Forecasting of Combat Activity in the Syrian Civil War Using Hot Spot

Kate Schaefer, Data Scientist, NGA Data Corps

4/17/2019

The overall classification of this presentation is:

Classified by:
Derived from:
Declassify on:

NATIONAL GEOSPATIAL NGA INTELLIGENCE AGENCY

Outline

- Hypothesis
- Syrian Civil War
- Hot Spot vs. Density
- Methodology
- Results
- Conclusion

Hypothesis

Test a methodology for forecasting spatio-temporal patterns of future violent events based on passed violent events.

Can past events serve as a predictor of future events?

Does grouping by time before grouping by space enhance the forecasting accuracy?

Syrian Civil War

- Started March 2011
- Still ongoing
- Global Database of Events, Language, and Tone (GDELT)
    - News media from print, broadcast, and web formats in over 100 languages, updated daily
    - Data from January 2014 – December 2015; 24-month test period
    - Used TimeMapper Visualizer to extract “Material Conflict” for all of Syria
        - Points are weighted at the city level per day
    - 297,127 discrete events

NGA

Hot Spot vs. Density

- Sometimes used interchangeably, not the same
- Density: clusters group of objects based on proximity
    - Can be used to see the “now”
- Hot Spot: Here refers to specific ArcPro tool *Optimized Hot Spot Analysis* which uses the Getis Ord GI* algorithm
    - Identifies statistically significant “hot” or “cold” clusters
- Hot Spot commonly used in crime mapping to plan for future distribution of resources
    - In the literature: Chainey, Tompson, Uhlig

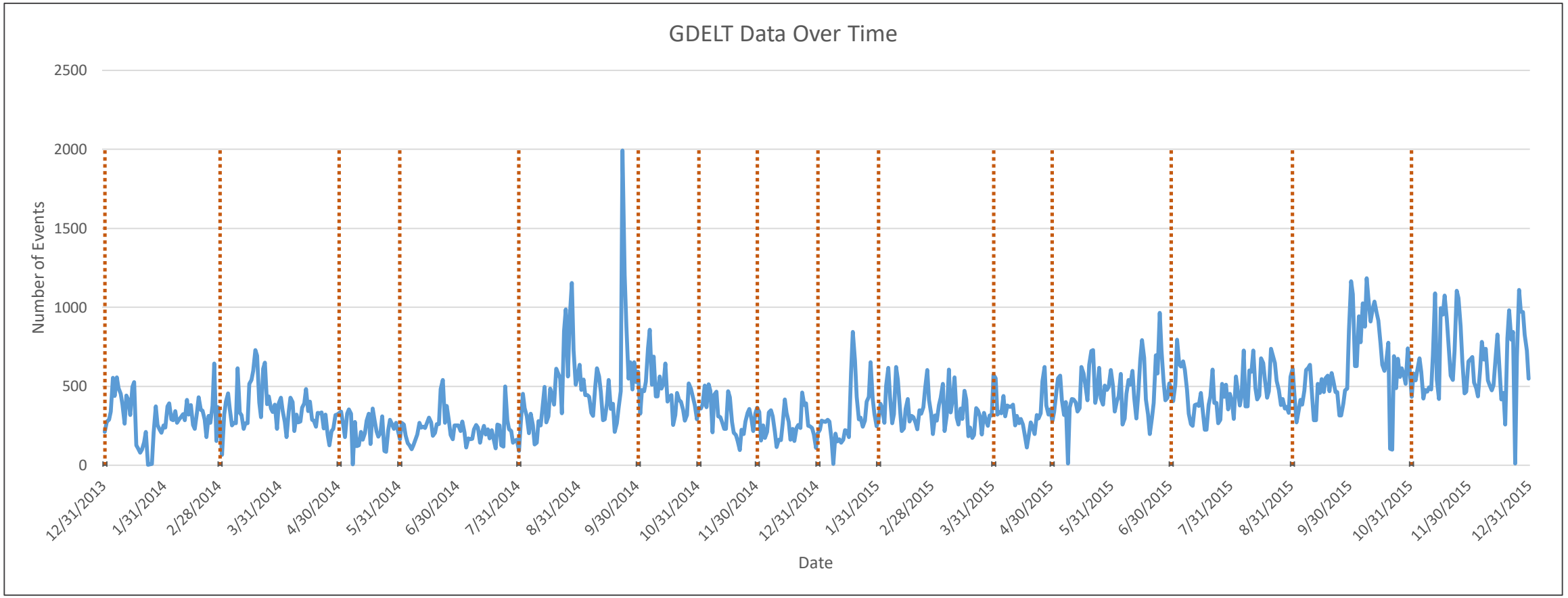
NGA

Methodology

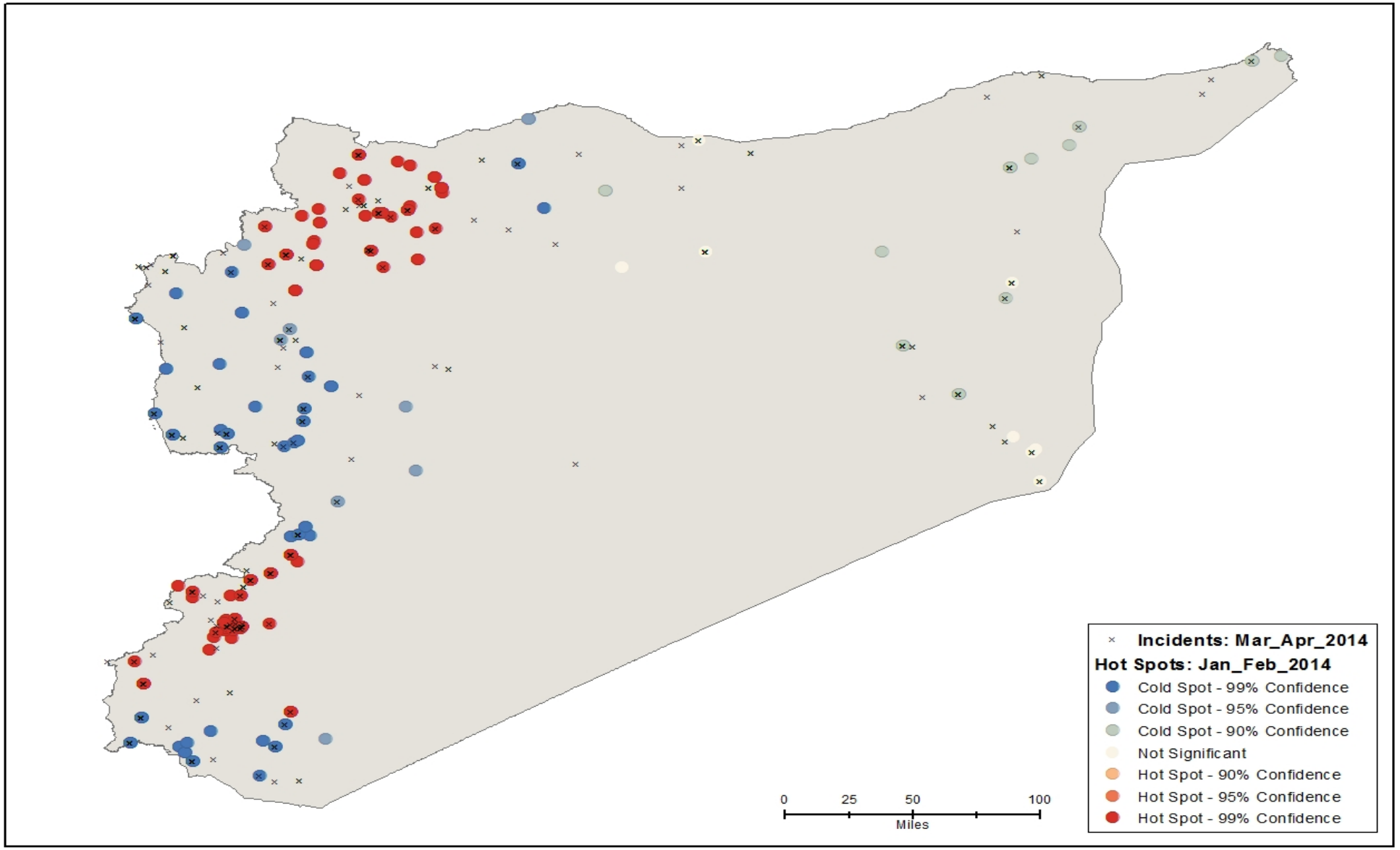
- Broke the dataset into different time sets, then tested to see if the past dataset predicted the future dataset.
- *Grouping Analysis*
  - Clusters data by natural breaks; in this case breaks in time
  - Calculates F-Statistic measuring group similarity/difference
- *Optimized Hot Spot Analysis*
  - Groups data by space
  - Calculates neighbor distance, incremental spatial autocorrelation, Getis Ord GI* statistic
- Assessing accuracy: Hit Rate (HR) is the percentage of new events that occur within the hot spots identified in preceding time period

NGA

Results



The blue line represents the GDELT hits per day. The orange dash lines represent the fifteen groups formed by the grouping analysis.



NGA

Results

- Hit rate range: 36.9 – 63.3%
- Average hit rate: 47.7%
- Average False Positive: 5.8%

- Average crime mapping hit rates: 8-20%
- Average hit rate for GDELT without the addition of the grouping analysis: 29.3%
    - Adding in the grouping method increased the average hit rate by 18.4%

NGA

Conclusion

Hot spot analysis commonly used in crime mapping; can also be applied to a war environment

Adding in the grouping analysis to the hot spot methodology increases the accuracy of forecasting future areas of violence

NGA

NGA