The Patterns of Conflict Emergence (PaCE)

Thomas Chadefaux
Trinity College Dublin

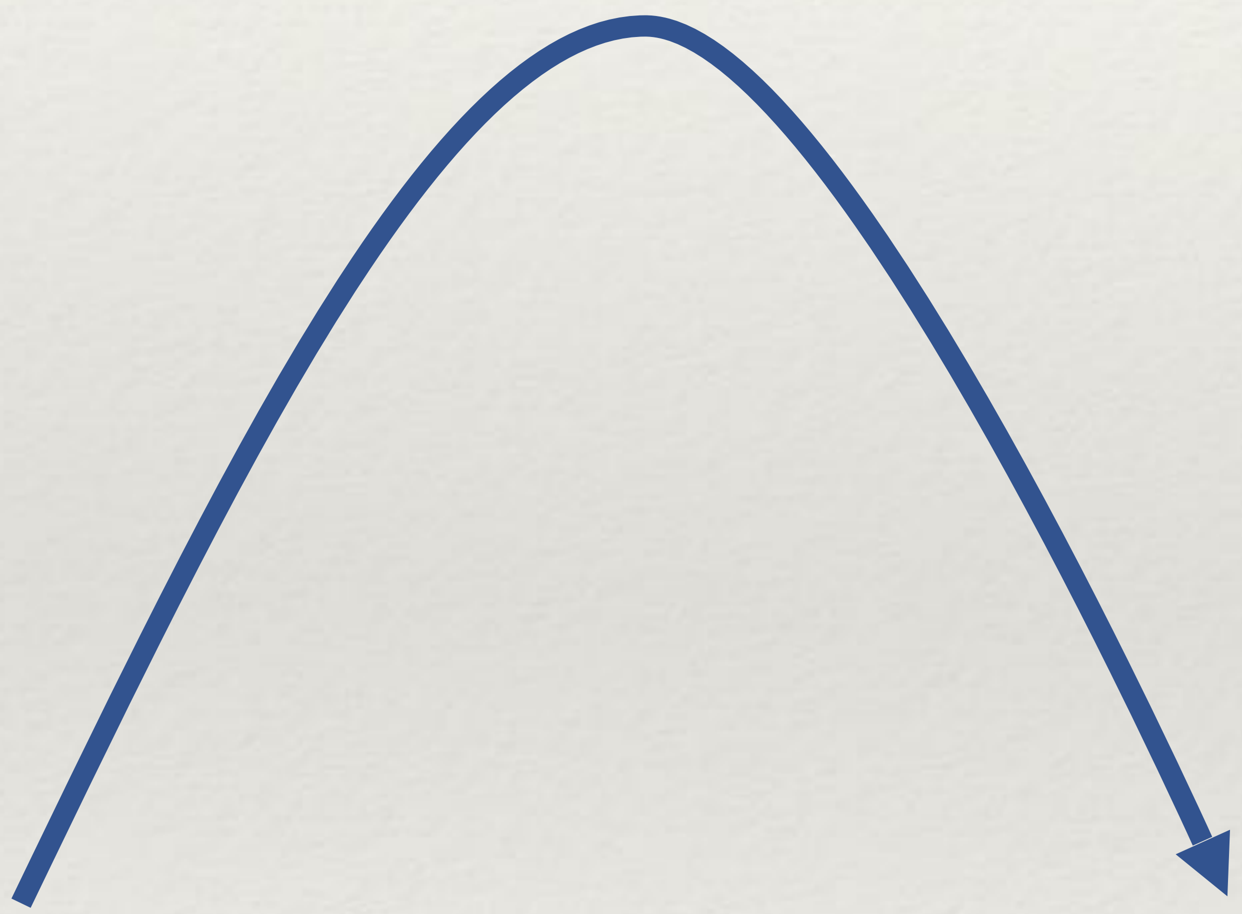
15 July 2022

PaCE

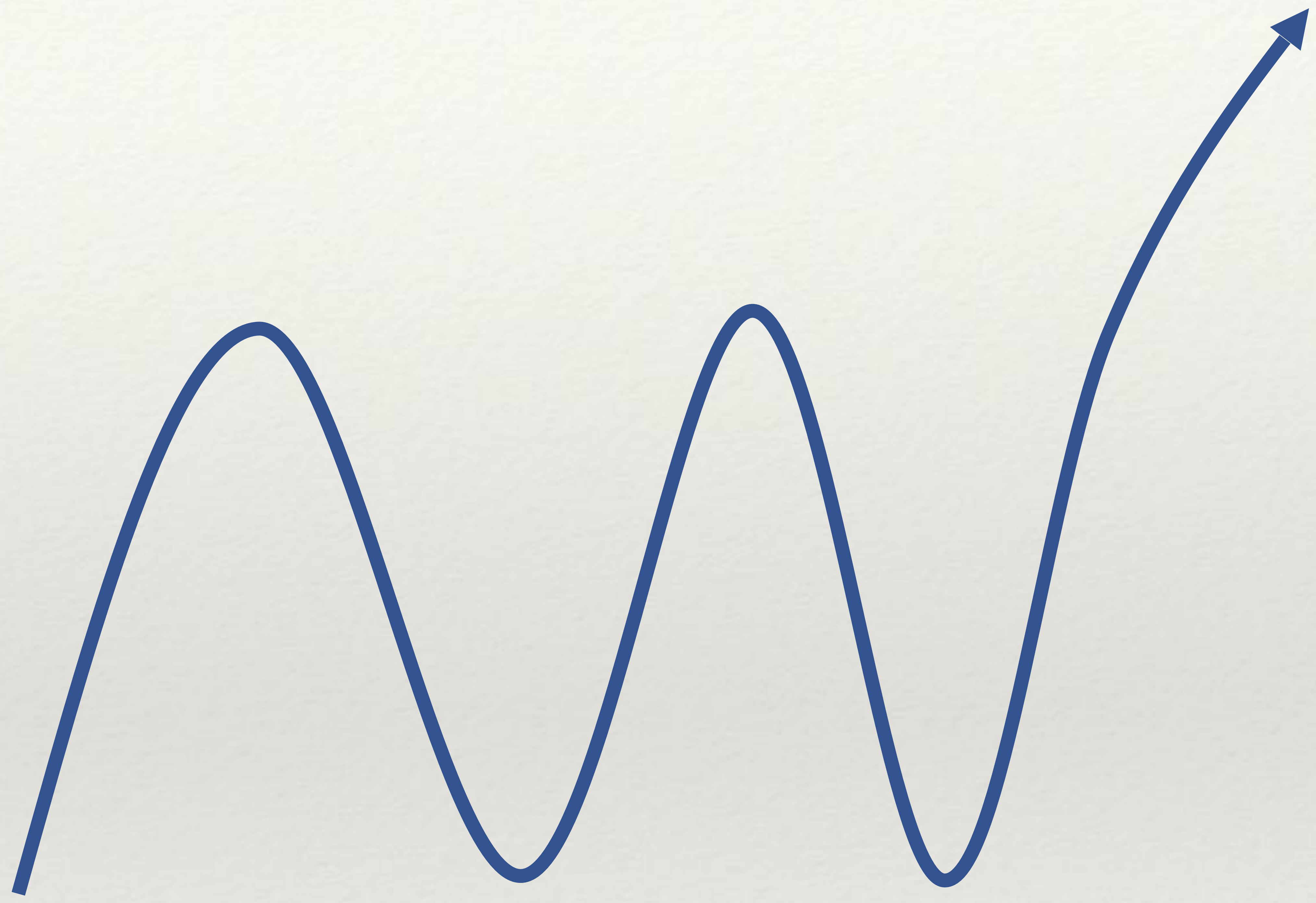
- ERC 2022-27
- Trinity College Dublin, Department of Political Science
- Team:
  - Political Science
  - Economics
  - Physics
  - Machine Learning

PaCE: Project overview

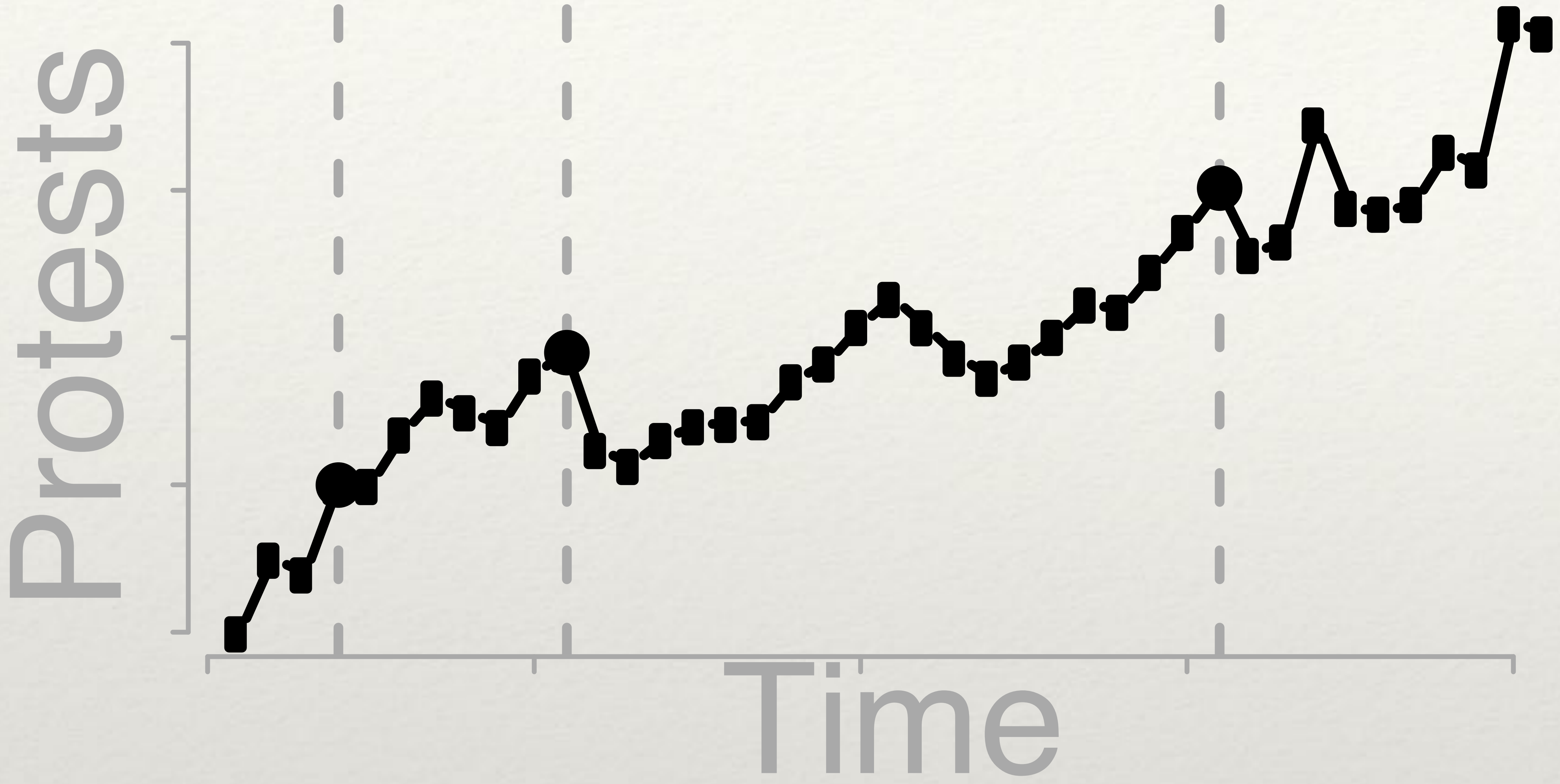
Stories



Tragedy



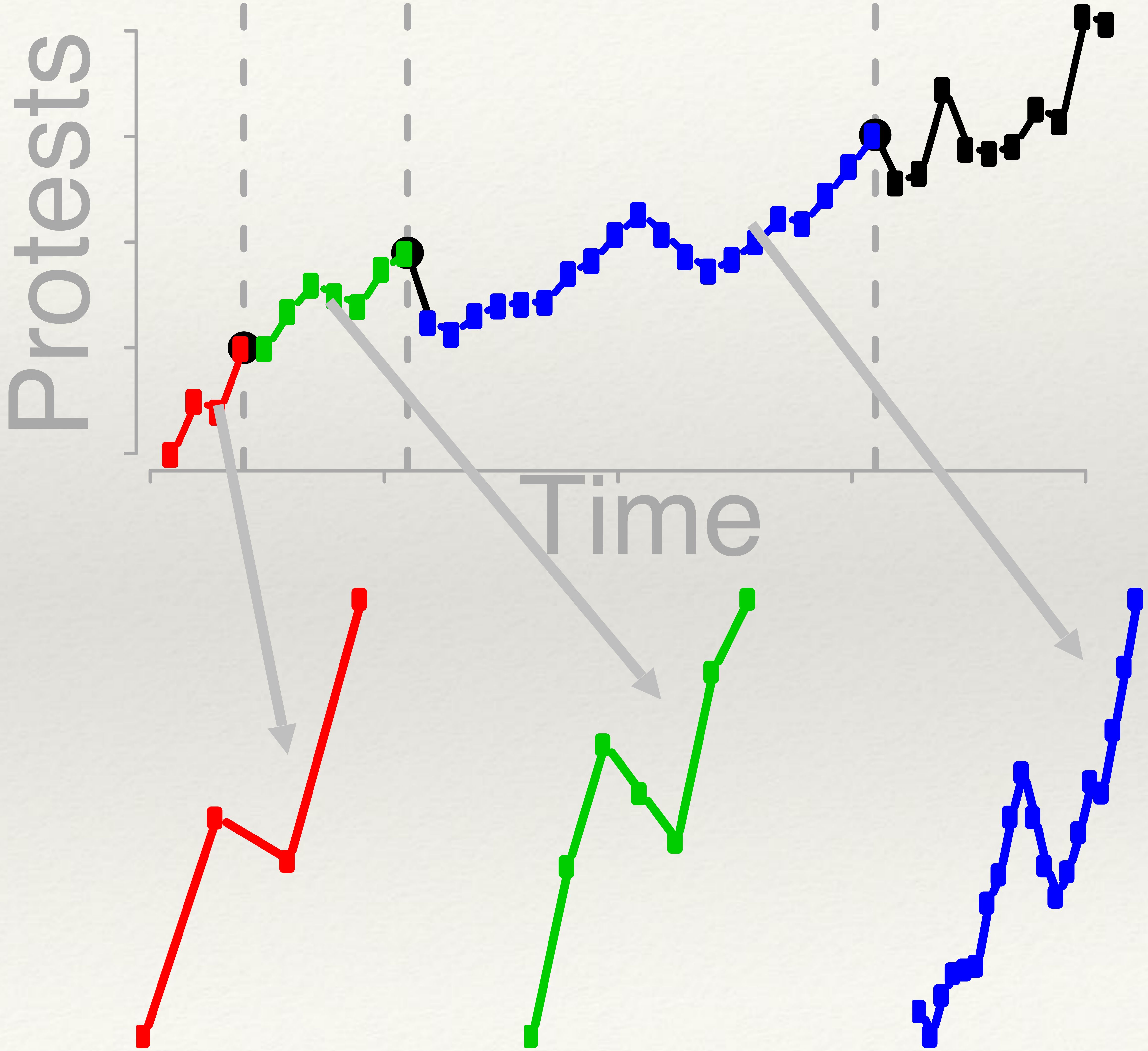
Seven point story



Protests
Time

Observations- as units of analysis

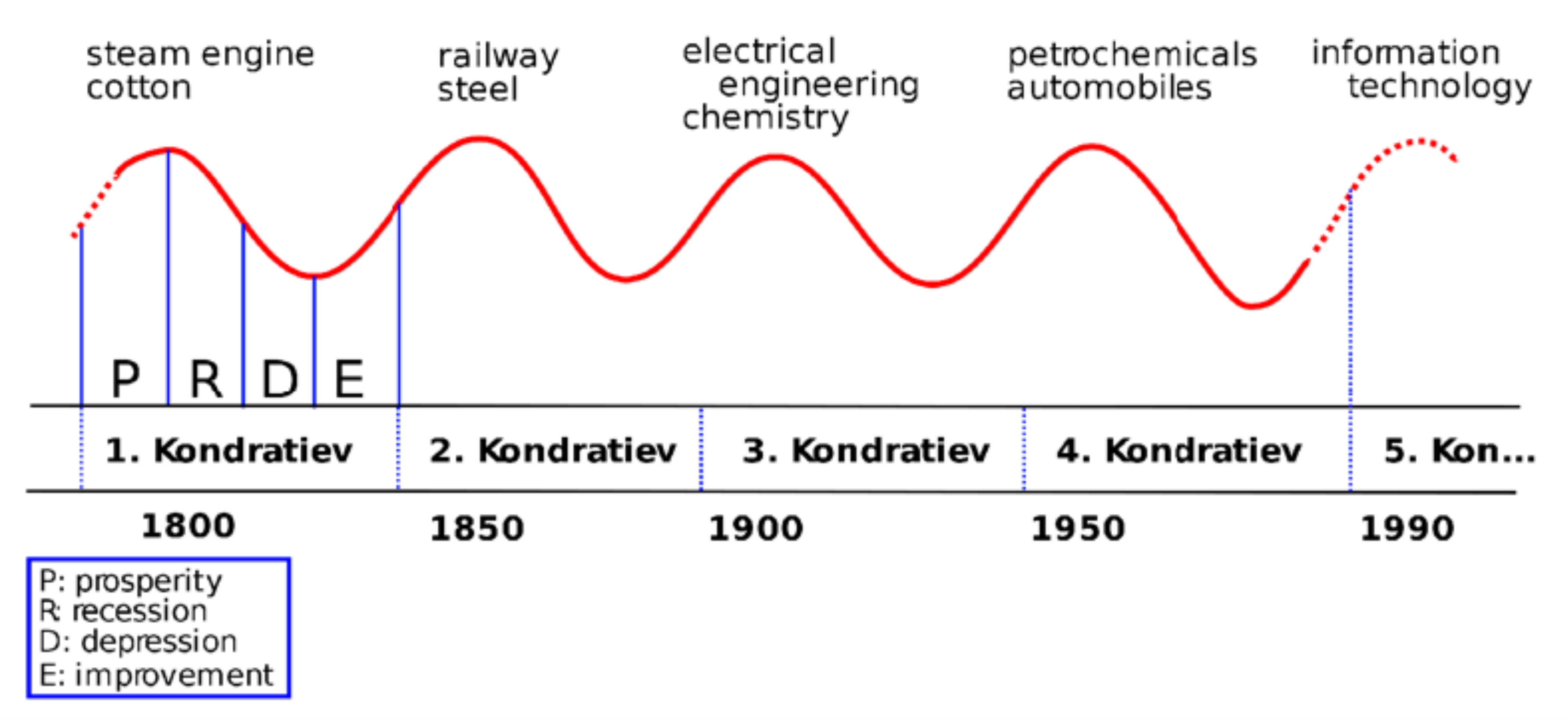
Sequences as units of analysis

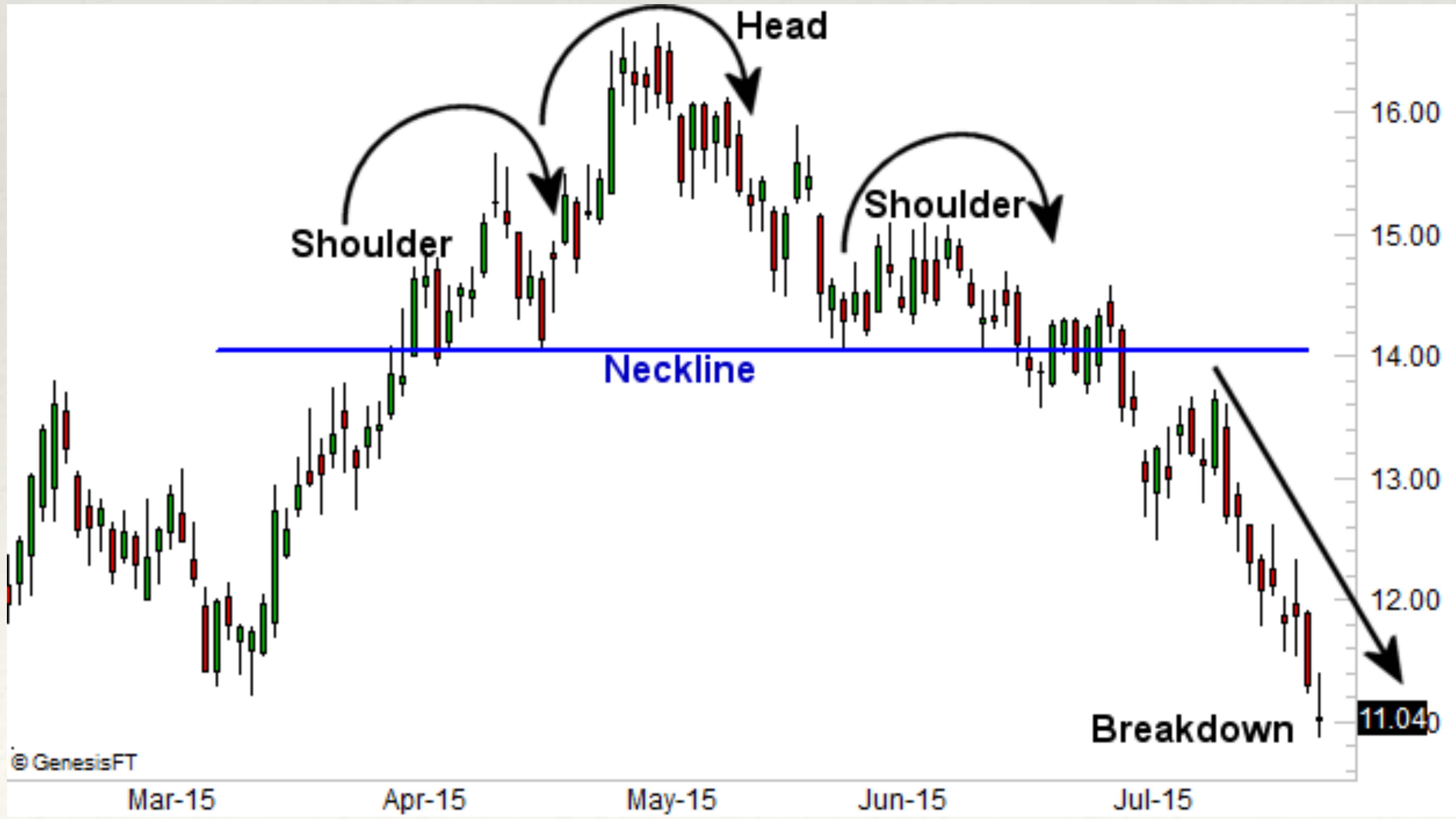


Sequences in other fields

In social sciences

- Patterns
  - Marx: Repeating motif over time
  - Kondratiev waves: alternating intervals of high sectoral growth and intervals of relatively slow growth
- No patterns
  - Efficient market hypothesis
  - Richardson





PaCE: One Big Question

Does history repeat itself?

PaCE: Three questions

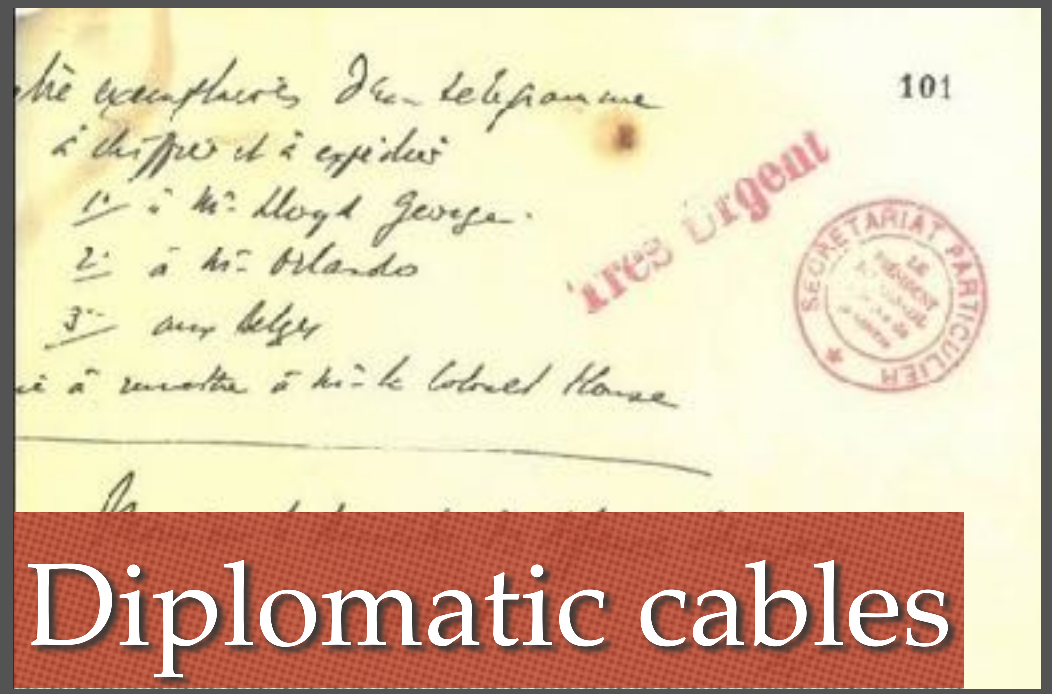
1. Are there patterns in the temporal sequences that lead to conflict?
2. Can we use these motifs to predict conflict?
3. Can we cluster these patterns to create new / inform existing theories of conflict

How?

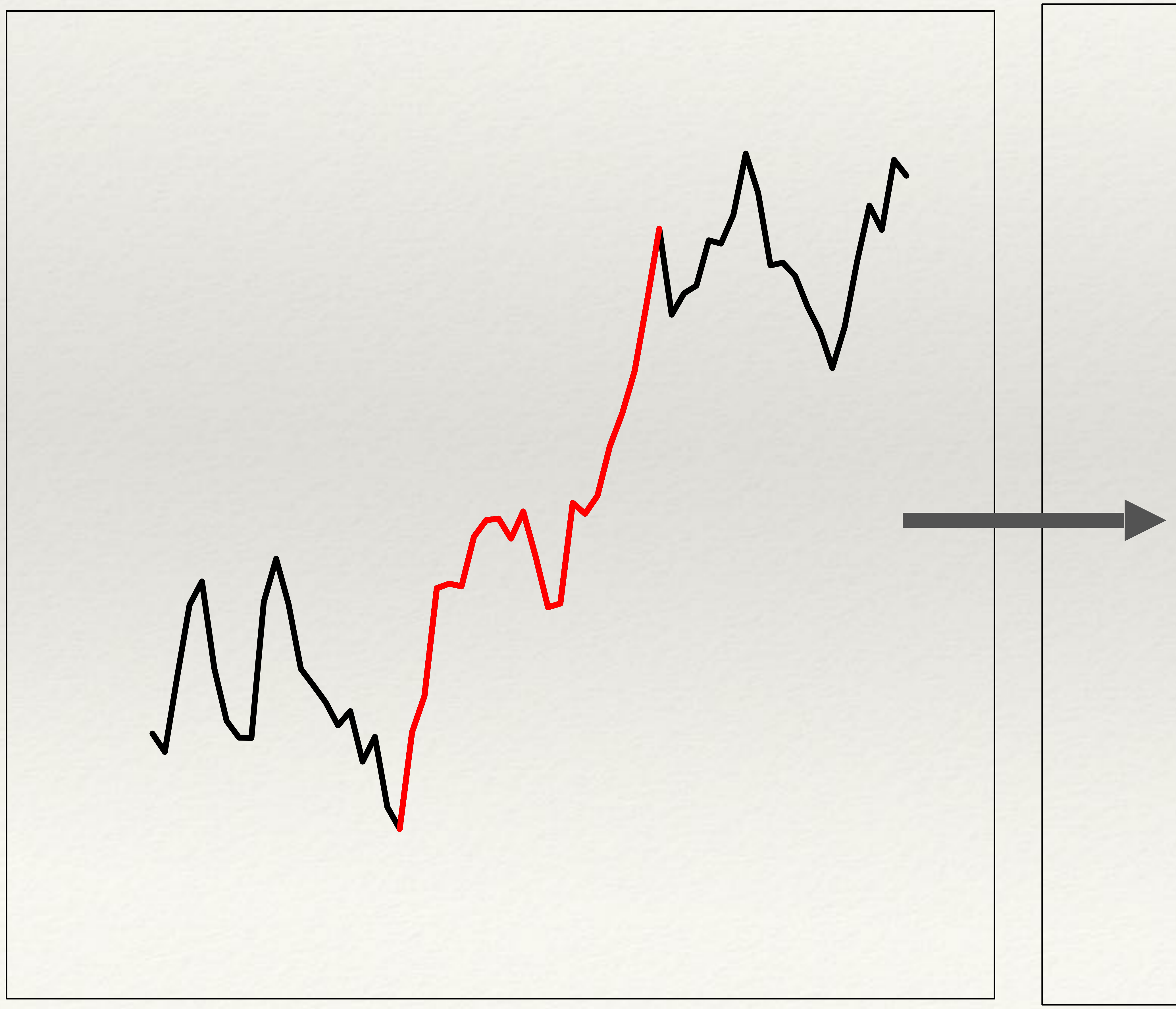
Predict…

- Interstate and Civil wars/ conflict
- Migration flows
- Protests
- Violence
- and more

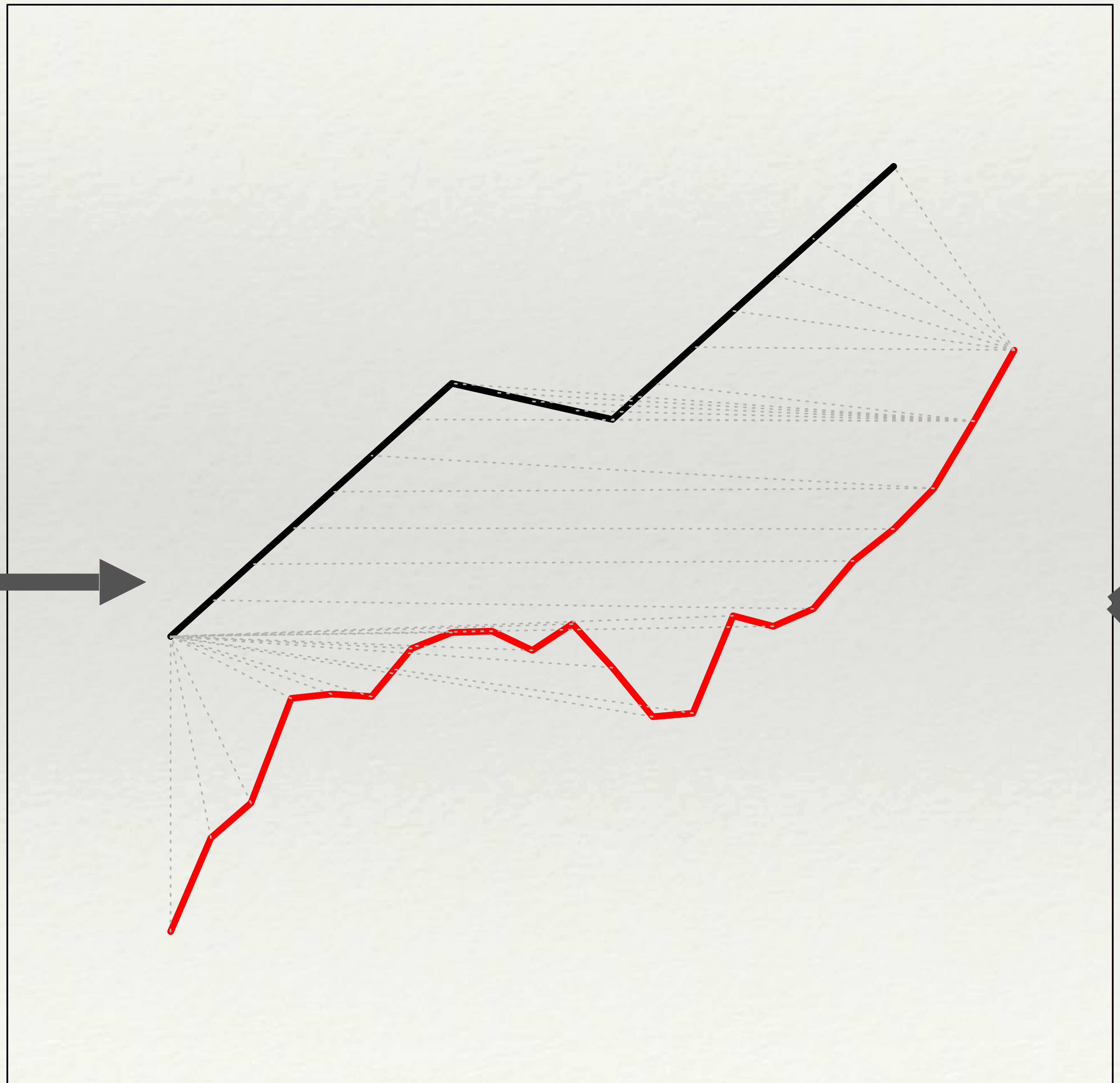
… using



Data acquisition
& processing



Distance
measures

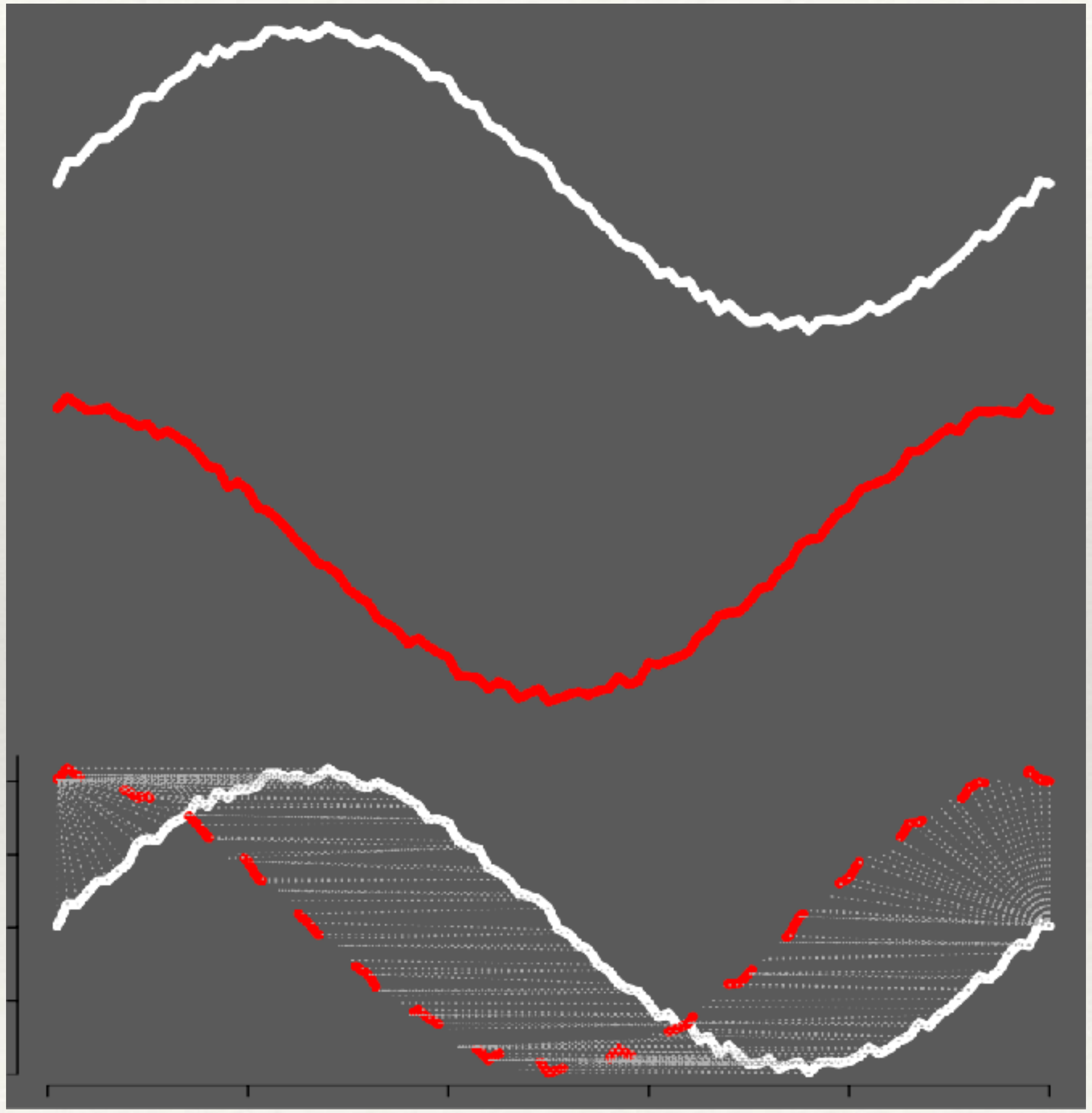


Classification

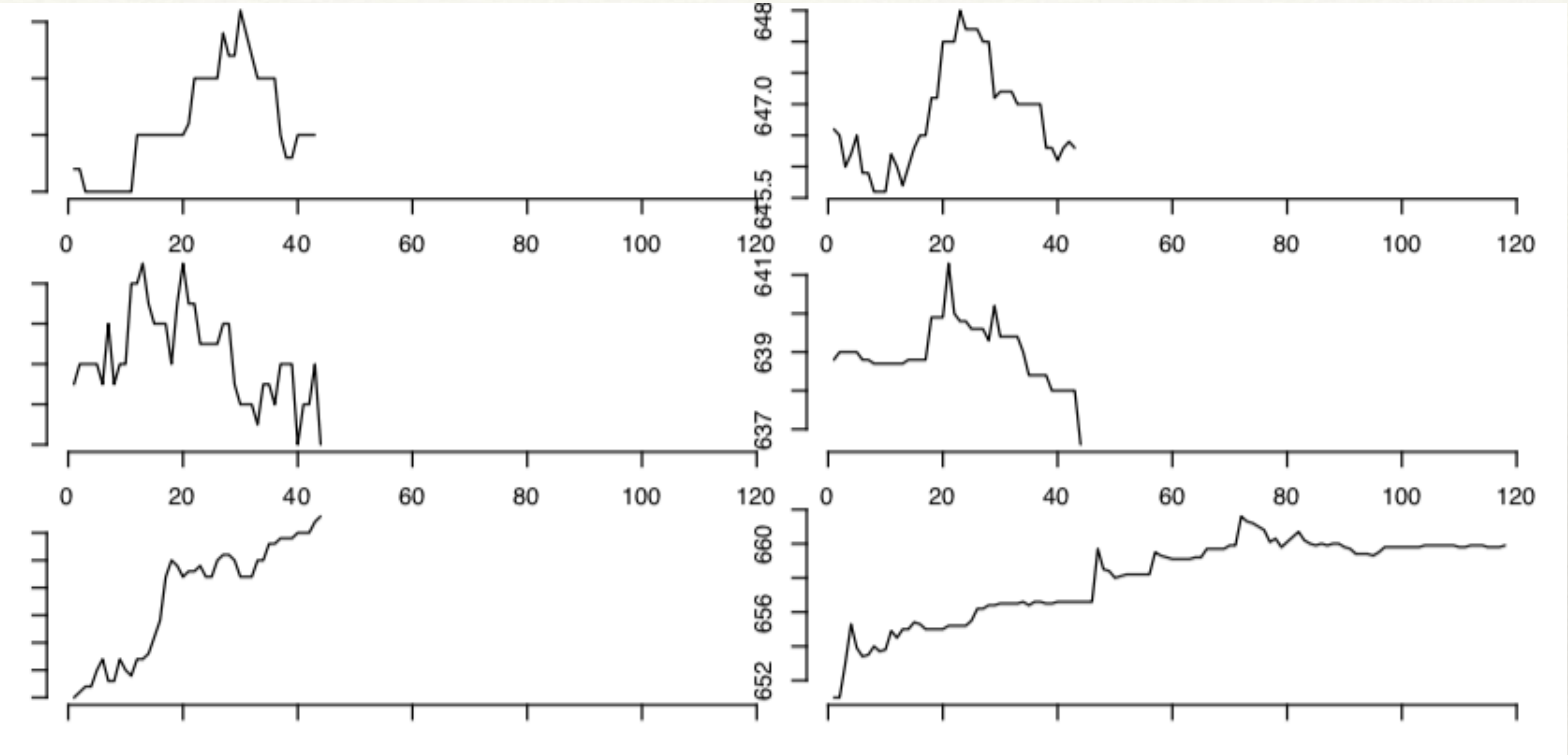
Conflict

Peace

Distance measures: An example

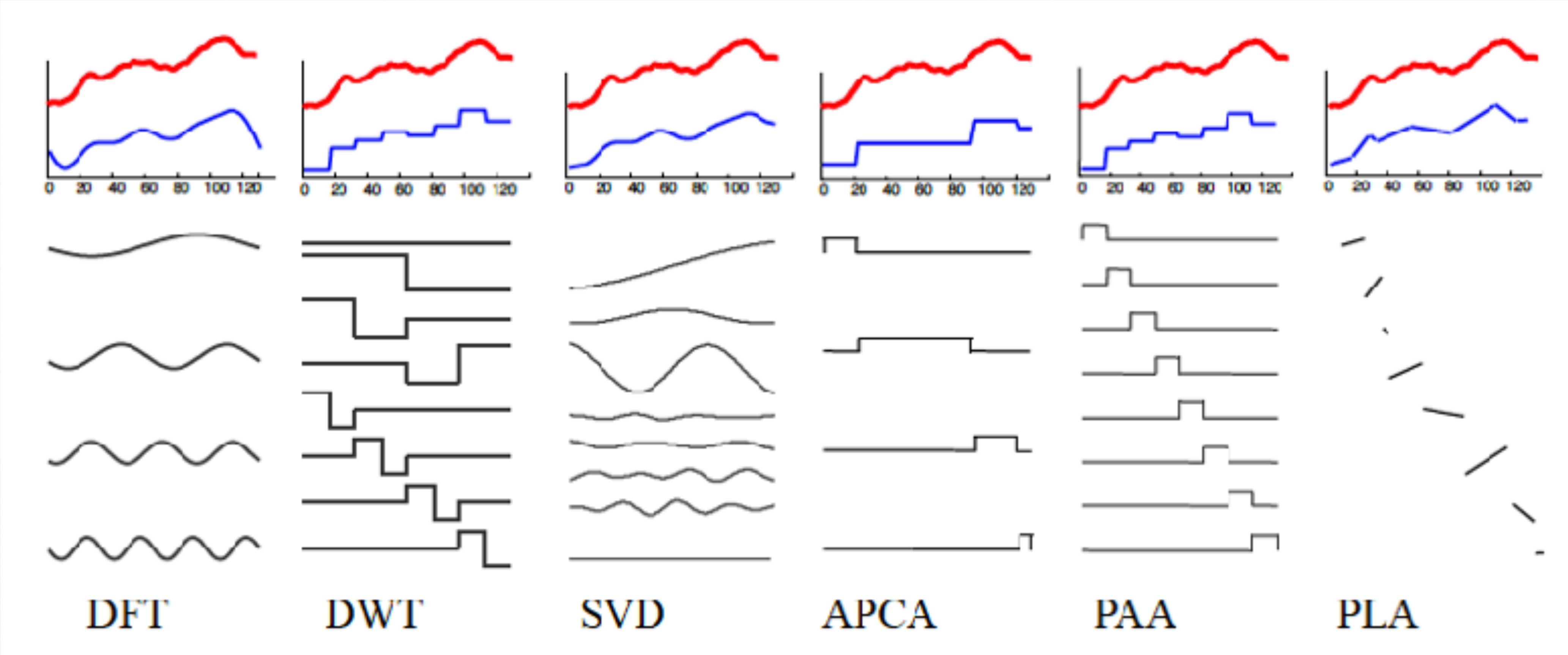


Dynamic Time Warping

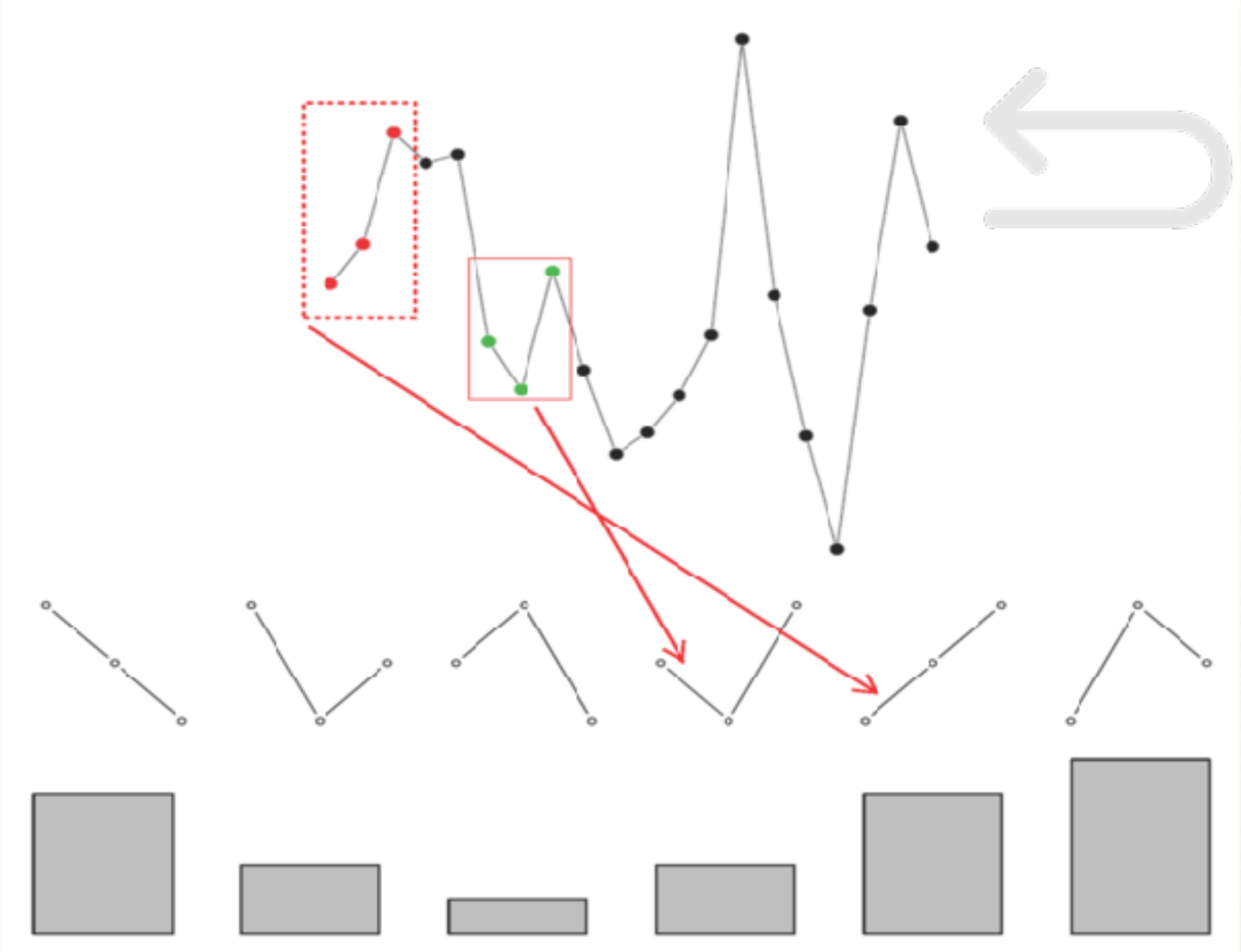


0
20
40
60
80
100
120
645.5
647.0
648
637
639
641
652
656
660

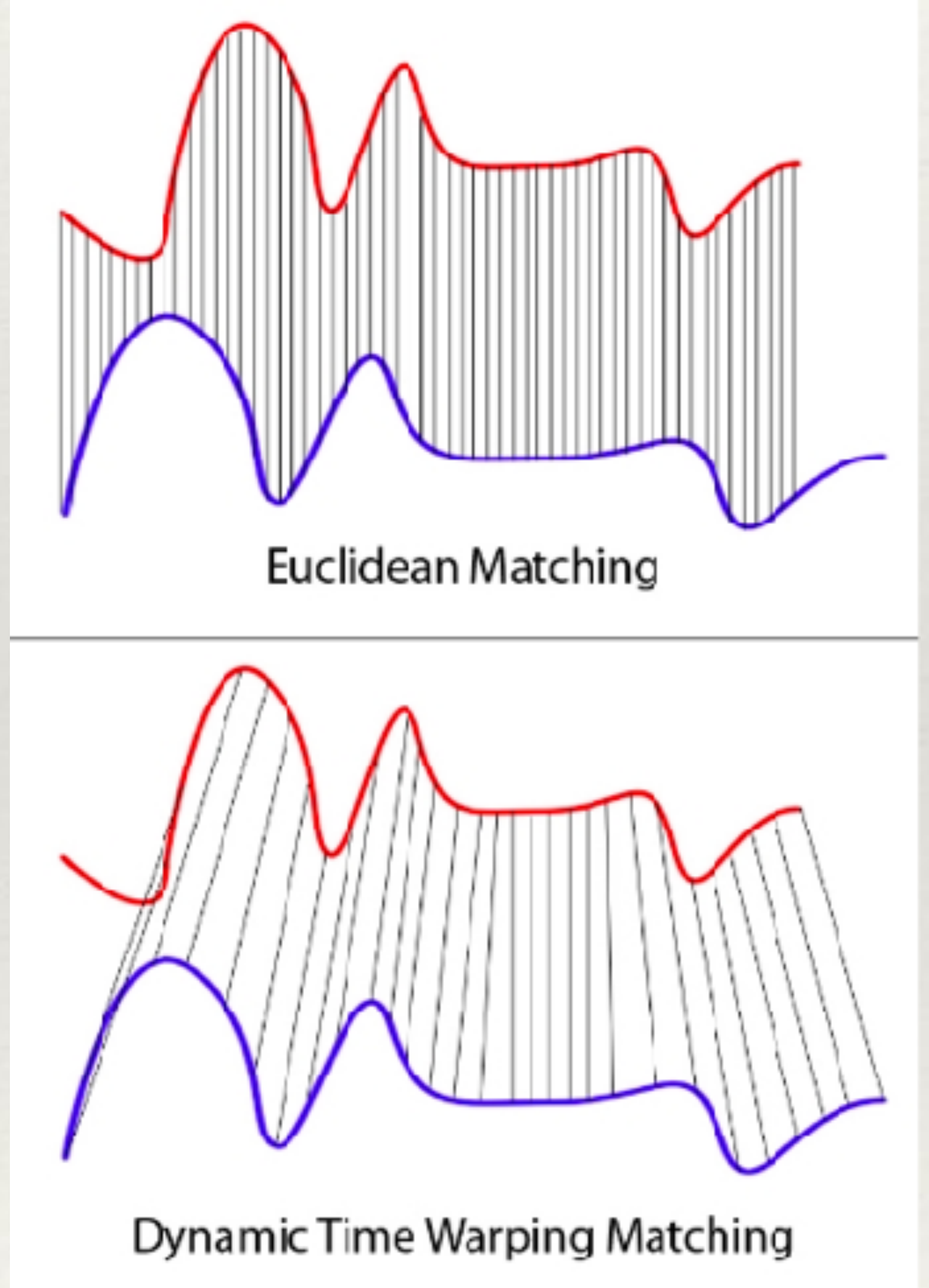
Methodology: Distance Measures



Dimensionality Reduction



Permutation Distribution Clustering



Dynamic Time Warping

Clustering

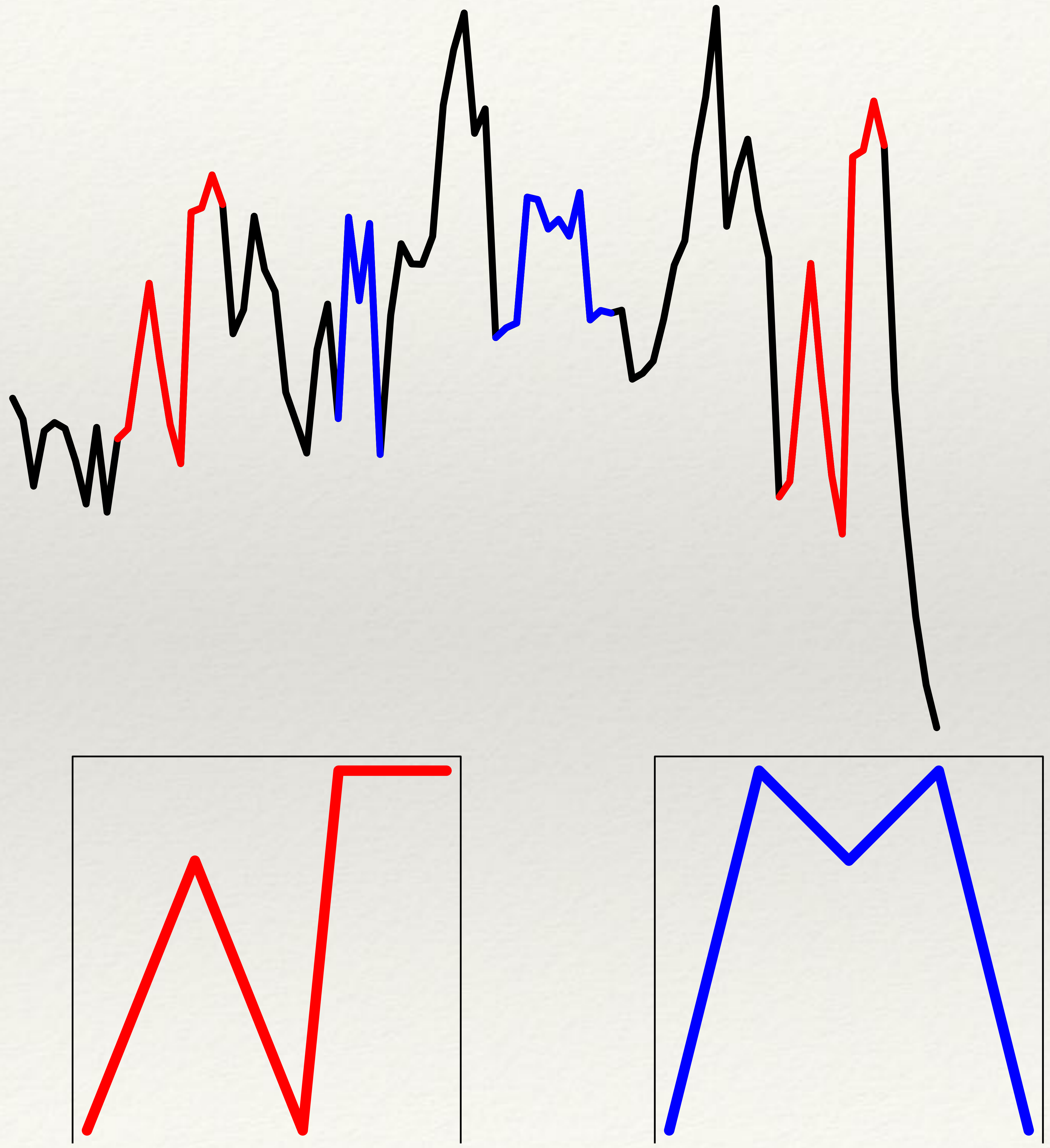
Predictions

Forecasting

- Backward forecasting
- Live forecasting

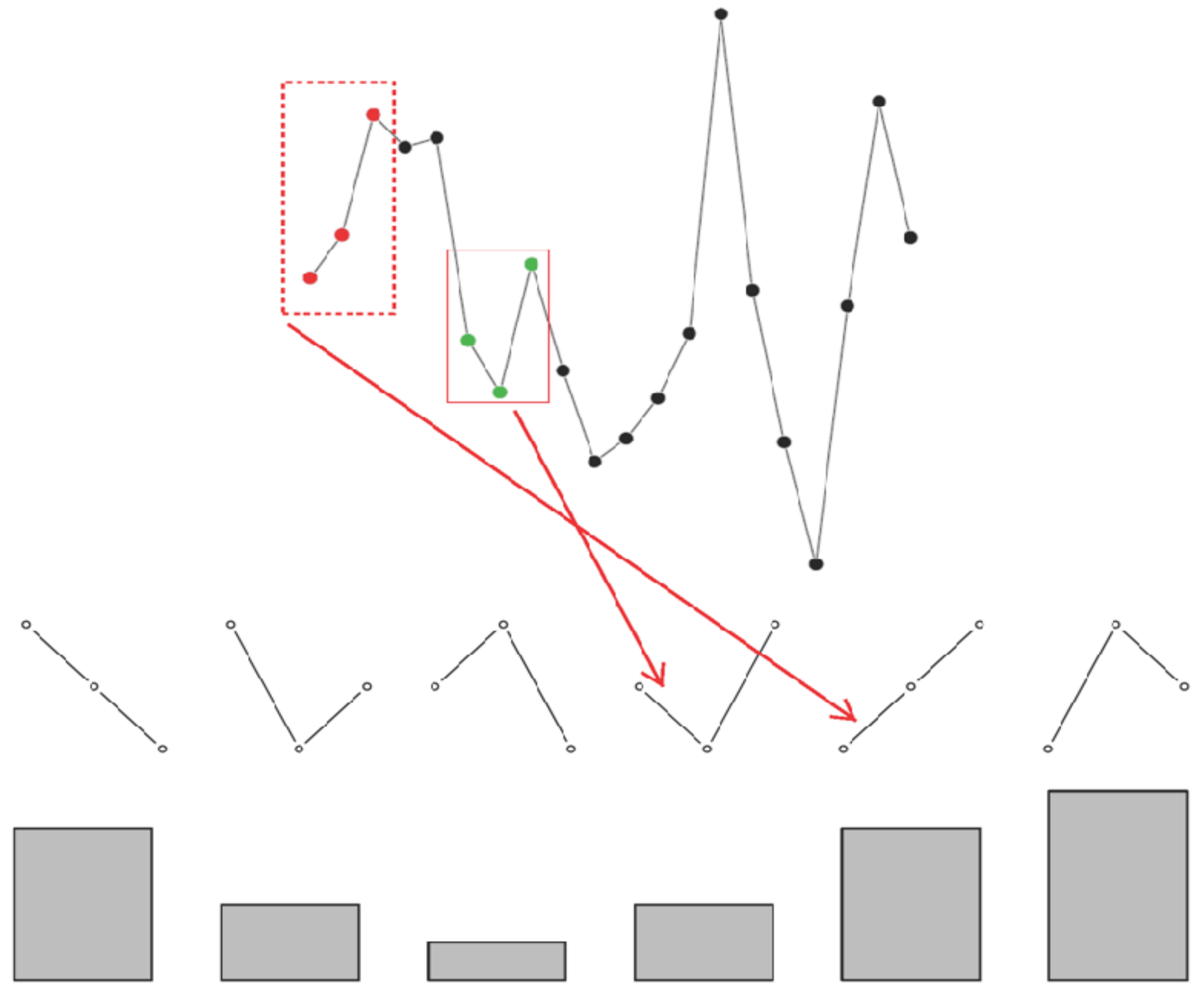
Theoretical implications 1

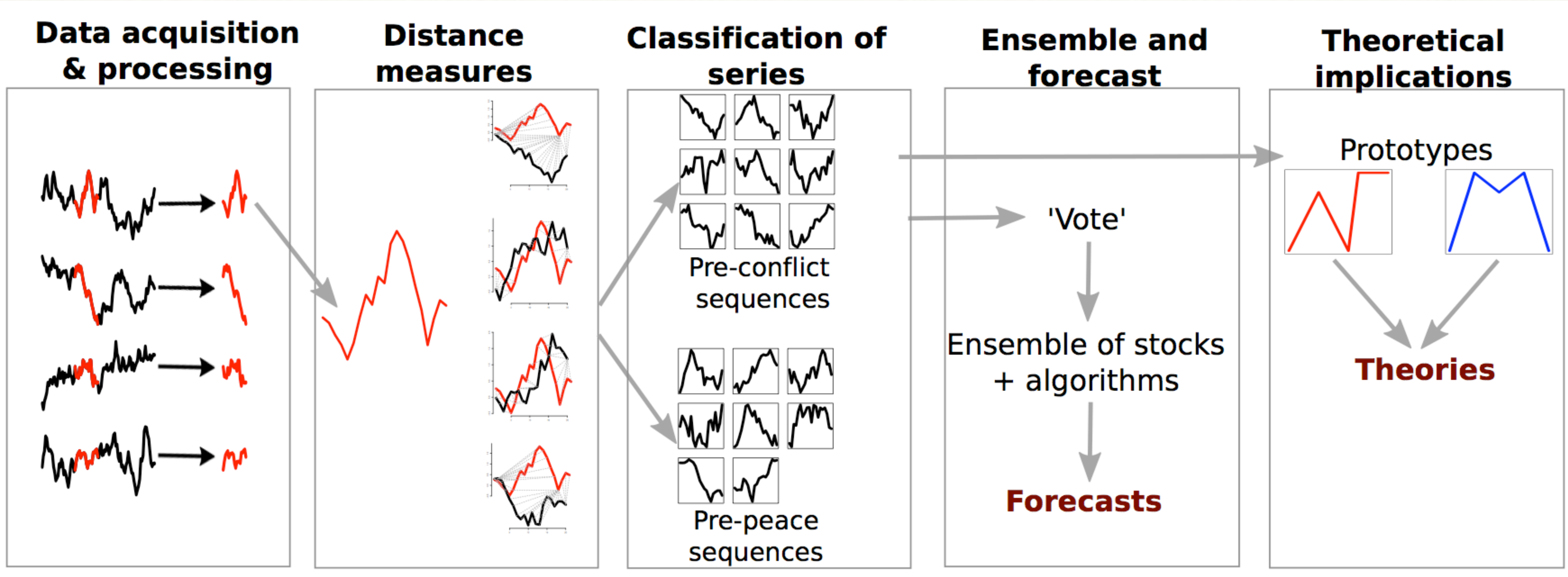
Prototypes for theory building



Theoretical implications 2

Fundamental limits to the predictability of conflict events?

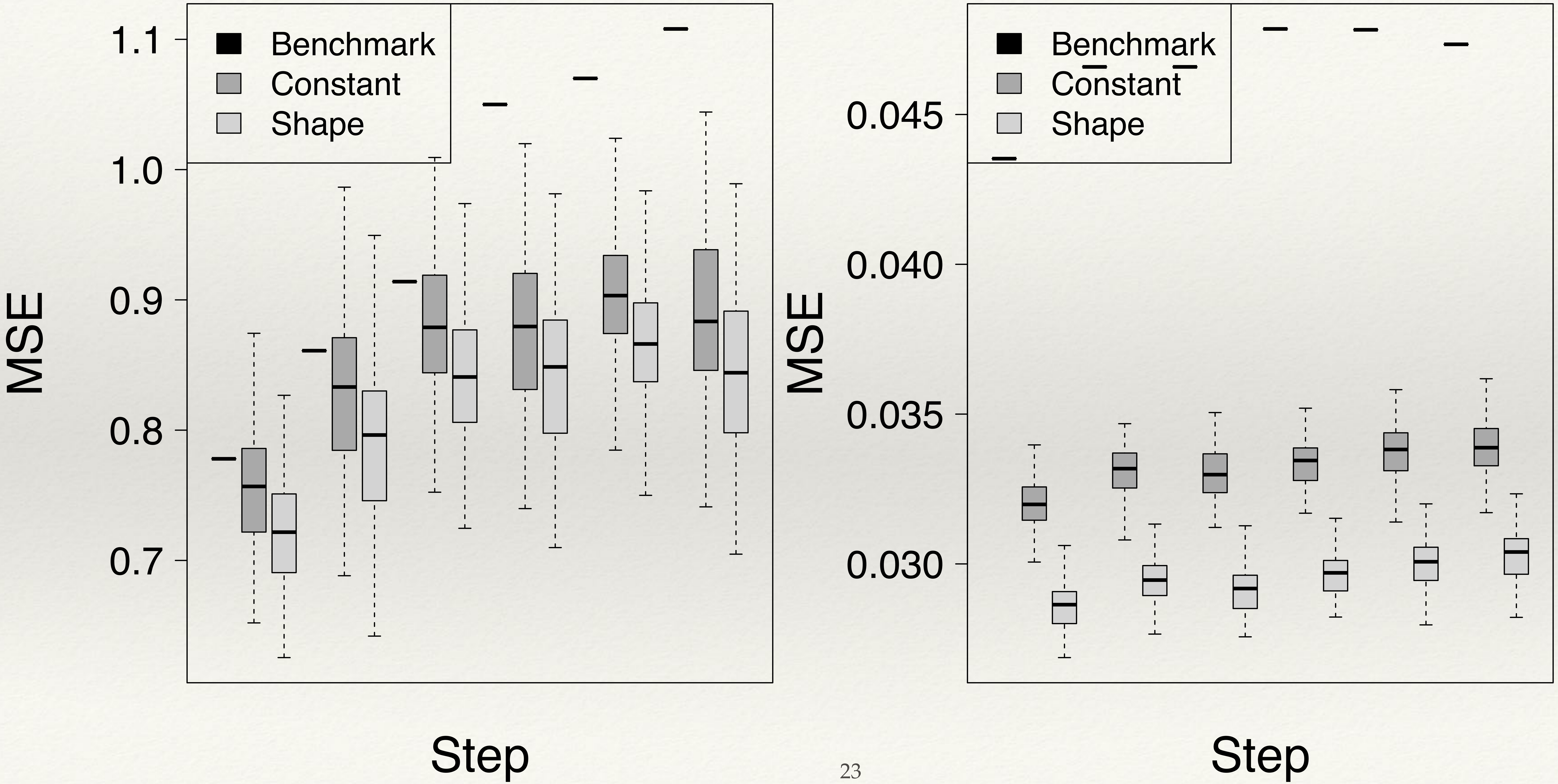




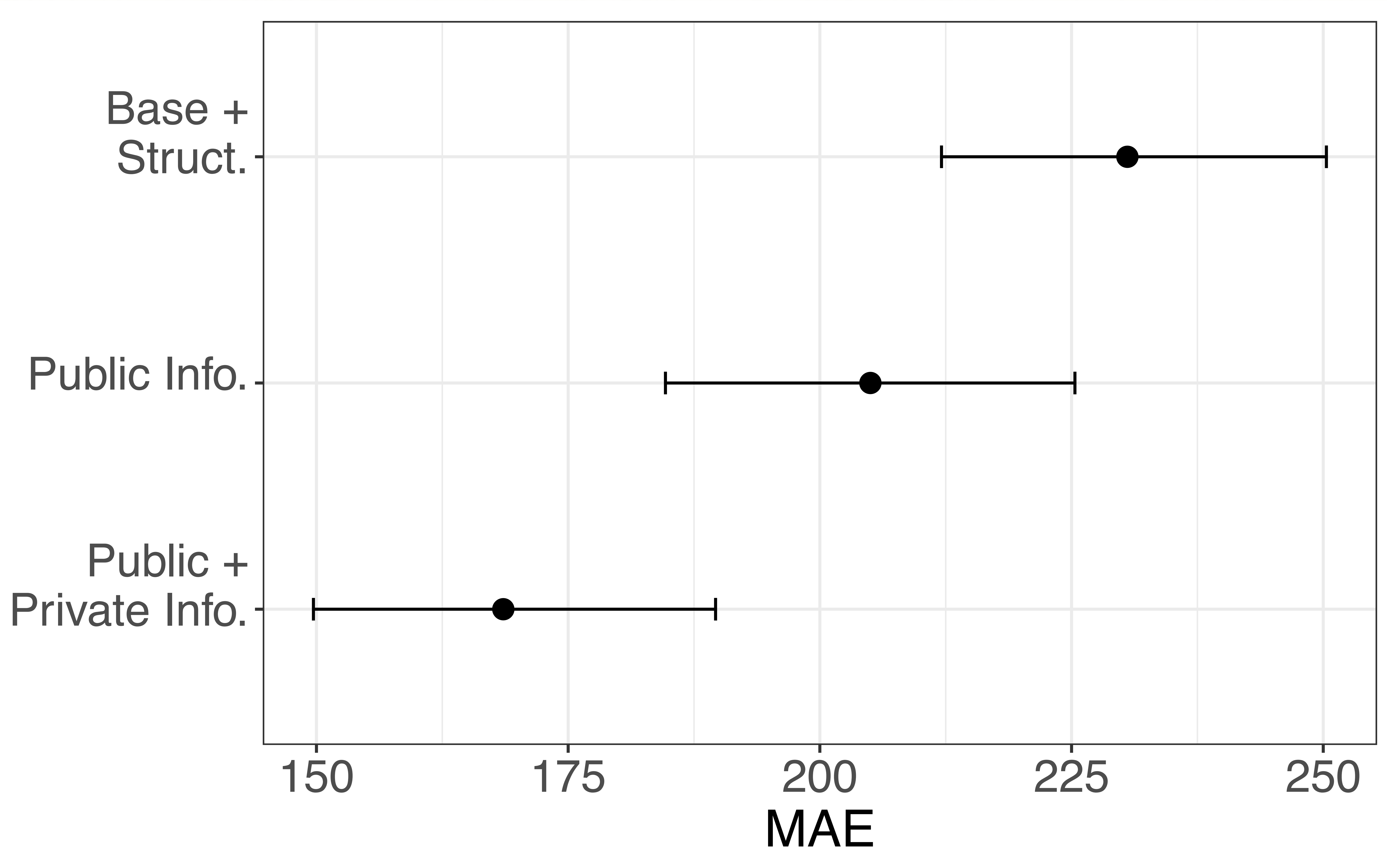
Data acquisition & processing
Distance measures
Classification of series
Ensemble and forecast
Theoretical implications
Pre-conflict sequences
Pre-peace sequences
Prototypes
'Vote'
Ensemble of stocks + algorithms
Forecasts
Theories

Some Preliminary results

Views Competition



Private and Public Information



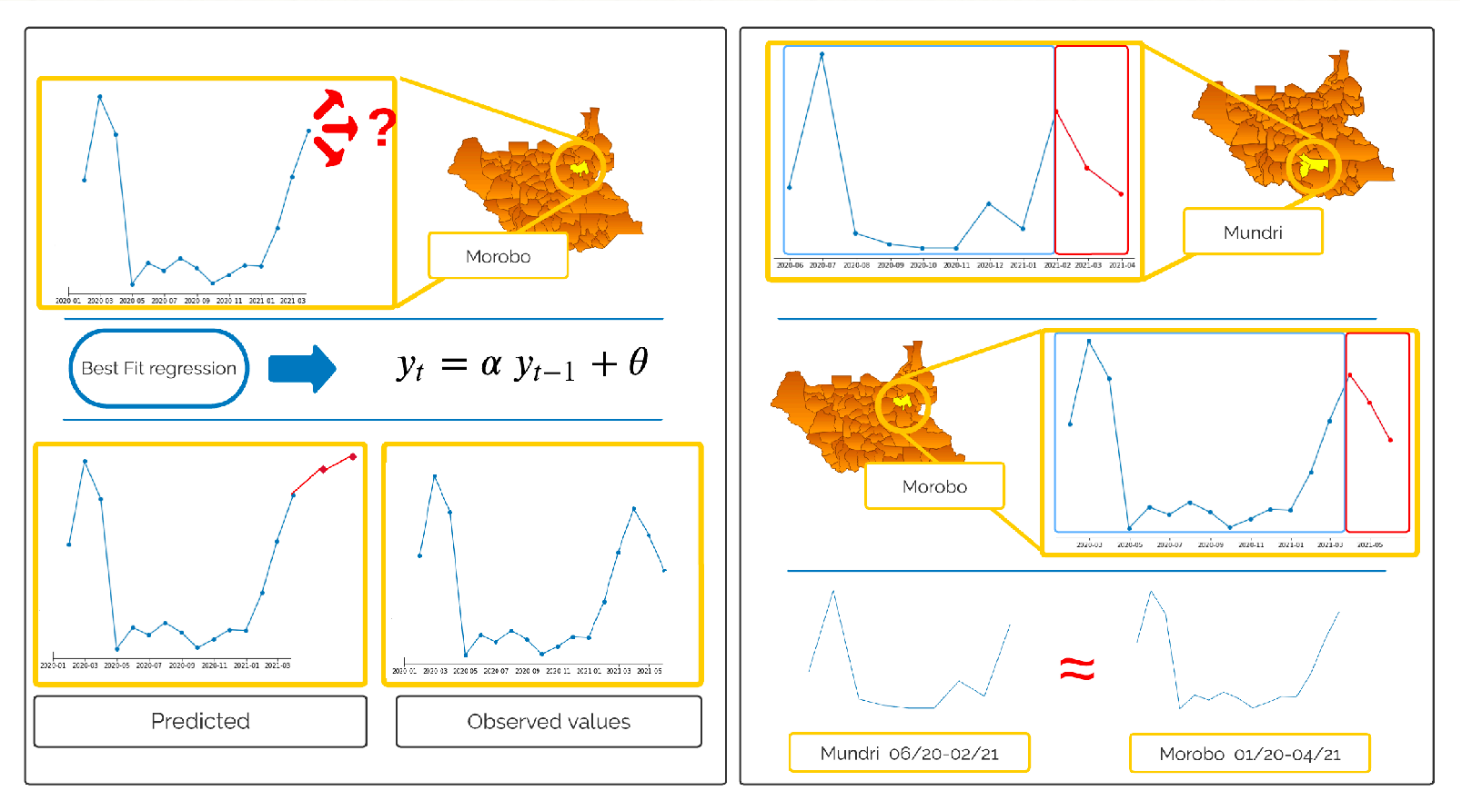
Initial projects (ongoing)

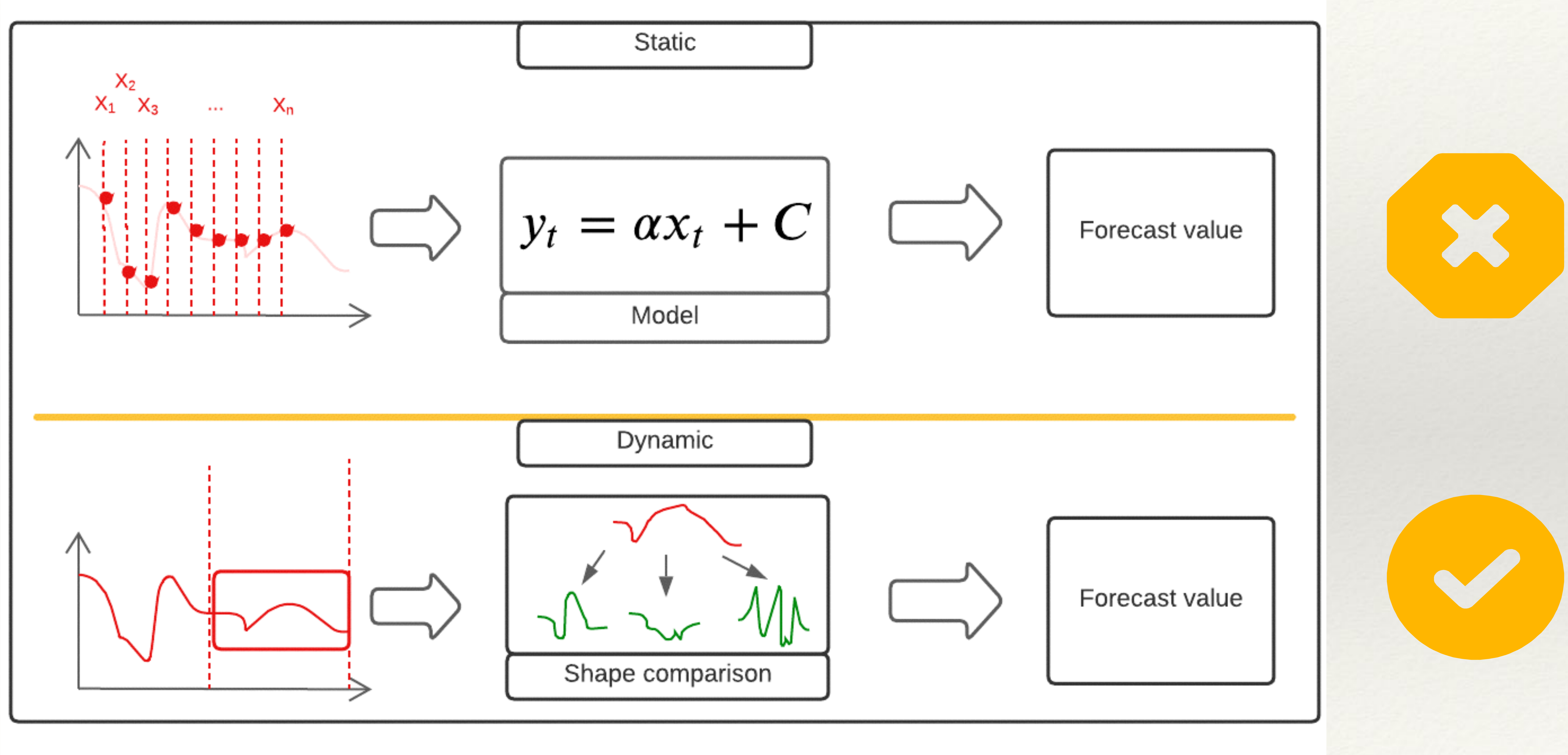
1. Temporal patterns in migration flows
2. Temporal patterns in protest data
3. Augmentation for conflict time series
4. Reducing uncertainty in conflict events using satellite data
5. Dynamic Synthetic Controls (if time allows)

Project 1
Temporal patterns in migration flows

Thomas Schincariol & Thomas Chadefaux

Example : Migration flow in Morobo - South Sudan

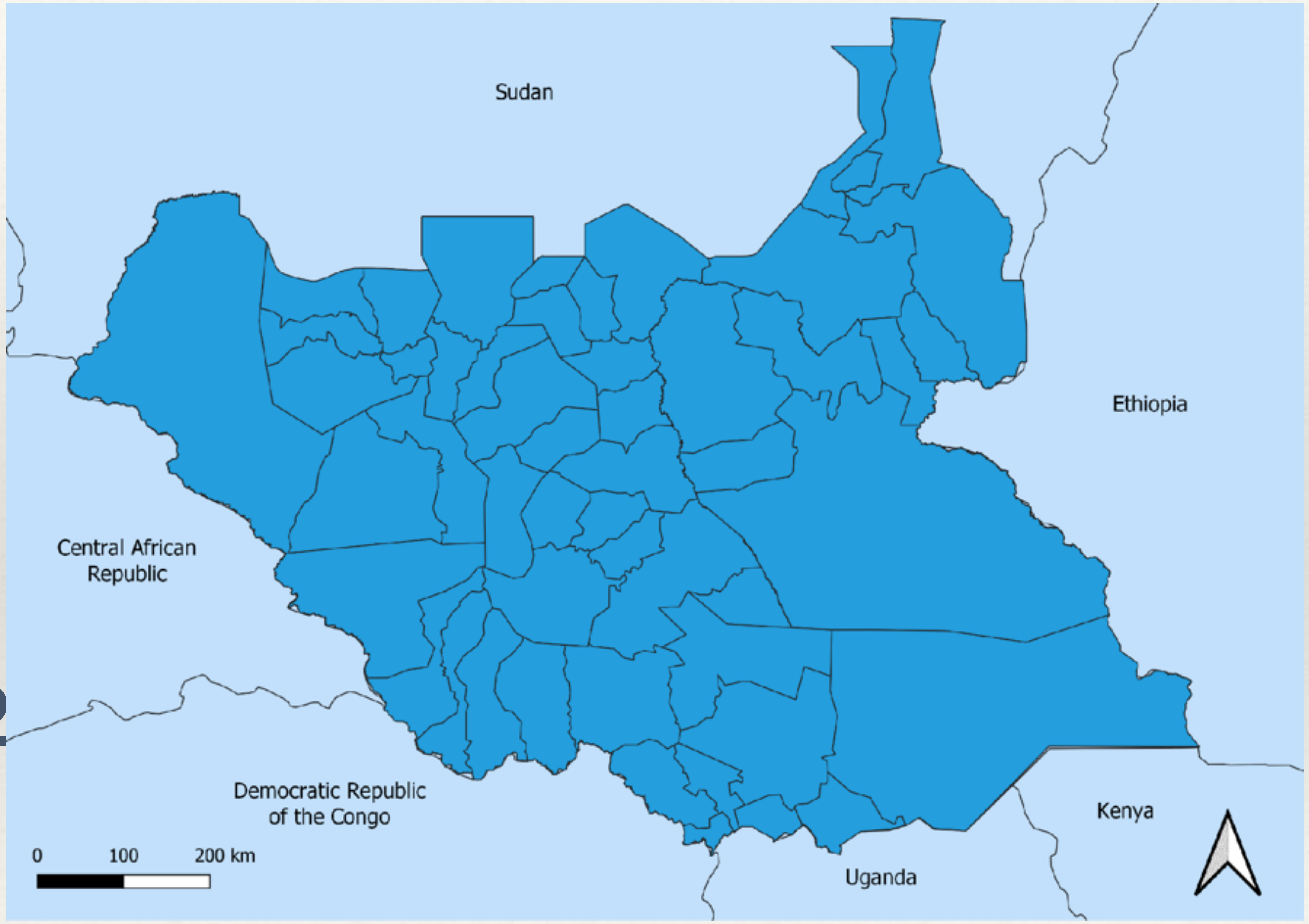



Static
X2
X1 X3 ... Xn
$y_t = \alpha x_t + C$
Model
Forecast value
Dynamic
Shape comparison
Forecast value

Study zone - South Sudan

- **48 regions**
  - Adm - 2 level
  - Regional merging if value < 10

  → ***72 to 48***



- **January 2020 to September 2**
  - **21 point time series**

Data Source

- **IOM-DTM flow monitoring surveys**
  - **31 key transit points within South Sudan and at its borders**

- **Focus on population leaving their region**
  - **Push** in the “push-pull” model
  - **Mostly economic, conflict or climate motivations**

Method : Benchmark of autoregressive models

Basic Model

Autoregressive model

$$y_t = c + \alpha * y_{t-1}$$

No parameters

→ no optimization

VS

Common Literature model

ARIMA model

$$y'_t = c + \sum_{i=1}^{p} \alpha_i * y'_{t-i} + \sum_{j=1}^{q} \beta_j * \epsilon_{t-j}$$

3 parameters (p,d,q)

VS

Dynamic model

ARIMAX

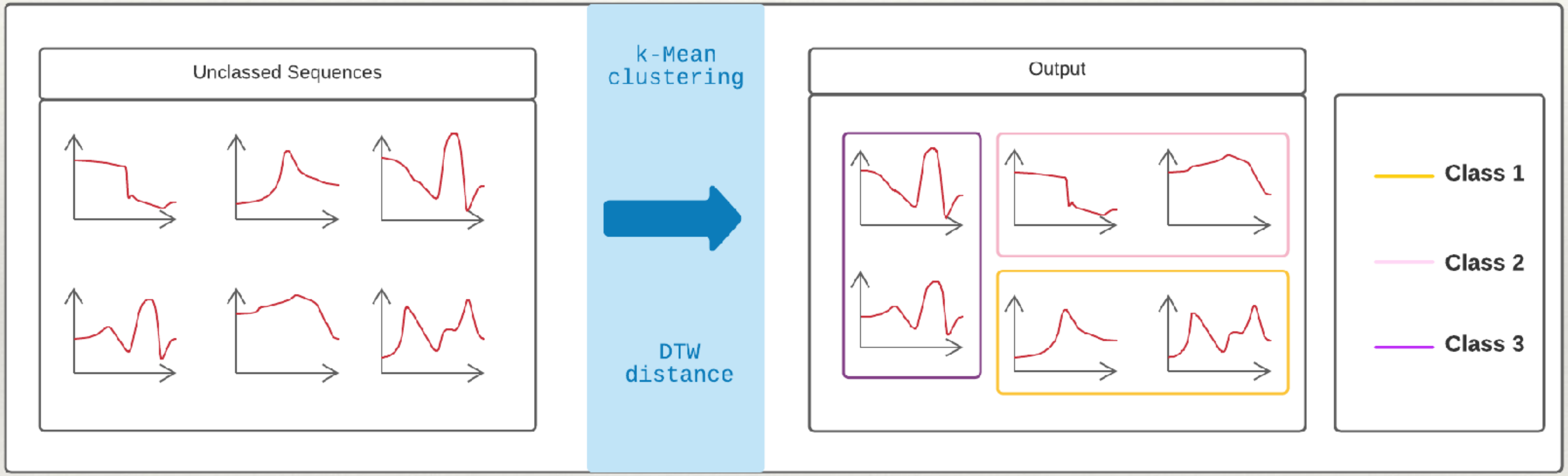
$$y'_t - c + \sum_{i=1}^{p} \alpha_i * y'_{t-i} + \sum_{j=1}^{q} \beta_j * \epsilon_{t-j} + \gamma_c$$

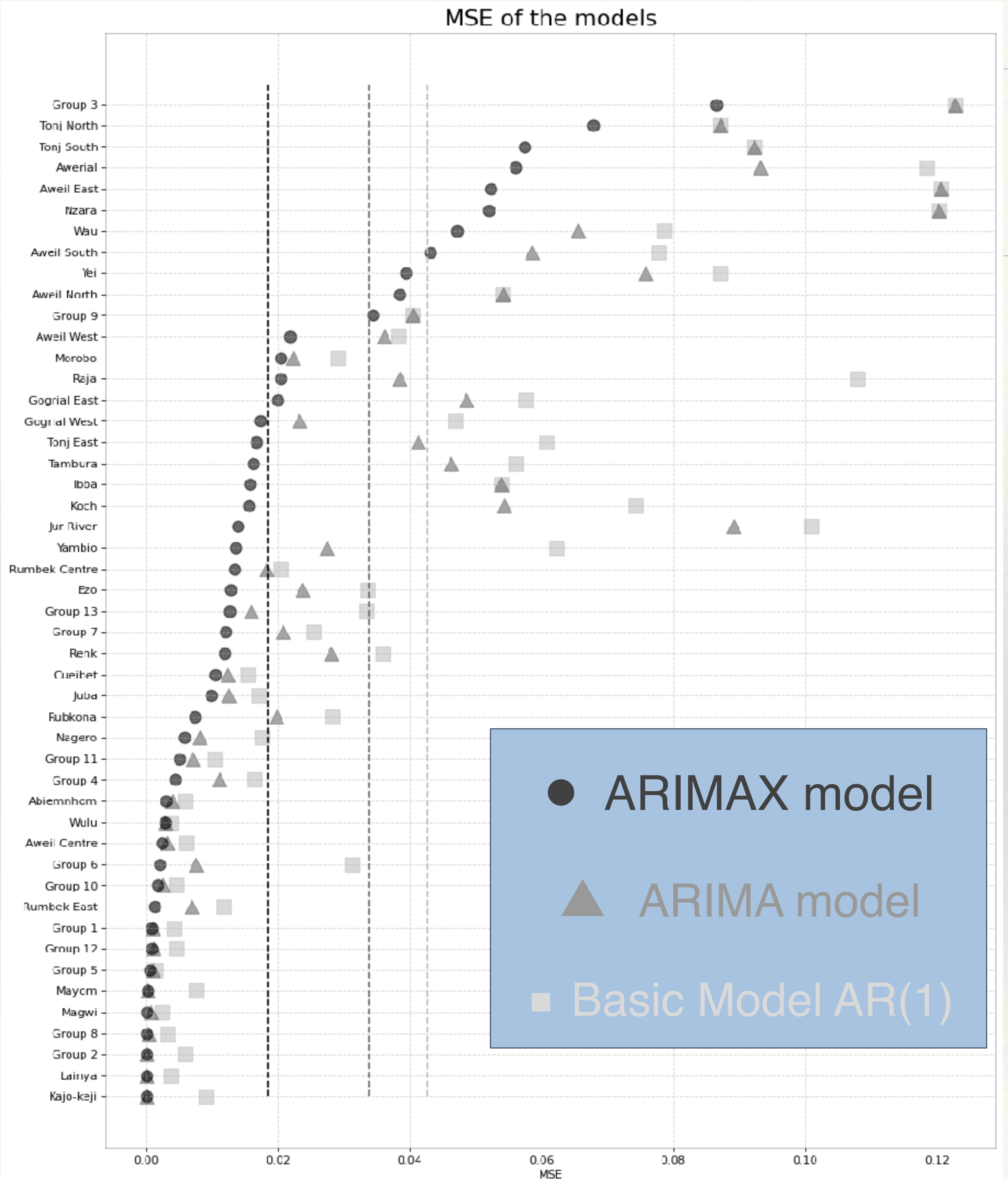
$$\gamma_c = \begin{cases} \gamma_1, & class_1 - 1 \\ \gamma_2, & class_2 = 1 \\ \gamma_3, & class_3 = 1 \end{cases}$$

5 parameters :

- **3 ARIMA (p,d,q)**
- **2 DTW classification**

Method : DTW - Classification application

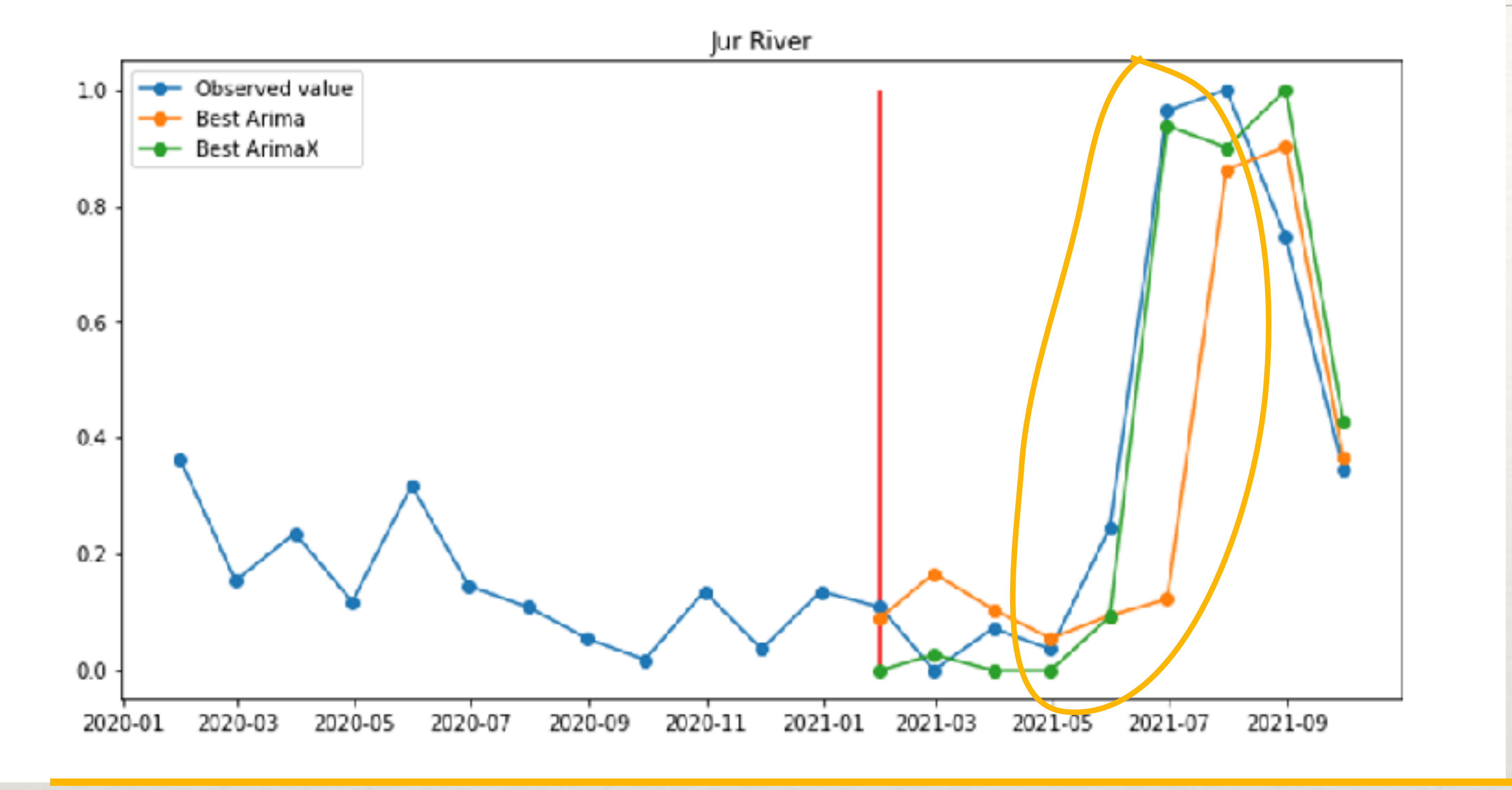




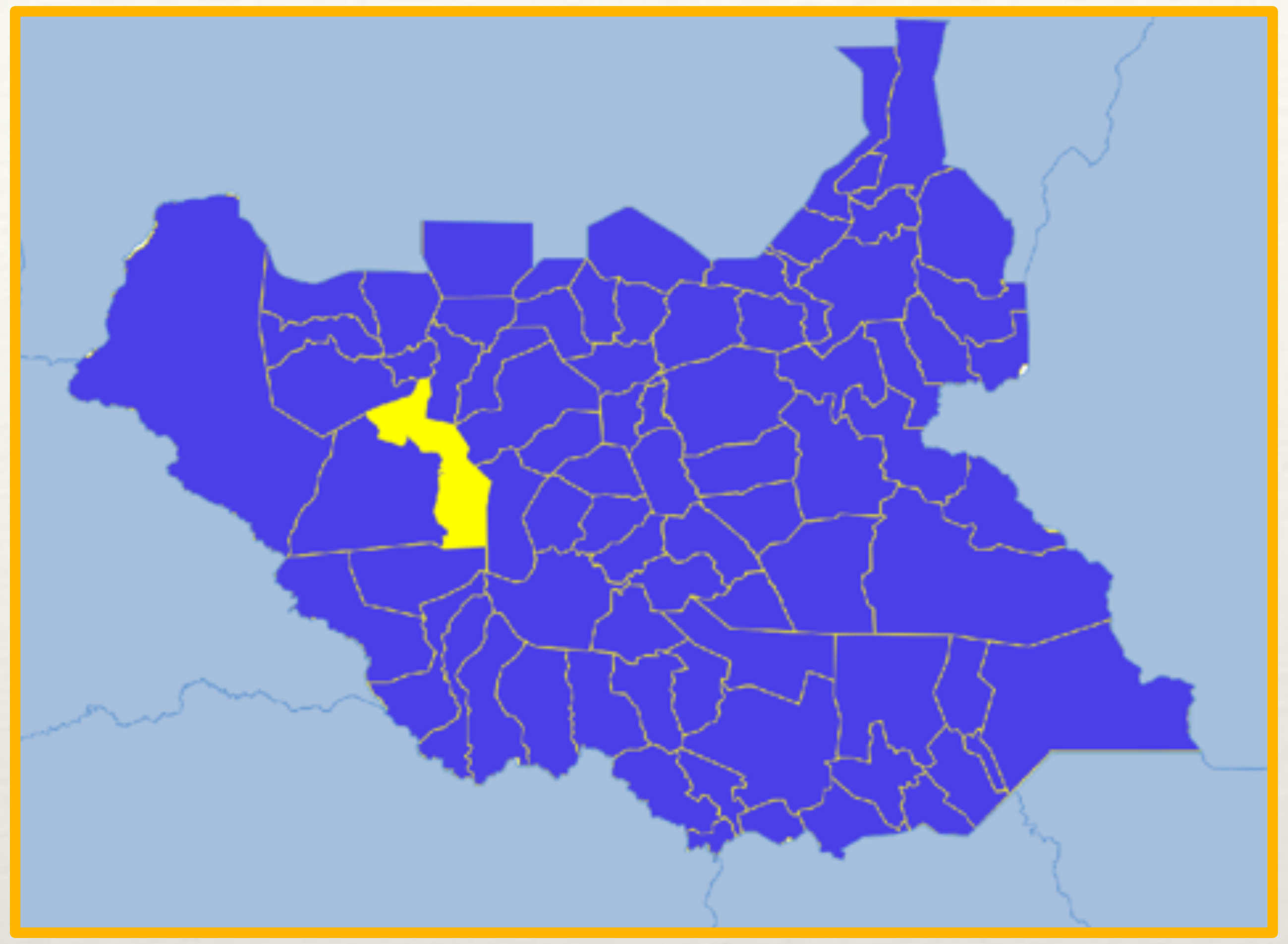
MSE Plots

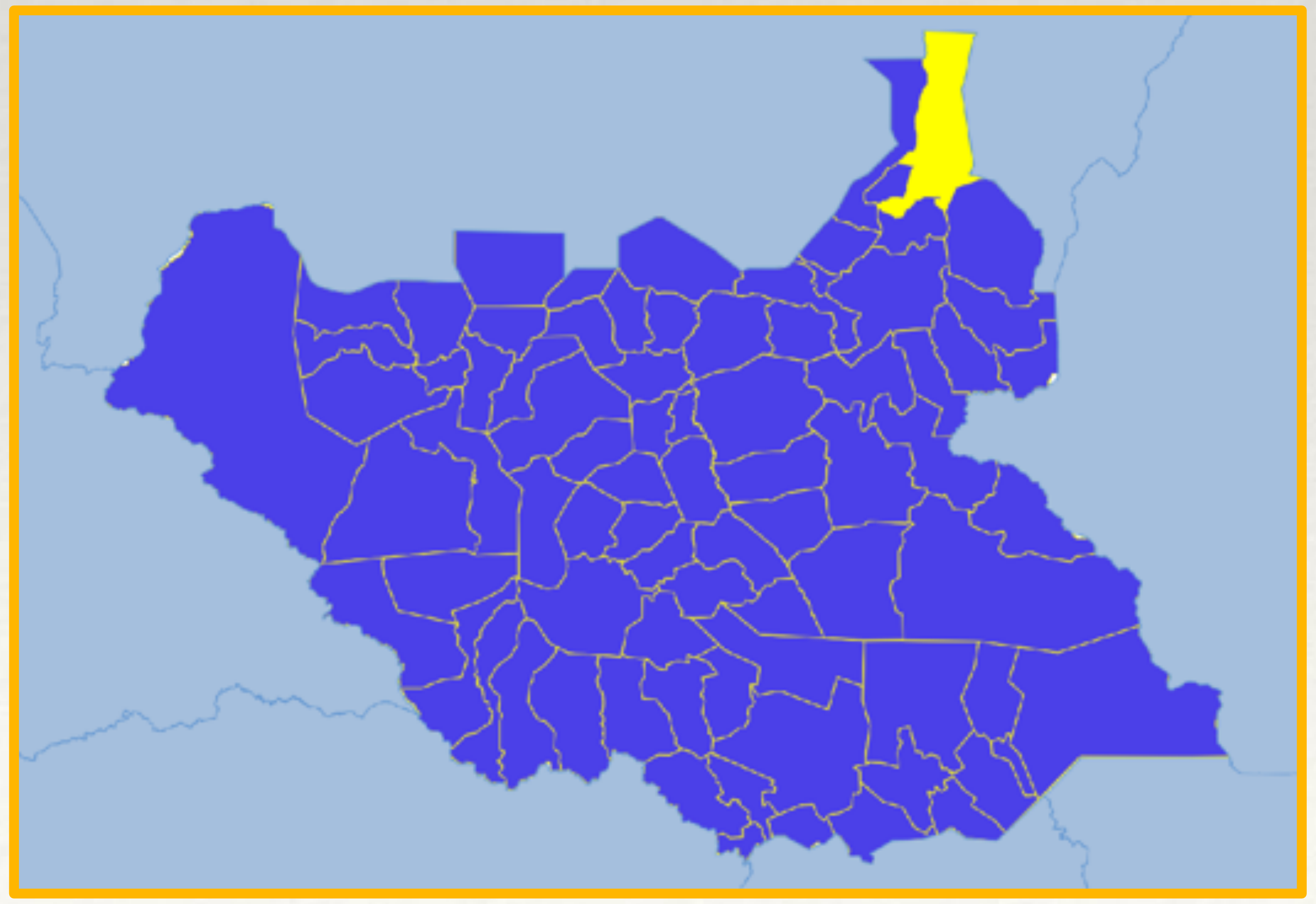
Mean value of MSE for each South Sudan regions

Region forecast examples : Jur River and Renk

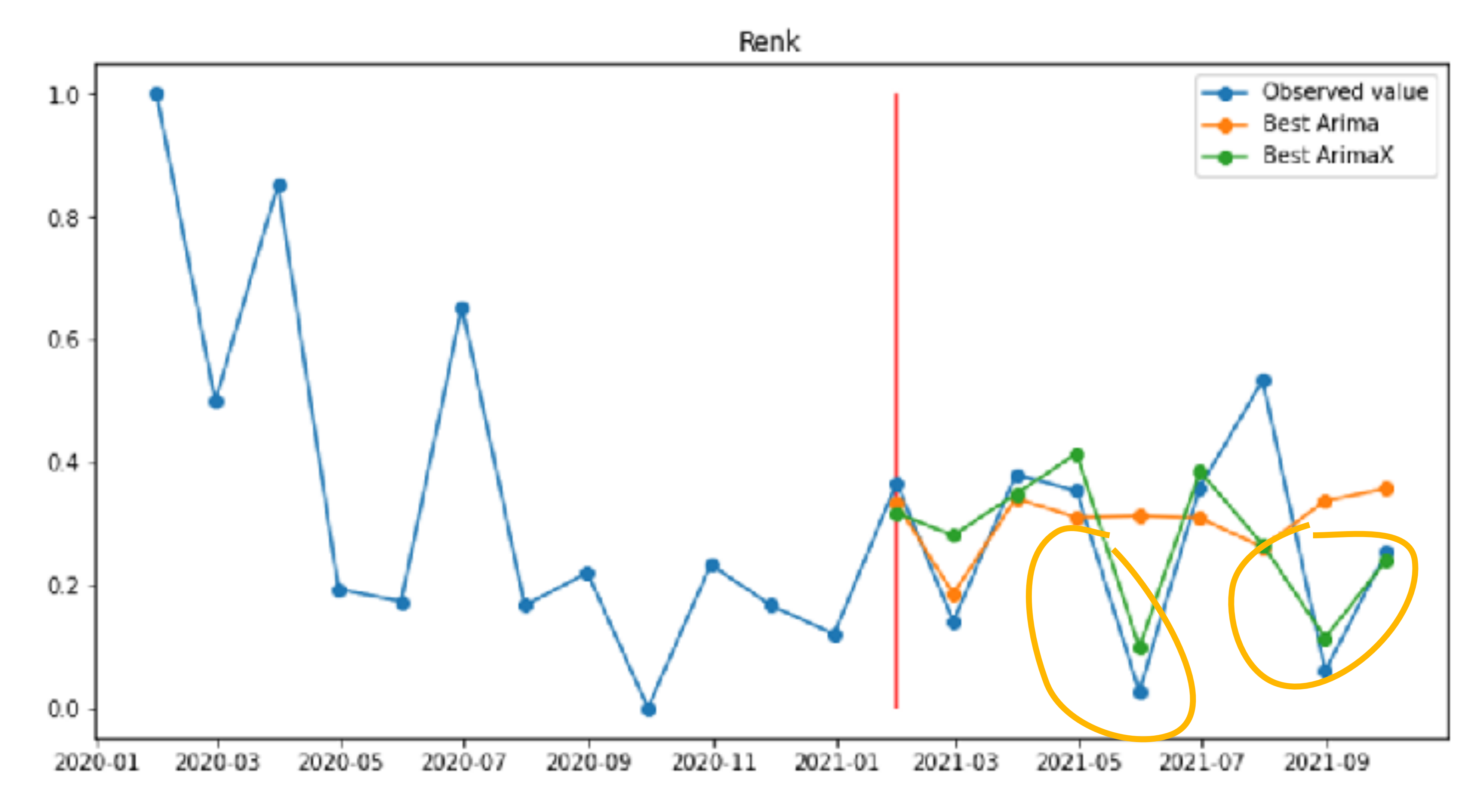


Jur River

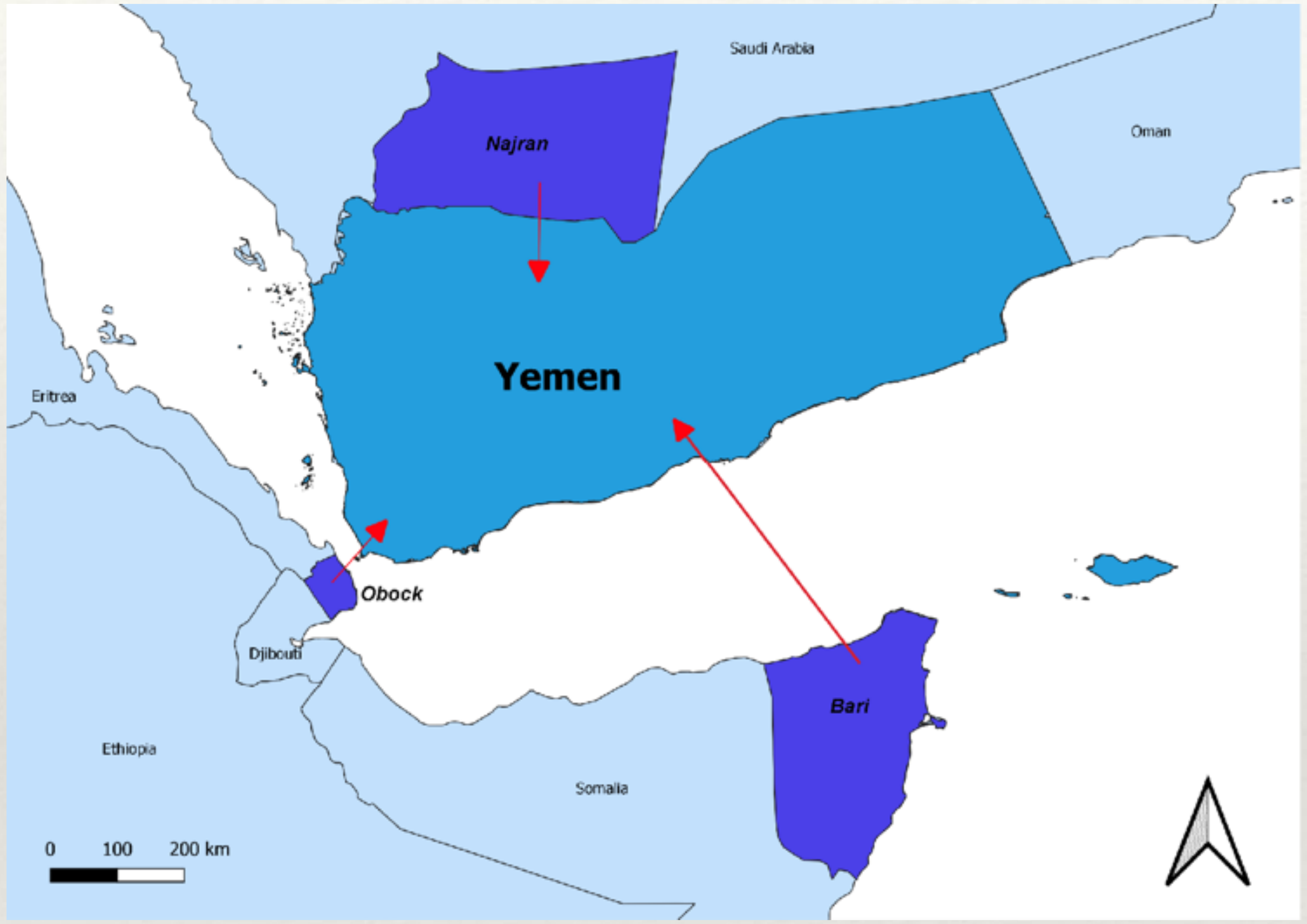




Renk



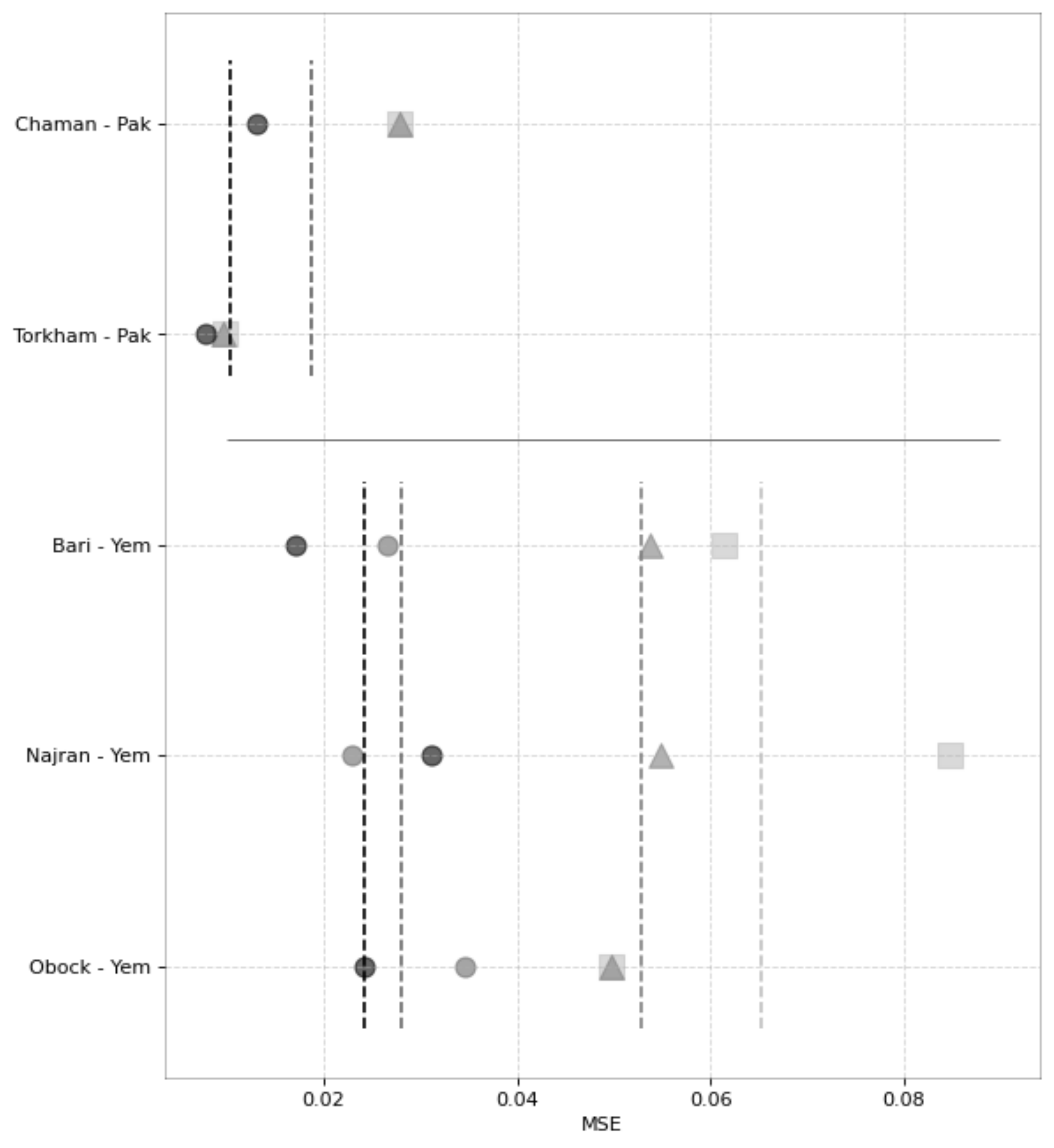
Beyond South Sudan : Yemen and Pakistan



- ***Bari (Somalia), Obock (Djibouti), Najran (South Arabia)***
- ***49 months time series (Jan 2018-Jan 2022)***

- *Torkham and Chaman points*
- *52 weeks time series*

Beyond South Sudan : Yemen and Pakistan



Mean value of MSE for Pakistan and Yemen

- ● ARIMAX model
- ● ARIMAX model *(South Sudan)*
- ▲ ARIMA model
- ■ Basic Model AR(1)

Conclusion

- Good results in South Sudan, Yemen and Pakistan

- Apply the method to other topics
  - Climate
  - Protest
  - Other ideas ?

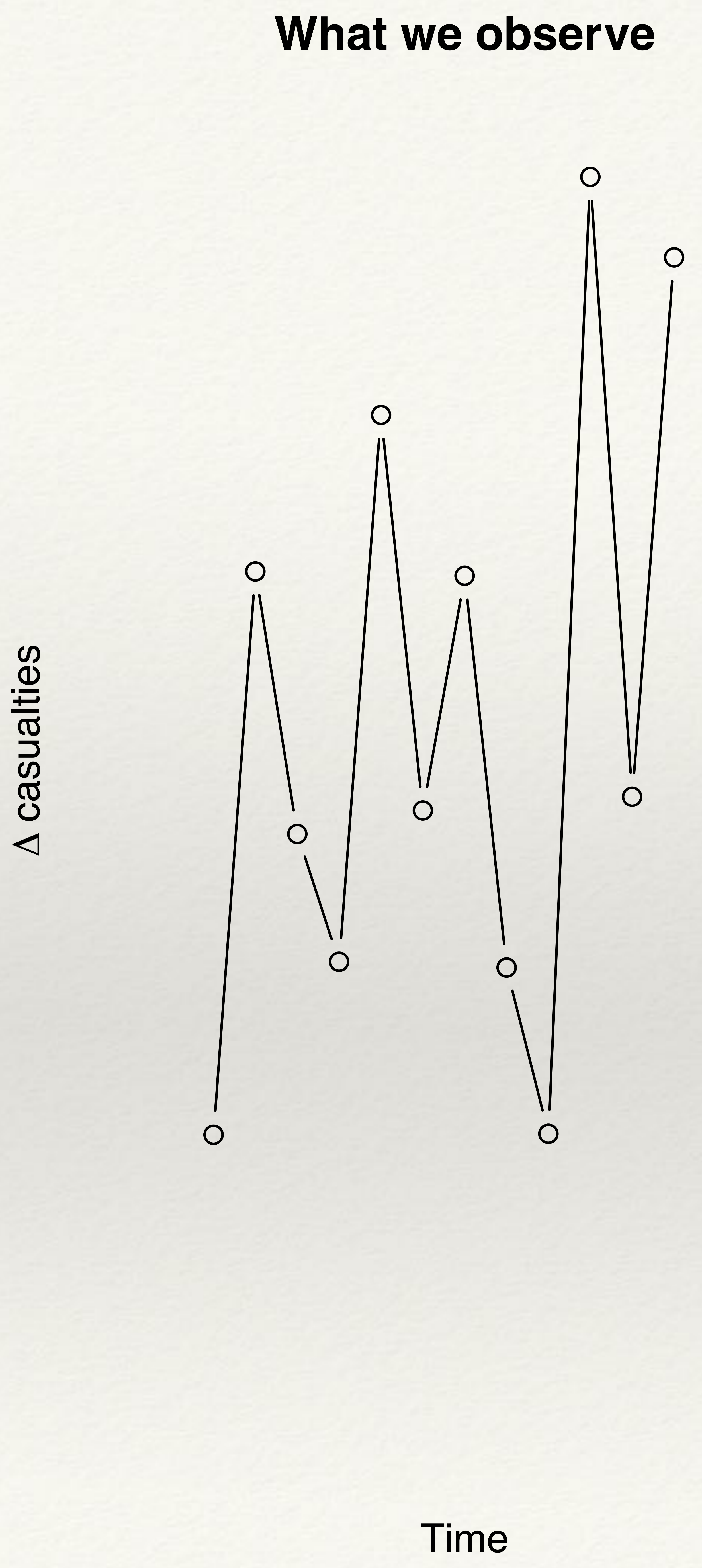
Project 2

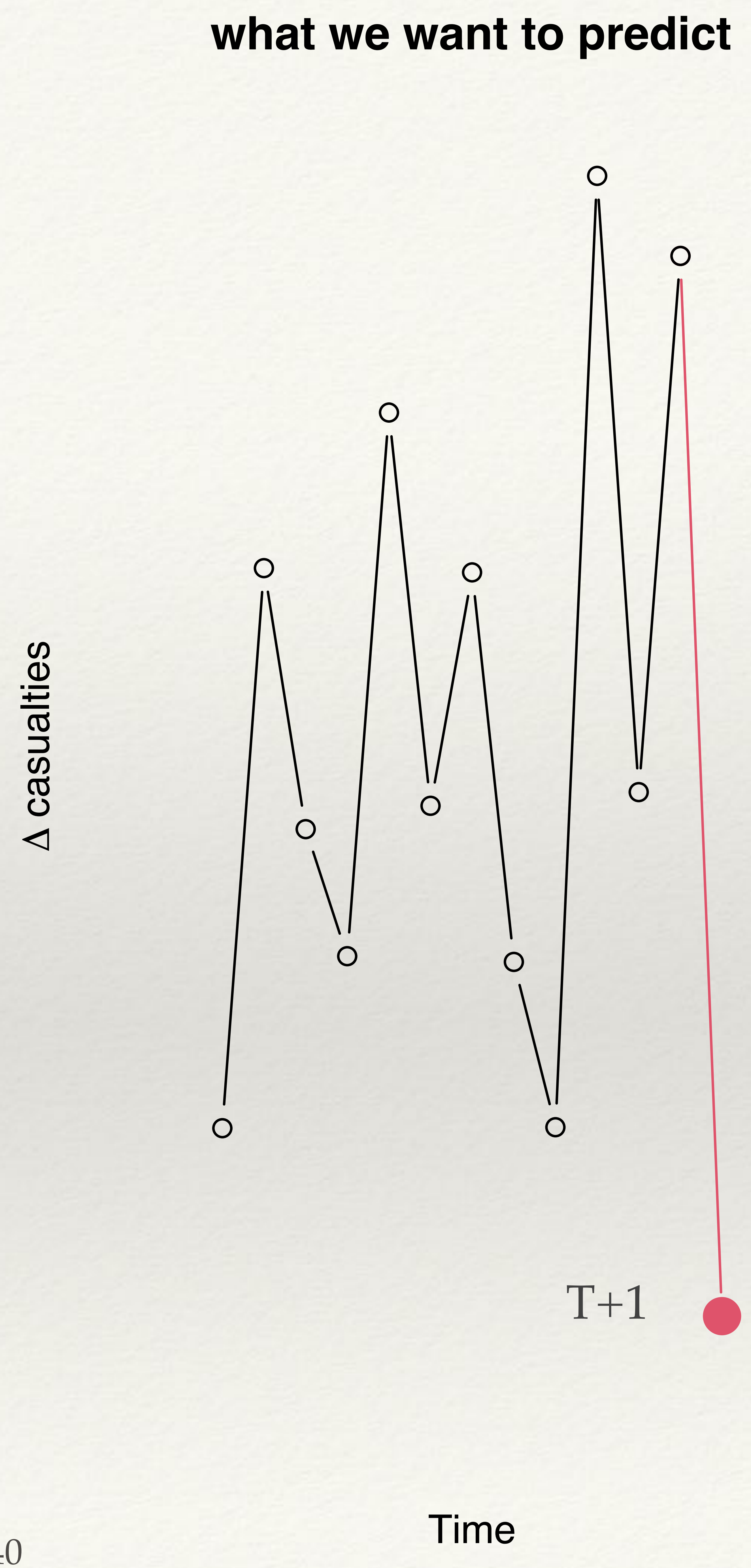
Temporal Patterns in Protest Data

Hannah Frank & Thomas Chadefaux

Project 3
Augmentation for time series

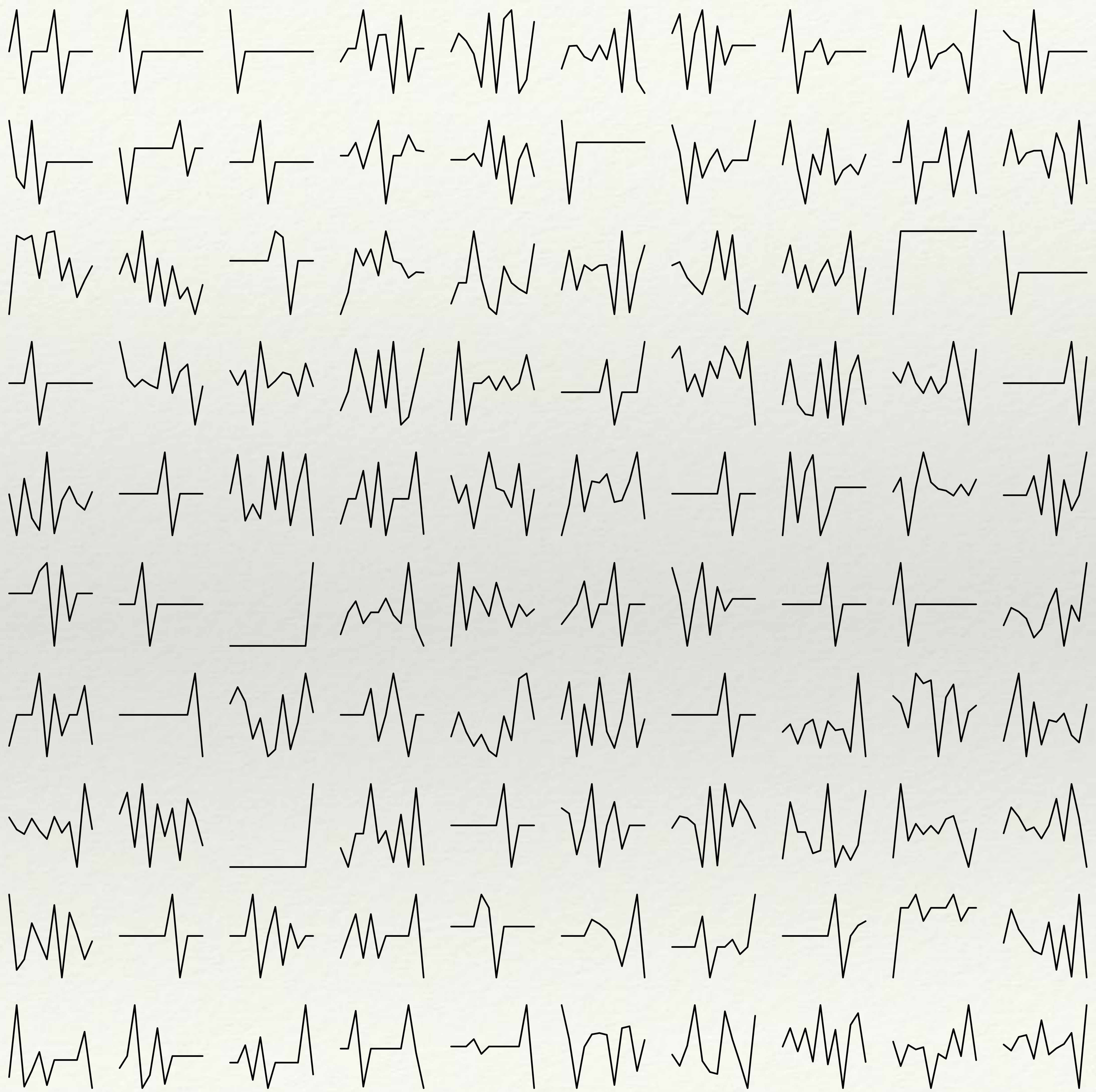
Thomas Chadefaux


What we observe
Δ casualties
Time


what we want to predict
Δ casualties
T+1
Time

Learning Set

Test set



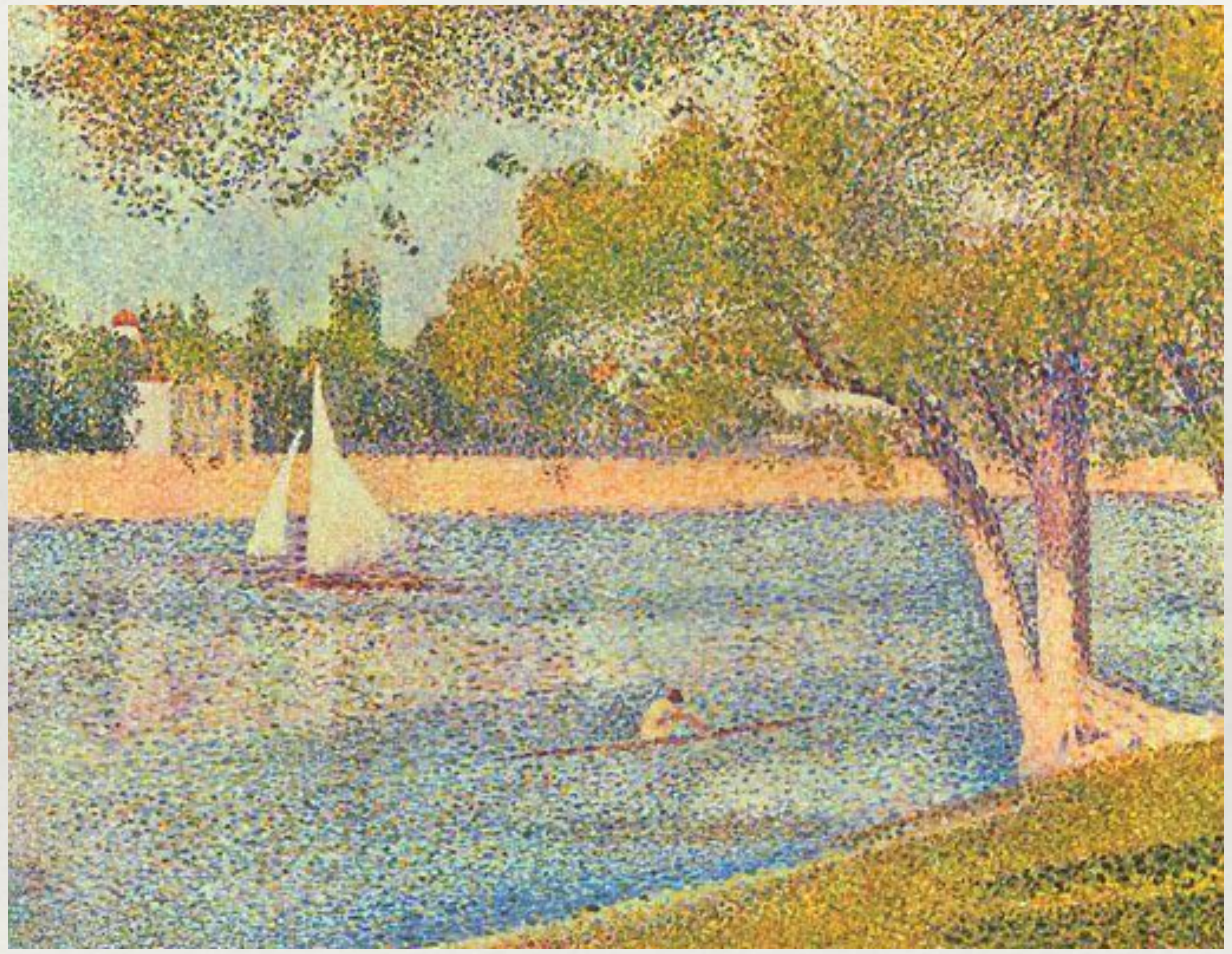
- Data is scarce
  - One world
- Acquiring data is expensive

Common methods of dealing with limited data in social sciences

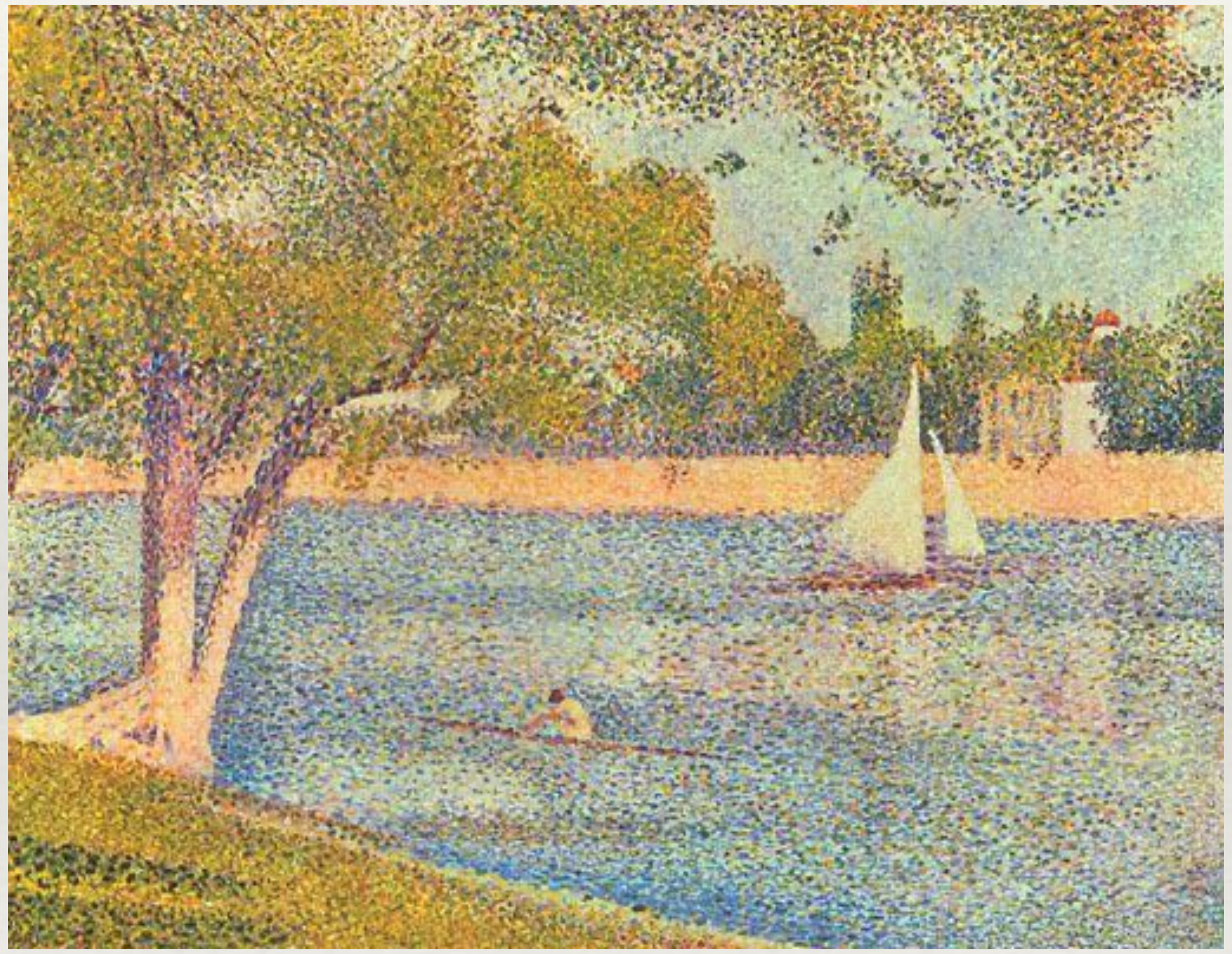
- Bootstrapping and Bagging: use subsets of the data.
- Imputation: fill gaps
- Create entirely new dataset: augmentation

Augmentation

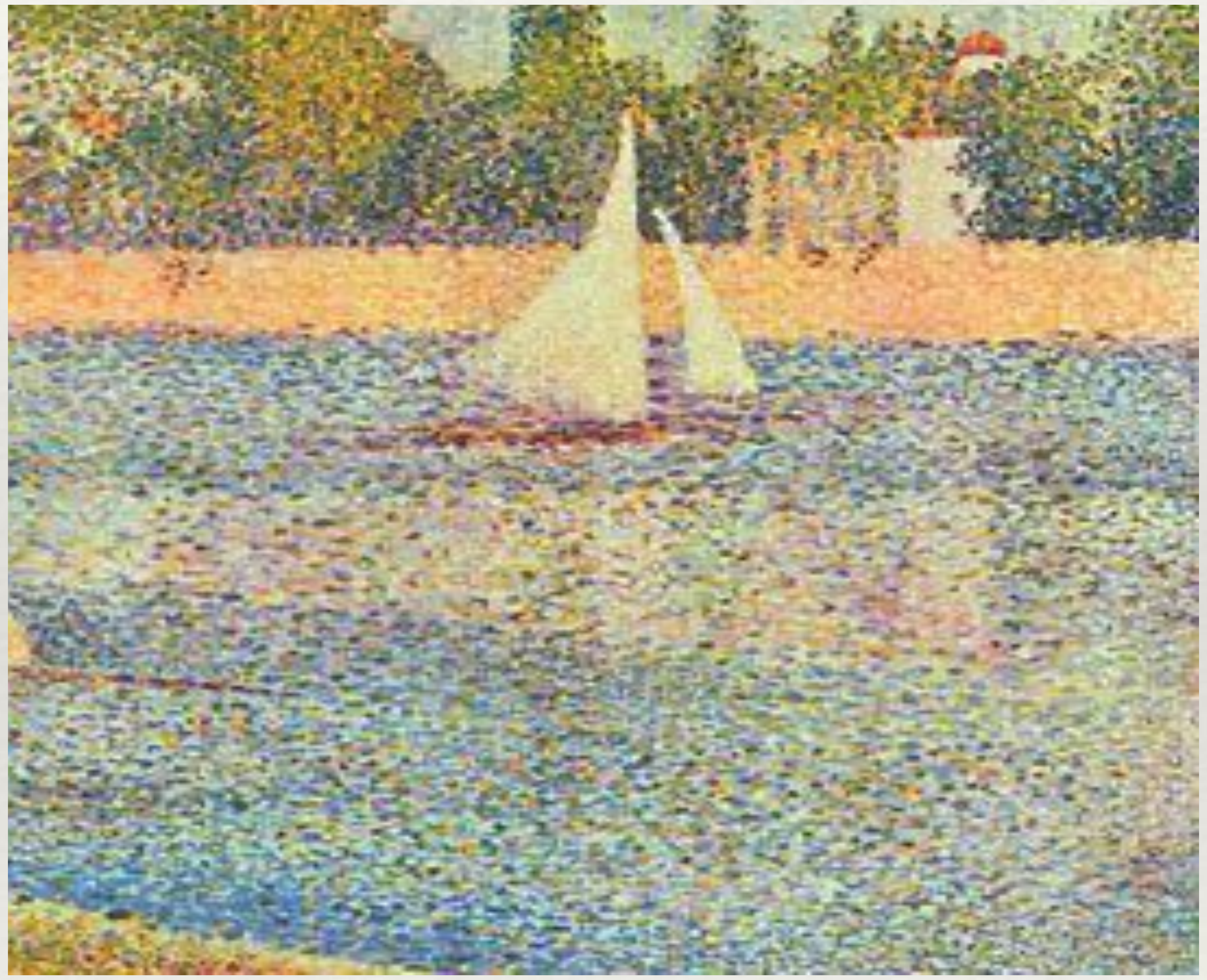
Original



Flipped

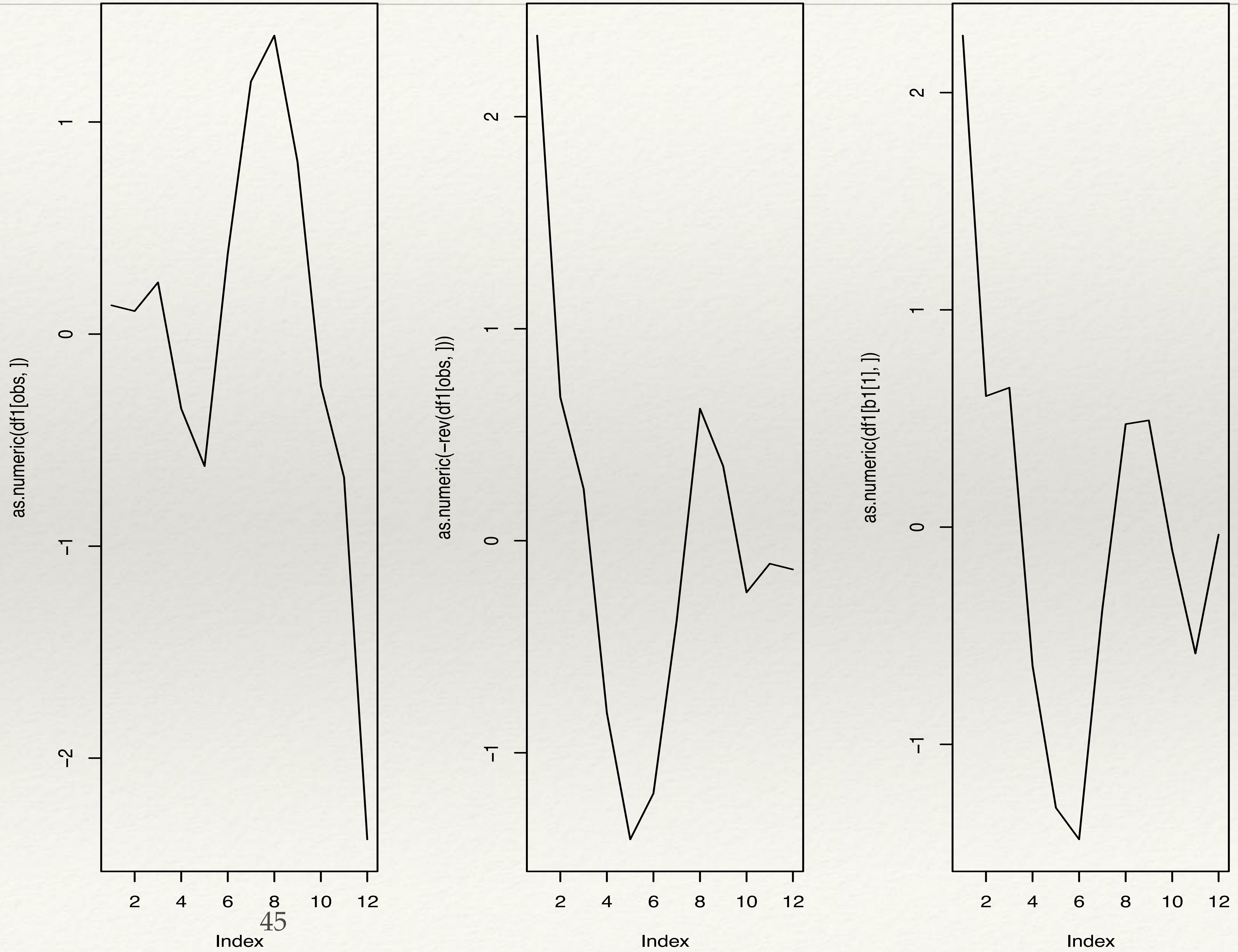


Cropped / stretched

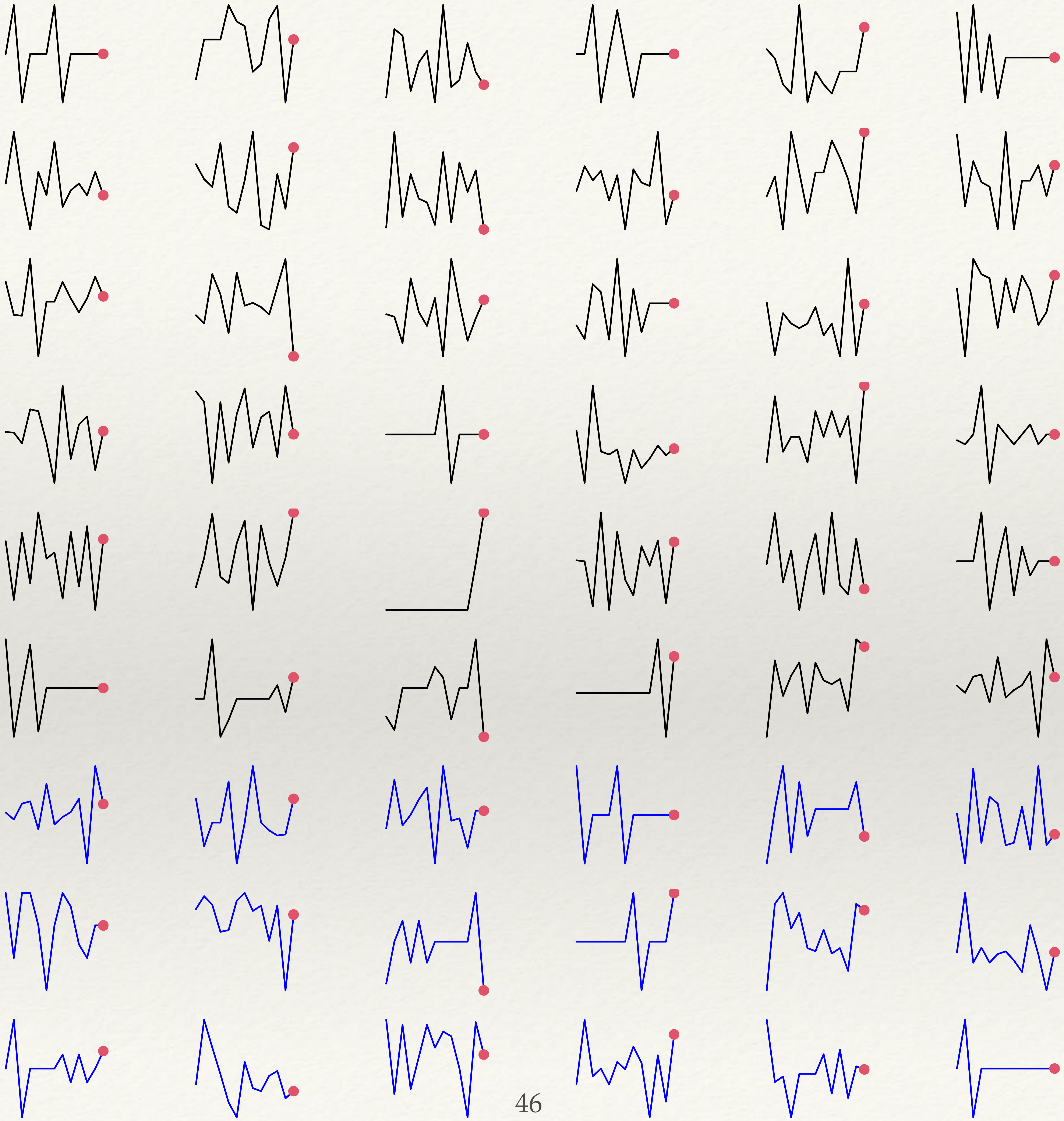


Basic augmentation for time series: Flipped series

- History repeats
  - In reverse
  - Escalation and de-escalation as mirrors



Augmented Set



Learning set

↓

RF / NN / LM

↓

Predict test set

↓

MSE

Augmented
Learning set

↓

RF / NN / LM

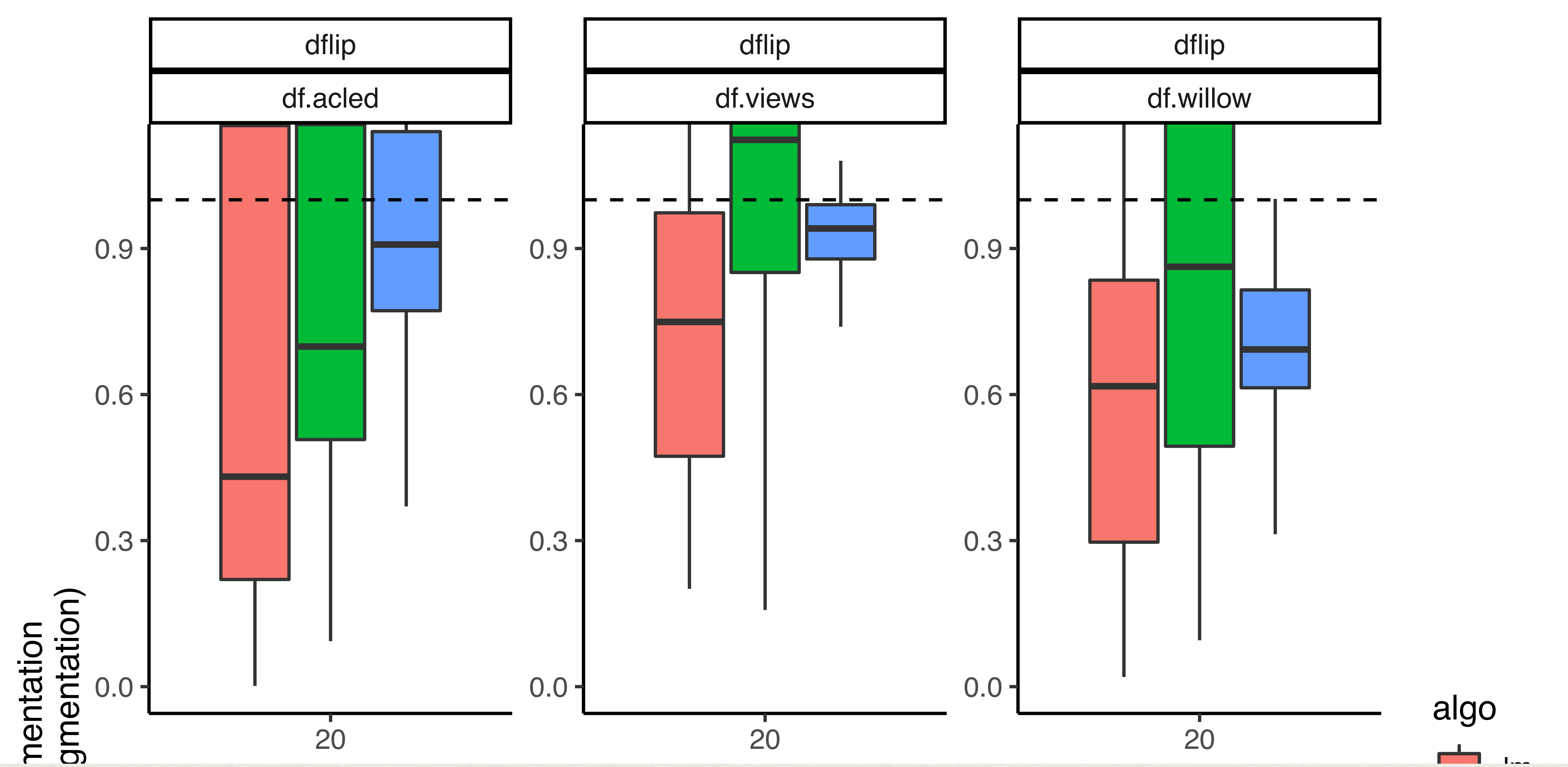
↓

Predict test set

↓

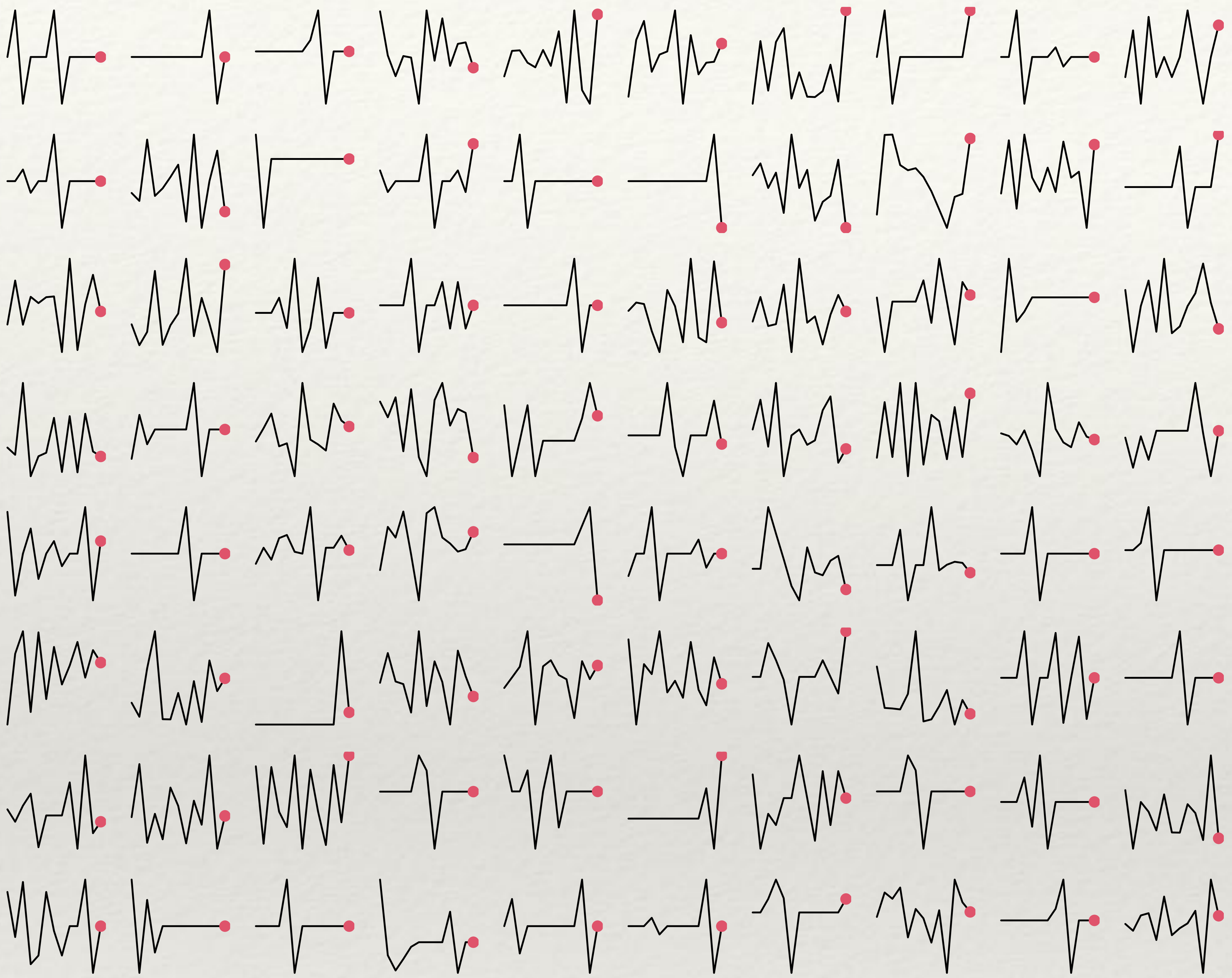
MSE

Results: double flip



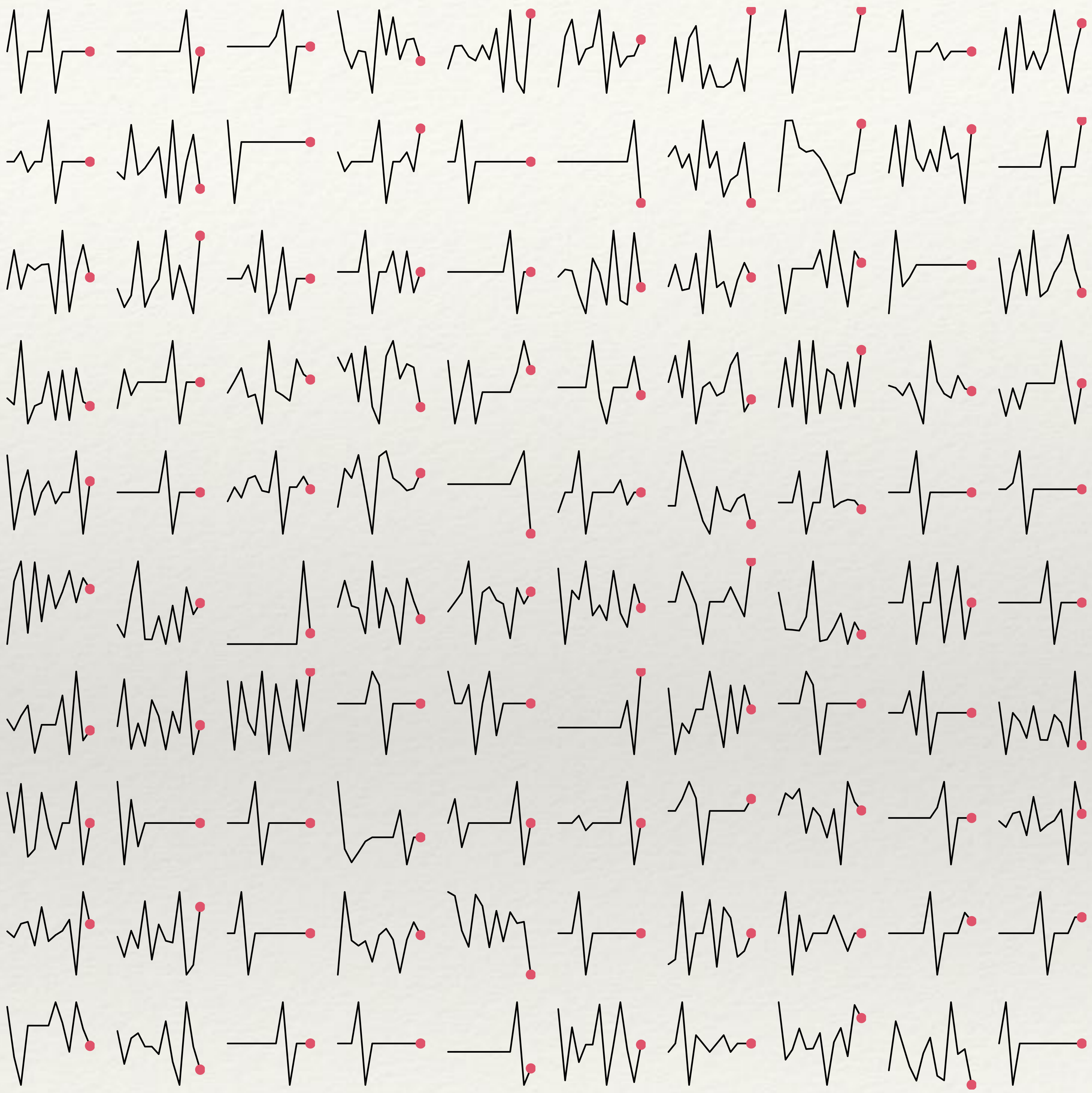
How much of an improvement is that really?

MSE with 80 obs



=

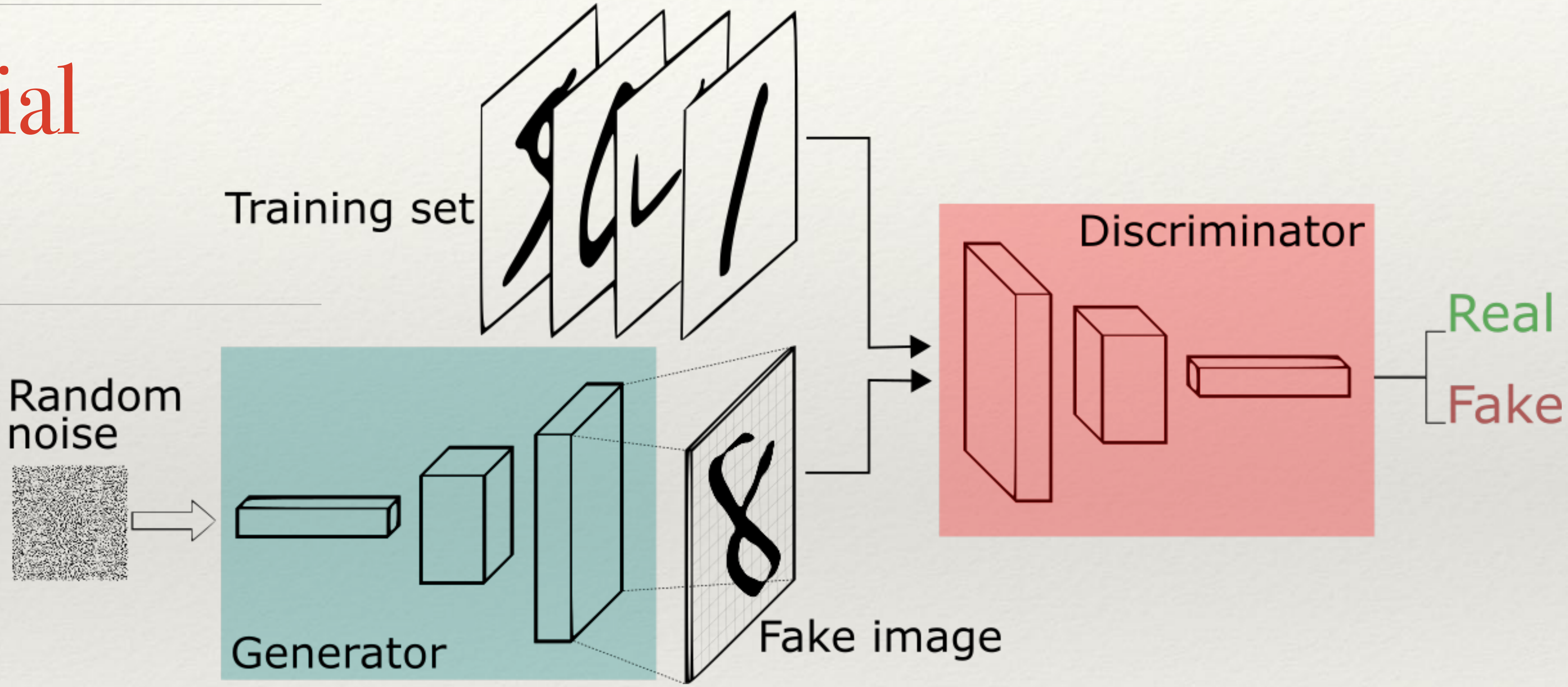
MSE with 100 obs



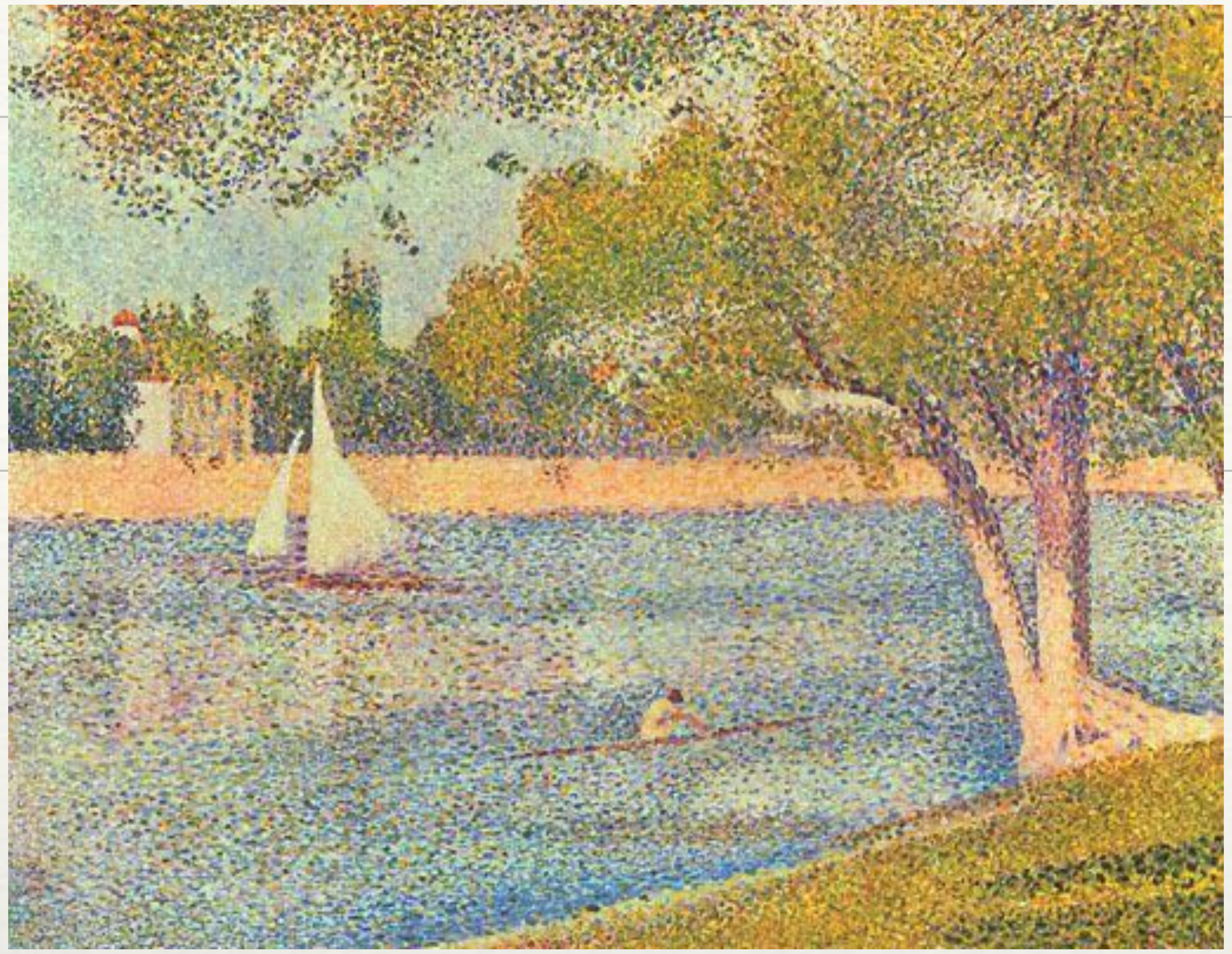
+ double flip Augmentation

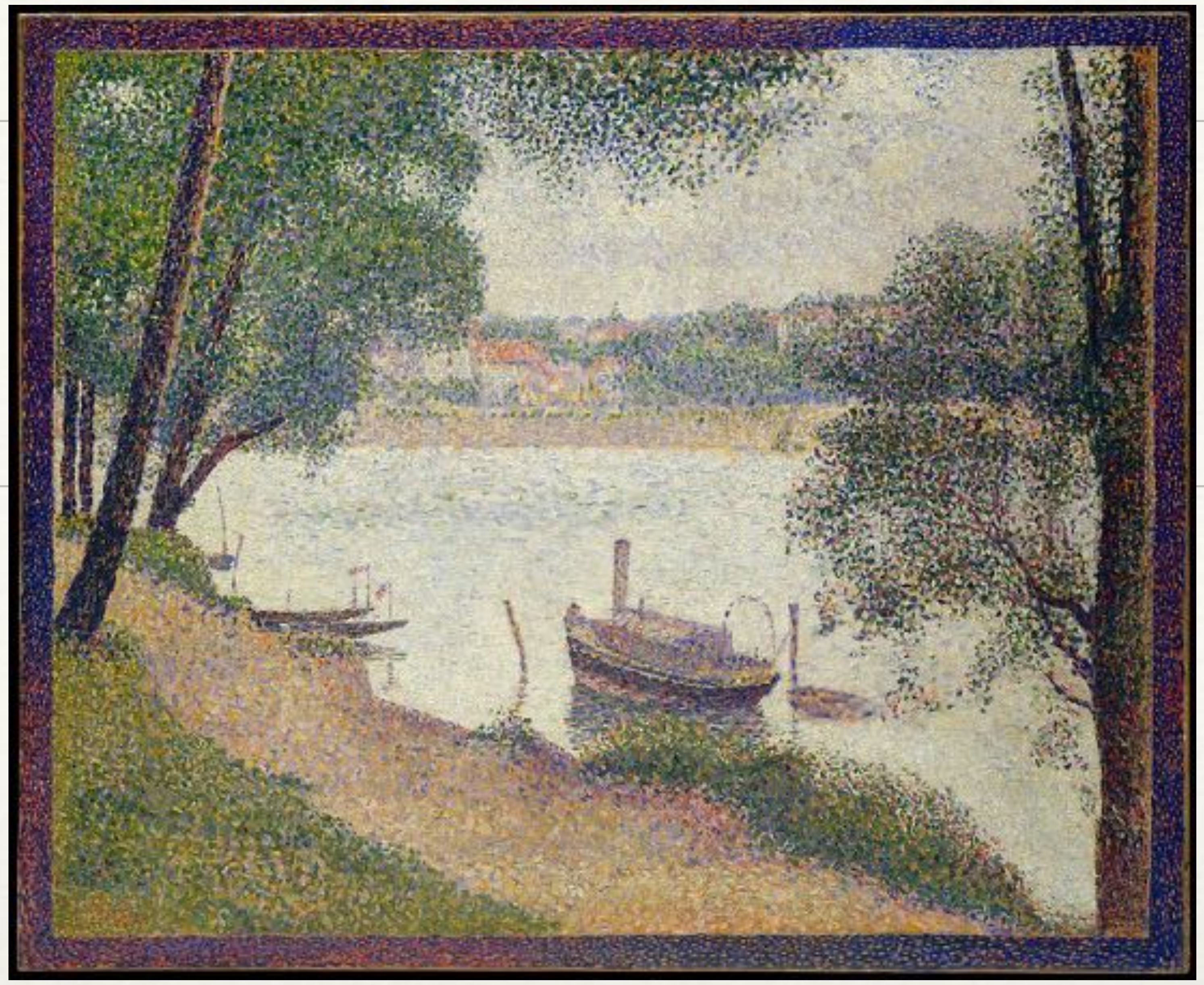
More complex augmentation methods

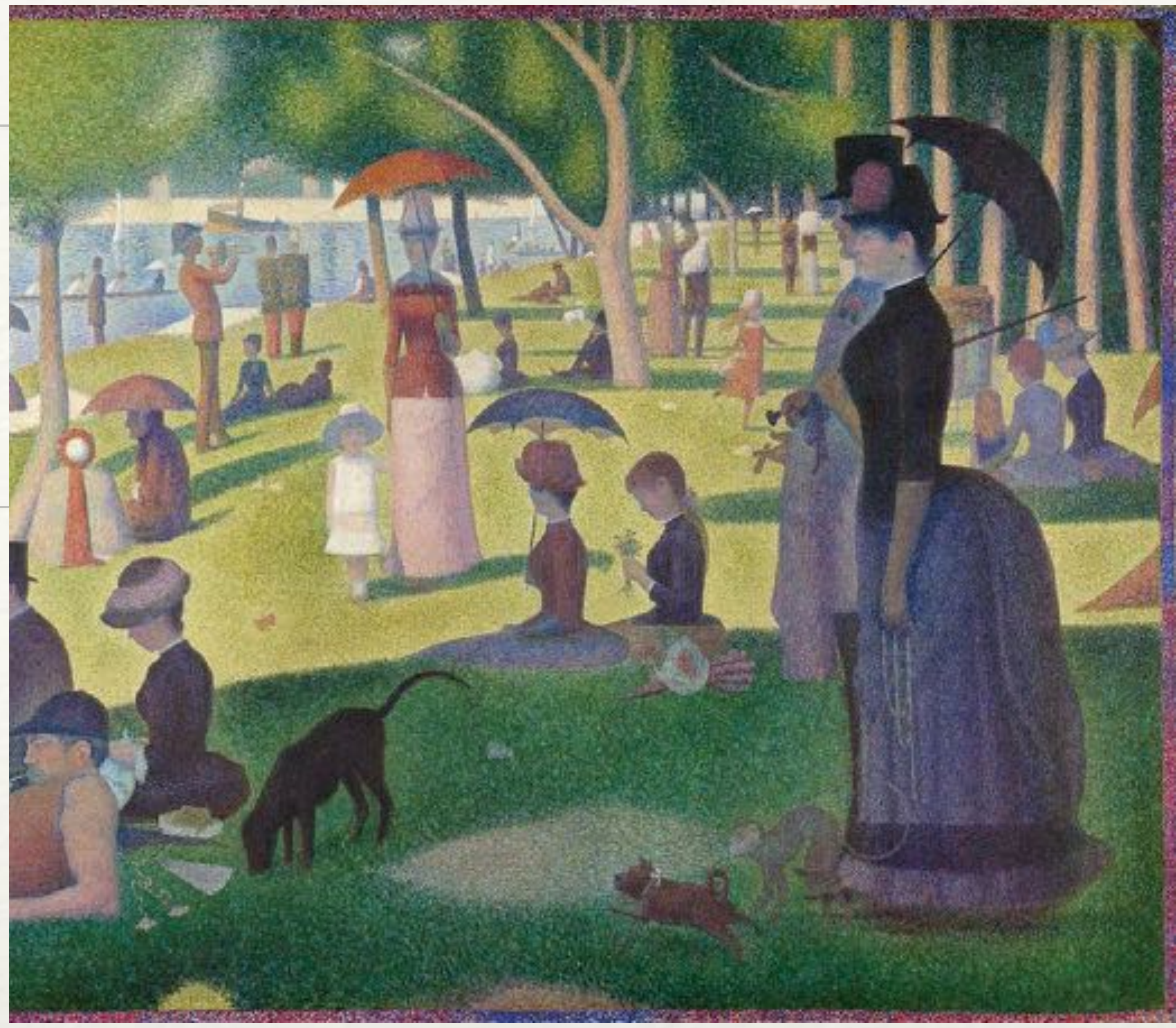
Generative adversarial networks



Seurat







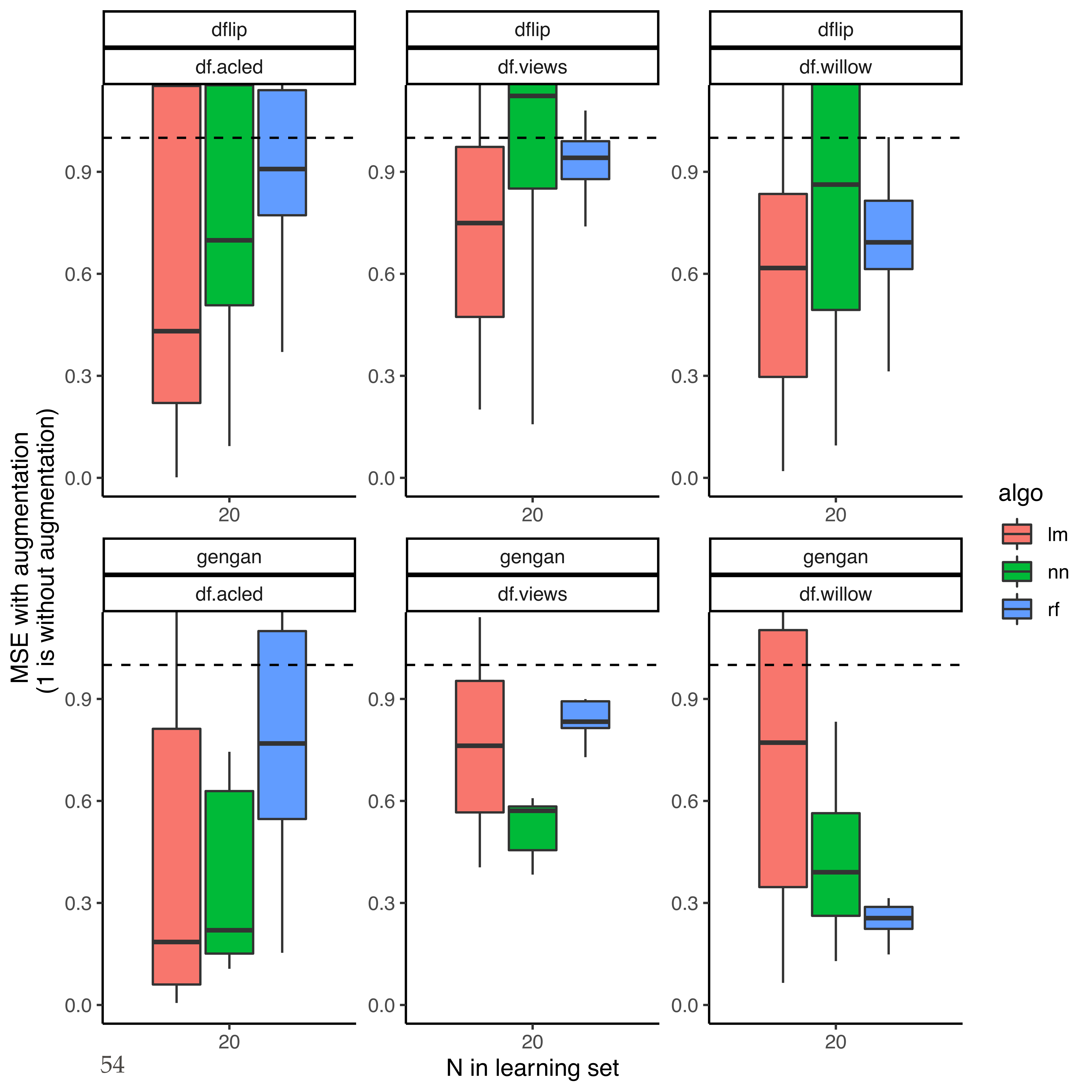
Not
Seurat

Augmentation: GAN

Fake Seurat painting

Lorem Ipsum Dolor

Results: GAN



Todo

- Multi-dimensional data – Panel
- Other datasets – suggestions?

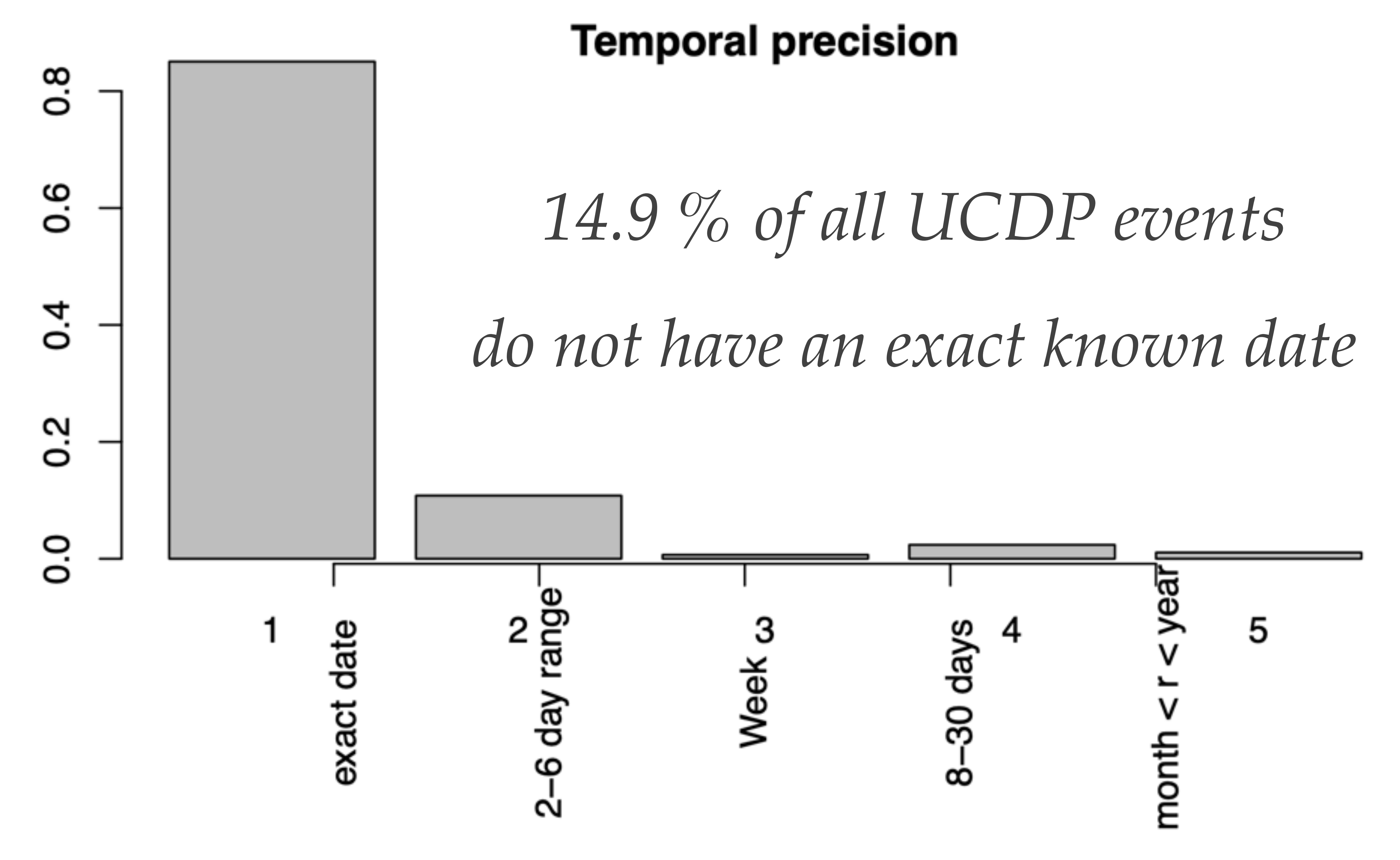
Project 4

Reducing Uncertainty In Conflict Events Using Satellite Data

Gareth Lomax & Thomas Chadefaux

Motivation

- News based datasets (e.g., UCDP, ACLED) often contain temporal and geographic uncertainty.
- Uncertainty is compounded by the severity of violence.
- Datasets do not report quantitative extent of non violent damage.



Satellite Imagery?

Contains a wealth of change detection literature.

Medium resolution data is well covered under permissive licences.

(Very) High Resolution data is expensive, unless you know where to look.

NASA FIRMS

Fire Information For Resource Management Service

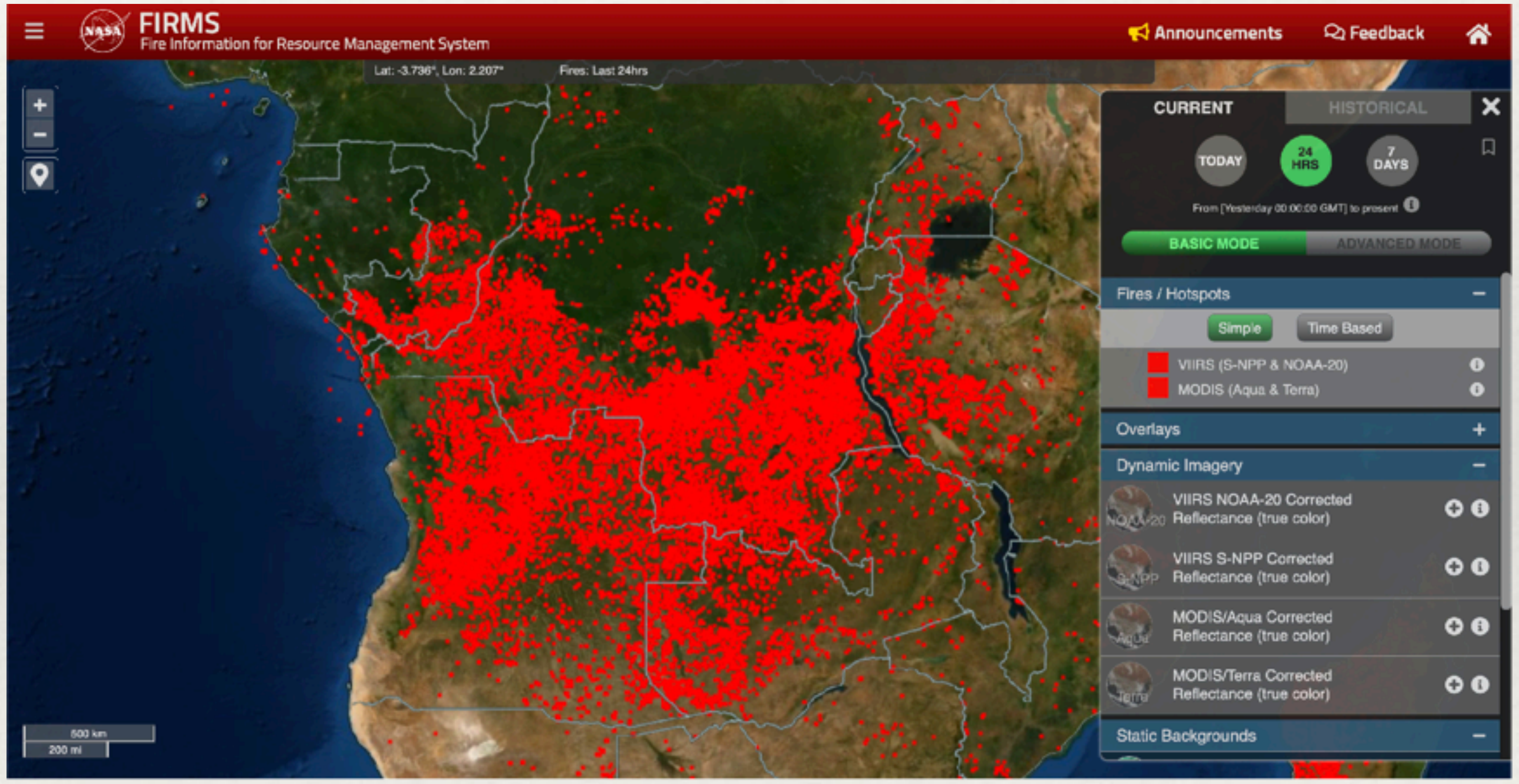
Data from MODIS and VIIRS sensors

MODIS resolution 1km, VIIRS resolution 375m

MODIS aboard Terra and Aqua EOS, acquisitions at 10:30 a.m. and 10:30 p.m MLT

VIIRS aboard Suomi NPP and NOAA-20

Total acquisitions 8 per day.



NASA FIRMS DASHBOARD OVER AFRICA 12/07/2022

Method

Use Firms Archive Data to identify statistically unlikely fires to reduce uncertainty in conflict data.

Myanmar

- Ethnic cleansing committed by the Tatmadaw (Military) from 2017 - 2018 lead to large scale displacement and destruction.
- Approximately 900,000 Rohingya displaced.
- In Rakhine State 30.8% of events have no exact date. 84% of events have no precise location.
- Open Source Projects e.g Ocelli Project Seek to reduce this uncertainty.

Ocelli Project

- Dataset of 613 (620) Rohingya Settlements destroyed between 2017 and 2018.
- Based on open source very high resolution satellite images (google earth) to identify village burnings manually.
- Mixture of satellite and aerial imagery.

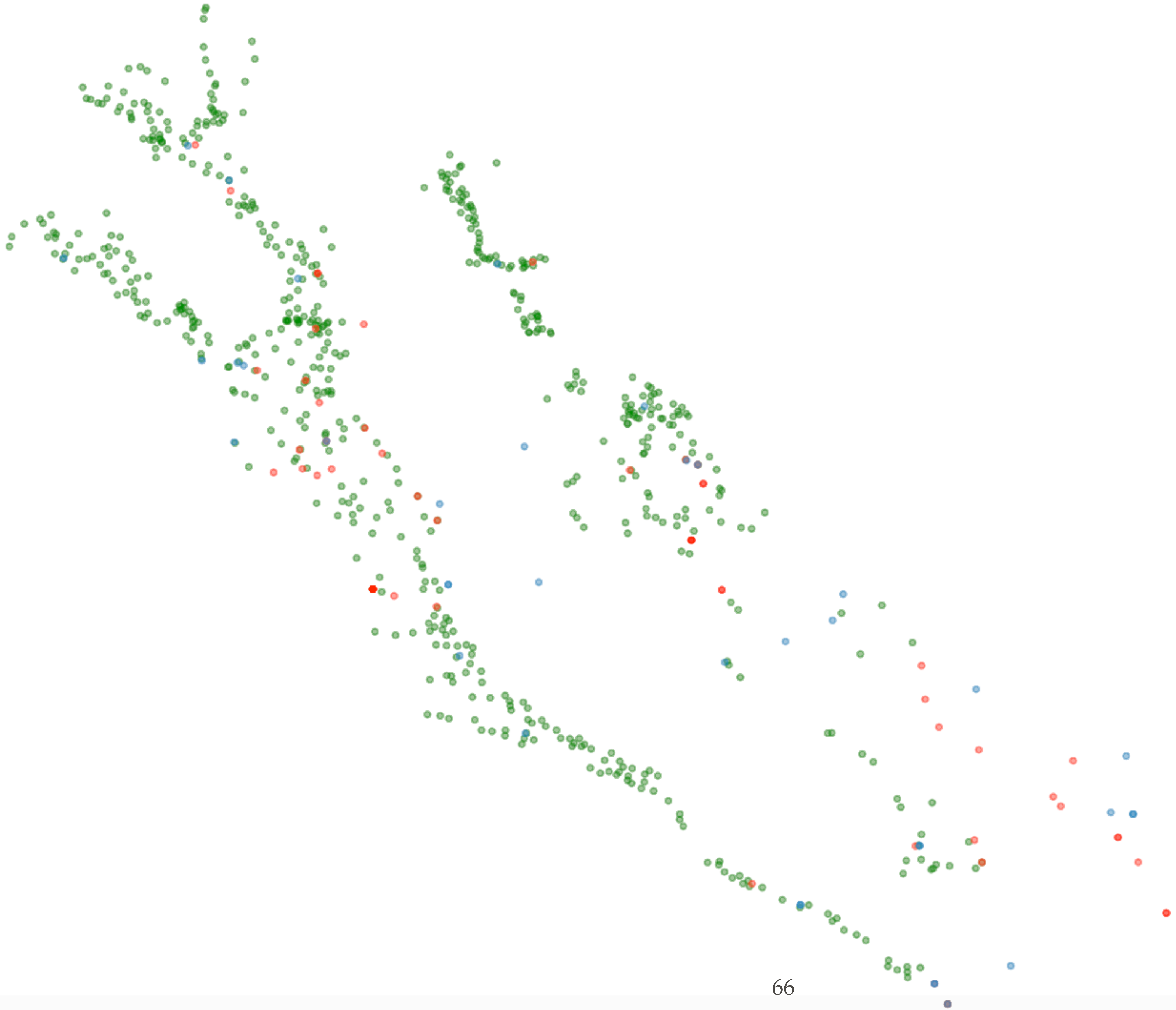
Anauk Myinhlut and Al Le Than Kyaw area

December 27, 2016

October 31, 2019

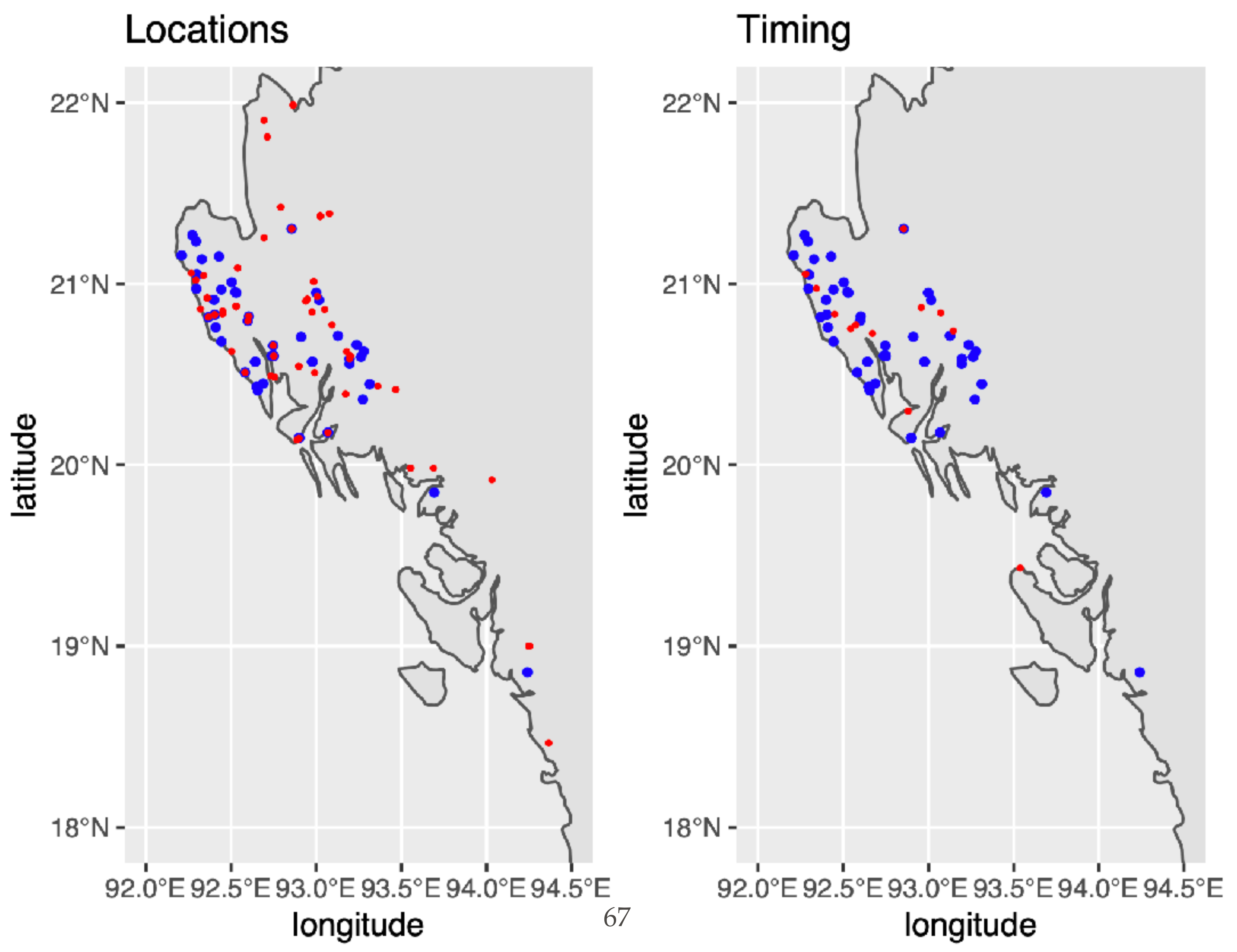

Image © 2021 Maxar Technologies
Google Earth


Image © 2021 Maxar Technologies
Google Earth

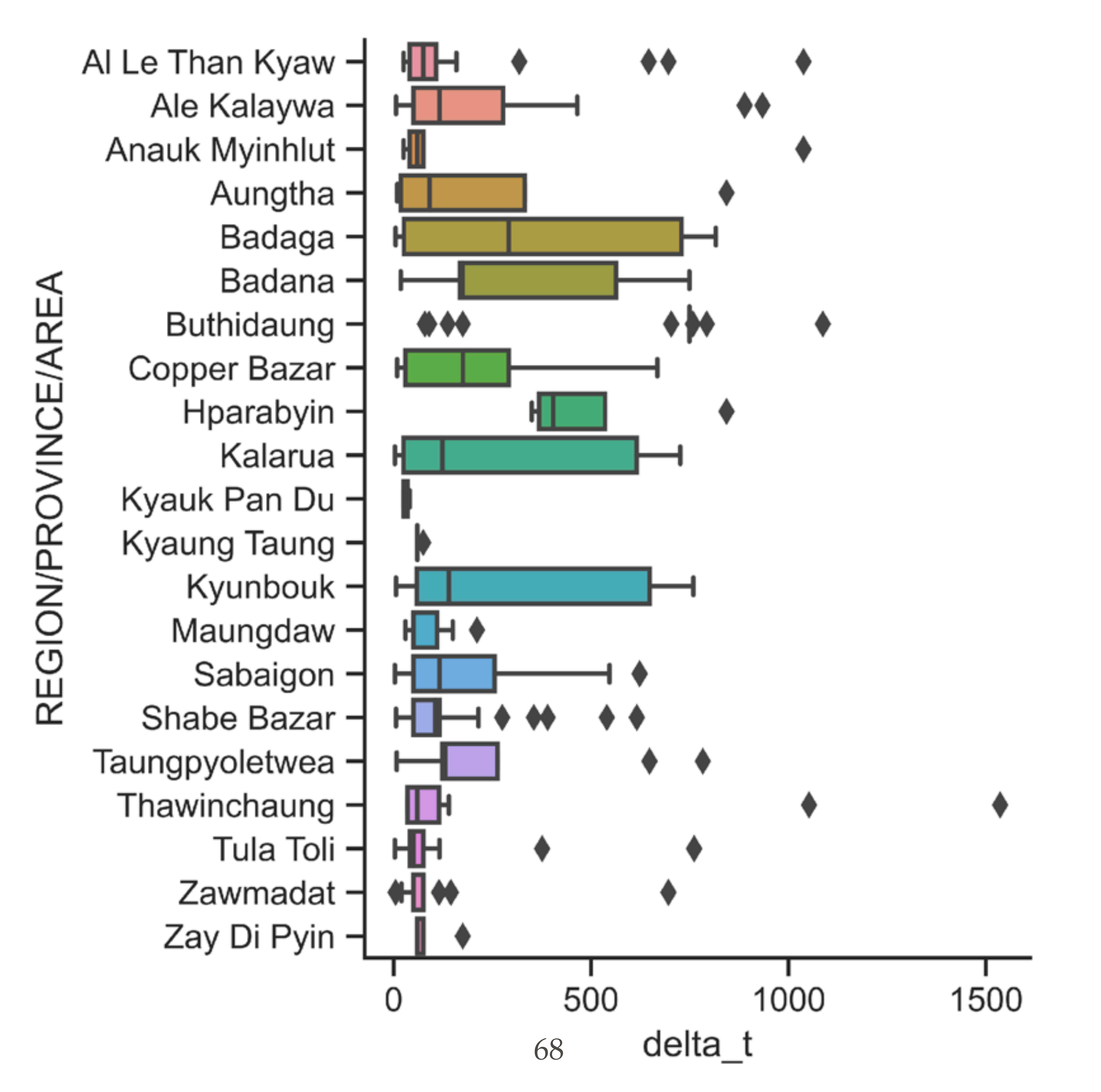


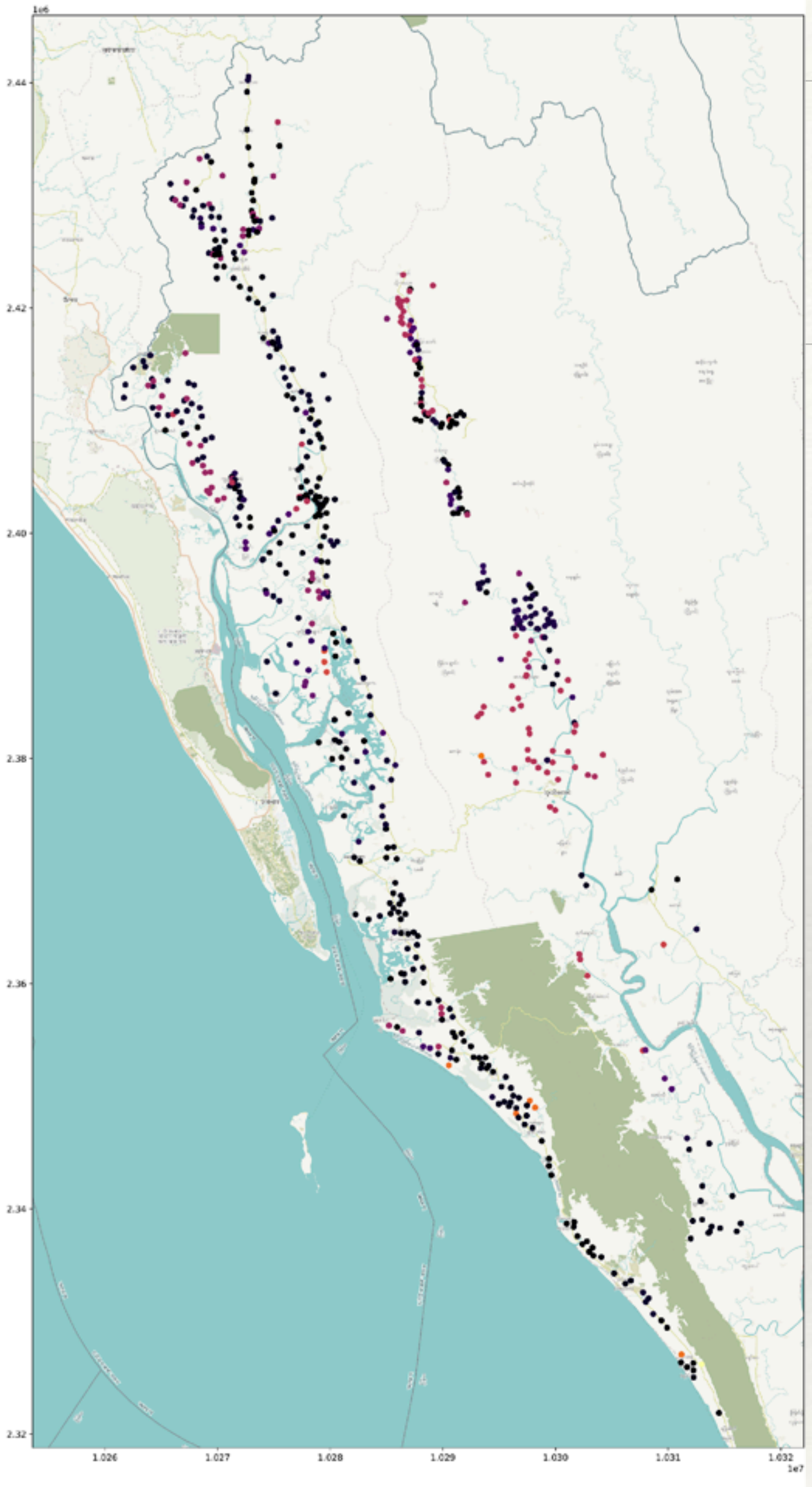
- Green: Ocelli
- Red: ACLED
- Blue: UCDP

Uncertainty in UCDP data (Rakhine)

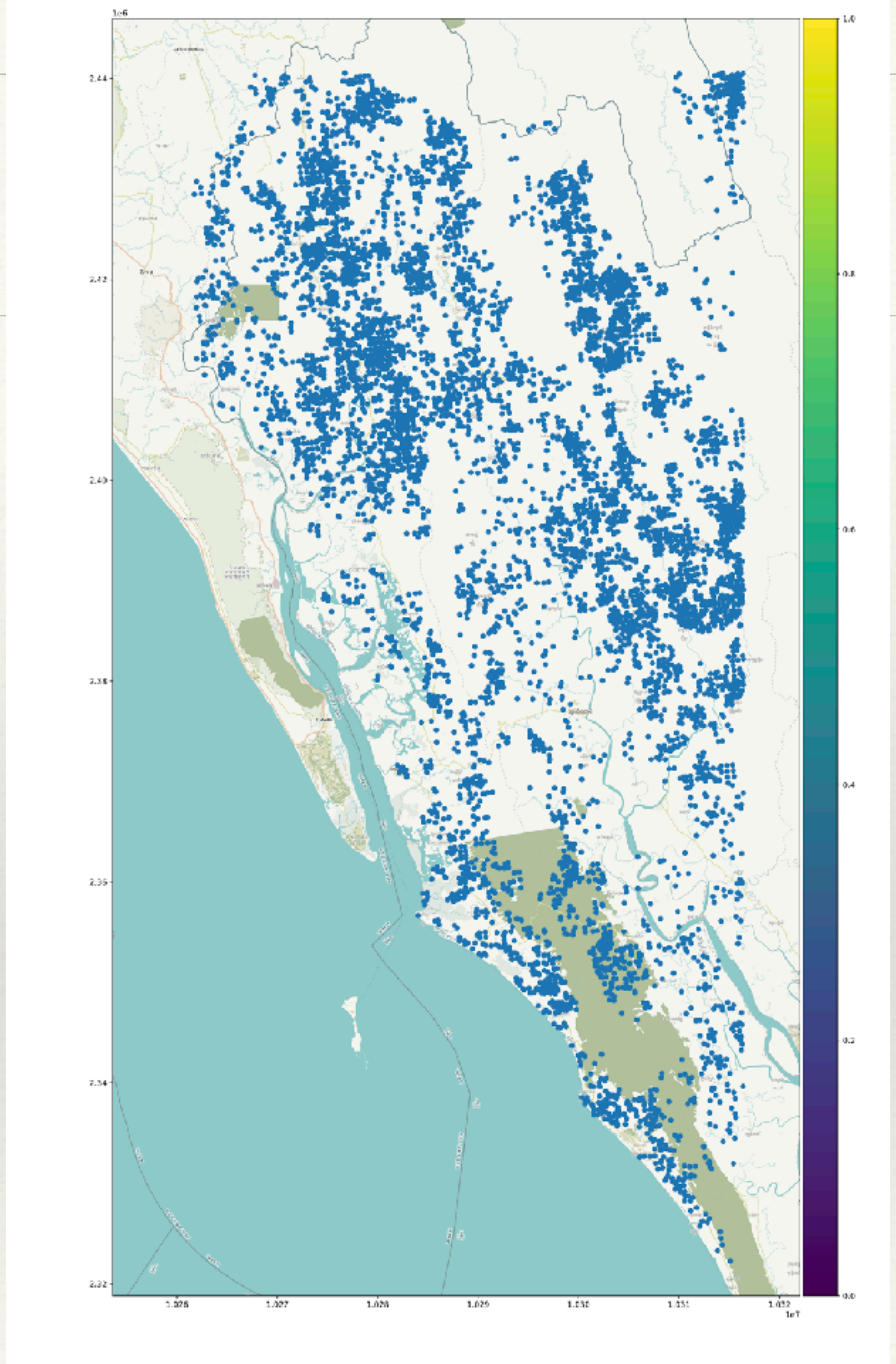


Uncertainty in Ocelli data (Rakhine)





Event data (Ocelli)



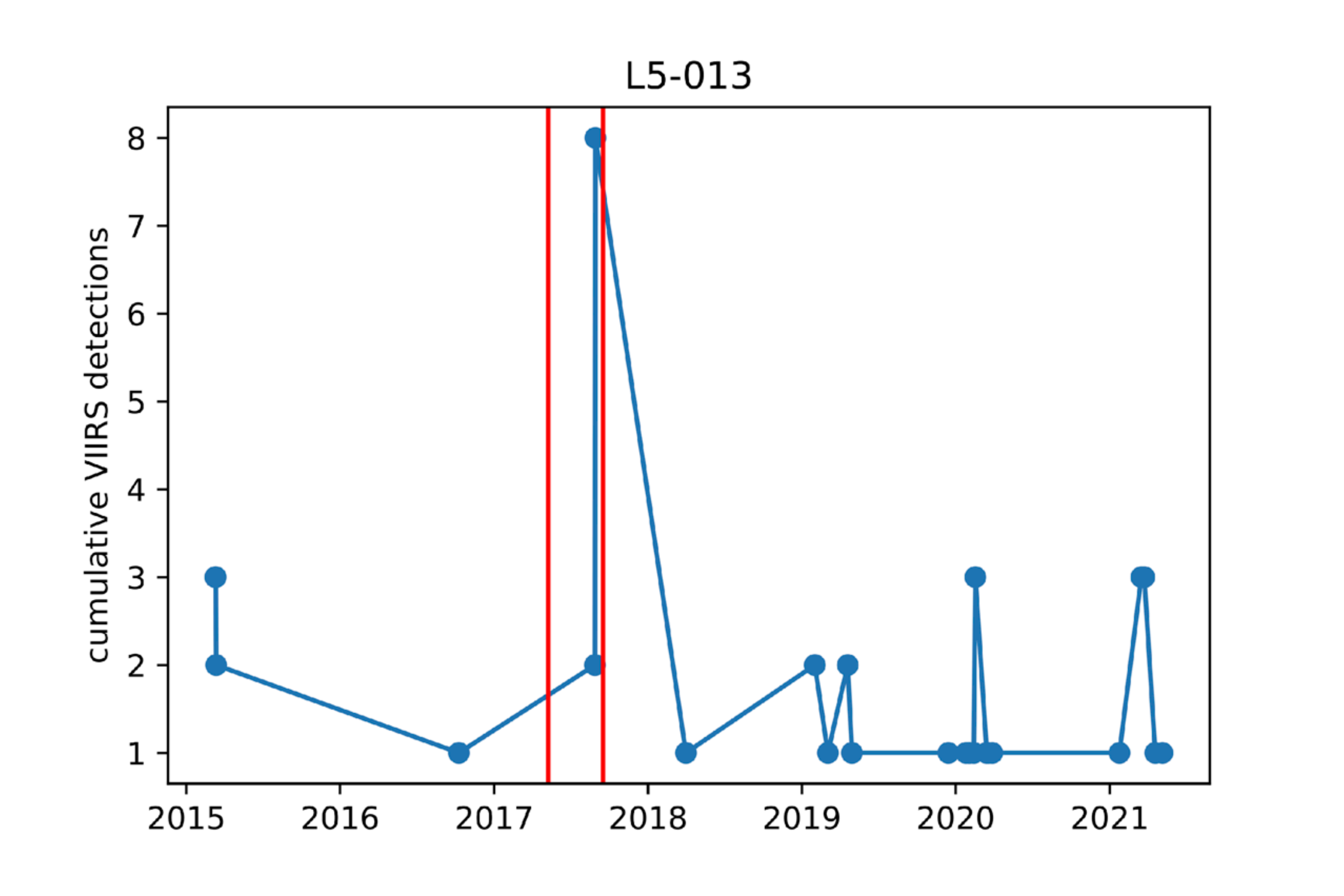
Fire data (FIRMS)

Method

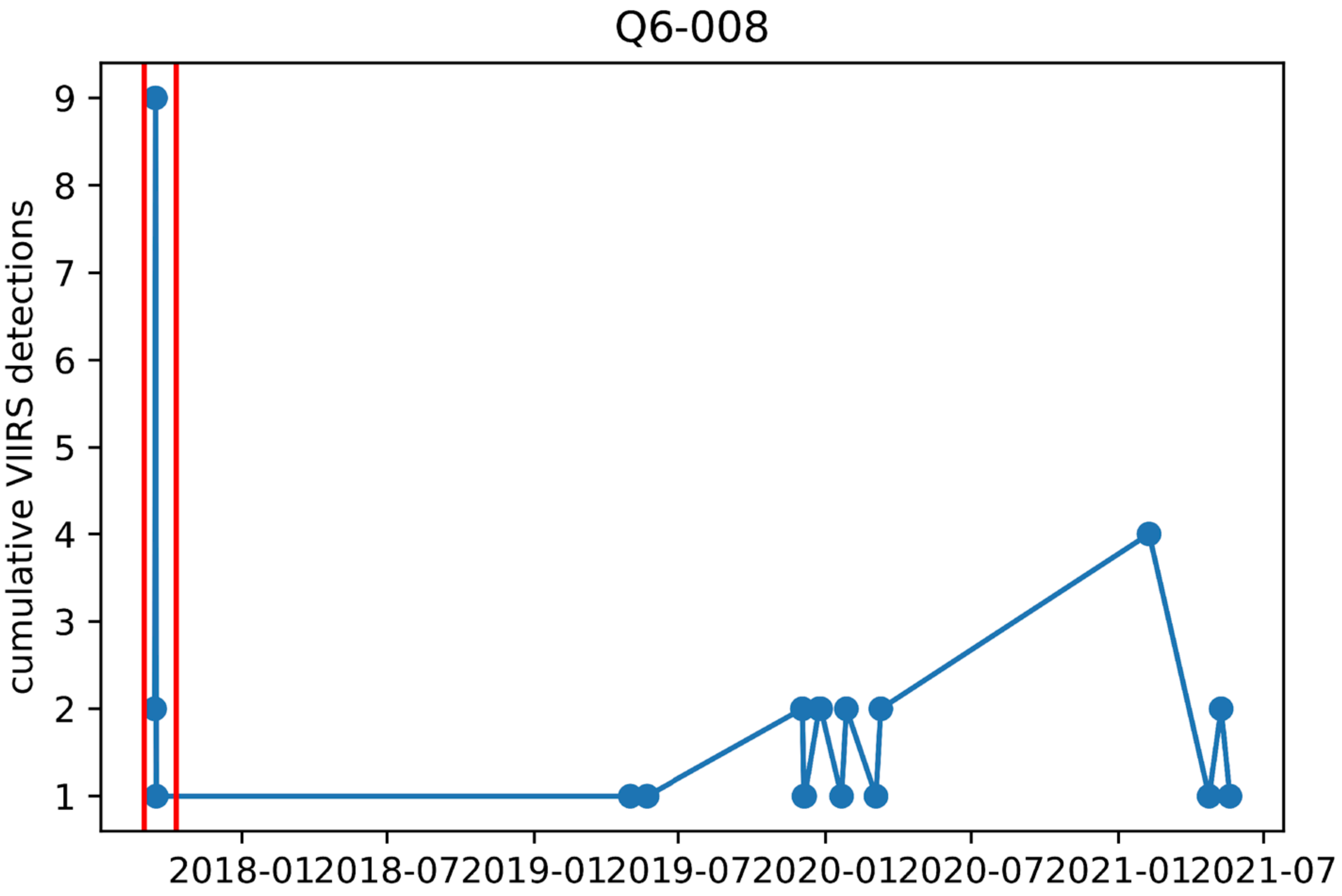
Fire data is highly seasonal due to regular wildfires and agricultural burnings.

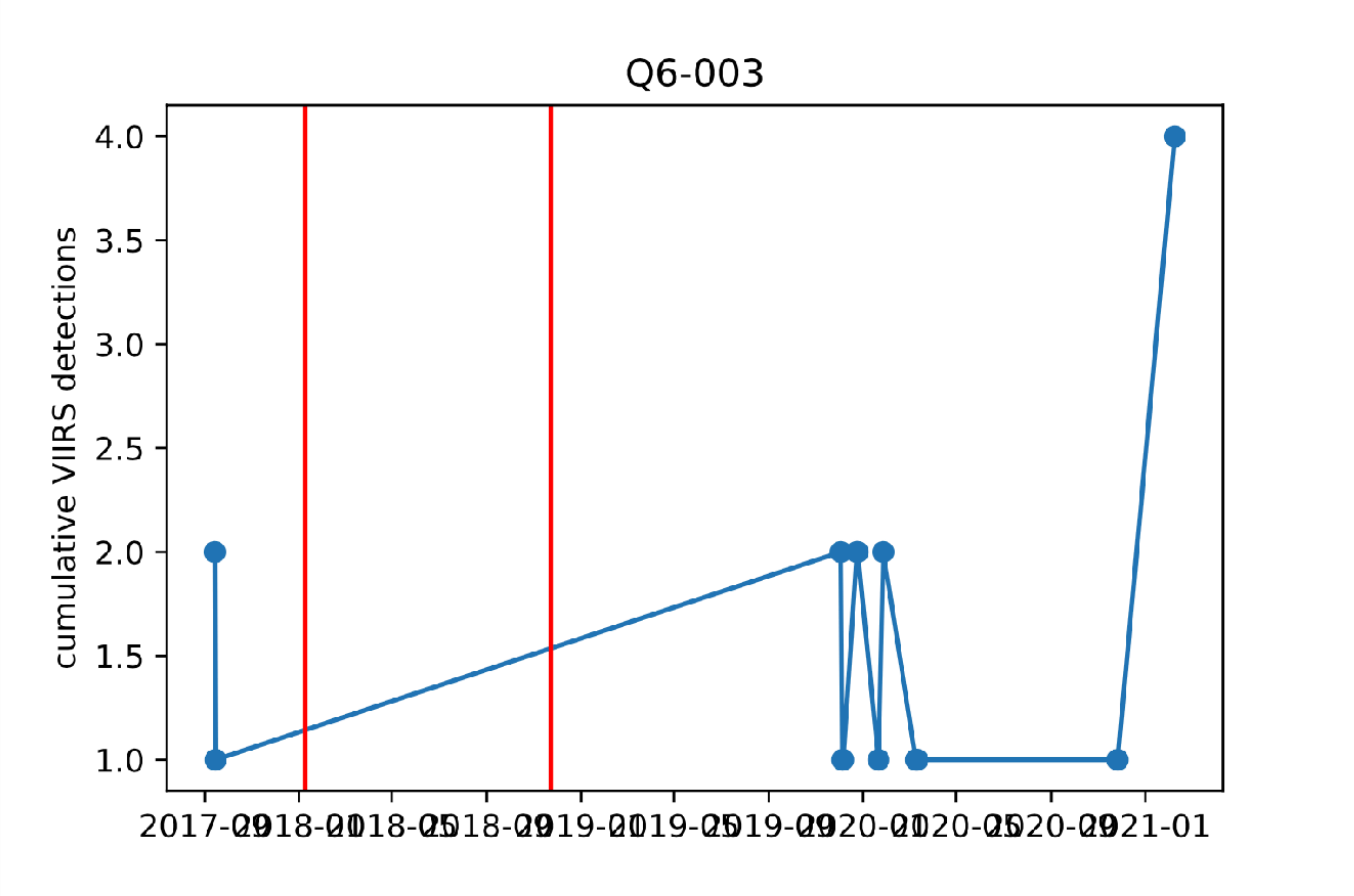
Extract localised burnings in radius about towns and compare to local minima.

Anomaly detection methods compared: Naive KDE's, ARIMAX residuals, Seasonally adjusted temporal KDE's.



L5-013
cumulative VIIRS detections
8
7
6
5
4
3
2
1
2015
2016
2017
2018
2019
2020
2021


Q6-008
cumulative VIIRS detections
9
8
7
6
5
4
3
2
1
2018-01
2018-07
2019-01
2019-07
2020-01
2020-07
2021-01
2021-07


Q6-003
cumulative VIIRS detections
4.0
3.5
3.0
2.5
2.0
1.5
1.0
2017-09
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09
2020-01
2020-05
2020-09
2021-01

Overall performance

Seasonal KDE accuracy:

- Pick a town-day at random. 13% of the time it will fall within an Ocelli window. I.e., a 13% chance of correctly “timing” an event.

- Using the anomalous fires, the same happens 41% of the time

- If we only consider only areas where burning actually took place, then 63%.

Why it matters

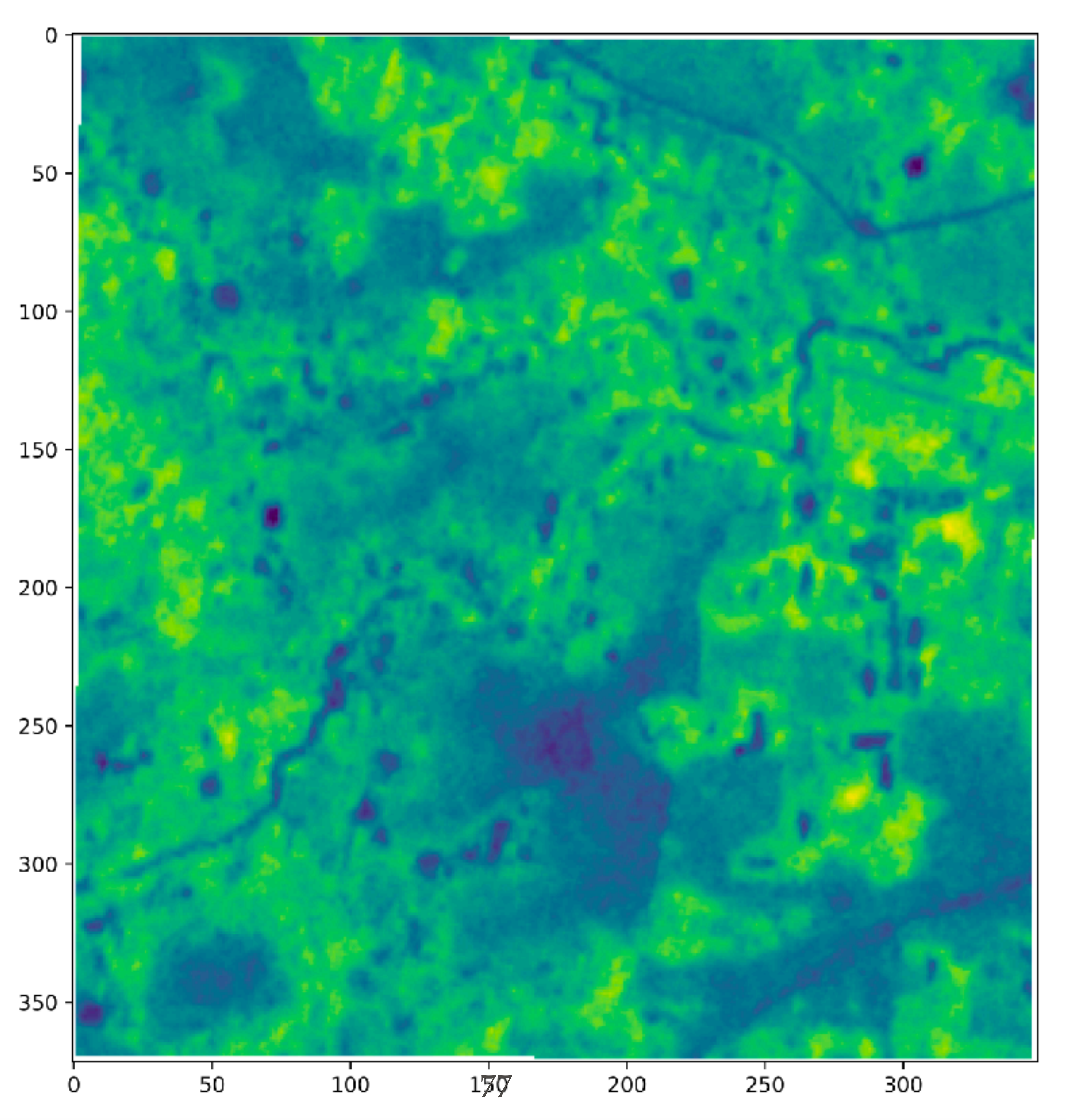
- News-based reporting often lacks precision. Also makes real-time reporting impossible
- Solution: satellite data. BUT expensive
- -> need to know where to look
    -

Fires: pros and cons

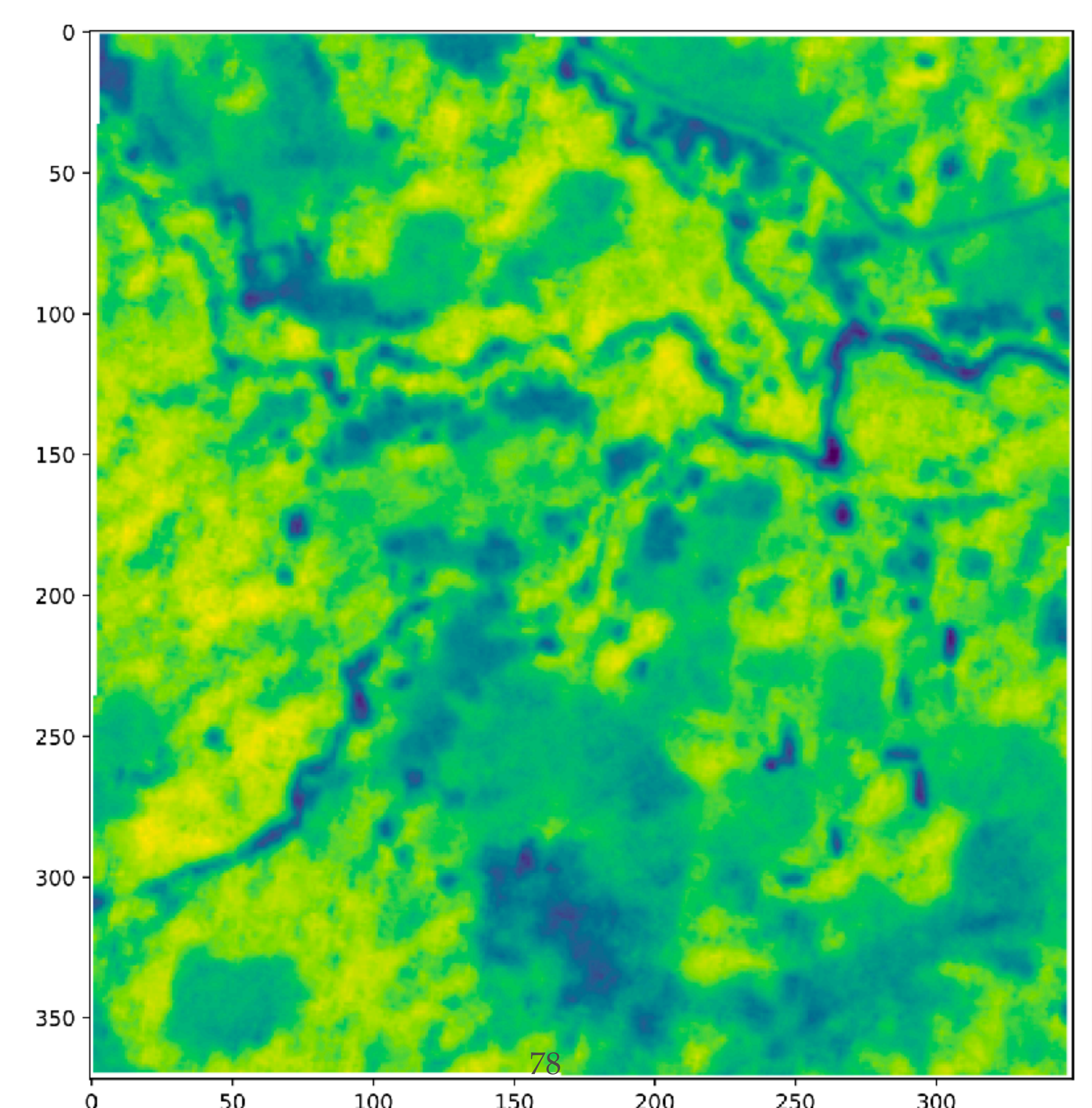
- Pros:
  - go back 20 years
  - cheap
  - instantaneous (3 hrs)
  - Allow to estimate news-based datasets' underreporting factor
-

- Cons:
  - Limited to events that cause fires
- Low accuracy for now

Next steps: Reflectance



Next steps: Reflectance



Next steps

- Assign fires to closest Ocelli event, as opposed to all within a radius -> accuracy will improve
- Remove false positives in uninhabited locations
- Earth mover distance
- Improve seasonal KDE

Project 5

Dynamic Synthetic Controls: Accounting for Varying Speeds in Comparative Case Studies

Jian Cao & Thomas Chadefaux

Research Question

One of the most popular comparative analysis methods, synthetic control, **cannot account for potential different speeds** in time series.

Research Question

One of the most popular comparative analysis methods, synthetic control, **cannot account for potential different speeds** in time series.

Can we design a method that can solve this problem?

Outline

1. Limitation of Synthetic Control

Outline

1. Limitation of Synthetic Control
2. Two-Fold Dynamic Time Warping

Outline

1. Limitation of Synthetic Control
2. Two-Fold Dynamic Time Warping
3. Simulation Study

Outline

1. Limitation of Synthetic Control
2. Two-Fold Dynamic Time Warping
3. Simulation Study
4. Re-run Abadie's Analyses

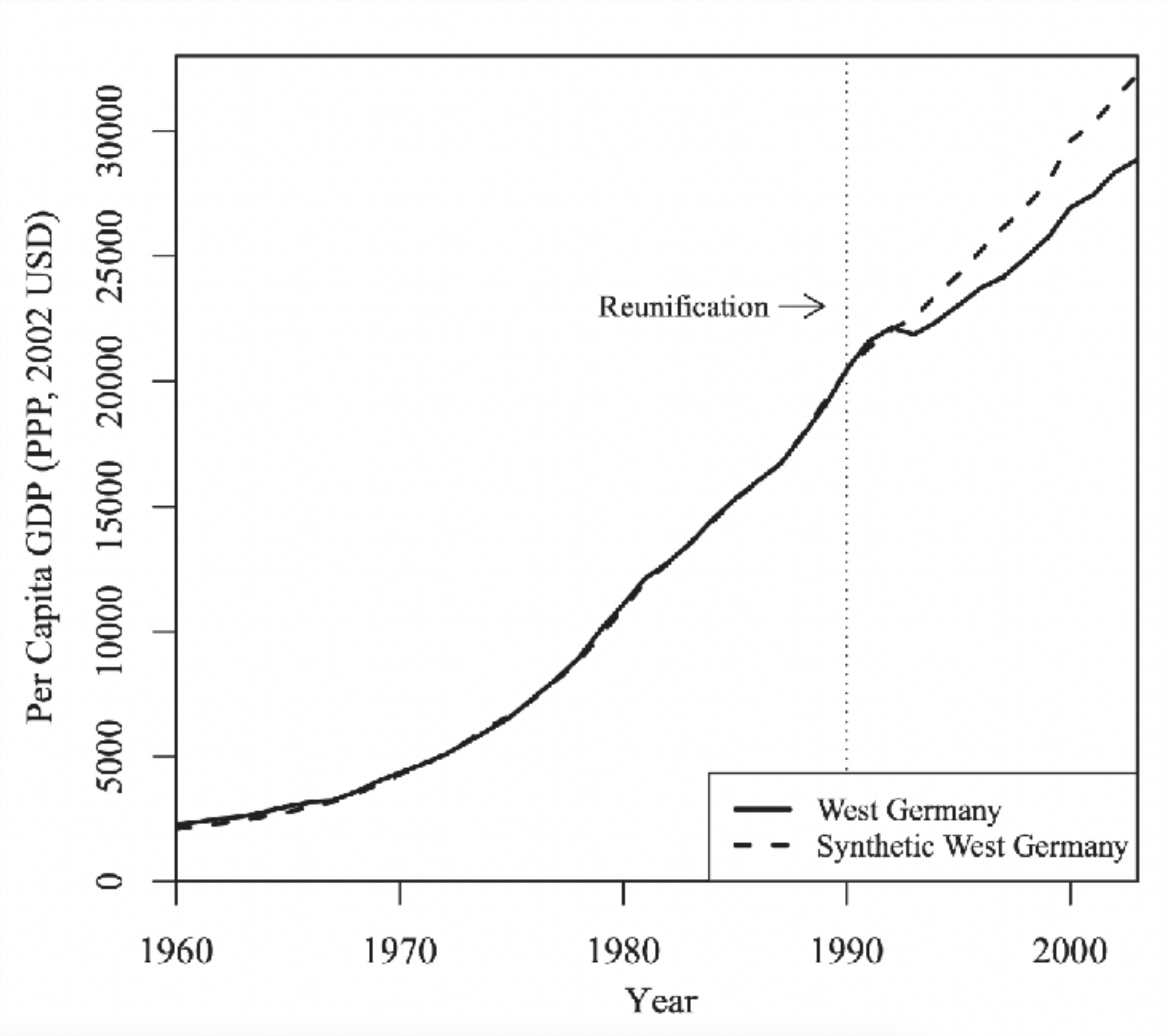
Outline

1. Limitation of Synthetic Control
2. Two-Fold Dynamic Time Warping
3. Simulation Study
4. Re-run Abadie's Analyses
5. Conclusion

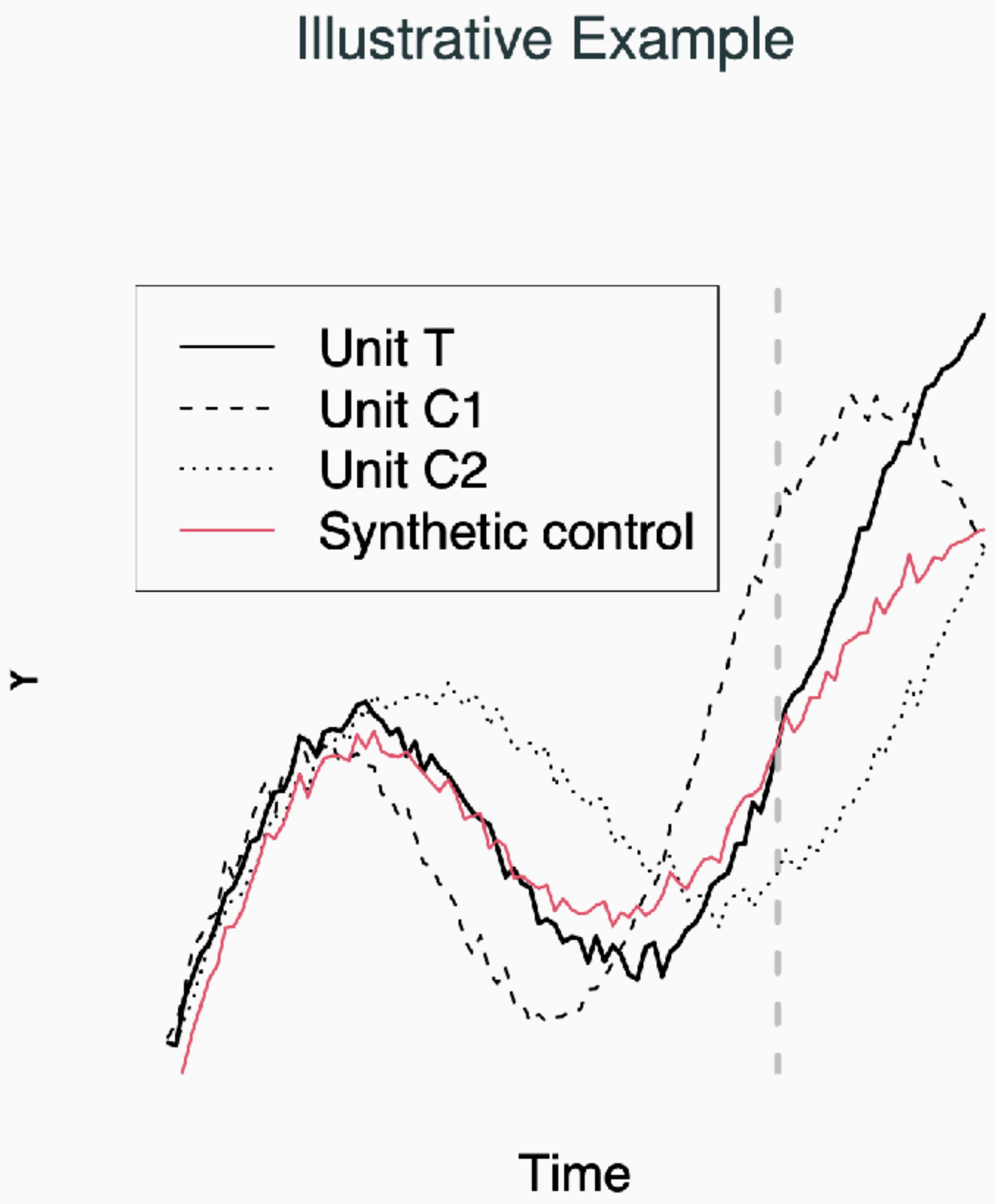
Limitation of Synthetic Control

Synthetic Control

The Economic Cost of the 1990 German Reunification



Synthetic Control (Different Speed)



Two-Fold Dynamic Time Warping

What is TFDTW?

- Two-Fold Dynamic Time Warping (TFDTW) is a new DTW method that is designed specifically for comparative case studies, and helps them account for different speeds in time series.

What is TFDTW?

- Two-Fold Dynamic Time Warping (TFDTW) is a new DTW method that is designed specifically for comparative case studies, and helps them account for different speeds in time series.
- It is a pre-processing method.

What is TFDTW?

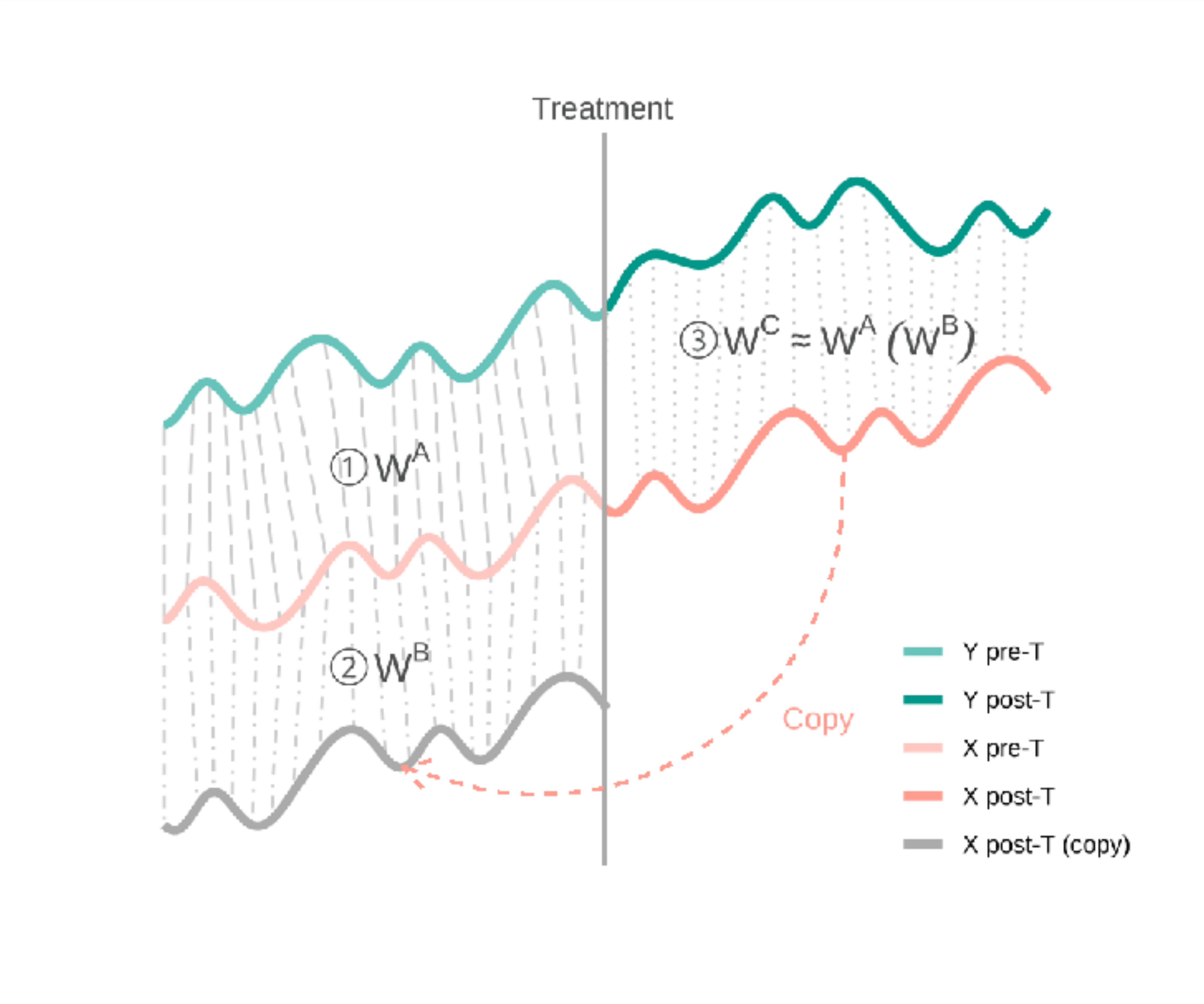
- Two-Fold Dynamic Time Warping (TFDTW) is a new DTW method that is designed specifically for comparative case studies, and helps them account for different speeds in time series.
- It is a pre-processing method.
- It warps time series to match the speed of the target time series.

What is TFDTW?

- Two-Fold Dynamic Time Warping (TFDTW) is a new DTW method that is designed specifically for comparative case studies, and helps them account for different speeds in time series.
- It is a pre-processing method.
- It warps time series to match the speed of the target time series.
- Any comparative case study methods, e.g. synthetic control, can be used after TFDTW as differences in speeds have been removed.

Process of TFDTW

Two-Fold Dynamic Time Warping



Simulation Study

Simulation Study

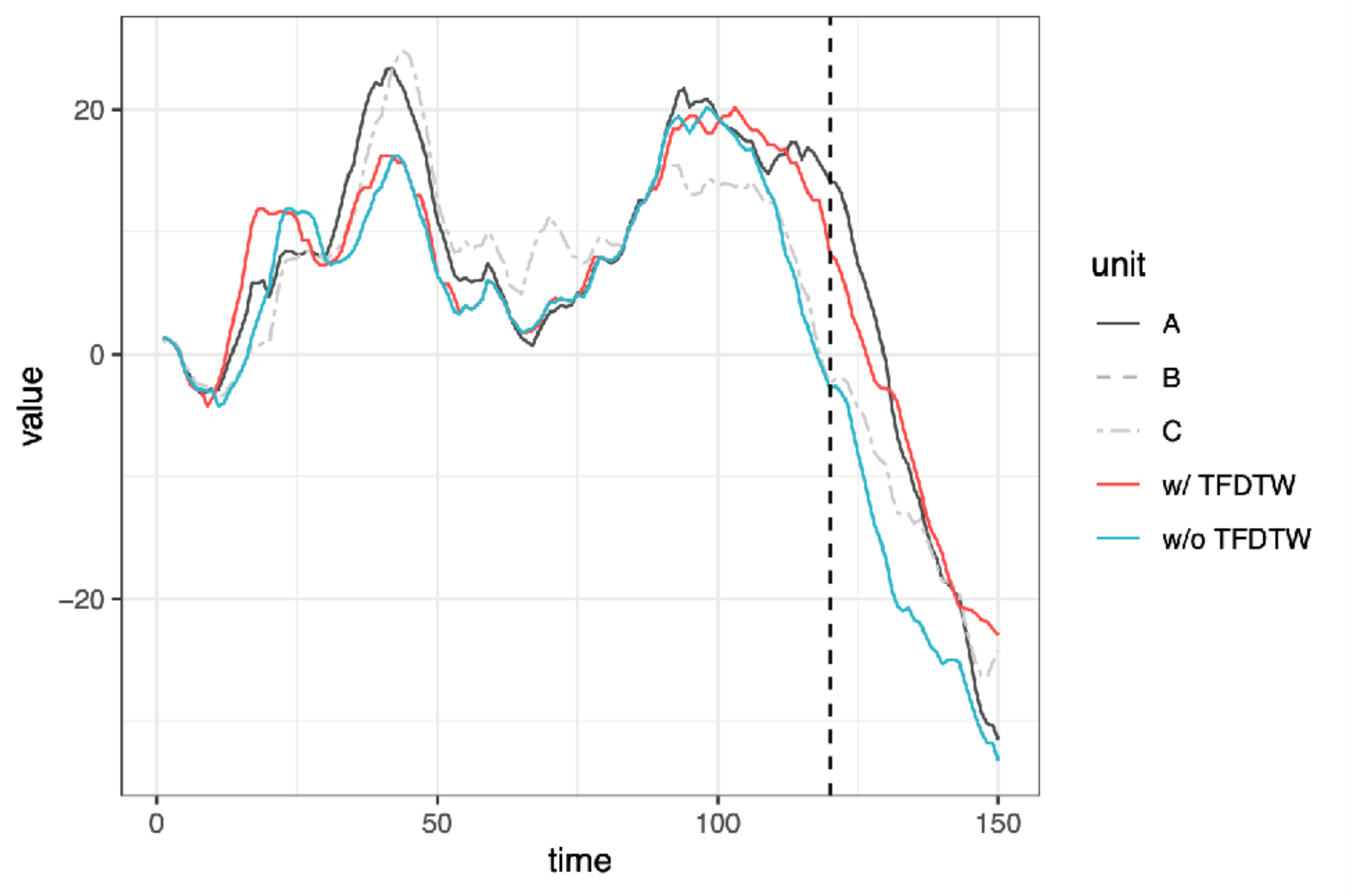
- Data generation process (DGP):

$$y_{i,t} = \sigma_t + \sum_{l=1}^{p} (\beta_{i,t,l} y_{i,t-l} + \phi_{i,t,l} x_{i,t-l}) + \varepsilon_{i,t}$$

- Simulate 1,000 data sets

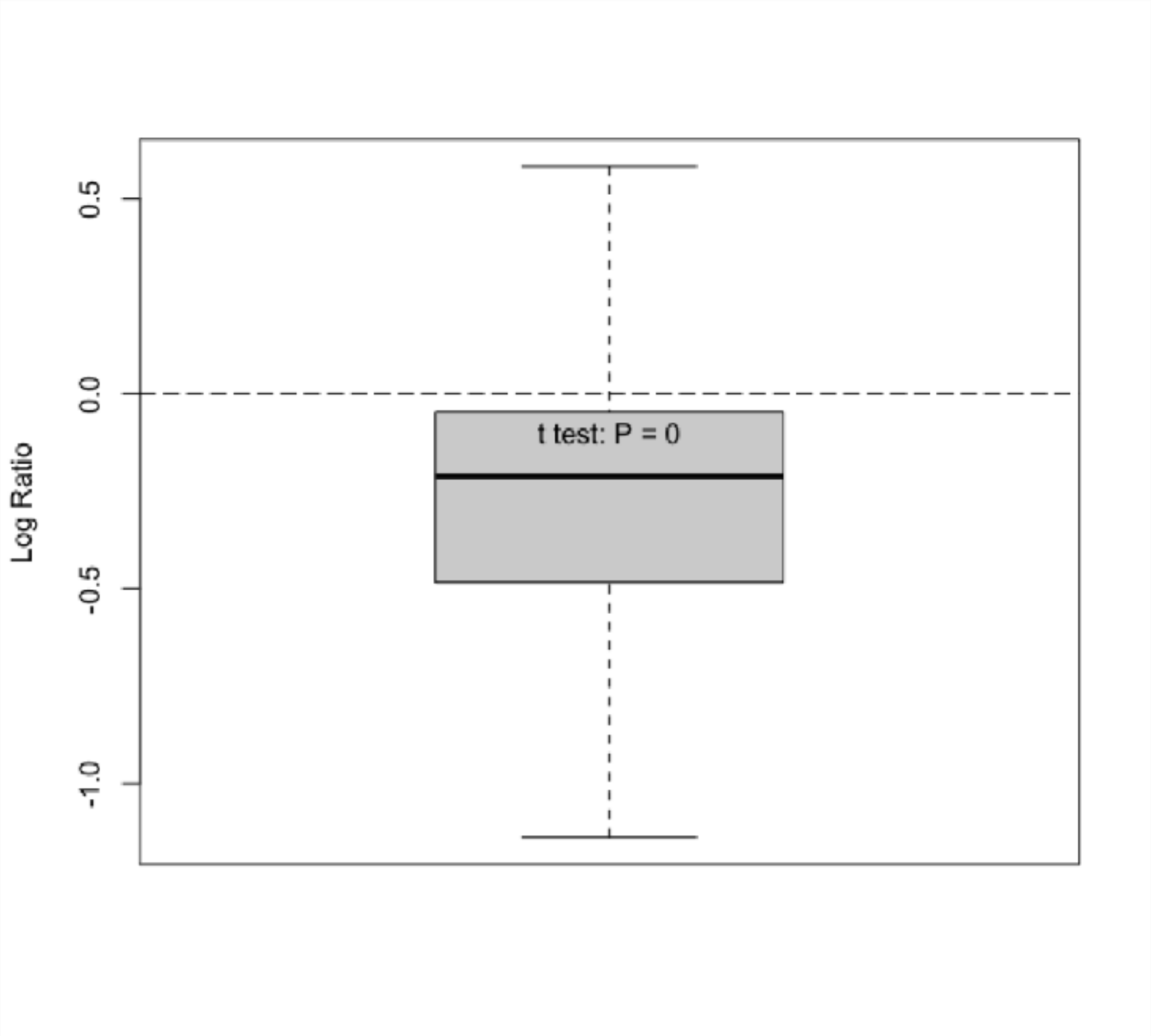
Simulation Study

Simulation Example



Simulation Study

Simulation Result: 1000 Data Sets



$$\text{Log Ratio} = log(MSE_{w/\ TFDTW}/MSE_{w/o\ TFDTW})$$

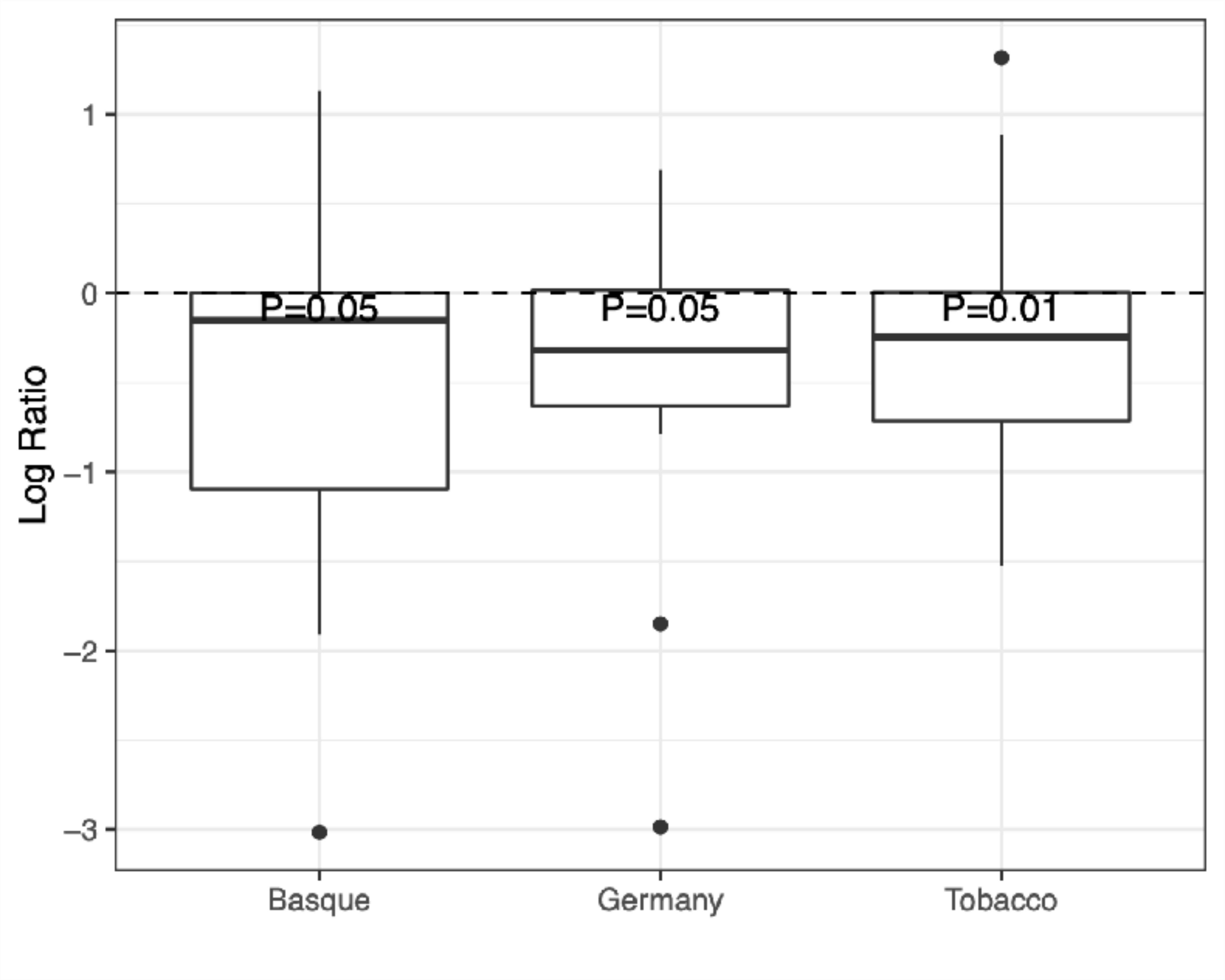
Re-run Abadie's Analyses

Abadie's Analyses

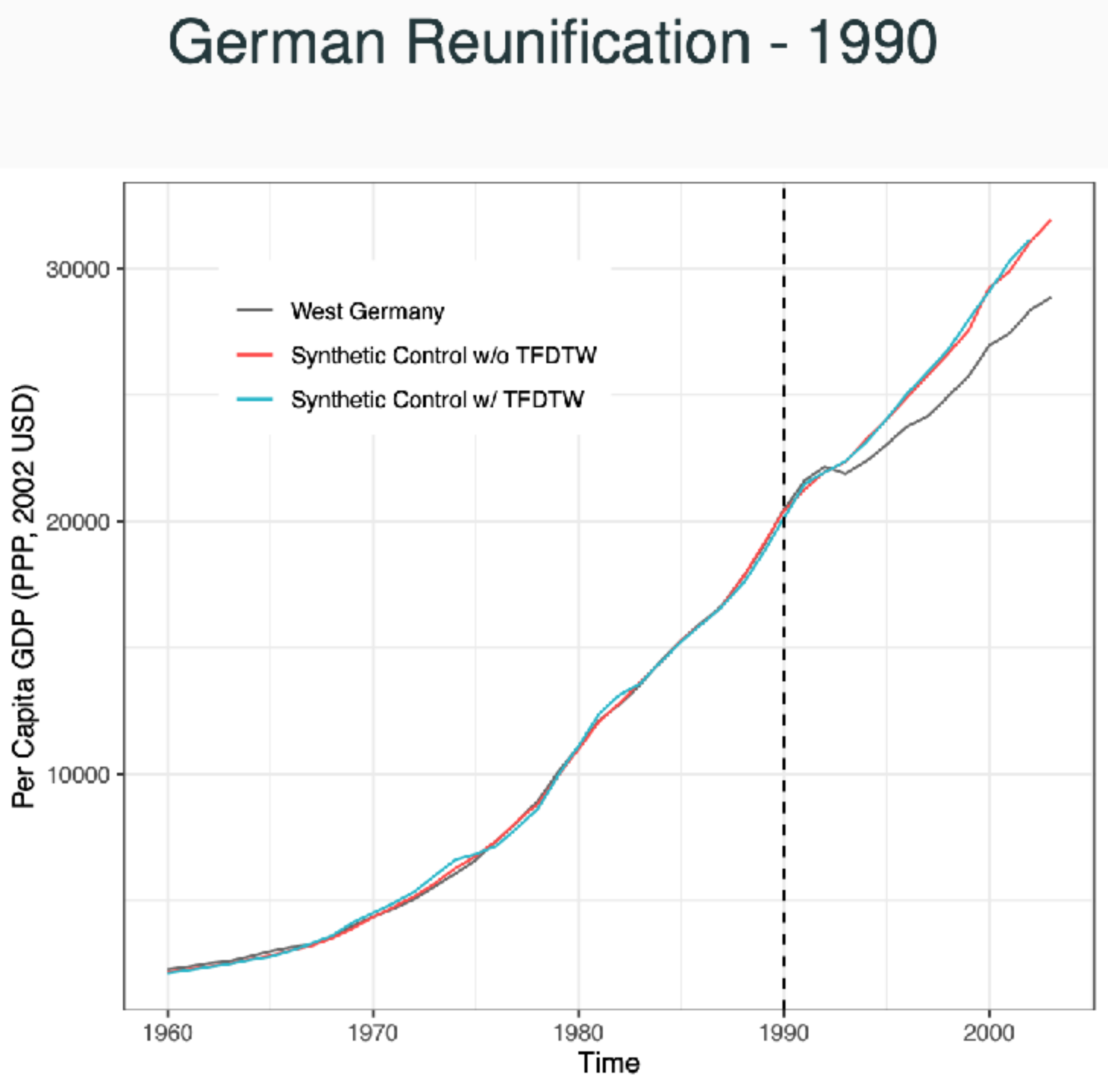
- German Reunification - 1990
- California Tobacco Law - 1989
- Terrorism in Basque Country, Spain - 1970

Results

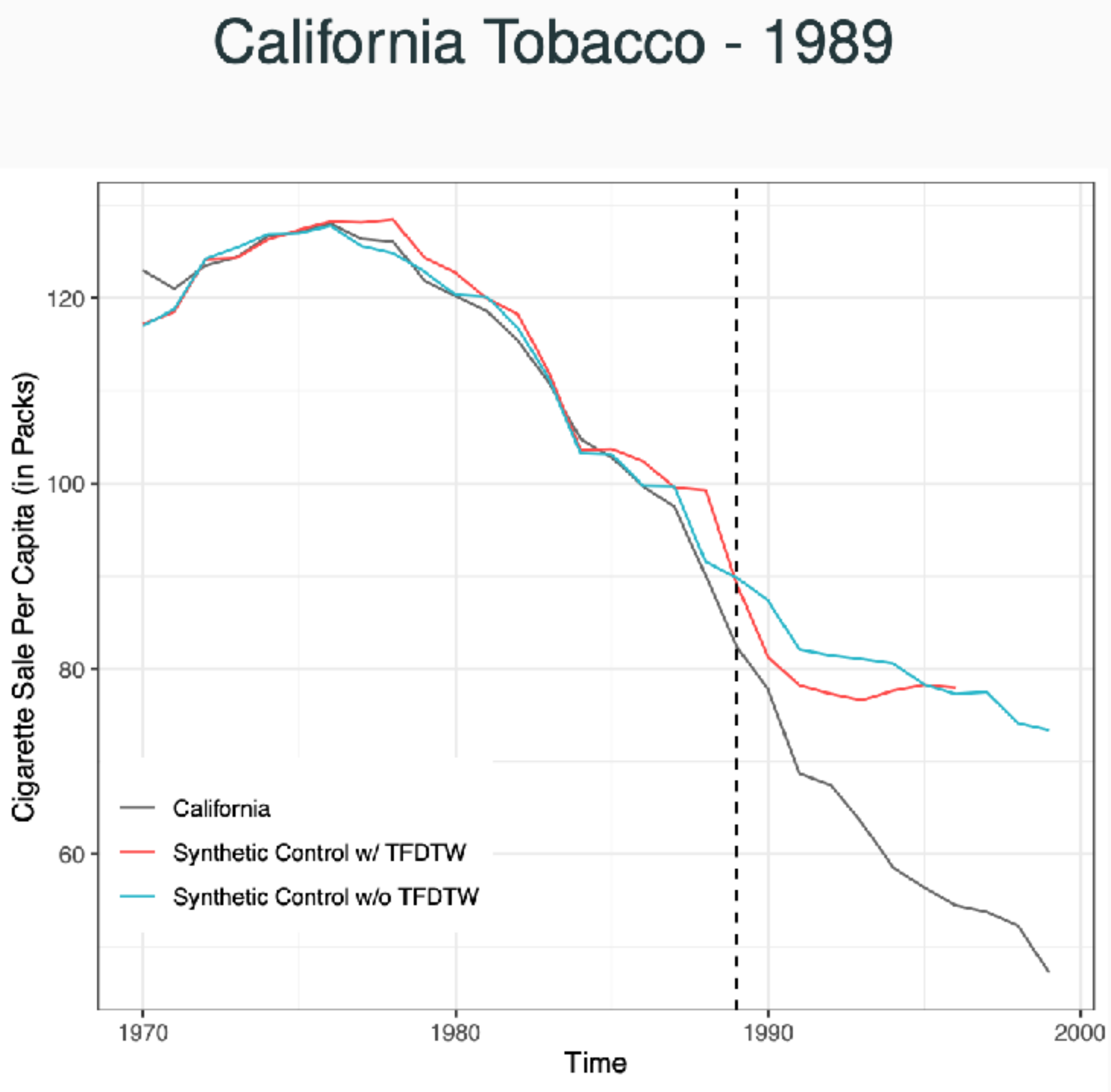
$$\text{Log Ratio} = log(MSE_{w/\ TFDTW}/MSE_{w/o\ TFDTW})$$



Does TFDTW Change Abadie's Results?

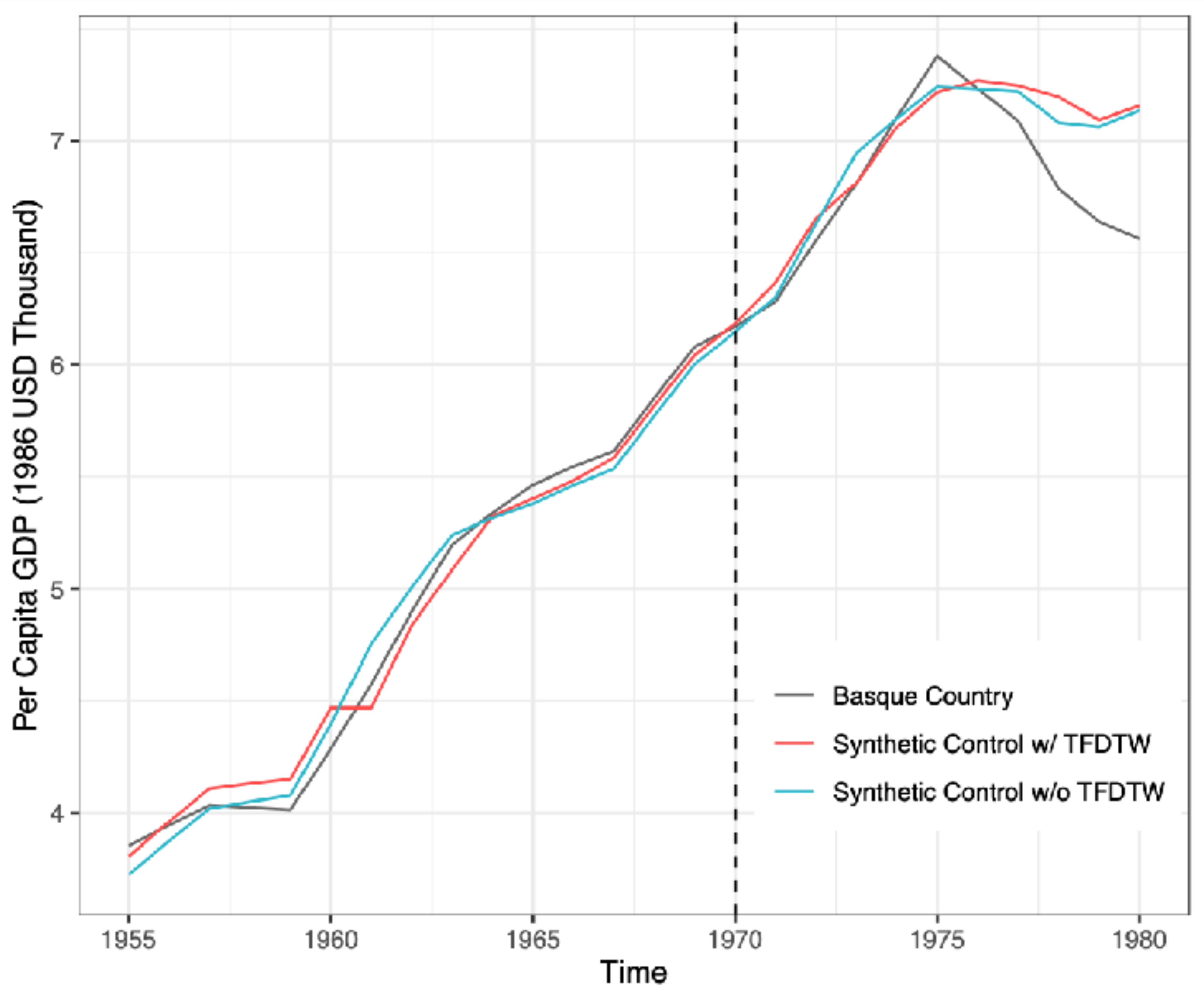


Does TFDTW Change Abadie's Results?



Does TFDTW Change Abadie's Results?

Basque Country, Spain - 1970



Conclusion

- TFDTW significantly improves synthetic control's ability to account for different speeds in time series.

Conclusion

- TFDTW significantly improves synthetic control's ability to account for different speeds in time series.
- After applying TFDTW in Abadie's analyses, the estimated causal effects are changed.
  - German reunification ↓
  - California tobacco law ↓
  - Basque country terrorism ↑