# Signals of Doubt: Text-Mining Climate Skepticism

Constantine Boussalis and Travis G. Coan*

Harvard Law School

October 24, 2013

Working paper presented at the "New Directions in Analyzing Text as Data Workshop", London School of Economics, 27-28 September 2013.

**Abstract**

Although agreement among scientists on anthropogenic climate change is clear, national surveys show that the American public's perceptions on the science of climate change diverge significantly from the "consensus view." Explanations for this divergence range from the influence of local weather events to the health of the economy, while many place blame on the efforts of climate change deniers to sway public opinion away from mainstream climate science. This study examines the dynamics of climate change "contrarianism" in the American context by analyzing the content of major conservative think tanks on the issue from the mid 1990s to the present. Specifically, we rely on an unsupervised text classification algorithm to produce time series measures of climate skepticism. Our results highlight how even simple uses of recent advances in natural language processing provide insight into key questions in the literature on climate change contrarianism.

*Keywords:* climate change, skeptics, text classification, latent Dirichlet allocation

## 1 Introduction

Climate scientists resoundingly agree that the Earth is getting warmer and that the rise in average temperature is predominantly due to human activity. The

* *Contact*: 410 Areeda Hall, Harvard Law School, 1545 Massachusetts Avenue, Cambridge, MA 02138. tcoan@law.harvard.edu

Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC) states that, "warming of the climate system is unequivocal", and that "it is extremely likely that human activities have exerted a substantial net warming influence on climate since 1750"(Solomon et al. 2007). Similar statements have been made by major scientific organizations. For example, in the United States, the National Academy of Sciences concurs, stating that, "there is a strong, credible body of evidence, based on multiple lines of research, documenting that climate is changing and that these changes are in large part caused by human activities" (National Research Council 2010). In their survey of a representative sample of Earth scientists, Doran and Zimmerman (2009) find that 96.2% of respondents who are active climate researchers agree that mean global temperatures relative to pre-1800s levels have risen, and 97.4% of the same group agree that human activity is a significant contributor to the changing average global temperature. The authors conclude by stating that, "it seems that the debate on the authenticity of global warming and the roles played by human activity is largely nonexistent among those who understand the nuances and scientific basis of long-term climate processes"(Doran and Zimmerman 2009, p. 23). Anderegg et al. (2010) find that among 1,372 active climate researchers, 97-98% agree with the main tenets of anthropogenic climate change as expressed in the IPCC Fourth Assessment Report. [1] In a recent survey of 11,944 academic articles on climate change over the period 1991-2011, Cook et al. (2013) find that among the studies that express an opinion on anthropogenic global warming, over 97% agree with the consensus view that increases in the average global temperature have been caused by human greenhouse gas emissions, whereas less than 2% reject anthropogenic global warming.

While a strong consensus among climate scientists regarding human-induced rising global temperatures appears to be a reality, perceptions among the American public on climate change diverge significantly from the "consensus view". In 2012, 41% of Americans believed that increases in the Earth's temperature are determined by "effects of natural changes in the environment that are not due to human activities", while 32% believed that scientists are "unsure" about whether global warming is occurring, and 42% held the view that the seriousness of global warming is "generally exaggerated" in the news (up from 30% in 2006) (Gallup News Service 2012). Americans also seem to be global leaders in opposing mainstream climate science. In 2010, a survey of 150 countries found that the United States had the largest share of respondents (47%) who primarily attribute rising global temperatures to natural causes (Ray and Pugliese 2011).

What explains this chasm in understanding of global warming between climate change experts and the general American public? This question has been explored extensively in the literature (e.g., Immerwahr 1999, Sterman and Sweeney 2002, Krosnick et al. 2006, Brody et al. 2008, Kellstedt et al. 2008, Borick and Rabe 2010). Explanations of the divergence between beliefs on cli-

[1]Specifically, that it is "very likely" that human-made greenhouse gas emissions have been responsible for "most" of the "unequivocal" warming of the average global temperature in the second half of the 20th century (Anderegg et al. 2010, p. 12107)

mate change held by scientists with those of large segments of the U.S. population abound. Some have argued that short-term weather phenomena can have significant impacts on lay opinion on climate science if affected people believe that recent local temperatures are abnormally high (Li et al. 2011, Egan 2012, Krosnick et al. 2006). Others argue that the state of the economy influences public concern for climate change. As economic conditions, such as unemployment, worsen, studies have shown that the public tends to attribute a lower priority to climate change (Guber 2003, Smith et al. 2011, Scruggs and Benegal 2012). Empirical evidence also suggests that information attribution and elite cues influence public opinion on global warming Guber (2013).

One prominent explanation–and the emphasis of the current study–is the influence of skeptical segments of the conservative movement on the public's understanding of climate science. In this paper, we apply automated text analytic methods to systematically understand what arguments climate change contrarians have put forth into the public sphere and how the prevalence of these arguments has evolved over time.

# 2 Understanding 21st Century Climate Change Contrarianism

This study focuses on the role of segments of the American conservative movement that have targeted climate science and environmental policy proposals which seek to address global warming.

McCright and Dunlap (2000; 2003) argue that a concerted effort on the part of an ideologically conservative countermovement to climate science is largely responsible for the troubling levels of skepticism among the American public regarding the integrity of climate science and the need for policies, such as carbon emission reductions, to combat global warming. McCright and Dunlap (2010) describe this conservative countermovement against climate science and policy as a network of conservative foundations, think-tanks, media outlets and public intellectuals that are heavily funded by conservative families and various corporations. The level of climate change denial varies within this network: in the extreme case, some continue to espouse the view that irregular warming is not occurring, while others concede the observed warming trend but deny any human cause. Some within the network agree that the warming is anthropogenic but strongly disagree with mitigation policies.

This movement is not a new phenomenon. Conservative activism against environmental policymaking emerged in parallel with the first steps toward environmental regulation during the late 1960s and early 1970s. The objective of these activists is tackling what many within the movement regard as anti-modernist scientific inquiries and policies—that is, science that seeks to evaluate the impacts and costs of modern industrialist production (Scnaiberg 1980, McCright and Dunlap 2010). The policy implications of this "impact science" is viewed as a threat to free-market capitalism by many conservative anti-environmentalists and, as such, it should be challenged in order to protect economic productivity and technological innovation.

Having gained significant political power with the arrival of Ronald Reagan to the White House, conservative activists pushed to weaken the authority of the Environmental Protection Agency and the Department of the Interior, albeit unsuccessfully (McCright and Dunlap 2010). Indeed, the push for curtailing environmental policies during the first Reagan administration was met with popular discontent (Gillroy and Shapiro 1986). Nonetheless, attempts to dismantle the environmental regulatory architecture, prevent substantive environmental protection policy, and reduce support for environmental research picked up during the Republican-led Congress of the mid 1990s, continued through the presidency of George W. Bush, and have persisted to the present. McCright and Dunlap (2010) review how, to achieve these ends, anti-environmentalist activist groups have relied on tactics such as:

1. Obfuscating and undermining consensus scientific findings by promoting contrarian viewpoints and selectively focusing on studies which deviate from the consensus view, while choosing to ignore or misrepresent mainstream climate science (see McCright and Dunlap 2000; 2003; 2010, Mooney 2005, see).

2. Manipulating, altering, and suppressing climate science reports produced by government agencies. For example, the Bush Administration effectively excluded the comprehensive Clinton-era report *National Assessment of Potential Consequences of Climate Variability and Change* from all official climate-related documents (see Mooney 2007).

3. Targeting climate scientists in academia and government agencies with intimidation and threats of sanctions (see McCright and Dunlap 2000; 2003, Oreskes and Conway 2010).

The tactic most relevant to the current study, however, is the concerted effort by the conservative countermovement to sway public opinion on climate science and policy through the media and public events. Regarding the use of media, the effort has largely been effective in generating what has come to be known as the "dueling scientists scenario" (McCright and Dunlap 2003), whereby, in an effort to uphold the journalistic norm of "objectivity," rigorous findings and unsubstantiated speculation are effectively equated and mixed together to produce a "confusing impression that scientists share no consensus of the probable magnitude, timing, and potential seriousness of the environmental and societal consequences of the documented and well-understood buildup of various greenhouse-enhancing gases in the atmosphere" (Schneider 1993). In their seminal study on the role of "balanced reporting" on climate change in the American "prestige" print press, Boykoff and Boykoff (2004) argue that journalistic norms such as objectivity, fairness, accuracy, and balance serve as a source of "informational bias" regarding coverage of global warming. Indeed, when it comes to science reporting, these journalistic norms act as [surrogates] for validity checks" since "the typical journalist, even one trained as a science writer, has neither the time nor the expertise to check the validity of claims herself" (Dunwoody and Peters 1992, Boykoff and Mansfield 2008). In effect, while providing a "balanced" view, many media outlets are really presenting a biased view of climate science by offering grossly disproportionate levels of attention to climate change contrarians. [2]

# 3 Learning about Climate Skepticism: An Unsupervised Approach

What does it mean to be a climate skeptic? What topics shape the conversation of the conservative counter-movement on global warming? These questions were explored in detail in McCright and Dunlap (2000)'s seminal study on climate skepticism and the conservative movement's counter-claims. The authors content analyze documents related to global warming for 14 major conservative think tanks over the period 1990-1997. After searching each organizations website, they gained access to 224 publications—the vast majority of which were produced during 1996 and 1997—and spent the summer of 1998 coding the documents. Overall, the content analysis suggests that climate skepticism during this period centered on three major counter-claims (see McCright and Dunlap 2000, pg. 510, Table 3):

1. *The evidentiary basis of global warming is weak or wrong.* Arguments falling within this counter claim tended to focus on the uncertainties associated with published scientific findings—i.e., that it is "junk" science—and expressed doubts whether there was indeed a "scientific consensus" when it came to global warming.

2. *Global warming would be beneficial if it was to occur.* The main arguments under this claim focused on the benefits in terms of weather, health, and

[2] The empirical record suggests that the American media do, in fact, allow for disproportionate levels of attention to contrarian viewpoints. Regarding print media, Boykoff and Boykoff (2004) estimate that for 1988-2002, about 53% of "prestige" newspaper coverage of global warming was "balanced"—that is, it provided "roughly equal" attention to the view that human activity is primarily responsible for global warming and also the opposing contrarian position that any warming is due to natural causes. In a similar vein, McCright and Dunlap (2003) found that, over the period 1994-1997, a handful of climate change contrarians were cited in the nation's most popular newspapers just as often as the leading climate scientists. In a comparative context, and perhaps unsurprisingly, studies have found that the United States "prestige" print press appears to be a global leader in hosting skeptical viewpoints about climate change. Dispensa and Brulle (2003), Brossard et al. (2004) show how, during the period of study, top American newspapers were much more likely to emphasize uncertainty in climate science findings relative to selected foreign newspapers. More recently, Painter and Ashe (2012) content analyze articles from major newspapers from the United States, United Kingdom, Brazil, China, France, and India for early 2007 and November 2009 - February 2010. The authors find that, relative to these other countries, American newspaper coverage is much more likely to voice uncontested skeptical views on climate change.

agriculture.

3. *Global warming policies would do more harm than good.* That is, the economic, security, and environmental costs are too high to justify action.

This paper builds upon the early work of McCright and Dunlap (2000) by employing a simple probabilistic model to "learn" the topics discussed by climate skeptics. What motivated our choice to rely on automated text analysis, as opposed to human coding of contrarian documents? One significant reason was reducing data collection cost. As discussed in Grimmer and Stewart (2013), human-coding methods can be very costly in terms of time and resources. As the size of the corpus intended to be coded increases, so does the cost, since each text needs to be read by a human coder. When working with very large corpora, as is the case in the current study, the cost can become prohibitive. A second motivation to utilize an automated text analytic method was that following McCright and Dunlap (2000), there has been no update to the systematic classification of contrarian arguments against climate science and policy. As will be discussed below, by utilizing an automated classification method, we are able to identify a range of new and unique topics that have sprouted among deniers since the late 1990s.

## 3.1 Building a Contrarian Corpus

To build a corpus of "contrarian texts," we scrapped the websites of 15 well-known conservative think tanks and organizations for information related to climate change. Our choice of organizations, to a large extent, mirrors that of McCright and Dunlap (2000). Table 1 displays the names of the organizations and think-tanks which we study along with the number of documents that are included in the corpus from each website. The total number of documents included in the corpus are 13,114. To retrieve the contrarian documents, we downloaded the HTML code of pages which were either classified by the respective organization as dealing with climate change or from within-site search results for the terms "climate change" or "global warming". Next, documents that contained the term "climate change" or "global warming" were retained from the larger set of retrieved documents. Documents in PDF format and audiovisual materials were a minority of the overall set of documents and were excluded in the current analysis.[3] Relevant text was extracted from the HTML source code using a set of regular expressions. The resulting text was then tokenized and filtered (i.e. stop words and punctuation were removed, tokens were stemmed using a the common Porter Stemmer).

[3]We are currently in the process of cleaning PDF reports. These documents (primarily "Policy Reports") will be added to the corpus shortly

| Organization Name | Number of Documents |
|---|---|
| American Enterprise Institute (AEI) | 642 |
| Cato Institute (CEI) | 301 |
| Competitive Enterprise Institute (CEI) | 937 |
| Fraser Institute | 63 |
| Global Warming Policy Foundation | 7,892 |
| Heartland Institute | 272 |
| Heritage Foundation | 220 |
| Hoover Institution | 24 |
| International Climate and Environmental Change Assessment Project (ICECAP) | 1,784 |
| George C. Marshall Institute | 139 |
| National Center for Policy Analysis (NCPA) | 43 |
| National Center for Public Policy Research (NCPPR) | 386 |
| Pacific Institute | 226 |
| Reason Foundation | 159 |
| **Total** | 13,114 |

**Table 1:** The number of documents currently in the corpus for 15 well-known conservative think tanks and organizations for information related to climate change.

## 3.2 Finding Topics

This section utilizes the well-known latent Dirichlet allocation (LDA) model originally proposed in Blei et al. (2003). LDA is a simple hierarchical Bayesian model which starts with the assumption that each word in a text is exchangeable, that a text in a corpus is a combination of a specific number of topics ($T_k$), and each specific topic is represented as a distribution of words ($w$) in a fixed vocabulary. The generative structure that produces each document in a corpus is represented as random mixtures of latent topics and their associated distributions of words. Specifically, the LDA assumes that documents are generated from the following probabilistic process:

1. Each of the $k$ topics are drawn from a topic distribution by

$$\theta \sim Dirichlet(\alpha)$$

2. The term distribution $\beta$ for each topic is represented by

$$\beta \sim Dirichlet(\eta)$$

3. For each of the $N$ words $w_n$:

   Randomly sample a topic $z_n \sim Multinomial(\theta)$.

   Choose a word $w_n$ from $p(w_n|z_n, \beta)$.

While this process provides a simple representation of the data generating process for corpus of texts, the model has been shown to preform well in a wide range of areas, from population biology to information retrieval(see Blei 2012, for an overview).

#### 3.2.1 Inferring Topic Structure

The next step is to infer the underlying structure (i.e., topics) from based on the model described above. We rely on the sparse Gibbs sampler described in Yao et al. (2009).[4] After a good deal of experimentation regarding model's hyperparameters, we found that the efficient hyperparameter optimization routine utilized in Wallach et al. (2009a) provided the most easily interpretable set of topics.[5]

LDA requires one to specify the number of topics *a priori*. This presents an obvious challenge, as researchers generally do not have strong prior information about the number of topics in a corpus. While a range of methods have been introduced in the literature to estimate the "optimal" number of topics based on the held-out likelihood (see Wallach et al. 2009b, for an overview), there remains considerable debate on the utility of data-driven approaches for generating interpretable topics. For instance, using a large number of human subjects, Chang et al. (2009) present evidence suggesting models which preform better in terms of held-out likelihood, may actually infer less meaningful topics. The results suggest the need to carefully examine the interpretability of the latent space when employing topic models and provide caution against blindly choosing the model that minimizes held-out likelihood.

In this study, we take balanced approach between "optimal" data-driven methods and a qualitative assessment of the interpretability of the latent space. First, in terms of data-driven methods, we rely on 10-fold cross-validation and examine changes in "perplexity" over a coarse grid of topic numbers (Blei et al. 2003).[6] Here, we looked primarily for major changes (i.e. reductions) in the estimated perplexity when moving across the coarse grid. This analysis suggests considerable changes in the estimated perplexity when moving from, say, 20 to 100 topics, but only minor gains thereafter. Moreover, our analysis generally conforms to the findings in Chang et al. (2009): simply minimizing the model's perplexity with respect to the number of topics—as is often suggested in the

[4]Our choice to rely on the sparse Gibbs sampler was driven primary by concerns over efficiency and computational convenience. Note that little changes when using other commonly employed sampling algorithms such as the variational expectation maximization approach discussed in Blei et al. (2003) or the (also quite efficient) collapsed Gibbs sampler discussed in Griffiths and Steyvers (2004).

[5]We optimize both the $\alpha$ and $\beta$ hyperparameter in the results provided below. It is important to note, however, that the topics described below display a good deal of stability across a wide range of specifications.

[6]The intuition associated with minimizing perplexity is that if the model is accurately inferring the structure of the corpus, it should also be able to accurately infer the structure of held-out data from the same corpus. Perplexity thus measures how confused (or "perplexed") the model is when seeing new data.

topic modeling literature (Wallach et al. 2009b)—led to a considerable number of topics that were quite difficult to interpret. Based on this assessment, we chose to estimate the LDA assuming a 100 topic model.

### 3.2.2 Topic Interpretation

Table 2 provides a list of the estimated topics, a descriptive label, and a list of the 5 most probable keywords for each topic. The validity of the assigned topic labels was assessed by small sample ($n = 10$) of articles assigned (with a high proportion) to each topic (Quinn et al. 2010). After removing 16 "junk" topics (AlSumait et al. 2009), the final list included 84 topics representing a range of issues related to global warming, from skepticism associated with climate science to energy policy.[7]

**Table 2:** Topics in the Climate Skepticism Corpus

| | Topic Name | Topic Family | Import-* ance | Keys |
|---|---|---|---|---|
| 1 | International Agreements | International Agreements | 0.036 | climat, countri, develop, copenhagen, nation |
| 2 | Climate Trends (Long-Term) | Climate Trends | 0.032 | warm, temperatur, global, climat, year |
| 3 | Energy Consumption | Economic Impacts | 0.030 | energi, bill, price, cost, per |
| 4 | Shale Gas (Hydraulic Fracturing) | Unconventional Energy | 0.026 | shale, frack, drill, water, well |
| 5 | Government Investment Renewable (UK) | Renewable Energy | 0.022 | govern, plan, new, invest, project |
| 6 | Shale Gas (Energy Independence) | Unconventional Energy | 0.022 | price, natur, shale, energi, product |
| 7 | Global Cooling | Climate Trends | 0.021 | winter, cold, snow, record, temperatur |
| 8 | Sea Level Rise | Climate Impacts | 0.019 | ice, sea, level, rise, arctic |
| 9 | General Policy Keywords | Environmental Policy | 0.019 | polici, more, problem, govern, need |
| 10 | Energy Policy (UK) | Energy Policy | 0.018 | energi, govern, minist, green, britain |
| 11 | Renewable Energy (General) | Renewable Energy | 0.017 | energi, renew, fuel, technolog, fossil |
| 12 | IPCC (Peer-Review) | Scientific Integrity | 0.017 | ipcc, report, review, panel, author |
| 13 | Solar Energy | Renewable Energy | 0.017 | solar, subsidi, energi, panel, industri |
| 14 | Climategate | Scientific Integrity | 0.017 | email, cru, climateg, jone, univers |
| 15 | Consensus Myth | Scientific Uncertainty | 0.016 | scienc, scientist, scientif, climat, consensu |
| 16 | Cap and Trade | Energy Policy | 0.015 | tax, carbon, trade, emiss, cap |
| 17 | US Debate of Kyoto | International Agreements | 0.014 | kyoto, treati, protocol, emiss, nation |
| 18 | Clean Air Act | Environmental Policy | 0.014 | epa, regul, air, act, greenhous |
| 19 | Wind (UK) | Renewable Energy | 0.013 | wind, turbin, farm, energi, plan |
| 20 | Anthropogenic Causes | Human Forces | 0.013 | global, warm, climat, scientist, caus |
| 21 | California AB32 | Energy Policy | 0.012 | california, cap, bill, state, energi |
| 22 | Obama | US Politics | 0.012 | obama, presid, democrat, republican, elect |
| 23 | Ice Age | Climate Trends | 0.011 | ice, period, year, temperatur, climat |
| 24 | Temperature Data | Skepticism (Data) | 0.010 | data, temperatur, station, record, adjust |
| 25 | Climate Trends (Short-Term) | Climate Trends | 0.010 | temperatur, year, global, warm, decad |
| 26 | Biofuel | Renewable Energy | 0.010 | food, ethanol, biofuel, crop, product |

*Continued on Next Page...*

[7] AlSumait et al. (2009) note that not all topics in an estimated topic model are of equal importance and it is not uncommon to have a set of "junk" topics that pick up common co-occurrences of words with little or no substantive meaning.

**Table 2:** (Continued)

| | | | | |
|---|---|---|---|---|
| 27 | Solar Forcing | Natural Forces | 0.010 | solar, sun, cycl, activ, earth |
| 28 | US Legislature | US Politics | 0.010 | senat, bill, hous, vote, energi |
| 29 | Wind | Renewable Energy | 0.010 | power, wind, electr, gener, renew |
| 30 | Climate Activism (Alarmists) | Scientific Integrity | 0.010 | global, warm, hansen, climat, nasa |
| 31 | Clouds | Natural Forces | 0.010 | atmospher, climat, effect, greenhous, cloud |
| 32 | Nuclear Power | Conventional Energy | 0.010 | nuclear, power, reactor, plant, japan |
| 33 | Government Funding Research | US Politics | 0.010 | fund, million, govern, billion, program |
| 34 | Aviation Emissions (Asian Retaliation to EU) | Environmental Policy | 0.010 | trade, airlin, european, china, aviat |
| 35 | Oil Production | Conventional Energy | 0.010 | oil, drill, barrel, product, pipelin |
| 36 | Carbon Reduction's Impact on Industry | Economic Impact | 0.010 | industri, cost, energi, manufactur, busi |
| 37 | Energy Policy (EU) | Energy Policy | 0.009 | european, europ, commiss, target, franc |
| 38 | Climate (General Discussion) | Buzz Words | 0.009 | climat, chang, impact, scienc, polici |
| 39 | Email Discussions | Scientific Integrity | 0.009 | public, inform, report, document, issu |
| 40 | Extreme Weather (General) | Extreme Weather | 0.009 | weather, extrem, event, climat, flood |
| 41 | Coal | Conventional Energy | 0.009 | coal, power, plant, electr, fire |
| 42 | Public Opinion | Society | 0.009 | percent, poll, survey, american, warm |
| 43 | Wind (Hurricane) | Extreme Weather | 0.009 | hurrican, storm, tornado, global, tropic |
| 44 | JUNK | JUNK | 0.008 | offic, global, warm, polici, prefix |
| 45 | Australian Carbon Tax | Energy Policy | 0.008 | australia, australian, govern, carbon, canada |
| 46 | It's Cold Outside | Weather | 0.008 | offic, met, forecast, winter, weather |
| 47 | Hockey Stick | Scientific Integrity | 0.008 | mann, hockey, stick, data, univers |
| 48 | UK Coverage of Climate | Society | 0.008 | bbc, sceptic, societi, climat, royal |
| 49 | Cost-Benefits of Emissions Reduction | Economic Impact | 0.008 | cost, percent, per, billion, econom |
| 50 | German Renewable Policy | Renewable Energy | 0.007 | germani, german, energi, merkel, green |
| 51 | Shale Gas (Mediterranean) | Unconventional Energy | 0.007 | shale, reserv, resourc, energi, cubic |
| 52 | Carbon Emissions (General) | Energy Policy | 0.006 | carbon, emiss, dioxid, greenhous, reduc |
| 53 | Coral Reefs | Climate Impacts | 0.006 | ocean, coral, more, reef, read |
| 54 | Electric Cars | Energy Policy | 0.006 | car, electr, vehicl, batteri, hybrid |
| 55 | Environmental Activists | Society | 0.006 | environment, group, green, campaign, environmentalist |
| 56 | Security (Asia) | International Relations | 0.006 | china, state, secur, unit, nation |
| 57 | Climate Models | Skepticism (Models) | 0.006 | model, climat, predict, forecast, comput |
| 58 | Ocean Temperature | Climate Impacts | 0.006 | ocean, temperatur, surfac, heat, climat |
| 59 | Ideological Debates | Society | 0.006 | book, polit, social, peopl, liber |
| 60 | Economic Growth | Development | 0.006 | econom, technolog, growth, develop, market |
| 61 | Carbon Taxes | Energy Policy | 0.006 | govern, peopl, american, tax, more |
| 62 | Pollution (Air) | Environmental Policy | 0.006 | environment, pollut, air, environ, qualiti |
| 63 | Green Jobs | US Politics | 0.006 | job, green, creat, economi, econom |
| 64 | Peer-Review (General) | Scientific Integrity | 0.005 | paper, research, studi, publish, review |
| 65 | Government Investment Green Tech | Renewable Energy | 0.005 | compani, busi, bank, invest, corpor |
| 66 | International Health | Health | 0.005 | health, malaria, diseas, death, risk |
| 67 | Polar Bears | Climate Impacts | 0.005 | bear, polar, arctic, popul, ice |
| 68 | Japan Methane Hydrate | Unconventional Energy | 0.005 | methan, hydrat, earth, water, japan |
| 69 | Fuel Standards | Energy Policy | 0.005 | fuel, car, vehicl, standard, effici |
| 70 | Melting Glaciers | Climate Impacts | 0.004 | glacier, melt, himalayan, india, indian |
| 71 | Religion and Environment | Society | 0.004 | human, earth, world, natur, environment |
| 72 | Al Gore | Al Gore | 0.004 | gore, truth, film, inconveni, nobel |
| 73 | Deforestation | Human Forces | 0.004 | forest, speci, tree, extinct, wood |
| 74 | Adaptation | Adaptation | 0.004 | ghg, polici, climat, control, chang |

*Continued on Next Page...*

**Table 2:** (Continued)

| | | | | |
|---|---|---|---|---|
| 75 | Less Developed Countries | Development | 0.004 | world, countri, develop, global, africa |
| 76 | Heartland Institute Incident | Heartland Institute Incident | 0.004 | fraud, document, investig, gleick, polic |
| 77 | Water (Drought/Flood) | Extreme Weather | 0.004 | water, drought, flood, precipit, rainfal |
| 78 | Environmental Law | Law | 0.003 | court, law, state, case, legal |
| 79 | Energy Security (Europe) | Energy Security | 0.003 | russia, russian, poland, gazprom, europ |
| 80 | Volcanic Activity | Natural Forces | 0.002 | ozon, earth, erupt, volcan, year |
| 81 | Disaster Insurance | Extreme Weather | 0.002 | disast, insur, risk, flood, peopl |
| 82 | White House | US Politics | 0.002 | administr, bush, presid, white, hous |
| 83 | IPCC (Uncertainty) | Scientific Uncertainty | 0.001 | scenario, uncertainti, ipcc, sensit, estim |
| 84 | Development Goals | Development | 0.001 | popul, lomborg, year, age, ehrlich |

*Caption*: This table provides the results of the 84 topic LDA, as descried above *We measure "importance" as the proportion of documents in which the topic has the highest posterior probability. We recognize that , clearly, this is an extremely rough measure of importance.

When looking across Table 2, we see many of the usual suspects discussed in previous studies but also some new topics. The topics produced by the model are generally consistent with those presented in McCright and Dunlap (2000), with a number of topics speaking directly to McCright and Dunlap (2000)'s "Counter-Claim One" (on the evidentiary basis for climate change) and "Counter-Claim Three" (questioning the utility of climate policies). Table 3 provides an overview of the specific themes outlined in McCright and Dunlap (2000, , pg. 510, Table 3), along with the topics most closely aligned with a given theme.

Skeptical voices related to the validity of mainstream climate science are common within the corpus. The notion that human activity, specifically the emission of greenhouse gases into the atmosphere, is leading to a rise in global temperatures (topic 20) has been labelled as the "man-made global warming scare" that is being pushed by "scare-mongers".[8] Appeals to long-term natural cycles in temperature (topic 2) are common support for arguments against anthropogenic global warming, such as the Roman and Medieval Warm Periods. Documents also focus on alternate climate forcing inputs such as solar (topic 27), clouds (topic 31), and volcanoes (topic 80) as more plausible explanatory factors for climate fluctuations than greenhouse gas emissions. The very existence of a true scientific consensus on anthropogenic warming (topic 15) continues to be denied and has been called "manufactured"[9] and "premature".[10] The predictive power of climate change models (topic 57) that are referenced in the IPCC assessments is often questioned. The validity and reliability of empirical data used in climate change studies (topic 24) to demonstrate global warming impacts are also cast into doubt. Further, the integrity of climate scientists is also frequently questioned, especially in relation to the peer-review process of the IPCC (topic 12) and the 2009 "climategate" email leakage incident (topic 14), Impacts of climate change on animals such as polar bears (topic 67) and coral (topic 53)

[8] http://icecap.us/index.php/go/new-and-cool/does_it_really_take_much_imagination_to_project_what_the_scare_mongers_will/

[9] http://judithcurry.com/2012/10/28/climate-change-no-consensus-on-consensus/

[10] http://www.hoover.org/publications/defining-ideas/article/138101

are routinely downplayed; as are environmental impacts such as sea level rise (topic 8) and melting glaciers (topic 70).

The results of the LDA model also demonstrate the breadth of topics discussed in documents referencing climate change with important issue linkages across both the domestic and international political economy. For instance, much critical discussion surrounds international mitigation policies (topics 1 and 17), which are typically refuted based upon expected detrimental economic impacts such as rising energy prices (topic 3) and reduced economic growth (topic 60). Renewable energy technologies such as solar (topic 13), wind (topic 19), and biofuels (topic 26) are almost always presented as wasteful and counter-productive. Unconventional sources of energy such as shale gas (topics 4, 6, and 51) and methane hydrates (topic 68), on the other hand, are discussed in positive terms, typically in relation to energy independence and technological innovation.

Climate change deniers also do not seem to hesitate when singling out specific individuals as either disingenuous or misguided. Al Gore (topic 72) is a favorite target, especially in relation to his environmental activism and campaigning. Much of the focus has been on Gore's documentary film *An Inconvenient Truth*, which has been characterized as, "a colorfully illustrated lawyer's brief for global warming alarmism and energy rationing"[11] and that screenings of the film in schools amounts to "indoctrination". Gore has also been labeled a hypocrite for, among other things, the "prodigious personal use of electricity at his Nashville mansion"[12] and has often been denigrated for accepting the Nobel Peace Prize for his environmental activism. Individual scientists have also been targeted. For example, Michael Mann's infamous "hockey stick graph" (topic 47), which depicted rapid growth of post-Industrial Revolution global temperature, drew heavy fire from climate change skeptics, and his research agenda has been branded as "the biggest taxpayer-financed gravy train for science and academia in decades."[13]

#### 3.2.3 Meta Topics and Topic "Families"

Even a cursory glance at Table 2 demonstrates a number of common themes across topics. As such, in addition to assigning each topic a more detailed label, we also coded each topic into a more general "topic family." These decisions were again guided by reading a small sample of articles. Based on this coding, the contrarian corpus is dominated by topics related to skepticism over climate science, discussions of energy policy (conventional, unconventional, and renewable), and domestic and international policy concerns (including specific climate agreements and domestic climate policies). As a further check on whether topics with similar labels are semantically similar, Figure 1 presents the results from a hierarchical agglomerative clustering (HAC) of the 84 topics (Quinn et al. 2010).

[11] http://cei.org/studies-other-studies/al-gores-science-fiction-skeptics-guide-inconvenient-truth

[12] http://www.foxnews.com/story/2008/06/19/junk-science-al-gore-epic-hypocrisy/

[13] http://icecap.us/index.php/go/political-climate/hockey_stick_creator_michael_mann_seeks_courts_help_to_ensure_no_inquiry_no/

| **Description** | **Topics** |
|---|---|
| *The evidentiary basis for global warming is weak or wrong.* | |
| 1. The scientific evidence for global warming is highly uncertain | 20, 24, 57 |
| 2. Mainstream climate research is "junk" science | 46, 47, 83 |
| 3. The IPCC intentionally altered its reports to create a "scientific" consensus" on global warming | 12, 15 |
| 4. Global warming is merely a myth or a scare tactic. | 42, 55 |
| 5. Global warming is merely a political tool. | 22, 72 |
| *Global warming policies would do more harm than good.* | |
| 1. Proposed policies would harm the economy. | 3, 36, 49, 69 |
| 2. Proposed action would weaken national security. | 56, 6 |
| 3. Proposed action would threaten sovereignty | 17 |
| 4. Proposed action would actually harm the environment | – |

**Table 3:** Counterclaim 1 and 3 from McCright and Dunlap (2000), as well as examples of topics in Table 2 that are consistent with each theme. We found few instances of McCright and Dunlap (2000)'s Counterclaim 2 (i.e. that a warmer climate would actually be beneficial.)

As shown in the figure, the analysis suggests a number of distinct themes. For instance, topics related to energy and policy, as climate trends and the consequences climate change tend to cluster together in the figure.

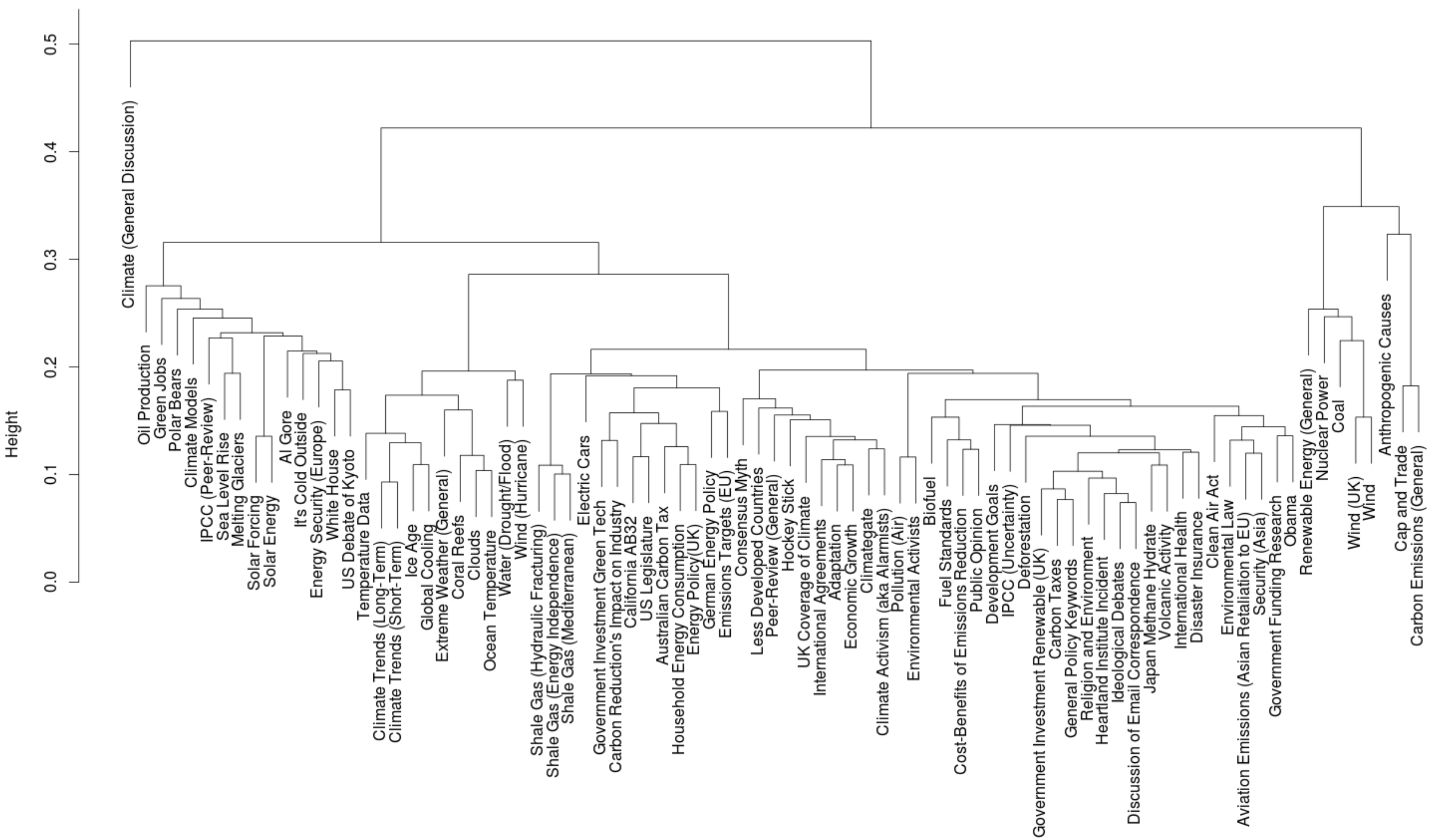


**Figure 1:** HAC results

# 4 Model Validation (Preliminary)

While the topics produced in Table 2 are largely consistent with the literature's understanding of climate skepticism, Grimmer and Stewart (2013) make a forceful case for the need to carefully validate text-analytic models. This section provides a first step toward achieving that objective.

## 4.1 Accuracy in a Small Random Sample of Contrarian Documents

Our first initial check on validity relies on a content analysis of a small ($n = 100$), random sample of documents from the contrarian corpus. The goal was to determine whether the topics that the model assigned for a document were similar to those that would be assigned by a human coder. To achieve this objective, we had a research assistant (blind to the study's objectives) code the "dominant" topic for each of the 100 documents. While the 84 topics in Table 2 provide a granular look of the issues discussed in the contrarian corpus, the coder was instructed to code at the more abstract "topic family" level. This decision was made for both practical and substantive reasons. Substantively, the literature on the contrarian counter movement and hypotheses associated with the dynamics of skepticism overtime tend to be specified at this more general level of analysis. Thus, our concern lies more in whether the model correctly classifies documents into these broader themes.

Table 4 provides descriptive statistics for the results. As demonstrated in the table, the accuracy score varies considerably across the topic families. For instance, while the model only accurately classified roughly half of articles for "Energy Policy," it did appreciably better for more specific aspects of energy such as "Renewable Energy" (6/6) and "Unconventional Energy" (7/7). While these results should obviously be interpreted with caution, they do at least suggest that data generated from the LDA may provide insight into important facets of the energy debate.

The results in Table 4 also suggest that the places were the model had difficulties. In particular, the model had difficulties sorting out the differences between "Scientific Uncertainty" and various aspects of skepticism over climate science (i.e., "Skepticism (Data)" and "Skepticism (Models)"). These topics are highly related (both substantively and semantically) and thus it not all that surprising that these areas posed challenges for the LDA.

## 4.2 Predictive Validity of Topic Families

As a next step in determining measurement validity, we rely on inferring the predictive validity of a given topic family (Quinn et al. 2010). That is, we are interested in determining to what extent a given topic family is sensitive to external events. We would expect that the proportion of words from a given topic

| **Topic Family** | **N (Documents)** | Accuracy |
|---|---|---|
| Al Gore | 4 | 0.500 |
| Climate Change Impacts | 3 | 0.667 |
| Climate Trends | 2 | 0.500 |
| Conventional Energy | 2 | 0.500 |
| Development | 3 | 0.333 |
| Economic Impact | 6 | 0.833 |
| Energy Policy | 24 | 0.500 |
| Extreme Weather | 2 | 1.000 |
| Human Forces | 2 | 0.000 |
| International Agreements | 4 | 1.000 |
| International Relations | 1 | 0.000 |
| Natural Forces | 3 | 1.000 |
| Other | 3 | 0.000 |
| Renewable Energy | 6 | 1.000 |
| Scientific Integrity | 6 | 0.833 |
| Scientific Uncertainty | 4 | 0.000 |
| Skepticism (Data) | 2 | 0.500 |
| Skepticism (Models) | 4 | 0.000 |
| Society | 7 | 0.714 |
| Unconventional Energy | 7 | 1.000 |
| US politics | 3 | 0.667 |
| Weather | 1 | 0.000 |

**Table 4:** Accuracy in a small random sample of documents.

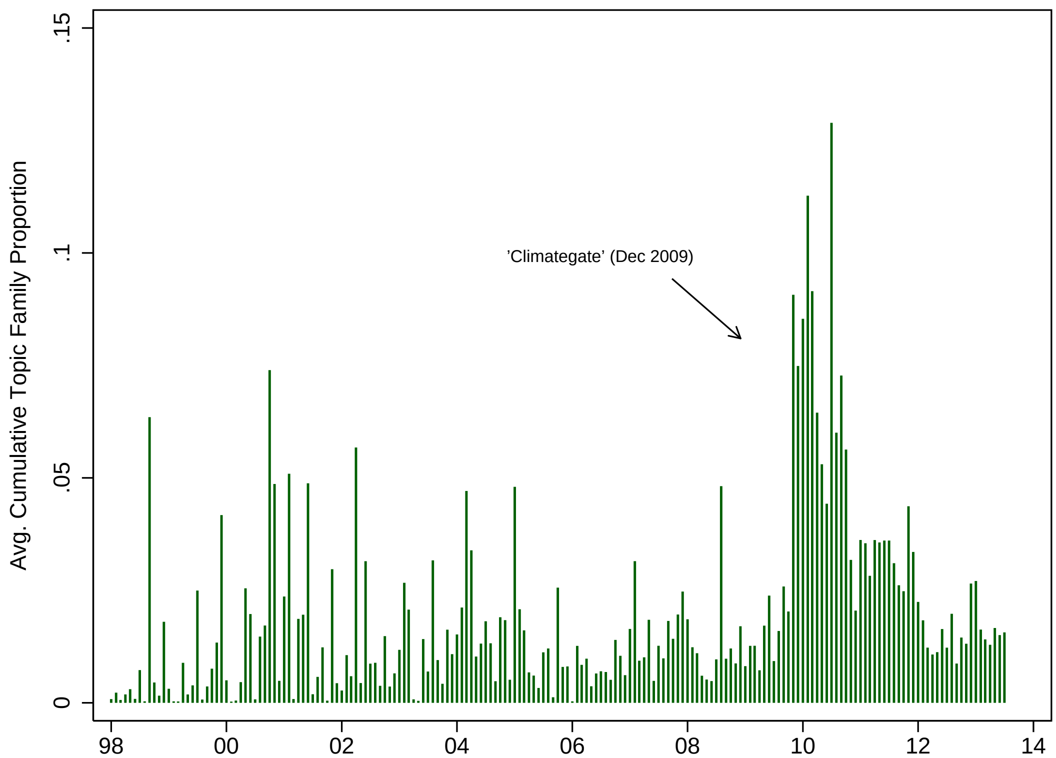


**Figure 2:** "Scientific Integrity" topic family monthly average proportions over 1998-2013.

should be more likely to be observed in the corpus when an external event related to that topic occurs. Figure 2 illustrates, for a given month, the expected proportion of words from a randomly chosen document that the model associates with the "scientific integrity" topic family. As we can see, the expected proportion ramps up in late 2009 and peaks in February 2010, when, on average, approximately 11 percent of the words in a randomly selected document are related to the "scientific integrity" topic family. This is expected since this period coincides with the so-called "Climategate" scandal where emails of researchers from the Climate Research Unit at the University of East Anglia were hacked, uploaded to the Internet, and subsequently scrutinized by climate skeptics.

As another example, Figure 3 plots the average monthly proportion of words associated with the "Al Gore" topic from a randomly selected document. Large spikes in expected topic proportions are observed in late 1998, while Gore was still Vice President and actively campaigned for international climate policy. Further, spikes are observed following his presidential campaign and in May

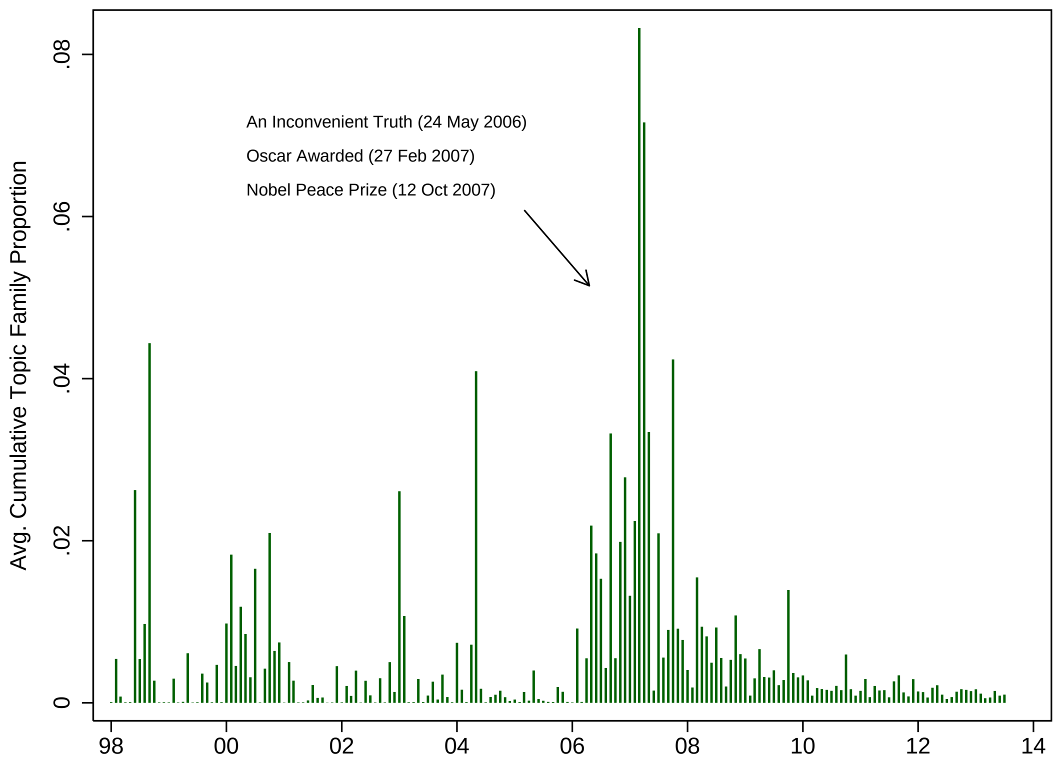


**Figure 3:** "Al Gore" topic monthly average proportions over 1998-2013.

2004 when he gave a scathing speech criticizing George W. Bush. It is only during 2006 and after that we observe sustained attention to Al Gore within the contrarian corpus. Specifically, the maximum value of the series occurs in March 2007 which is around the period when *An Inconvenient Truth* received the Academy Awards for Best Documentary Feature and Best Original Song. A substantial increase in the expected topic proportion is also observed in October 2007 when Al Gore and the scientists of the IPCC were awarded the Nobel Peace Prize. What is also interesting to note is the sustained attention given to Al Gore following early 2006, whereas, in contrast, there were relatively long periods prior to this point that the expected proportion for the "Al Gore" topic remained marginal.

## 5 Conclusion

While the debate on the determinants of public opinion on climate change continues, most would agree that climate skepticism, as expressed by conservative think-tanks and organizations, has at the least some discernible impact on public perceptions of climate change science and policy. Questions, however, do remain as to how large of an impact climate science skeptic groups have had on public opinion and what type of contrarian argument has had the greatest impact. A significant limitation in the literature which has, to a certain extent, hampered our ability to answer these difficult questions is a lack of a valid time series of climate change skepticism which can be used for hypothesis testing.[14] By employing automated text analytic approached to the study of contrarianism, we are able to generate data series which can help us study fundamental questions, such as whether skeptics have shifted away from anti-science to more anti-policy positions, and, if so, when and why this shift occurred.

Another advantage of using text-mining methods to the study of climate change skepticism is that we are better able to update our knowledge of contrarian arguments due to the significant reduction of required time and resources needed for coding. McCright and Dunlap (2010, p. 114), for example, recognize that there are certain limits to our understanding of current contrarian discourse due to a lack of updated systematic knowledge. We believe that text-mining offers the opportunity to maintain a more current picture of the discourse across multiple platforms (i.e., social media, blogs, broadcast and print media) Further, we are able to significantly increase the number of documents employed in the learning process from a few hundred as done in previous studies to thousands. Again, using OCR and other text processing methods, text-mining methods can be applied to all kinds of contrarian publications, such as books, and are not limited to Internet materials.

In future work, we plan to apply the results from the current study as well as supervised and unsupervised classification methods to study contrarianism in US print and television outlets; with the overarching objective being to study how topics have evolved over time, platforms and outlets.

## References

AlSumait, L., Barbara, D., Gentle, J., and Domeniconi, C. (2009). Topic significance ranking of lda generative models. In *ECML*.

Anderegg, W., Prall, J., Harold, J., and Schneider, S. (2010). Expert credibility in climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 107:12107–9.

[14]For example, Scruggs and Benegal (2012) operationalize contrarianism as frequency of *New York Times* articles with skeptical references. We believe that, while this measure approximates contrarianism, a more valid measure is possible.

Blei, D. M. (2012). Probablistic topic models. *Communications of the ACM*, 55(4):77–84.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Borick, C. and Rabe, B. (2010). A reason to believe: Examining the factors that determine individual views on global warming. *Social Science Quarterly*, 91(3):777–800.

Boykoff, M. and Boykoff, J. (2004). Balance as bias: global warming and the us prestige press. *Global Environmental Change*, 14:125–136.

Boykoff, M. and Mansfield, M. (2008). 'ye olde hot aire': Reporting on human contributions to climate change in the uk tabloid press. *Environmental Research Letters*, 3:1–8.

Brody, S., Vedlitz, S. Z. Z., and Grover, H. (2008). Examining the relationship between physical viulnerability and public pperception of global climate change in the united states. *Environment and Behavior*, 40:72–95.

Brossard, D., Shanahan, J., and McComas, K. (2004). Are issue-cycles culturally constructed. *Mass Communication & Society*, 7:359–77.

Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., and Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *NIPS*.

Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs, P., and Skuce, A. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, 8:024024.

Dispensa, J. and Brulle, R. (2003). Media's social construction of environmental issues. *International Journal of Sociology and Social Policy*, 24(10):74–105.

Doran, P. and Zimmerman, M. (2009). Examining the scientific consensus on climate change. *Eos, Transactions American Geophysical Union*, 90(3):22–23.

Dunwoody, S. and Peters, H. (1992). Mass media coverage of technologcal and environmental risks. *Public Understanding of Science*, 1:199–230.

Egan, P., . M. M. (2012). Turning personal experience into political attitudes: The effect of local weather on americans' perceptions about global warming. *The Journal of Politics*, 74(3):796–809.

Gallup News Service (2012). Gallup poll social series: Environment. Technical report.

Gillroy, J. and Shapiro, R. (1986). The polls: environmental protection. *Public Opinion Quarterly*, 50(2):270–279.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101:5228–5235.

Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):267–297.

Guber, D. L. (2003). *Grassoots of the Green Revolution: Polling America on the Environment*. Cambridge: MIT Press.

Guber, D. L. (2013). A cooling climate for change? party polarization and the politics of global warming. *American Behavioral Scientist*, 57:93–115.

Immerwahr, J. (1999). *Waiting for a signal: Public attitudes toward global warming, the environment and geophysical research*. American Geophysical Union.

Kellstedt, P., Zahran, S., and Vedlitz, A. (2008). Personal efficacy, the information environment, and attitudes toward global warming and climate change in the united states. *Risk Analysis*, 28(1):113–126.

Krosnick, J., Holbrook, A., Lowe, L., and Visser, P. (2006). The origins and consequences of democratic citizens' policy agendas: A study of popular concern about global warming. *Climatic change*, 77(1-2):7–43.

Li, Y., Johnson, E., and Zaval, L. (2011). Local warming: daily temperature change influences belief in global warming. *Psychological Science*, 22(4):454–459.

McCright, A. and Dunlap, R. (2000). Challenging global warming as a social problem: An analysis of the conservative movement's counter-claims. *Social Problems*, 47(4):499–522.

McCright, A. and Dunlap, R. (2003). Defeating kyoto: The conservative movement's impact on u.s. climate change policy. *Social Problems*, 50(3):348–373.

McCright, A. and Dunlap, R. (2010). Anti-reflexivity: The american conservative movement's success in undermining climate science and policy. *Theory, Culture & Society*, 27(3):100–133.

Mooney, C. (2005). *The Republican War on Science*. New York: Basic Books.

Mooney, C. (2007). An inconvenient assessment. *Bulletin of the Atomic Scientists*, 63(6):40–7.

National Research Council (2010). *Advanicing the Science of Climate Change*. National Academies Press.

Oreskes, N. and Conway, E. (2010). *Merchants of Doubt; How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming.* New York: Bloomsbury.

Painter, J. and Ashe, T. (2012). Cross-national comparison of the presence of climate skepticism in the print media in six countries, 2007-10. *Environmental Research Letters*, 7:1–8.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., and Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1).

Ray, J. and Pugliese, A. (2011). Worldwide, blame for climate change falls on humans: Americans among least likely to attribute to human causes.

Schneider, S. (1993). Degrees of certainty. *Research and Exploration*, 9(2):173–190.

Scnaiberg, A. (1980). *The Environment.* New York, NY: Oxford University Press.

Scruggs, L. and Benegal, S. (2012). Declining public concern about climate change: Can we blame the great recession. *Global Environmental Change.*

Smith, T., Marsden, T., Hout, M., and Kim, J. (2011). General social surveys, 1972-2010.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., editors (2007). *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.

Sterman, J. and Sweeney, L. (2002). Cloudy skies: assessing public understanding of global warming. *System Dynamics Review*, 18(2):207–240.

Wallach, H. M., Mimno, D., and Mccallum, A. (2009a). Rethinking lda: Why priors matter. NIPS.

Wallach, H. M., Murray, I., Salakhutdinov, R., and Mimno, D. (2009b). Evaluation methods for topic models. In *ICML.*

Yao, L., Mimno, D., and McCallum, A. (2009). Streaming inference for latent dirichlet allocation. kdd, 2009. *KDD.*