Meteorological Event Identification using National Weather Service Forecast Discussions

Kaylin Bugbee[1], Rahul Ramachandran[2], Brian Freitag[1], Manil Maskey[2], Jeffrey Miller[3]

1 – University of Alabama in Huntsville; 2 – NASA Marshall Space Flight Center; 3 – Climate Forecast Applications Network

IMPACT Interagency Implementation and Advanced Concepts Team

Introduction

With the expected increased in data volume of Earth science datasets over the next decade, innovative data mining techniques are required to ensure data usability. One such approach includes using non-traditional data sources to extract relevant spatiotemporal information that can be used to access larger, more memory intensive datasets. If successful, information extracted from alternative data sources can be leveraged by the greater Earth science community to accelerate scientific discovery.

Methods

GOAL: Demonstrate the feasibility of using alternative data sources, specifically National Weather Service (NWS) area forecast discussions (AFD), to generate a curated database designed for accelerated scientific exploration.

NWS Area Forecast Discussions

The NWS issues AFDs about 4 times a day for each weather forecasting office (WFO). Forecast discussions are written by experienced meteorologists and include short-term and long-term outlooks with detailed discussion on upcoming weather events. Thus, this text data serves as an expertly curated dataset that provide sub-daily snapshots of the most important weather impacts for a given region. Archived NWS AFDs are provided by the Iowa Environmental Mesonet site hosted by Iowa State University. AFDs from 2001 to 2017 are downloaded for CONUS (122 of 126 NWS WFOs). Each AFD is parsed for terms contained in the AMS Glossary of Meteorology. Identified terms are extracted from the text and stored in a local database.
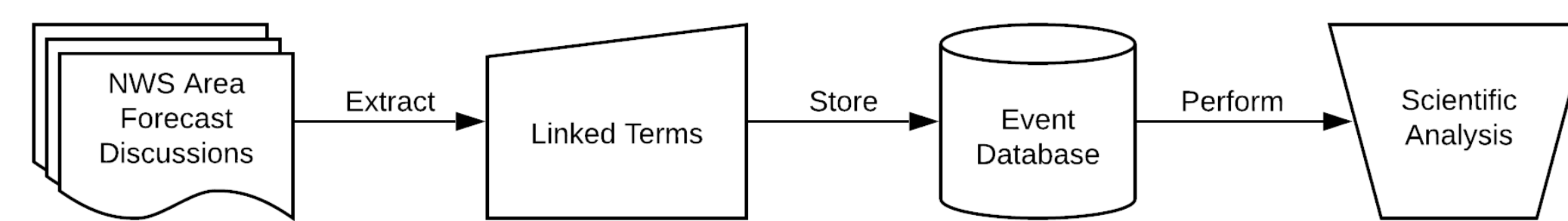

Figure 1 Workflow diagram for extracting information from NWS AFDs.

Data Analysis

More than 2 million unique NWS AFDs are compiled to create an alternative Earth science database more than 50 GB in size. As a means to test feasibility, a 10-year period from 2007 to 2017 is used to extract 20 meteorological terms (Figure 2).

- accumulation, atmospheric river, bomb, bow echo, bright band, derecho, downburst, downslope wind, flooding, fog, freeze, gap wind, hard freeze, hurricane, microburst, sea breeze, smoke, snow, supercell, tornado

General trends in term usage can be used for several scientific applications:

1) Case identification from spatial and temporal anomalies
2) Trend analysis of term usage for climatological application
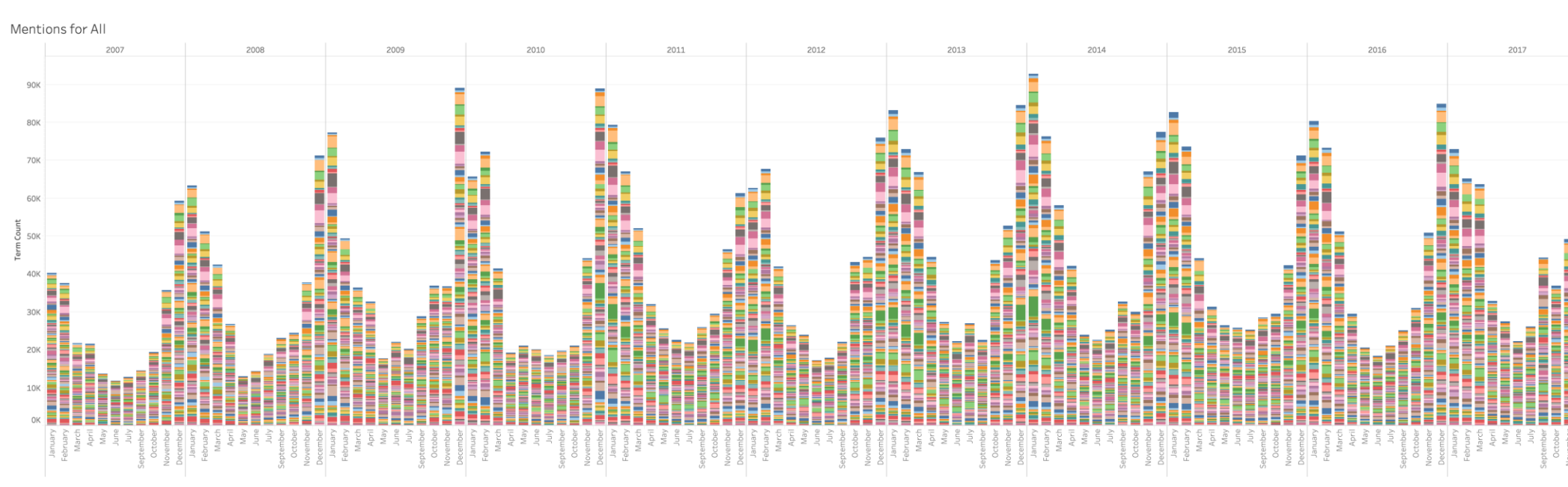3) Evaluation of consistency in communication of scientific information to the general public

Figure 2 Total number of terms extracted for all 122 offices from 2007 – 2017.

Distribution of the selected terms indicate the most common phenomena discussed in the AFDs is snow in the winter and flooding in the summer months. Thus, we leverage these two terms to show the feasibility of the science applications listed above.

Results

We use the term "snow" and evaluate the time series from 2007 – 2017 over the southeastern U.S. (Figure 3). An anomalous peak can be seen in 2015 with more than 4000 extractions of "snow" from AFDs occurring in the month of February.
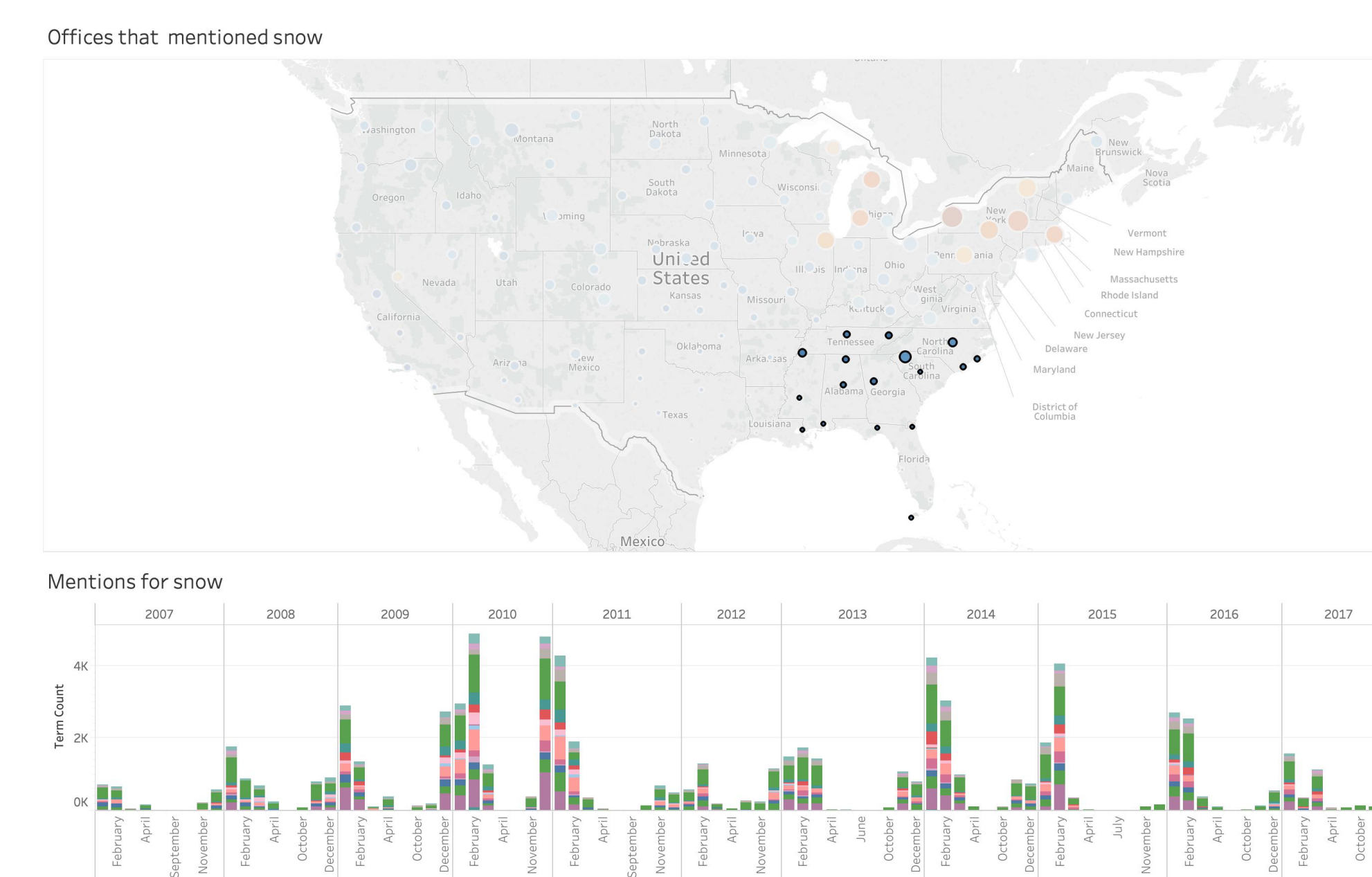

Figure 3 Total extractions for the term "snow" for 2007 – 2017 for the southeastern U.S.

A closer look at term extractions by day for February 2015 reveals two distinct peaks where term usage exceeds 250 counts per day (Figure 4). The peak on the 24th and 25th is associated with a snowstorm across the southeastern U.S. that broke snowfall records in Arkansas, Mississippi, Alabama, Georgia, and the Carolinas.
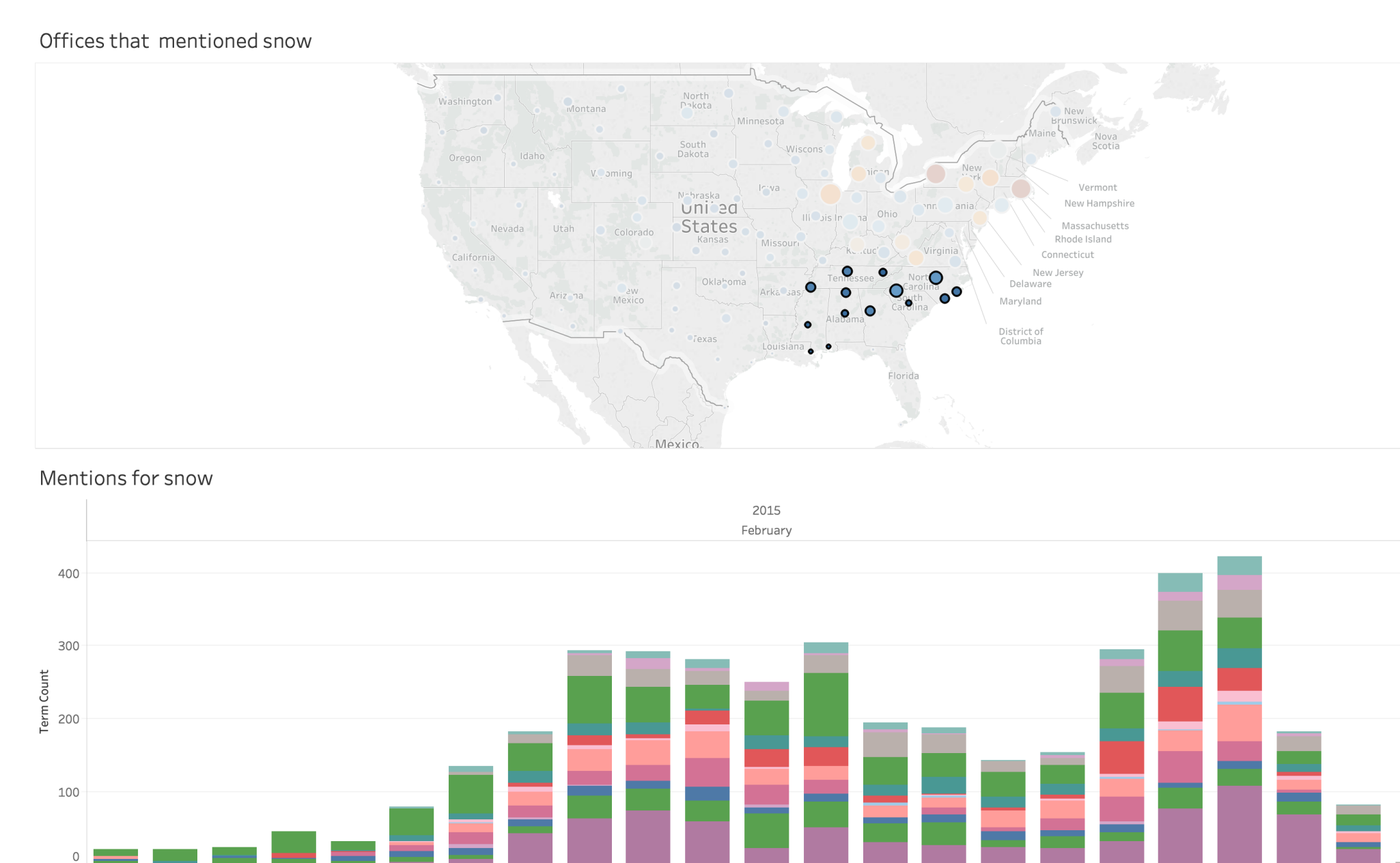

Figure 4 Counts of extractions with "snow" by day for February 2015 for the southeastern U.S.

Another approach useful for extracting meaningful information from the AFDs is to look at term usage as a function of time. With sea levels rising in response to global warming and melting land ice, coastal flooding events are projected to increase in the future. Analysis of "flooding" from offices along the Gulf of Mexico indicates that coastal flooding events are possibly increasing (Figure 5). In 2007, no month exceeds 200 mentions of flooding, whereas each of the years 2015 – 2017 have 7 months exceeding 200 and 2 exceeding 400 counts.
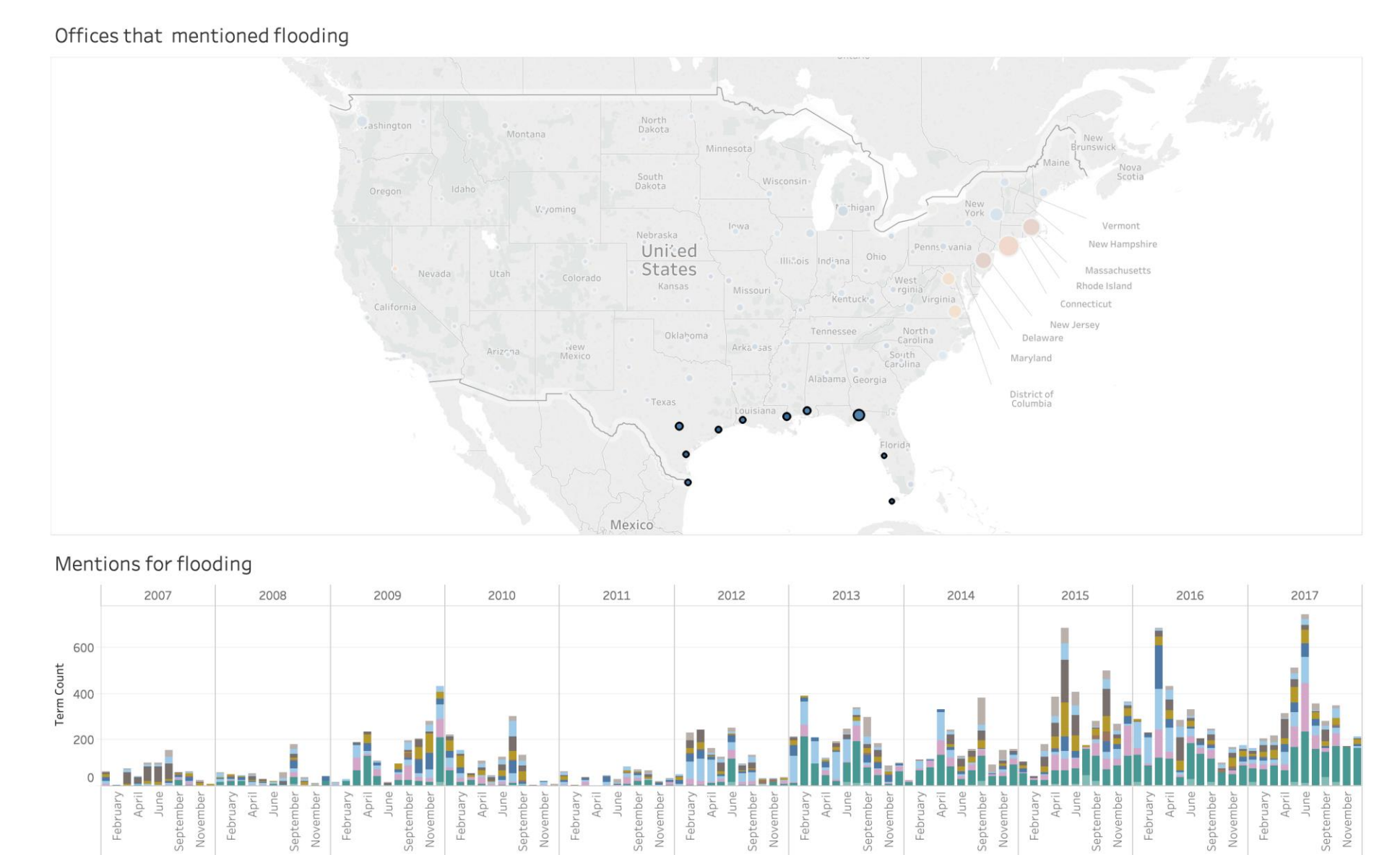

Figure 5 Counts of extractions with "flooding" from 2007 – 2017 for Gulf Coast NWS offices.

Conclusions

We demonstrate the ability of using alternative data sources to develop a database that can be used to perform preliminary scientific analysis. Potential applications of this dataset include: 1) spatial and temporal patterns for case identification; 2) climate studies from extended time-series; and 3) social science studies to improve communication of scientific information.

Contact: brian.freitag@nsstc.uah.edu

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

NASA