# Data Mining Twitter for Augmenting NASA Precipitation Research and Application

NASA/Goddard EARTH SCIENCES DATA and INFORMATION SERVICES CENTER (GES DISC)

**William Teng[1,2], Arif Albayrak[1,2], and George Huffman[1]**

[1]NASA Goddard Space Flight Center; [2]ADNET Systems, Inc.

*Email: William.L.Teng@nasa.gov*

## Motivation

- The Twitter social microblogging database, which recently passed its 10th anniversary, is potentially a rich source of real-time and historical, global information for science applications, beyond the by-now fairly familiar use of Twitter for natural hazards monitoring (e.g., USGS earthquake monitoring's Did You Feel It? http://earthquake.usgs.gov/data/dyfi/).
- We have been exploring the feasibility of extracting from the Twitter data stream useful information for application to NASA precipitation research, in particular, augmenting existing validation programs, with both "passive" and "active" participation by the twitterers.

## Crowd-sourcing/Citizen Science

- Our general crowd-sourcing strategy is to *not* require participants to explicitly "sign up" or install some app to contribute as citizen scientists.
- As much as possible, let routine users of Twitter to continue to do so.
- This is a more robust approach. To effectively crowd-source, a large source of crowd, obviously, is needed. The Twitter data stream is such a source of crowd.
- Twitter "follow" feature is a build-in recursive mechanism for recruiting potential new citizen scientists.
- Sentiment analysis is key concept. See review [1] and as applied to, e.g., politics [2], climate change [3], and tourism [4].

### Passive

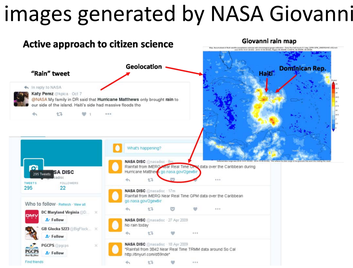
In the passive case, we have experimented with

- Listening to the Twitter stream in real time for "precipitation" and related tweets (in different languages).
- Applying basic filters for exact phrases.
- Extracting location information.
- Mapping the resulting tweet distributions.

### Active

In the active case, we have evaluated different methods of engaging with potential participants, e.g.,

- Replying to "precipitation tweets" with Global Precipitation Measurement (GPM) images generated by NASA Giovanni, centered on the tweet locations



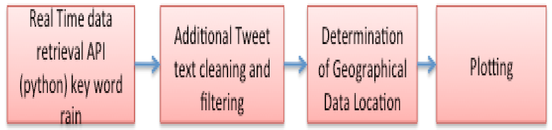

- Filter and extract "precipitation tweet" (Oct. 7, 2016) about Hurricane Matthew.
- Determine geolocation of tweet.
- Use Giovanni to generate GPM rain map.
- "Reply" to "precipitation tweet" by tweeting to @nasadisc, a special Twitter account created for the experiment.

## Potential Science Applications

- Validating satellite precipitation estimates across the full range of weather regimes is challenging, because many regions lack data.
- To improve this situation, the GPM mission has pursued dedicated field campaigns and exchanges with a range of international partners.
- Mining the Twitter stream could augment GPM's validation program.
- Time-varying set of "precipitation tweets" ~ organic network of rain gauges.
- Science rationale for mining "precipitation" and related tweets is to potentially obtain a widespread view of precipitation occurrence.
- Precipitation/no-precipitation boundary is subject to considerable uncertainty.
- "Precipitation tweets" could help constrain this boundary.
- Potential exists for tweets to help tune existing GPM algorithms.
- Twitter stream processing system is not GPM-specific and is applicable to other missions, e.g., Soil Moisture Active Passive (SMAP).

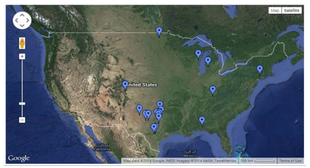
## Architecture of Twitter Stream Processing and Analysis

### Prototype system

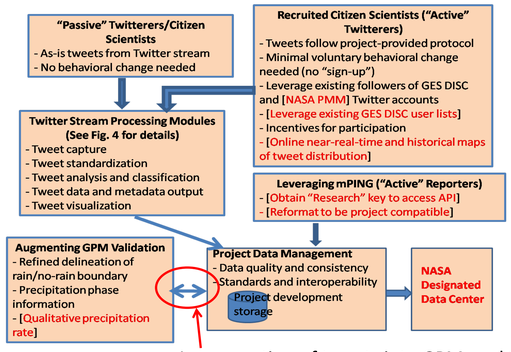



Basic architectural components of prototype system

Map of tweets that was part of a Maryland high school science fair project supported by our prototype.



### Proposed system

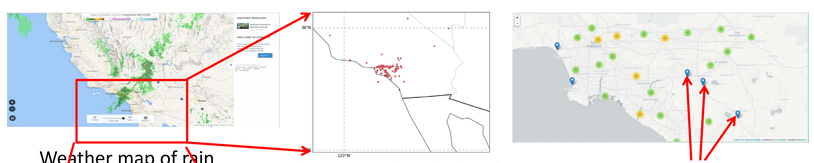



Overall flow of tweets from Twitter stream to data archive of proposed system

Incorporation of tweets into GPM could improve its validation. Conversely, performance of tweets in such improvement could help improve the tweet processing system.

## Preliminary Examples

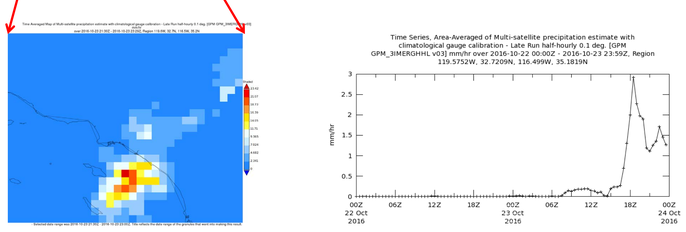
### Rain event in Los Angeles area, CA, October 23, 2016



Weather map of rain event

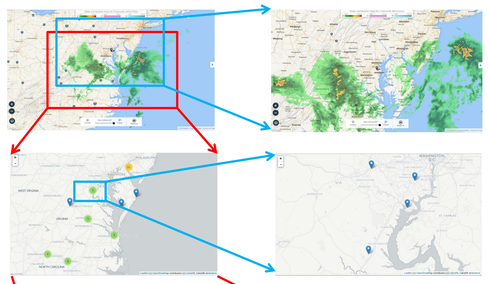
Corresponding distribution of tweets over ~1.5 hour …

… down to individual locations



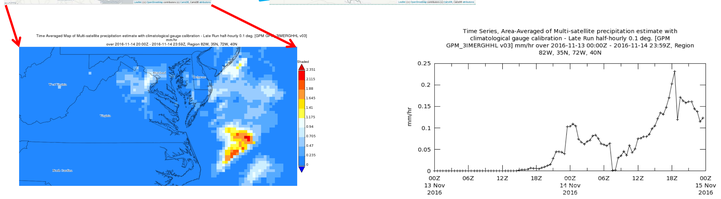

NASA Giovanni visualization of GPM map (left) and time series (right) of the LA area rain event

### Rain event in Delaware-Maryland-Virginia, November 14, 2016



Weather map of rain event at 1520 EST (left) and 1855 EST (right)

Corresponding distribution of tweets over ~3.5 hour …




NASA Giovanni visualization of GPM map (left) and time series (right) of the DE-MD-VA area rain event

### Filtering and extracting tweets: Some numbers

| Los Angeles area | |
|---|---|
| Total # "rain" tweets collected (global) (A) | 63,428 |
| Of (A), # tweets w/ exact geolocation (global) (B) | 1,124 |
| Of (A), # tweets w/ "polygon" geolocation (global) | 965 |
| Of (B), # tweets w/in [-130, 30, 37, -110] (incl. Los Angeles area) (C) | 147 |
| Of (C), # amateur weather station tweets | 30 |
| # tweets w/ phrase, "rain" | 120 |
| # tweets w/ phrase, "rainy" | 12 |
| # tweets w/ phrase, "it is raining" | 15 |

| Delaware-Maryland-Virginia area | |
|---|---|
| Total # "rain" tweets collected (global) (A) | 64,994 |
| Of (A), # tweets w/ exact geolocation (global) (B) | 2,503 |
| Of (A), # tweets w/ "polygon" geolocation (global) | 1,286 |
| Of (B), # tweets w/in [-80, 35, 42, -75] (DE-MD-VA area) (C) | 40 |
| Of (C), # amateur weather station tweets | 25 |
| # tweets w/ phrase, "rain" | 30 |
| # tweets w/ phrase, "rainy" | 5 |
| # tweets w/ phrase, "it is raining" | 1 |
| # tweets w/ phrase, "pouring" | 4 |

## Summary

- Our prototype system has demonstrated the potential for extracting from the Twitter data stream useful information for science applications.
- We will soon begin to implement our proposed system, in a new project under the NASA Citizen Science for Earth Systems Program.
- Potential exists to significantly extend the application realm of Twitter, as a platform for citizen science, beyond natural hazards monitoring to science applications.

References

[1] Liu, B., 2012. Sentiment analysis and opinion mining, San Rafael, CA: Morgan & Claypool Publishers, 167 pp.

[2] Rill, S., D. Reinel, J. Scheidt, and R.V. Zicari, 2014. PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis, Knowledge-Based Systems, *69*, 24-33, doi: 10.1016/j.knosys.2014.05.008.

[3] Kirilenko, A.P., T. Molodtsova, and S.O. Stepchenkova, 2015. People as sensors: Mass media and local temperature influence climate change discussion on Twitter, Global Environmental Change, *30*, 92-100, doi: 10.1016/j.gloenvcha.2014.11.003.

[4] Kirilenko, A.P. and S.O. Stepchenkova, 2017. Sochi 2014 Olympics on Twitter: Perspectives of hosts and guests, Tourism Management, *in* review.