sci 9(1):1–13