

# Geolocating tweets via spatial inspection of information inferred from tweet meta-fields

GeomobTLV  
April 2022

# Introduction

- Twitter - a robust platform for distributing messages
- More than 500 million users worldwide
- 280 characters only in a tweet
- Sent via web pages, mobile devices & 3<sup>rd</sup> party apps
- Only 1-3 % of the tweets are geolocated (GPS/GeoIP)



# What are the goals?

- Developing a spatial approach for georeferencing tweets and verifying potential newsworthy messages



- Analyzing natural hazards in Israel and evaluating their severity using Twitter

# Motivations

- Georeferencing tweets spatially and temporally
- Verifying accuracy/veracity using spatial approaches
- Eyewitness/Newsworthy tweets
- Studying Natural Hazards using social media
- Information for decision makers during a catastrophe

**1969:** by 2020 we're going to have flying cars, all diseases cured, and everyone will be rich

**2020:**



Joe Biden ✓  
@JoeBiden

I can't believe I have to say this,  
but please don't drink bleach.

יש ציוצים  
הומוריסטים/סרקסטיים...



זהבה גלאון ✓ @zehavagalon · Sep 17, 2019

**משה תמצץ** פעם אחת אם אתה מוחזק בבלפור **פעמיים** אם אתה בקיסריה



Moshe Kahlon - **משה כחלון** @KahlonMoshe · Sep 17, 2019

מנהיג תנועת הליכוד הוא ראש הממשלה בנימין נתניהו והוא המועמד הבלעדי מטעמה להמשיך ולכהן כראש ממשלה.

233

510

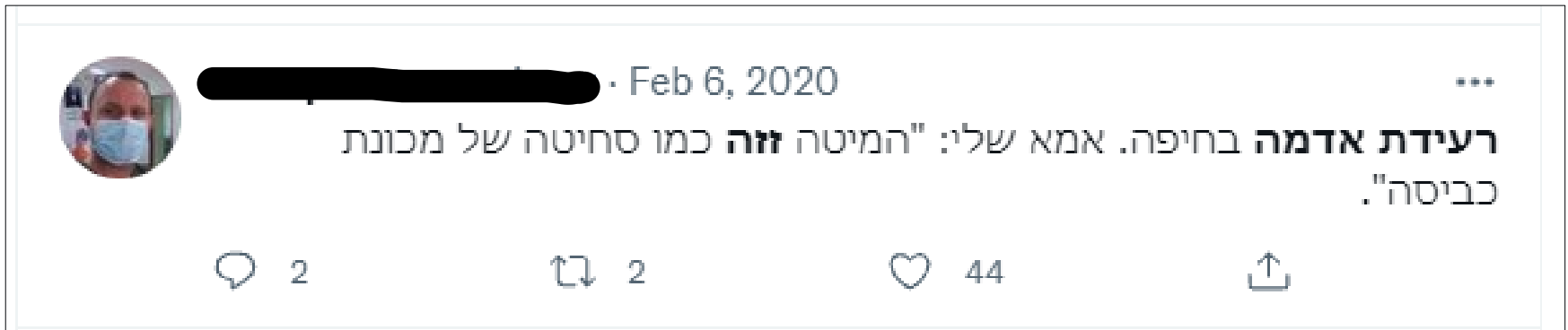
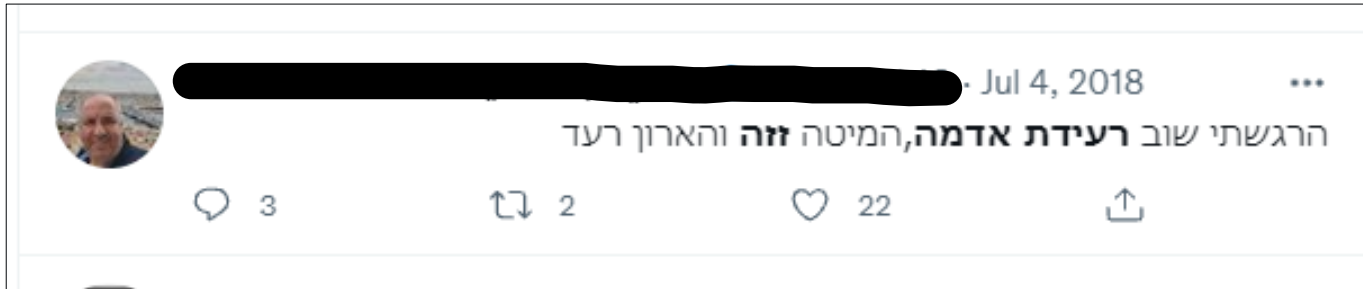
5.7K



יש כאלו  
שאינם רלוונטיים...



אבל יש צינצים  
השווים המון!



“

*Where is the location the tweet  
was sent from and where is the  
location of the phenomena  
described?*

”



# Georeferencing tweets

Three major approaches:

- GPS/GeoIp of the device (relevant for 1-3% of all tweets)
- Network-related associations (e.g., followers, retweets)
- Georeferencing by matching **tokens** to lists of known **locations** (i.e., **gazetteers**)

## Tweet (text)



**President Biden** ✓ @POTUS · 20m

United States government official

My Administration is working overtime to show that our climate commitment is about action — not words.

That's why on my very first day in office, I took action to rejoin the Paris Agreement. And we've been continuing our work every day since.

586

495

2.7K



## Profile (user location & user description)



**President Biden** ✓

1,808 Tweets



Follow

**President Biden** ✓

@POTUS

United States government official

46th President of the United States, husband to @FLOTUS, proud dad & pop.

Tweets may be archived: [whitehouse.gov/privacy](https://whitehouse.gov/privacy)

[WhiteHouse.gov](https://whitehouse.gov) Joined January 2021

12 Following 14.6M Followers



Followed by Real News IL, Ran Israeli, and 16 others you follow

Tweets

Tweets & replies

Media

Likes

# Locations proxies

- The device location (true GPS-based location) >> the location of the reporting user
- Phenomena location >> the location of the described phenomena

Cases:

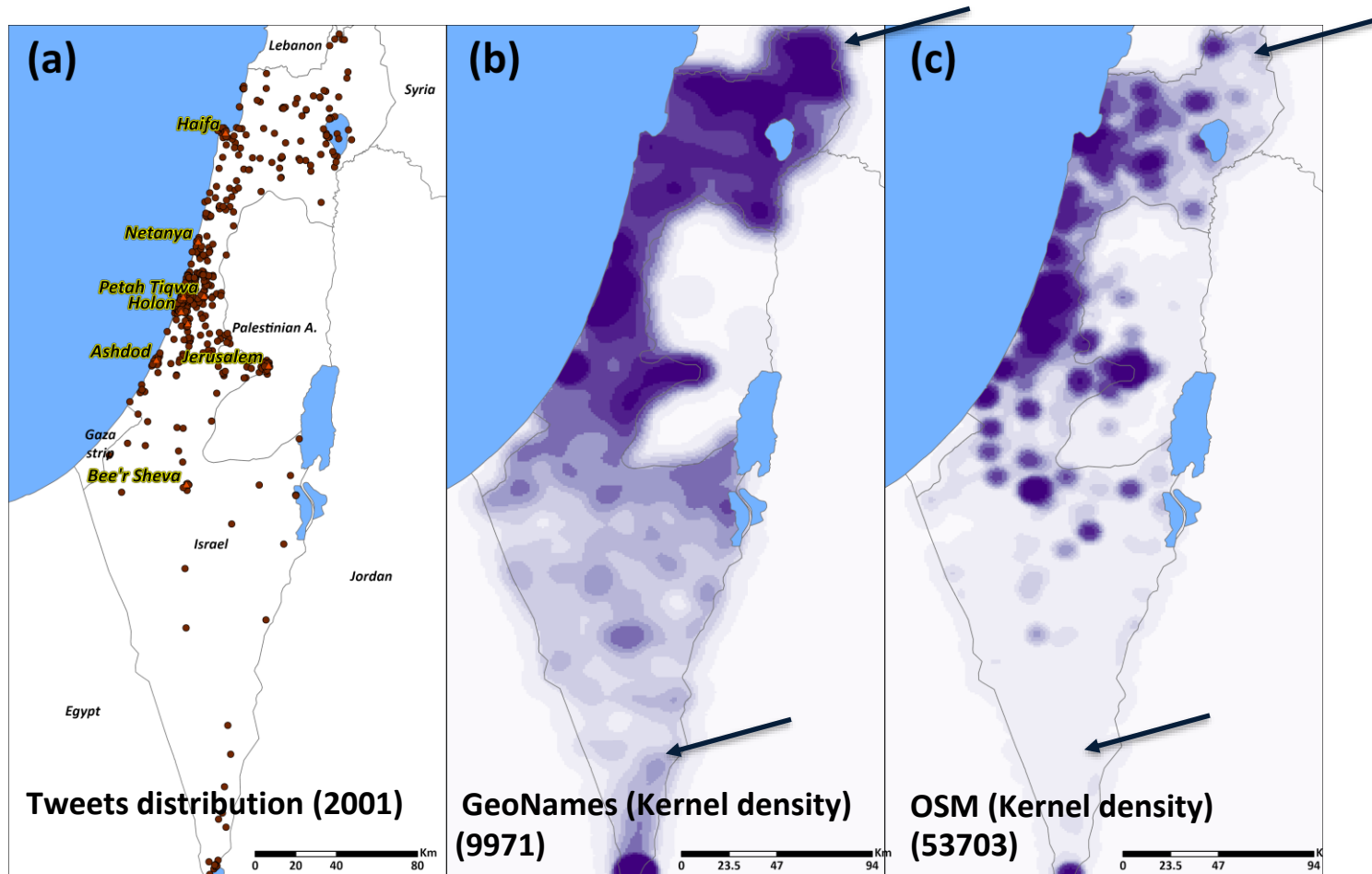
- New York (text)-New York (UL) (while in New York)
- New York (text)-New York (UL) (while in London)
- New York (text)-Berlin (UL) (while in London)

The structure  
of a tweet  
(json): the  
**User** object



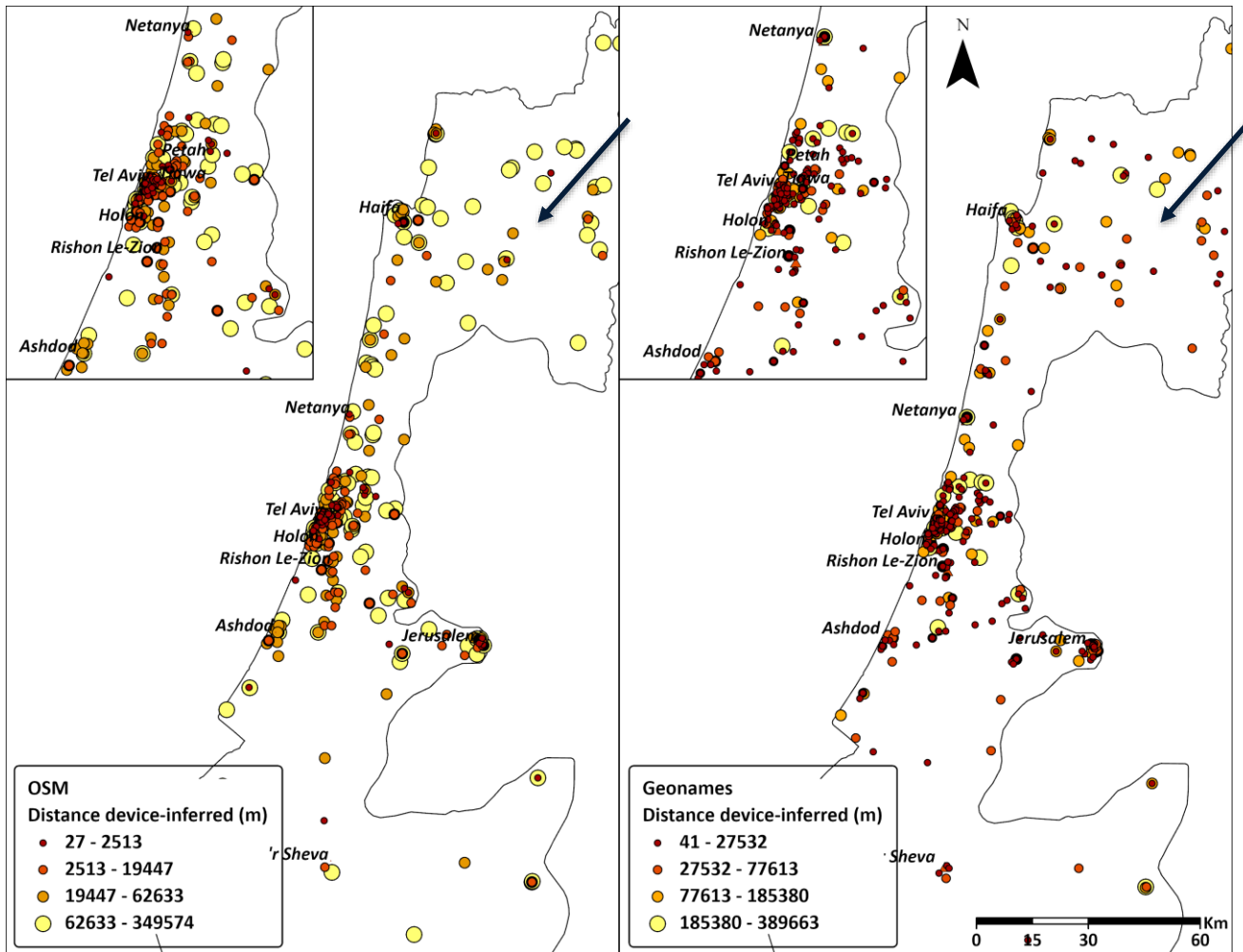
```
{
  "id": 6253282,
  "id_str": "6253282",
  "name": "Twitter API",
  "screen_name": "TwitterAPI",
  "location": "San Francisco, CA",
  "profile_location": null,
  "description": "The Real Twitter API. Tweets about API changes, service issues",
  "url": "https:\\\\t.co\\8IkCzCDr19",
  "entities": {
    "url": {
      "urls": [{
        "url": "https:\\\\t.co\\8IkCzCDr19",
        "expanded_url": "https:\\\\developer.twitter.com",
        "display_url": "developer.twitter.com",
```





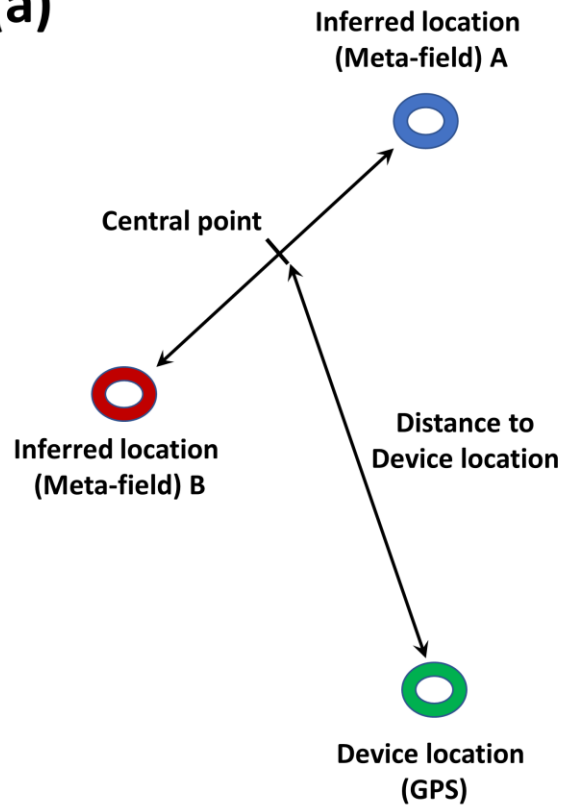
# Matching tokens-localities

- Tokenizing the **text**, **user location** & **user description** fields
- Matching tokens (tweet) >> localities (gazetteer) using the Levenshtein Distance algorithm
- Extracting a location (x/y/z) + the score (quality) of matching
- Benchmark – comparing to geolocated (GPS) tweets

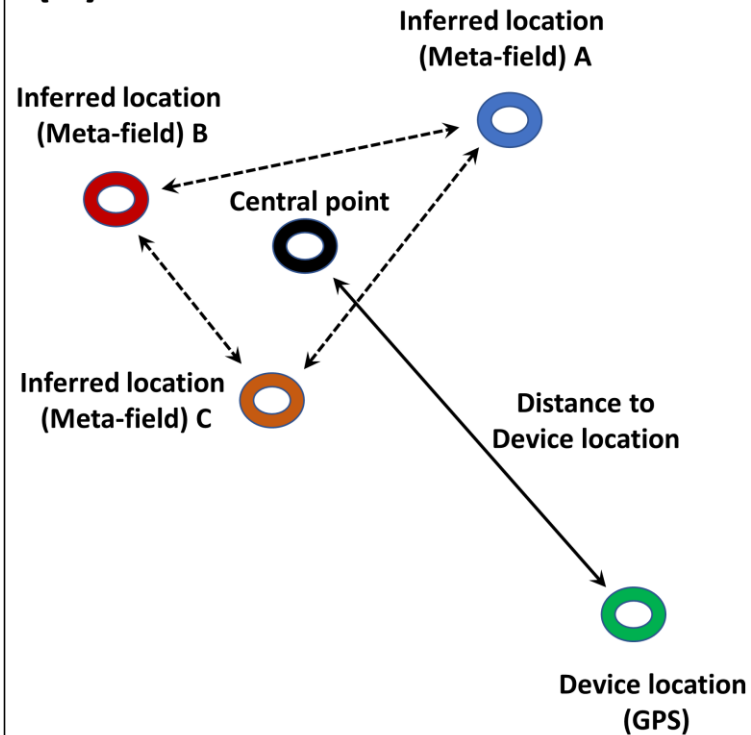


## Proxy inference

(a)

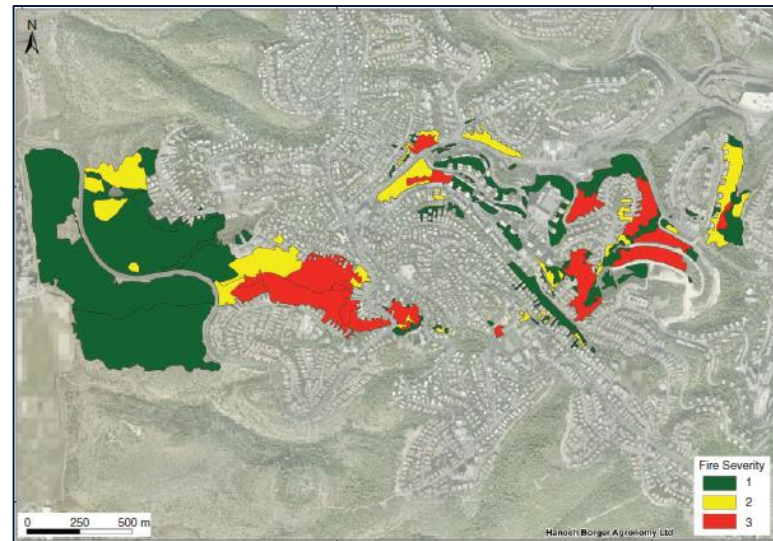
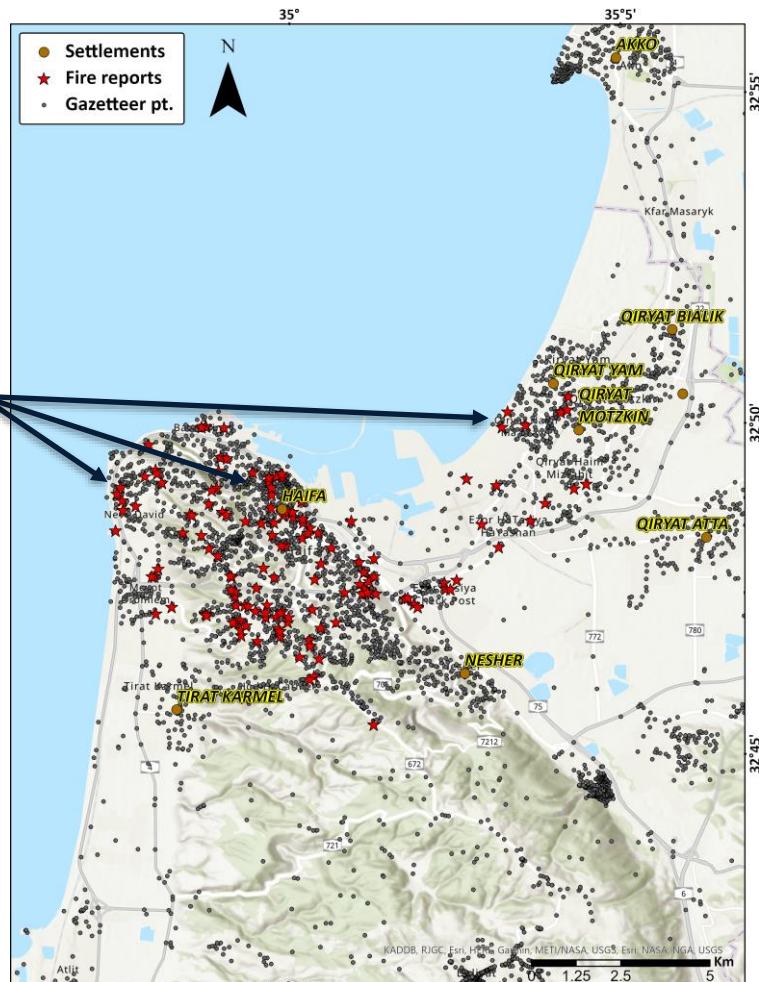


(b)

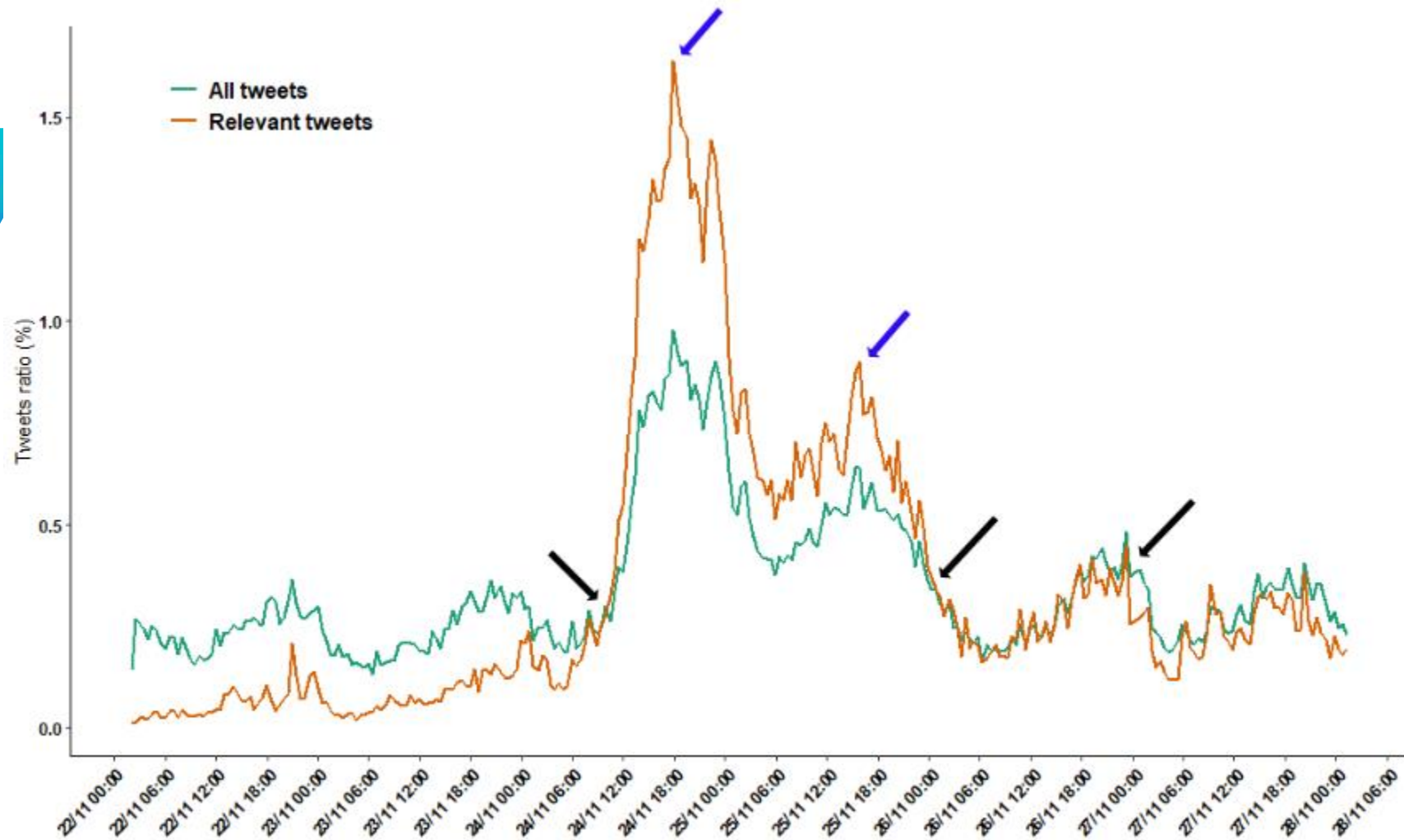


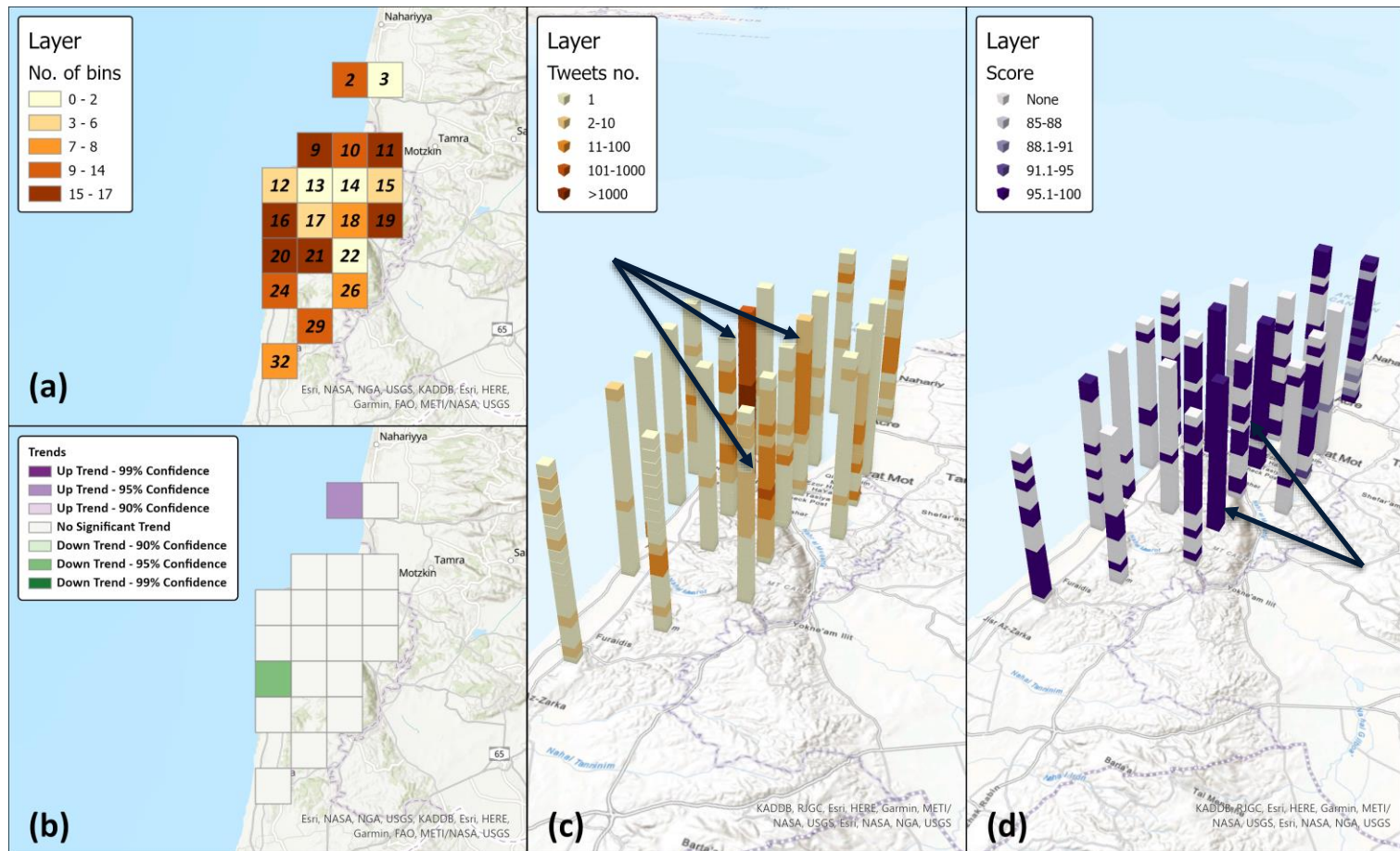


IFR calls



Fire severity in Haifa (credit: Tessler et al., 2019)

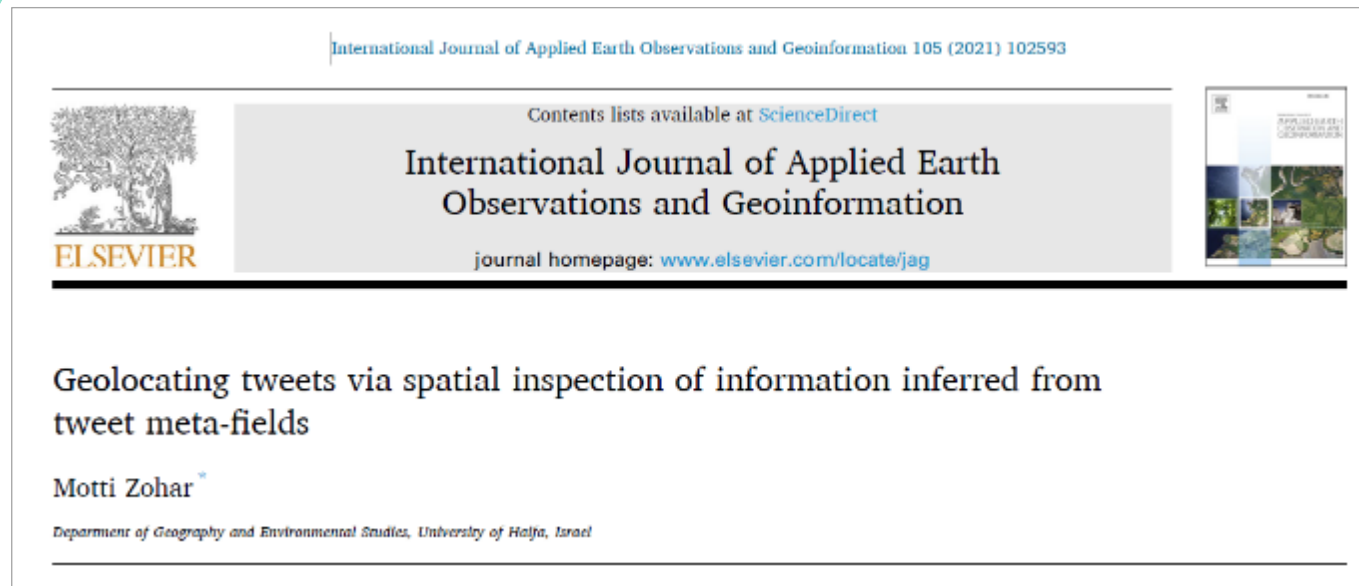




# Summary

- Geonames/OSM datasets for tweets georeferencing
- Meta-fields of tweets as sources of information
- Device location vs. phenomena location
- Proxies/indications for newsworthy tweets

# Thank You!



<https://www.sciencedirect.com/science/article/pii/S0303243421003007>

***The research was funded by the Israel Ministry of Science and Technology and the National  
Emergency Management Authority of the Ministry of Defense (Grant #3-14737)***