A Data Mining Framework for Analyzing

Geospatial-Temporal Data

Raunak Sarbajna

Department of Computer Science Lamar University

Beaumont, Texas, USA Email: [rsarbajna@lamar.edu](mailto:rsarbajna@lamar.edu)

Sujing Wang

Department of Computer Science Lamar University

Beaumont, Texas, USA Email: [sujing.wang@lamar.edu](mailto:sujing.wang@lamar.edu)

***Abstract*—. Change analysis and automatic storytelling are essential techniques in understanding patterns and trends in multifaceted, time-series geospatial-temporal data. In this paper, we introduce a new data mining framework for analyzing the chang patterns of spatio-temporal data. It includes change analysis techniques and automatic storytelling methodology for spatio-temporal data. We evaluate the effectiveness of our framework through case studies involving Twitter emotion data and North American Drought data. The experimental results show that our framework can discover interesting change patterns and useful information from spatial-temporal data.**

***Index Terms*—change analysis, storytelling, sentiment analysis,**

**polygon, spatio-temporal data.**

I. INTRODUCTION

Analyzing change in spatial-temporal data is critical for many applications including developing early warning systems that monitor environmental conditions, detecting political unrest and crime monitoring. Change analysis models are essential in understanding larger patterns and trends in multifaceted, time-series geographic data. The purpose of this study is to detect spatiotemporal changes within sequential (time-series) geospatial-temporal data.

Modern technology digitizes wide sources of information

constantly, with hour after hour of data being stored. Most of it is unprocessed raw data. This is especially true with remote sensing networks and resultant geospatial-temporal data. Most of this data has meta- information associated with it which can reveal patterns or trend if properly processed nd analyzed. Performing basic change analysis on this data gives us insights that would otherwise go unrecognized.

A common understanding is that most big data available

today is either archival, media or web scrapes [25]. However, a large source of that data is actually from Geographic Information Systems (GIS), and the tools available to interpret

this kind of data are lacking [26]. The multiplicity of APIs have standardized access and structuring, but they limit much of the meta-data associated with them. Most publicly available (i.e. non-governmental) ’big data’ sources with spatial components revolve around data scraped from mobile software platforms, including twitter, Instagram, Snapchat and reviews on mapping apps [26]. The recent API changes by Google Maps, for instance, aptly show that data content, and thus its meaning, is subject to regulation that is outside the control of researchers. The goal of this research project is to detect and analyze how the patterns of features change over time and space in spatio-temporal data and automatically story-telling based on the time-serials of spatio-temporal data.

Our approach provides a change monitoring framework which creates a change graph that captures the changes in spatial clusters and a change summarization framework that creates specific change summaries based on the change graph based on the change story types.

Our research contributions are summarized as follows:

1) A novel change analysis framework for spatio-temporal data

2) New change predicates that are data agnostic and can work on a large spectrum of data

3) A new measure of interestingness to aid in generating automated storytelling based on the change analysis results

The rest of the paper is structured as follows. Section 2 reviews previous literature on the subject and discusses related work. Section 3 introduces our data mining framework and lays out the methodology in detail. Section 4 evaluates the framework with case studies on drought datasets and twitter data.

Section 5 provides a conclusion and discusses potential future expansions to the framework.

II. RELATED WORK

Storytelling techniques are effective summarization method to succinctly organize extensive information. Traditional storytelling has been mostly successful on news articles, blogs, as well as structured databases. However, traditional storytelling techniques tend to perform poorly on social media content, such as Twitter, where text lacks proper form and function [11]. Moreover, the ability to support dynamic storylines as they evolve is critical to modelling fast moving social media streams such as Twitter. Dos Santos et al. [21] introduced a set of methods to automatically derive stories over linked entities in tweets. They model a story as a graph of entities propagating through spatial regions in a temporal sequence, and controls search space complexity by suggesting regions of exploration. They developed algorithms to conduct storytelling to model tweets over space and time, reasoning over spatio- temporal features, and devise spatio-temporal storylines based on connectivity strength.

Kumar et al. [14] proposed an efficient storytelling imple- mentation that embeds the CARTwheels [15] redescription mining algorithm which utilizes induced classification trees to model redescriptions in an A\* search procedure, using the CARTwheels to supply next move operators on search branches to the A\* search procedure. Vocht et al. [15] pro- posed the implementation of an optimized algorithm con- trolling the pathfinding process to obtain more homogeneous search domain and retrieve more links between adjacent hops in each path to improve the semantic relatedness of concepts mentioned in a story by increasing the relevance of links between nodes through additional domain delineation and refinement steps. Chen et al. [20] proposed a multimodal imitation learning via generative adversarial networks (MIL- GAN) method to directly model users’ interests as reflected by various data by imitating users’ demonstrated storylines. MIL- GAN model is designed to learn the reward patterns given user-provided storylines and then applies the learned policy to unseen data. Santos et al. [21] combined storytelling and Spatio-logical Inference (SLI) to generate rules of interaction among entities and measure how well they forecast a real- world event.

Hossain et al [13] introduced Google Fusion Tables(GFT) that offers collaborative data management in the cloud for data scientists to enable the integration of increasingly complex geospatial data to support storytelling. The paper focused on introduction of overview of map processing in GFT, the architecture overview of GFT, and how to scale to large datasets, massive and complex polygon datasets. GFT provides a useful tool for storytelling through interactive maps.

Kumar et al. [14] formulated storytelling as a generalization of redescription mining. Stories are defined as chains of re- descriptions. They proposed an efficient storytelling algorithm as A\* search around the outputs of a CARTwhells redescrip- tion mining algorithm. The efficiency and scalability of the proposed algorithm were evaluated by three application case studies: word overlaps in large English dictionaries, exploring connections between gene sets in a bioinformatics data set, and relating publications in the PubMed index of abstracts.

Hossain et al. [19] proposed an approach to automatically construct stories between entities in large document collections that can help from directed chains of relationships, with

*N Distance*(*C entroid*(*c*)*,C entroid*(*m*))

support for co-referencing, evidence marshaling, and imposing

syntactic constrains on the story generation process. A new optimization techniques based on concept lattice mining is used to rapidly construct stories on massive datasets.

Chen et al. [20] introduced an approach, multimodal imitation learning via generative adversarial networks (MIL-GAN) for generating storyline on unseen data. It can directly model users’ interests as reflected by various data. This approach is used to learn the reward patterns given user-provided storylines and then applies the learned policy to unseen data.

Santos et al. [21] introduced three methods of association analysis, Distance-based Byesian Inference, Spatial Associa- tion Index, and Spatio-logical inference, to capture relatedness among real-world events in high data volumes, and to model similar events that are described disparately under high data variability. It takes as input a set of geotemporally-encoded text streams about violent events called “storylines”. This study demonstrated that spatio-temporal storytelling is able to capture important associations among violent events reported in social media and traditional datasets.

III. METHODOLOGY

*A. Change Analysis Predicates*

We begin by creating our change predicates. There are three change functions that are necessary for complete change predicates, we define them as follows:

*• Agreement*(*c, m*) = (*area*(*c ∩ m*))*/*(*area*(*c ∪ m*))

*• C ontainment*(*c, m*) = (*area*(*c ∩ m*))*/*(*area*(*m*))

*• Overlap*(*s, p*) = *area*(*s ∩* (*p*1 *∩ . . . ∩ pm* ))*/area*(*s*)

Here, *c* is a spatial cluster polygon from the batch timed *ti* and *m* is another spatial cluster polygon from a batch timed *ti*+*x* , with *x* indicating a time increase. We denote cluster

intersection as *∩* and union as *∪*. *Area*(*x*) denotes the area

covered by the polygon *x*, upto a desired level of accuracy.

The change function *Agreement*(*c, m*) is a measure of similarity. *C ontainment*(*c, m*) measures whether individual polygons coincide. To measure concurrence between one poly- gon and a set of polygons, we use the *Overlap*(*s, p*) functions. Here *p*1 *, ..., pm* are spatial clusters from a batch *b*, and *s* is a cluster from a different batch *bl* . Other basic functions include *C entroid*(*x*), which indicates the Centorid of a polygon as defined as:

"£*N −*1

7) *Shif ting ↔ i*=1

Implementations of these change predicates can be

*N*

found in the Section IV.

*B. Polygon Generation from Point Data Sources*

We improve the change analysis framework called Aconcagua [6]. The system expects an input of emotion polygons annotated with emotion assessment scores, with +1 representing a very high positive emotion and -1 representing a very high negative emotion. While this method does lead to inconsistencies in locations due to geo-referencing inaccuracies, we find that the 8 10m precision [8] works well within city limits.

There are several ways of using the numerous point data we

have obtained from this step into an actual polygonal map:

1) Creating closed contour lines for contour lines that lie on the boundary of the observation area.

2) Creating a convex hull from points with similar scores. We elaborate on creating convex hulls in the following

section.

*C. Change Analysis for Polygonal Map Data Sources*

Our approach use three primary set operations: union, intersection and erase. We calculate the area of each individual polygon within each map layer. We then execute a union operation and calculate area. The union layer now contains the original areas of both layers and the areas of the overlapping polygons - we now need to query them properly to prepare for calculating the change percentage and tabulating intersection. To outline the polygon, we examine several different methods:

1) We find features common to either of the layers but not both, essentially performing a symmetrical difference.

2) We erase the larger of the polygons from the smaller, thus retaining only the growth, and do vice-verse for shrinking.

3) We perform simple intersection and then invert selection to get changed regions.

Our approach then combined several techniques:

1) **Data Pre-Processing:** This involves curation of datasets with obvious geo-referencing errors. This would prefer- able be done by minimizing the root mean square error. We initially.

*Cx* = 1

6*A*

*Cy* = 1

6*A*

*i*=0 (*xi* + *xi*+1 )(*xi yi*+1 *− xi*+1 *yi* )

*i*=0 (*yi* + *yi*+1 )(*xi yi*+1 *− xi*+1 *yi* )

"£*N −*1

2) **Parametrization of polygons:** Calculate shape and area parameters for each individual polygon with each map

where *xN* = *x*0 and A = the area of the polygon. We

also use a simple Manhattan distance function *Distance*(*c, m*)

which is tuned to the required georeferenced accuracy. We use all these functions to defined these seven change predicates:

1) *S − C ontinuing*(*c, m*) *↔ Agreement*(*c, m*) *≥* 0*.*8

2) *B − C ontinuing*(*c, b*) *↔ Overlap*(*c, b*) *≥* 0*.*8

3) *Growing*(*c, m*) *↔ C ontainment*(*c, m*) *≥* 0*.*9

4) *Shrinking*(*c, m*) *↔ Growing*(*m, c*)

5) *Disappearing*(*c*) *↔ ∃i*(*belong − to*(*c, i*)

6) *N ovel*(*c*) *↔ ∃i*(*belong − to*(*c, i*)*and*(*i* = 1*ornot*(*B −*

*C ontinuing*(*c, i −* 1))

layer.

3) **Analysis through Symmetrical Difference:** Extract fea- tures common to either of the layers.

4) **Polygon Union Computation:** Union sequential layers to contain the original areas of both layers and the areas of overlapping polygons.

5) **Polygon Erase Operation:** Erase larger polygons from smaller (or vice-versa) for detection of growth/shrinking.

6) **Polygon Intersection/Invert:** Selecting and then la- belling the changed regions.

The framework can be found in Figure 1.

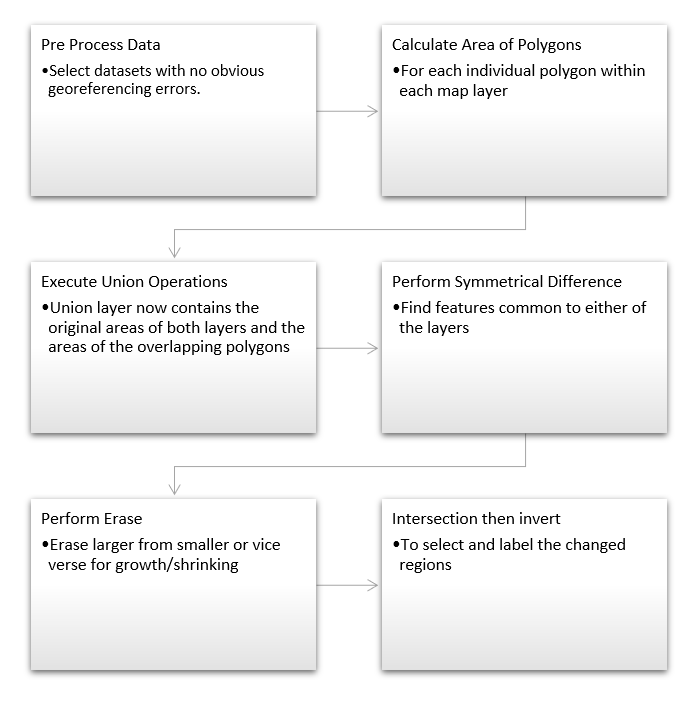


Fig. 1. The Framework Architecture

*D. Storytelling Techniques for Geospatial Data*

In order to tell a coherent spatio-temporal data story from the change analysis output, We propose using an interestingness function for this task. .[11]

We choose a set of change polygons SCP. It not only contains those polygons but also their associated characteristics. For example, a SCP could contain a set of spatial clusters represented by polygons, their average drought score, total area of each polygon, centroid coordinates of each polygon, and other summaries for each spatial cluster (polygon). We define the function as:

*f* : *SC P →* [0 *, ∞*)

We define a threshold *ω*, which ensures that a narrative will only be generated an object *p ∈ SC P* such that *f* (*p*) *≥ ω*.

Sample parameters for *ω* include

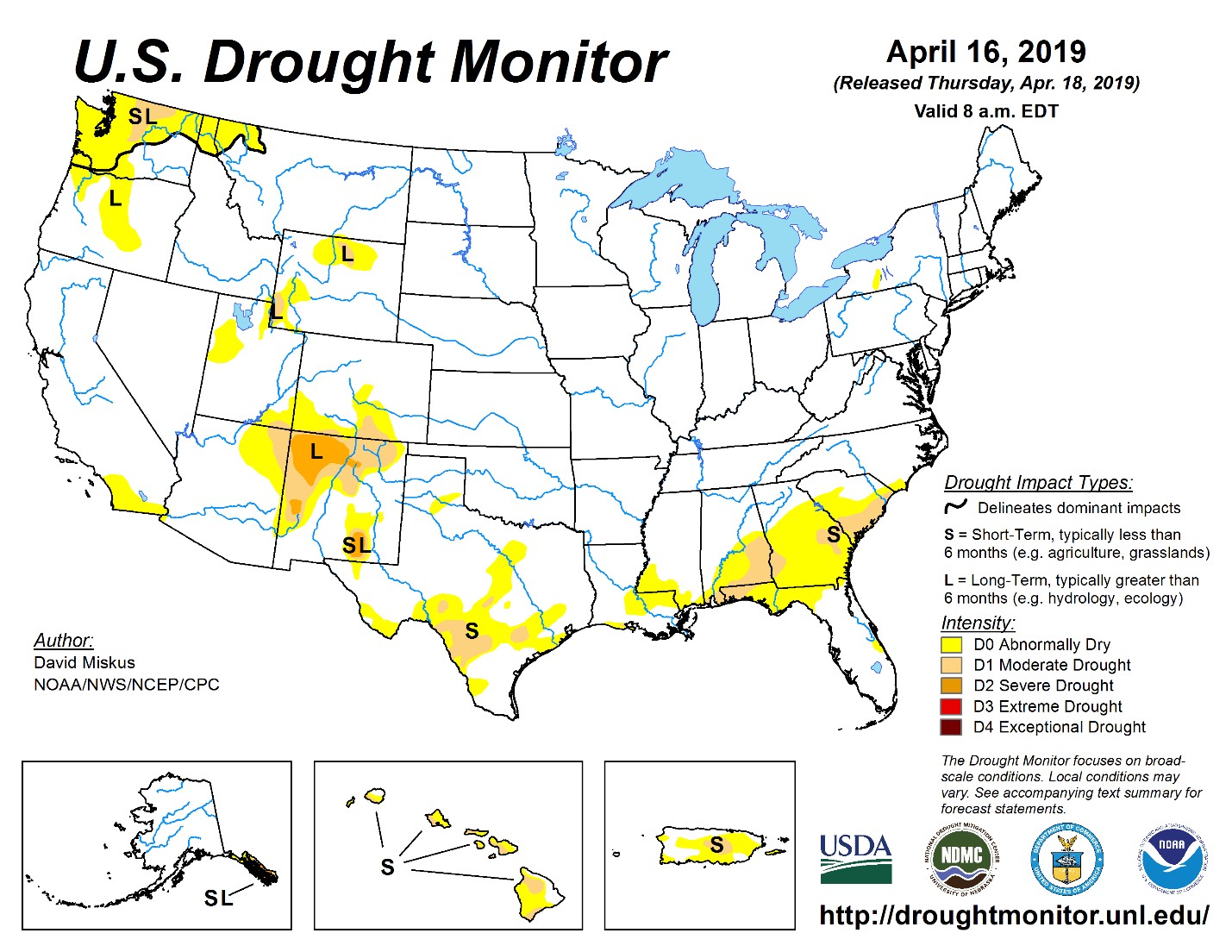


Fig. 2. United Sates Drought Monitor Data

*• M ax* (*P ercentage C hange in P olygon*)

*•* Largest shift in polygon centroids

The threshold parameters need to be finely tuned so as to not exclude those polygons who fall through exceptions. Once we have a suitable selection of polygons and chose a threshold value, we can create a summary narrative based on that.

IV. CASE STUDY

*A Data Sets.*

*We use two datasets for our case study, drought dataset from the North American Portal [??]and Twitter posts from the United States [9]. The Drought data is defined by the USDA as shown in Figure 2. The spatial reference for the Drought shapefiles is summarized in Table 1.*

The Twitter data contains the timestamp, longitude, latitude and the text of each tweet, which are then processed and tokenized. The spatial reference for the twitter data can be seen in

Table 2.

b. Dataset Preprocessing

Our initial approach to this problem was to store all shape- files of XX data in a postgres database with a GIS addon and perform operations in python. We used psycopg2 and osgeo libraries to import, process and visualize maps. However, this leads to many problems with interconversions between georeferencing schemes, while converting from WKT geometry to PostGIS geography.

We start with basic point data, which contains latitude/- longitude, along with metadata identifying value of interest such asdrought level or emotion value. We insert the contents of the shapefile into a PostGreSQL database using the shp2pgsql toolkit that comes along with the PostGIS extension. The results can be seen in Figure 3.

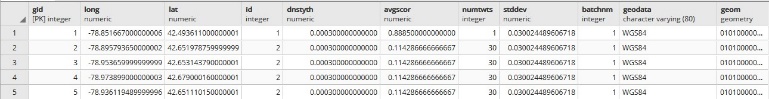


Fig. 3. Twitter shapefile data inserted into Postgres db

not points or lines.

1) Our query for this purpose was:

INSERT INTO p u b l i c . j u n e 1 p o l y ( a v g s c o r , n u m t w t s , g e o d a t a , i d , b a t c h n m , geom )

( SELECT d . a v g s c o r , d . n u m t w t s , d . g e o d a t a , d . i d , d . b a t c h n m , S T C o n v e x H u l l

( S T C o l l e c t ( d . geom ) )

FROM p u b l i c .”2014 *−*06 *−*01 ” AS d

GROUP BY ( d . i d , d . a v g s c o r , d . n u m t w t s ,

d . g e o d a t a , d . b a t c h n m )

TABLE I

LAYER SPECIFICATION FOR DROUGHT DATA

Name Type Width Precision long Real 24 15

lat Real 24 15 id Integer 9 0 dnstyTh Real 24 15 avgScor Real 24 15 numTwts Integer 9 0 stdDev Real 24 15 batchNm Integer 9 0 geoData String 80 0

HAVING S T G e o m e t r y T y p e ( S T C o n v e x H u l l

( S T C o l l e c t ( d . geom ) ) ) = ’ S T P o l y g o n ’ )

2) Then we insert the centroid of each polygon into the table using the query:

UPDATE p u b l i c . j u n e 1 p o l y

SET c e n t r o i d = S T C e n t r o i d ( geom )

3) We repeat this process for every shapefile needed.

*B. Drought Data Sources*

For three change predicates that we discussed in Section III:

1) To detect polygons that are increasing in size, we check for similar IDs, intersection and then the rate of overlap. We initially check whether the polygons intersect at all before querying for amount of overlap. This leads to faster processing as it discards the many combinations where the polygons don’t touch each other. Our query is structured as:

SELECT DISTINCT j 2 . *∗*

FROM p u b l i c . j u n e 1 p o l y j 1 ,

p u b l i c . j u n e 2 p o l y j 2

TABLE II

LAYER SPECIFICATION FOR TWITTER DATA

Name Type Width Precision

FIPS ADMIN String 4 0

GMI ADMIN String 7 0

ADMIN NAME String 42 0

FIPS CNTRY String 2 0

GMI CNTRY String 3 0

CNTRY NAME String 40 0

POP ADMIN Integer 9 0

TYPE ENG String 26 0

TYPE LOC String 50 0

SQKM Real 16 2

SQMI Real 16 2

COLOR MAP String 2 0

WHERE ST INTERSECTS ( j 1 . geom , j 2 . geom ) AND

( ST AREA ( ST INTERSECTION ( j 2 . geom , j 1 ) )

/ s t a r e a ( j 2 . geom ) ) *>* . 8 5

2) To detect polygons that are shrinking in size, we check for similar IDs, and lower rates of overlap. This can be modified based on need.

SELECT DISTINCT j 2 . *∗*

FROM p u b l i c . j u n e 1 p o l y j 1 ,

p u b l i c . j u n e 2 p o l y j 2

WHERE ST INTERSECTS ( j 1 . geom , j 2 . geom ) AND

( ST AREA ( ST INTERSECTION ( j 2 . geom , j 1 ) )

/ s t a r e a ( j 2 . geom ) ) *<* . 2 5

3) To detect polygons that have shifted, we compare their centroids and check if they have moved over 75km. Remember these polygons are created through a convex hull of points, which cannot ensure the centroid will lie within the polygon itself. Which is why we are taking a sufficiently large bounding value for the polygon.

SELECT S T D i s t a n c e S p h e r o i d

( j 1 . c e n t r o i d , j 2 . c e n t r o i d ,

’ SPHEROID [ ”WGS 8 4 ” , 6 3 7 8 1 3 7 , 2 9 8 . 2 5 ] ’ ) , j 1 . i d FROM p u b l i c . j u n e 1 p o l y j 1 ,

p u b l i c . j u n e 2 p o l y j 2

WHERE j 1 . i d = j 2 . i d AND

S T D i s t a n c e S p h e r o i d ( j 1 . c e n t r o i d , j 2 . c e n t r o i d ,

’ SPHEROID [ ”WGS 8 4 ” , 6 3 7 8 1 3 7 , 2 9 8 . 2 5 ] ’ ) *>* 7 5 0 0 0 ;

Figure 5 shows the generated maps based on the change predicate.

*C. Emotion Spatial Clusters from Twitter*

The tool used for creating the emotion clusters of Twitter data was the K2 framework [7].

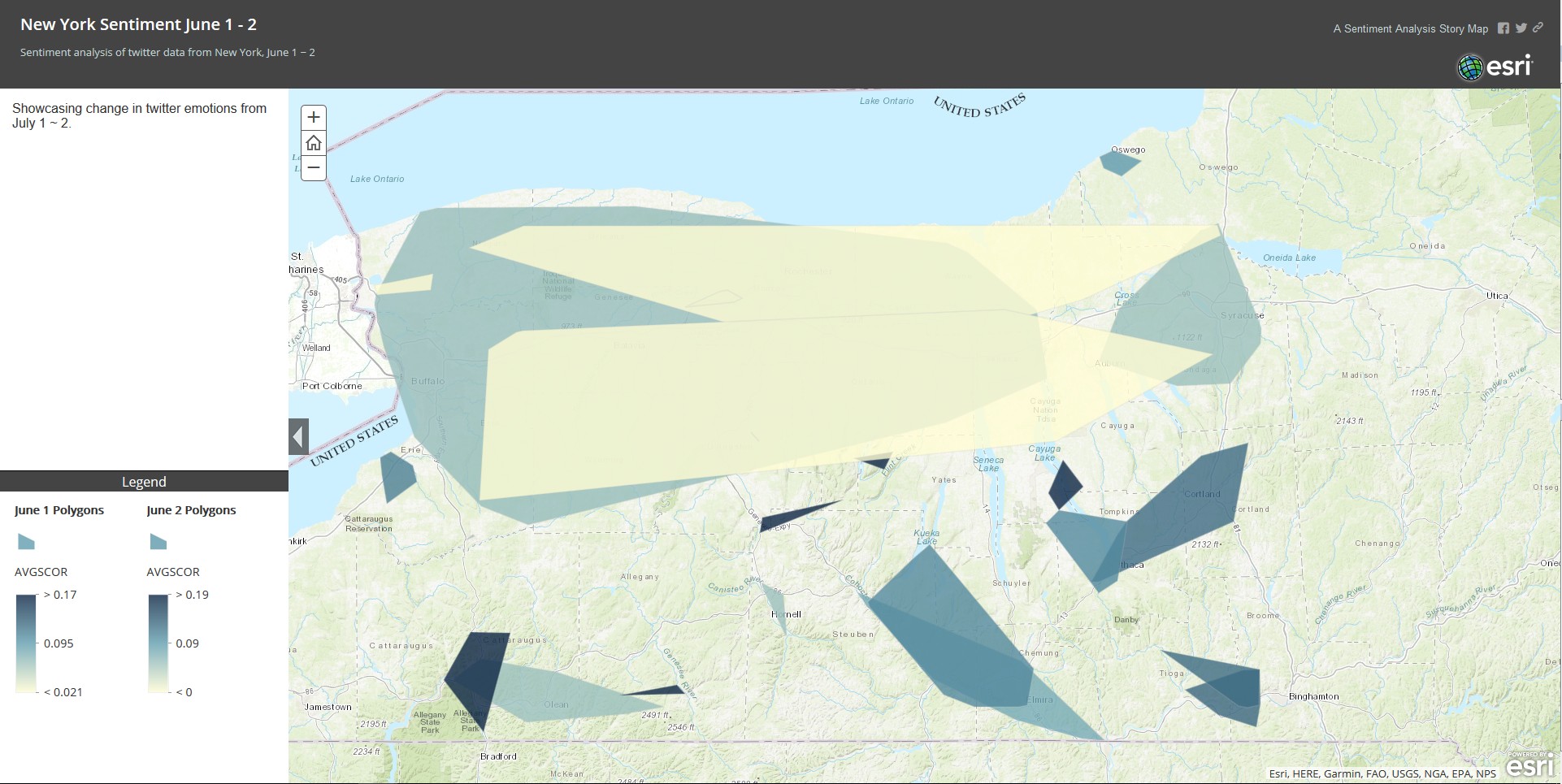


Fig. 5. Polygons generated from Sentiment analysis

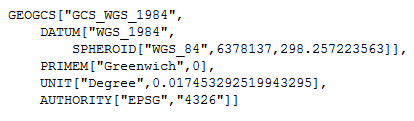


Fig. 6. Spatial Reference for Drought data

According to K2, the best package for the emotion score was the Valence Aware Dictionary and Sentiment Reasoner (VADER) system [10], whose analyser parses the tokenized text and checks within a lexicon for words with strong sentiment. The final score is created from the weighted average of all sentimental words and lies within the range [-1, 1], as is required for our Aconcagua implementation.

The original polygons for the drought datasets, before our analysisare shown in Figure 7.

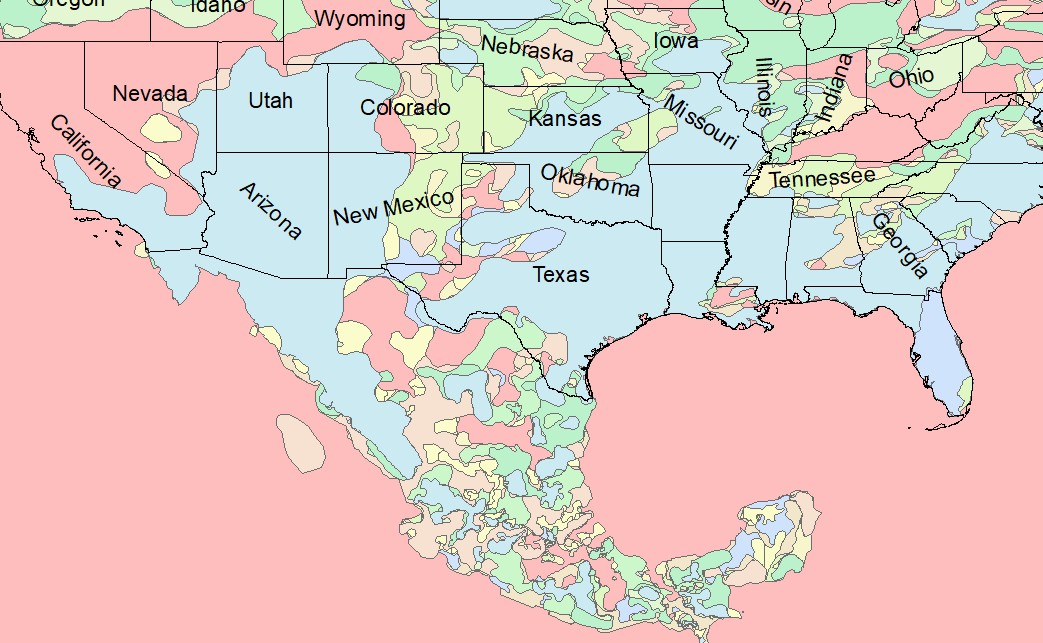


Fig. 7. Continental USA Original Drought polygon Distribution for 2017

We focus on two specific regions for this case study. First, figure

8 shows the regions where areas of drought grew in Cali- fornia after the wildfires in 2017. We notice patterns of increasing drought surrounding the regions that were burnt down, with especially large period upstream from rivers. Our procedure works well dealing with the large number of small polygons in the region that occur due to isolated wildfires.

However, we had difficult dealing with convex polygons that intersected with each other multiple times.

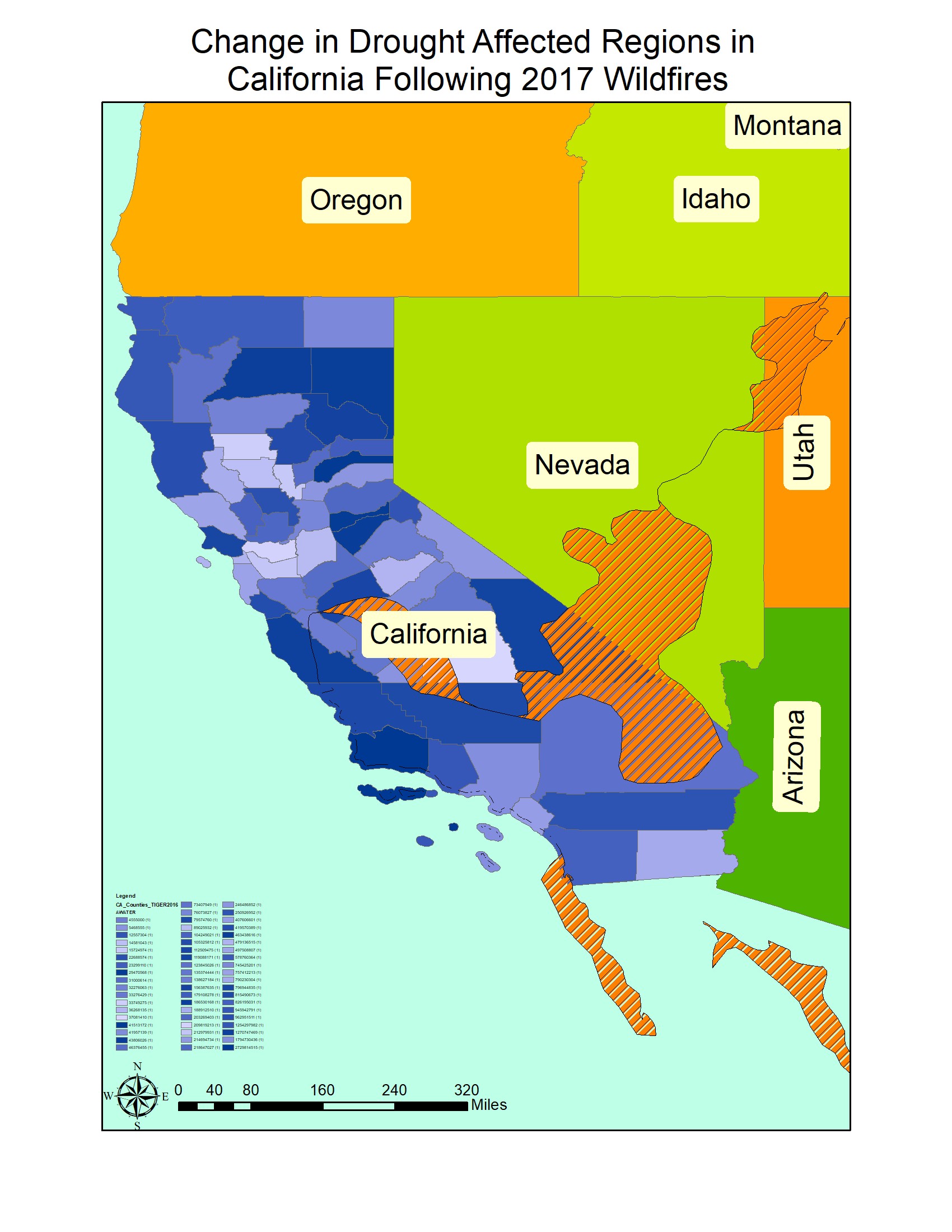


Fig. 8. Change in drought affected regions in California following 2017 wildfires.

Next, we inspect the region in Texas after Hurricane Harvey. Figure 9 shows the regions where drought affected regions increased following the disaster and Figure 10 shows the regions decreased.

We observe that drought affected regions decrease at a high rate around the South Eastern Texas and Louisiana region. However, there is no clear relation between drought prone regions and river basins close to the coast. We believe that high amount industrial regions create a micro climate that affects the water content entering the soil.

According to our results, the drought prone regions in- creased substantially around the West Texas region both during and after Hurricane Harvey, with a larger increase as winter came around. However, this could be an artefact of the fact that the original dataset had very few polygons in the Texas area, which leads to broader conclusions. Our technique is still severely dependent on the resolution of the input imagery.

The results from the analyzing the twitter emotion maps can be seen in Figure 11.

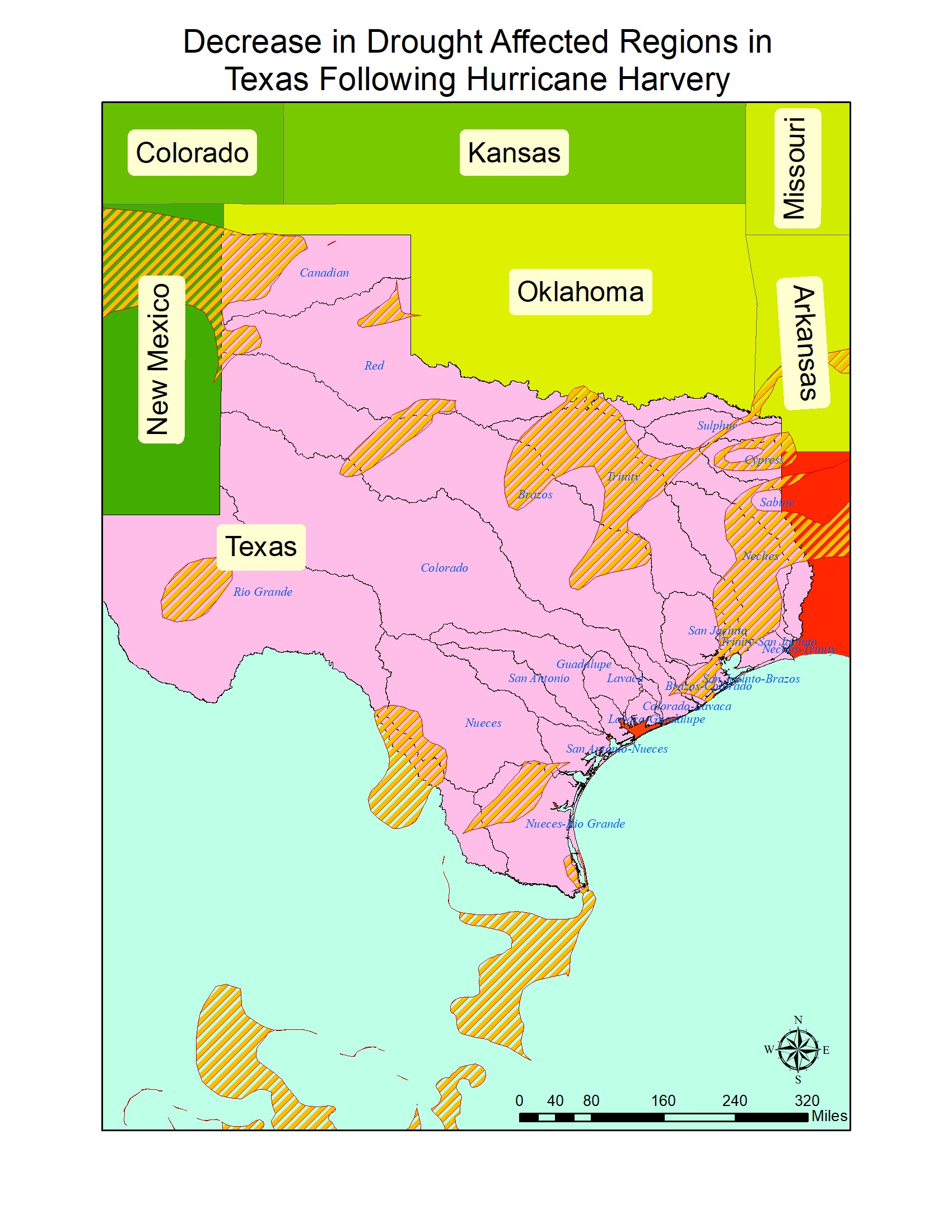


Fig. 9. Decrease in Drought Affected Regions in Texas following Hurricane

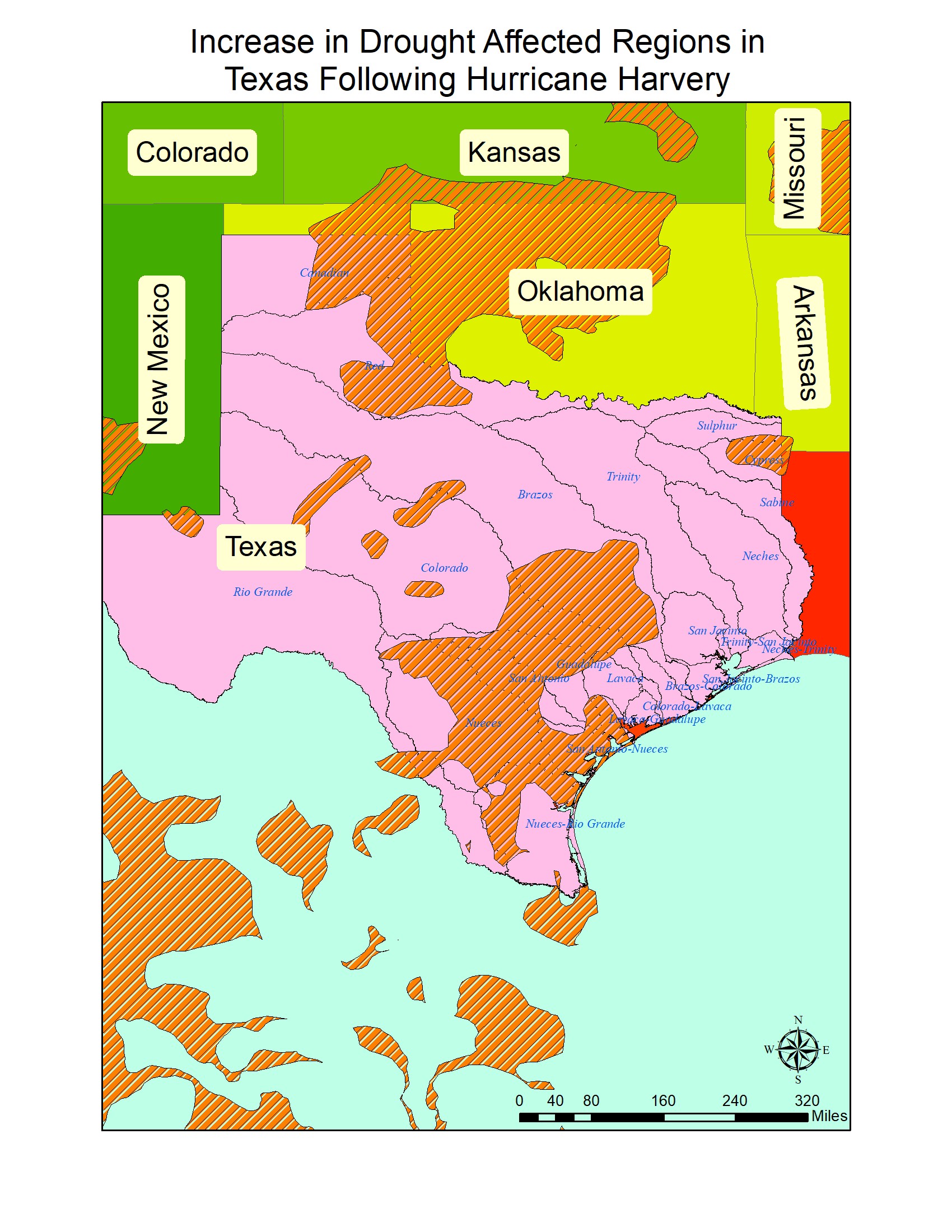
Harvey

V. CONCLUSION AND FUTURE WORK

Our experimental studies show that our change detection and analysis framework can successfully detect changes in spatiotemporal datasets. our change predicates can work on a variety of data sources, including both polygonal and point datasets. Ourframe- work is able to handle both geographically large and small map sources without any significant geo-referencing distortions being introduced.

There is a lot work left in dealing with irregularly shaped convex polygons that appear in poorly georeferenced real world data. We believe this is because the sensitivity of the geographic operations in PostGIS or ArcGIS. We are currently working on extending our framework to deal with similar cases.

Fig. 10. Increase in Drought Affected Regions in Texas following Hurricane



Harvey

ACKNOWLEDGMENT

This research is supported by Lamar University Research

Enhancement Grant.

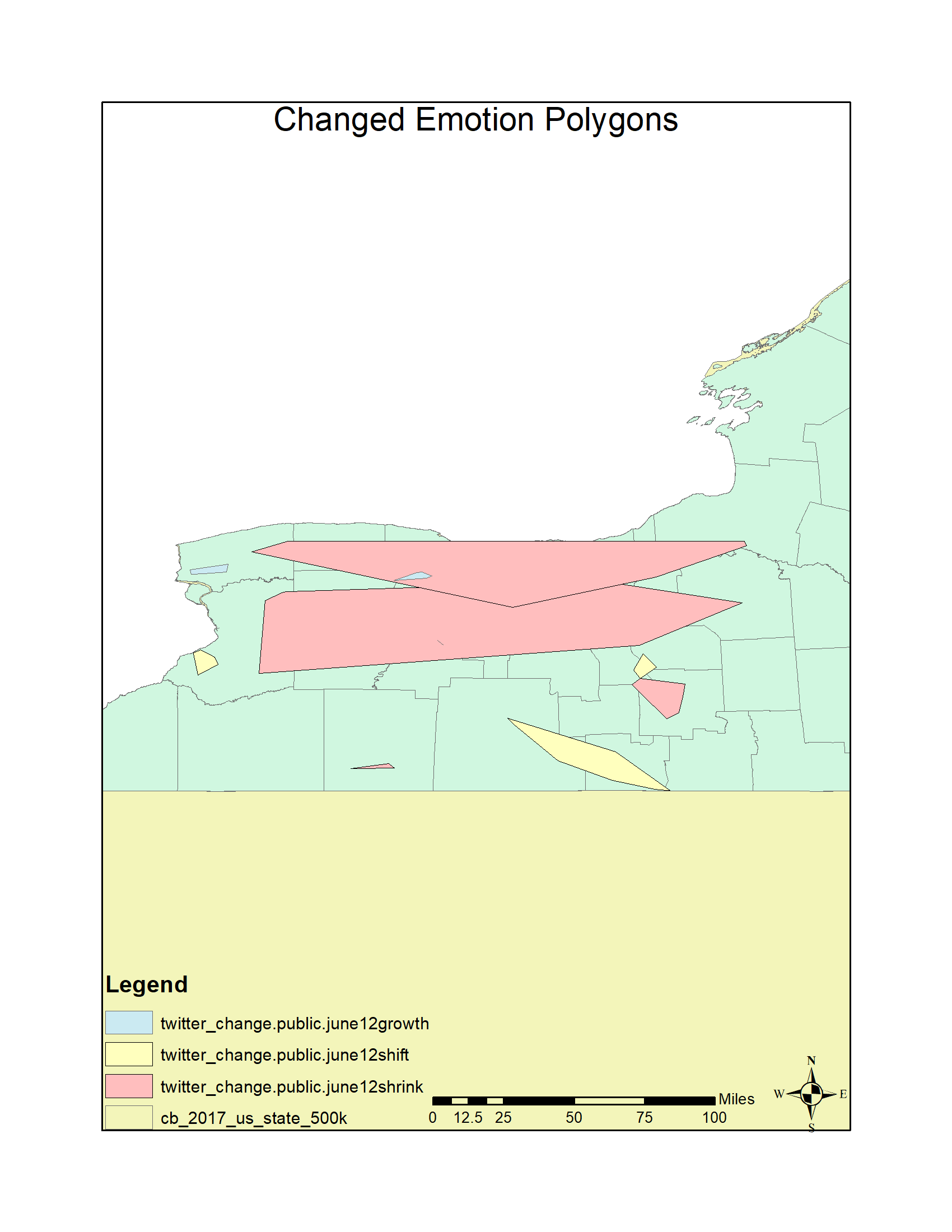


Fig. 11. Change in emotion Polygons in New York State

REFERENCES

[1] J. Im and J. Jensen. 2005. A change detection model based on neighborhood correlation image analysis and decision tree classi- fication. Remote Sensing of Environment 99, 3 (2005), 326–340. [DOI:http://dx.doi.org/10.1016/j.rse.2005.09.008](http://dx.doi.org/10.1016/j.rse.2005.09.008)

[2] Suming Jin, Limin Yang, Patrick Danielson, Collin Homer, Joyce Fry, and George Xian. 2013. A comprehensive change detection method for updating the National Land Cover Database to circa

2011. Remote Sensing of Environment 132 (May 2013), 159–175. [DOI:http://dx.doi.org/10.1016/j.rse.2013.01.012](http://dx.doi.org/10.1016/j.rse.2013.01.012)

[3] D. Lu, P. Mausel, E. Brond´ızio, and E. Moran. 2004. Change detection techniques. International Journal of Remote Sensing 25, 12 (June 2004),

2365–2401. [DOI:http://dx.doi.org/10.1080/0143116031000139863](http://dx.doi.org/10.1080/0143116031000139863)

[4] Merrill K. Ridd and Jiajun Liu. 1998. A Comparison of Four Algorithms for Change Detection in an Urban Environment. Remote Sensing of En- vironment 63, 2 (1998), 95–100. [DOI:http://dx.doi.org/10.1016/s0034-](http://dx.doi.org/10.1016/s0034-)

4257(97)00112-0

[5] S. Minu and Amba Shetty. 2015. A Comparative Study of Image Change

Detection Algorithms in MATLAB. Aquatic Procedia 4 (March 2015),

1366–1373. [DOI:http://dx.doi.org/10.1016/j.aqpro.2015.02.177](http://dx.doi.org/10.1016/j.aqpro.2015.02.177)

[6] K. Elgarroussi, S. Wang, R. Banerjee, and C. F. Eick, “Aconcagua: A Novel Spatiotemporal Emotion Change Analysis Framework,” in Proceedings of the 2Nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, New York, NY, USA, 2018, pp. 54–61.

[7] R. Banerjee, K. Elgarroussi, S. Wang, A. Talari, Y. Zhang, and C. F. Eick, “K2: A Novel Data Analysis Framework to Understand US Emotions in Space and Time,” Int. J. Semantic Computing, vol. 13, no. 01, pp.

111–133, Mar. 2019.

[8] Geomenke, “How Accurate is the GPS on my Smart Phone? (Part 2),” Community Health Maps, 07-Jul-2014. .

[9] W. Z.-M.-L. I. for the S. S. Person, “Geotagged Twitter posts from the

United States: A tweet collection to investigate representativeness,” Jan.

2016.

[10] C. J. Hutto, VADER Sentiment Analysis. VADER (Valence Aware Dic- tionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social

.. 2019.

[11] X. Huang, A. An, and N. Cercone, “Comparison of Interestingness Func- tions for Learning Web Usage Patterns,” in Proceedings of the Eleventh International Conference on Information and Knowledge Management, New York, NY, USA, 2002, pp. 617–620.

[12] W. Wang, Lu Chen, K. Thirunarayan, and Amit P. Sheth, Harnessing Twitter ”Big Data” for Automatic Emotion Identifcation, in Proc. of the IEEE 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing.

[13] R. F. Dos Santos, S. Shah, F. Chen, A. Boedi-hardjo, P. Butler, C.T. Lu and N. Ramakrishnan, ”Spatiotemporal storytelling on twitter”, Virginia Tech Computer Science Technical Report, 2015, [online] Available: [http://vtechworks.lib.vt.edu/handle/10919/24701.](http://vtechworks.lib.vt.edu/handle/10919/24701) [20] D. Kumar, N. Ra- makrishnan, R. F. Helm and M. Potts, Algorithm for storytelling, in IEEE Transactions on Knowledge and Data Engineering, 2008, Volume

20 Issue 6, pp. 736-751.

[14] N. Ramakrishnan, D. Kumar, B. Mishra, M. Potts and R. R. Helm, Turn- ing CARTwheels: an alternating algorithm for mining redescriptions, in Knowledge Discovery and Data Mining, 2004, pp. 266-275.

[15] L. D. Vocht, C. Beecks, R. Verborgh, T. Seidl, E. Mannens and Rik Van De Walle, Improving Semantic Relatedness in Paths for Storytelling with Linked Data on the Web, in Revised Selected Papers of the ESWC 2015

Satellite Events on The Semantic Web: ESWC 2015 Satellite Events, May 31-June 04, 2015.

[16] Z. Chen, X. Zhang, A. P. Boedihardjo, J. Dai, Chang-Tien Lu, Mul- timodal storytelling via generative adversarial imitation learning, in Proceedings of the 26th International Joint Conference on Artifcial Intelligence, August 19-25, 2017, Melbourne, Australia.

[17] R. D. Santos, S. Shah, F. Chen, A. Boedihardjo, Chang-Tien Lu, N.

Ramakrishnan, Forecasting location-based events with spatio-temporal storytelling, in Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, pp. 13-22, November

04-04, 2014, Dallas/Fort Worth, Texas.

[18] Jayant Madhavan, Sreeram Balakrishnan, Kathryn Brisbin, Hector Gon- zalez, Nitin Gupta,Alon Halevy, Karen Jacqmin-Adams, Heidi Lam, Anno Langen, Hongrae Lee,Rod McChesney, Rebecca Shapley, Warren Shen, “Big Data Storytelling through Interactive Maps”

[19] M. Shahriar Hossain1, Patrick Butler1, Arnold P. Boedihardjo2, Naren Ramakrishnan, Storytelling in Entity Networks to Support Intelligence Analysts, KDD ’12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, Pages 1375-1383

,Beijing, China , August 12 - 16, 2012

[20] Zhiqian Chen , Xuchao Zhang , Arnold P. Boedihardjo , Jing Dai , Chang-Tien Lu, Multimodal storytelling via generative adversarial imi- tation learning, Proceedings of the 26th International Joint Conference on Artificial Intelligence, August 19-25, 2017, Melbourne, Australia

[21] Raimundo F. Dos Santos, Jr. , Arnold Boedihardjo , Sumit Shah , Feng Chen , Chang-Tien Lu , Naren Ramakrishnan, The big data of violent events: algorithms for association analysis using spatio-temporal storytelling, Geoinformatica, v.20 n.4, p.879-921, October 2016

[22] Deept Kumar, Naren Ramakrishnan, Richard F. Helm, and Malcolm

Potts. 2008. Algorithms for Storytelling. IEEE TKDE20, 6 (2 2008),

[736–751.DOI:http://dx.doi.org/10.1145/1188913.1188915](http://dx.doi.org/10.1145/1188913.1188915)

[23] “GRACE Tellus Data,” GRACE Tellus. [Online]. Available:

https://grace.jpl.nasa.gov/data/get-data. [Accessed: 26-Jul-2019].

[24] “SARAL - eoPortal Directory - Satellite Missions.” [Online]. Available: https://directory.eoportal.org/web/eoportal/satellite-missions/s/saral. [Accessed: 26-Jul-2019].

[25] “Top 5 sources of big data — Artificial Intelligence — Data Science —,”

26-Nov-2017. [Online]. Available: [https://www.allerin.com/blog/top-5-](http://www.allerin.com/blog/top-5-) sources-of-big-data. [Accessed: 26-Jul-2019].

[26] J. Thatcher, “Big Data, Big Questions— Living on Fumes: Digital Footprints, Data Fumes, and the Limitations of Spatial Big Data,” International Journal of Communication, vol. 8, no. 0, p. 19, Jun. 2014.

[27] “Livehoods.” [Online]. Available: [http://livehoods.org/research.](http://livehoods.org/research) [Ac- cessed: 27-Jul-2019].