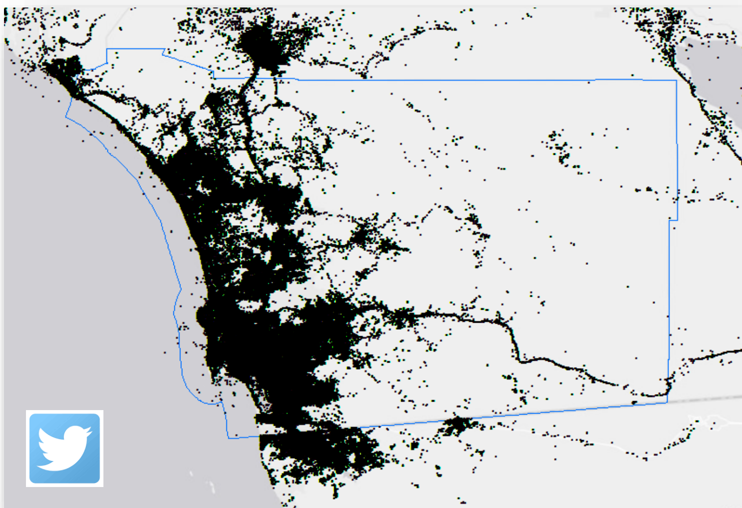
Spatiotemporal Data Analytics on Large Event Data

Atsushi Nara
Assistant Professor, Department of Geography
Associate Director, Center for Human Dynamics in the Mobile Age (HDMA)
San Diego State University

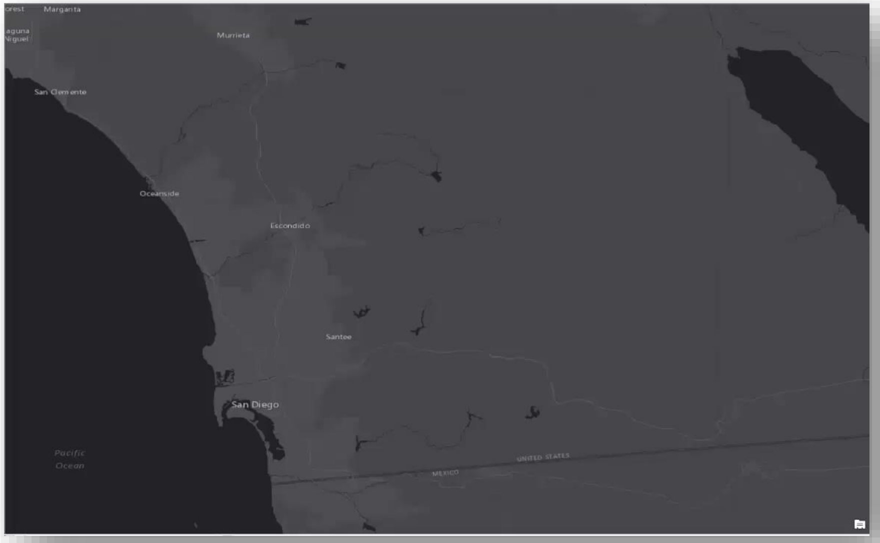
Introduction

The Age of Big Spatiotemporal Data

SPACE × TIME



Geo-referenced tweets



Extracting movements from tweets

The Age of Big Spatiotemporal Data

SPACE × TIME

Example: Tracking Movements by GPS devices

1 person per hour for a year:	$1 \times 24 \times 365 = 8{,}760$
1 person per minute for a year:	$1 \times 60 \times 24 \times 365 = 525{,}600$
1 person per second for a year:	$1 \times 60 \times 60 \times 24 \times 365 = 31{,}536{,}000$

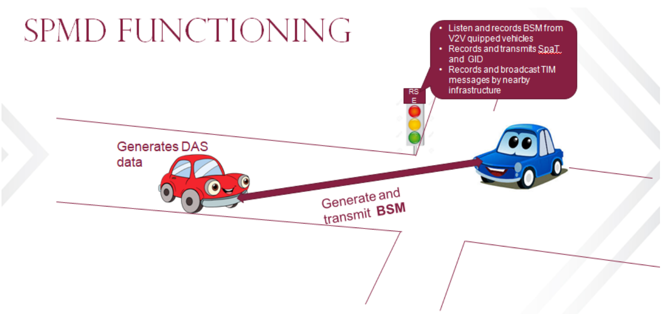
Introduction

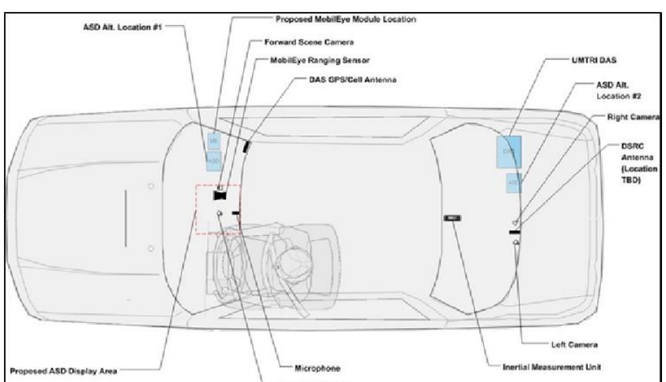
The Age of Big Spatiotemporal Data

Chen, Y., Tsou, M-H., Nara, A. (2019)

SPACE × TIME

Example: Safety Pilot Model Deployment (SPMD) Data





Subject:	2,836 Vehicles
Duration:	1 month
Frequency:	**≈ 10 Hz**
Data points:	**> 1.5 billion GPS points**
File size:	205 GB in CSV
Attributes:	speed, location, direction, yaw rate, heading, etc.

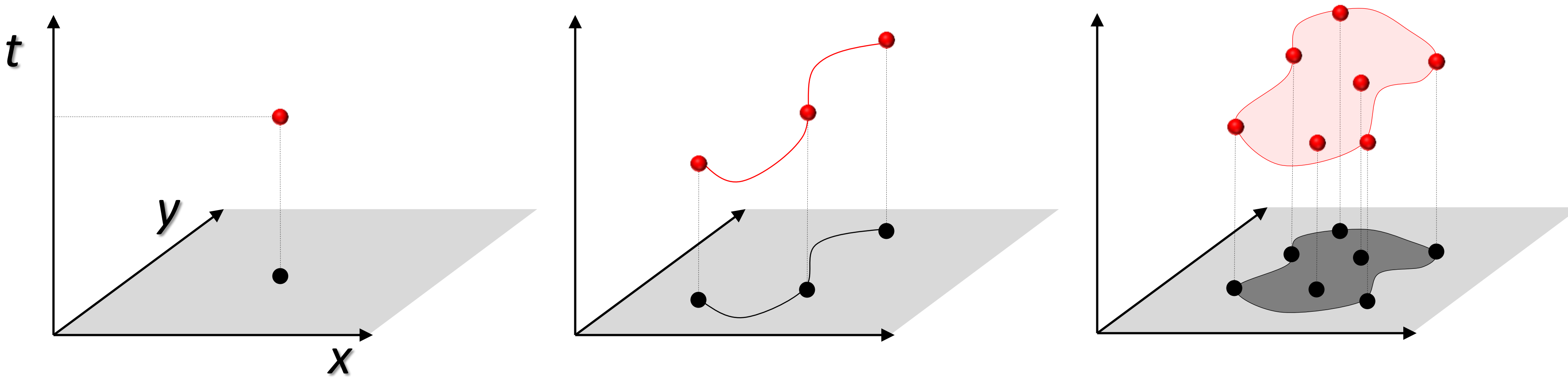
SAN DIEGO STATE UNIVERSITY HDMA @SDSU CALIFORNIA AIR RESOURCES BOARD NSF

Introduction

$$Event = (S, T, A)$$

S: Spatial Attributes (point, line, polygon, pixel, …)
T: Temporal Attributes (start-end time, duration)
A: Event Attributes (id, name, …)

Time Geography Representation (Vector)



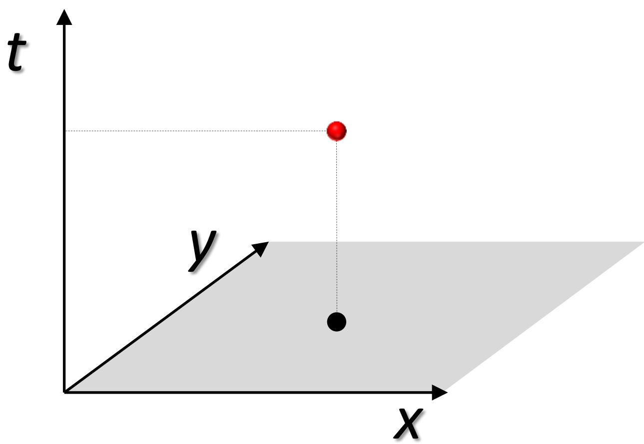
$$e=(x_1, y_1, t_1, A)$$

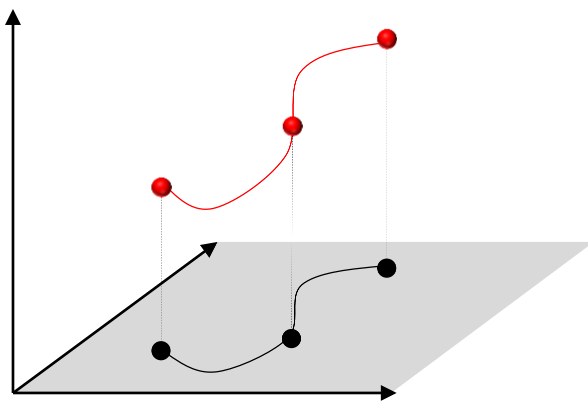
$$e=\{ (x_1, y_1, t_1), (x_2, y_2, t_2), (x_3, y_3, t_3), A \}$$

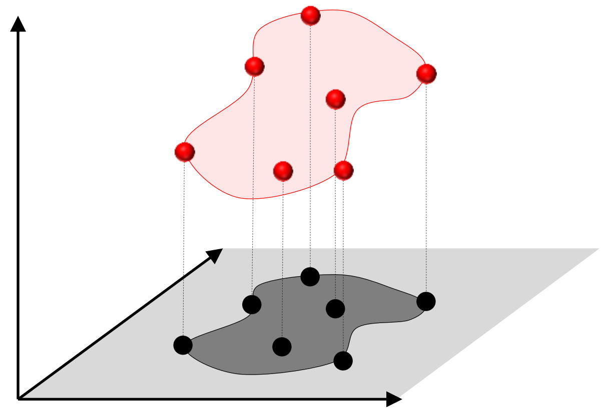
$$e=\{ (x_1, y_1, t_1, a_1), (x_2, y_2, t_2, a_2), \ldots, (x_n, y_n, t_n, a_n) \}$$

Introduction

Time Geography Representation (Vector)





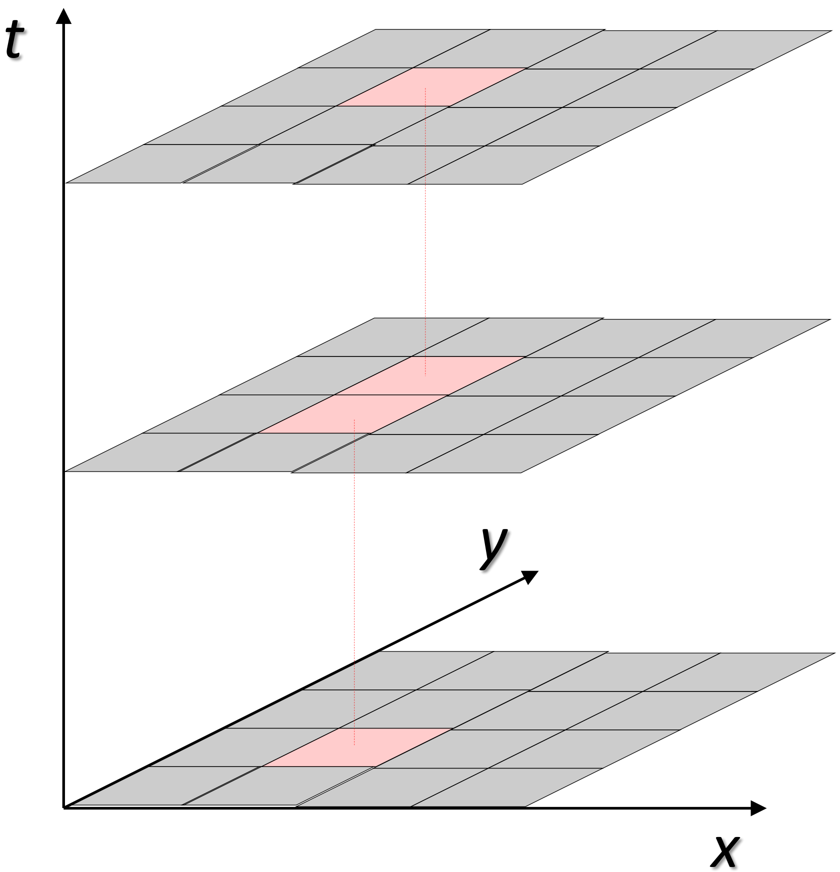


$e=(x_1, y_1, t_1, A)$

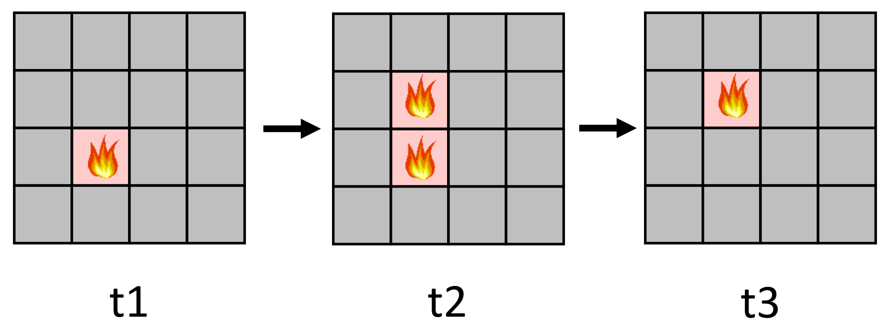
$$e=\{ (x_1, y_1, t_1), (x_2, y_2, t_2), (x_3, y_3, t_3), A \}$$

$$e=\{ (x_1, y_1, t_1, a_1), (x_2, y_2, t_2, a_2), \dots (x_n, y_n, t_n, a_n) \}$$

Time Geography Representation (Raster)

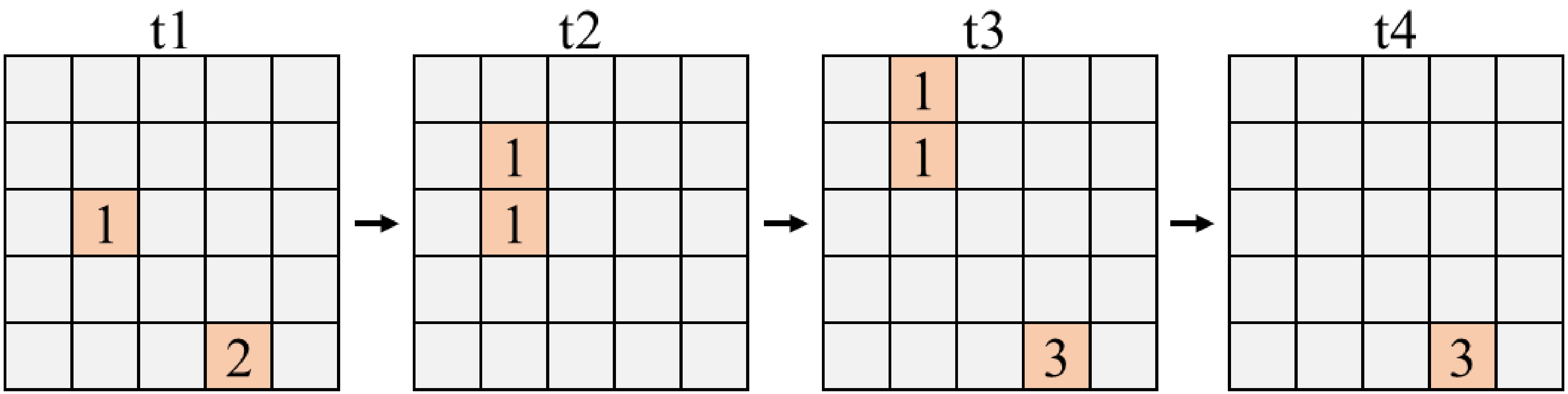


Single Spatiotemporal Event



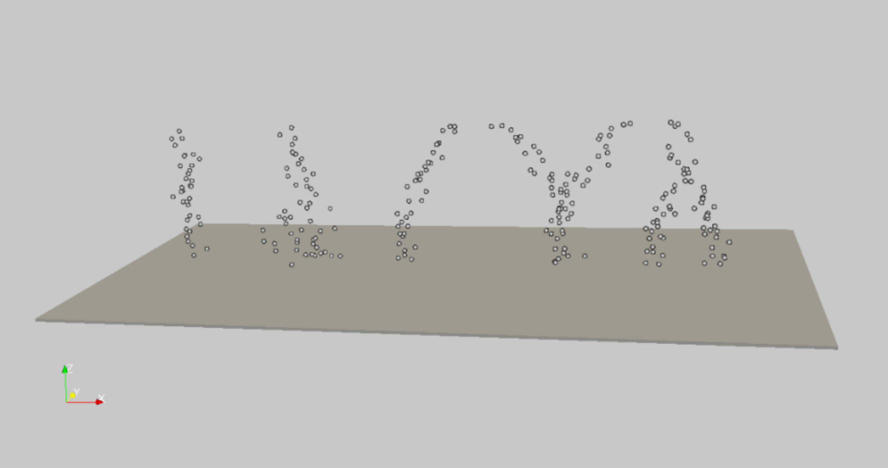
Time Geography Representation (Raster)

Three Spatiotemporal Events

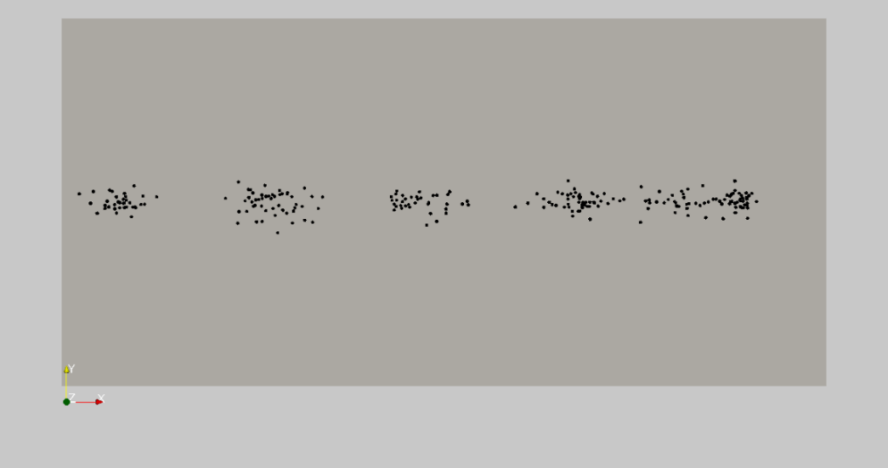


Introduction

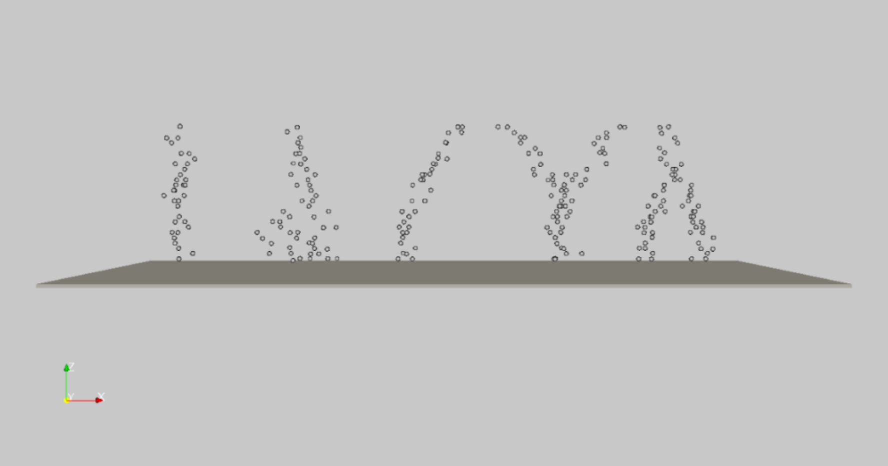
ST Points



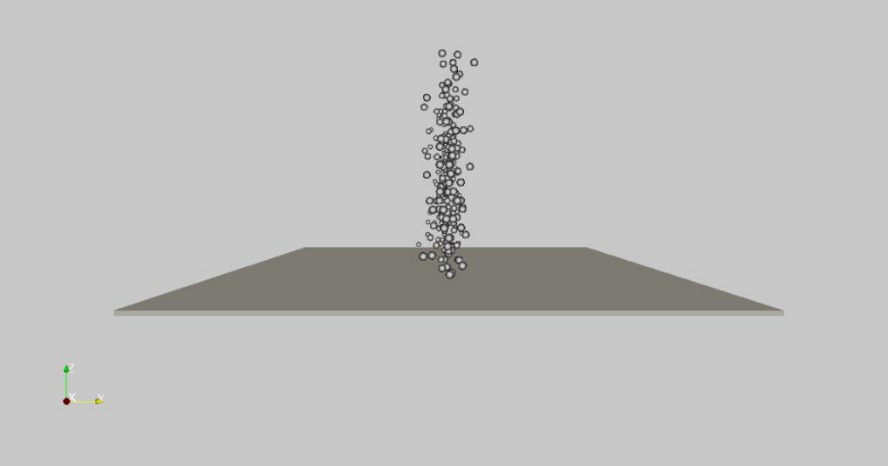
2D Points



ST Points (X-T)



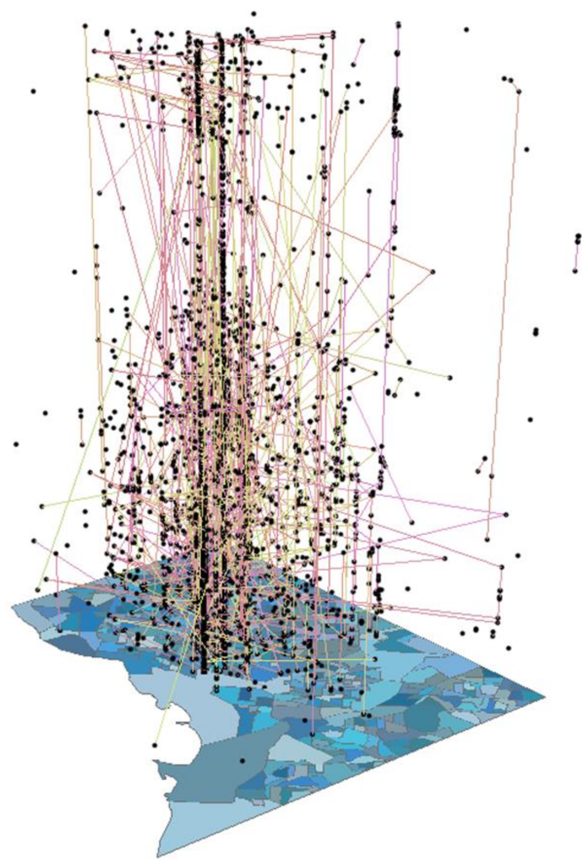
ST Points (Y-T)



Collaborative work with **Dr. May Yuan**, U of Texas at Dallas

Research Objective

- Explore spatiotemporal trends of large event datasets effectively

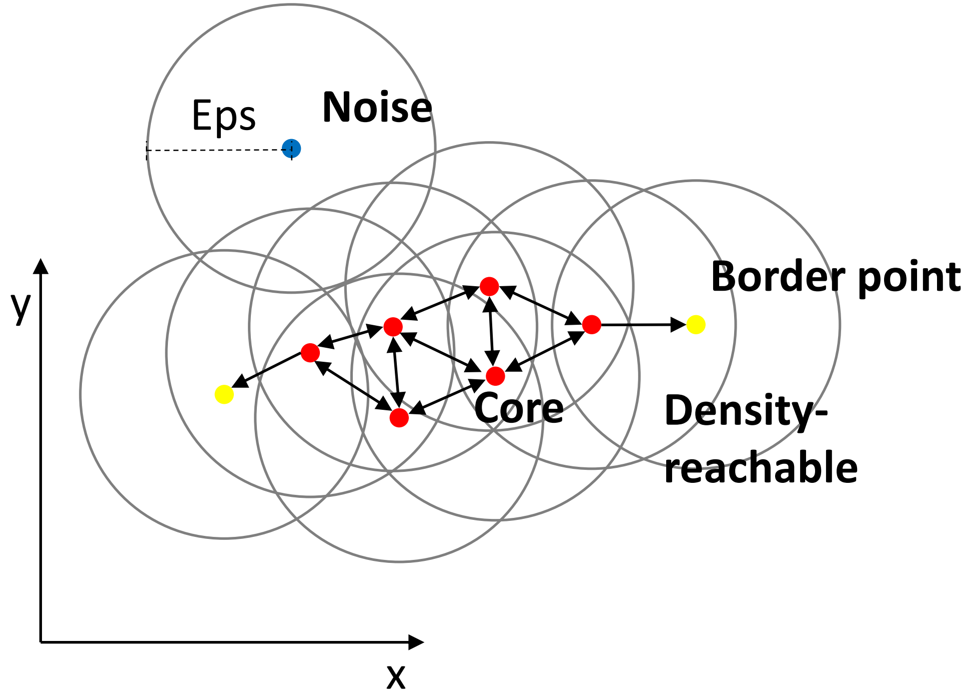


Methodology

- Develop a **parallel** implementation of **ST-DBSCAN** to identify spatiotemporal event clusters

DBSCAN

Min Points = 4

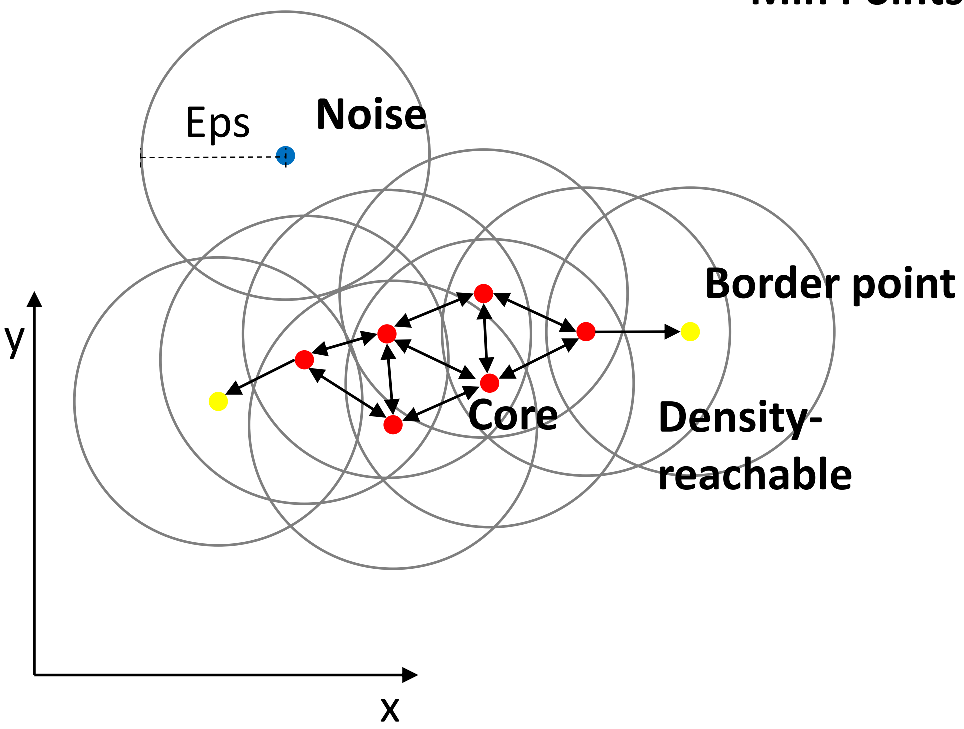


DBSCAN (2D)

Methodology

DBSCAN

Min Points = 4



DBSCAN (2D)

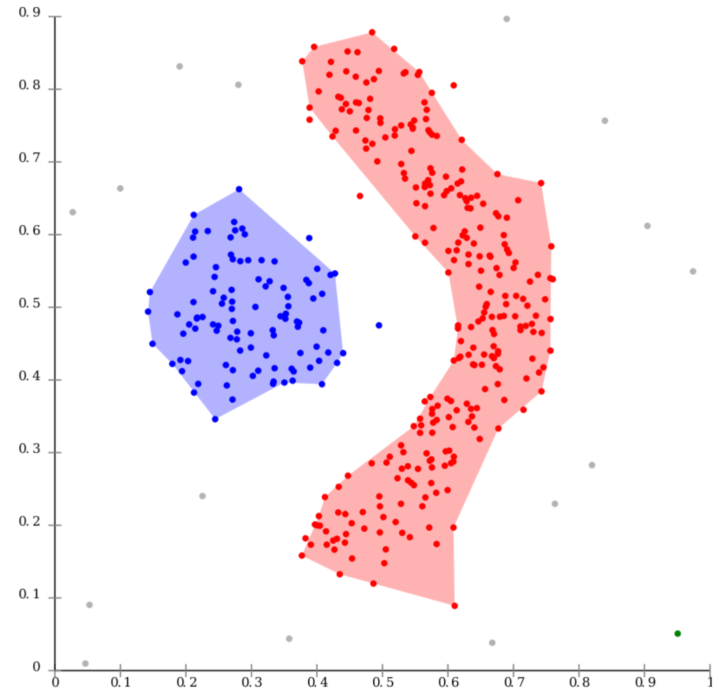


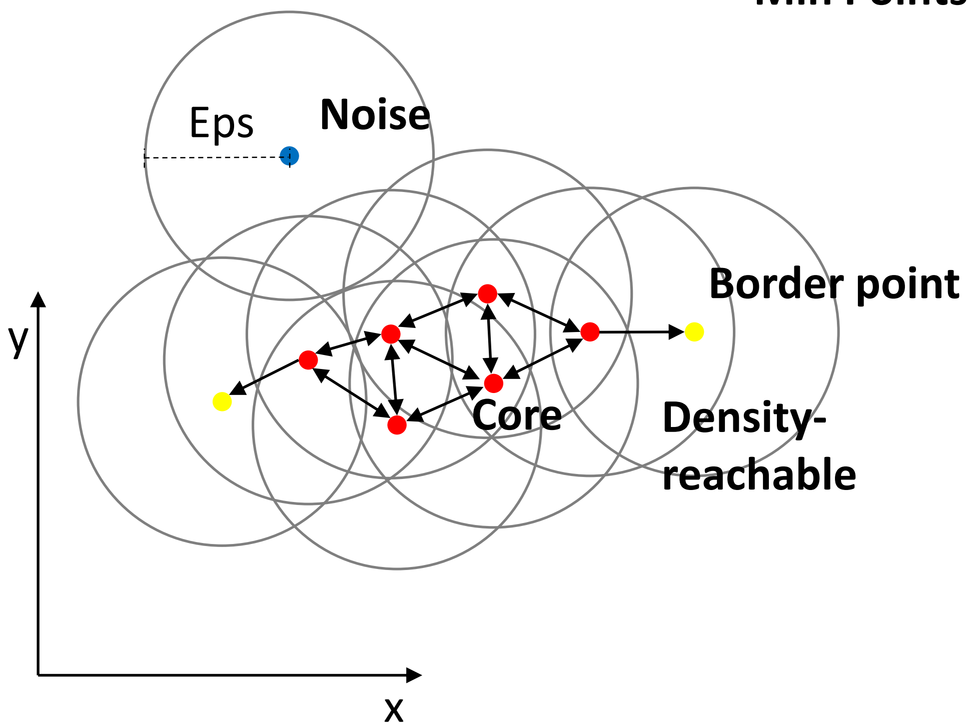
Image from Wikipedia

San Diego State University HDMA @SDSU California Air Resources Board NSF

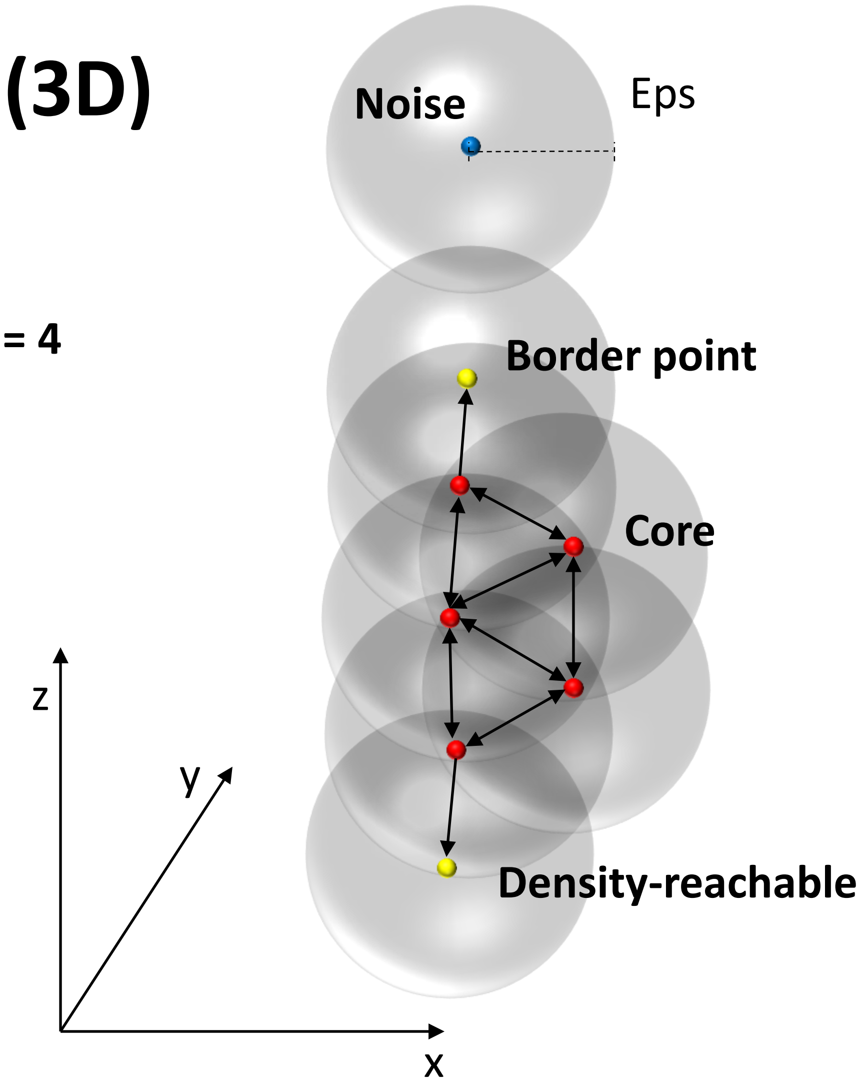
Methodology

DBSCAN (3D)

Min Points = 4



DBSCAN (2D)



DBSCAN (3D)

DBSCAN Runtime Complexity

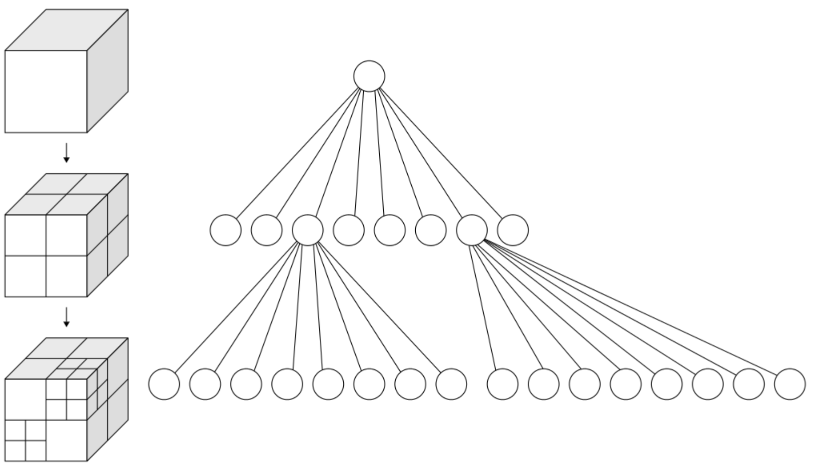
Average: **$O(N*\log(N))$**

Worst case scenario: **$O(N^2)$**

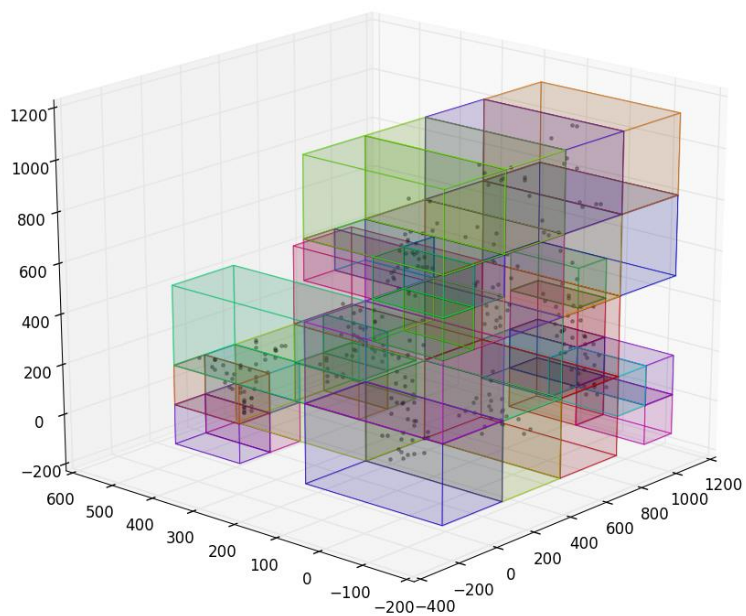
N: the size of the input data set

ST-DBSCAN Parallel Implementation

- Multi-core processing
- Partition data using Octree
- ST-DBSCAN on partitioned data
- Merge cluster results



Octree
(Image from Wikipedia)



Octree Partition with a test datasets

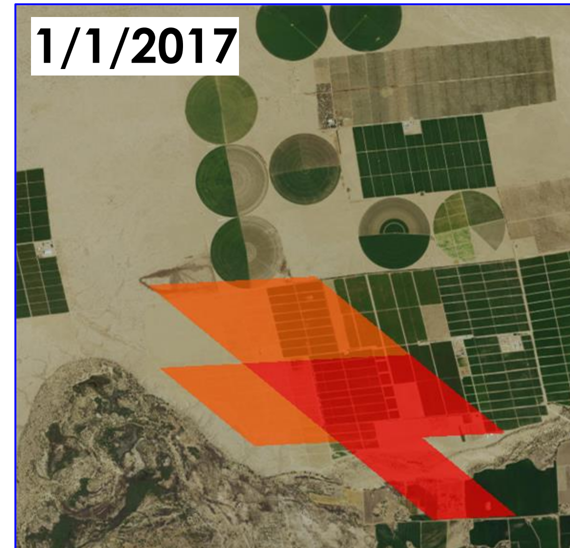
SAN DIEGO STATE UNIVERSITY HDMA @SDSU CALIFORNIA AIR RESOURCES BOARD NSF

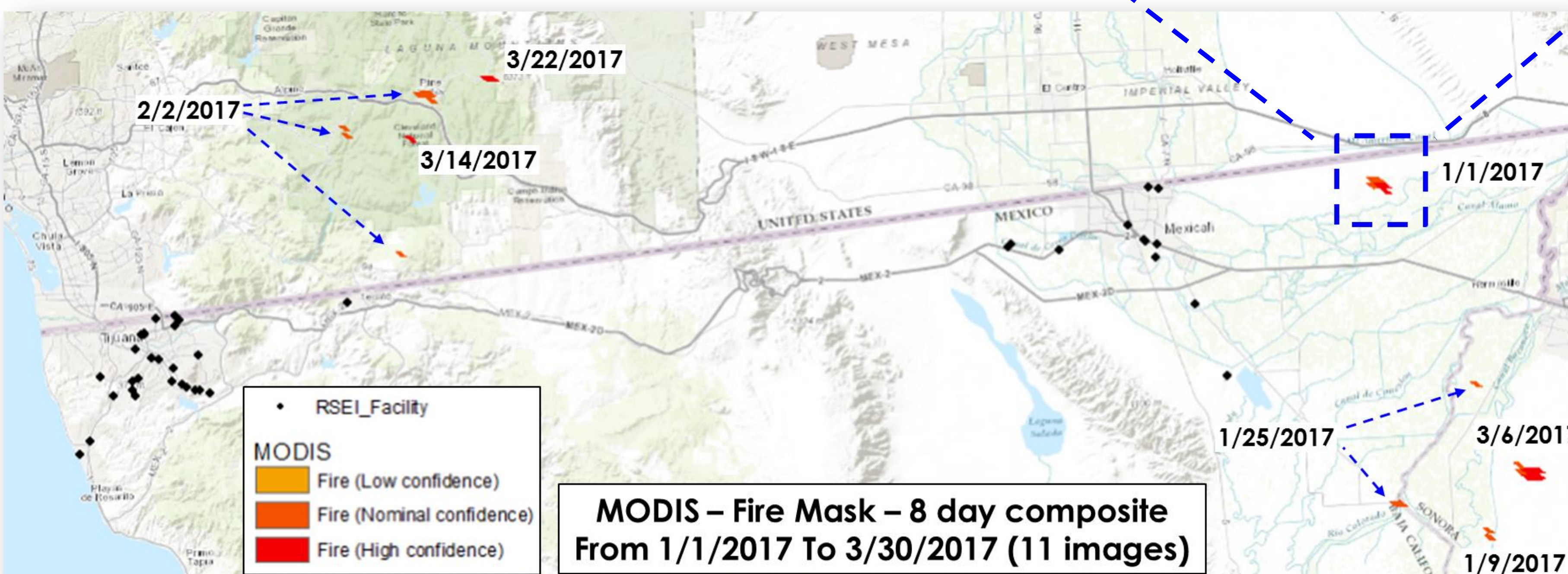
Case Study

Find burning activities such as agricultural burns in the US-Mexico border area

Data Source: MODIS – Active Fire Product

- Thermal Anomalies & Fire
- Cell size: ~1,000m x ~1,000m
- Cycle: 1 day x 2 (Terra & Aqua)
- 1,545 HDF files from 2000 – 2018



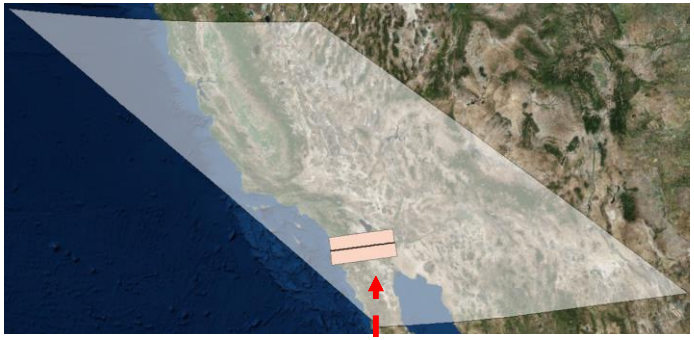


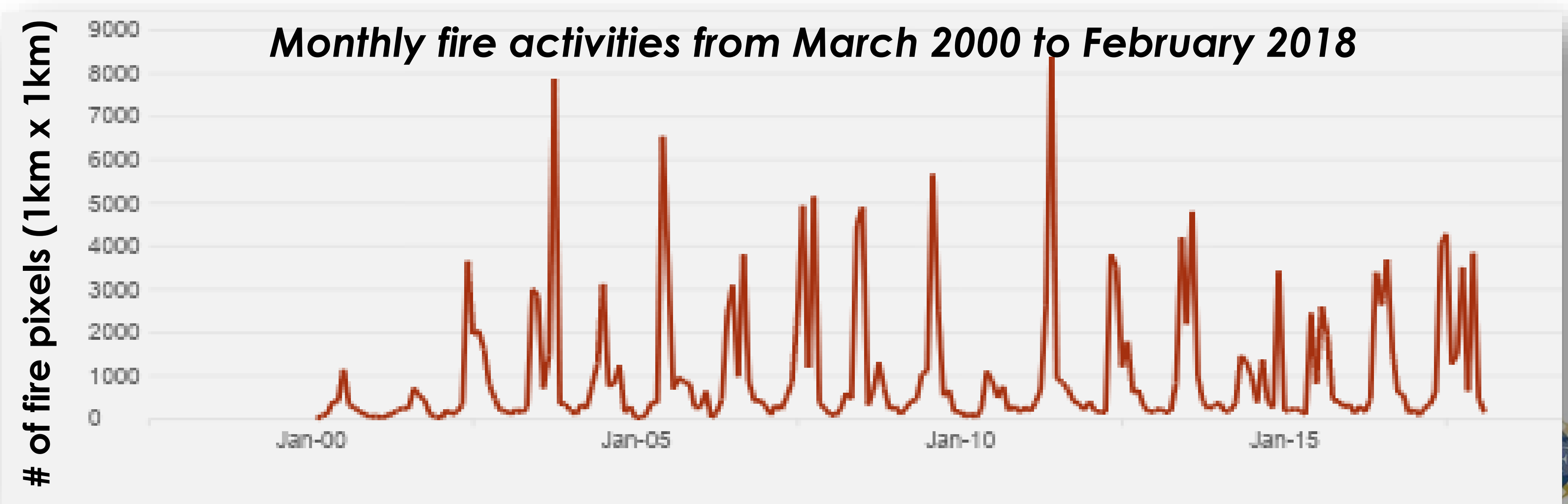
Case Study

Find burning activities such as agricultural burns in the US-Mexico border area

Data Source: MODIS – Active Fire Product

- Thermal Anomalies & Fire
- Cell size: ~1,000m x ~1,000m
- Cycle: 1 day x 2 (Terra & Aqua)
- 1,545 HDF files from 2000 – 2018





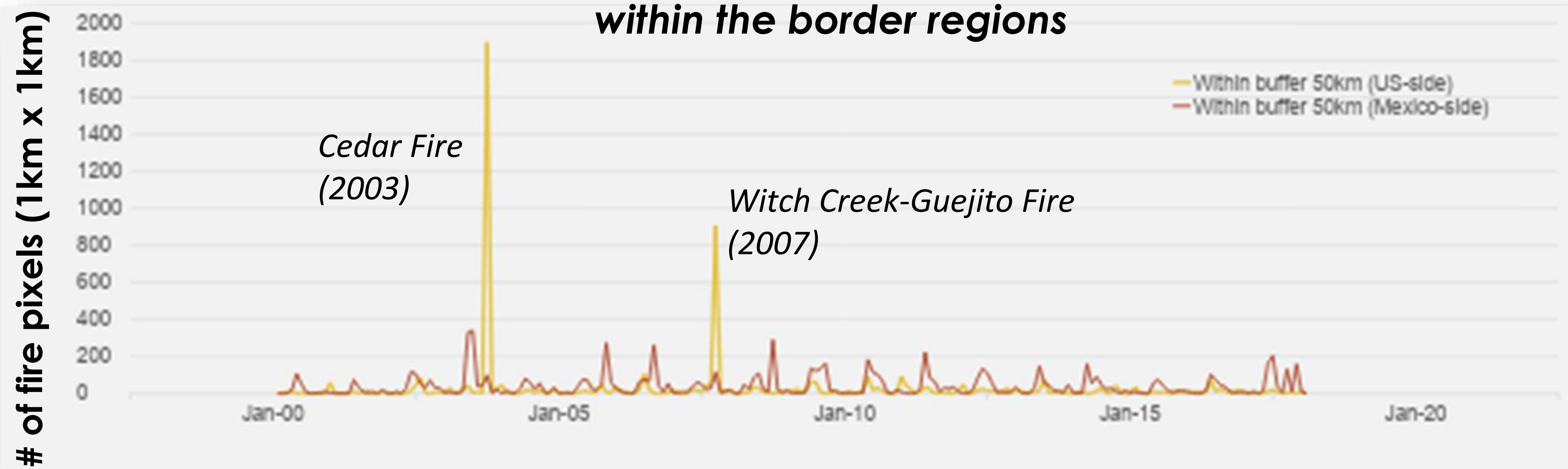
Case Study

Find burning activities such as agricultural burns in the US-Mexico border area

Data Source: MODIS – Active Fire Product
- Thermal Anomalies & Fire
- Cell size: ~1,000m x ~1,000m
- Cycle: 1 day x 2 (Terra & Aqua)
- 1,545 HDF files from 2000 – 2018

Fire pixels within the border regions

Monthly fire activities from March 2000 to February 2018 within the border regions



Case Study

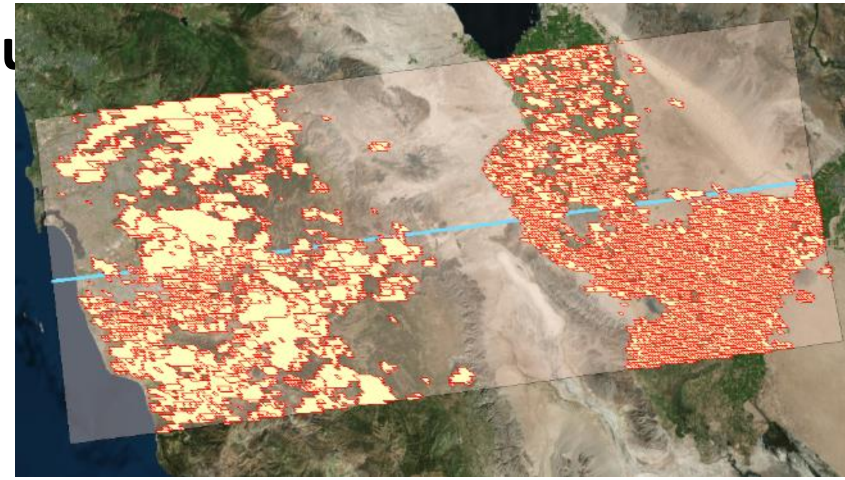
Find burning activities such as agricultural burns in the US-Mexico border area

Identify fire events using ST-DBSCAN

- eps-space = 1,000 m
- eps-time = 1 day
- min-points = 1 pixel
- spatially and temporally contiguous fire pixels

Case Study

Find burning activities such as agricultu burns in the US-Mexico border area

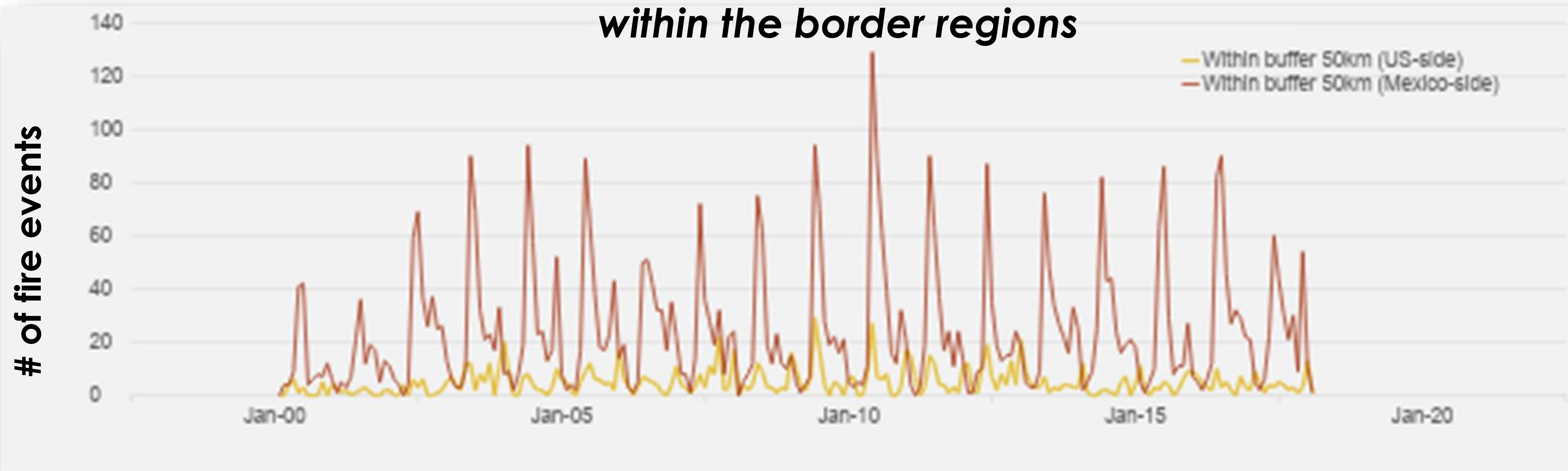


Fire events within the border regions

Identify fire events using ST-DBSCAN

- eps-space = 1,000 m
- eps-time = 1 day
- min-points = 1 pixel
- spatially and temporally contiguous fire pixels

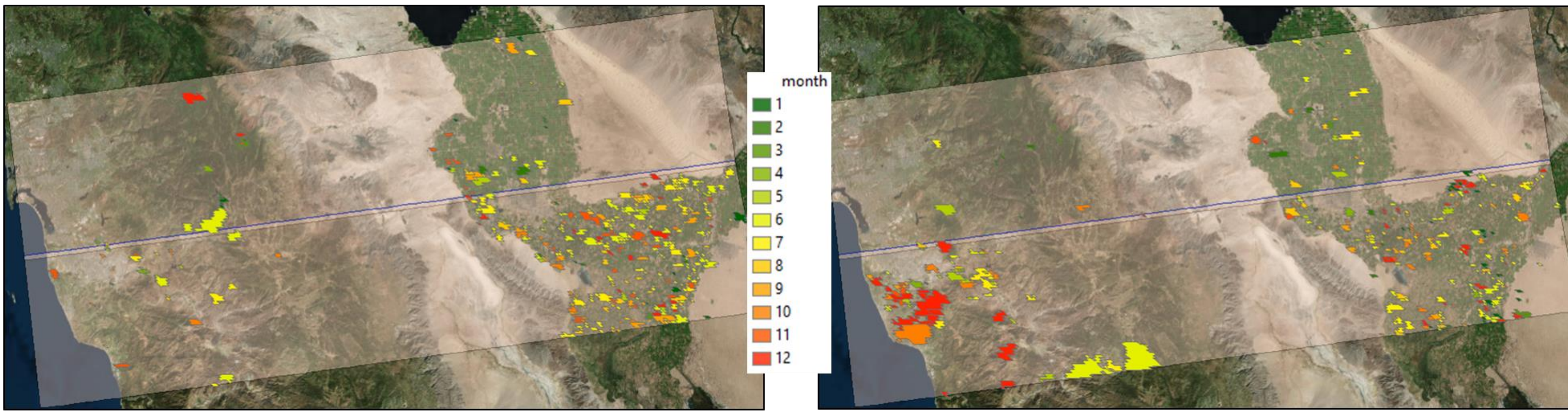
Monthly fire activities from March 2000 to February 2018 within the border regions



Case Study

Find burning activities such as agricultural burns in the US-Mexico border area

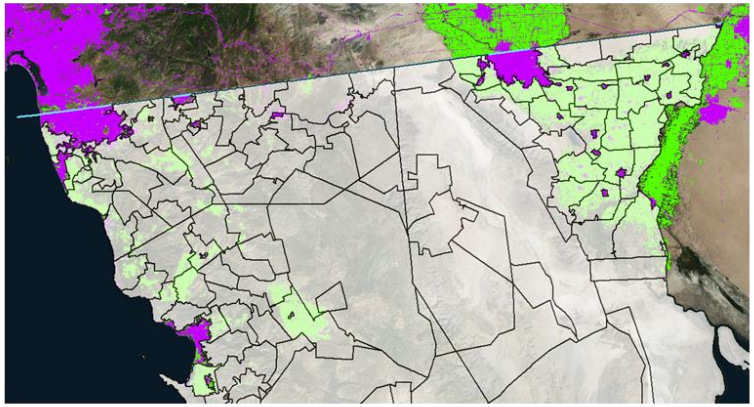
Fire Events



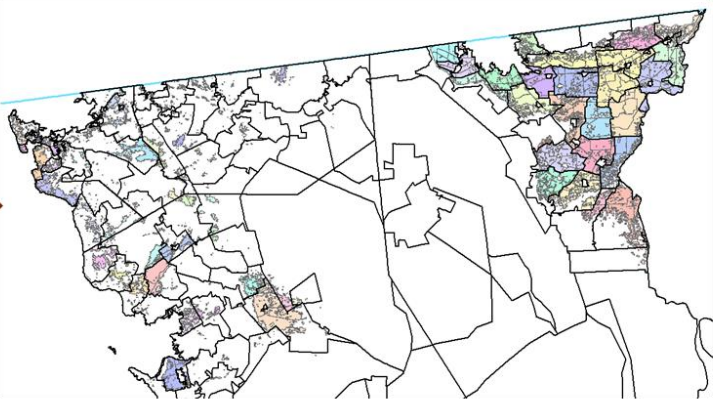
2016 **2017**

Case Study

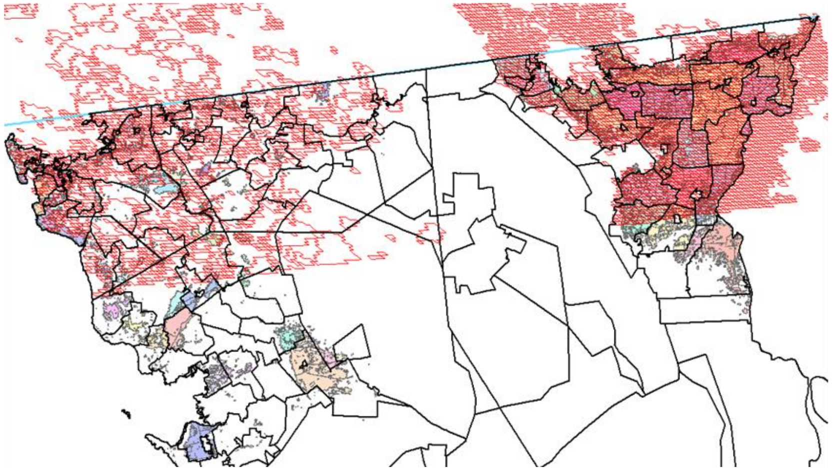
Summarize agricultural burn events (fires on crop fields) by INEGI rural areas in BC



Overlay crop fields with rural areas



Dissolve crop fields by rural areas

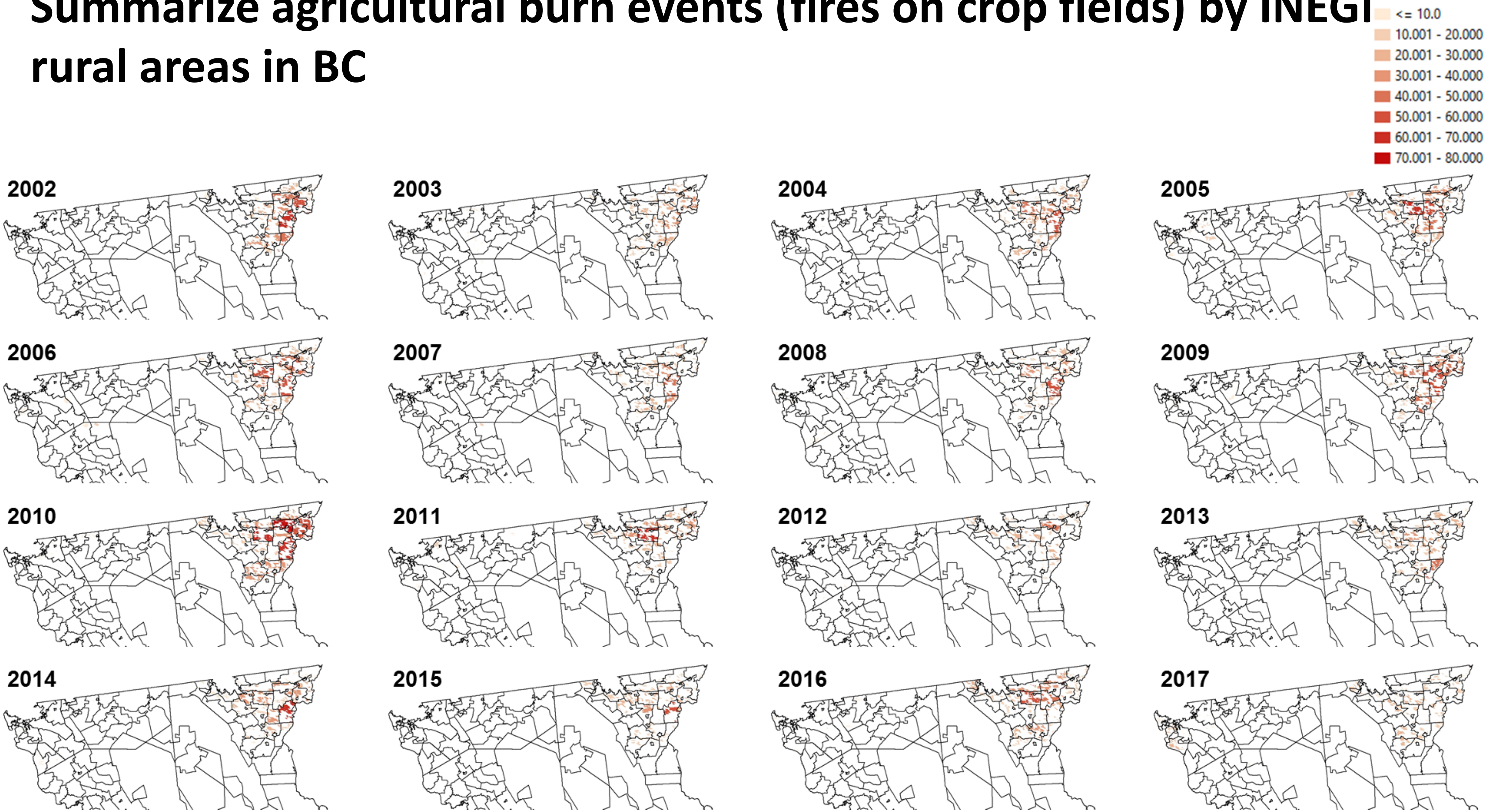


Overlay Crop Fields by Rural Areas with Fire Activities

Calculate total areas of crop field burned by rural areas

Case Study

Summarize agricultural burn events (fires on crop fields) by INEGI rural areas in BC

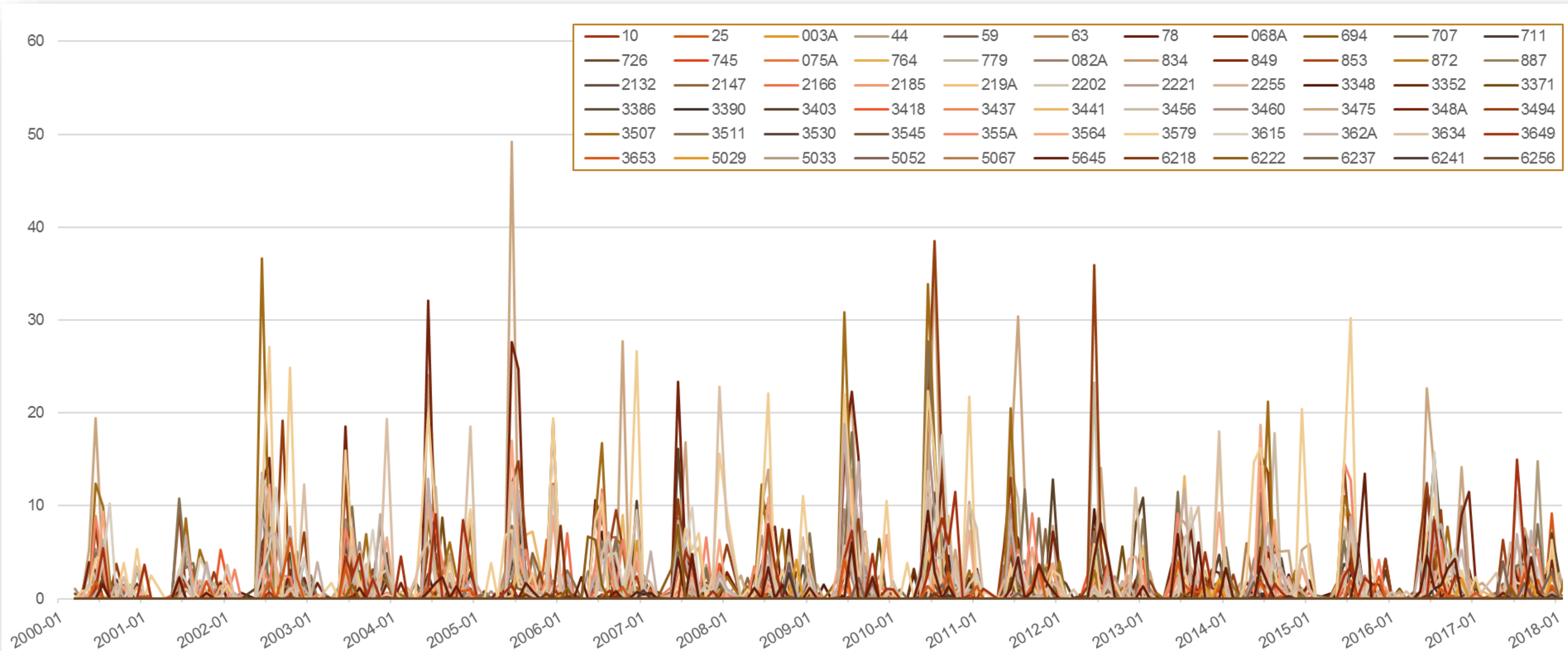


Yearly total area of crop field with fire activities by INEGI rural area (km^2)

SAN DIEGO STATE UNIVERSITY
HDMA @SDSU
CALIFORNIA AIR RESOURCES BOARD
NSF

Case Study

Summarize agricultural burn events (fires on crop fields) by INEGI rural areas in BC

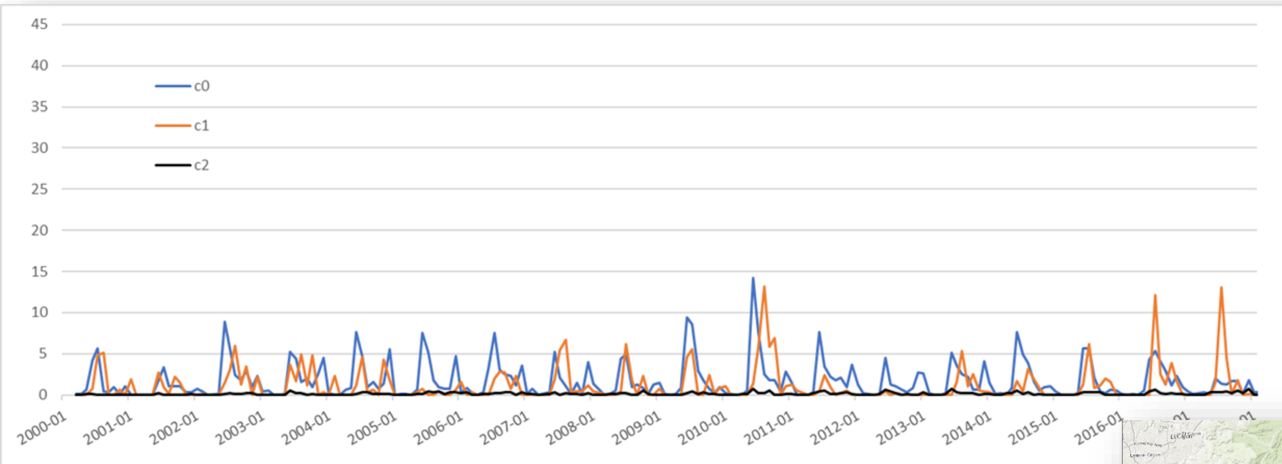


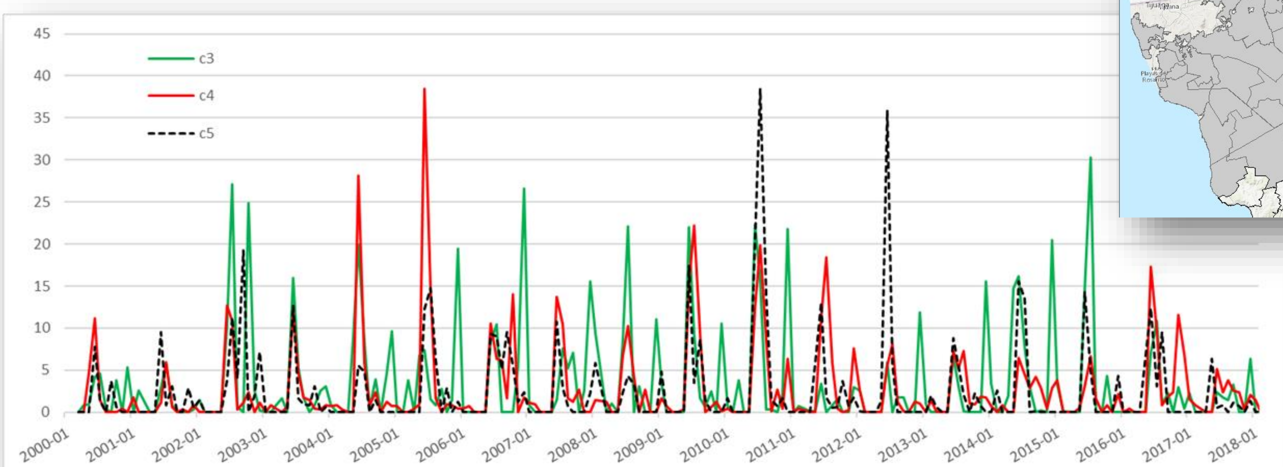
Monthly total area of crop field with fire activities by INEGI rural area (km^2)

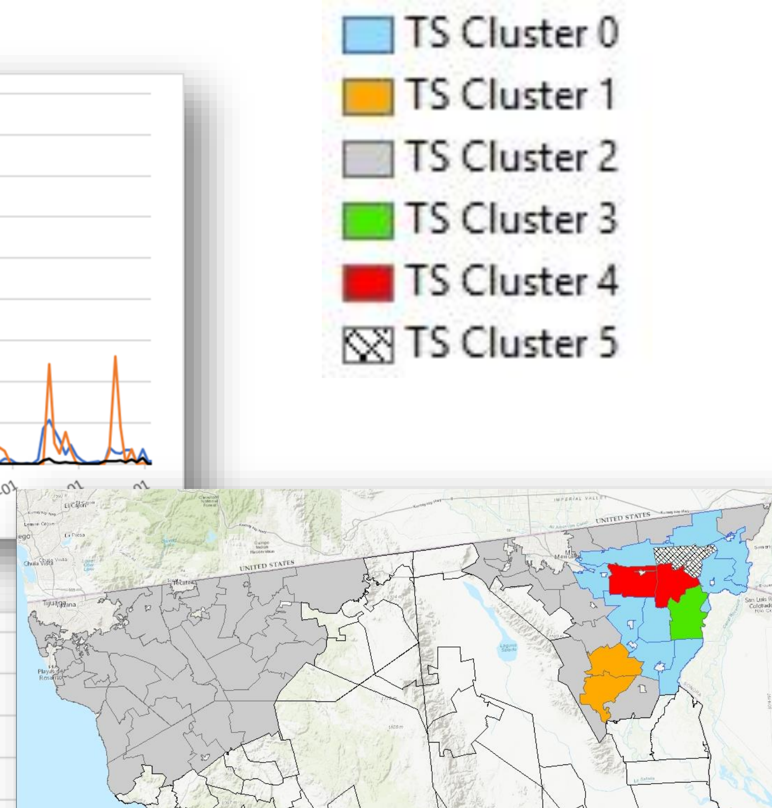
San Diego State University HDMA @SDSU CALIFORNIA AIR RESOURCES BOARD NSF

Case Study

Time series clustering to find distinct time-space patterns of crop fields with fire activities (Monthly)







San Diego State University

HDMA @SDSU

CALIFORNIA AIR RESOURCES BOARD

NSF

Future Works

- Verifying detected fire activities
- Explore non-agricultural urban fires
- Conduct a systematic performance test
- Apply ST-DBSCAN to other large event data
  - Crime Incidents
  - Social media
  - Transportation data

Acknowledgements

- This material is partially based upon work supported by **the California Air Resource Board (#16RD010),** *"Improving the CalEnviroScreen score at the US-Mexico border"* and the **National Science Foundation under Grant No. 1634641**, IMEE project titled *"Integrated Stage-Based Evacuation with Social Perception Analysis and Dynamic Population Estimation*". Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

- The Center for Human Dynamics at the Mobile Age at SDSU
- Dr. May Yuan, University of Texas at Dallas

SAN DIEGO STATE UNIVERSITY

HDMA @SDSU

CALIFORNIA AIR RESOURCES BOARD

NSF

Thank you for participation

Any Questions?

Atsushi Nara
anara@sdsu.edu