Modelling a Pandemic: A Comparative Study

Zhong Ruoyu (C200201@e.ntu.edu.sg) & Li Xingjian (XLI072@e.ntu.edu.sg)

Course CZ1115 Lab Group FSP6 Team 1

School of Computer Science and Engineering,
Nanyang Technological University

April 2021

Photo by John Moore | Copyright (c) 2020 Getty Images

The Models

The SIRD Model

- Short for Susceptible-Infectious-Recovered-Deceased
- A compartmental model that simplifies mathematical modelling of infectious diseases
- Describes the virus transmission

Reference
Hasan, A. et al. (2020). A new estimation method for COVID-19 time-varying reproduction number using active cases

The ARIMA Model

- Short for Auto-Regressive Integrated Moving Average
- One of the easiest, general and effective machine learning algorithm for forecasting a time series
- Predicted value depends on recent values

References
Nau, R. (2020). Introduction to ARIMA: nonseasonal models
Khot, V. (2018). Get a glimpse of future using time series forecasting using Auto-ARIMA and Artificial Intelligence

Which model is better for predicting infectious cases during the developing stage of a pandemic, SIRD or ARIMA?

Modelling a Pandemic: A Comparative Study

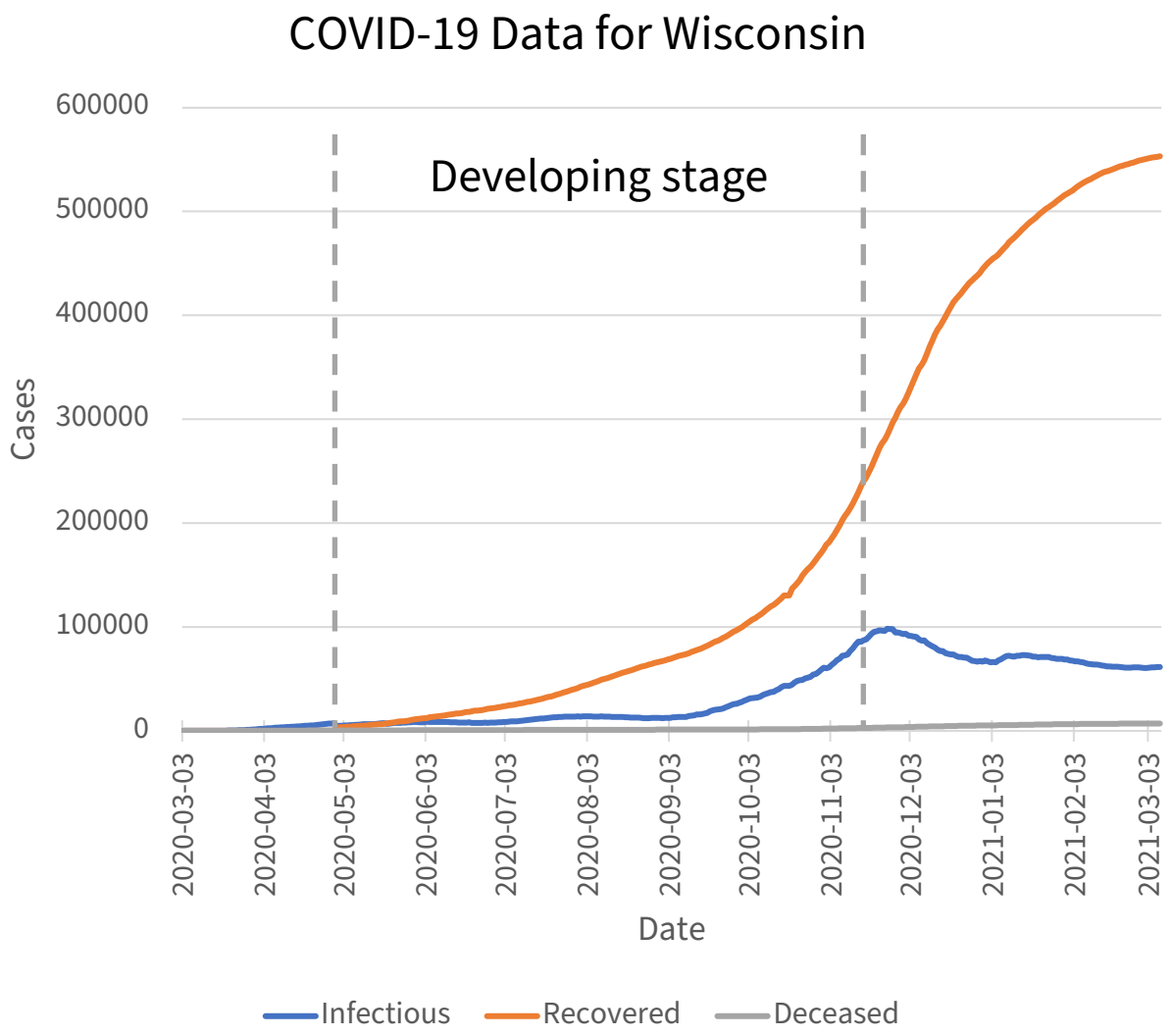
Wisconsin:
A Case Study

- Population density: 28th in US
- Inland state
  Less population movement

Reference
Statista Research Department. (2021). Population density in the U.S. by federal states including the District of Columbia in 2020

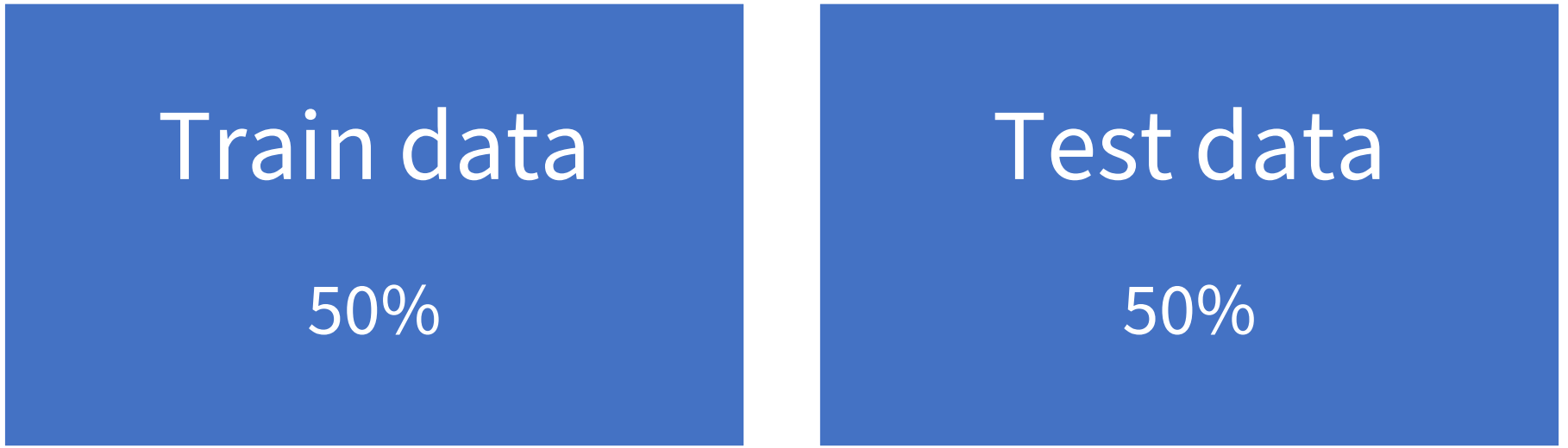
Data Preparation

- Data sources
  - (Major) COVID-19 statistics: The COVID Tracking Project
  - State population: The Census Bureau of the United States
  - State codes: Statistics Canada
- Obtain data from the abovementioned data sources
- Utilising pandas and NumPy to apply data cleaning and extraction
- With the help of the data definitions provided by the COVID Tracking Project API, extract the numbers of susceptible, infectious, recovered, and deceased cases for analysis



Exploratory Data Analysis – Wisconsin

Training the Models



Training the Models

The SIRD Model

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta IS}{N}, \\ \dfrac{dI}{dt} = \dfrac{\beta IS}{N} - \gamma I - \mu I, \\ \dfrac{dR}{dt} = \gamma I, \\ \dfrac{dD}{dt} = \mu I. \end{cases}$$

S	Susceptible
I	Infectious
R	Recovered
D	Deceased
N	Population
β	Rate of infection
γ	Rate of recovery
μ	Rate of mortality

Reference
Hasan, A. et al. (2020). A new estimation method for COVID-19 time-varying reproduction number using active cases

Training the Models
The SIRD Model

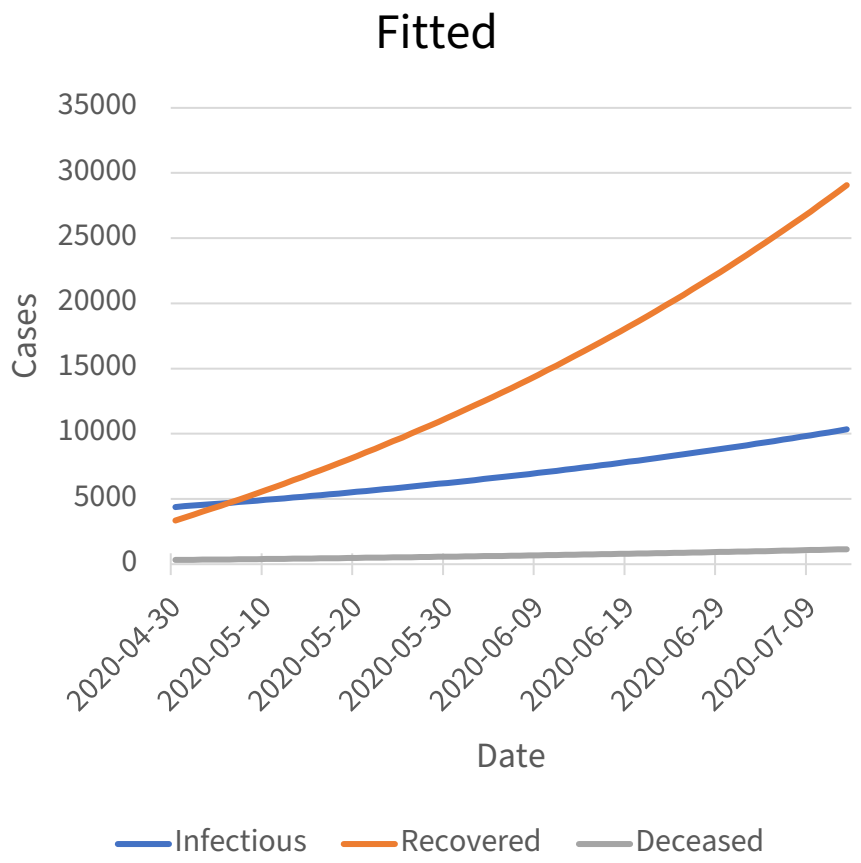
$$\begin{cases} \frac{dS}{dt} = -\frac{\beta IS}{N}, \\ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \mu I, \\ \frac{dR}{dt} = \gamma I, \\ \frac{dD}{dt} = \mu I. \end{cases}$$

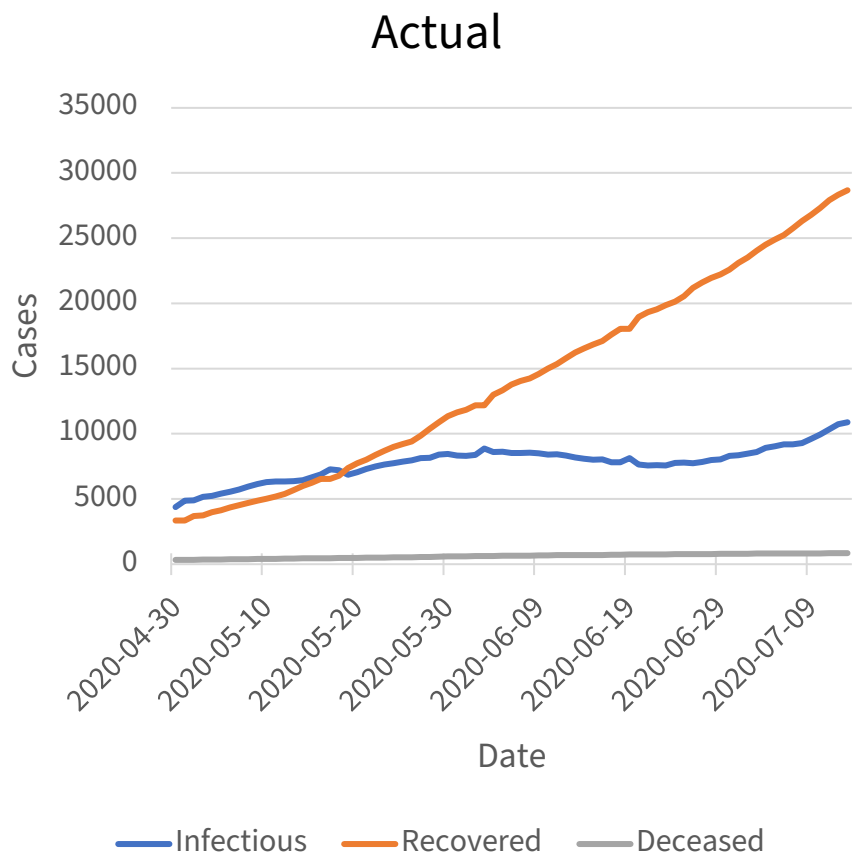
S	Susceptible
I	Infectious
R	Recovered
D	Deceased
N	Population
β	Rate of infection
γ	Rate of recovery
μ	Rate of mortality

- With the help of statistical analysis libraries …
- Integrate the system of ordinary differential equations (ODEs) to obtain the model
- Apply Powell's method on the root-mean-square error, the loss function
- Obtain minimised parameters (rates), and thus the fitted model

Training the Models

The SIRD Model – Wisconsin





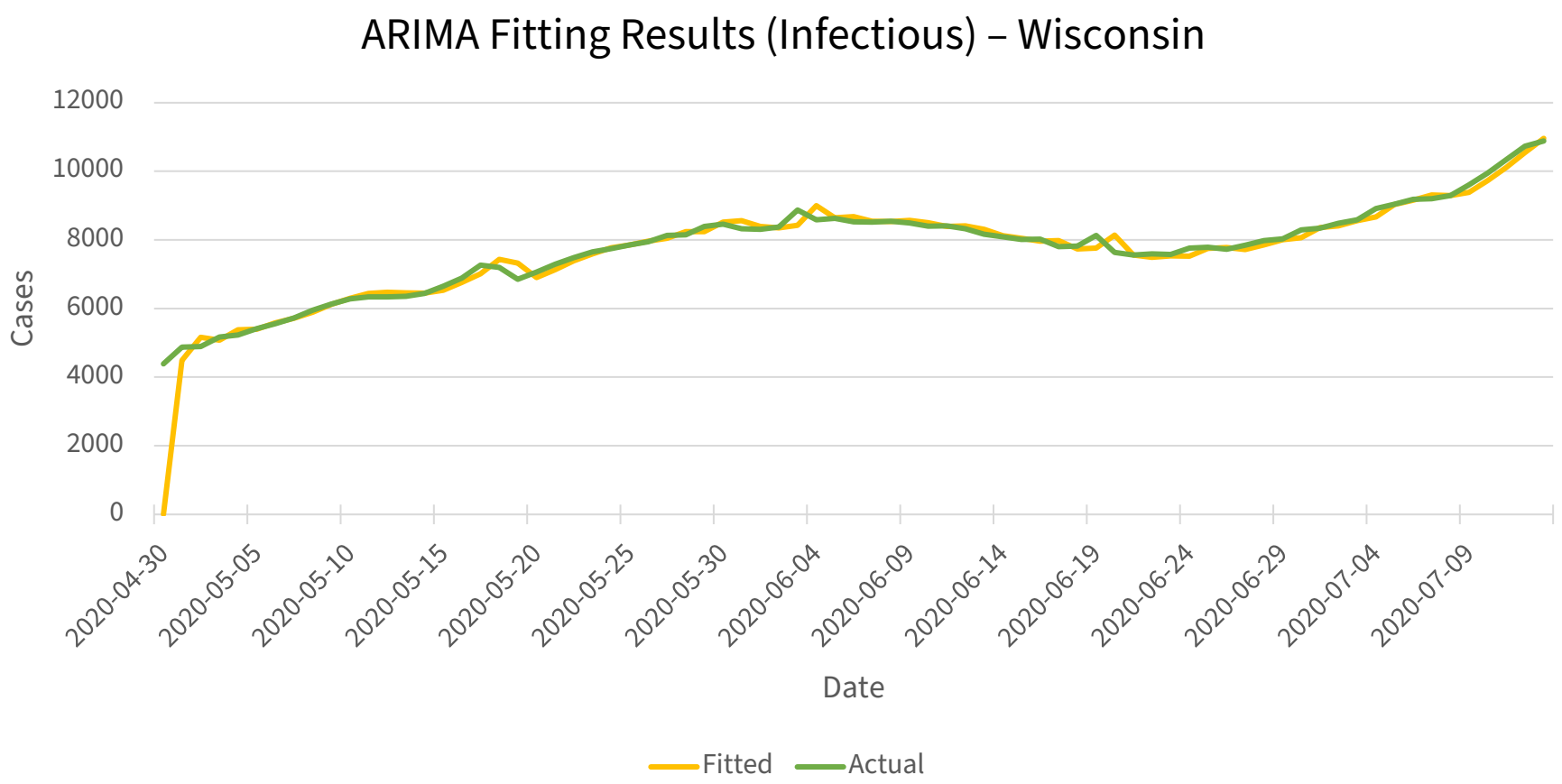
Training the Models

The ARIMA Model

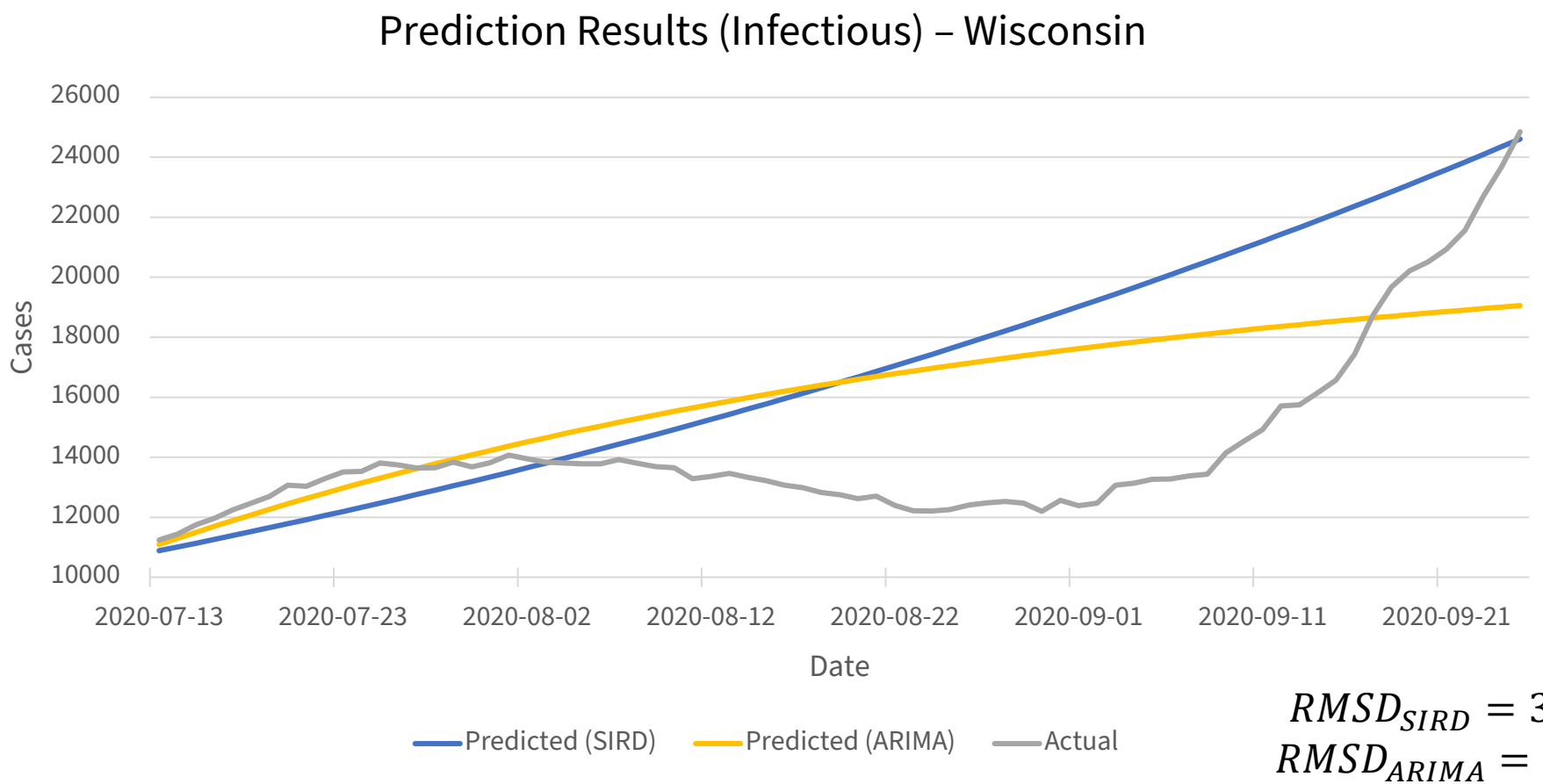
- Grid search for the best parameters
- Initialise a machine learning ARIMA model with the parameters found
- Train the model using the train data set, with the help of readily available machine learning libraries

Training the Models

The ARIMA Model – Wisconsin



Predicting Test Data – Wisconsin



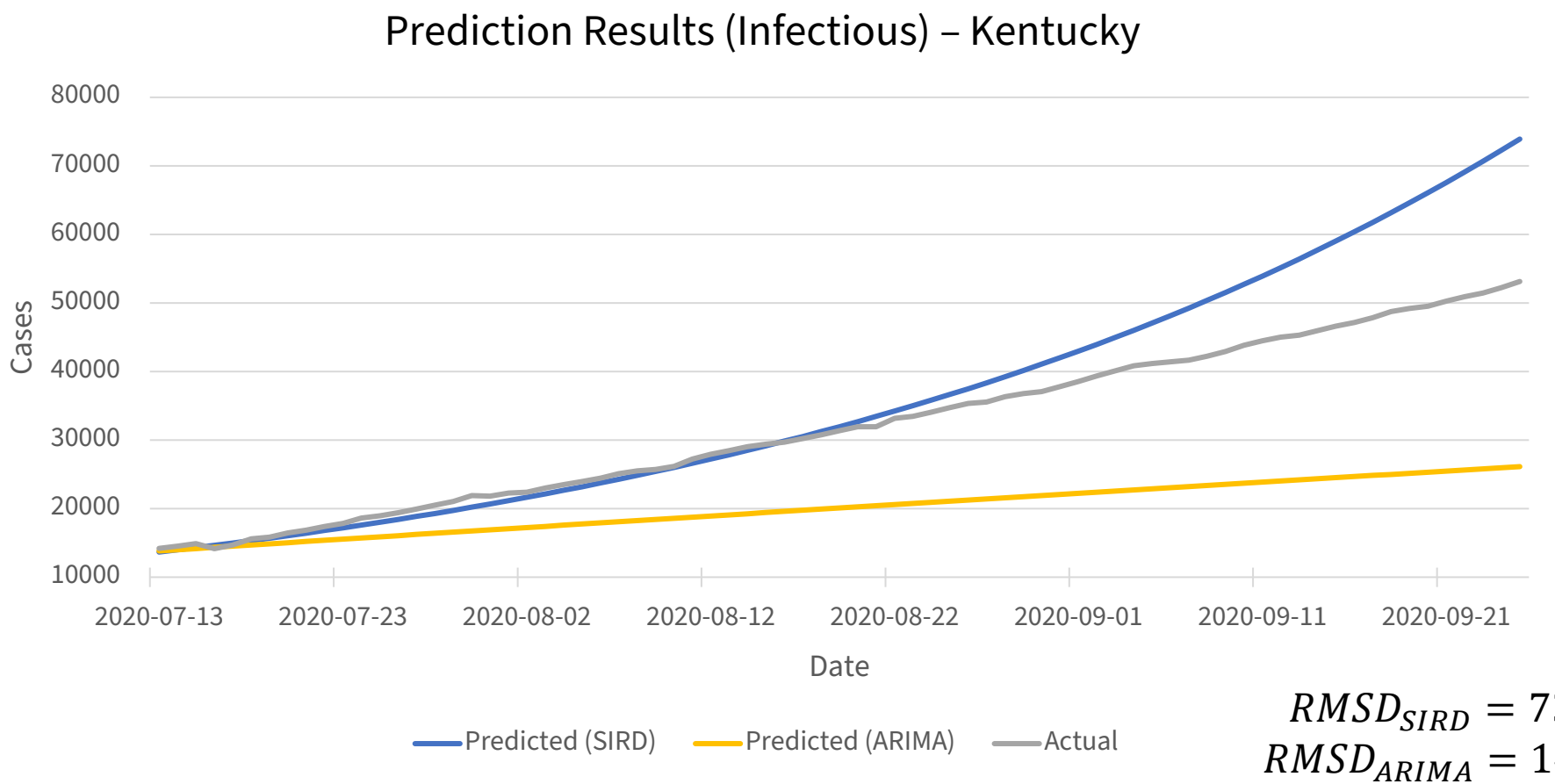
$$RMSD_{SIRD} = 3880.80$$
$$RMSD_{ARIMA} = 3053.45$$

Kentucky: Another Case Study

- Population density: 21st in US
- Inland state
  Less population movement
- Well-collected data

Reference
Statista Research Department. (2021). Population density in the U.S. by federal states including the District of Columbia in 2020

Predicting Test Data – Kentucky



$$RMSD_{SIRD} = 7372.63$$
$$RMSD_{ARIMA} = 14331.90$$

Conclusion

The SIRD Model

- Gives more reliable predictions for pandemic statistics during its developing stage
  - The virus transmission is well described with the mathematical model
- Gives more meaning to the data acquired, and hence is better for predicting infectious cases during the developing stage of a pandemic

The ARIMA Model

- Gives good short term predictions in certain cases
  - The machine learning method focuses more on historical values
- Not good at making long term predictions
  - Only takes the infectious cases into consideration
  - Treats the data as general numbers with no special meanings

References

- Hasan, A. et al.. 2020. *A new estimation method for COVID-19 time-varying reproduction number using active cases*. URL
- Khot, V.. 2018. *Get a glimpse of future using time series forecasting using Auto-ARIMA and Artificial Intelligence*. URL
- Nau, R.. 2020. *Introduction to ARIMA: nonseasonal models*. URL
- Statista Research Department. 2021. *Population density in the U.S. by federal states including the District of Columbia in 2020*. URL
- Statistics Canada. 2019. *List of U.S. States with Codes and Abbreviations*. URL
- The Census Bureau of the United States. 2021. *National Population Totals and Components of Change: 2010-2019*. URL
- The COVID Tracking Project. 2021. *Data API*. URL
- Vincent, T.. 2017. *A Guide to Time Series Forecasting with ARIMA in Python 3*. URL

Work Allocation

Zhong Ruoyu

- Technical details
  - Problem formulation
  - Data preparation
  - Model implementation
  - Data analysis
- Project presentation
  - Script Editing
  - Visualisation

Li Xingjian

- Overall idea & structure
  - Idea generation
  - Model selection
  - Information presentation
  - Conclusion
- Project presentation
  - Analysis
  - Scripts Drafting

Thank you

Modelling a Pandemic: A Comparative Study

Zhong Ruoyu (C200201@e.ntu.edu.sg)
Li Xingjian (XLI072@e.ntu.edu.sg)