Optimizing Anomaly Detection in COVID-19 Mortality Time Series: Modeling Epidemic Waves in Brazil

Amal Bilal[1,2], Xin Wang[1], Tengfei Ma[3]

Affiliations: Stony Brook University College of Engineering and Applied Sciences[1], Simons STEM Scholars Program[2], Stony Brook University Department of Biomedical Informatics[3]

Machine Learning for Healthcare (MIT - HST.925)

- Introduced to core biomedical informatics topics
  - clinical data
  - time-series modeling
  - healthcare data ethics and challenges
- Examined real-world examples and case studies

Course Page: https://mlhcmit.github.io/2024/index.html

Background

- Brazil experienced three distinct waves of COVID-19 which resulted in major increases in incidenece and mortality
- Severity of cases were influenced by factors such as:
  - vaccine coverage
  - public health measures
  - population density

Goals

1. Understand and predict state-level anomalies characterized by sudden significant changes in metrics
2. Visualize the severity of anomalies across the three waves of COVID-19 in Brazil
3. Tune hyperparameters to maximize alignment precision within waves without overfitting.

Methodology

- **Data Preprocessing**
  1. Perform basic data analysis to understand the data
  2. Aggregate the COVID-19 mortality data at weekly levels
  3. Define the three known wave periods in Brazil

- **Anomaly Detection**
  1. q_up: detects abrupt increases
  2. q_h, q_trend: detects patterns
  3. Tune parameters to find best alignment to the wave periods

Methodology

- **Visualization**
  1. Visualize wave periods over time
  2. Plot new deaths per week
  3. Detect anomalies inside and outside of waves
- **Evaluation**
  1. Calculate the percentage of detected anomalies within the wave periods
  2. Adjust parameters to visualize the balance between sensitivity and specifiity

Data

- Over 70,000 daily entries including...
    a. location
    b. date
    c. total number of cases and deaths
    d. number of new cases and deaths
    e. vaccine information (dosages)
    f. hospital information (count, deaths, length of stay)

Data

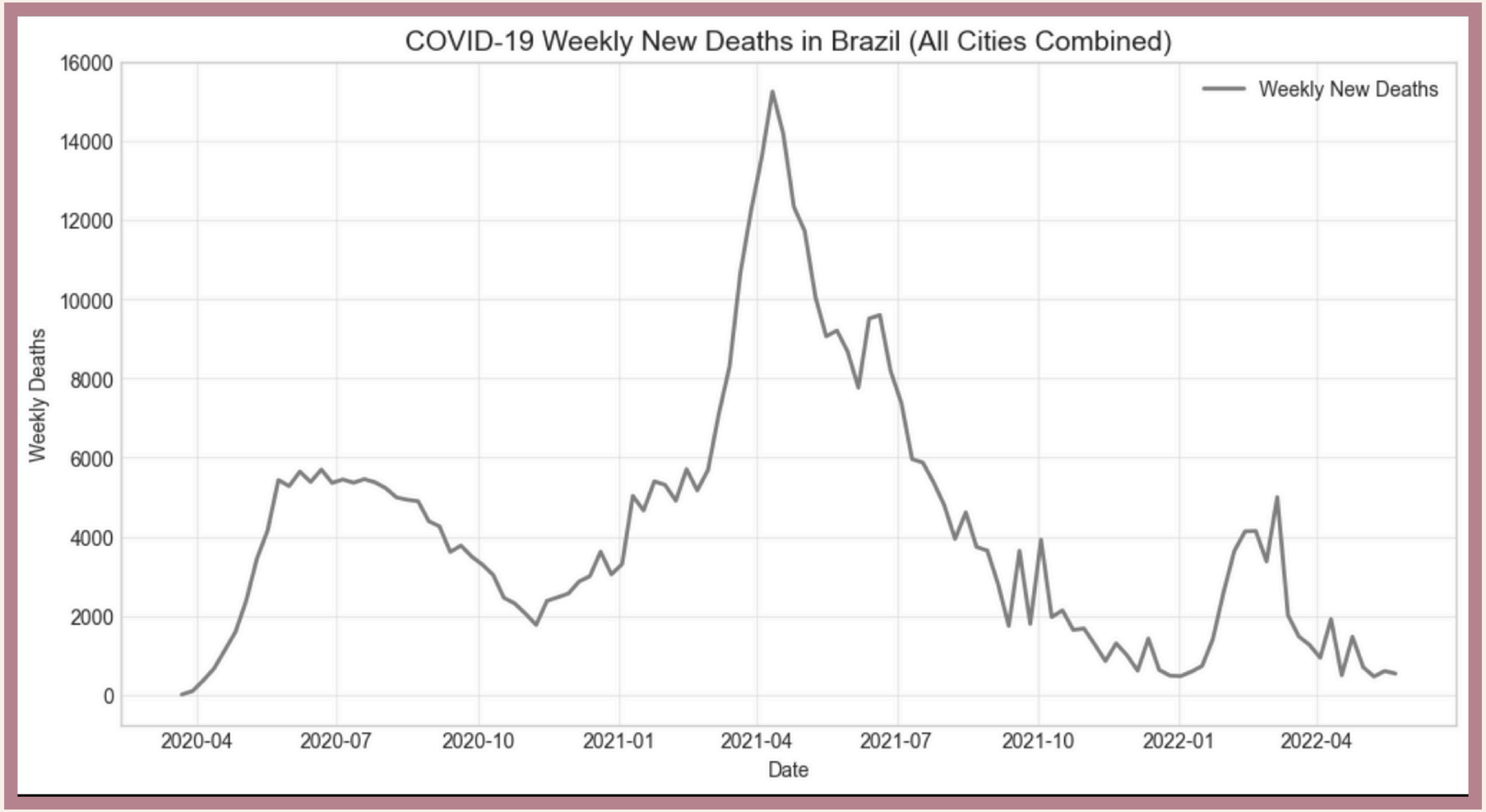
- Sample Data (random five rows)

	City	State	Date	Ct_Value	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalCases_100k_inhab	TotalDeaths_100k_inhab	...
691490	Venda Nova do Imigrante	ES	2022-02-28	NaN	8182	0	58	0	0.08182	0.00058	...
147920	Carmópolis	SE	2020-11-11	NaN	296	1	14	0	0.00296	0.00014	...
114721	Cachoeira do Sul	RS	2021-04-09	NaN	5429	82	84	2	0.05429	0.00084	...
98597	Brasília	DF	2020-08-08	26.22893	121824	1921	1712	30	1.21824	0.01712	...
343126	Joaçaba	SC	2020-07-16	NaN	179	6	1	0	0.00179	0.00001	...

5 rows × 21 columns

Time Series Data

- Weekly deaths over time



Wave Periods

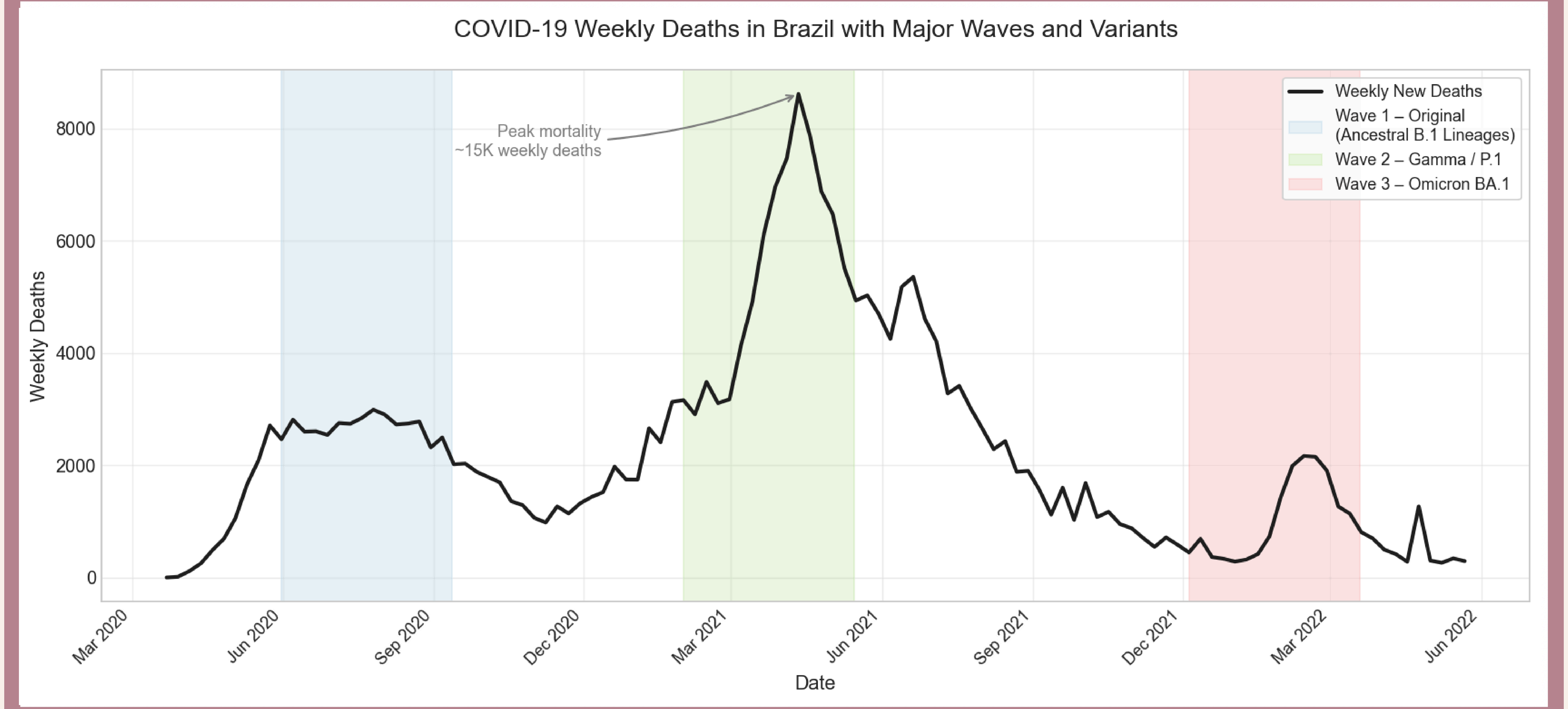
- Defining the wave periods

```
wave_periods = [
    ('2020-05-31', '2020-09-12', 'Wave 1 □ Original\n(Ancestral B.1 Lineages)'),
    ('2021-01-31', '2021-05-15', 'Wave 2 □ Gamma / P.1'),
    ('2021-12-05', '2022-03-19', 'Wave 3 □ Omicron BA.1'),
]
```

Source: https://www.mdpi.com/1999-4915/15/10/1997

Weekly Deaths

- Weekly deaths over wave periods
- Peak weekly deaths (15,254) occurred on April 11, 2021



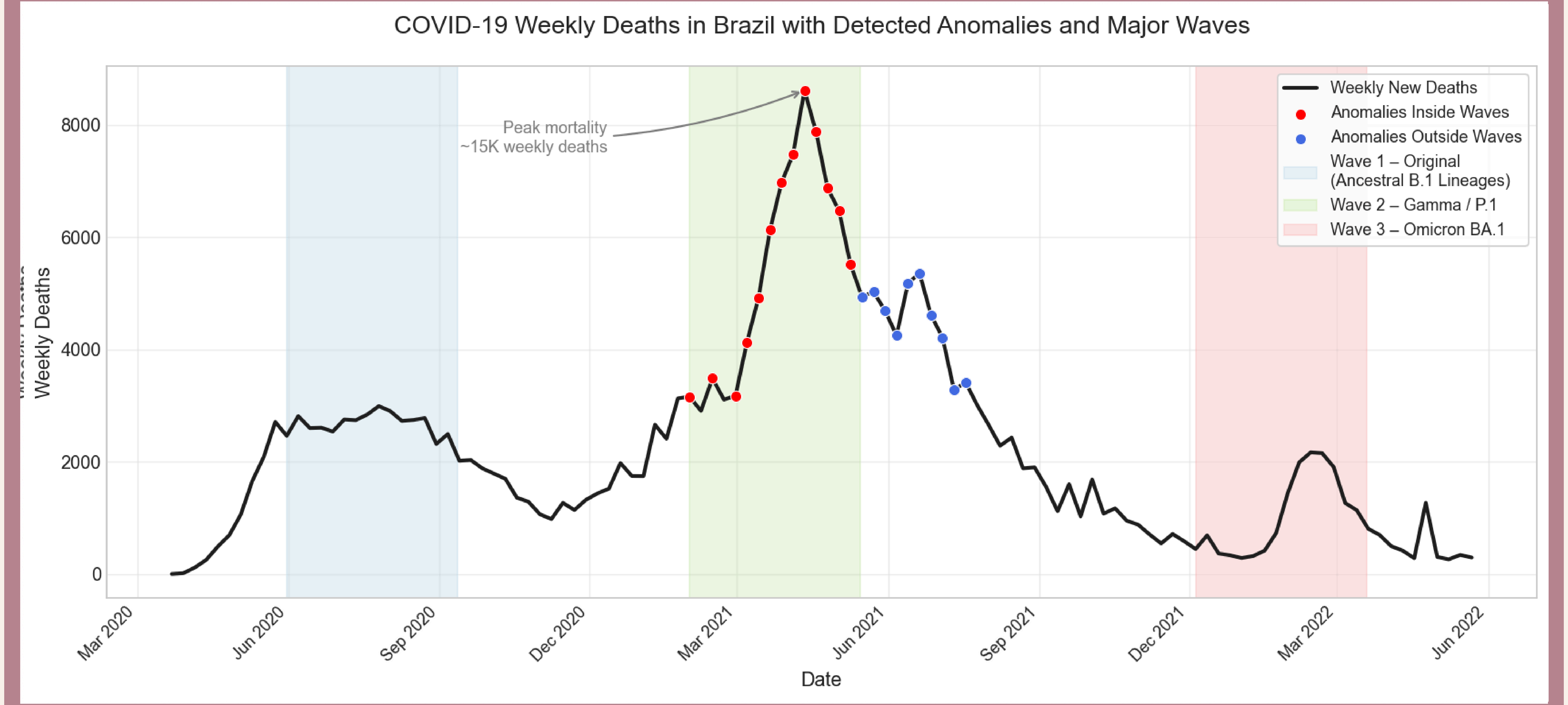
Anomaly Detection

Import anomaly detection functions:

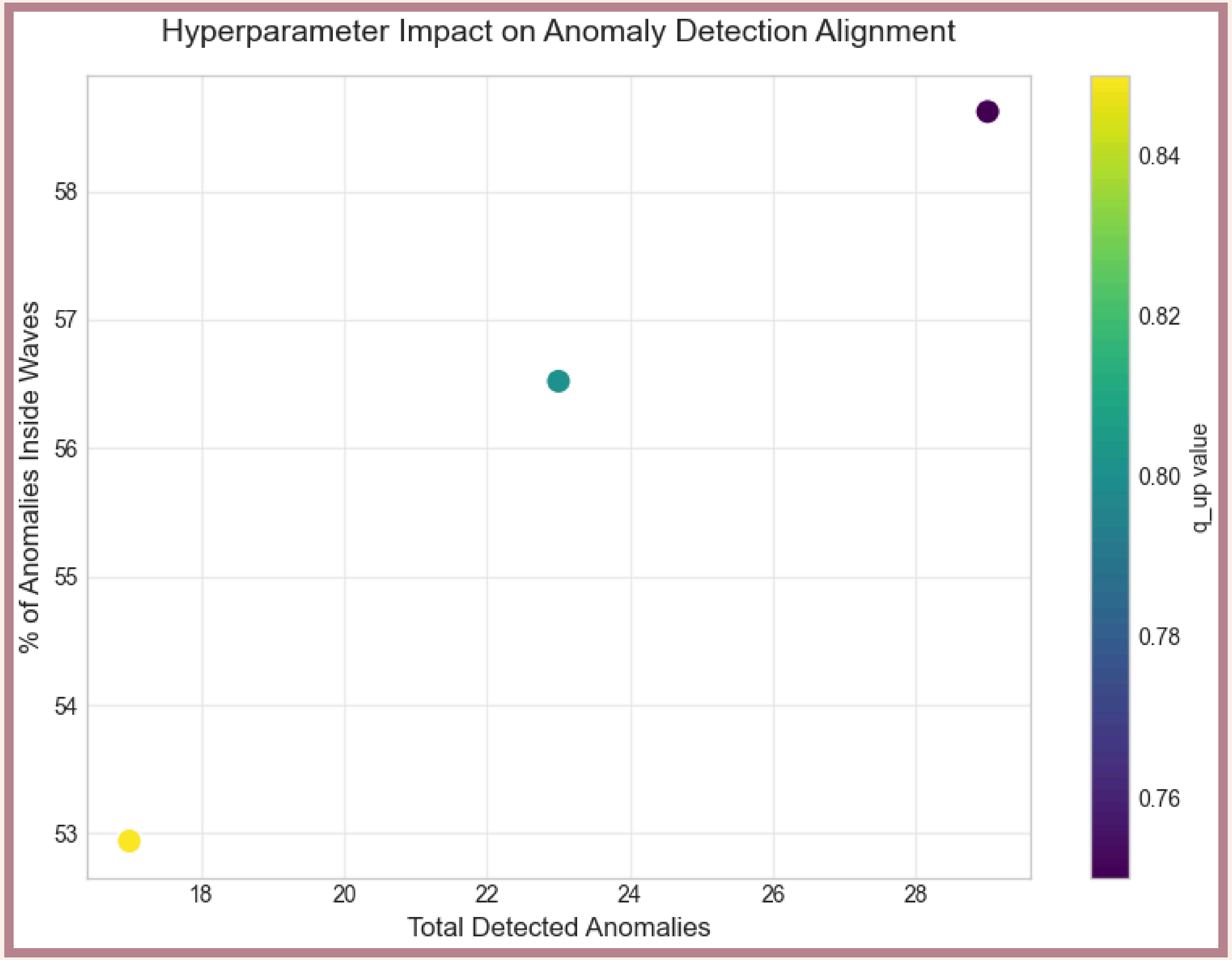
1. detect_type1_spike_up(r, q_up=0.8)
2. detect_type2_turn_patterns(r, w_pre=3, w_post=3, q_h=0.8, q_trend=0.7)

Anomaly Detection

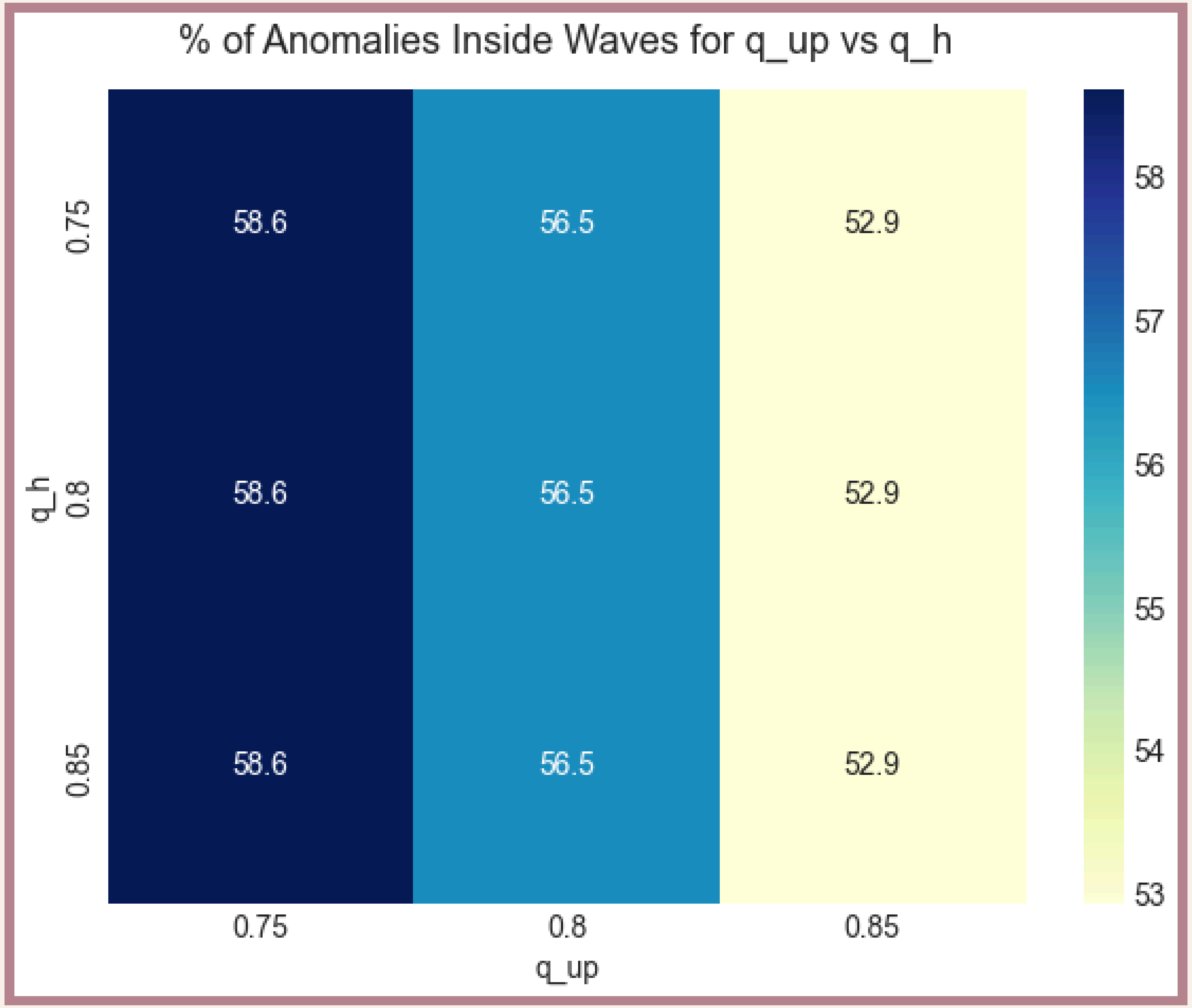
- 56.52% of 23 detected anomaly weeks fall within the three major COVID-19 waves.



Hyperparameter Tuning



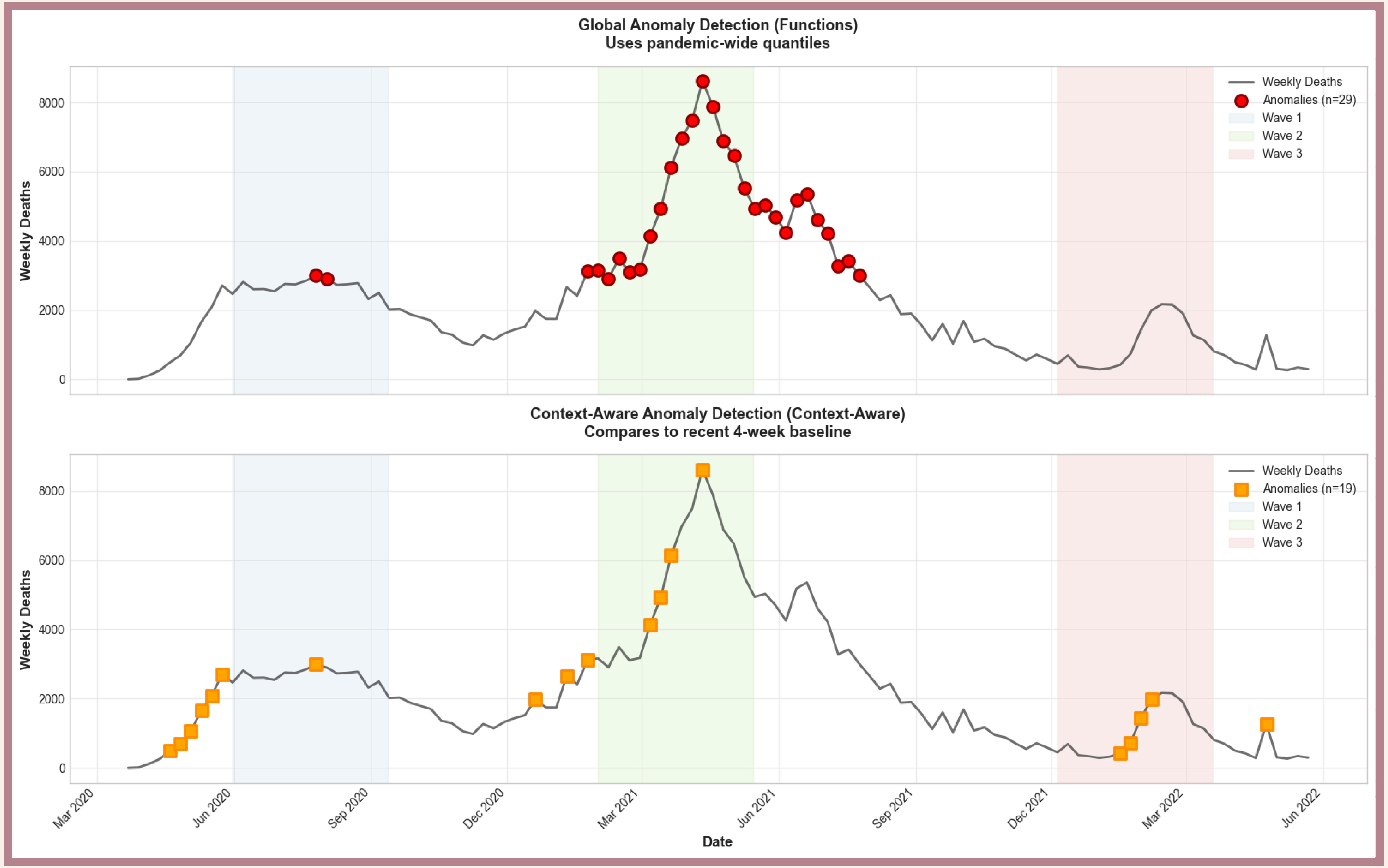
Hyperparameter Evaluation



Best Hyperparameters

q_up	0.75000
q_h	0.75000
q_trend	0.65000
Total_Anomalies	29.00000
Pct_Inside_Waves	58.62069

Global Anomaly Detection



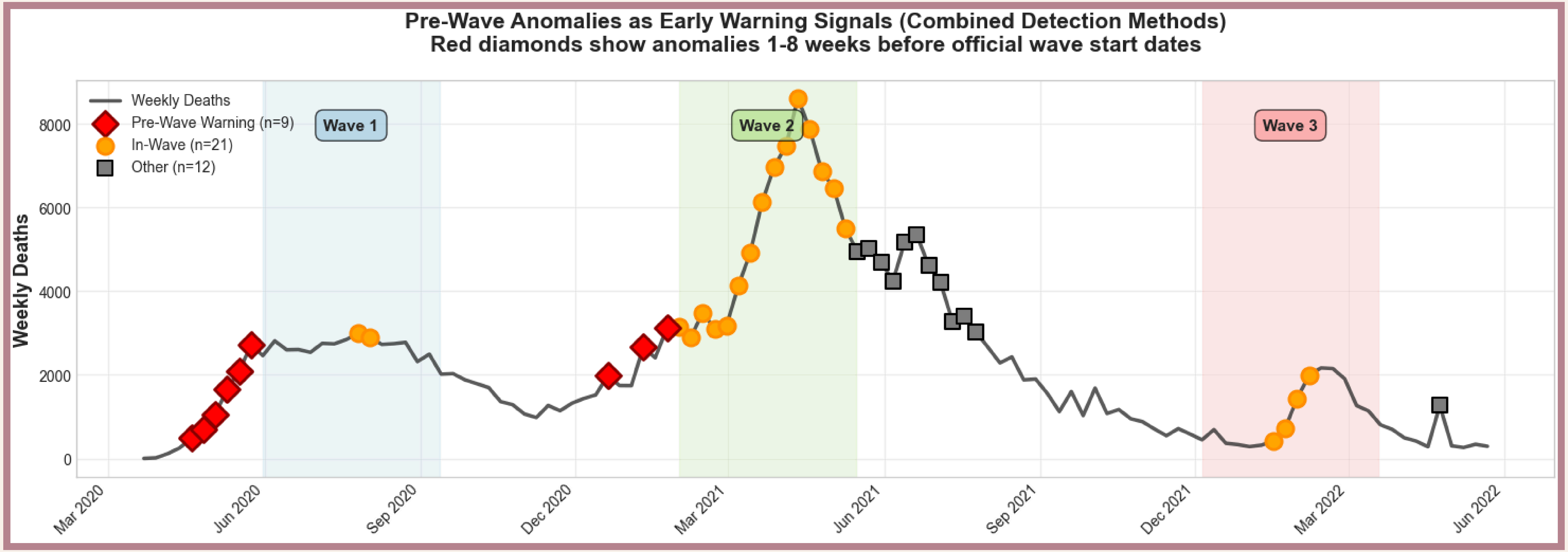
Anomaly Methods

```
Method Counts:
  Global Method (Functions):      29 anomalies
  Local Method (Context-Aware):   19 anomalies

Overlap Analysis:
  Detected by both methods:       6 anomalies
  Only by Global method:          23 anomalies
  Only by Local method:           13 anomalies

TOTAL UNIQUE ANOMALIES:           42 anomalies
```

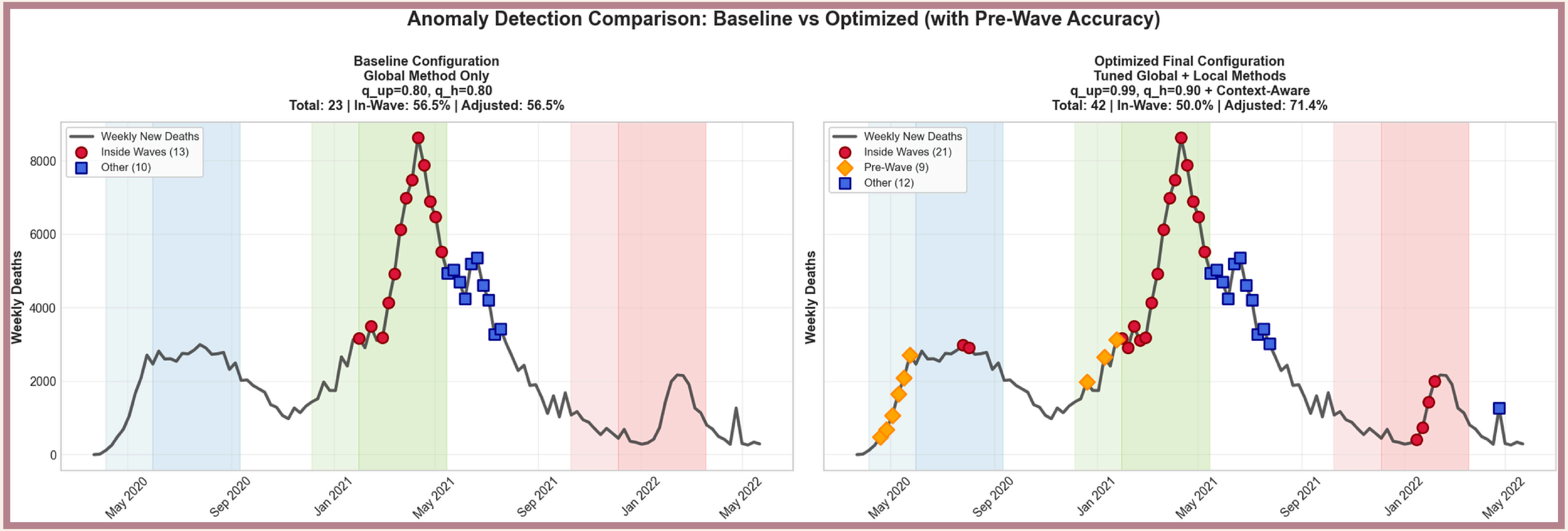
Predicting Future Waves

```
Wave 1:
  8 weeks before: 6 anomalies
  4 weeks before: 4 anomalies
  Closest anomaly: 2020-05-24 (7 days before wave start)
  Dates: [datetime.date(2020, 4, 19), datetime.date(2020, 4, 26), datetime.date(2020, 5, 3), datetime.date(2020, 5, 10),

Wave 2:
  8 weeks before: 3 anomalies
  4 weeks before: 2 anomalies
  Closest anomaly: 2021-01-24 (7 days before wave start)
  Dates: [datetime.date(2020, 12, 20), datetime.date(2021, 1, 10), datetime.date(2021, 1, 24)]

Wave 3:
  8 weeks before: 0 anomalies
  4 weeks before: 0 anomalies
  ⚠ No pre-wave anomalies detected
```

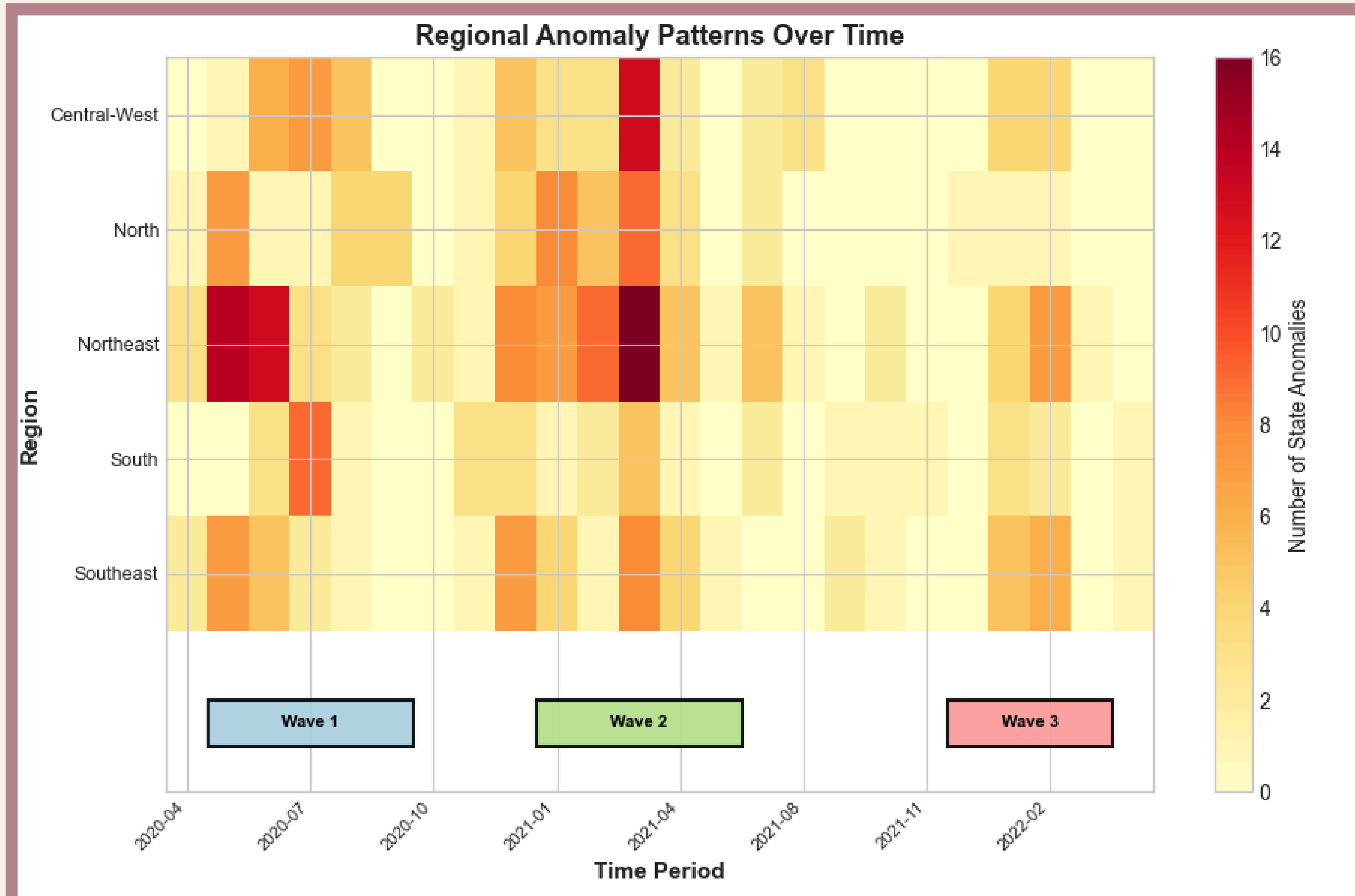
Predicting Future Waves

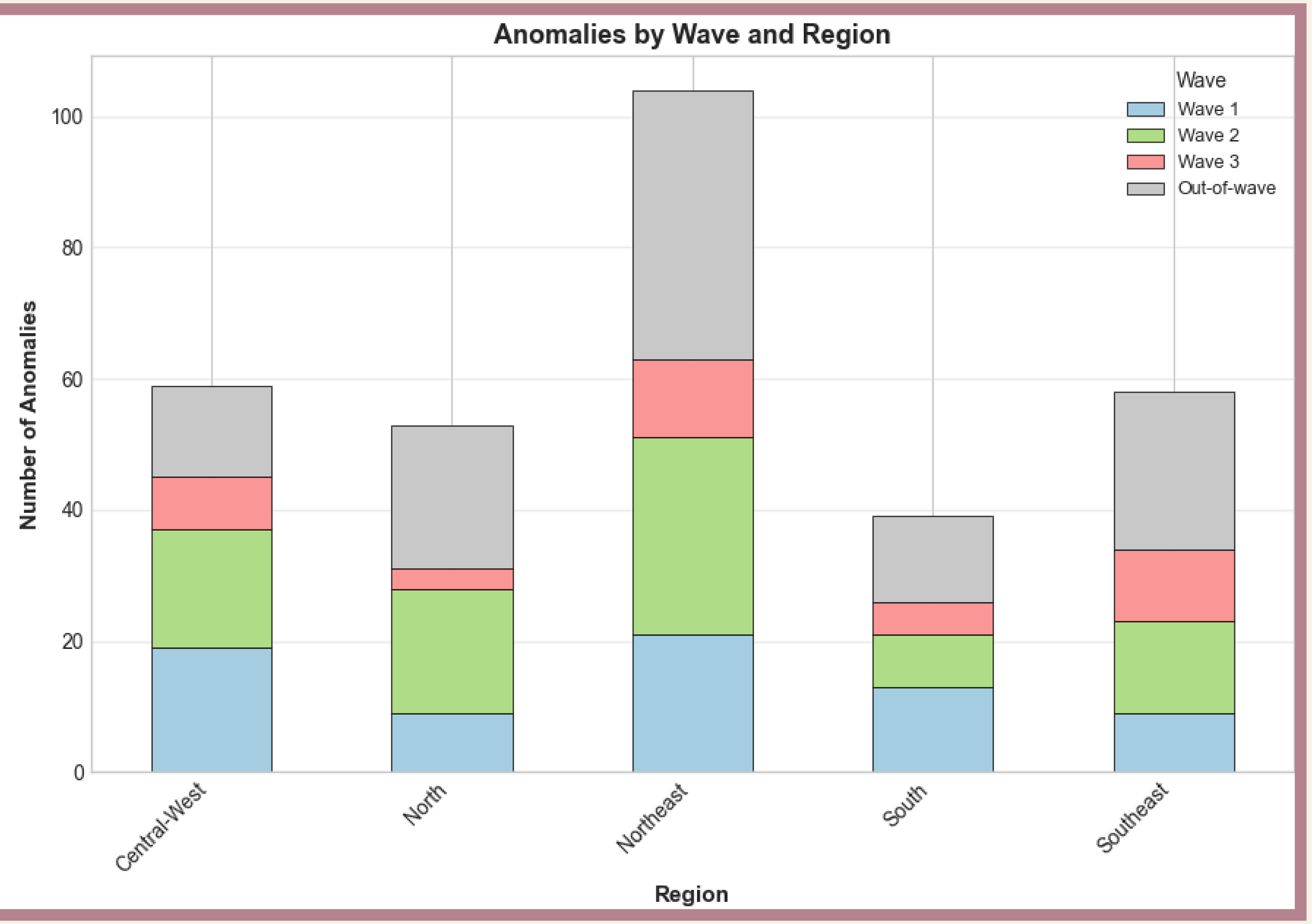


Final Comparison

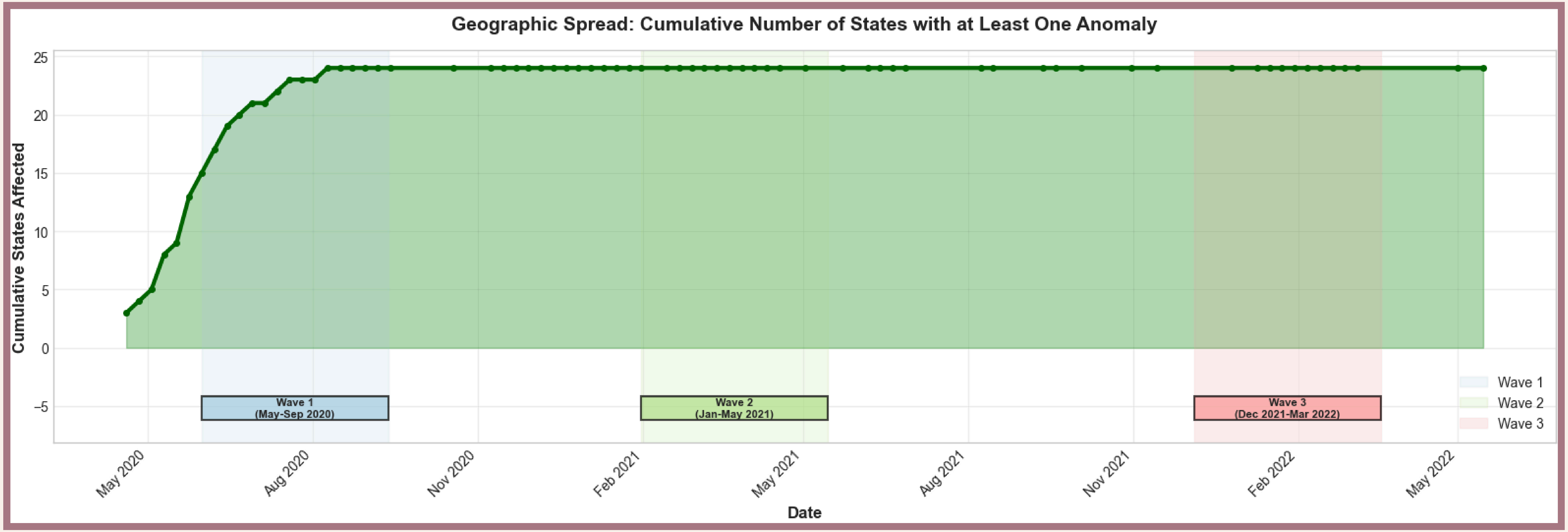


Regional Trends

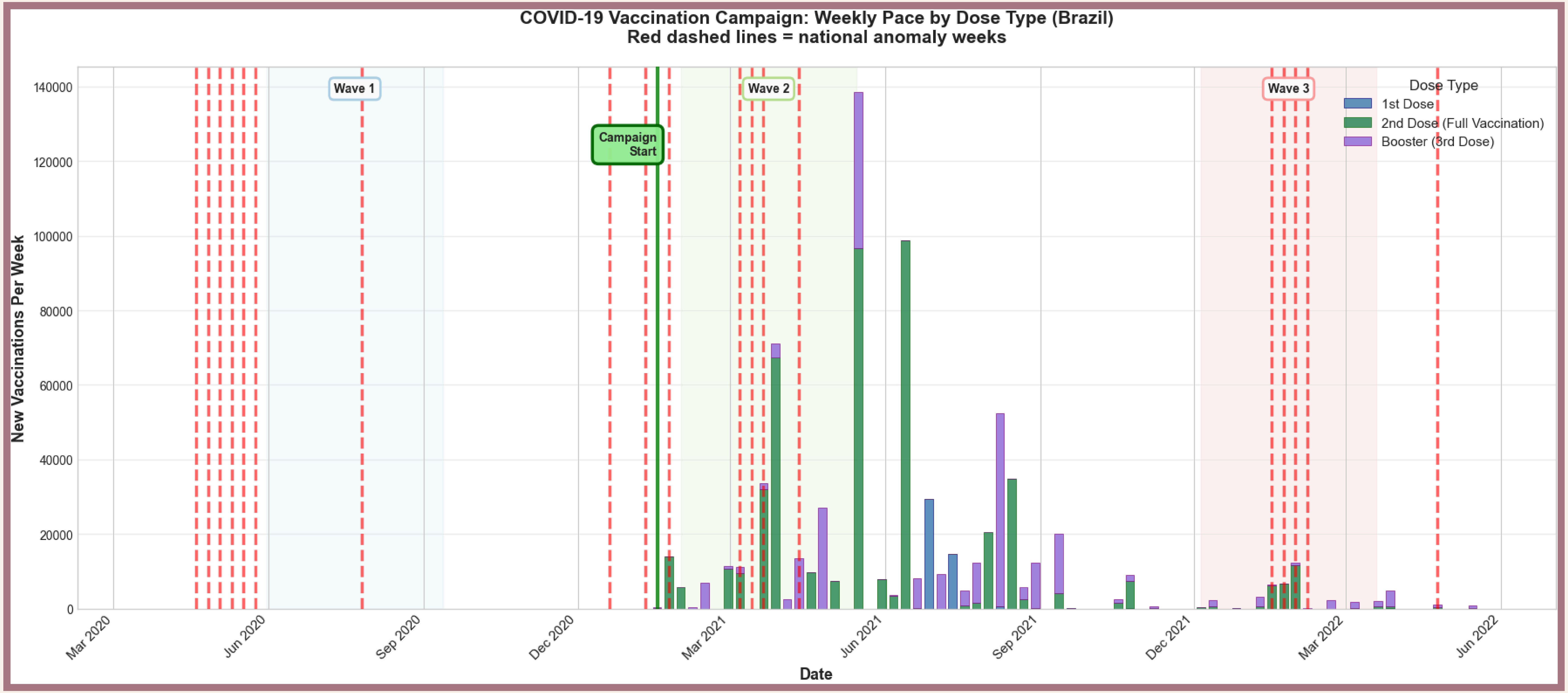




Regional Trends



Vaccination Trends



Analysis

- Weekly mortality patterns align with clearly defined COVID-19 waves in Brazil
- Baseline anomaly detection captured 23 anomalies, 56% of which were inside wave periods
- Hyperparameter tuning improved model alignment (29 anomalies and higher precision)
- Global and local anomaly detection methods were combined to reveal 42 unique anomaly weeks
- Pre-wave anomaly clusters appeared 4-8 weeks before Waves 1 and 2, highlighting potential for early warnings

Conclusion

- Tuned anomaly detection methods improves accuracy in identifying mortality spikes in wave periods
- Anomalies before major waves indicate value for early-warning systems
- Vaccination status and regional differences help explain anomaly patterns
- Future Work: train predictive models based on time-series findings, vaccination status, hospital information, and geographic information

Acknowledgements

- Principal Investigator: Dr. Tengfei Ma
- Mentor: Xin Wang

- Simons STEM Scholars Program

Thank You!

Any Questions?