January 2014

Linking Local Weather To Climate Change: One Year Of Twitter In The US

Tatiana Molodtsova

LINKING LOCAL WEATHER TO CLIMATE CHANGE: ONE YEAR OF TWITTER IN THE US

A Thesis

Submitted to the Faculty

Of the

University of North Dakota

by

Tatiana Molodtsova

for the Degree

of

Master of Science

Grand Forks, North Dakota
May
2014

This thesis, submitted by Tatiana Molodtsova in partial fulfillment of the requirements for the Degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Dr. Andrei Kirilenko

Dr. Svetlana Stepchenkova

Dr. Rebecca Romsdahl

Dr. Xiaodong Zhang

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Dr. Wayne Swisher
Dean of the School of Graduate Studies

Date

PERMISSION

Title	Linking local weather to climate change: one year of Twitter in the US
Department	Earth System Science and Policy
Degree	Master of science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Tatiana Molodtsova
Date 5/8/2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Andrei Kirilenko for the continuous support of my research, great patience and encouragement to further my academic career. Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Svetlana Stepchenkova, Dr. Rebecca Romsdahl, Dr. Xiaodong Zhang for review and guidance for the work. I will forever be thankful to my former research advisor, Dr. Nikolay Dronin from Moscow State University, for introducing me to the world of scientific research and changing my life forever. I would like to thank the Department of Earth System Science and Policy (ESSP) at the University of North Dakota (UND), for providing the opportunity to acquire the essential skills for the research and resources and facilities throughout working towards my Master's degree. I thank my colleagues in ESSP for help and advice, and the UND professors who shared their knowledge and expertise during the classes I took. Finally I wouldn't be able to accomplish my work without the support of my loving family and friends.

ABSTRACT

There is a high level of scientific consensus on climate change. Nevertheless for climate change research to have any practical value, to develop public support for climate policies, the climate research results must find the way to general public. That is why it is important to understand how the public perception of climate change forms.

During the last decades there have been a number of studies on the factors affecting the level of public concern on climate change. Two major groups of factors are hypothesized to have the biggest influence on the level of public concern on climate change: extreme weather events and the mass media topic coverage.

Local studies confirm that the weather events experienced by people in certain locations might be related to climate change. In 1998 James Hansen hypothesized that two weather parameters' variations, namely, temperature and precipitation, exceeding one standard deviation should be noticeable by people and result in increase of the level of public concern on the phenomena. Nevertheless no previous studies were able to test this hypothesis and demonstrate that people truly use the information about local weather to make assumptions about climate change. The other studies on public perception of climate change are generally based on the agenda-setting theory, stating that the level of public concern on the issue is a reflection of the extent and prominence of media coverage of the topic.

The previous studies on how public perception of climate change forms are mainly based surveys, which is an active approach to collect social data. With the development of social media, however, a passive surveying of public perceptions on climate change has become possible. In this thesis the change in climate change microblogging intensity in Twitter was used as a proxy of change in the level of concern on the issue.

The objectives of the study were to utilize the Twitter, a currently the most popular microblogging platform, as a source of public salience data to test if the changes in weather parameters and in media coverage result in changes of the level of public concern on climate change. For this purpose the multiple linear regression and multi-model inference statistical techniques were used on three geographical levels of data aggregation.

The results clearly show that changes in weather parameters have significant effect on the level of public concern on climate change on the national, regional and local scales. The mass media topic coverage was also positively associated with the level of public concern on the national level. The study demonstrated that the social media data provides unprecedented opportunities for public opinion research.

CHAPTER I

INTRODUCTION

Weather is the condition of the atmosphere at any particular time and place. Climate is usually defined as the "average weather" over a period of time. In scientific literature climate encompasses the statistics of meteorological measurements in a given region over 30 years. There are numerous climatic datasets and scientific publications, evidencing the climate change. For example Global Historical Climatology Network (Smith & Reynolds, 2005) and National Aeronautics and Space Administration’s (NASA) Goddard Institute for Space Studies (GISS; Hansen, 2001) datasets of the land-surface air temperature, balloon-borne (Karl et al., 1996) and satellite measurements (Hadley Centre Radiosonde Temperature Data Set; Parker, 1997) show similar warming rates and are consistent within their respective uncertainties. Moreover, constantly developing data acquisition and analysis methods, for example Radio Occultation, delivering high quality observations of the atmosphere (Kursinski et al., 1997), and collaborative scientific research facilitate the progress in understanding how climate is changing. Increases in global average air and ocean temperatures, widespread melting of snow and ice, and rising global average sea level are the terms, in which a climate scientist would describe the climate change (Intergovernmental Panel on Climate Change (IPCC), 2007), however, these phenomena are not quite observable and interpreted correctly by public. For example, in the United States, when talk-show hosts and television reporters asked people on the street what they think about climate change, a typical response was that a few degrees warmer might not be so bad (Corbett and Durfee, 2004). It is not surprising that researches show a huge gap between scientific consensus and public understanding of climate change (Newport, 2010; Weber, 2011). Moreover, climate change has become a political issue and a “hot” topic in mass media that only adds the complexity to forming the public opinion (Kirilenko and Stepchenkova, 2012; McCright, and Dunlap, 2011; Weingart et. al, 2000).

For climate change research to have any practical value, the results must find the way to general public. Scientists should establish effective communication and operate in scientific terms, yet understandable by people, to develop public support for climate policies. As it is common for people to perceive the latest climate fluctuation as long-term climate change (Hansen et al., 1998), several attempts to design an objectively measured climate indicator, which can be felt by people living in a certain territory were made. One of them is a “common-sense climate index” (CSCI), proposed by Hansen et al. (1998) and intended to be a simple measure of the degree to which climate change is occurring in one particular area, that will be observable by people, thus helping them to understand the climate variability. The index is based on easily observable weather parameters such as temperature and precipitation; the main hypothesis was that the change in climate becomes noticeable by the public when the change in these parameters is consistently observed and large enough. These parameters change would be interpreted by public as “abnormal weather conditions” and associated with climate change (Hansen et al., 1998). This hypothesis, however, has never been tested.

One of the ways to learn about the weather events that the public truly associates with climate change is to conduct a survey, which is an active approach to collect social data. During the last decades there have been a number of studies on public perception of climate change based on the public opinion poll data (Howe et al., 2012; Chambliss et al., 2012). However, the active approach to data collection has its drawbacks, for example it requires effort and engagement by both a surveyor and a respondent. Moreover, these data are usually based on manual counts, and therefore it is labor intensive. Finally, active approach is characterized by the limited scoping, because some groups of population are difficult to reach to.

With the development of social media, however, a passive surveying of public perceptions on climate change has become possible. The Internet and modern technology allow for real-time, continuous monitoring of public opinion on various topics. The social media, including social networks (e.g. http://facebook.com) and blogging platforms (e.g. http://livejournal.com) have broad, diverse audience, represented by users from many countries. There are also the drawbacks of passive data collection, e.g. the data always need to be manually filtered. If the data are spatially distributed, a geolocation resolving algorithm has to be developed. Finally there are privacy issues, associated with the personal information use (Tavani, 1999). In this thesis Twitter was used as a source of data on public concern on climate change.

1.1 Twitter

Twitter is currently the most popular microblogging platform. In December 2012, Twitter announced it had surpassed 200 million monthly active users from all over the world (Fiegerman, 2012). According to Smith and Brenner (2012), about eight percent of the Americans use Twitter on a typical day. Twitter also has broad geography. According to Kulshrestha et al.(2012), Twitter is most popular in the US, Europe and Asia (mainly Japan); Tokyo, New York and San Francisco are the major cities where user adoption of Twitter is high.

The scholars interested in monitoring of natural and social phenomena have adopted the new concept of viewing Twitter users as a large network of sensors that react to external events by tweeting (Howe at al., 2012; Wang et al., 2012). This approach seems specifically valuable for studying the social processes in the developing world, as social media platforms have become a forum for giving a voice to the masses in those countries.

1.2 Objectives

In our study I modify Hansen's concept of an objectively measured subjective climate change indicator, which can relate public feeling that the climate is changing to the observed meteorological parameters. Clearly, the yearly index, consisting of many weather parameters, is too rough to sense the connection between weather anomalies and climate change perception. Therefore our analysis was done on a weekly basis, taking all the potentially important influencing factors, which were suggested by climate change public perception studies, as independent variables. The base period for computing the anomalies was changed from 1951-1980 to 1971-2000 and time lag component was included in the analysis.

The specific objectives of this thesis are as follows:

- Develop a linear regression model of Twitter microblogging activity (dependent variable) using weather and media indicators (independent variables).
- Test the model at three levels of aggregation: national, regional, and local

For the purposes of the study, the entire 2012 population of Twitter microblogging activity on climate change topic was collected, accumulating over 1.8 million separate records (tweets) globally. The geographic location of the tweets was identified and associated daily and weekly intensity of tweeting with the following parameters of weather for these locations: temperature anomalies, "hot" temperature anomalies, "cold" temperature anomalies, precipitation anomalies, rain and snow events. To account for the mass media influence the articles on climate change from the "prestige press" (Stovall and Solomon, 1984) were included, which comprises the newspapers considered to be the most influential (Boykoff and Boykoff, 2004).

The main goal of the thesis is to examine if change in objective weather parameters affected subjective climate change public discourse in Twitter in 2012 in the United States.

CHAPTER II

LITERATURE REVIEW

2.1 Weather and climate perception

Previous studies show that the general public has difficulty distinguishing between weather and climate (Bostrom et al., 1994; Read et al., 1994; Palutikof et al., 2004; Weber, 2010). It means that often people are using the information about *local weather* to make assumptions about *global climate.* According to Read et al (1994), a failure to recognize that climate is a statistical concept with low correlation with local weather events may contribute to weather-related fluctuations in public concern about global warming". For example, it is a well-known fact that several "hot" summers during the 80s greatly intensified public fears about climate change (Read et al., 1994).

Moreover, the connection between the personal experience (weather) and the perception of climate change, has its "twist": it is not rare that people already have some preconceived beliefs about climate change and tend to use short-term weather phenomena to support them. Expectations of climate change (or stability) play a significant role in people's ability to detect climate trends in the area where they live. In 1982 Kupperman confirmed this assertion by a study of one historic example (Kupperman, 1982): English settlers who arrived in North America in the early colonial period assumed that climate was a function of latitude. Newfoundland, located south of London, was expected to have a moderate climate. Despite repeated cold temperatures, which resulted in deaths and crop failures, colonists stayed loyal to their expectations and generated complex explanations for these deviations. In another study by Weber (1997), farmers in Illinois were asked to recall salient temperature or precipitation statistics during the growing season of seven preceding years. The surveys showed that those farmers who believed that their region was undergoing climate change recalled temperature and precipitate on trends consistent with the warming trend, while those farmers who believed in a stable climate, recalled temperatures and precipitations consistent with that belief. Interestingly, both groups showed about equal amounts of error.

From the previous studies (Bostrom et al., 1994; Read et al., 1994; Palutikof et al., 2004; Weber, 2010) on how people relate weather to climate the following conclusions were drawn:

- People are using the personal weather experience to judge about climate change;
- It is not obvious what weather parameters are usually associated with climate change.

Therefore, it is reasonable to use a set of weather parameters when trying to link the public climate change perception and experienced weather, and work on multiple hypotheses (taking all the potentially important influencing factors, which were suggested by climate change public perception studies, as independent variables).

2.2 The agenda-setting theory

The agenda-setting role of the mass media is their influence on the salience of an issue and on specific opinions about this issue (McCombs, 2013). It is considered, that agenda-setting theory was formally developed by McCombs and Shaw (1972). In their presidential election case study, McCombs and Shaw were able to evaluate the degree to which the media determined the most important election issue and public salience of it. Nowadays the theory continues to be regarded as relevant.

There were studies on how long an issue will remain salient in people's minds in agenda-setting research (Winter and Eyal, 1981; Wanta and Hu, 1994). Winter and Eyal (1981) confirmed the agenda-setting theory and concluded that the "optimal effect span", which is the peak association between media emphasis and public emphasis of an issue, is between 4 and 6 weeks. Wanta and Hu (1994) examined time lags for agenda setting for five news media. They found time lags of 1 week for national network newscasts, 2 weeks for local newscasts, 3 weeks for a regional newspaper, 4 weeks for a local newspaper, and 8 weeks for a national news magazine, while a combination of the five news media produced an optimal time lag of 3 weeks.

Nevertheless, more recent studies conducted in the Internet era, suggest that the time lag effect of the agenda setting has substantially decreased, as the Internet has drastically changed the ways in which many people receive news and information. In Roberts et al. study the time lag varied between 1 and 7 days (Roberts et al., 2002). In 2011 Meraz also conducted a study based on the agenda-setting theory, where optimal time lags were tested. The results showed that one-day lag interval was supported by the data (Meraz, 2011). Therefore it is reasonable to be concerned about the time frame over which media coverage has the most impact on public opinion. Moreover, time-lag analysis is important because it might demonstrate the time-varying causal effects. Logically, the time lag for traditional media, such as “prestige press”, to affect online discussions should be relatively short. Thus, our analysis traced the influence of news media coverage for time lags ranging from 1 day to 1 week.

2.2.1 Mass media and public perceptions of climate change

It is considered that anthropogenic climate change first emerged on the public agenda in the early 1950s, when the newspaper Saturday Evening Post published an article entitled “Is the World Getting Warmer?”, which explored relations between the temperature change, agricultural shifts and rising sea levels (Abarbanel and McClusky, 1950). The peak of media coverage on climate change came in 1957, which was proclaimed the “International Geophysical Year” by the International Council of Scientific Unions (Boykoff and Rajan, 2007). One of the most prominent articles of that year entitled “Are Men Changing the

Earth's Weather?" was published by Robert C. Cowen in the Christian Science Monitor (Cowen, 1957).

Throughout the 1960s and 1970s media coverage of climate science remained sparse and only a few articles were published in newspapers. In the 80s the mass media discourse was mainly focused on the scientific findings and reports, e.g. published by the IPCC, severe extreme events and high-level policy meetings (Weber, 2012). Among the most noticeable events of that decade was the statement of NASA scientist James Hansen for the US Congress that there is 99% certainty that "warmer temperatures were caused by the burning of fossil fuels and not solely a result of natural variation", which generated substantial media coverage in 1988 (Boykoff and Rajan, 2007). Later other environmental issues, e.g. 1988 drought and 1989 Exxon Valdez oil spill, resulted in a dramatic decrease in public discussion of climate change.

The interface of climate science and mass media has become an increasingly politicized in the 1990s. This decade might be characterized by the emergence of a group of "climate sceptics", who were often funded by carbon-based industries (Boykoff and Boykoff, 2004) and by the Kyoto meeting in December 1997, when the representatives of the US and other nations met in an effort to combat global warming by signing an international treaty to limit greenhouse gas emissions. These resulted in debate in the media about whether or not climate change was occurring at all and temporarily pushed the issue of climate change into the national media spotlight.

In 2000 Krosnick et al. assessed the impact of this debate on the public perception of climate change (Krosnick et al., 2000). The authors conducted surveys before and after the media campaign. Interestingly, the authors found no evidence of news media agenda-setting. Although there was an increase in media coverage, there was no change in the proportion of participants who thought that global warming was likely to be an extremely serious national problem. The authors note that this may be because most past studies of agenda-setting have focused on judgments of a problem's current seriousness, whereas in this survey people were asked about seriousness in the future. Nevertheless it may also be that agenda-setting is not simply the result of the volume of the problem coverage.

Another study based on the computer-assisted content analysis of mass media articles was conducted to identify the major discussion themes within the climate change domain (Kirilenko and Stepchenkova, 2012). The authors suggested, that not only the volume of climate change publications change with time, there is also a significant qualitative shift since 1980s. The data were obtained from The New York Times (NYT), which frequently plays an agenda-setting role for other news media (McCombs 2004). According to the authors' findings, the major change in the coverage of climate change is the sharp decline in the coverage of science of climate change. Another feature is the general politicization of the topic. This is consistent with the results of another study by Weber and Morris (2010): nowadays the economic and political instruments and the possible consequences of climate change are mostly discussed in the mass media.

Another study on public perception of climate change by Brulle et al. (2012) was based on the data from 74 separate surveys over a 9-year period. The authors defined five potentially significant factors, influencing public concern on the phenomena, namely, extreme weather events, exposure to and understanding of scientific information, media coverage, advocacy groups and elite cues. The authors found that weather events do not influence the overall level of public concern. The results indicated that the promulgation of scientific information about climate change has a small but significant effect, while the political communications appear to be more important. Agenda-setting theory was also confirmed: media coverage of climate change directly affected the level of public concern. The comparison of the results found by Brulle et al. (2012) and the results of this study can be found in section 6.1.

2.3 Utilizing Twitter data in public opinion studies

Twitter was launched in July 2006 by Jack Dorsey, Evan Williams, Biz Stone and Noah Glass and by July 2006. The service rapidly gained worldwide popularity, with 500 million registered users in 2012, who posted 340 million tweets per day (Fiegerman, 2012). In 2011 Bruns concluded that Twitter is “the second most important social media platform” after Facebook. As a social media platform, Twitter facilitates broader public discussions, helping to bridge the gap between policy-makers, scientists and general public (Ausserhofer et al., 2013).

Secondly, Twitter provides the unique opportunities for public opinion studies. To the best of my knowledge, as Twitter is a very young social media platform, it has only been used as a data source on public opinion in political and social science (Puschmann and Burgess, 2013; Ausserhofer et al., 2013). For example Bollen et al. (2011) conducted the mood (namely, tension, depression, anger, vigor, fatigue, and confusion) analysis of Twitter data. The authors were able to demonstrate the significant influence of socioeconomic factors on fluctuations of the mood levels. O’Connor et al. (2010) discussed the feasibility of using Twitter data as a substitute for traditional polls. The authors found that the presidential approval polls exhibited correlation with Twitter sentiment data, which makes it a valuable source of public opinion data on political preferences.

Nevertheless the use of Twitter data in research, especially in relation to climate science, is still a unique experience, demanding the development of new approaches, discussed in current thesis.

CHAPTER III

DATA

3.1 Twitter data

Traditionally polls and surveys have been used to take “snapshots” of public opinion on the question of interest. But many events, especially extreme weather phenomena, tend to unfold rapidly, giving the researches no time to prepare and conduct a survey. The microblogging platform like Twitter provides a unique opportunity to keep up with changes in the public opinion.

On Twitter, the registered users make friends and share their status, or “make posts”, within a limit of 140 characters. Each Twitter user has a brief profile about him. The public profile usually includes the full name, the location, a web page, a short biography, and the number of tweets of the user. Twitter contains the enormous amount of data not only due to the number of registered users. Compared to regular blogging, microblogging is characterized by faster mode of communication, which allows for the high frequency of update. These features make Twitter a unique social data source, popular among scholars (Huberman et al., 2008; Zhao et al., 2009; Kwak et al., 2010; Pak and Paroubek, P., 2010).

Twitter as a source of data, however, has its drawbacks. Despite Twitter users have an option to include their primary location into their profiles, and Twitter has features that allow users to share their current location, not all users choose to do so. This requires additional data processing procedures from a researcher, who wants to use this data. For example in 2007 a study by Java et al. showed that for the 76K users in the author’s data collection about 39K had specified locations that could be parsed correctly and resolved to their respective latitude and longitudinal coordinates. Nevertheless, geo-located Twitter data were considered valuable for many research applications, for example urban management and planning (Frias-Martinez et al. 2012) , public health assessment (Ghosh and Guha 2013) and tourism management (Hawelka et al., 2013).

In 2012 the tweets containing the key words “climate change” or/and “global warming” were collected in four languages: English, German, Russian and Spanish. The total number of 1,853,392 tweets were collected using the python code. For this study the United States subset of tweets was used.

A special GeoNames API – based code, which allows resolving the user’s status location, was developed (Kirilenko and Stepchenkova, 2014). The locations were manually filtered to exclude nonsensical or generalistic locations (e.g. “Earth”, “Moon”), sparse populated places with the population of less than 1000 and of less than 100 tweets originated within the study period, and places. A tweet was also excluded from the analysis if the time zone discrepancy between the user-listed time zone and the time zone of the resolved tweet location of greater than one hour was detected. Additional filtering was conducted to eliminate the erroneous tweets, for example collected due to presence of the search words in a URL link. After the filtering there were 664,226 tweets in the database. As the tweets are GMT time-stamped, the data were adjusted by hourly and weekly tweeting intensity, allowing for correcting the minor errors related to a few Internet service outages in 2012. The study on the global Twitter dataset and the data collection technique was published by Kirilenko and Stepchenkova (2014).

The sentiment analysis of tweets is beyond the scope of this thesis. It was assumed that if the change in weather pattern results in changes in public concern about climate change, more tweets contain the key words “climate change” or “global warming”. The set of weather parameters was chosen based on previous studies and discussed in the sections below.

3.2 Weather parameters

Ideally, this set of factors thought to be involved in the process of interest is chosen before data collection. For the purposes of the study the weather parameters that public might relate to climate change must be identified, based on the existing studies.

In 2001 Vedwan and Rhoades examined how apple farmers in the western Himalayas of India perceive climate change. The choice of the group of survey respondents was clear, as apple farmers in the Kully Valley heavily depend on climatic conditions and are aware of weather fluctuations. The authors found that changes in snowfall patterns were associated with climate change in the region the most. Participants in the study perceived a definite reduction in snowfall over time. Specifically, snowfall patterns were thought to change in two ways: (1) reductions in the intensity of snowfall and (2) shift in the timing of snowfall. The most common method people used to describe the changes was the recollection of memorable events, such as the largest snowfall in a decade. Participants however reported no discernible change in the rainfall intensity, but mentioned a shift in timing of rain events. Respondents said the monsoon rains were slightly displaced to the period beyond mid-August. Interestingly these changes were seen as a consequence of increasing amounts of late snowfall. The periodicity of temperature was also believed to be influenced by the timing of snowfall. For example late snowfall was implicated as a cause of cooler temperatures in March and April. Thus snowfall was the weather parameter apple farmers in the western Himalayas of India associated with climate change the most.

Another study by Maddison (Maddison, 2007) examined farmer' perception of climate change in ten countries in Africa. The author also compared farmer' responses with real climatic data, collected from the nearest ground stations. The results show that farmer' perception of climate change varied between the countries, but most of the farmers believed average temperature had increased. By contrast almost none believed the average temperature had decreased, apart from Ethiopia. Notably, in Cameroon farmers didn't see significant change in temperature at all. In Senegal and Kenya farmer's climate change concerns were primarily associated with decreased rainfall levels. The truthfulness of farmer' assumptions about climate change heavily depended on the respondent's years of farming experience. According to Hansen et al. (Hansen et al., 1998) temperature and precipitation are climate indicators noticeable by people, and the sense of changes expected to accompany climate change are well defined. The authors also note that records of temperature and precipitation are often longer and have a better chance of revealing a detectable change than alternative climate variables such as cloud cover, winds, and humidity. The Hansen's composite climate index is the average of a temperature index and a moisture index, and the scale of this index is based on standard deviation during the 30-year base period. The standard deviation is a measure of the typical year-to-year fluctuations of the given quantity, and a value of 1 (or -1) is great enough to be noticeable, because a value that large or larger would normally (for the base period) occur only about 15% of the time. For example, if the summer is warm enough to yield an index of 1 or greater at a given location, most people who had been living at that location for a long time would agree that it was a ''hot'' summer. The Hansen's hypothesis, however, was not supported by any field studies.

Based on the previous studies two groups of parameters that roughly represent the weather condition in study area were chosen. The first group consists of temperature-based parameters:

1) Temperature, *t*;

2) Temperature anomaly, *T:*

$$T = \frac{t - tnormal}{\sigma(t)}, \tag{1}$$

where t_{normal} is the mean daily temperature collected for the same date throughout the base period (1971-2000).

3) Absolute temperature anomaly, T_{abs}:

$$T_{abs} = abs(T) \tag{2}$$

4) Extreme temperature anomaly, $T_{extreme}$:

$$T_{extreme} = \begin{cases} abs(T), abs(T) > 1 \\ 0 \end{cases} \tag{3}$$

5) "Hot" temperature anomaly, T_{hot} :

$$T_{hot} = \begin{cases} T, T > 1 \\ 0 \end{cases} \tag{4}$$

6) "Cold" temperature anomaly, T_{cold}:

$$T_{cold} = \begin{cases} abs(T), T < -1 \\ 0 \end{cases}, \tag{5}$$

The second group consists of the following precipitation-based parameters:

1) Precipitation, p:

2) Precipitation anomaly, P:

$$P_i = \frac{p - p_{normal}}{\sigma(p)}, \tag{6}$$

where p_{normal} is the mean daily precipitation collected for the same date throughout the base period (1971-2000).

3) Extreme precipitation anomaly, $P_{extreme}$:

$$P_{extreme} = \begin{cases} P, P > 1 \\ 0 \end{cases} \tag{7}$$

4) Rain;

5) Snow;

6) Rain anomaly, $Rain_{anom}$;

$$Rain_{anom} = \frac{Rain - Rain_{normal}}{\sigma(Rain)}, \tag{8}$$

where $Rain_{normal}$ is the mean daily liquid precipitation collected for the same date throughout the base period (1971-2000).

7) Snow anomaly, $Snow_{anom}$

$$Snow_{anom} = \frac{Snow - Snow_{normal}}{\sigma(Snow)}, \tag{9}$$

where $Snow_{normal}$ is the mean daily solid precipitation collected for the same date throughout the base period (1971-2000).

The National Weather Service Summary of the Day available from the National Climate Data Center (NCDC) for 31,944 stations in the United States was chosen as the source of temperature and precipitation data (http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv=&georegionabbv=). Climate normals used in the parameter estimation were also obtained from NCDC.

3.3 Climate change publications data

The NYT is a daily newspaper, founded and continuously published in New York City since September 18, 1851. Its website is one of the most popular news sites in the United States, receiving more than 30 million unique visitors per month as officially reported in January 2011.

Following McCombs (2004), the NYT, which frequently plays an agenda-setting role for other news media, was chosen as a source of mass media climate change topic coverage data. The data was obtained on the daily basis and number of climate change publications per day, *Npub*, was used as a proxy of climate change related events of the national level, like presidential speeches and release of major scientific reports. There were 2706 publications, related to climate change. The number of publications significantly increased after the hurricane Sandy (Figure 1).

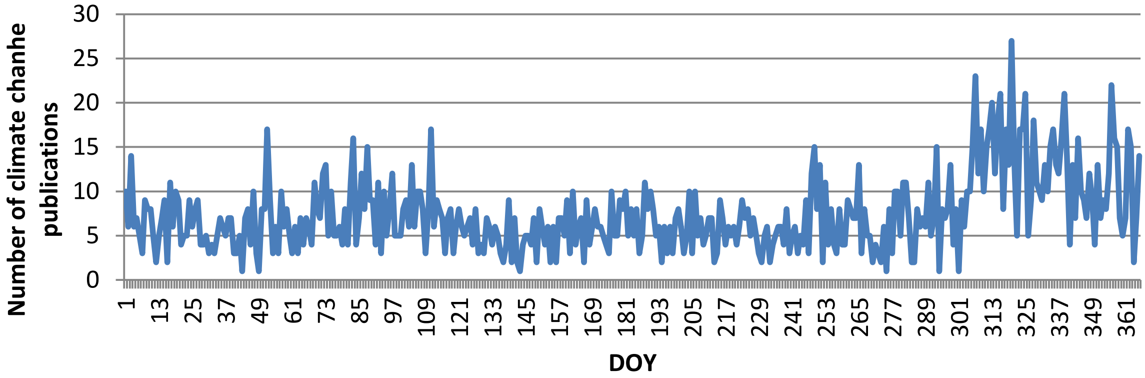

Figure 1 The number of climate change publications in NYT in 2012.

CHAPTER IV

METHODOLOGY

The objectives of the study were to utilize the climate change microblogging intensity in Twitter as a proxy of public salience data to test if the changes in weather parameters and in media coverage result in changes of the level of public concern on climate change. For this purpose the multiple linear regression and multi-model inference statistical techniques were used on three geographical levels of data aggregation.

4.1 Multiple linear regression

The multiple linear regression model is at least as widely-used in the time series context as in classical statistics, for example the common research task is to model the relationship between mortality rate and air pollution parameters (Wyzga, 1978; Shumway et al., 1988). Similarly, in economics multiple linear regression is used to identify which socio-economic factors might have influence on a variable of interest like crime rate or unemployment rate (Corman and Mocan, 1996; Raphael and Winter-Ebmer, 2001). The main idea is to express a response series, say x, as a linear combination of explanatory variables, say $y_1, y_2, \ldots, y_n$:

$$x=\beta_0 + \beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_2 y_n + \varepsilon \qquad (10)$$

Estimating coefficients $\beta_{1,} \beta_{2,\ldots,} \beta_n$ allows modeling x in terms of the inputs.

In this study the multiple regression model has the following structure: number of tweets is a dependent variable, influenced by weather parameters and number of newspaper articles on climate change. Any form of regression, however, relies on certain assumptions.

4.1.1 Assumptions

There are four principal assumptions for linear regression models. The first basic assumption is the linearity of the relationship between dependent and independent variables. Nonlinearity is usually most evident in a scatterplot of the dependent variable versus independent variable or a plot of residuals versus predicted values, which are a part of standard regression output. Independence of the errors (no autocorrelation) is a second assumption and could be a serious problem in time series regression models. The Durbin-Watson statistic provides a test for the data autocorrelation.

Another assumption is the homoscedasticity (constant variance) of the errors, which often arises in time series models due to the effects of inflation and/or real compound growth (Montgomery et al., 2012).

Finally the violation of the normality of the error distribution may arise either because of the distributions of the variables used are themselves significantly non-normal, and/or the linearity assumption is violated. For examination of the distribution of the variables Kolmogorov-Smirnov (for large number of observations) or Shapiro-Wilkinson (for small sample size) tests are used.

Additionally there must be no collinearity among the predictors, which can be assessed using Variance Inflation Factor (VIF) analysis. For this study the data transformation (weekly averaging and first order differencing) was applied to meet the required assumptions. The assumptions were checked for every model described in the results chapter.

After the data were collected and transformed, I aggregated the variables on three geographical levels.

4.2 Geographical levels

The analysis of climate change microblogging intensity was conducted at three different scale levels, namely, national, regional and local. This allowed for better understanding the microblogging intensity patterns and their relation to the local, regional and large-scale weather patterns. The number of climate change publications in the NYT remained the same on three different geographical levels.

For the national level analysis the station-based weather data collected from NCDC and transformed were aggregated on the 1,5°×1,5° geographical latitude and longitude grid and then averaged for the entire country. The total number of climate change tweets for the US was treated as a dependent variable.

For the regional level analysis the nine climate regions as defined by NCDC were used (Fig.2). The regions are defined based on the monthly temperature and precipitation averages that have been obtained from the stationary data and are regularly used in climate summaries (Karl and Koss, 1984; Gleason et al., 2007). For the regional level analysis weather data were aggregated on the 1,5°×1,5° geographical latitude and longitude grid and then averaged for the each climate region. For each climate region the separate analysis was conducted, using total number of tweets coming from each climate region as a dependent variable.

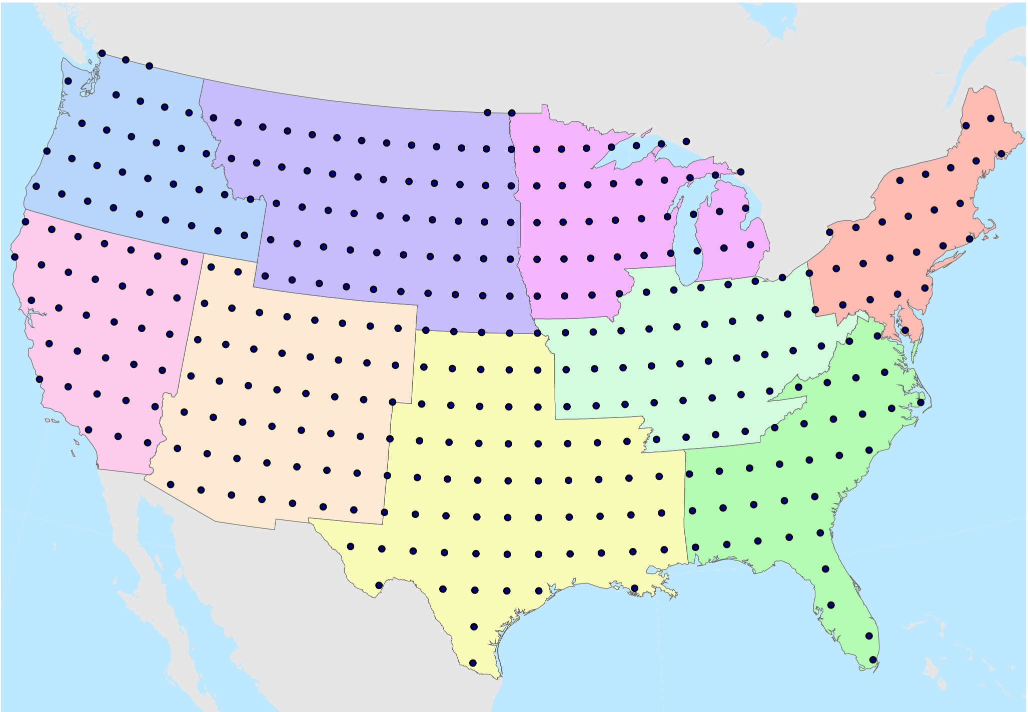

Figure 2 The nine climate regions as defined by NCDC and the 1.5°×1.5° geographical latitude and longitude grid used for data aggregation.

The urban areas in the United States as defined by the U.S. Census Bureau were chosen as a domain for the local level analysis (http://www.gpo.gov/fdsys/pkg/FR-2012-03-27/pdf/2012-6903.pdf). The list of 497 urban areas was filtered, so that only the urban areas with the population of more than 1000 and of more than 100 tweets originated within the study period were taken into analysis. After the filtering there were 245 urban areas (Appendix A, Table 11). The 2012 data series on weather and number of tweets in the radius of 0.5 degree for each location were collected. The time series from urban areas were merged in one matrix variable by variable, so that the data from the entire country would be used for the hypotheses testing, but based on finer resolution local scale that would allow picking up local weather anomalies on the contrast with the country level analysis. Additionally the weight was assigned to each urban area depending on the number of urban areas in each climate region, so that each climate region was equally represented in a final model using the formula:

$$Weight = \frac{Average\ number\ of\ urban\ areas\ in\ climate\ regions}{Number\ of\ urban\ areas\ in\ climate\ region} \tag{11}$$

The weights assigned for each urban areas could be found in Appendix A, Table 11.

4.3 Multiple working hypotheses

Having identified the parameters suitable for the purposes of the study, the functions that could mimic the relationship between independent variables (weather parameters and mass media topic coverage) and the response variable (microblogging intensity) had to be defined in terms of mathematical operators. The appropriate methods could have been found in literature or borrowed from other disciplines. For this thesis the multiple linear regression was chosen as the basic modeling technique. Nevertheless there are too many possibly useful predictors (six temperature-based variables, seven precipitation-based variables and number of climate change publications) to work with, which implies multiple working hypotheses.

(each hypothesis can be formulated as a linear model and tested separately). The concept of "multiple working hypotheses" was developed by Chamberlin (Chamberlin, 1965). In this concept there is no null hypothesis, instead, there are several scientifically justified hypotheses, equivalently, models. Relevant empirical data are collected and analyzed with the expectations that the results will tend to support one or more hypotheses, while rejecting other hypotheses. The concept is relevant for this study. Working with multiple hypotheses usually includes the model selection process, which is finding the most statistically significant predictor or a combination of predictors. There are several model selection approaches, described in literature.

4.3.1 Model selection approaches

Model selection is a process of finding the most statistically significant model(s) from the set of competing ones, where every model has a different predictor or a combination of predictors. The most popular approach of model selection is either a step-wise (forward) or a skip-wise (backward) sequential testing. However, when many parameters are present in the global model, sequential testing becomes a problem, as too many tests are to be made (Westfall et al., 1993).

Cross-validation is another option for model selection (Zucchini, 2000). For cross-validation the data are divided into two parts- for model fitting and for model validation. The whole process must be repeated hundreds of thousands of times. It is a computer-intensive technique and is rarely used when more than 15 models are to be evaluated or when dealing with large data sets.

In this study, multi-model inference (MMI), which is a modeling approach often used to compare competing candidate models, evaluate how well each is supported by data, and identify the best supported model(s), was used. MMI is based on the Akaike information criterion (AIC), which is a popular measure of the relative goodness of fit of a model, which was derived by Akaike (Akaike, 1973) as:

$$AIC = 2k - 2\ln(L), \tag{12}$$

where k is the number of parameters in the model, and L is the maximized value of the likelihood function for the model. The individual AIC values are not interpretable as they contain arbitrary constant and are much affected by the number of observations. To rescale AIC values the following equation is used:

$$\Delta AIC_i = AIC_i - AIC_{min}, \tag{13}$$

where AIC_{min} is the minimum of the AIC values, computed for all the tested candidate models. This transformation forces the best model to have ΔAIC of zero, while the rest of the models have ΔAIC of greater than zero. Usually the models having ΔAIC of ≤ 2 have substantial data support, and therefore should be concerned in further analysis. The detailed description of this statistic technique and examples can be found in (Akaike, 1973; Burnham and Anderson, 2004).

The code allowing computing ΔAIC values for the competing models was developed in R statistical software, allowing comparing all possible combinations of the predictors and testing multiple hypotheses. Two models supported by the data the most were retained each for the national and local levels, nine separate models were retained for the each climate region for regional level analysis. All the assumptions were checked for the each final model. The results are presented by the geographical levels.

CHAPTER V

RESULTS

The primary goal of this study was to explore how the weather patterns experienced in a certain location were translated into the public salience on climate change. On the national level the change in country-averaged temperature was positively associated with the US climate change microblogging intensity. In 2012 in the United States the temperature increase had a positive feedback on the change of the number of tweets on climate change.

The regional level analysis showed that in the Southwest and the West North Central climate regions the “cold” temperature anomalies were negatively associated with the climate change microblogging intensity, while in the Southeast and Central climate regions the abnormal as compared to climatological averages cooling has a positive effect on the number of climate change tweets. Perhaps this is the result of different preconceived beliefs about climate change in different parts of the country. In the Central and Northwest climate regions the precipitation increase had a positive effect on the climate change microblogging intensity. The Central climate region experienced the precipitation peaks in the late spring and late fall (due to Superstorm Sandy), which was reflected in the number of climate change tweets. The Northwest climate region experienced high precipitation in the early 2012 and in the end of the year matched by the increased climate change microblogging intensity.

The local level analysis didn’t bring more understanding in the relation between the weather parameters and climate change public perception. When all the urban areas were taken into account, the change in number of climate change publications in the NYT and abnormally “hot” weather were associated with the change in number of tweets, which is consistent with the results of the national-level part of the study.

The study showed that the regional-level analysis provided more statistically significant models. The explanation of this might be in the fact that the noticeable weather anomalies have usually regional geographical extent. The correlation of temperature anomaly time series for neighboring stations was illustrated by Hansen and Lebedeff (1987) as a function of station separation for different latitude bands: the average correlation coefficient was shown to remain above 50% to distances of about 1200 km at most latitudes, but in the tropics the correlation falls to about 35% at station separation of 1200 km.

The results are presented in detail by the geographical levels in the following sections (5.1, 5.2, 5.3).

5.1 Country level

A total of 664,226 tweets on climate change from the US in 2012 were obtained. The average number of tweets/day was 1,814.825, with the highest number of tweets (4,564) on the 31st of October, and with the lowest number of tweets (7) on the 27th of December. Microblogging intensity was also unevenly distributed in space (Fig.3).

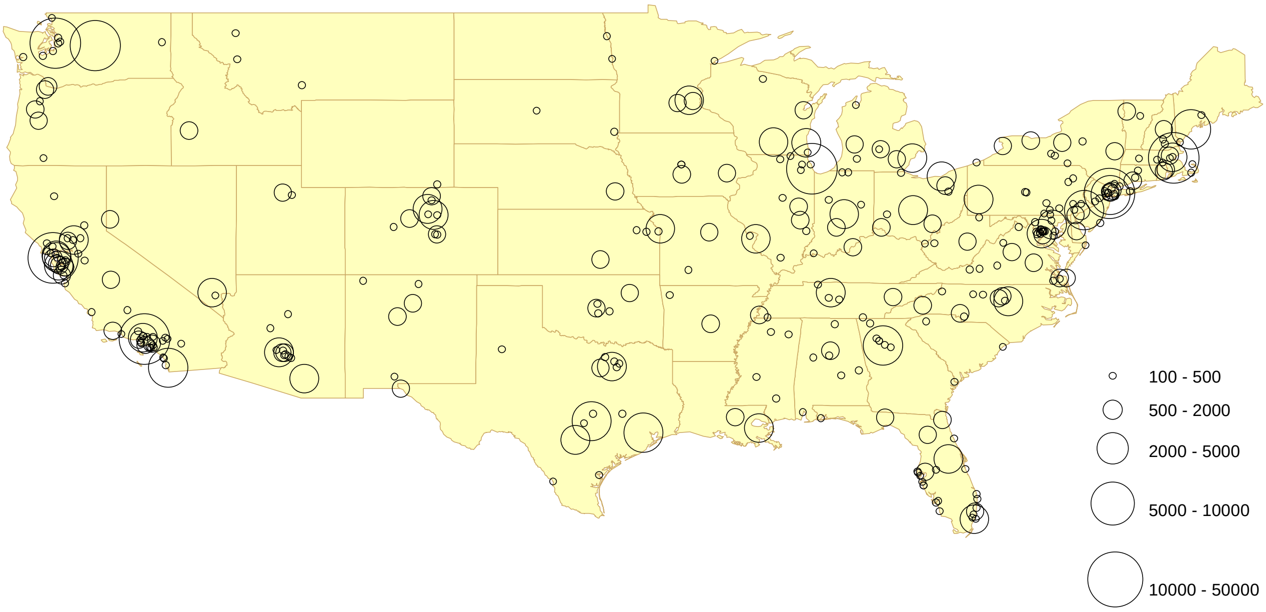

Figure 3 Distribution of tweets on climate change in the USA (year 2012). Only locations with population > 1000 and the number of tweets > 100 are shown

For the national level analysis all the data series were averaged for the entire country. The final model selected for the country level was chosen using MMI. Two explanatory variables significantly predicted the weekly change in number of tweets. It was found that weekly changes in number of climate change publications in the NYT, *Npub* (β = 1409.8, p < .05) and temperature t (β = 725.82, p < .05) were significant predictors. Hence, dependent variable i.e. change in climate change microblogging intensity (*Ntweets*), can be estimated using the following formula:

$$Ntweets = \beta_0 + \beta_1\, Npub + \beta_2\, t + \varepsilon \qquad (14)$$

The results of the regression indicated the two predictors explained 16% of the variance (R^2=0.16, RSE: 0.043 on 48 df, p<0.05). The final model statistics are summarized in table 1.The time series graph of the selected variables is shown in Figure 4.

Table 1 Final model parameters for country level

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.215	0.071	-0.037	0.970
N_{pub}	1.42	0.056	2.423	0.019
t	0.73	0.024	2.128	0.038

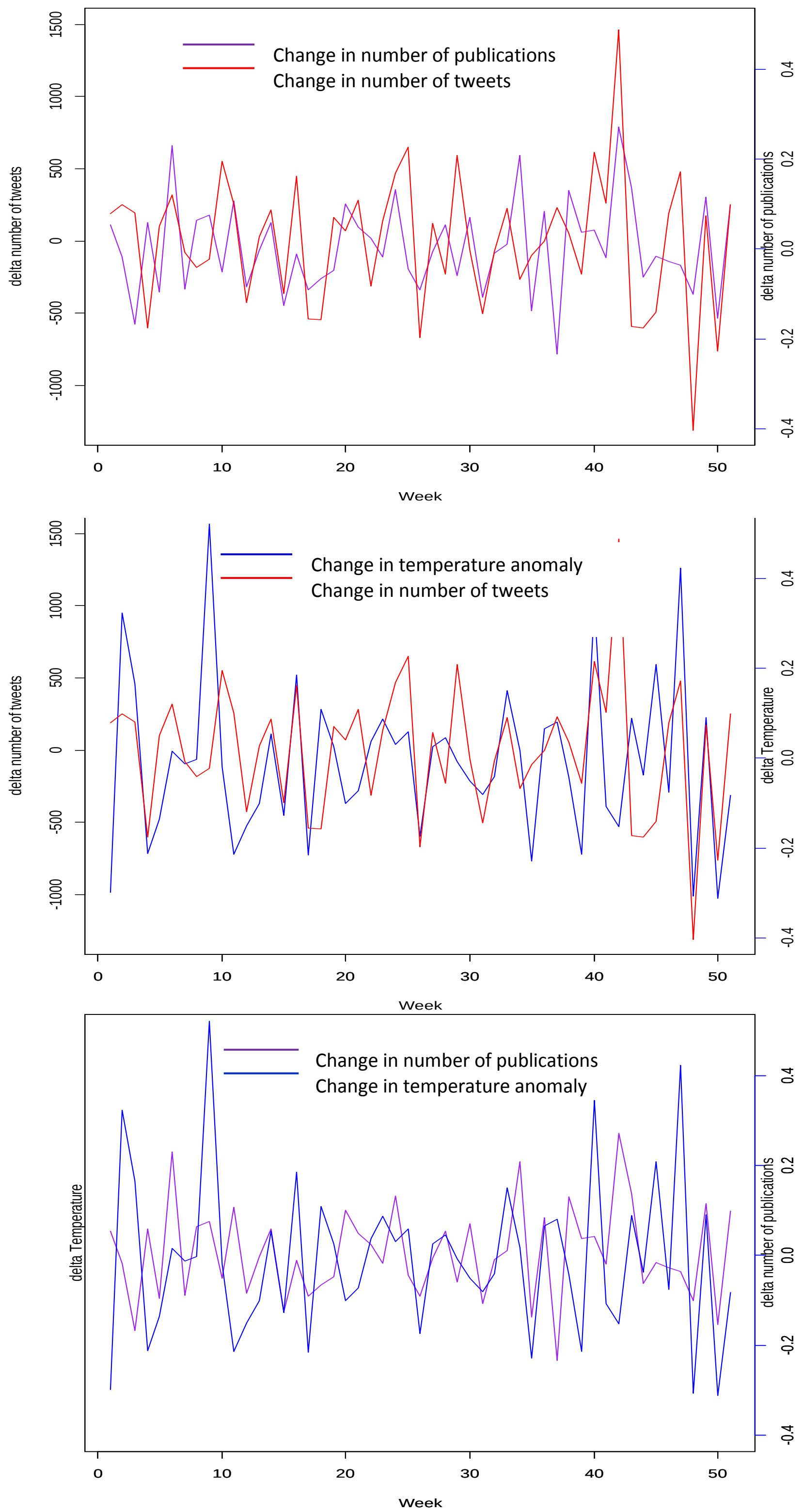

Figure 4 Time series of weekly change in number of climate change tweets, number of climate change publications in NYT and temperature in 2012 in the USA

The positive effects of temperature and number of climate change publications in the NYT on number of climate change tweets can clearly be seen in scatterplots of Figure 5, while no clear relation between the explanatory variables is seen. The time lag analysis was done. The outputs of the number of tweets didn't depend on lagged values of any other series.

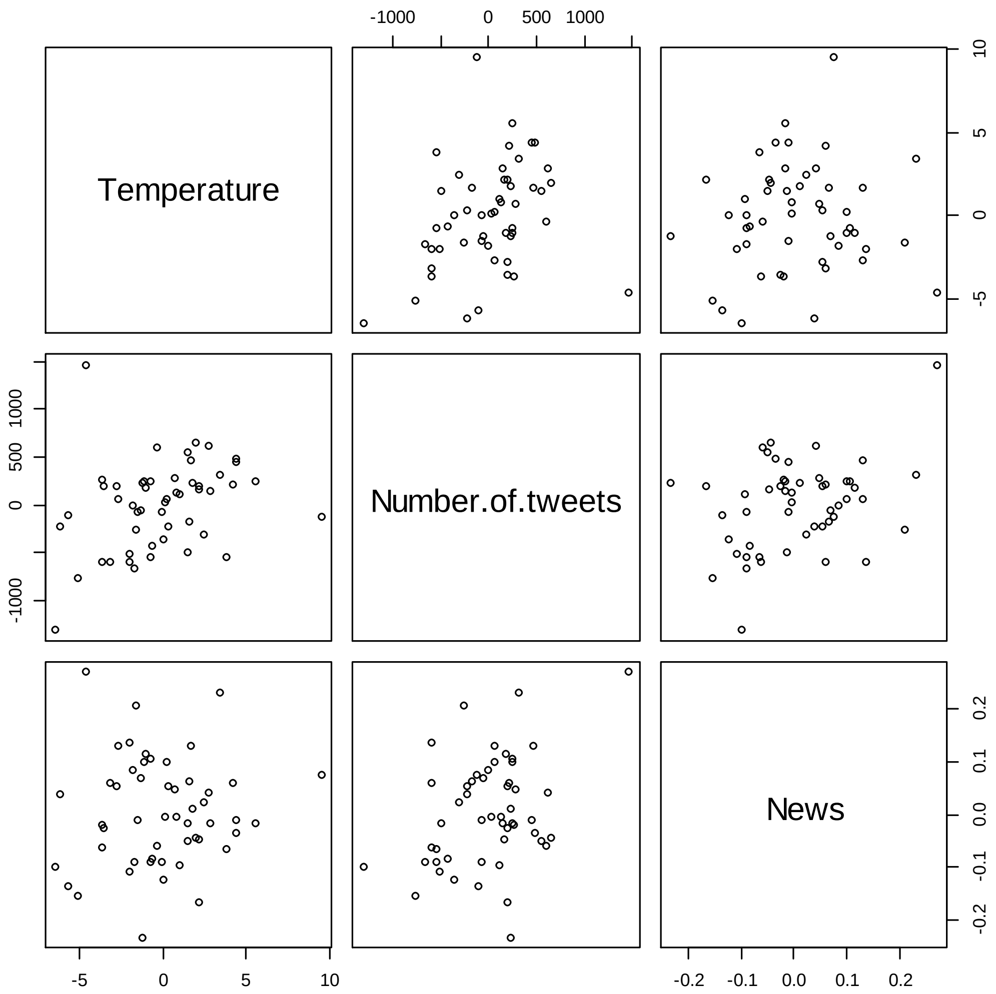

Figure 5 Scatterplot matrix showing plausible relations between the weekly change in number of climate change tweets, temperature and number of climate change publications in New York Times.

It should be noted that according to NOAA scientists, the globally averaged temperature for 2012 marked the 10th warmest year since record keeping began in 1880. In 2012, in the United States, warmer-than-average temperatures prevailed across much of the country. In 2012, the contiguous United States had its warmest March and April on record. The record-high July temperatures and warmer-than-average June and August, brought the contiguous United States its second hottest summer on record. In addition to the summer being hot for a large part of the country, it was also dry, resulting in a drought footprint comparable to the drought episodes of the 1950s.

5.2 Regional level

The nine climate regions as defined by NCDC were used for the study on the regional level (Fig.6). The regions are defined based on the monthly temperature and precipitation averages that have been obtained from the stationary data and are regularly used in climate summaries (Karl and Koss, 1984; Gleason et al., 2007). For the regional level analysis all the series were averaged for the each climate region. On the regional level the separate models were selected for each climate region using MMI.

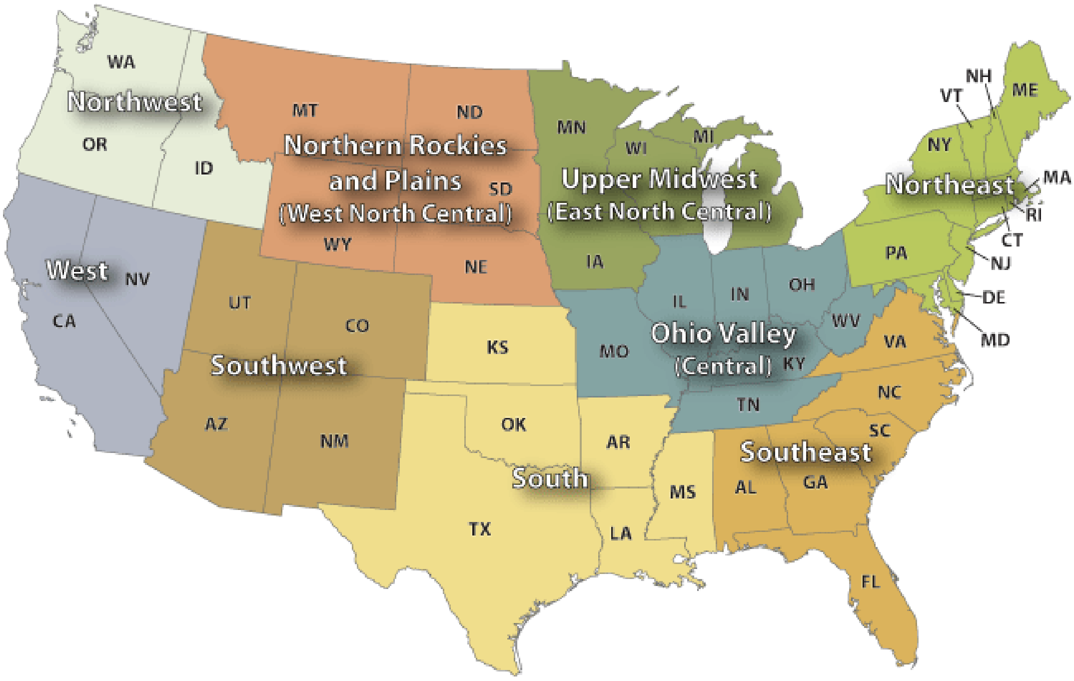

Figure 6 The nine regions as defined by the NCDC. Modified from

http://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php.

In 2012 the climate regions varied greatly in climate change microblogging intensity (Fig.7). The Northeast (NE) climate region had the highest total number of tweets (121375 tweets), and the West North Central (WNC) had the lowest total number of tweets (2620 tweets). When the population size was taken into account, the Northeast (NE) region had the highest number of 7.7 of tweets per 1000 persons, and the WNC had the lowest number of 1.7 tweets per 1000 persons in 2012 (Fig.8).

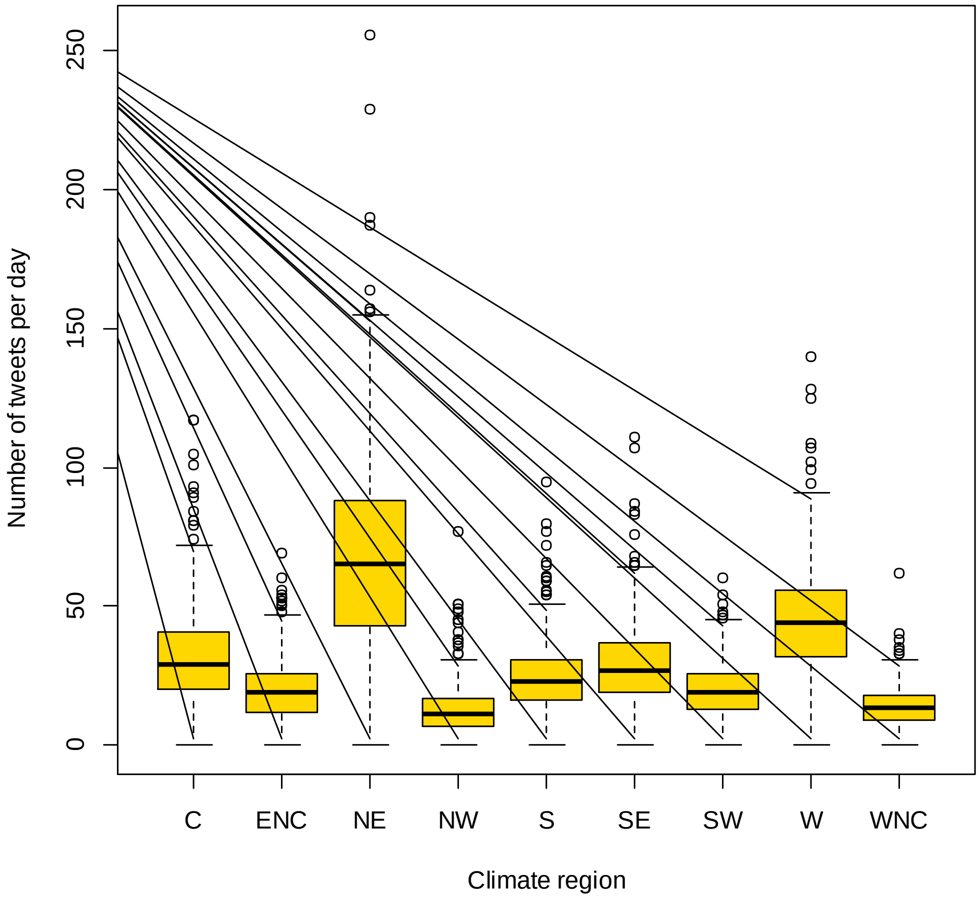

Figure 7 Number of climate change tweets in climate regions per day in 2012. C- Central, ENC-East North Central, NE- Northeast, NW- Northwest, S- South, SE- Southeast, SW- Southwest, W- West, WNC- West North Central climate regions.

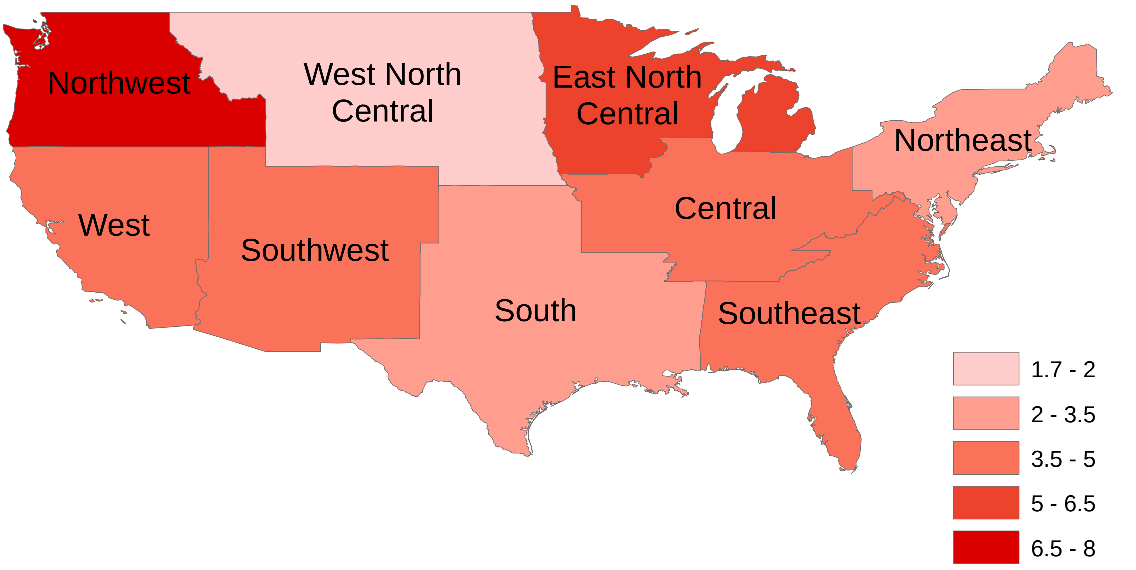

Figure 8 Number of climate change tweets per 1000 persons in 2012 in climate regions.

The results are reported by climate regions. The time lag analysis was included in the regional-level study. The results indicated that the outputs of the number of tweets didn't depend on lagged values of any other series.

5.2.1 Northwest climate region (NW)

For the NW climate region the final model selected for the country level was chosen using MMI. It contained two explanatory variables, which significantly predicted the weekly change in number of tweets. It was found that change in "hot" temperature anomaly T_{hot} (β = 0.01, p < .05) and precipitation, p (β = 0.07, p < .05) were significant predictors. Hence, dependent variable i.e. climate change microblogging intensity (*Ntweets*), can be estimated using the following formula:

$$Ntweets = \beta_0 + \beta_1\, Thot + \beta_2\, p + \varepsilon \qquad (14)$$

The results of the regression indicated the two predictors explained 13% of the variance (R^2=0.13, RSE: 0.4778 on 48 df, p<.05). The final model statistics are summarized in table 2. The positive effects of the predictor variables can clearly be seen in scatterplots of Figure 9, while no clear relation between the explanatory variables is seen.

Table 2 Final model parameters for NW climate region

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	0.009	0.066	0.143	0.886
p	0.068	0.034	2.015	0.049
T_{hot}	0.009	0.066	0.143	0.027

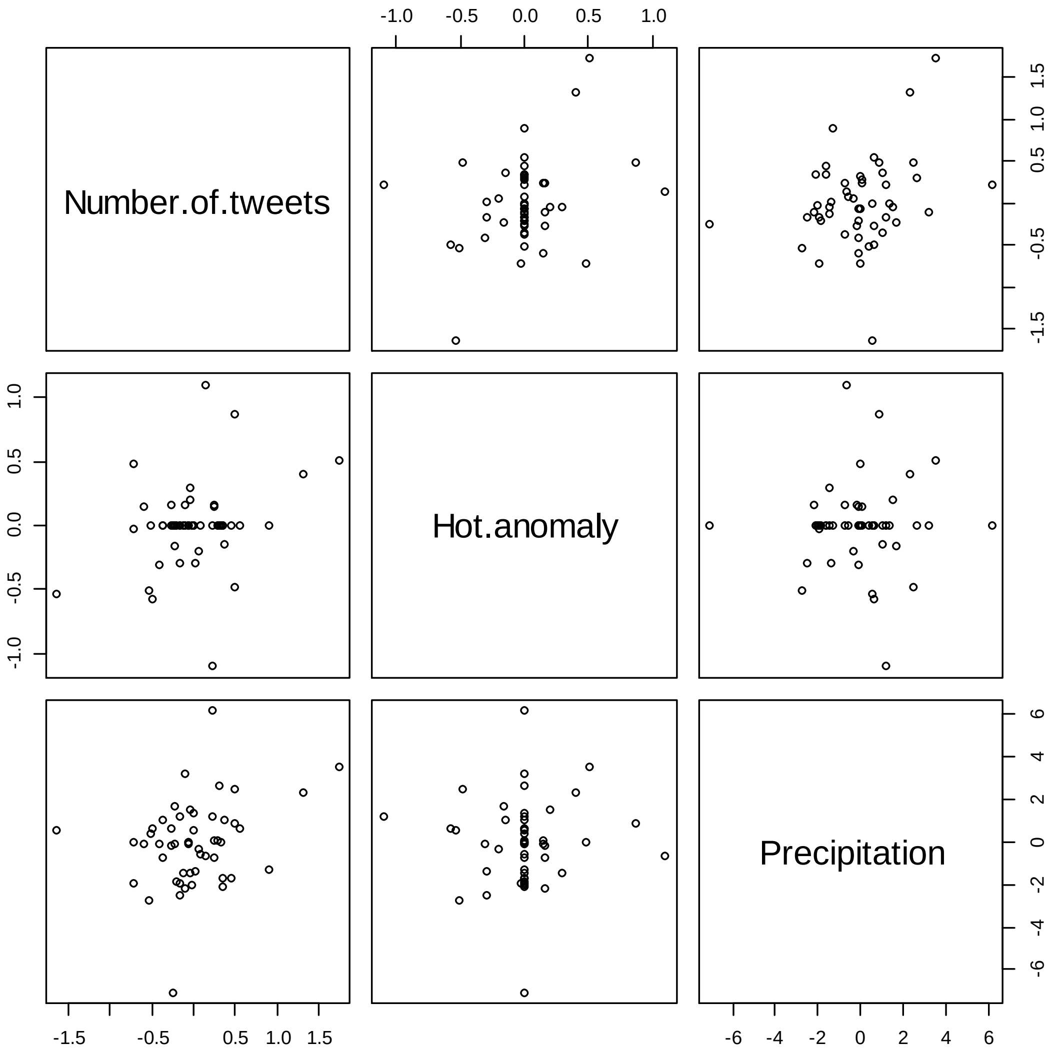

Figure 9 Scatterplot matrix showing plausible relations between the weekly change in number of climate change tweets, hot temperature anomaly and precipitation.

5.2.2 West climate region (W)

For the West climate region the final model contained one explanatory variable, which explained 11% of the variance (R^2=0.11, RSE: 0.6114 on 49 degrees of freedom, $p<.05$) in the dependent variable- the weekly change in number of tweets. It was found that change in temperature anomaly T ($\beta = 0.35$, $p < .05$) is a significant predictor. Hence, the climate change microblogging intensity (*Ntweets*), can be estimated using the following formula:

$$Ntweets = \beta_0 + \beta_1 T + \varepsilon \qquad (15)$$

The final model statistics are summarized in table 3. The positive effect of the change in temperature anomaly on number of tweets can clearly be seen in scatterplot (Figure 10).

Table 3 Final model parameters for the West climate region

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	0.001	0.085	0.017	0.986
T	0.350	0.131	2.667	0.010

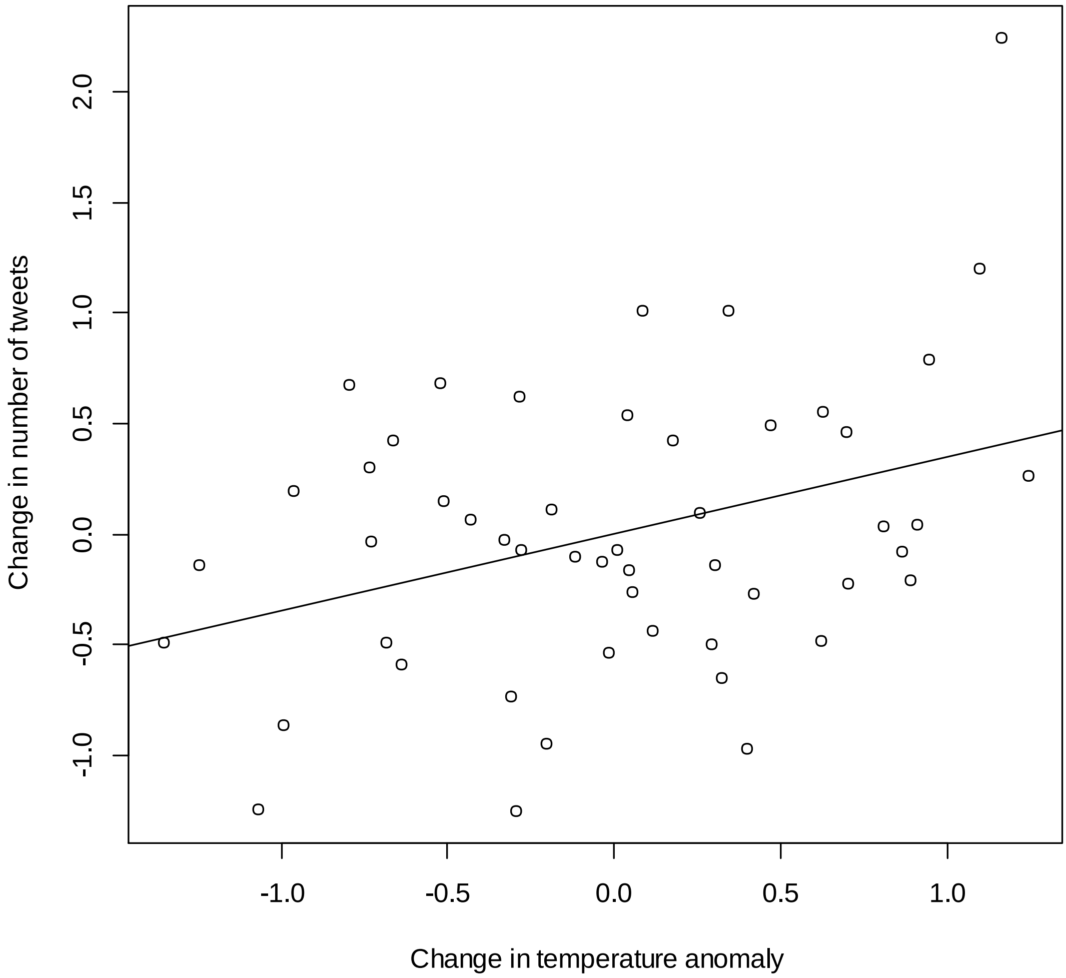

Figure 10 The change in temperature anomaly plotted against the change in number of climate change tweets in West climate region.

5.2.3 Southwest climate region (SW)

The results for the Southwest climate region indicate that the change in temperature anomaly *T* has significant ($p<0.01$) positive effect on the change of number of tweets, N_{tweets}. The change in this weather parameter explains 13% of the microblogging intensity variability in the SW climate region (RSE: 0.5297 on 49 df, R^2=0.13). The final model formula is:

$$Ntweets = \beta_0 + \beta_1 T + \varepsilon \tag{16}$$

The final model statistics are summarized in table 4. The scatterplot of Figure 11 illustrates the positive effect of the change in temperature anomaly on number of tweets.

Table 4 Final model parameters for the Southwest climate region

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	0.014	0.085	0.164	0.870
T	0.642	0.304	2.109	0.040

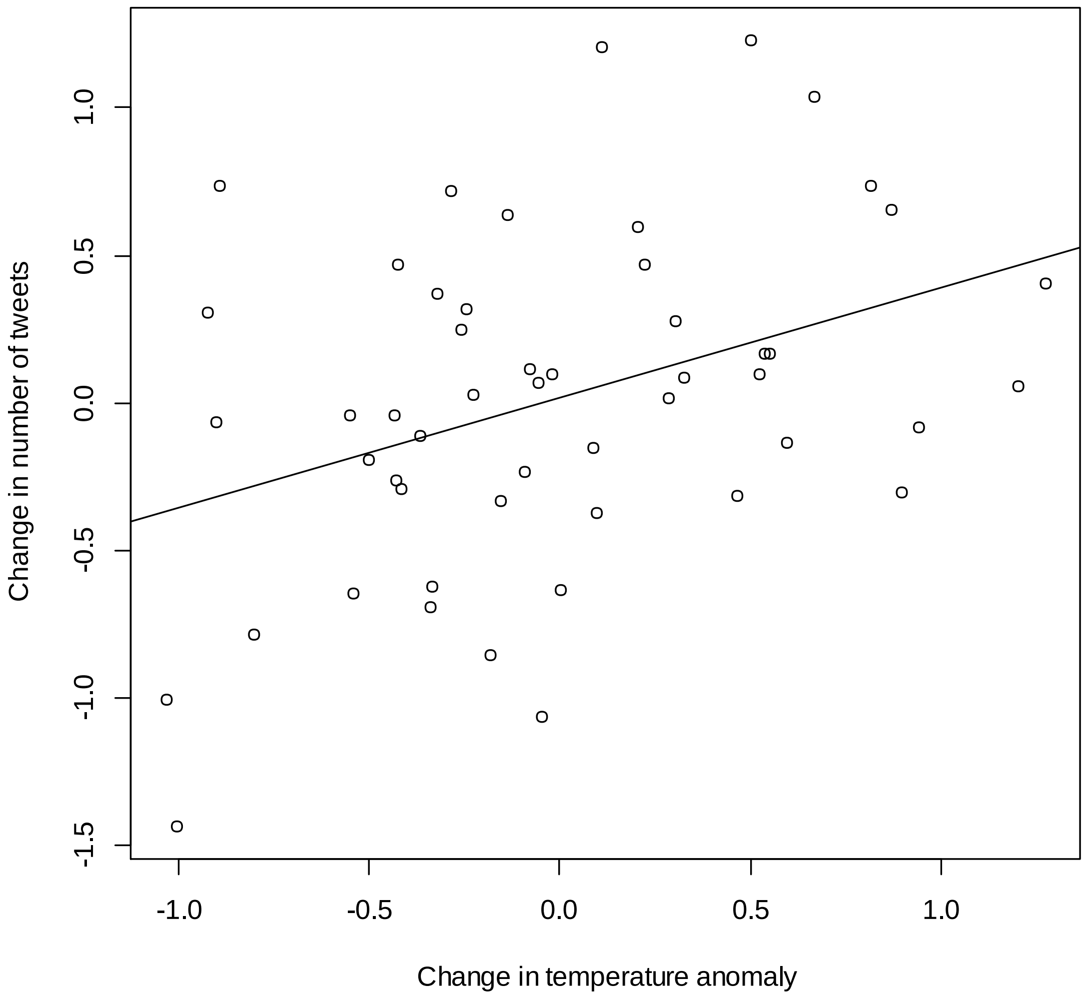

Figure 11 The change in temperature anomaly plotted against the change in number of climate change tweets in the Southwest climate region.

In some cases there is more than one statistically significant final model. For the SW climate region the second final model contained the change in "cold" weather anomalies T_{cold} as a predictor for the change in number of climate change tweets:

$$Ntweets = \beta_0 + \beta_1\, Tcold + \varepsilon \qquad (17)$$

This model was more statistically significant than the first one. The predictor variable explained 21% of variability in the dependent variable (RSE: 0.6134 on 15 degrees of freedom, R^2=0.21, p-value= 0.03). The effect of the "cold" temperature anomaly, however, was negative. The cooling was associated with the decrease in climate change microblogging intensity (Fig.12). The final model statistics are summarized in table 5.

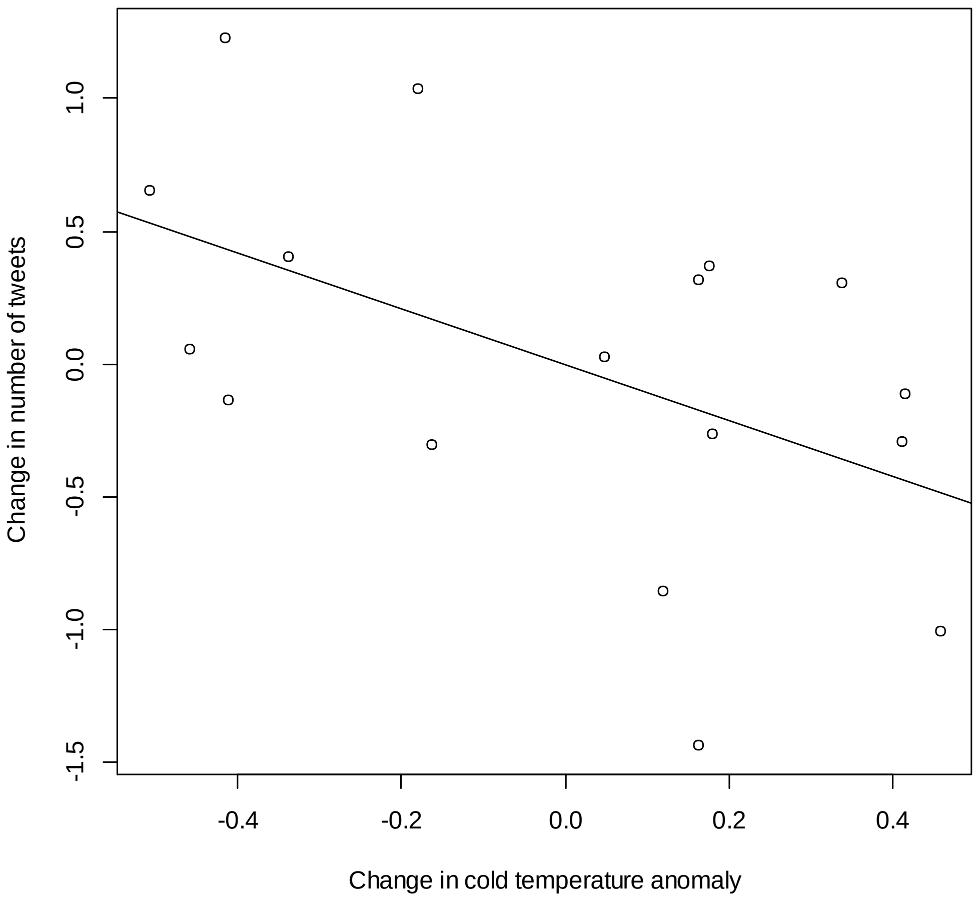

Figure 12 The change in “cold” temperature anomaly plotted against the change in number of climate change tweets in the Southwest climate region.

Table 5 Final model parameters for the Southwest climate region

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.001	0.148	-0.007	0.993
T_{cold}	-1.06	0.460	-2.292	0.036

5.2.4 West North Central climate region (WNC)

For the WNC climate region the “cold” temperature anomaly, T_{cold}, was also the most significant predictor of the number of climate change tweets. The change in single weather parameter explained 27% of microblogging intensity variability (RSE: 0.575 on 12 df, R^2=0.27, p-value<0.05). The final model formula is:

$$Ntweets = \beta_0 + \beta_1 \, Tcold + \varepsilon \qquad (18)$$

The final model statistics are summarized in table 6. The negative linear effect of the independent parameter can be seen in scatterplot of Figure 13.

Table 6 Final model parameters for the WNC climate region

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.028	0.153	-0.183	0.857
T_{cold}	-0.877	0.365	-2.4	0.033

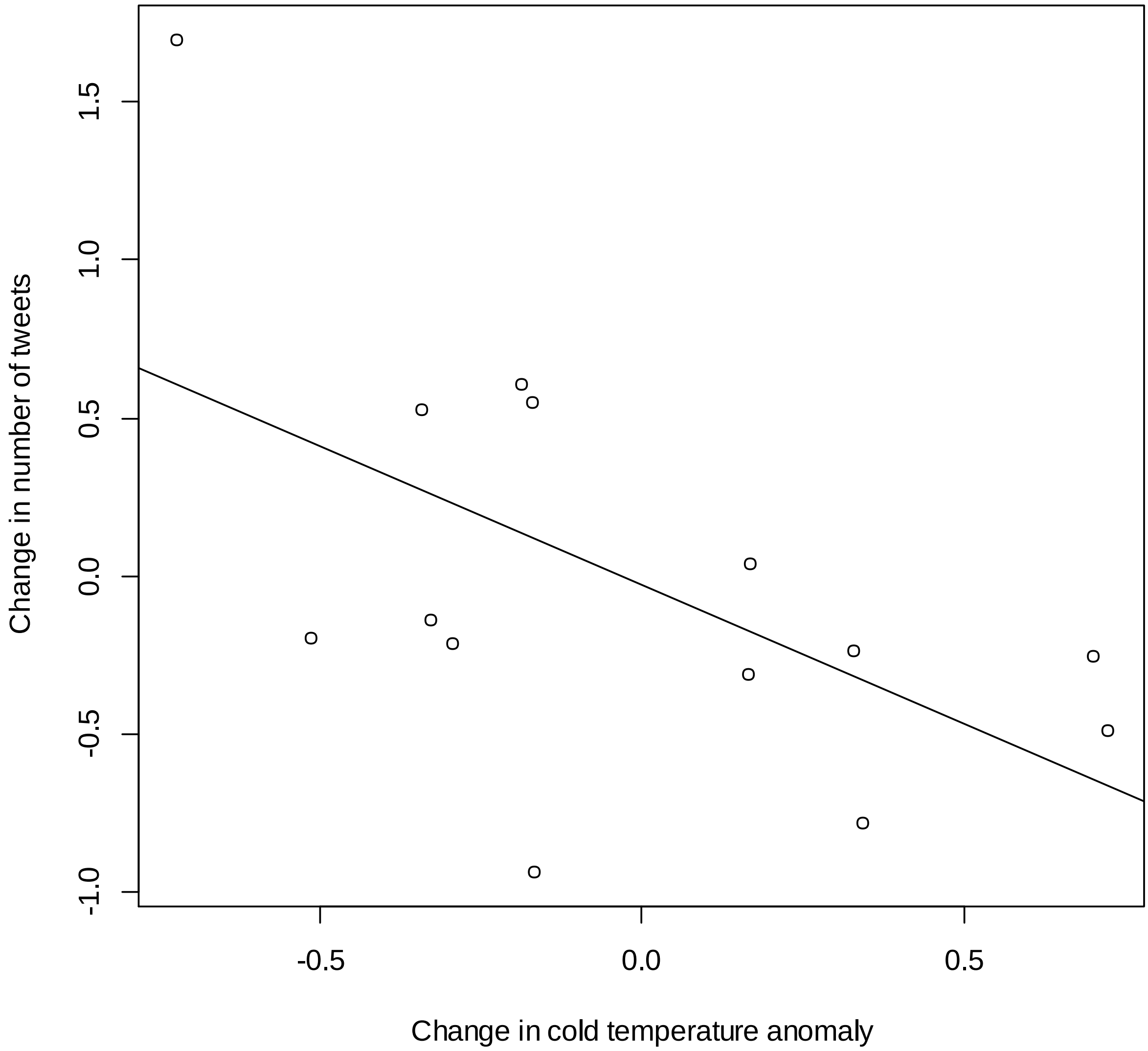

Figure 13 The change in “cold” temperature anomaly plotted against the change in number of climate change tweets in the WNC climate region.

5.2.5 East North Central climate region (ENC)

The MMI outcome for the ENC climate region was a final model:

$$Ntweets = \beta_0 + \beta_1\, Thot + \varepsilon \qquad (19)$$

Where the “hot” temperature anomaly, T_{hot}, alone explained 17% of variability of the dependent variable, N_{tweets} (RSE: 0.52 on 48 df, R^2= 0.17, p-value<0.005). The final model statistics are summarized in table 7.The effect of the change of “hot” temperature anomaly on the change in climate change microblogging intensity was positive. The warming weather was associated with the increase in climate change microblogging intensity (Fig.14).

Table 7 Final model parameters for the ENC climate region

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.012	0.075	-0.171	0.864
T_{hot}	0.513	0.155	3.298	0.001

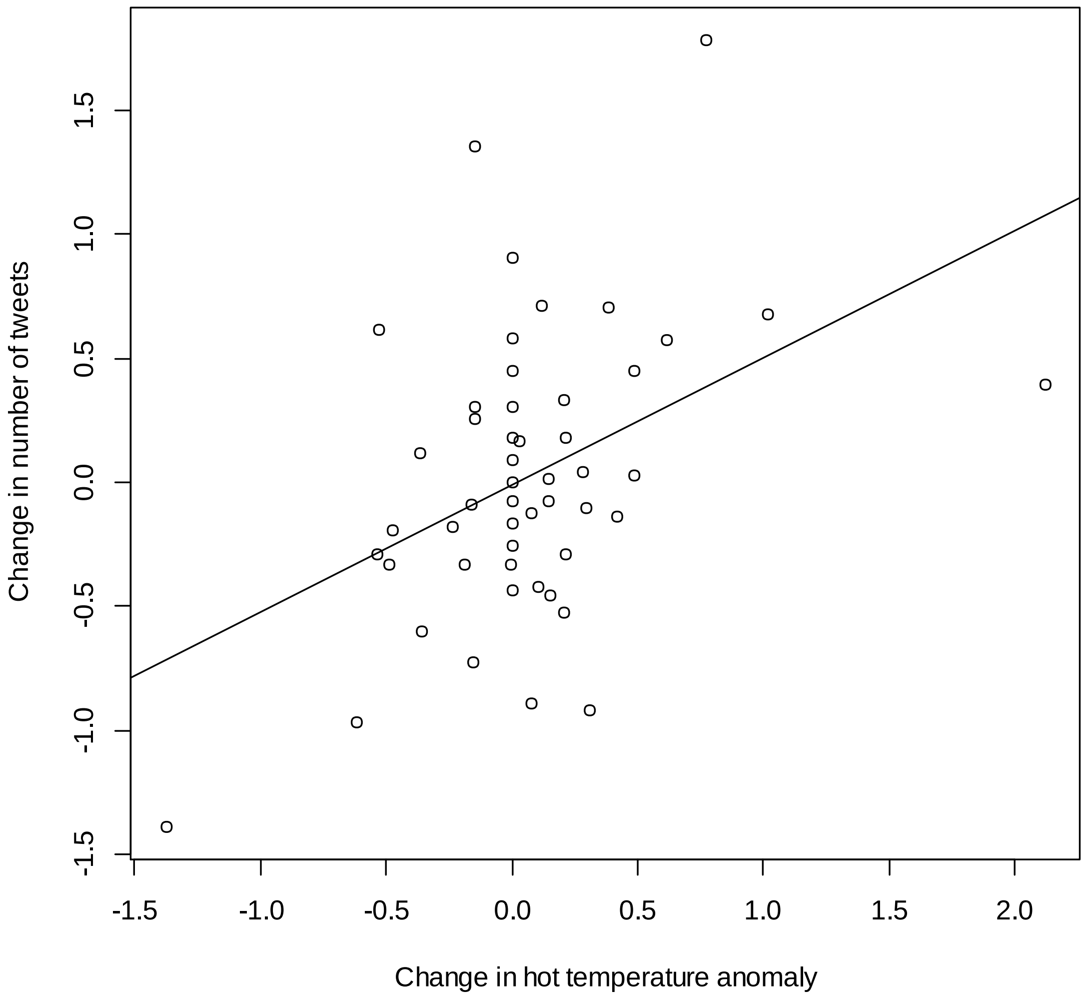

Figure 14.The change in “hot” temperature anomaly plotted against the change in number of climate change tweets in the ENC climate region.

5.2.6 Central (C)

The MMI analysis indicated that from weather parameters that were taken into account, two of them ("cold" temperature anomaly and precipitation anomaly) worked the best in describing the microblogging intensity variability on climate change in 2012 in the Central climate region. The final model formula is:

$$Ntweets = \beta_0 + \beta_1 Tcold + \beta_2 P + \varepsilon \quad (20)$$

The two parameters together described 10% of the dependent parameter variability (RSE: 0.7131 on 48 df, R^2=0. 1, p-value< 0.05). The final model statistics are summarized in table 8. As can be seen from the scatter plots (Fig. 15), each of the predictors matches closely with the dependent variable in some parts of the year, while no obvious relation between predictors themselves.

Table 8 Final model parameters for the Central climate region

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	0.005	0.099	0.053	0.958
T_{cold}	1.023	0.454	2.251	0.029
P	0.119	0.060	1.976	0.043

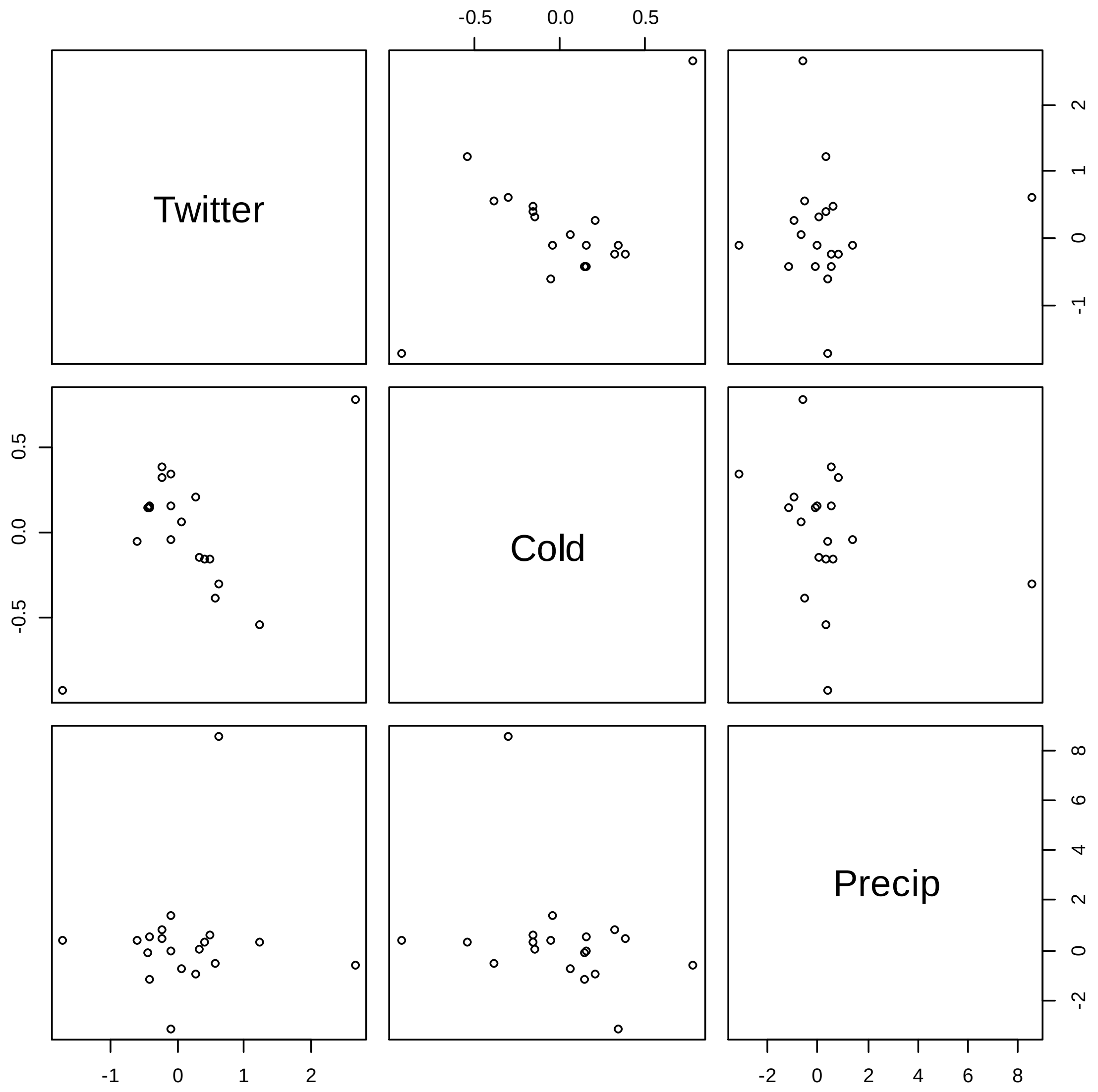

Figure 15 The change in “cold” temperature anomaly and precipitation anomaly plotted against the change in number of climate change tweets in the Central climate region.

5.2.7 Southeast (SE)

For the Southeast climate region the change in cold temperature anomaly as an independent variable was included into the final model:

$$Ntweets = \beta_0 + \beta_1\, Tcold + \varepsilon \tag{21}$$

The final model statistics are summarized in table 9. The change in cold temperature anomaly explained about 10% of the variability in microblogging intensity (RSE: 0.611 on 49 df,

R^2=0.08, p-value<0.05). The variables were positively associated, with no time lag (Figure 16).

Table 9 Final model parameters for the SE climate region

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	0.014	0.085	0.164	0.87
T_{cold}	0.642	0.304	2.109	0.040

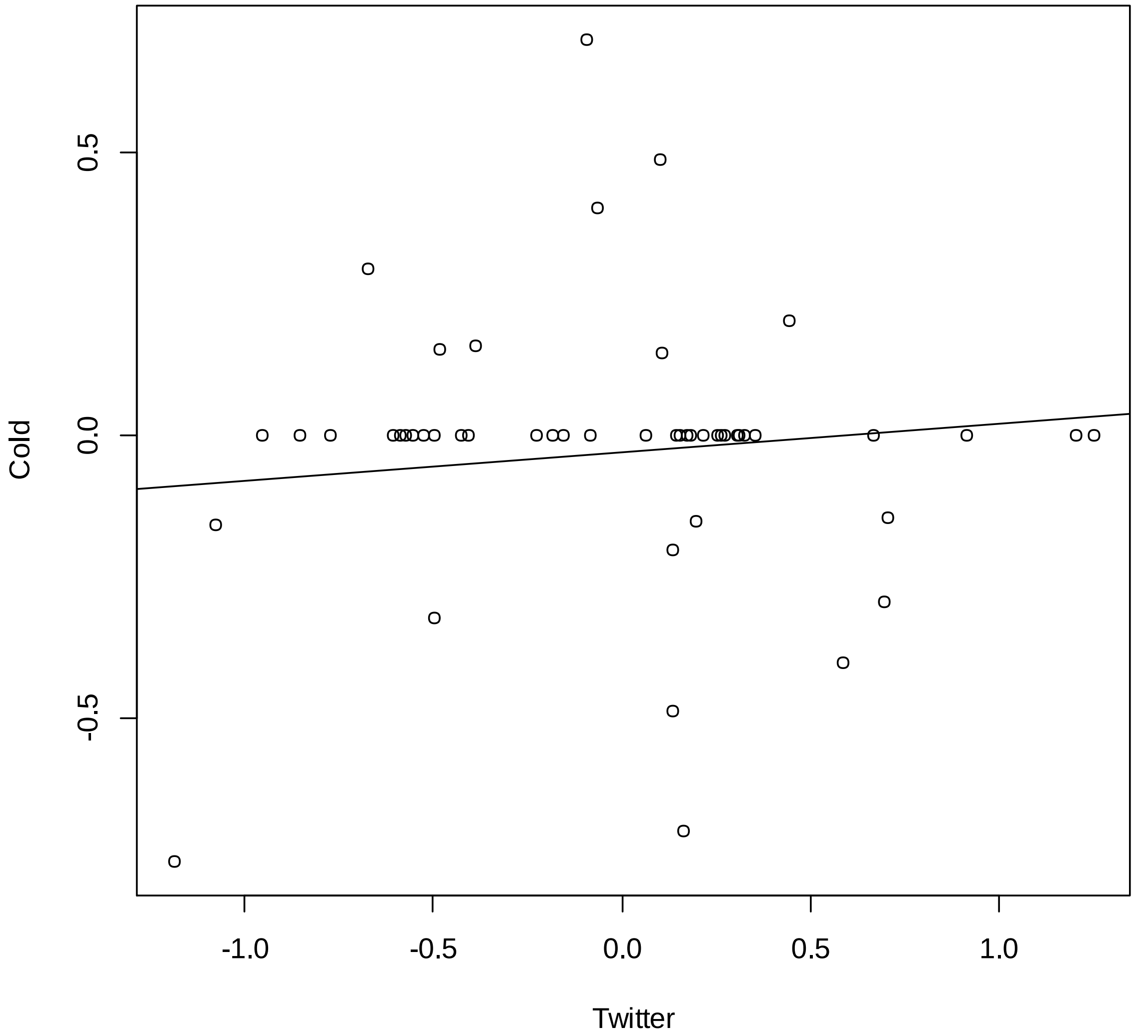

Figure 16 The change in “cold” temperature anomaly plotted against the change in number of climate change tweets in the SE climate region.

The regional level analysis showed which weather anomalies affected the climate change microblogging intensity in the different parts of the country in 2012 (Figure 17). No statistically significant models were found for the Northeast (NE) and South (S) climate regions. The temperature anomalies of 2012 were not reflected in numbers for climate change tweets, which is similar to the mismatch of the dependent variable with precipitation parameters. There was also no significant influence of climate change newspapers publications on the number of tweets.

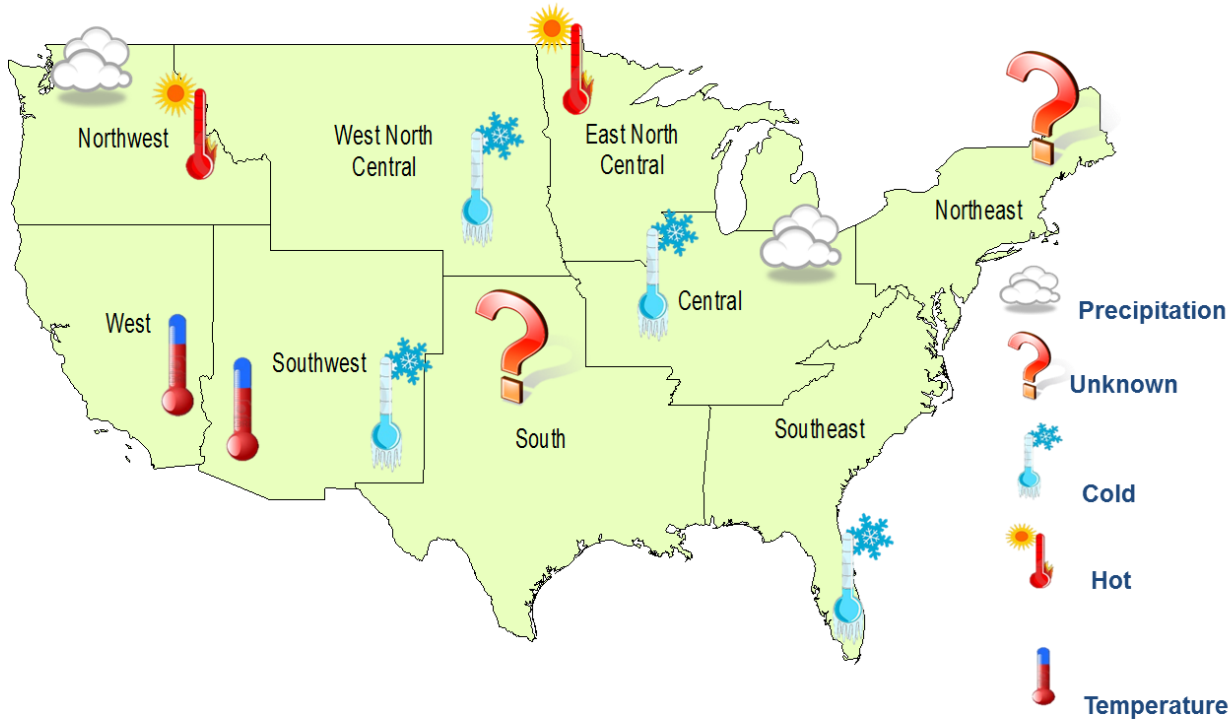

Figure 17. The weather phenomena affecting climate change microblogging intensity in 2012 by climate region

<u>5.3 Local level</u>

For the local level analysis the data series from urban areas were merged to find the weather parameters that describe the variability in climate change microblogging intensity in 2012. The MMI analysis determined the final model with the most statistically significant parameters:

$$Ntweets = \beta_0 + \beta_1\, Thot + \beta_2\, Npub + \varepsilon \qquad (21)$$

From table 10 we can see that the temperature increase, or a heat wave, and the change in number of climate change publications in the NYT were positively associated with the change in number of tweets in urban areas in 2012. The predictor variables, however, explained only 7% of the microblogging intensity variance (RSE: 2.252 on 9339 df, R^2=0.07 , p-value: < 2.2e-16). This might be explained by the large number of urban areas included into the model, experiencing different weather anomalies in 2012.

Table 10 Final model parameters for the urban areas

	β	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.00446	0.023306	-0.191	0.848
T_{hot}	0.471573	0.038586	12.221	<0.00001
N_{pub}	1.06841	0.227824	-4.69	<0.00001

CHAPTER VI

DISCUSSION, LIMITATIONS AND CONCLUSIONS

6.1 Discussion

The climate change is occurring and posing multiple risks for human and natural systems. The impacts of climate change can be alleviated and through adaptation and mitigation policy. For the promotion of climate change policy and research it is important to translate the substantial knowledge on the phenomenon and its potential threats to general public, as failure to take public values and views into consideration might result in misunderstanding and opposition by the electorate.

The public perception of climate change, however, is a complicated issue itself. The studies show that it is not obvious how the public awareness of climate change forms and how public opinion on climate change can be shifted towards more scientific understanding of it (Read et al., 1994; Vedwan and Rhoades, 2001; Weber and Stern, 2011). In scientific literature two major groups of factors were hypothesized to have the biggest influence on the level of public concern on climate change: extreme weather events and the mass media topic coverage. In 1998 James Hansen hypothesized that the weather parameters' variations, namely, temperature and precipitation, exceeding one standard deviation should be noticeable by people and result in increase of the level of public concern on the phenomena. Nevertheless no previous studies were able to test this hypothesis and demonstrate that people truly use the information about local weather to make assumptions about climate change. The other studies on public perception of climate change are generally based on the agenda-setting theory, stating that the level of public concern on the issue is a reflection of the extent and prominence of media coverage of the topic. These factors were tested previously using survey data (Vedwan and Rhoades, 2001; Weber and Stern, 2011; Brulle et al., 2012). Moreover, there have been no studies that tested these two groups of factors in a single model.

This study was set to test whether weather conditions and media activity together or separately influence public perceptions of climate change using Twitter data. This unique social media source of data allows for real-time, continuous monitoring of public opinion on various topics (O'Connor et al., 2010). The Twitter has broad, diverse audience, represented by users from many countries, which provides new opportunity for public opinion research (Java et al., 2009).

Specifically, the study developed a model of Twitter microblogging activity using weather parameters, described in section 3.2, and the number of media messages in NYT as an indicator of media activity about climate change. For this purpose the multiple linear regression and multi-model inference statistical techniques were used on three geographical levels of data aggregation, namely, national, regional and local.

The results indicate that on the national level both the temperature increase and increase in number of climate change publications had a positive feedback on the change of the number of tweets on climate change in 2012 in the United States. It should be noted that according to NOAA scientists, the globally averaged temperature for 2012 marked the 10th warmest year since record keeping began in 1880. Specifically in the United States, warmer-than-average temperatures prevailed across much of the country. In 2012, the contiguous United States had its warmest March and April on record. The record-high July temperatures and warmer-than-average June and August, brought the contiguous United States its second hottest summer on record.

The regional level analysis showed that in the Southwest and the West North Central climate regions the “cold” temperature anomalies were negatively associated with the climate change microblogging intensity, while in the Southeast and Central climate regions the abnormal as compared to climatological averages cooling has a positive effect on the number of climate change tweets. Perhaps this is the result of different preconceived beliefs about climate change in different parts of the country. In the Central and Northwest climate regions the precipitation increase had a positive effect on the climate change microblogging intensity. The Central climate region experienced the precipitation peaks in the late spring and late fall (due to Superstorm Sandy), which was reflected in the number of climate change tweets. The Northwest climate region experienced high precipitation in the early 2012 and in the end of the year matched by the increased climate change microblogging intensity. No statistically significant models were found for the Northeast and South climate regions. This might be due to other yet unknown factors, influencing public perception of climate change.

The local level analysis showed that the change in number of climate change publications in the NYT and abnormally “hot” weather were associated with the change in number of tweets, which is consistent with the results of the national-level part of the study.

The study showed that the regional-level analysis provided more statistically significant models. The explanation of this might be in the fact that the noticeable weather anomalies have usually regional geographical extent. The correlation of temperature anomaly time series for neighboring stations was illustrated by Hansen and Lebedeff (1987) as a function of station separation for different latitude bands: the average correlation coefficient was shown to remain above 50% to distances of about 1200 km at most latitudes, but in the tropics the correlation falls to about 35% at station separation of 1200 km.

For the first time in scientific literature the results clearly show that changes in weather parameters have significant effect on the level of public concern on climate change in contrast with the results obtained by Brulle et al. (2012). This discrepancy in the results might be explained by the fact that our study is based on the unique social media data, allowing for passive data collection and continuous monitoring, which is especially valuable for finding the link between fast-unfolding weather events and immediate public reaction.

The agenda-setting theory was also confirmed, which is consistent with the finding of Brulle et al. (2012): the mass media topic coverage was positively associated with the level of public concern on the national level. Finally it was investigated how long the issue of climate change remains salient in people's minds. In this study based on the weekly data no time lag between the newspaper topic publications number and number of climate change tweets was found, which is consistent with the more recent studies conducted in the Internet era (Meraz, 2011), which shows that the Internet has drastically changed the ways in which many people receive news and information This findings can be used for the future public opinion studies based on the Twitter data.

6.2 Limitations

One of the limitations for this study comes with the use of social media data. The huge amount of entries demand constant filtering out the erroneous texts that do not have relation to the climate change phenomena, inaccurate geographical locations and duplicates. Twitter, a microblogging service less than six years old, is a new source of data and no universal and effective method of processing this type of data was developed and described in literature. The results also indicate that there must be other yet unknown factors influencing climate change microblogging intensity. The weather anomalies and media coverage could explain only up to 30 percent of the variability in the tweeting time series on the topic.

6.3 Conclusions

The results of the study indicated that the variations in weather parameters were able to explain up to thirty percent of variance in climate change microblogging intensity. For the first time the results of the study showed that changes in weather parameters have significant effect on the level of public concern on climate change. The relation was demonstrated on the national, regional and local scales.

Mass media topic coverage was positively associated with the level of public concern, which is in agreement with the agenda-setting theory. Nevertheless the connection between the mass media and public salience is far from straightforward, as no statistically significant "agenda-setting" was found on the regional level. Perhaps that can be explained by the fact that the topic coverage does not necessarily determine public engagement, but rather shapes the possibilities for engagement (Boykoff, 2008).

The previous studies demonstrated the existence of the time lag between the peak of media emphasis and public emphasis of an issue. Nevertheless no time lag between the changes in media coverage, and changes in climate change microblogging intensity was found. This is accordant with the more recent studies that suggest that the time lag effect of the agenda setting has substantially decreased, as the Internet has drastically changed the ways in which many people receive news and information.

The study demonstrated that the social media data provides unprecedented opportunities for research. The passive data collection allows for real-time, continuous monitoring of the level of public concern on various topics. The social media audience is diverse and growing every day, which ensures its important role in future scientific research.

APPENDIX A

Table 11 Urban areas with weight

Urban area	Climate region	lat	lon	weight
Akron	Central	-81.49	41.06	0.82
Blountville	Central	-82.41	36.49	0.82
Charleston	Central	-81.64	38.35	0.82
Chattanooga	Central	-85.20	35.05	0.82
Chicago	Central	-87.71	41.83	0.82
Cincinnati	Central	-84.43	39.26	0.82
Clarksville	Central	-87.37	36.56	0.82
Cleveland	Central	-81.55	41.50	0.82
Columbus	Central	-82.99	40.01	0.82
Dayton	Central	-84.18	39.75	0.82
Evansville	Central	-87.54	37.99	0.82
Fort Wayne	Central	-85.11	41.11	0.82
Frankfort	Central	-84.86	38.19	0.82
Hammond	Central	-87.42	41.56	0.82
Hendron	Central	-88.64	37.03	0.82
Independence	Central	-94.40	39.05	0.82
Indianapolis	Central	-86.12	39.80	0.82
Jefferson City	Central	-92.20	38.57	0.82
Joliet	Central	-88.06	41.52	0.82
Kansas City	Central	-94.54	39.00	0.82

Knoxville	Central	-84.02	35.96	0.82
Lexington	Central	-84.50	38.02	0.82
Louisville	Central	-85.68	38.20	0.82
Memphis	Central	-89.90	35.12	0.82
Naperville	Central	-88.16	41.76	0.82
Nashville	Central	-86.73	36.19	0.82
Rockford	Central	-89.05	42.26	0.82
Rockton	Central	-89.05	42.46	0.82
South Bend	Central	-86.21	41.69	0.82
South Parkersburg	Central	-81.55	39.24	0.82
St. Louis	Central	-90.43	38.67	0.82
Toledo	Central	-83.62	41.64	0.82
Ann Arbor	East North Central	-83.69	42.26	1.76
Burlington	East North Central	-88.26	42.68	1.76
Cedar Rapids	East North Central	-91.63	42.02	1.76
De Pere	East North Central	-88.09	44.43	1.76
Des Moines	East North Central	-93.61	41.60	1.76
Detroit	East North Central	-83.20	42.47	1.76

Flint	East North Central	-83.68	43.00	1.76
Grand Rapids	East North Central	-85.67	42.93	1.76
Green Bay	East North Central	-88.06	44.52	1.76
Lansing	East North Central	-84.54	42.72	1.76
Madison	East North Central	-89.40	43.07	1.76
Milwaukee	East North Central	-87.89	43.06	1.76
Minneapolis	East North Central	-93.43	44.97	1.76
Oshkosh	East North Central	-88.57	44.01	1.76
Saint Paul	East North Central	-93.11	44.94	1.76
Albany	Northeast	-73.80	42.71	0.77
Allentown	Northeast	-75.47	40.61	0.77
Annapolis	Northeast	-76.52	38.98	0.77
Augusta	Northeast	-69.78	44.31	0.77
Baltimore	Northeast	-76.68	39.28	0.77
Boston	Northeast	-71.11	42.31	0.77

Bridgeport	Northeast	-73.21	41.19	0.77
Buffalo	Northeast	-78.78	42.93	0.77
Cambridge	Northeast	-71.12	42.38	0.77
Clifton	Northeast	-74.20	40.81	0.77
Concord	Northeast	-71.51	43.21	0.77
Dover	Northeast	-75.52	39.15	0.77
Erie	Northeast	-80.09	42.11	0.77
Harrisburg	Northeast	-76.82	40.26	0.77
Hartford	Northeast	-72.69	41.76	0.77
Irondequoit	Northeast	-77.60	43.22	0.77
Lowell	Northeast	-71.28	42.63	0.77
Manchester	Northeast	-71.44	43.05	0.77
Manhattan	Northeast	-73.95	40.82	0.77
Montpelier	Northeast	-72.57	44.26	0.77
New Haven	Northeast	-72.92	41.33	0.77
Philadelphia	Northeast	-75.12	40.07	0.77
Pittsburgh	Northeast	-79.97	40.46	0.77
Providence	Northeast	-71.43	41.85	0.77
Rochester	Northeast	-77.59	43.14	0.77
Rumson	Northeast	-74.14	40.10	0.77
Springfield	Northeast	-72.57	42.15	0.77
Stamford	Northeast	-73.56	41.05	0.77
Syracuse	Northeast	-76.15	43.05	0.77
Trenton	Northeast	-74.73	40.23	0.77

Washington	Northeast	-77.00	38.94	0.77
Waterbury	Northeast	-73.03	41.56	0.77
Worcester	Northeast	-71.81	42.25	0.77
Yonkers	Northeast	-73.86	40.94	0.77
Bellevue	Northwest	-122.16	47.64	1.76
Boise City	Northwest	-116.25	43.62	1.76
Eugene	Northwest	-123.11	44.06	1.76
Kent	Northwest	-122.31	47.42	1.76
Marietta-Alderwood	Northwest	-122.51	48.79	1.76
Newport Hills	Northwest	-122.13	47.56	1.76
Olympia	Northwest	-122.89	47.03	1.76
Opportunity	Northwest	-117.32	47.65	1.76
Pine Lake	Northwest	-122.03	47.57	1.76
Portland	Northwest	-122.62	45.54	1.76
Prairie Ridge	Northwest	-122.14	47.15	1.76
Salem	Northwest	-123.01	44.92	1.76
Tacoma	Northwest	-122.47	47.22	1.76
Vancouver	Northwest	-122.63	45.65	1.76
West Pasco	Northwest	-119.14	46.24	1.76
Abilene	South	-99.75	32.44	0.73
Amarillo	South	-101.86	35.19	0.73
Arlington	South	-97.09	32.69	0.73
Austin	South	-97.76	30.33	0.73
Baton Rouge	South	-91.13	30.47	0.73

Beaumont	South	-94.13	30.08	0.73
Ciudad Jußrez	South	-106.44	31.75	0.73
Corpus Christi	South	-97.41	27.74	0.73
Dallas	South	-96.74	32.89	0.73
Del City	South	-97.52	35.50	0.73
Denton	South	-97.12	33.21	0.73
Dewey	South	-95.93	36.73	0.73
El Paso	South	-106.38	31.79	0.73
Fort Polk South	South	-93.21	31.05	0.73
Fort Worth	South	-97.28	32.79	0.73
Houston	South	-95.42	29.80	0.73
Irving	South	-96.97	32.88	0.73
Jackson	South	-90.20	32.32	0.73
Killeen	South	-97.72	31.11	0.73
Lafayette	South	-92.03	30.21	0.73
Lenexa	South	-94.78	38.90	0.73
Little Rock	South	-92.36	34.76	0.73
Lubbock	South	-101.88	33.56	0.73
McAllen	South	-98.23	26.22	0.73
Metairie	South	-90.18	29.98	0.73
New Orleans	South	-90.05	29.99	0.73
Norman	South	-97.46	35.22	0.73
Overland Park	South	-94.68	38.91	0.73
San Antonio	South	-98.51	29.48	0.73

Shreveport	South	-93.76	32.46	0.73
Topeka	South	-95.70	39.02	0.73
Tulsa	South	-95.92	36.10	0.73
Waco	South	-97.15	31.54	0.73
Wichita	South	-98.52	33.90	0.73
Wichita Falls	South	-97.33	37.69	0.73
Alexandria	Southeast	-77.25	38.86	0.61
Athens	Southeast	-83.40	33.96	0.61
Atlanta	Southeast	-84.37	33.77	0.61
Birmingham	Southeast	-86.89	33.48	0.61
Bithlo	Southeast	-81.18	28.54	0.61
Buena Ventura Lakes	Southeast	-81.35	28.33	0.61
Bunche Park	Southeast	-80.26	25.83	0.61
Cape Coral	Southeast	-81.99	26.64	0.61
Carrollwood Village	Southeast	-82.47	28.01	0.61
Cary	Southeast	-78.82	35.76	0.61
Charlotte	Southeast	-80.82	35.20	0.61
Cocoa	Southeast	-80.80	28.41	0.61
Columbia	Southeast	-80.98	34.02	0.61
Durham	Southeast	-78.89	35.99	0.61
Eden	Southeast	-79.72	36.51	0.61
Fayetteville	Southeast	-78.94	35.06	0.61
Fruit Cove	Southeast	-81.63	30.09	0.61
Gainesville	Southeast	-82.37	29.66	0.61

Greensboro	Southeast	-79.85	36.08	0.61
Gulf Breeze	Southeast	-87.09	30.38	0.61
Huntsville	Southeast	-86.63	34.72	0.61
Iona	Southeast	-81.94	26.49	0.61
Jacksonville	Southeast	-81.74	30.32	0.61
Lake Buena Vista	Southeast	-81.47	28.31	0.61
Macon	Southeast	-83.59	32.86	0.61
Middleburg	Southeast	-81.93	30.08	0.61
Mobile	Southeast	-88.12	30.71	0.61
Montgomery	Southeast	-86.26	32.37	0.61
Newport News	Southeast	-76.44	37.09	0.61
Norfolk	Southeast	-76.20	36.83	0.61
North Charleston	Southeast	-80.03	32.90	0.61
Oldsmar	Southeast	-82.73	27.89	0.61
Orlando	Southeast	-81.34	28.44	0.61
Pine Hills	Southeast	-81.47	28.58	0.61
Poinciana Place	Southeast	-81.47	28.14	0.61
Port St. Lucie	Southeast	-80.33	27.29	0.61
Raleigh	Southeast	-78.63	35.80	0.61
Savannah	Southeast	-81.11	32.05	0.61
Tallahassee	Southeast	-84.28	30.45	0.61
Tangelo Park	Southeast	-81.46	28.43	0.61
Upper Grand Lagoon	Southeast	-85.76	30.17	0.61
Winston-Salem	Southeast	-80.27	36.10	0.61

Winter Springs	Southeast	-81.38	28.77	0.61
Albuquerque	Southwest	-106.62	35.09	1.46
Aurora	Southwest	-104.87	39.70	1.46
Chandler	Southwest	-111.87	33.29	1.46
Colorado Springs	Southwest	-104.78	38.85	1.46
Denver	Southwest	-105.01	39.78	1.46
Flagstaff	Southwest	-111.63	35.20	1.46
Fort Collins	Southwest	-105.08	40.56	1.46
Gilbert	Southwest	-111.76	33.35	1.46
Los Chaves	Southwest	-106.77	34.73	1.46
Mesa	Southwest	-111.75	33.42	1.46
Peoria	Southwest	-89.62	40.74	1.46
Phoenix	Southwest	-112.02	33.37	1.46
Provo	Southwest	-111.66	40.23	1.46
Pueblo	Southwest	-104.62	38.27	1.46
Sandy	Southwest	-111.91	40.67	1.46
Santa Fe	Southwest	-105.97	35.66	1.46
Tempe	Southwest	-111.93	33.38	1.46
Tucson	Southwest	-110.91	32.22	1.46
Antioch	West	-121.80	37.99	0.71
Bakersfield	West	-119.02	35.35	0.71
Carson City	West	-119.75	39.17	0.71
Chula Vista	West	-117.05	32.62	0.71
Corona	West	-117.56	33.88	0.71

Fairfield	West	-122.03	38.26	0.71
Fallon Station	West	-118.71	39.42	0.71
Fresno	West	-119.79	36.78	0.71
Henderson	West	-114.97	36.02	0.71
Lancaster	West	-118.15	34.68	0.71
Las Vegas	West	-115.16	36.15	0.71
Los Angeles	West	-118.12	33.94	0.71
Mexicali	West	-115.48	32.67	0.71
Modesto	West	-120.99	37.66	0.71
Moreno Valley	West	-117.24	33.93	0.71
Oceanside	West	-117.24	33.13	0.71
Oxnard	West	-119.18	34.19	0.71
Palmdale	West	-118.08	34.57	0.71
Rancho Cucamonga	West	-117.59	34.08	0.71
Reno	West	-119.78	39.52	0.71
Richmond	West	-122.15	37.73	0.71
Riverside	West	-117.40	33.95	0.71
Roseville	West	-121.28	38.75	0.71
Sacramento	West	-121.45	38.52	0.71
Salinas	West	-121.64	36.69	0.71
San Bernardino	West	-117.26	34.09	0.71
San Diego	West	-117.08	32.79	0.71
San Francisco	West	-122.43	37.75	0.71
San Jose	West	-122.05	37.39	0.71

Santa Clarita	West	-118.55	34.42	0.71
Santa Rosa	West	-122.71	38.44	0.71
Simi Valley	West	-118.74	34.27	0.71
Stockton	West	-121.30	37.97	0.71
Thousand Oaks	West	-118.87	34.18	0.71
Tijuana	West	-116.98	32.55	0.71
Visalia	West	-119.32	36.33	0.71
Bismarck	West North Central	-100.78	46.81	3.29
Cheyenne	West North Central	-104.80	41.14	3.29
Helena	West North Central	-112.03	46.60	3.29
Lincoln	West North Central	-96.67	40.80	3.29
Lockwood	West North Central	-108.52	45.79	3.29
Omaha	West North Central	-96.05	41.24	3.29
Pierre	West North Central	-100.35	44.36	3.29
Sioux Falls	West North Central	-96.74	43.53	3.29

REFERENCES

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

Abarbanel, A., McClusky, T. (1950). Is the world getting warmer? *Saturday Evening Post*, 1 Jul, p22.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60(2), 255-265.

Ausserhofer, J., & Maireder, A. (2013). National politics on twitter: structures and topics of a networked public sphere. *Information, Communication & Society*, *16*(3), 291-314.

Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. Retrieved from http://arxiv.org/abs/0911.1583.

Bostrom, A., Morgan, M.G., Fischhoff, B. & Read, D. (1994) What do people know about global climate change? *Risk Analysis,* 14, no. 6, 959-970.

Box, G. E., & Jenkins, G. M.(1970). Time Series Analysis: Forecasting and Control. Golden-D. iv, San Francisco, California, USA

Boykoff MT, Boykoff JM (2004) Balance as bias: global warming and the US prestige press. *Glob Environ Change,* 14: 125–136.

Boykoff MT, Rajan SR. Signals and noise. Mass-media coverage of climate change in the USA and the UK. *EMBO Rep*. 2007;8(3):207–211.

Boykoff, M. T. (2008). The cultural politics of climate change discourse in UK tabloids. *Political Geography*, 27(5), 549-569.

Boykoff, M. T., & Boykoff, J. M. (2004). Balance as bias: global warming and the US prestige press. *Global environmental change*, 14(2), 125-136.

Bravo-Marquez, F., et al. (2012). Opinion Dynamics of Elections in Twitter. Web Congress (LA-WEB), 2012 Eighth Latin American, IEEE.

Brulle, R. J., Carmichael, J., & Jenkins, J. C. (2012). Shifting public opinion on climate change: an empirical assessment of factors influencing concern over climate change in the US, 2002–2010. *Climatic change*, 114(2), 169-188.

Burnham, K. P., & Anderson, D. R. (2002). Model selection and multi-model inference: a practical information-theoretic approach. Springer, Fort Collins, Colorado, USA.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2), 261-304.

Chamberlin, T. C. (1965). The method of multiple working hypotheses. *Science*, 148(3671), 754-759.

Chambliss, E. L. and B. V. Lewenstein (2012). "Establishing a climate change information source addressing local aspects of a global issue. A case study in New York State." *JCOM* 11: 3.

Corbett, J. B., & Durfee, J. L. (2004). Testing public (un) certainty of science media representations of global warming. *Science Communication*, 26(2), 129-151.

Corman, H., & Mocan, H. N. (1996). A time-series analysis of crime and drug use in New York City (No. w5463). National Bureau of Economic Research.

Cowen RC (1957) Are men changing the earth's weather? *Christian Science Monitor*, 4 Dec, p13.

Fiegerman, S. (2012). "Twitter Now Has More Than 200 Million Monthly Active Users."

Finn, S. and E. Mustafaraj (2012). Real-Time Filtering for Pulsing Public Opinion in Social Media. FLAIRS Conference.

Frias-Martinez, V., Soto, V., Hohwald, H., & Frias-Martinez, E. (2012, September). Characterizing Urban Landscapes using Geolocated Tweets. In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom) (pp. 239-248). IEEE.

Ghosh, D., & Guha, R. (2013). What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartography and Geographic Information Science*, 40(2), 90-102.

Gleason, K. L., Lawrimore, J. H., Levinson, D. H., Karl, T. R., & Karoly, D. J. (2008). A revised US climate extremes index. *Journal of Climate*, 21(10), 2124-2137.

Golbeck, J. and D. Hansen (2011). Computing political preference among twitter followers. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.

Hansen, J. E., & Sato, M. (2001). Trends of measured climate forcing agents. Proceedings of the National Academy of Sciences, 98(26), 14778-14783.

Hansen, J., & Lebedeff, S. (1987). Global trends of measured surface air temperature. *Journal of Geophysical Research*: Atmospheres (1984–2012), 92(D11), 13345-13372.

Hansen, J., et al. (1998). "A common-sense climate index: Is climate changing noticeably?" Proceedings of the National Academy of Sciences 95(8): 4113-4120.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2013). Geo-located Twitter as the proxy for global mobility patterns. arXiv preprint arXiv:1311.0680.

Howe, P. D., Markowitz, E. M., Lee, T. M., Ko, C. Y., & Leiserowitz, A. (2013). Global perceptions of local temperature change. *Nature Climate Change*, 3(4), 352-356.

Huberman, B. A., Romero, D. M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. arXiv preprint arXiv:0812.1045.

Java, A., Song, X., Finin, T., & Tseng, B. (2009). Why we twitter: An analysis of a microblogging community. *In Advances in Web Mining and Web Usage Analysis* pp. 118-138. Springer Berlin Heidelberg.

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in ecology & evolution*, 19(2), 101-108

Jones, P., et al. (1999). The use of indices to identify changes in climatic extremes. *Climatic Change* 42(1): 131-149.

Karl, T. R., et al. (1996). "Indices of climate change for the United States." Bulletin of the American Meteorological Society 77(2): 279-292.

Kirilenko, A. P., & Stepchenkova, S. O. (2012). Climate change discourse in mass media: application of computer-assisted content analysis. *Journal of Environmental Studies and Sciences*, 2(2), 178-191.

Krosnick, J. A., Holbrook, A. L., & Visser, P. S. (2000). The impact of the fall 1997 debate about global warming on American public opinion. *Public Understanding of Science*, 9(3), 239-260.

Kulshrestha, J., Kooti, F., Nikravesh, A., & Gummadi, P. K. (2012, June). Geographic Dissection of the Twitter Network. In ICWSM.

Kupperman, K. O. (1982). The puzzle of the American climate in the early colonial period. *The American Historical Review*, 87(5), 1262-1289.

Kursinski, E. R. (1997). The GPS radio occultation concept: Theoretical performance and initial results, Ph.D. thesis, California Institute of Technology, Pasadena, California, 1997.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (pp. 591-600). ACM.

Kwon, E. S. and Y. Sung (2011). Follow me! Global marketers Twitter use. *Journal of Interactive Advertising*, 12(1): 4-16

Liu, X., Lindquist, E., & Vedlitz, A. (2011). Explaining media and congressional attention to global climate change, 1969-2005: an empirical test of agenda-setting theory. *Political Research Quarterly*, 64(2), 405-419.

Maddison, D. (2007). The perception of and adaptation to climate change in Africa (Vol. 4308). World Bank Publications.

McCombs, M. E. (2004). Setting the agenda: The mass media and public opinion. Malden, MA:Blackwell.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public opinion quarterly*, 36(2), 176-187.

McCright, A. M., & Dunlap, R. E. (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. *The Sociological Quarterly*, 52(2), 155-194.

Meraz, S. (2011). Using time series analysis to measure intermedia agenda-setting influence in traditional media and political blog networks. *Journalism & Mass Communication Quarterly*, 88(1), 176-194

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). Wiley, Hoboken, New Jersey, USA.

Moser, S. C. (2010). Communicating climate change: history, challenges, process and future directions. *Wiley Interdisciplinary Reviews: Climate Change*, 1(1), 31-53.

Newport, F. (2010). Americans' global warming concerns continue to drop. Gallup.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11, 122-129.

Pak, A., & Paroubek, P. (2010, May). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In LREC.

Palutikof, J. P., Agnew M.D. & Hoar, M.R. (2004) Public perceptions of unusually warm weather in the UK: impacts, responses and adaptations. *Climate Research* 26, 1, 43-59.

Parker, D. E., Gordon, M., Cullum, D. P. N., Sexton, D. M. H., Folland, C. K., & Rayner, N. (1997). A new global gridded radiosonde temperature data base and recent temperature trends. *Geophysical Research Letters*, 24(12), 1499-1502.

Puschmann, C., Burgess, J. (2013). The Politics of Twitter Data. *HIIG Discussion Paper Series* No. 2013 http://papers.ssrn.com

/abstract=2206225.Raphael, S., & Winter- Ebmer, R. (2001). Identifying the Effect of Unemployment on Crime. *Journal of Law and Economics*, 44(1), 259-283.

Read, D., Bostrom, A., Morgan, M.G., Fischhoff, B. & Smuts, T. (1994) What do people know about global climate change? 2. Survey studies of educated laypeople. *Risk Analysis* 14, no. 6 (1994): 971-982.

Roberts, M., Wanta, W., & Dzwo, T. H. D. (2002). Agenda setting and issue salience online. *Communication Research*, 29(4), 452-465.

Shumway, R. H., Azari, A. S., & Pawitan, Y. (1988). Modeling mortality fluctuations in Los Angeles as functions of pollution and weather effects. *Environmental Research*, 45(2), 224-241.

Smith, A., & Brenner, J. (2012). Twitter use 2012. Pew Internet & American Life Project.

Smith, T. M., & Reynolds, R. W. (2005). A global merged land-air-sea surface temperature reconstruction based on historical observations (1880-1997). *Journal of Climate*, 18(12), 2021-2036.

Stone, G. C., & McCombs, M. E. (1981). Tracing the time lag in agenda-setting. *Journalism Quarterly*, 58(1), 51-55.

Stovall, J. G., & Solomon, J. H. (1984). The poll as a news event in the 1980 presidential campaign. *Public Opinion Quarterly*, 48(3), 615-623.

Tan, Y., & Weaver, D. H. (2012). Agenda diversity and agenda setting from 1956 to 2004: What are the trends over time?. *Journalism Studies*, (ahead-of-print), 1-17.

Thomas R. Karl and Walter James Koss, 1984: Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895-1983. *Historical Climatology Series* 4-3, National Climatic Data Center, Asheville, NC, 38 pp.

Vedwan, N., & Rhoades, R. E. (2001). Climate change in the Western Himalayas of India: a study of local perception and response. *Climate Research*, 19(2), 109-117.

Wanta, W., & Hu, Y. W. (1994). Time-lag differences in the agenda-setting process: An examination of five news media. *International Journal of Public Opinion Research*, 6(3), 225-240.

Weber, E. U. (1997). Perception and expectation of climate change. *Psychological perspectives to environmental and ethical issues in management*, 314-341.

Weber, E. U. (2010). What shapes perceptions of climate change?. *Wiley Interdisciplinary Reviews: Climate Change*, 1(3), 332-342.

Weber, E. U., & Morris, M. W. (2010). Culture and Judgment and Decision Making *The Constructivist Turn. Perspectives on Psychological Science*, 5(4), 410-419.

Weber, E. U., & Stern, P. C. (2011). Public understanding of climate change in the United States. *American Psychologist*, 66(4), 315.

Weingart, P., Engels, A., & Pansegrau, P. (2000). Risks of communication: discourses on climate change in science, politics, and the mass media. *Public understanding of science*, 9(3), 261-283.

Westfall, P. H., Young, S. S., & Wright, S. P. (1993). On adjusting P-values for multiplicity. *Biometrics*, 49(3), 941-945.

Wilson, K. M. (2000). Communicating climate change through the media. *Environmental risks and the media*, 201-217.

Winter, J. P., & Eyal, C. H. (1981). Agenda setting for the civil rights issue. *Public Opinion Quarterly*, 45(3), 376-383.

Wyzga, R. E. (1978). The effect of air pollution upon mortality: a consideration of distributed lag models. *Journal of the American Statistical Association*, 73(363), 463-472.

Zhao, D., & Rosson, M. B. (2009). How and why people Twitter: the role that micro-blogging plays in informal communication at work. In Proceedings of the ACM 2009 international conference on Supporting group work (pp. 243-252). ACM.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, 44(1), 41-61.