Surveillance Advances

Data for action: The role of data science in public health surveillance

November 28, 2023 12:00 – 1:00pm (CT) / 1:00 – 2:00pm (ET)

Speakers

Dr. Jude Kong — Assistant Professor, York University
Executive Director, ACADIC, AI4PEP

Dr. Nathaniel Osgood — Professor, University of Saskatchewan
Director, Computational Epidemiology & Public Health Informatics Laboratory

National Collaborating Centre for Infectious Diseases
Centre de collaboration nationale des maladies infectieuses

Public Health Agency of Canada
Agence de la santé publique du Canada

Land Acknowledgment: NCCID

The National Collaborating Centre for Infectious Diseases is hosted by the University of Manitoba, on the original lands of Anishinaabe, Cree, Oji-Cree, Dakota and Dene peoples, and on the homeland of the Métis Nation.

At NCCID, we strive to honor the lands and their original caretakers in our work. We acknowledge that we are on Treaty One land. We recognize that this and other treaties, have been implemented as part of the process of colonization intended to benefit some while harming others. We are committed to working with our partners towards reconciliation.

Housekeeping

Seminar recording and presentation slides will be available shortly after the seminar at the NCCID website: https://nccid.ca/.

If you have technical problems with Zoom, please email us at nccid@umanitoba.ca.

The chat box for participants has been disabled for this session.
We will use the chat box to share additional information.

Please use the Q&A tab to submit your questions for our speakers.

You can "like" other people's questions to push them up in priority.

Image by macrovector on Freepik

Accreditation

Surveillance Advances is a self-approved group learning activity (Section 1) as defined by the Maintenance of Certification Program of the **Royal College of Physicians and Surgeons of Canada**.

The seminar series is also approved by the Council of Professional Experience for professional development hours for members of the **Canadian Institute of Public Health Inspectors**.

If you would like a letter of participation, please complete the survey which will be shared after the seminar.

Image by freepik on Freepik

Land Acknowledgment: PHAC

I would like to take this time to acknowledge the land that I live and work on is the traditional territory of the Wendat, the Anishnaabeg, Haudenosaunee, Métis, and the Mississaugas of the Credit First Nation.

It is home to many First Nations, Métis, and Inuit peoples. I am grateful for the opportunity to share their home.

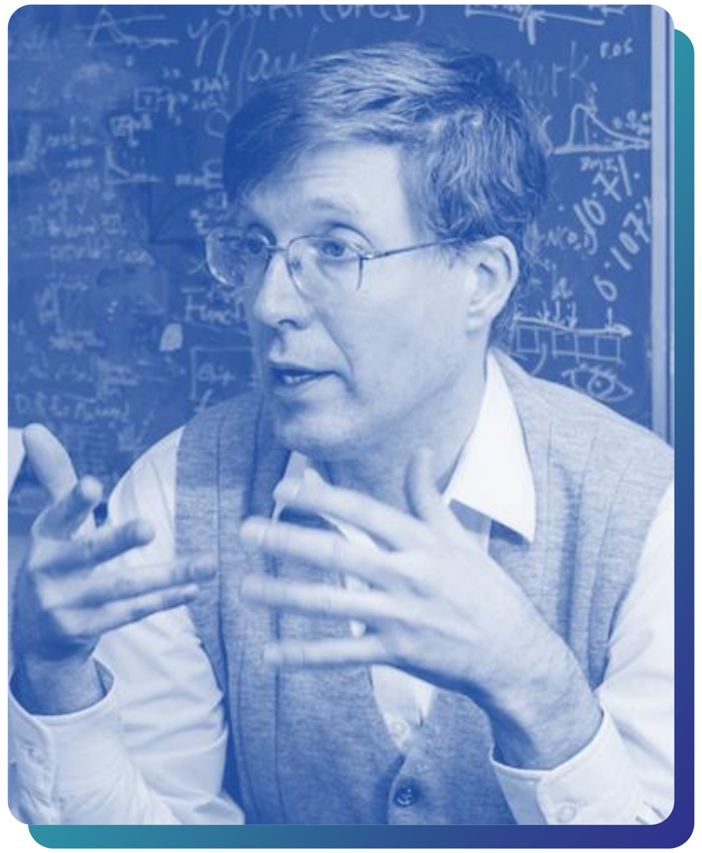
Today’s Speakers

Dr. Jude Kong Phd

Executive Director, Africa-Canada Artificial Intelligence & Data Innovation Consortium (ACADIC), Resilience Research Atlantic Alliance on Sustainability, Supporting Recovery and Renewal (AI4PEP)

Assistant Professor, York University

jdkong@yorku.ca



Dr. Nathaniel Osgood Phd

Professor, University of Saskatchewan

Director, Computational Epidemiology & Public Health Informatics Laboratory

nathaniel.osgood@usask.ca

Learning Objectives

By the end of this seminar, you will be able to:

- Understand how AI, data science, and mathematical solutions can be used responsibly to improve public health surveillance and response to emerging and re-emerging infectious disease outbreaks
- Understand the feasibility and benefits of using AI to enable real-time updates of transmission modeling with diverse incoming data streams
- Highlight the accuracy benefits secured by incorporating wastewater data sources among the real datasets used to ground models
- Note the likely strong opportunities from jointly updating models of multiple pathogens sharing common risk factors with cross-pathogen surveillance data

Image by freepik on Freepik

AI-Epidemix: A disease outbreak detection and response tool supported by AI and a multi-source real-time data collection platform

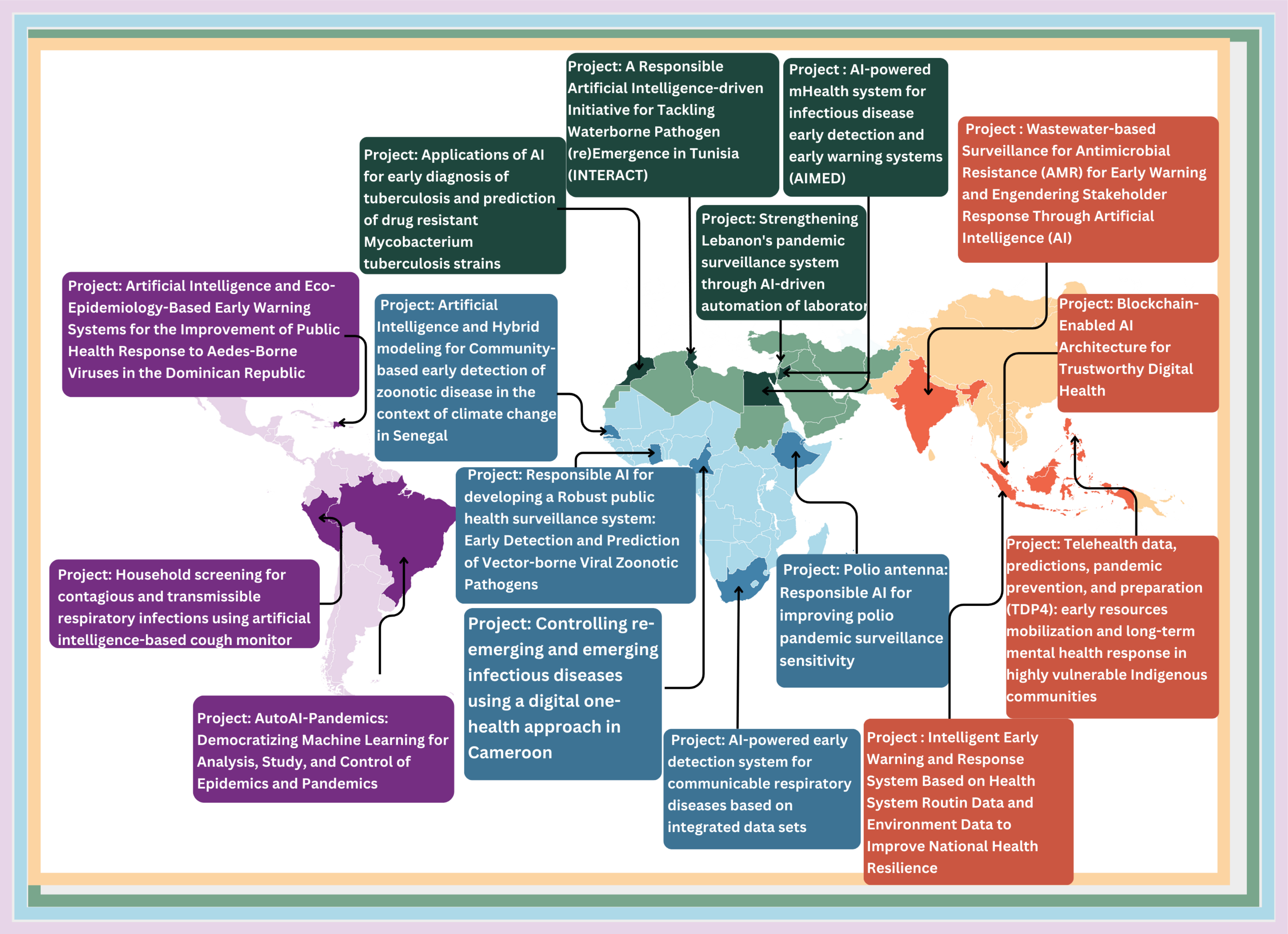
Dr. Jude Kong PhD

Executive Director, Africa-Canada Artificial Intelligence & Data Innovation Consortium

Executive Director, Global South AI for Pandemic and Epidemic Preparedness and Response Network (AI4EP)

Assistant Professor, Department of Mathematics & Statistics, York University

Twitter: @dzevela


Project: A Responsible Artificial Intelligence-driven Initiative for Tackling Waterborne Pathogen (re)Emergence in Tunisia (INTERACT)
Project : AI-powered mHealth system for infectious disease early detection and early warning systems (AIMED)
Project : Wastewater-based Surveillance for Antimicrobial Resistance (AMR) for Early Warning and Engendering Stakeholder Response Through Artificial Intelligence (AI)
Project: Applications of AI for early diagnosis of tuberculosis and prediction of drug resistant Mycobacterium tuberculosis strains
Project: Strengthening Lebanon's pandemic surveillance system through AI-driven automation of laborator
Project: Artificial Intelligence and Eco-Epidemiology-Based Early Warning Systems for the Improvement of Public Health Response to Aedes-Borne Viruses in the Dominican Republic
Project: Artificial Intelligence and Hybrid modeling for Community-based early detection of zoonotic disease in the context of climate change in Senegal
Project: Blockchain-Enabled AI Architecture for Trustworthy Digital Health
Project: Responsible AI for developing a Robust public health surveillance system: Early Detection and Prediction of Vector-borne Viral Zoonotic Pathogens
Project: Telehealth data, predictions, pandemic prevention, and preparation (TDP4): early resources mobilization and long-term mental health response in highly vulnerable Indigenous communities
Project: Household screening for contagious and transmissible respiratory infections using artificial intelligence-based cough monitor
Project: Polio antenna: Responsible AI for improving polio pandemic surveillance sensitivity
Project: Controlling re-emerging and emerging infectious diseases using a digital one-health approach in Cameroon
Project: AutoAI-Pandemics: Democratizing Machine Learning for Analysis, Study, and Control of Epidemics and Pandemics
Project: AI-powered early detection system for communicable respiratory diseases based on integrated data sets
Project : Intelligent Early Warning and Response System Based on Health System Routin Data and Environment Data to Improve National Health Resilience

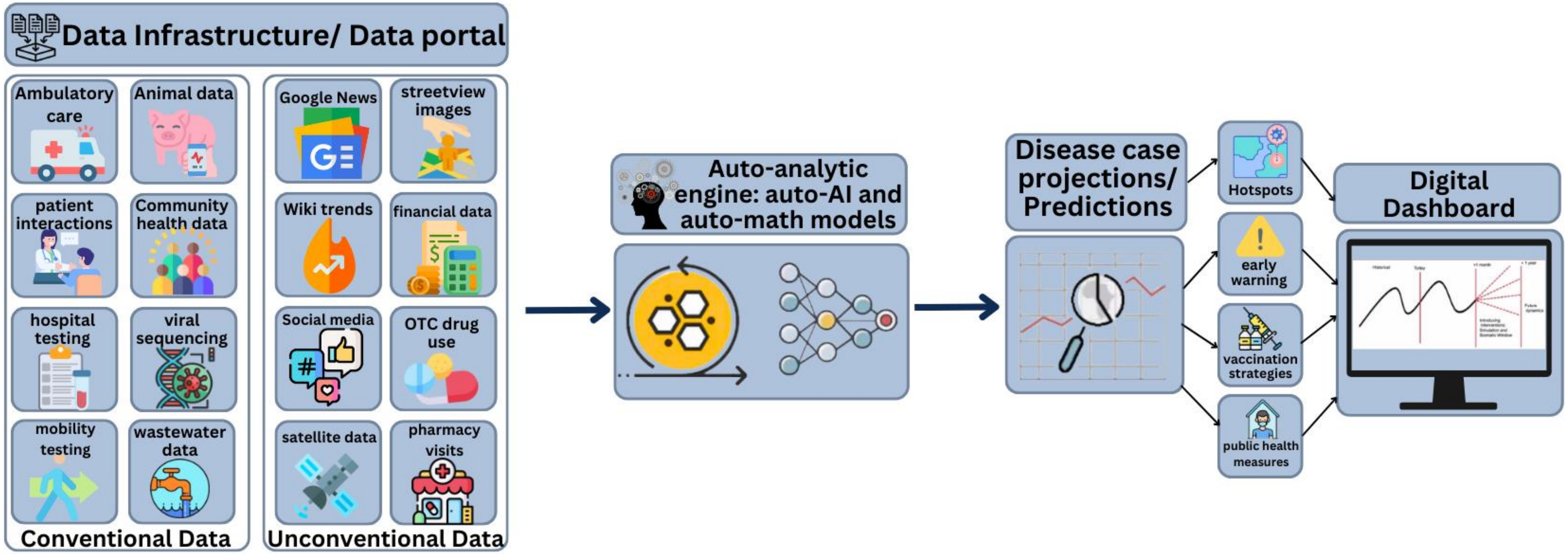
IDRC · CRDI

International Development Research Centre

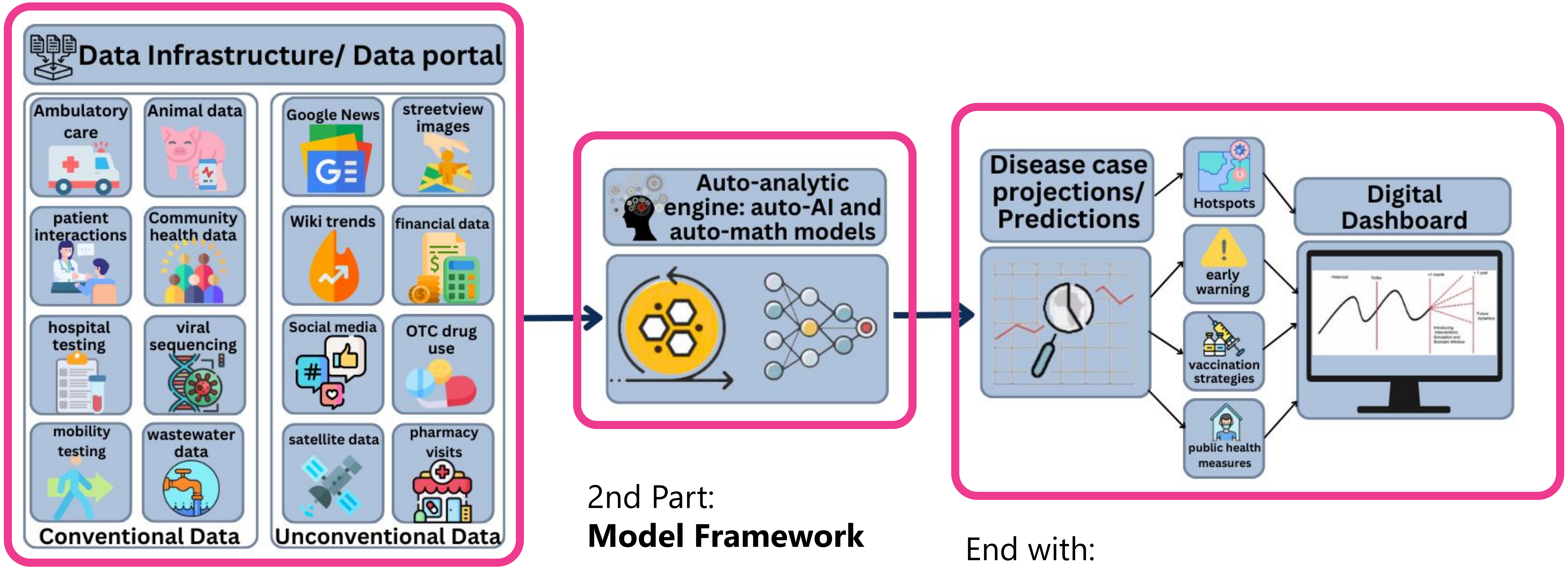
Centre de recherches pour le développement international

Canada

AutoAI Epidemix



AutoAI Epidemix

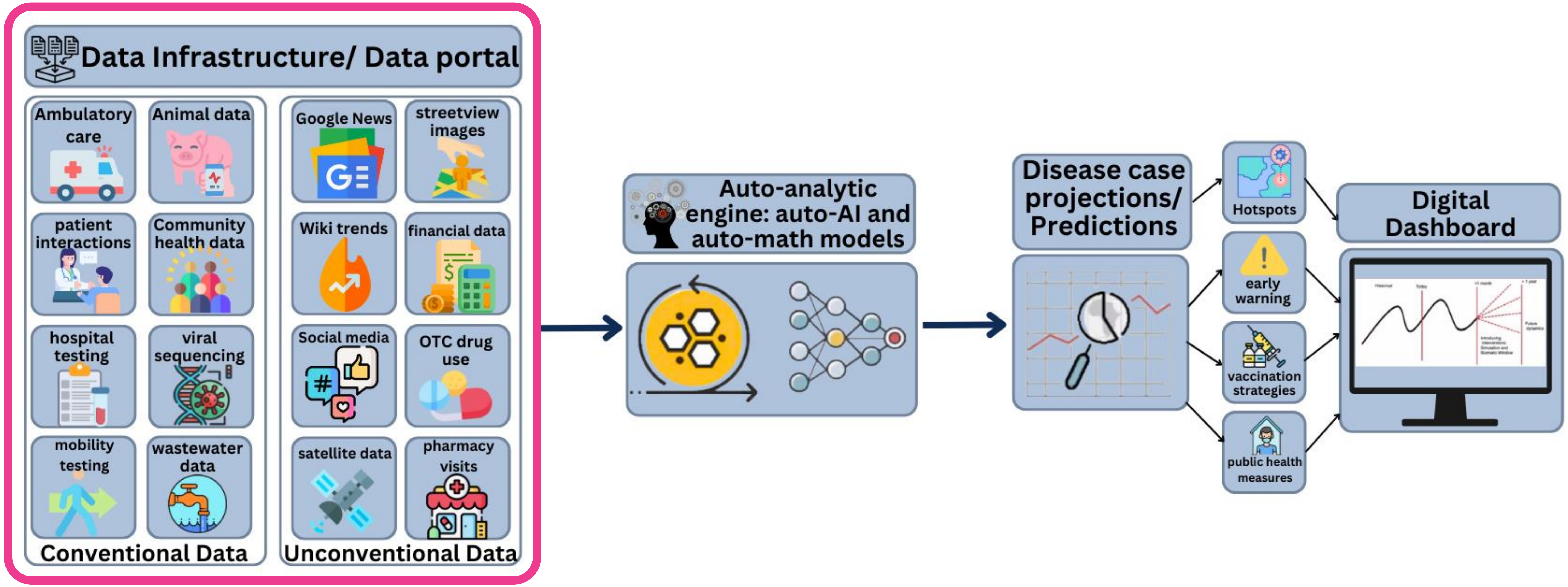


1st Part of the Talk:
Data Sources and Their Correlation with Diseases

2nd Part:
Model Framework

End with:
Examples (Influenza, Lyme, Covid)

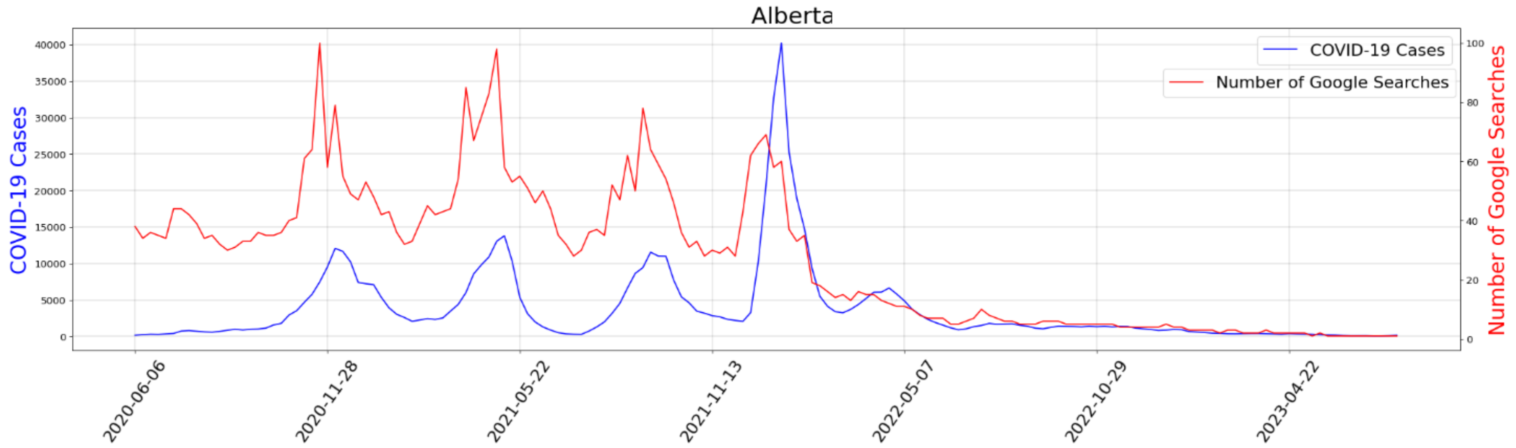
AutoAI Epidemix

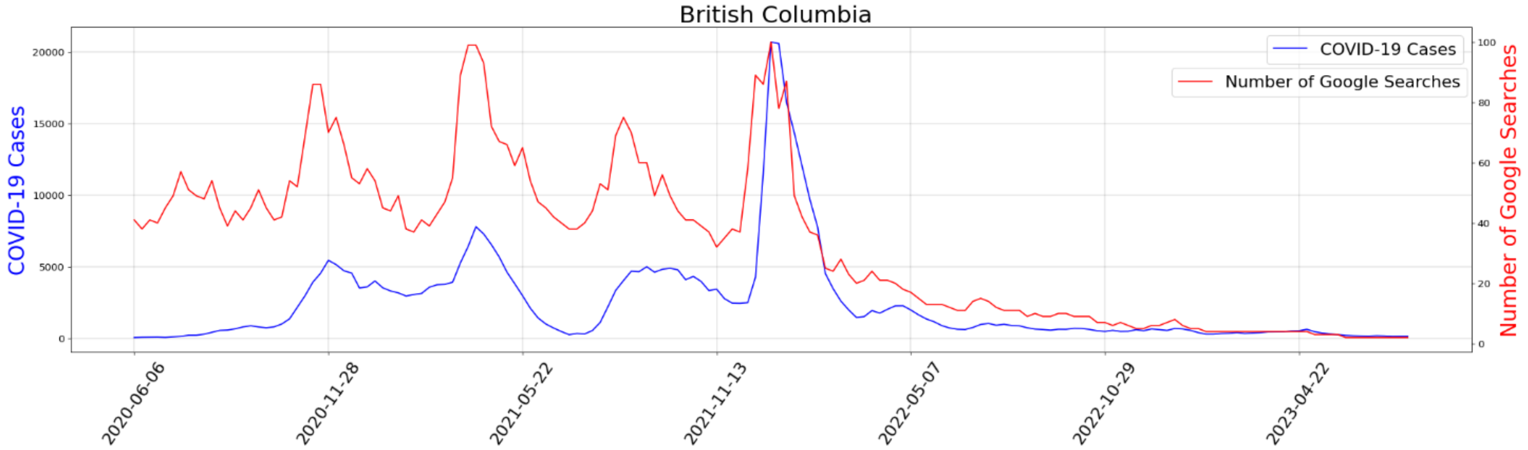


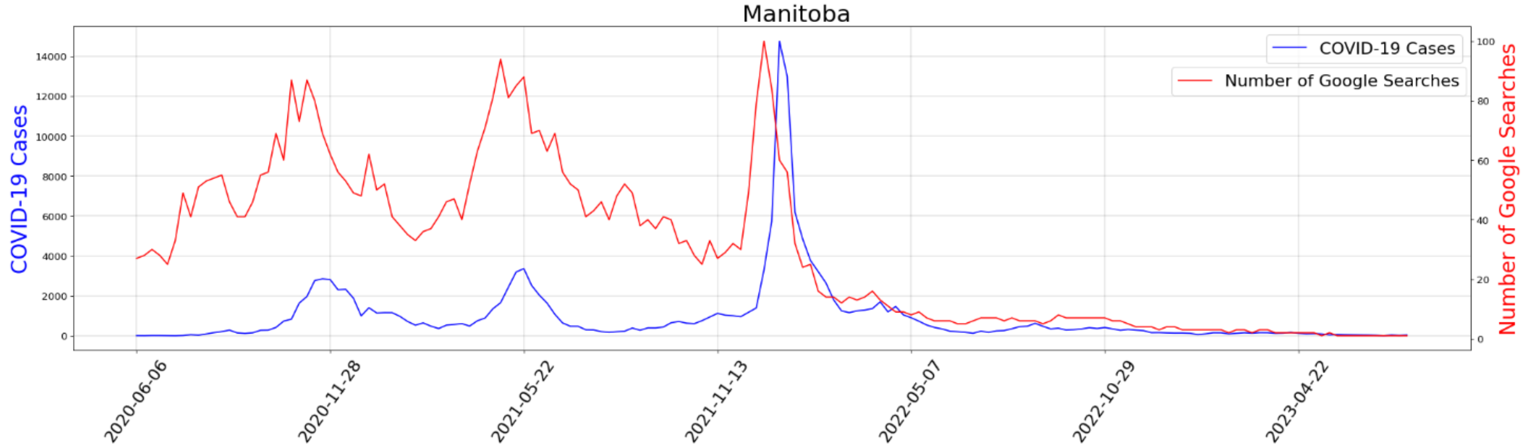
1st Part of the Talk:
Data Sources and Their Correlation with Diseases

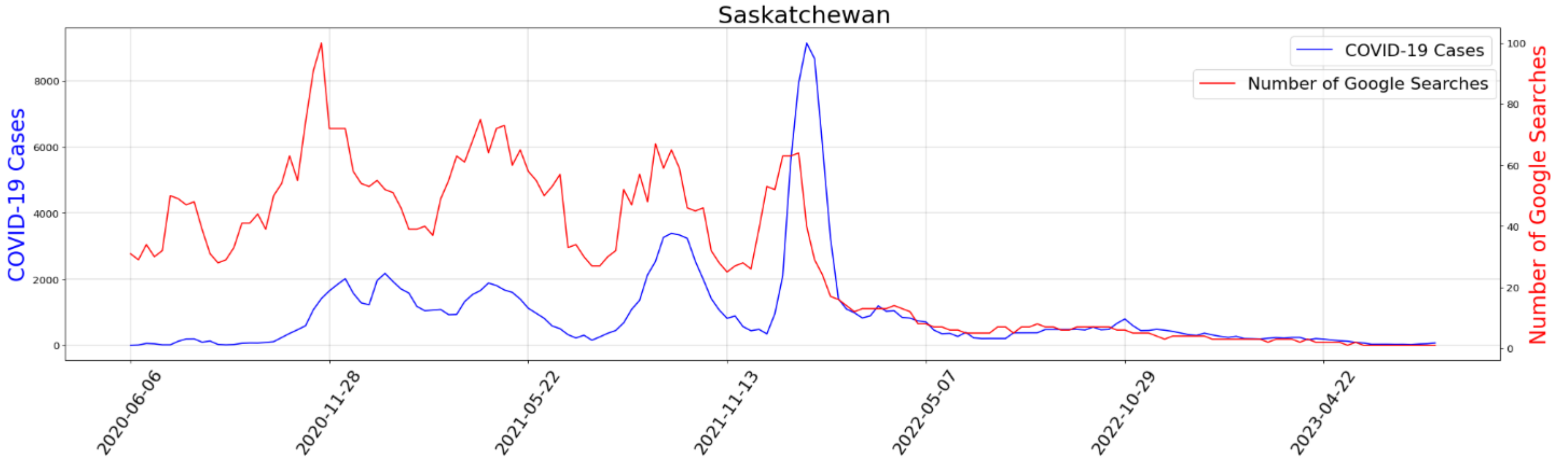
Google Trends

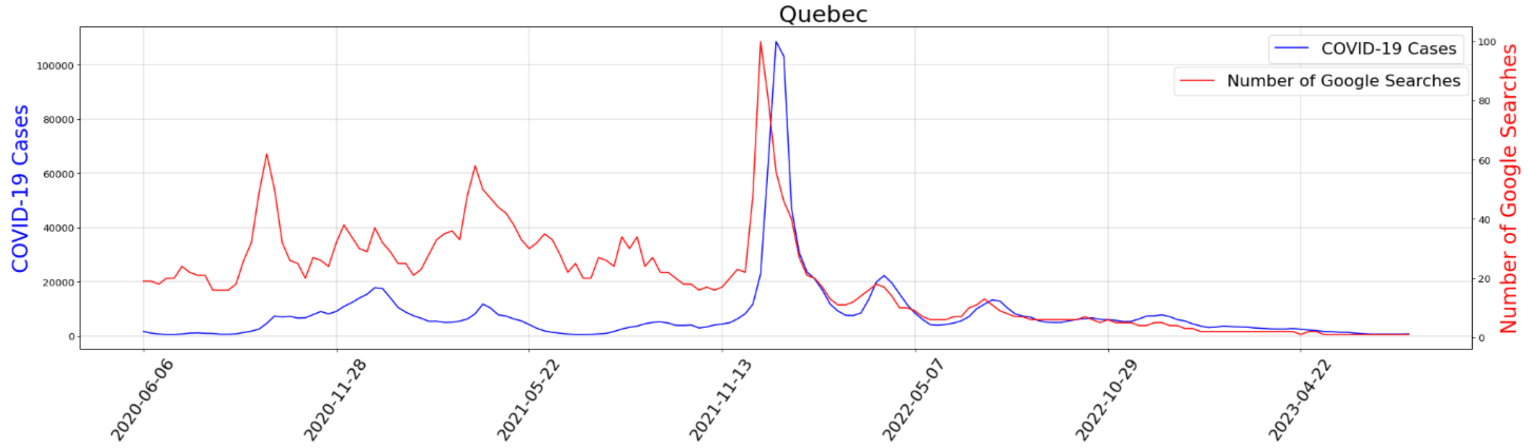
- A great source of data is the number of times people from a particular province/community have searched Google for a specific disease topic.
- Google trends in a particular region is strongly correlated with the number of disease cases in that area, in most cases.
- In most cases, Google trends even peak earlier than the actual disease cases.





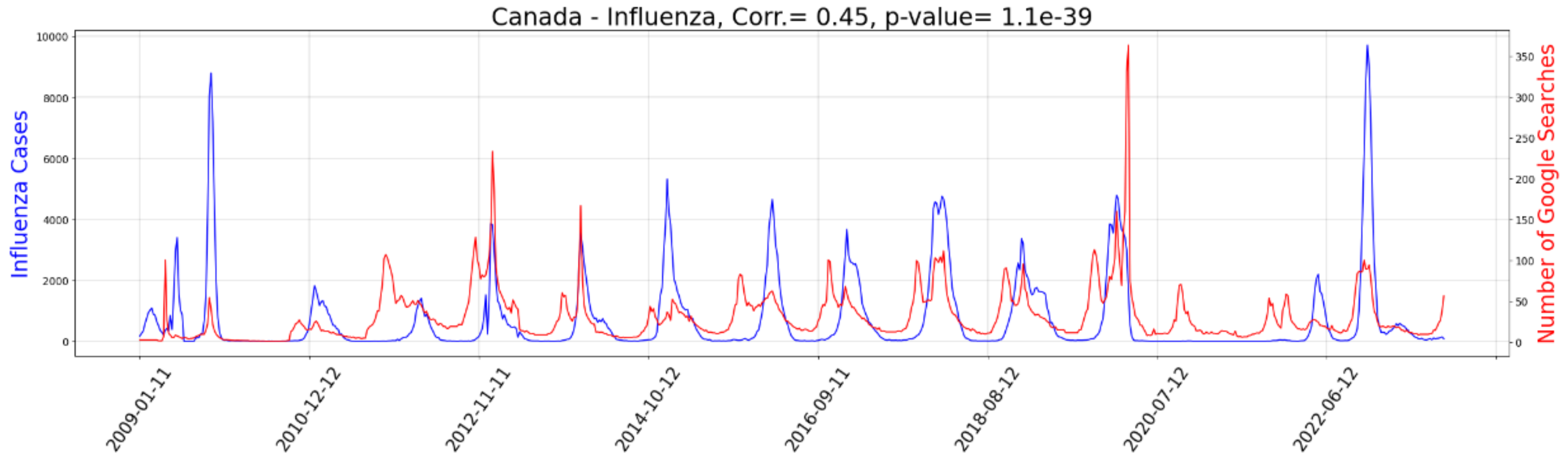


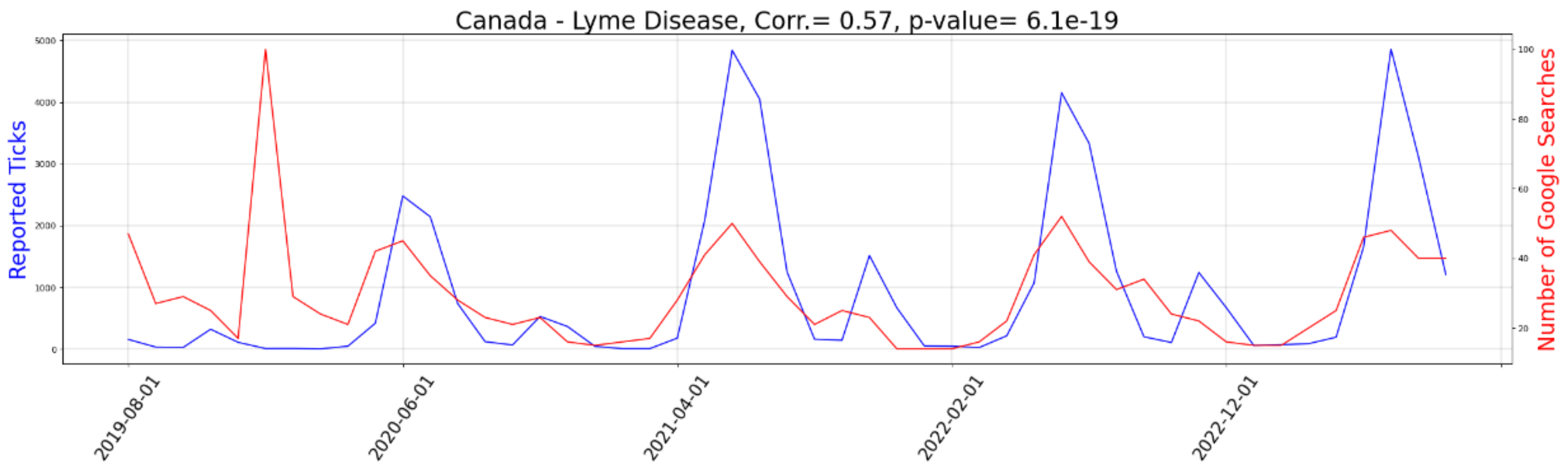


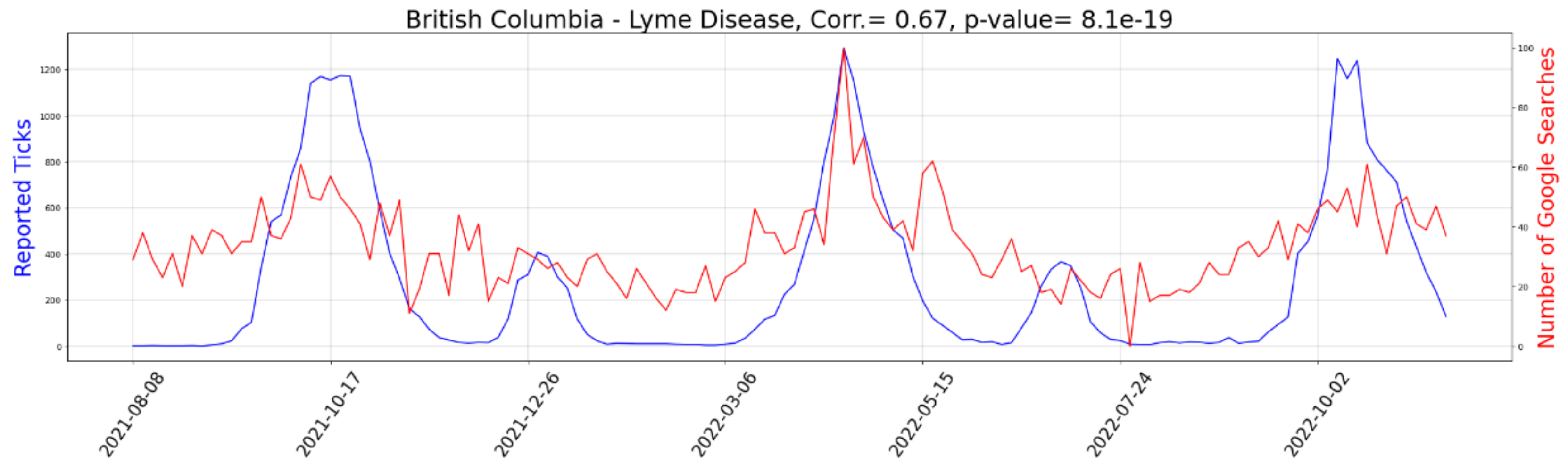


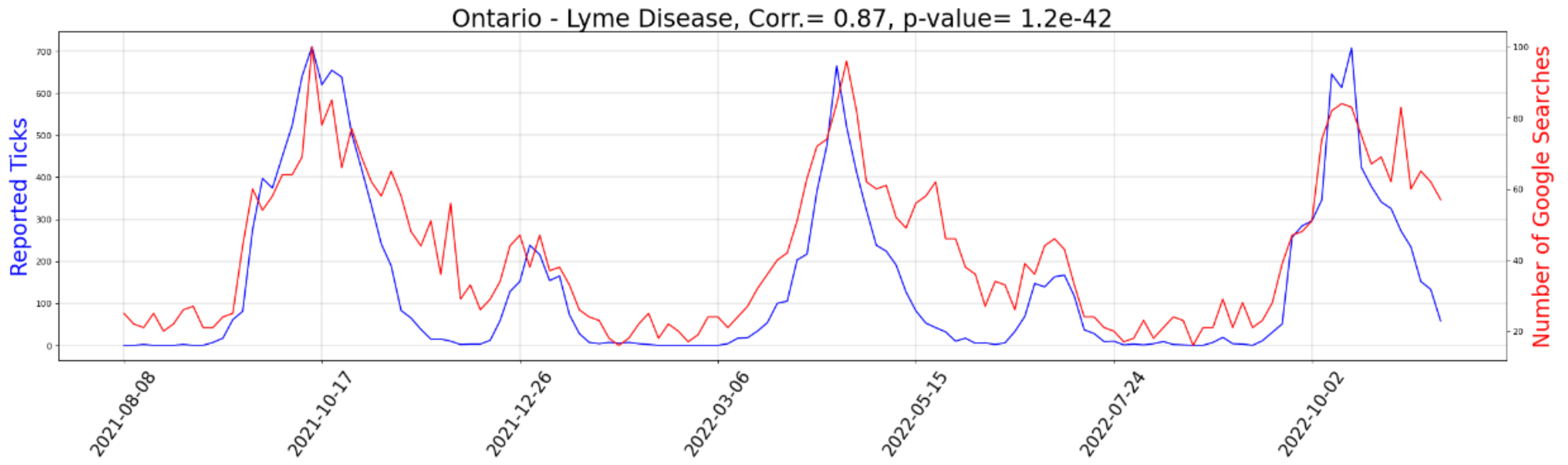
Google Trends

- Number of Google searches on a particular disease topic is almost always strongly correlated with the number of cases.







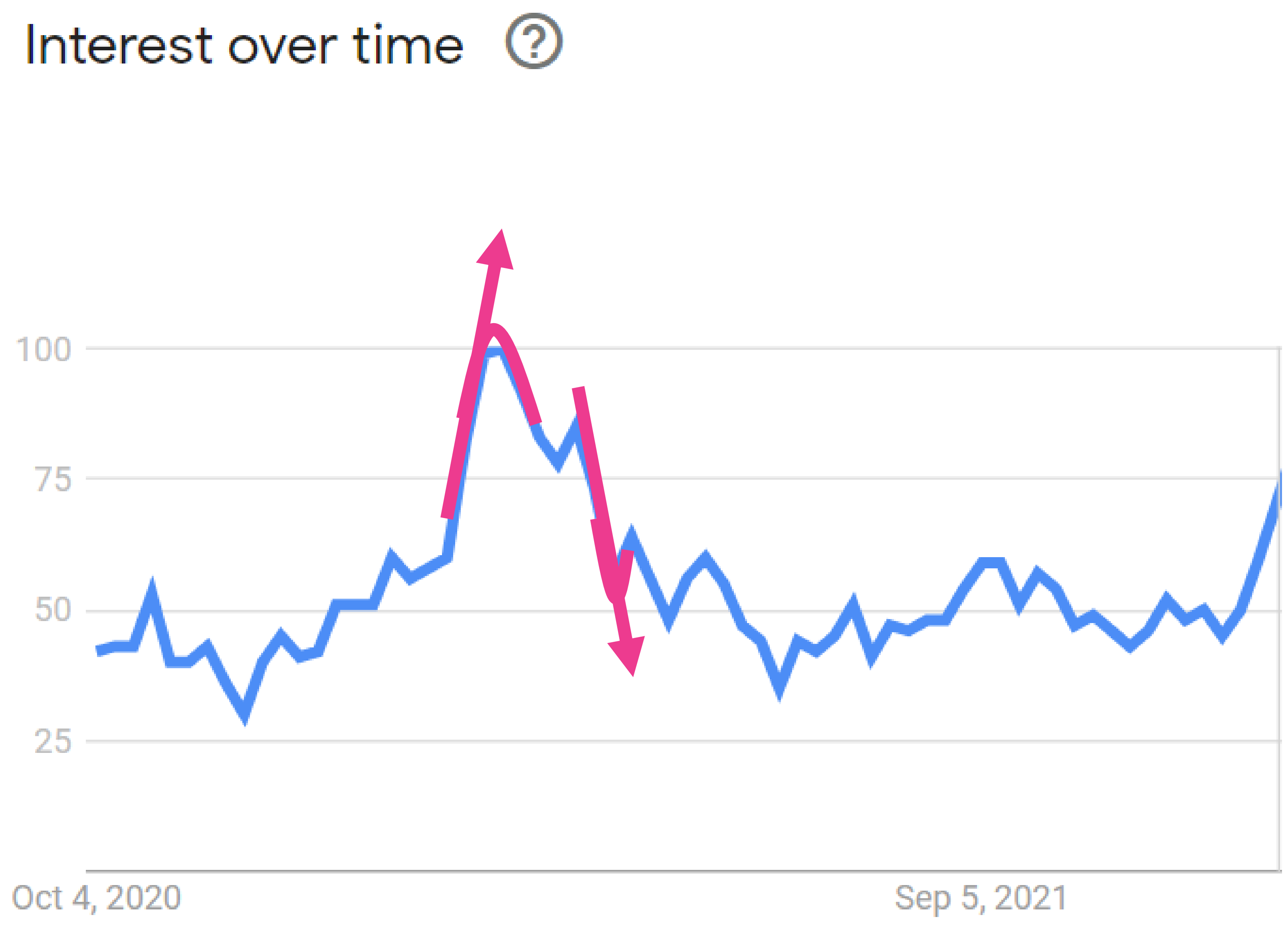


Google Trends

- In addition to the number of searches, the rate of its increase or decrease is another indicator of the number of cases rising or falling.
- Moreover, the second order rate, or the concavity and convexity of the time series is a great indicator of the number of cases increasing or decreasing.

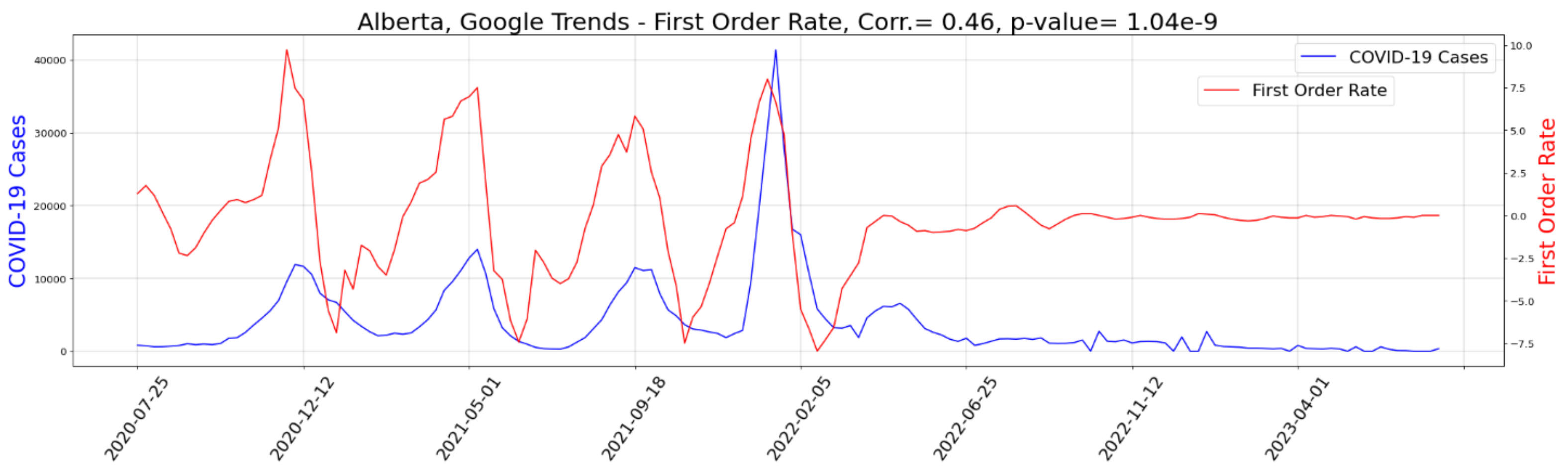
Google Trends

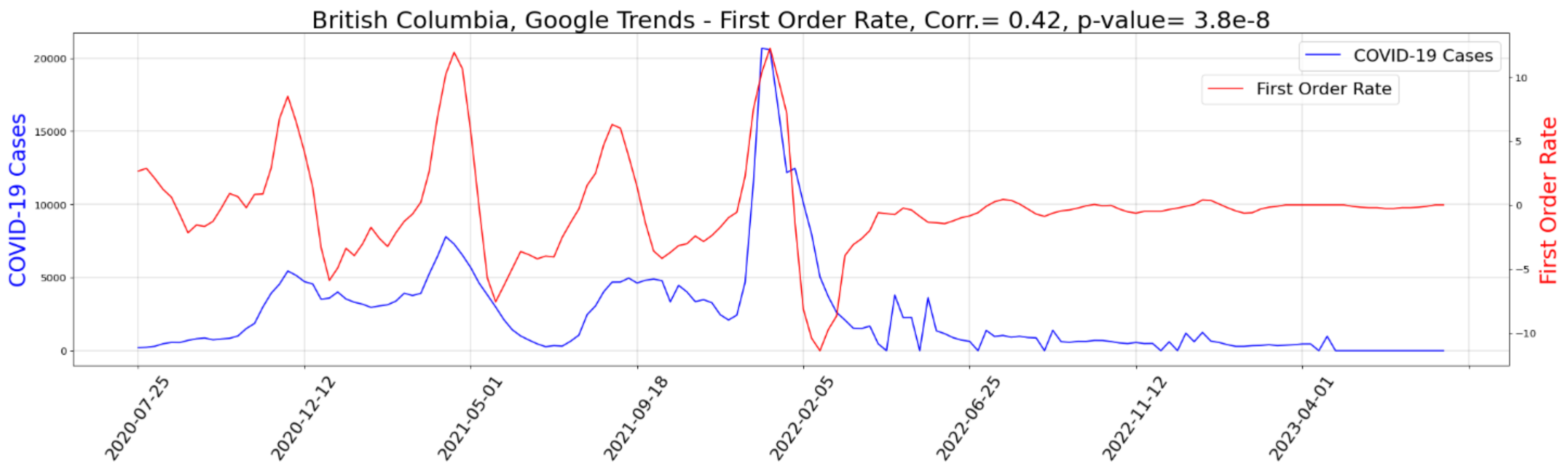
- In addition to the number of searches, the rate of its increase or decrease is another indicator of the number of cases rising or falling.
- Moreover, the second order rate, or the concavity and convexity of the time series is a great indicator of the number of cases increasing or decreasing.

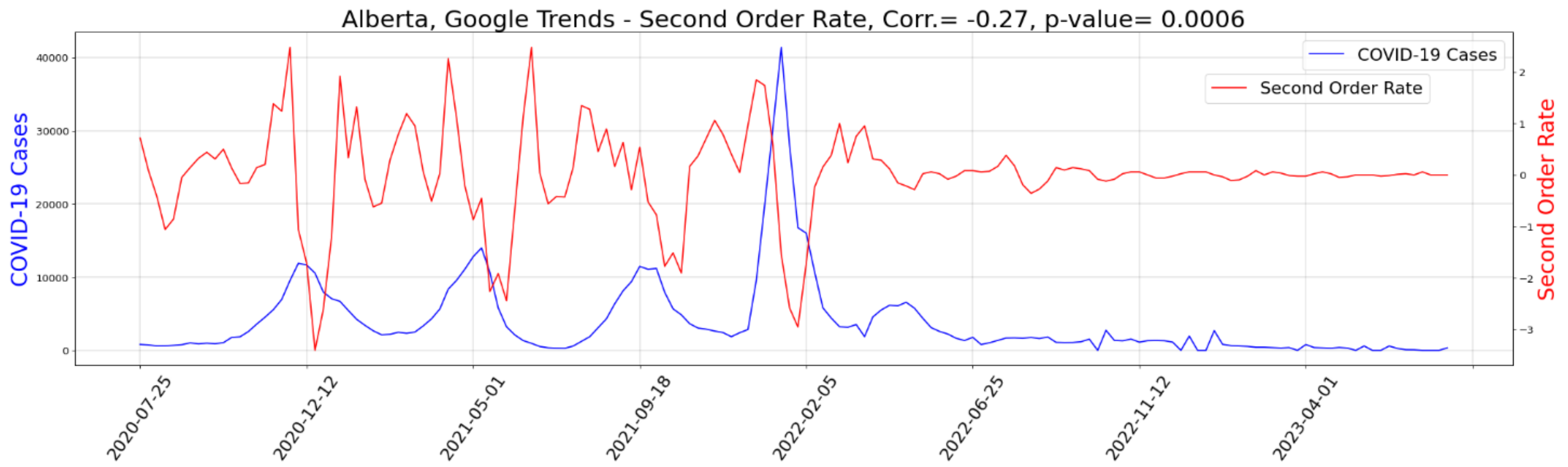


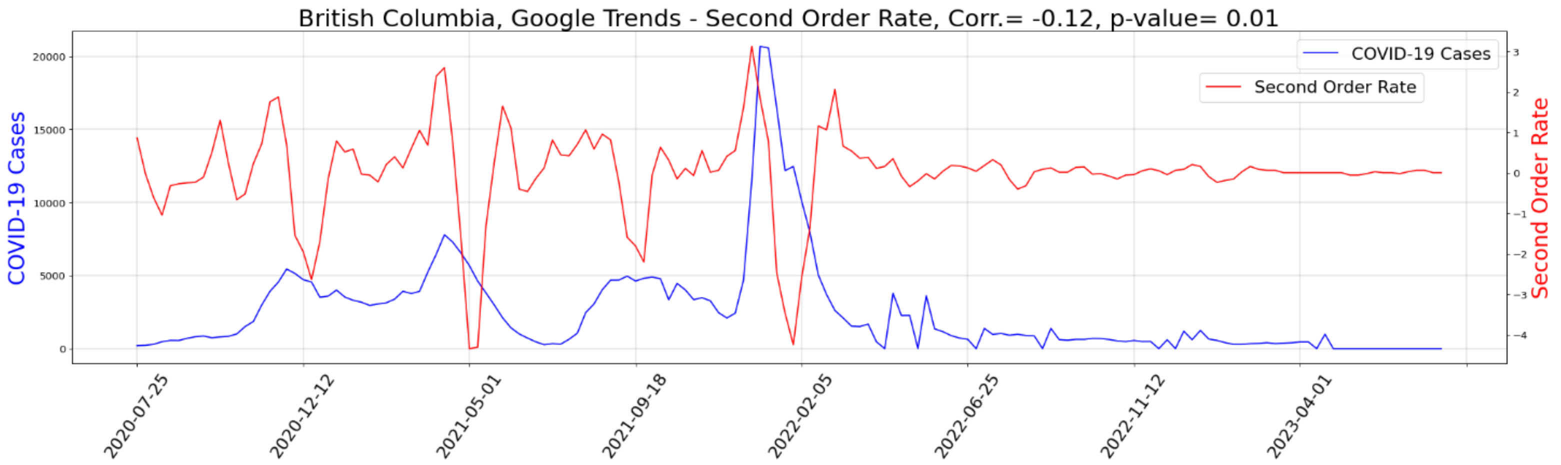
Google Trends

- The first and second order rates of the Google trends are also correlated with the number of cases.



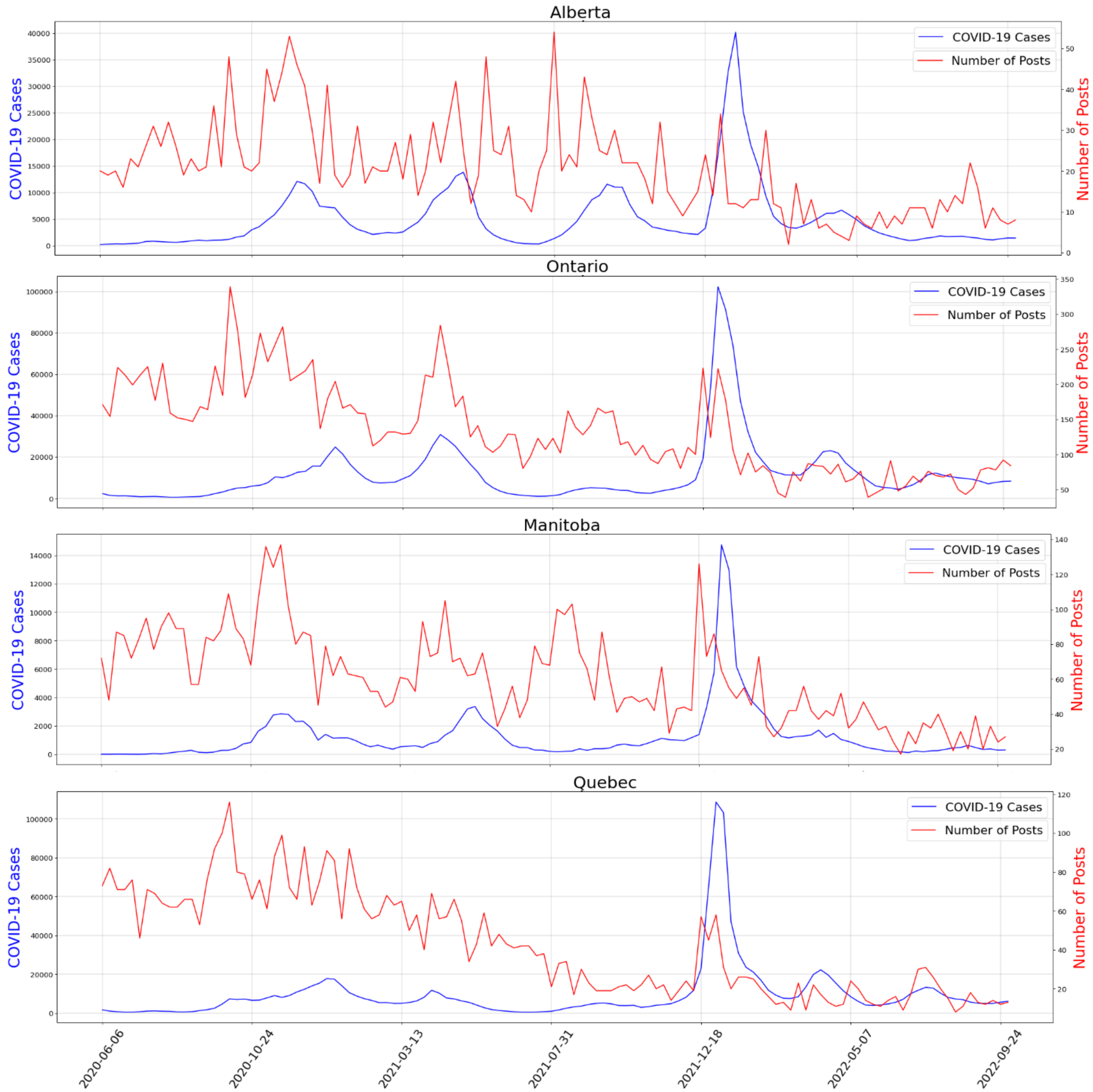






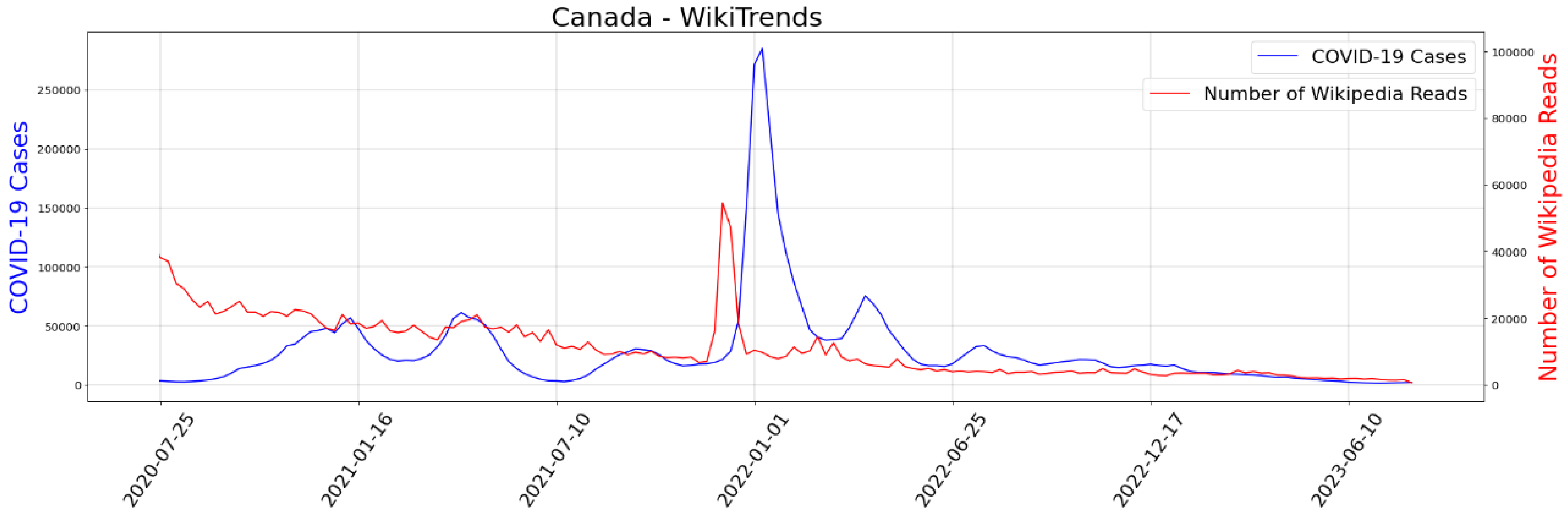
Social Media

- People are increasingly are using social media to share their opinions and experiences.
- As number of cases increase, people discuss the outbreak more on social media, therefore, social media platforms such as Reddit and Twitter are also a great source of predicting outbreaks.



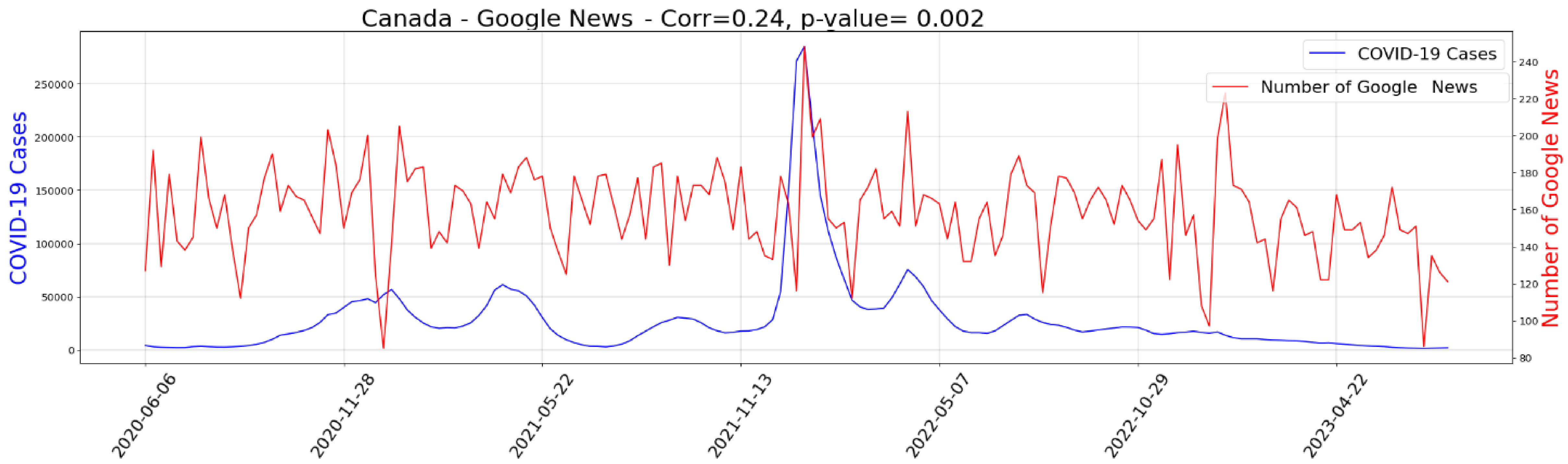
Wiki Trends

- For some disease (e.g. COVID-19), a separate Wikipedia page is available for different countries.
- The number of views on a specific page which could be retrieved using Wiki Trends rest API is also an indicator for the number of cases.
- For most countries Wikitrends is also well correlated with the number of cases.
- Sometimes Wikitrends peaks earlier than the number of cases.



News

- Volume of news released on a particular disease could also be used as an indicator of the number of cases.
- The number of Google news released in a particular country for a certain keyword, which could be retrieved using Google news API is mostly well correlated with the number of cases.

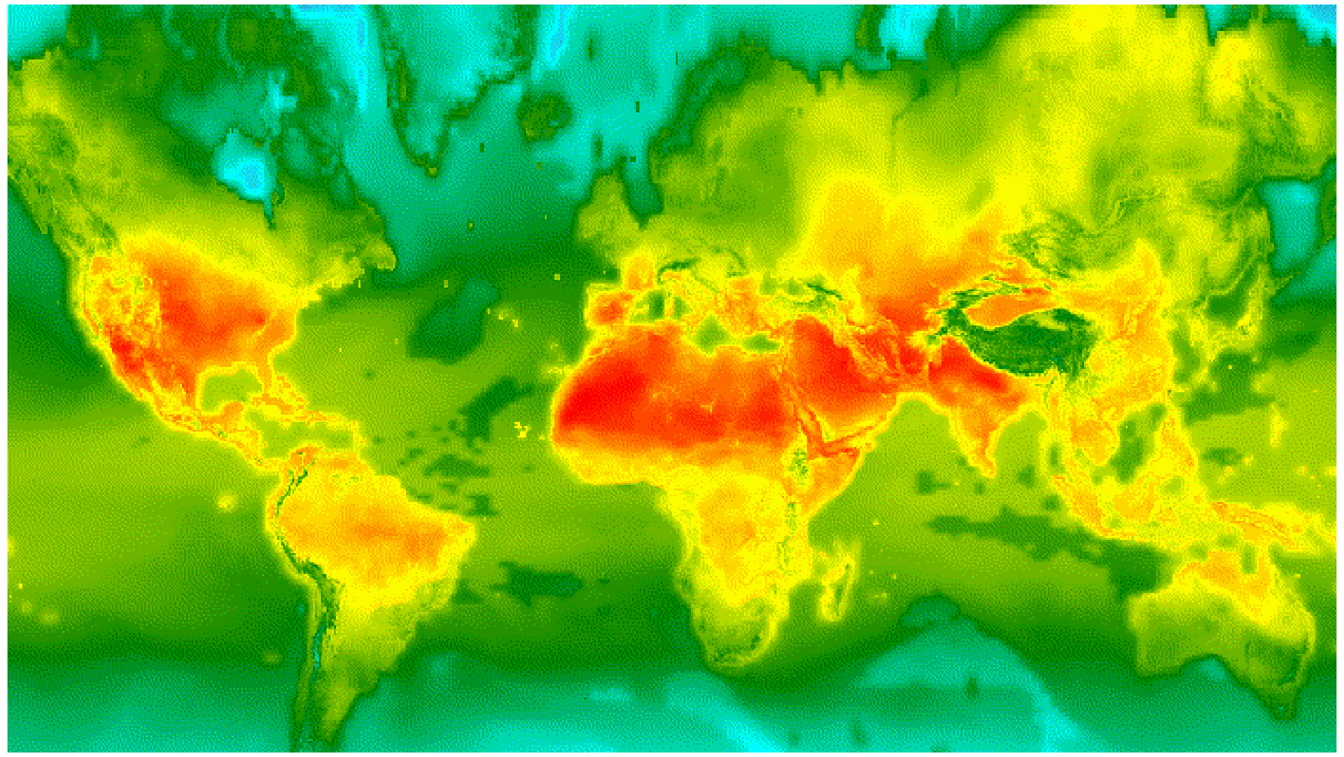


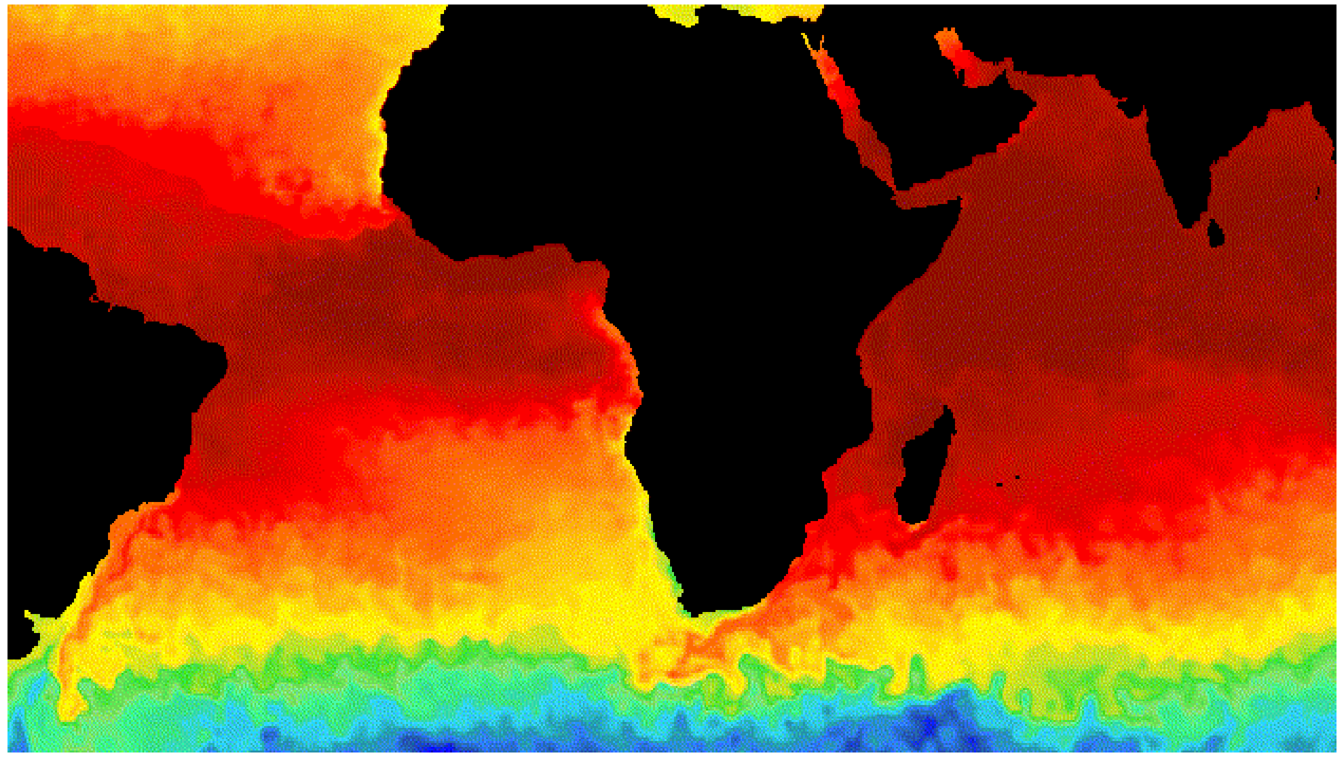
Satellite Data

- Multiple parameters could be obtained using satellite data including (but not limited to):

  - **Climate**: factors such as annual temperature, annual rainfall, isothermality, diurnal range temperatured etc could be obtained using climate data.

  - **Surface temperature**: sensors provide surface temperature and emissivity information. Vector born diseases could be well correlated with surface temperature data.



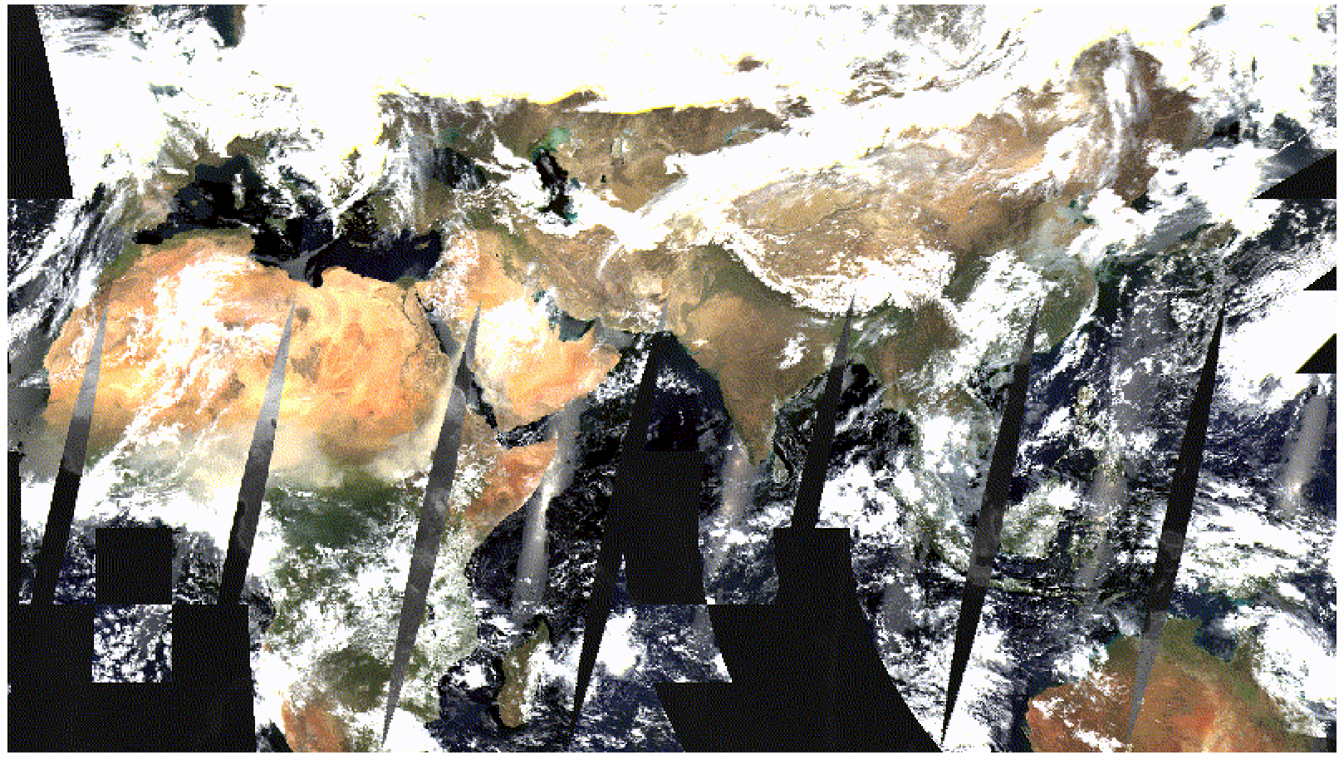


Satellite Data

- Multiple parameters could be obtained using satellite data including but not limited to:

  - **Weather data**: factors such as temperature, precipitation, snowfall, wind direction, gust, humidity, etc. could be obtained using weather data.

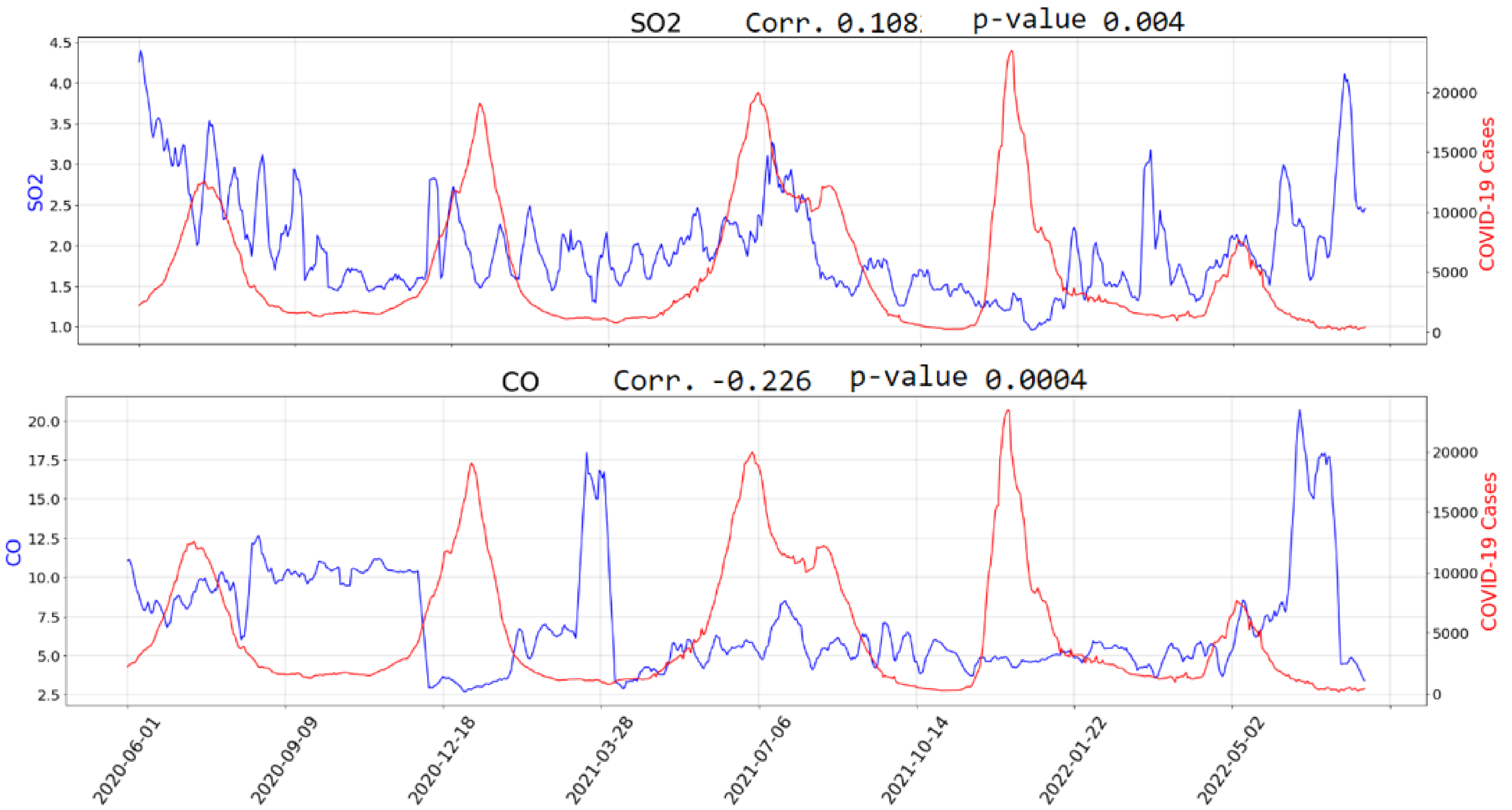
  - **Air quality data**: factors such as the concentration of NO_2, CO, SO_2, O_3, CH_4 could be obtained using SENTINEL-5 mission, which is part of the European Earth Observation Program.

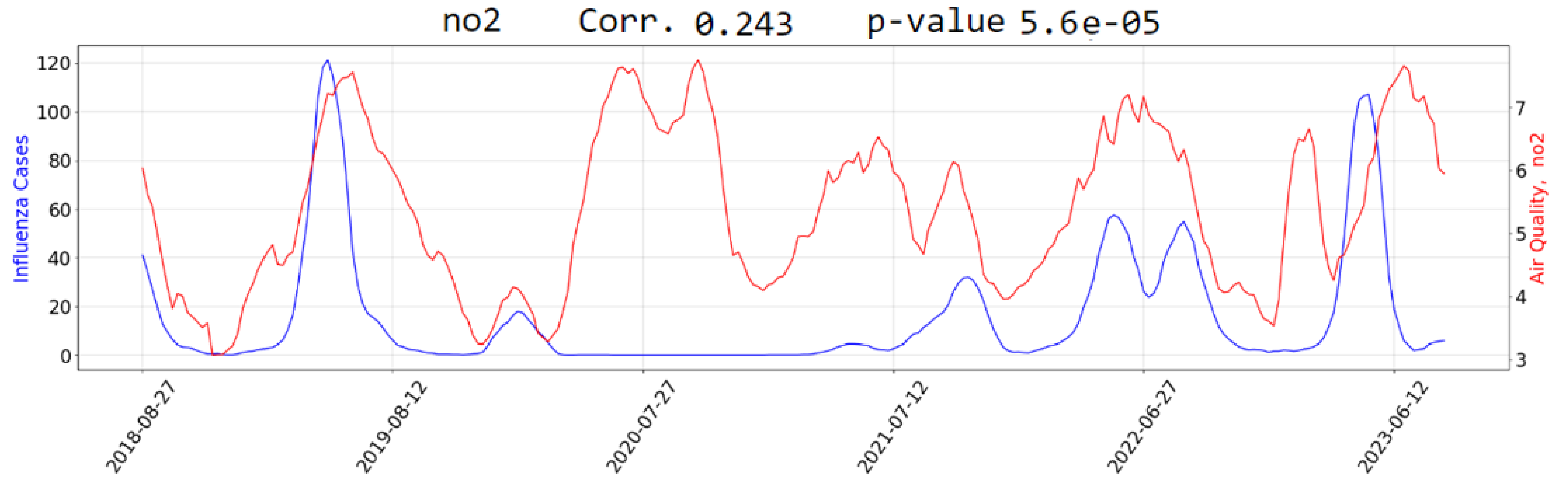


Air Quality Data

- Air quality parameters include concentration of elements such as CO, CO_2, NO_2, O_3, CH_4

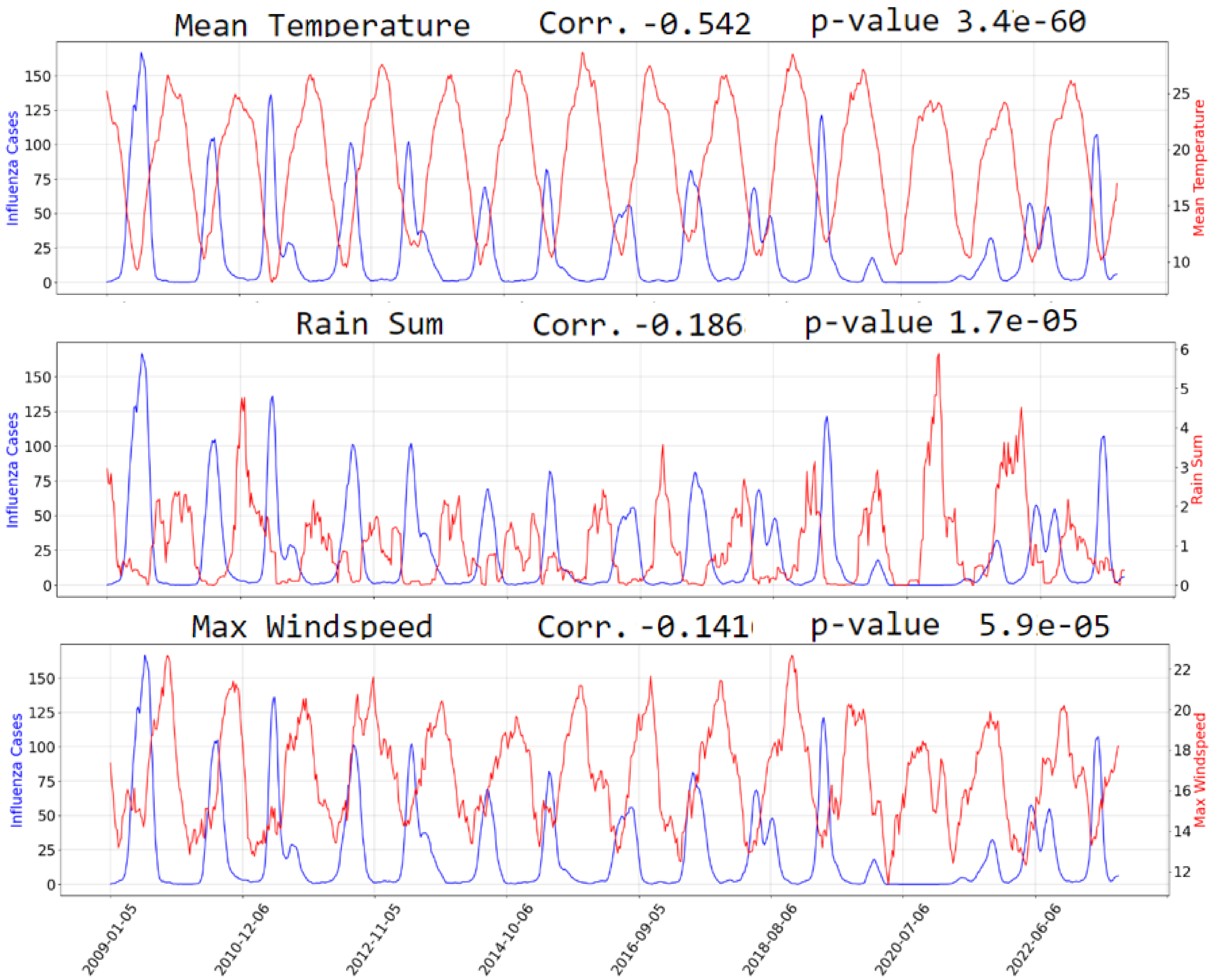
- Air quality parameters is correlated with respiratory diseases such as COVID-19 and Influenza.



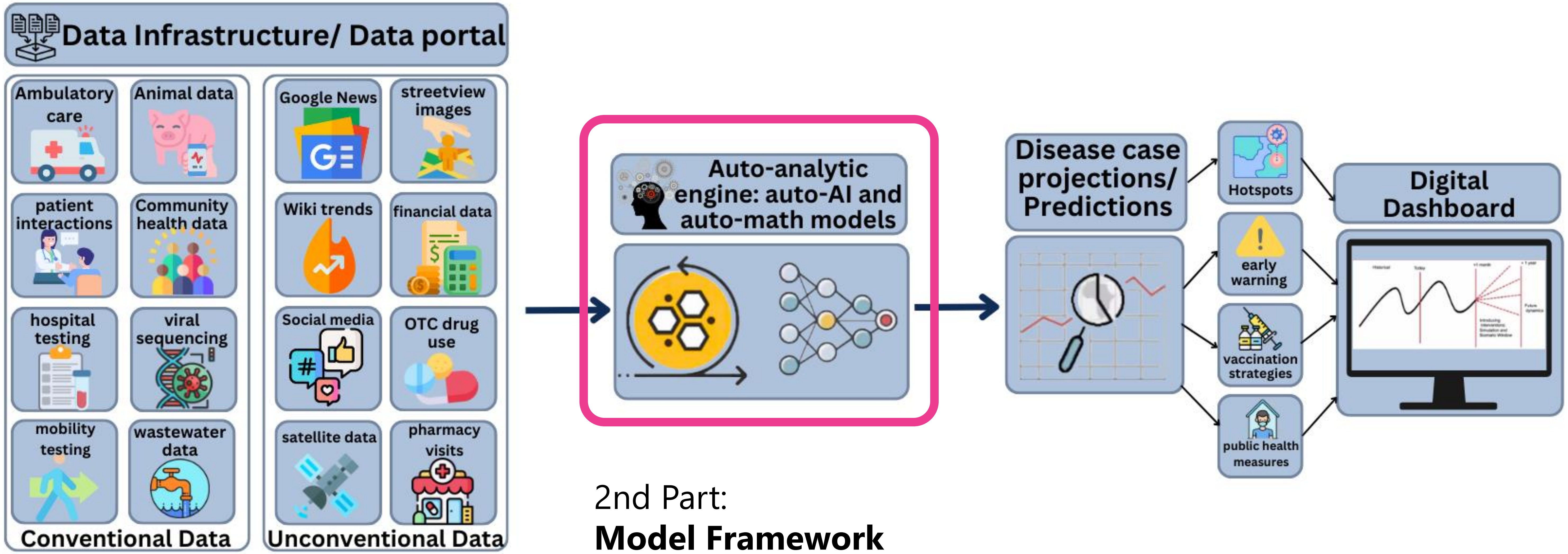


Weather Data

- Weather data includes parameters such as temperature, humidity, rainfall, wind speed, gust, etc.
- Weather data is correlated with Influenza cases.



AutoAI Epidemix



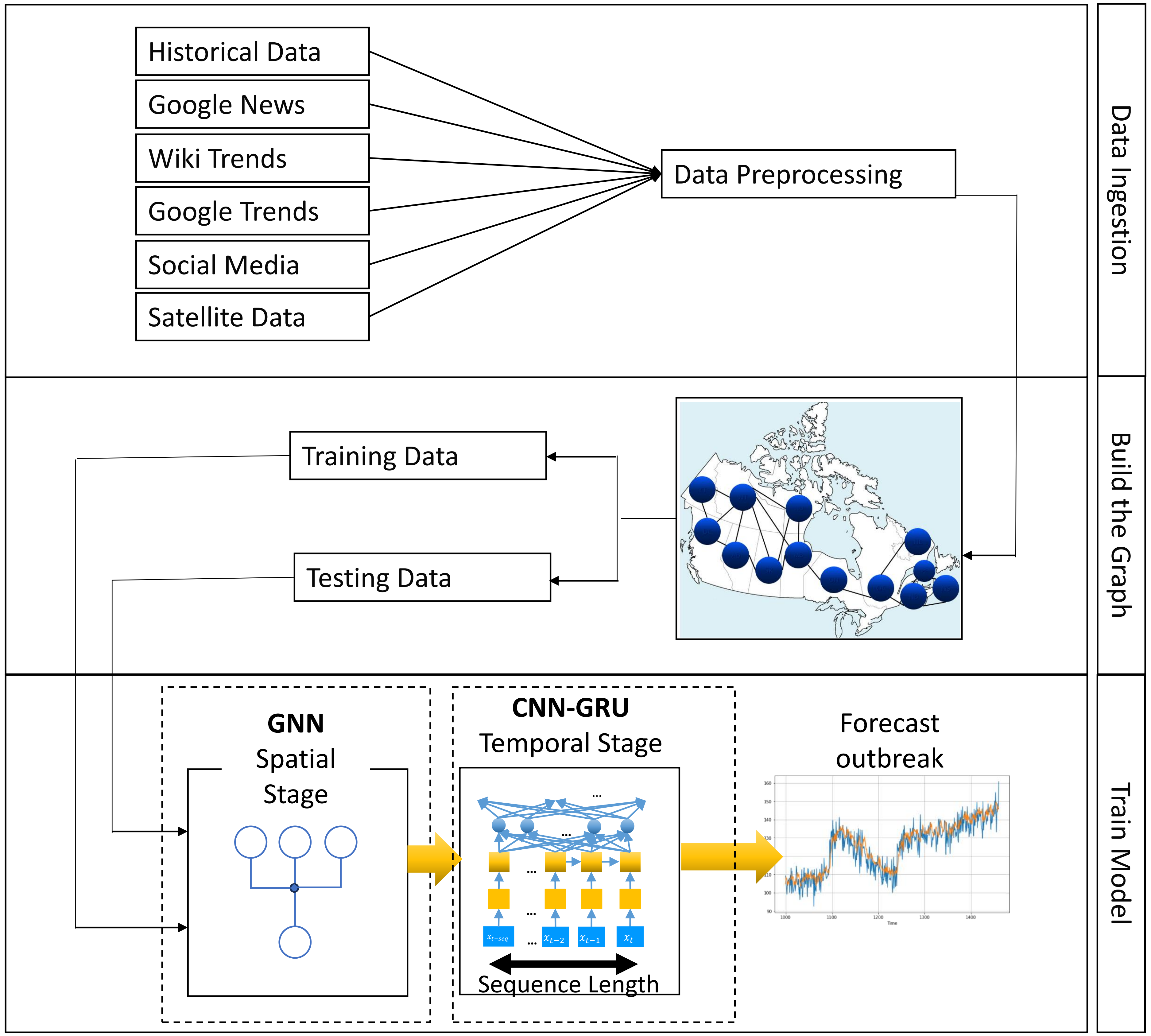
2nd Part:
Model Framework

Community Health Data

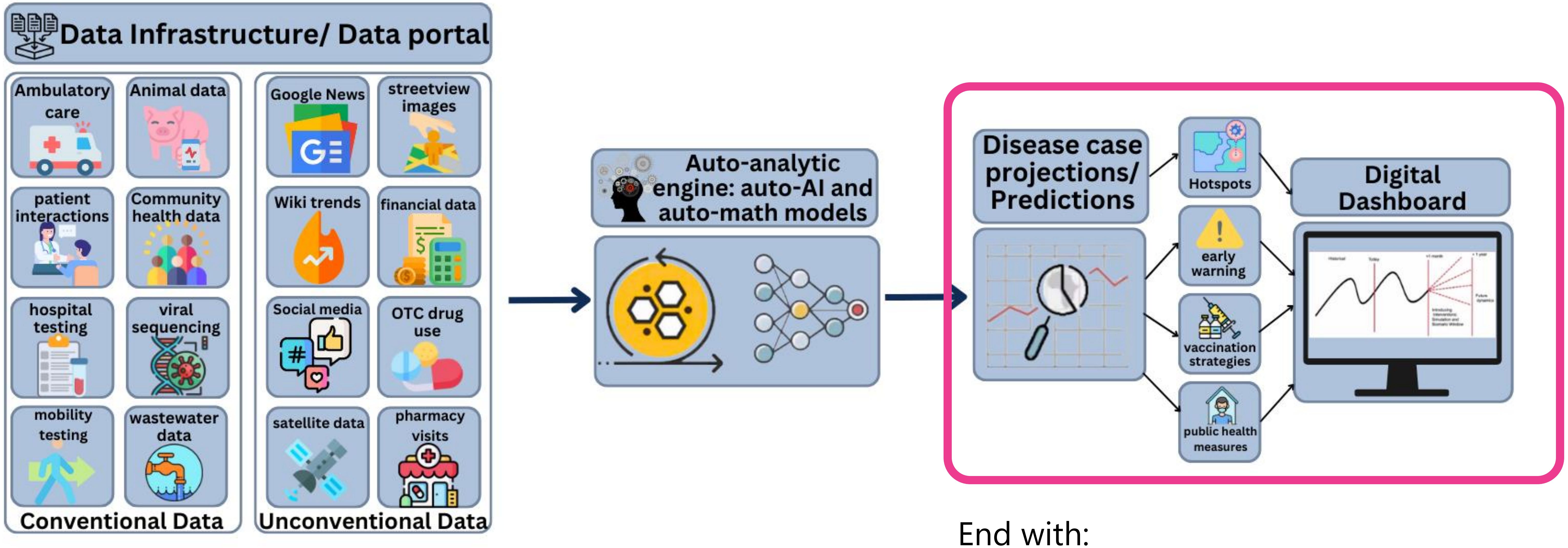
Flow-Diagram

The flow diagram includes three blocks:

1. **Data Ingestion:**
   - For each province, multiple sources of data are treated as time-series, and stored into tensors.
   - The preprocessing includes replacing missing data with zero, specifying the features and the labels, and finally, center and scaling the values.
2. **Build the Graph:**
   - The nodes and their data, the edges and weights are defined.
   - The train and test time-series of each node is prepared.
3. **Train Model:**
   - In each node, the data of the neighbors are combined and the result goes through an RNN model which is CNN-GRU.
   - The model is trained and then tested.



AutoAI Epidemix

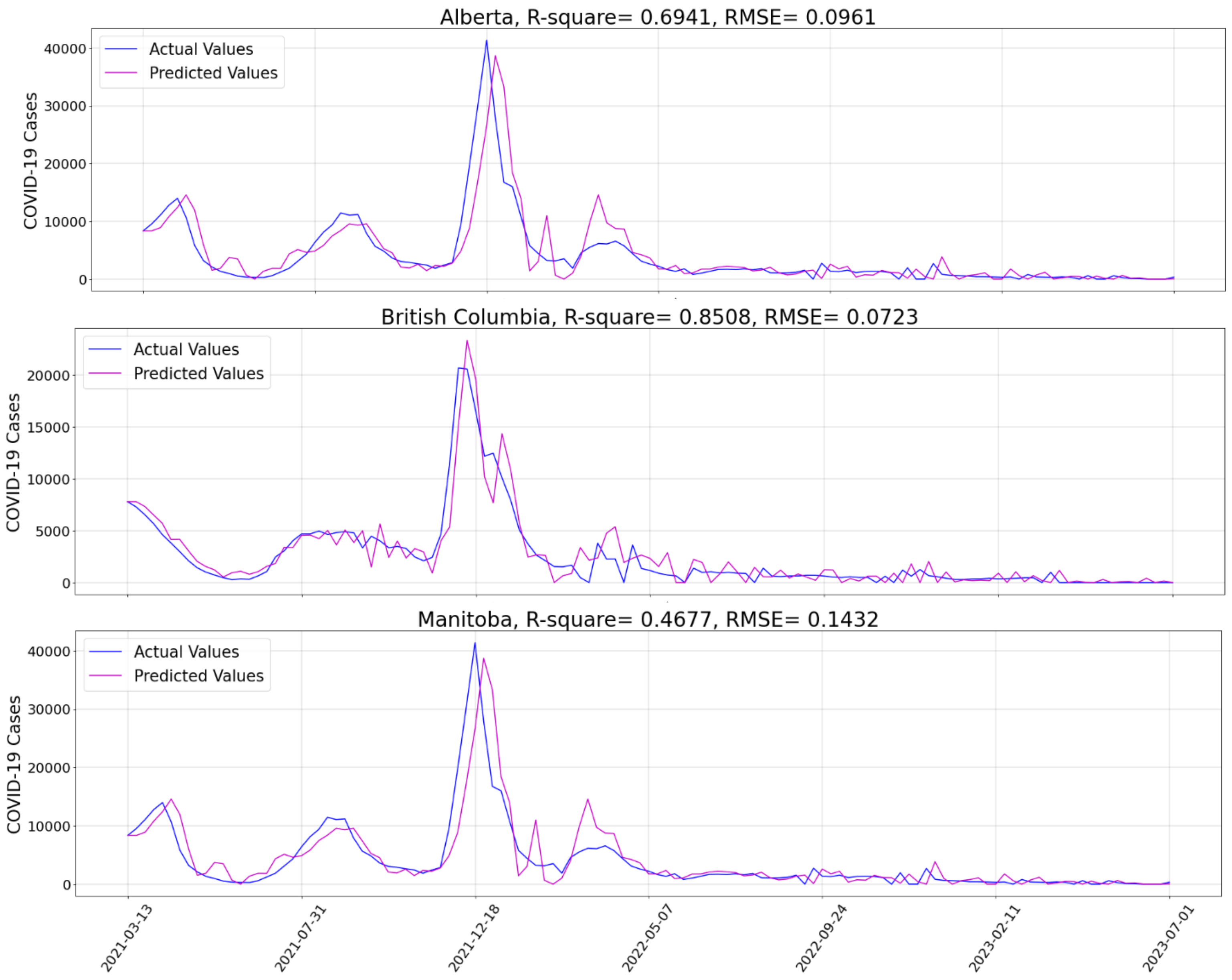


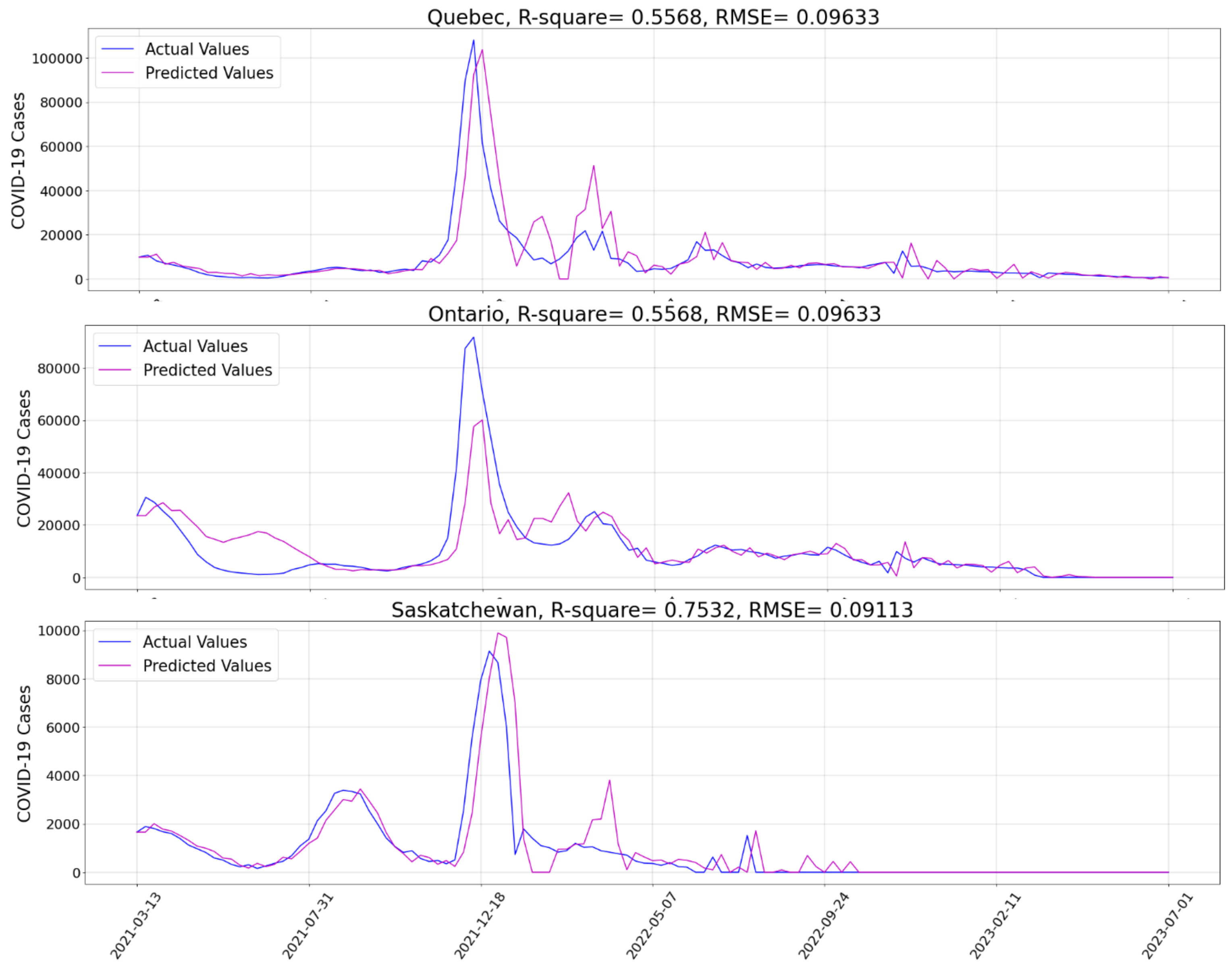
End with:

Examples (Influenza, Lyme, Covid)

Canada: 14 Step-Ahead Prediction

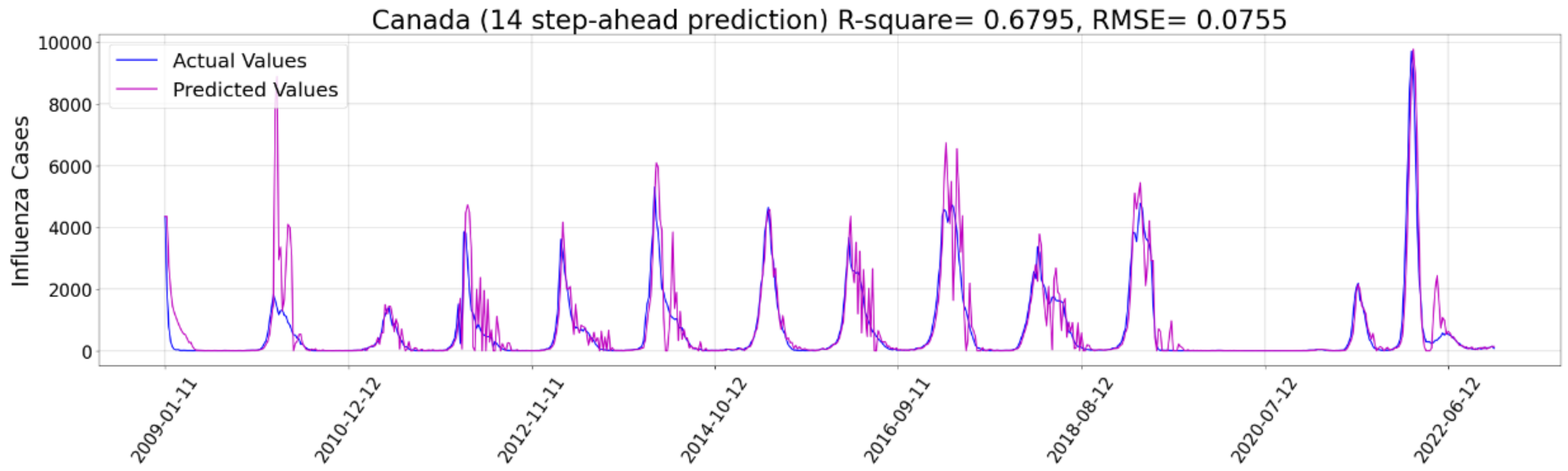
- COVID-19 cases of different provinces of Canada are available on weekly basis.
- The final model predicts COVID-19 waves of all the provinces very well.

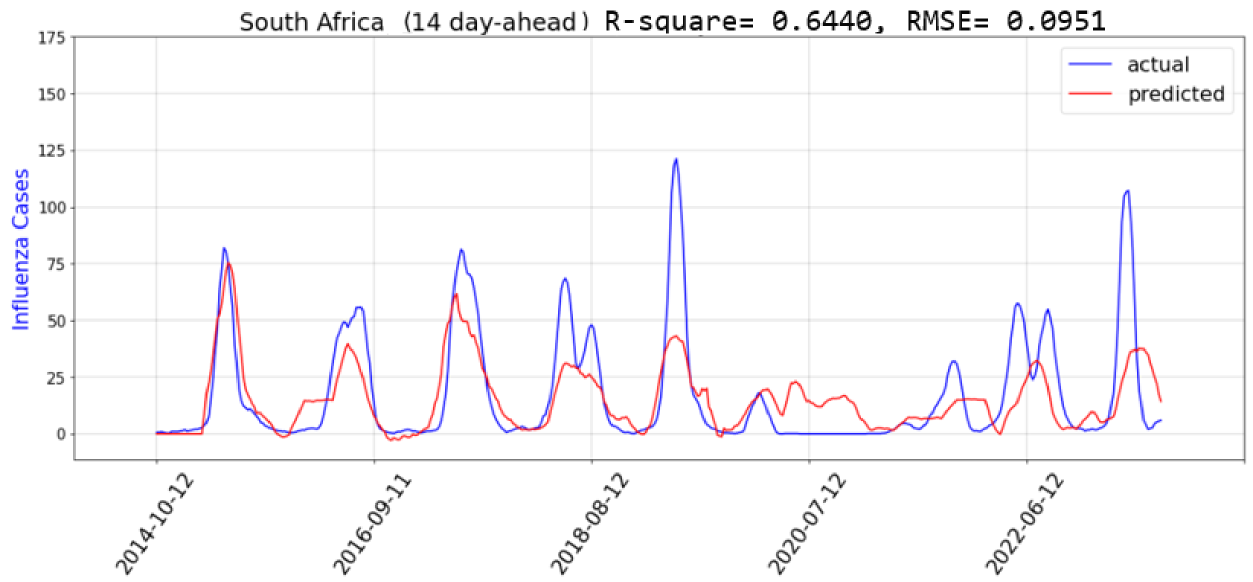


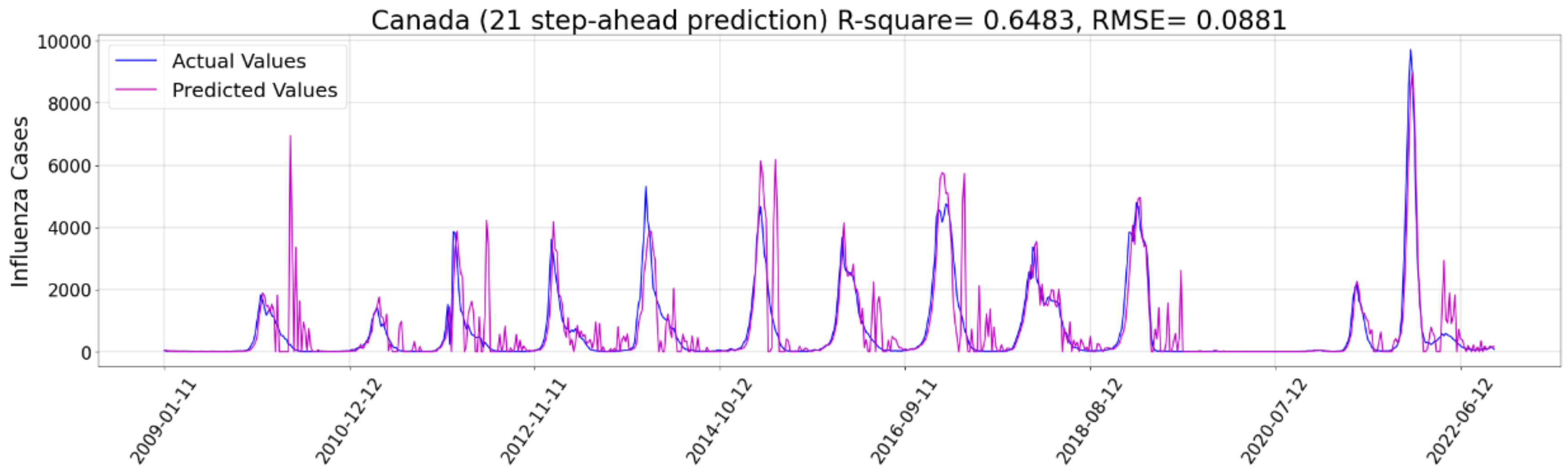


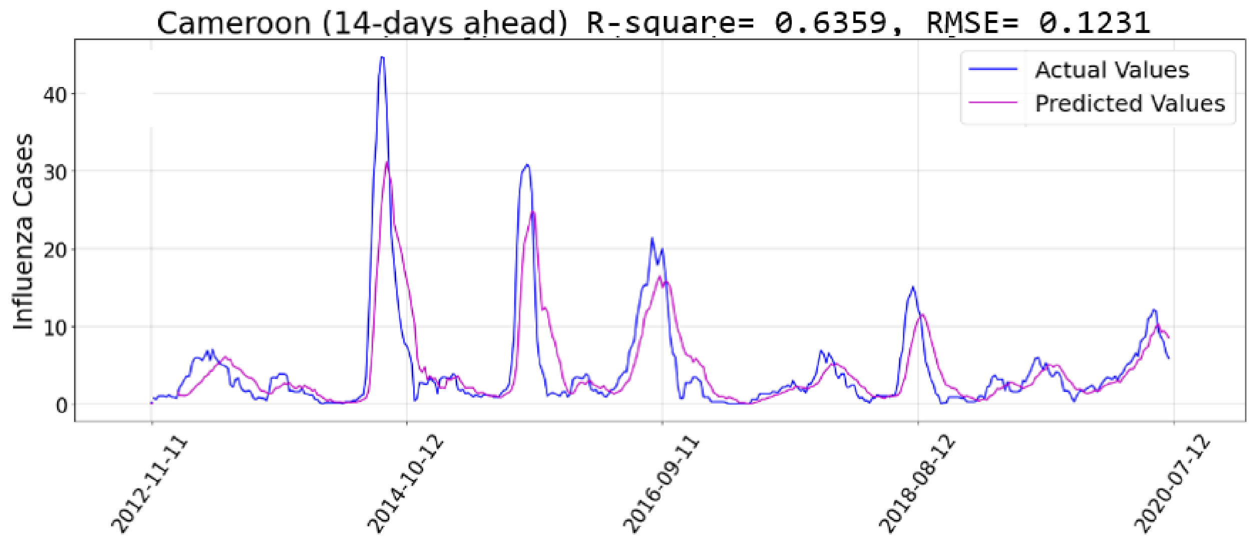
Forecasting Influenza

- Influenza waves have been predicted for different countries with an outstanding accuracy.



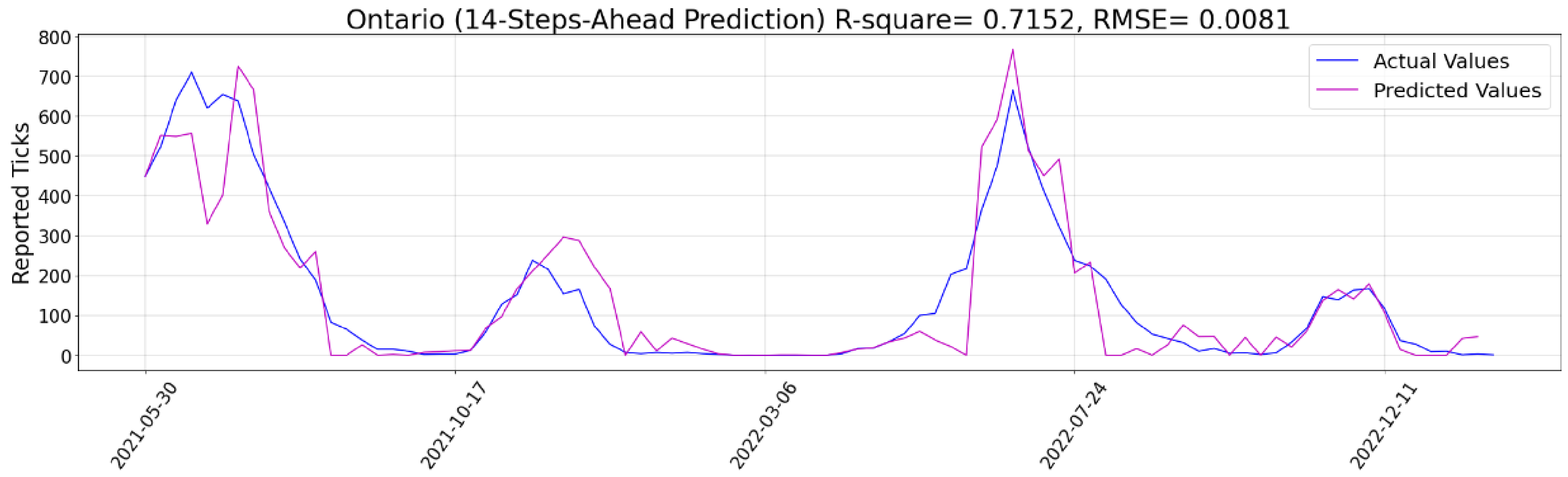


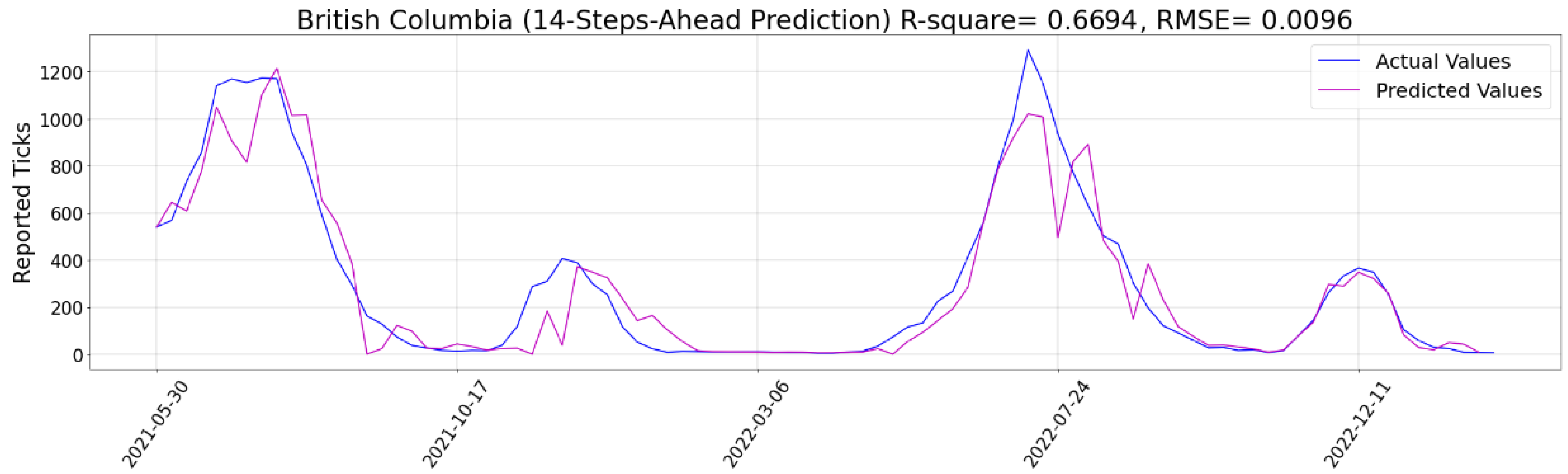




Forecasting Lyme Disease

- The etick dataset was used as an indicator of Lyme diseases prevalence in different provinces.
- Our model is able to predict the volume of ticks for different provinces with an outstanding accuracy.



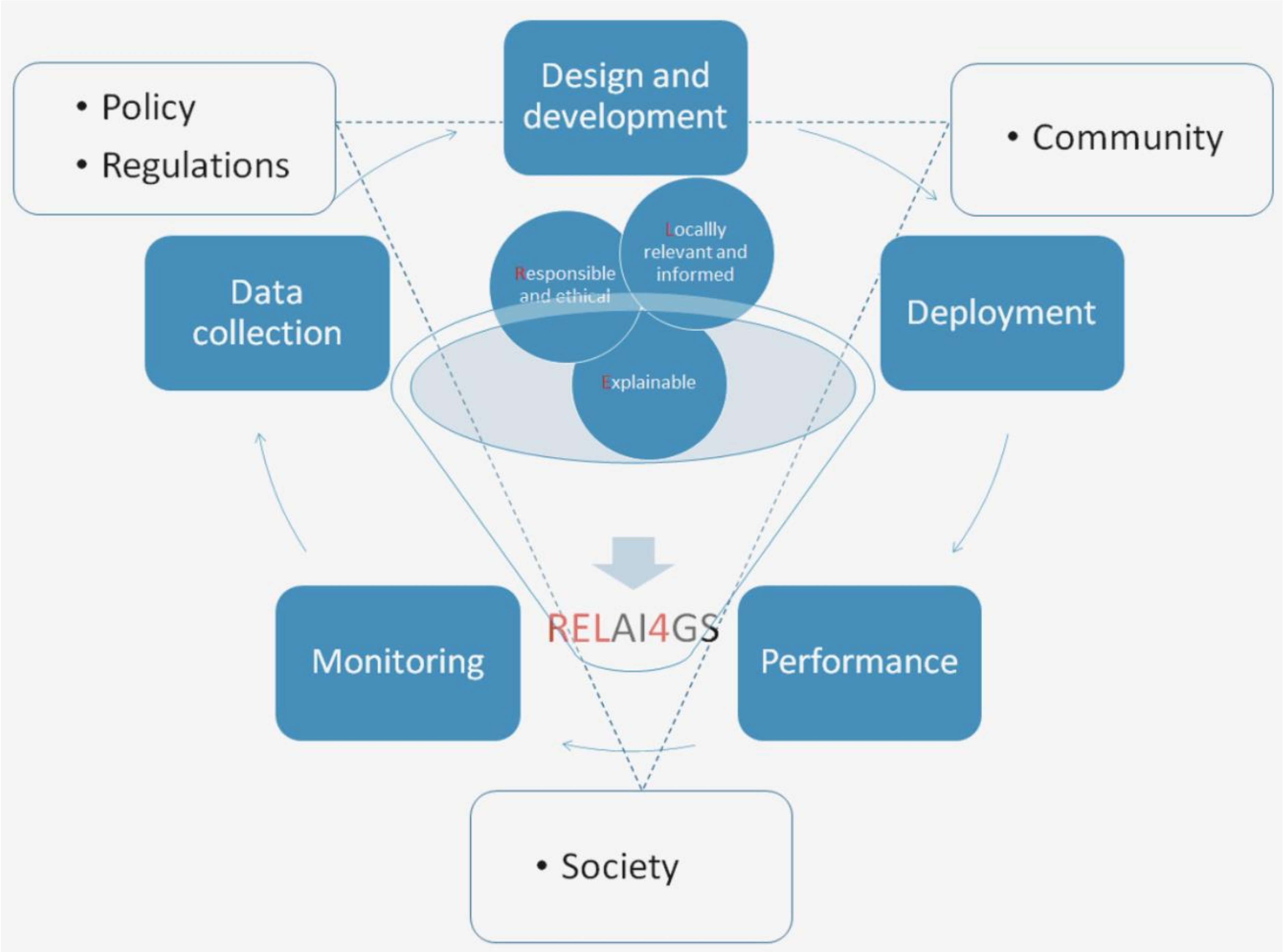


Resemblance between Historical Colonialism and Current Data Colonialism

Historical Colonialism	Current Data Colonialism
Appropriation of natural resources	**Appropriation** [and quantification] of human life (through datafication)
Expropriation of land, resources, bodies	**Expropriation** of social life (e.g., social media) and bodies (e.g, IoT is upcoming) ▪ People are "just there" for capital to "discover" and exploit
Exploitation through industrial capitalism	**Exploitation** through AI capitalism (commodification of human life)

Our Framework

- **Responsible**: Accountable, auditable, compliant, ethical, respectful, safe, and secure
- **Explainable**: Equitable, fair, interpretable, reliable, reproducible, transparent, trustworthy, unbiased
- **Local**: Autonomous, caring, connecting, decolonized, human- and community-centred, inclusive, intentional, intersectional, just, practical, protecting, process-based, sustainable




Kong Research Group 2022
Kevin Hui
Mahakprit Kaur
Wisdom Avusuglo
Jude Dzevela Kong
Haleema Ahmed
Sherif Shuaib
Blessing Ogbuokiri
Andrew Omame
Mahnaz Alavinejad
Selin Tahir
Qing Han
Cameron Arboine
YORK U
UNIVERSITÉ
UNIVERSITY
IDRC · CRDI
Canada
NSERC
CRSNG
Zahra Movahedinia
J
Jared Wieclawek
AFRICA - CANADA
ARTIFICIAL INTELLIGENCE AND DATA INNOVATION CONSORTIUM
Sida

Dr. Zahra Nia

Acknowledgment: Funders

NSERC CRSNG

New Frontiers in Research Fund
Fonds Nouvelles frontières en recherche

IDRC · CRDI
International Development Research Centre
Centre de recherches pour le développement international

Canada

Sida

YORK U
UNIVERSITÉ
UNIVERSITY

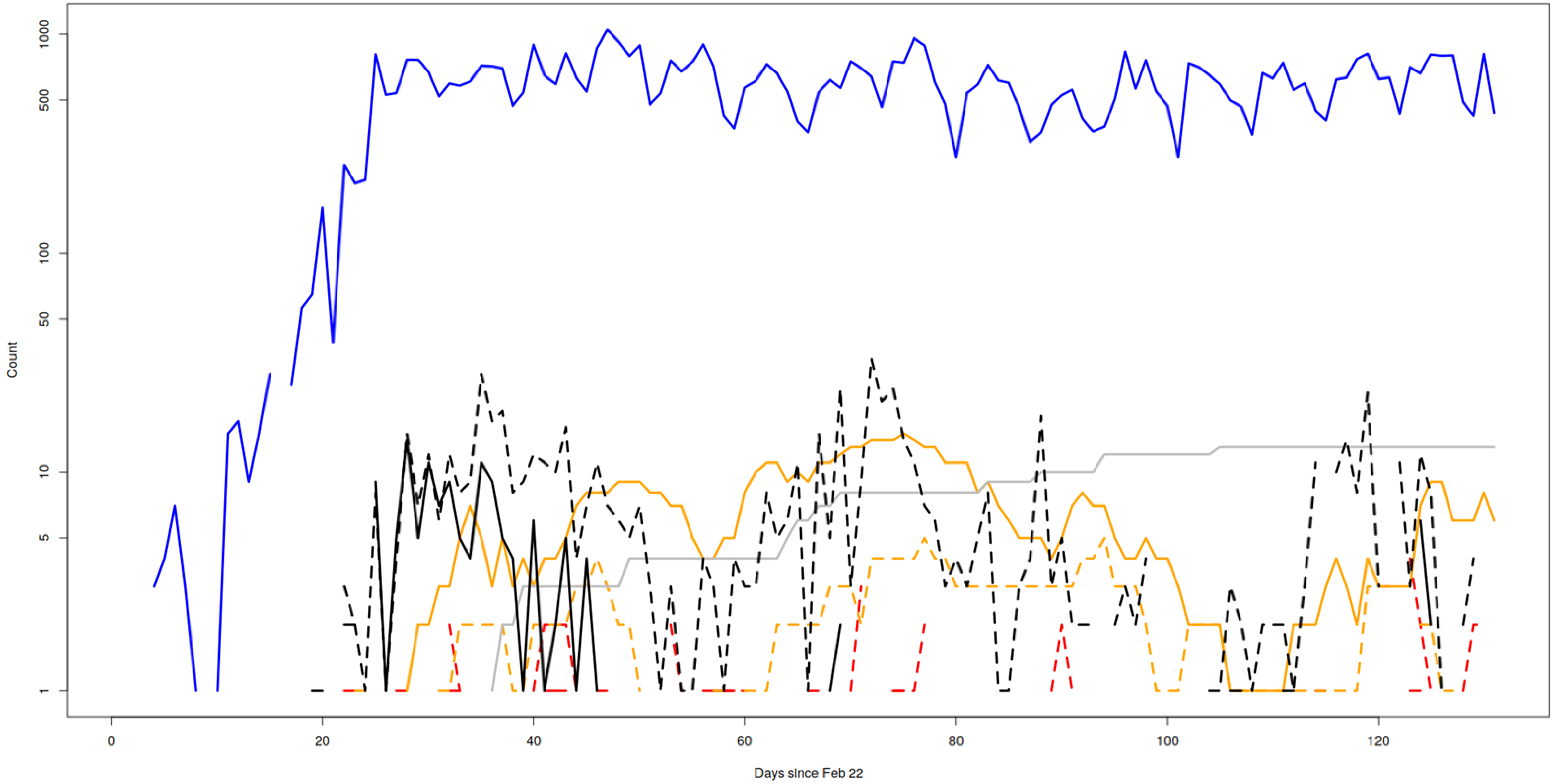
Conclusion

- An agile early warning, alert, and response system (like AI-Epidemix) is paramount for controlling and containing infectious disease.
- Multiple sources of data (e.g. Google trends, social media, satellite data, and street view images) are collected on regional/community level.
- Machine learning methods, particularly Recurrent Neural Networks and Graph Neural Networks are utilized to forecast emerging and re-emerging diseases.
- The framework will assist policy-makers, health officials, as well as physicians by accurately forecasting various disease outbreaks.

Service Delivery from Models: Real-time Multipathogen Epidemiology & Acute Care Demand Monitoring and Nowcasting via PMCMC-Leveraged Transmission Models

Dr. Nathaniel Osgood PhD
Computational Epidemiology & Public Health Informatics Laboratory
University of Saskatchewan

Diverse Reported Data Sources



Wastewater Epidemiology

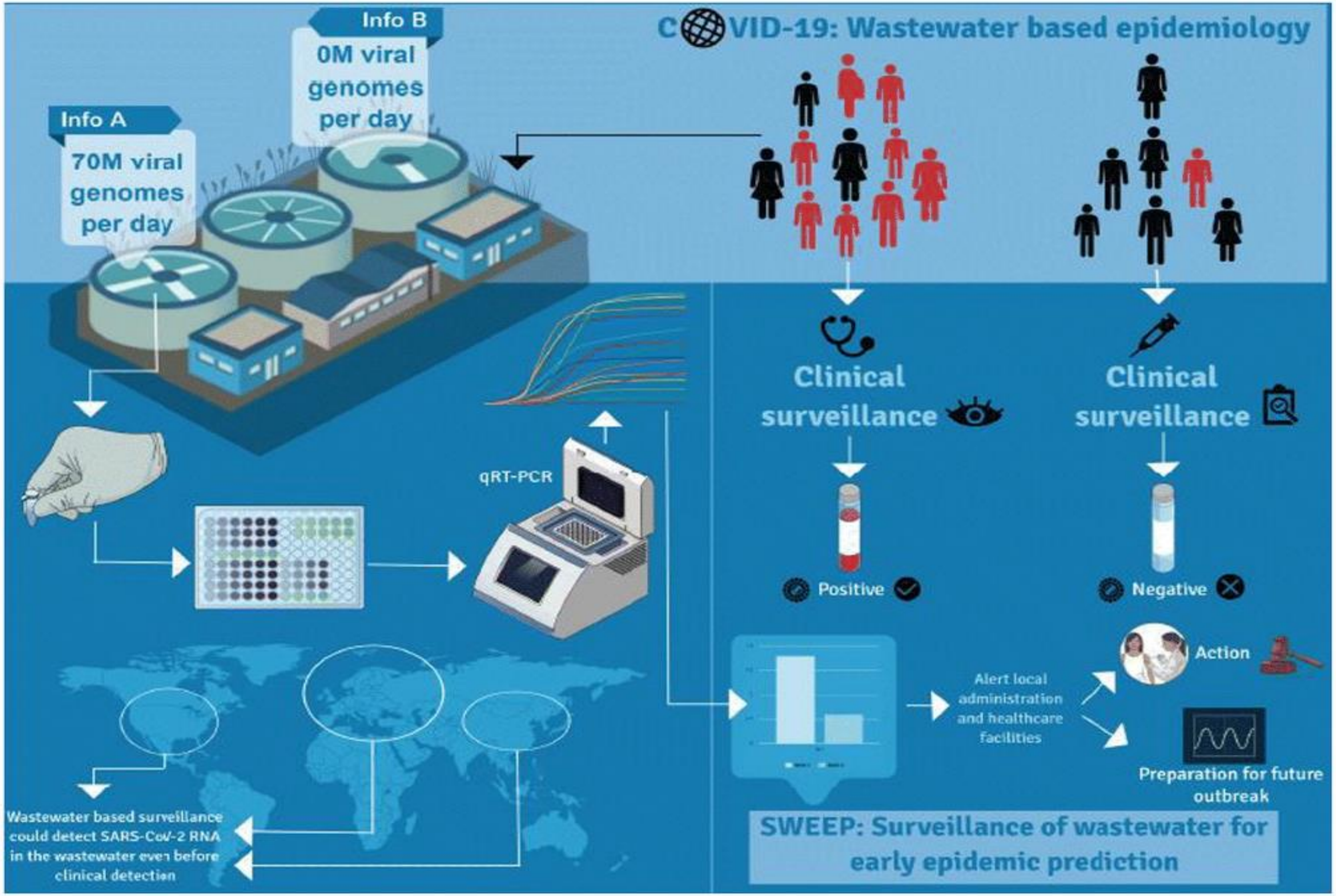
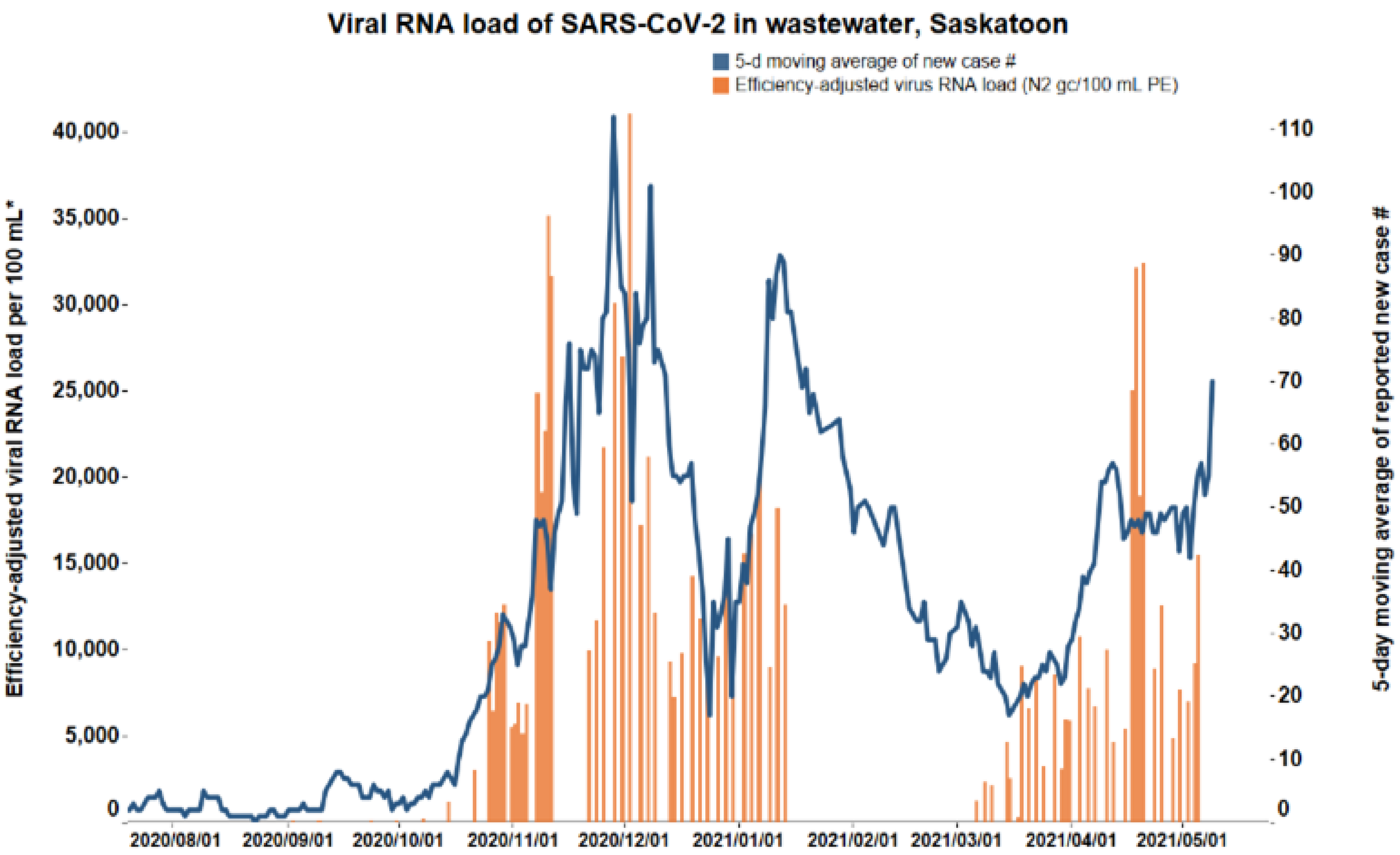


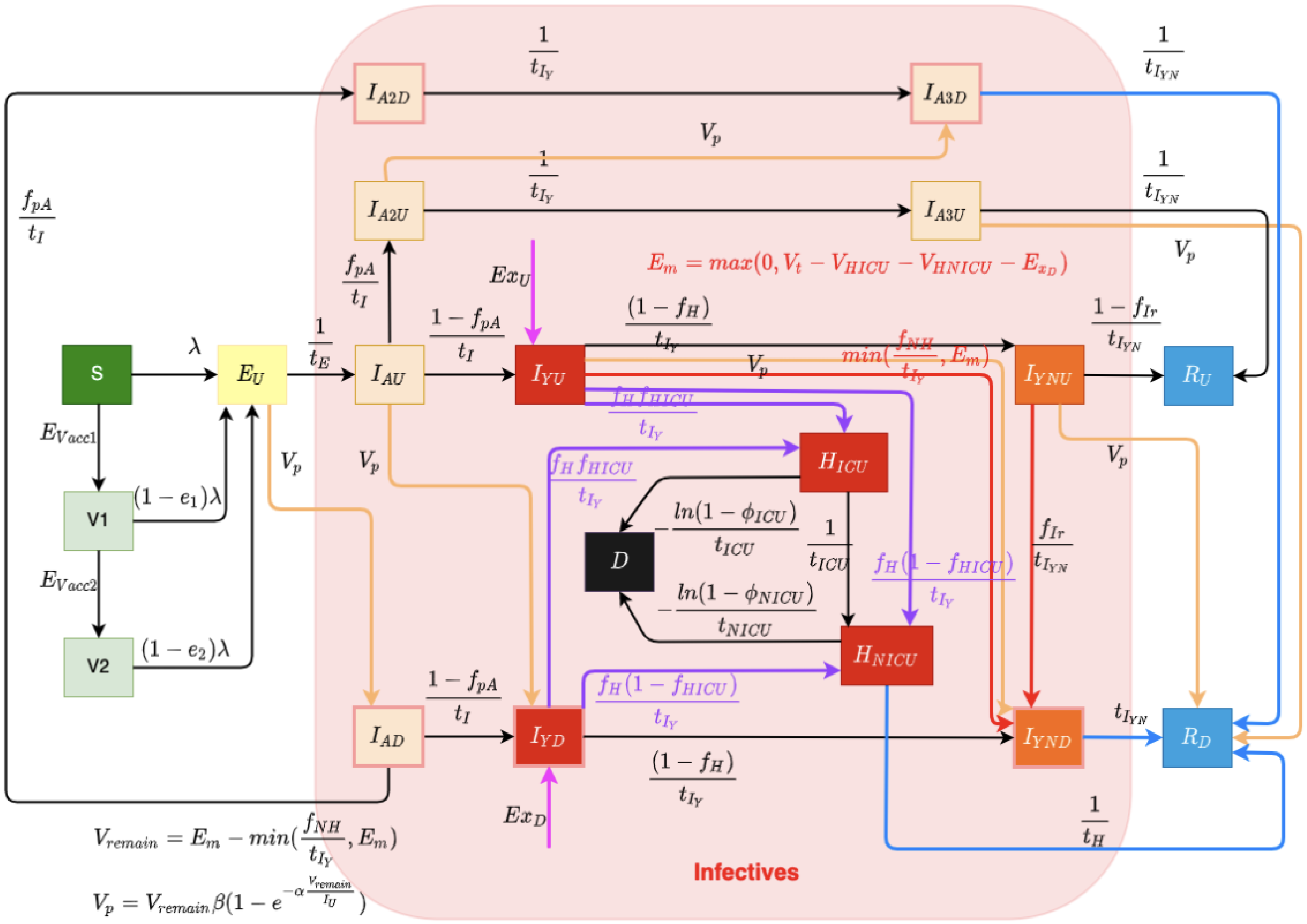
Image source:
Kumar, M., Mohapatra, S., Mazumder, P. et al. Making Waves Perspectives of Modelling and Monitoring of SARS-CoV-2 in Aquatic Environment for COVID-19 Pandemic. Curr Pollution Rep 6, 468–479 (2020). https://doi.org/10.1007/s40726-020-00161-5

Wastewater Data



Source:
University of Saskatchewan Toxicology Centre via COVID-19 Early Indicators,
https://water.usask.ca/covid-19/#MeasuringVirusIndicatorsinWastewaterasanEarlyWarningofCOVID19Outbreaks

Transmission Models



Reflections

- Unassisted, all models diverge from empirical situation as time passes.
  - Divergence between model state & empirical state
  - Some relevant challenges: Stochastics, exogenous changes, approximations, omissions, heterogeneity …
- Divergence can strongly limit effectiveness of model evaluation of intervention tradeoffs.
- Merely adjusting parameters will not support ongoing alignment between model & empirical data.

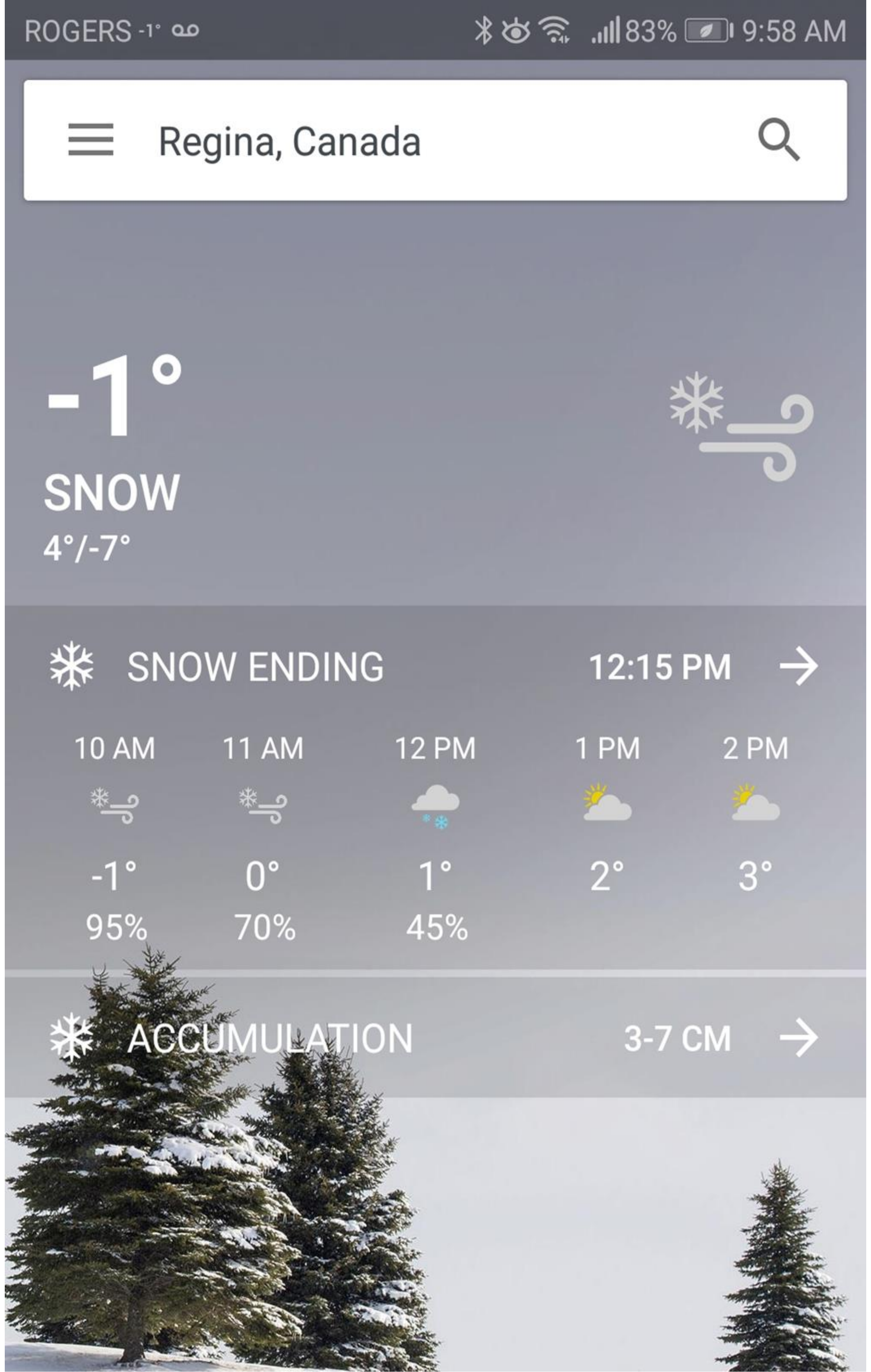
Making Sense of the Evidence: Dynamic Models as Always-Updated Services

- Render current understanding of natural history of infection & diverse incoming data sources into evolving integrated picture of underlying current epidemiology
    - Changes in behaviour
    - Count of undiagnosed infectives
    - Force of infection
    - Regional distribution of cases
    - Effective reproductive number

- Projection forward over days & weeks from current situation

- Capacity to evaluate policy scenarios for public health & acute care needs, looking forward from current situation

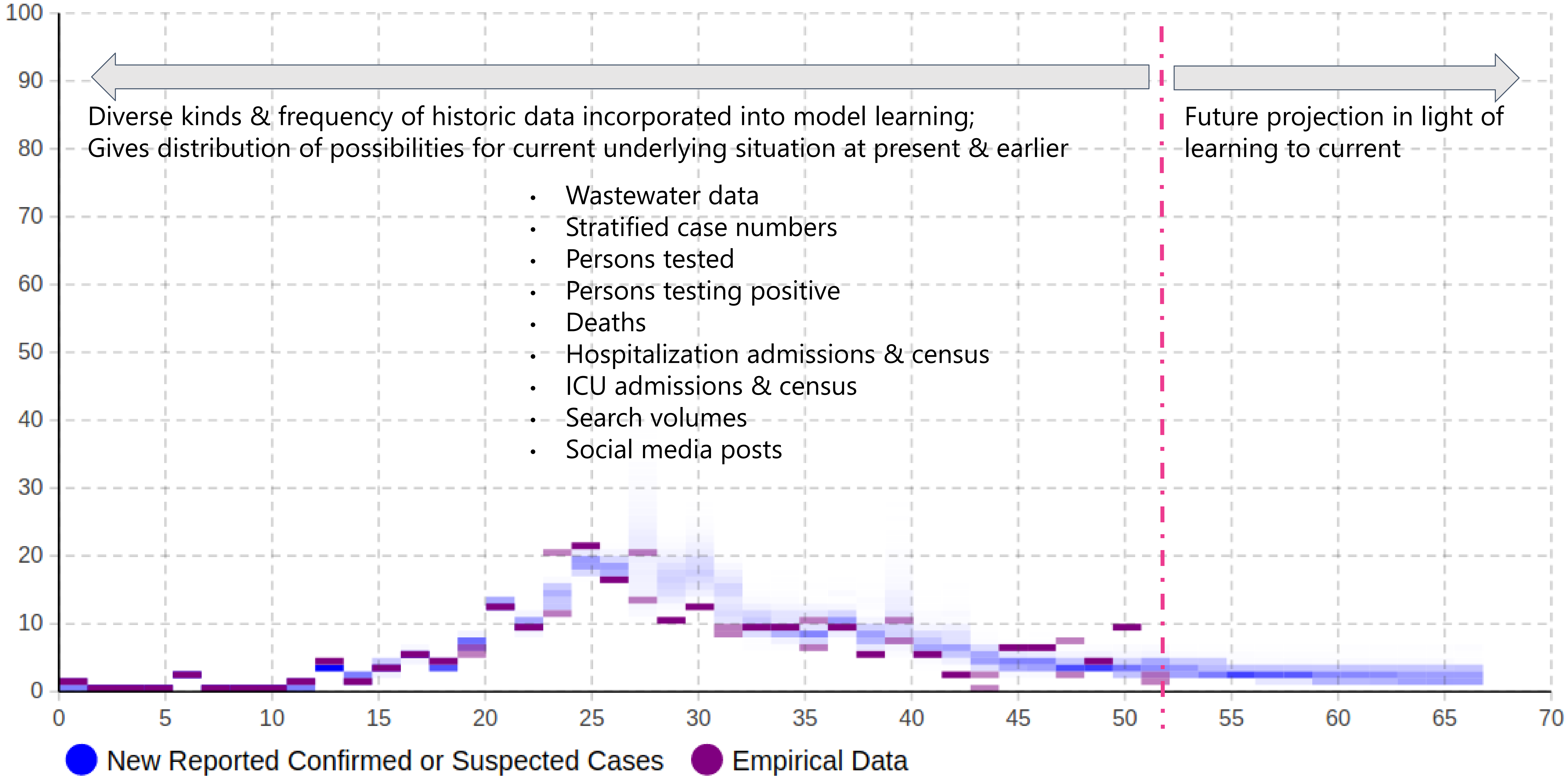
As with weather forecasts, we keep our models updated to reflect the latest evidence and use that to anticipate future state (intervention scenarios, project forward).



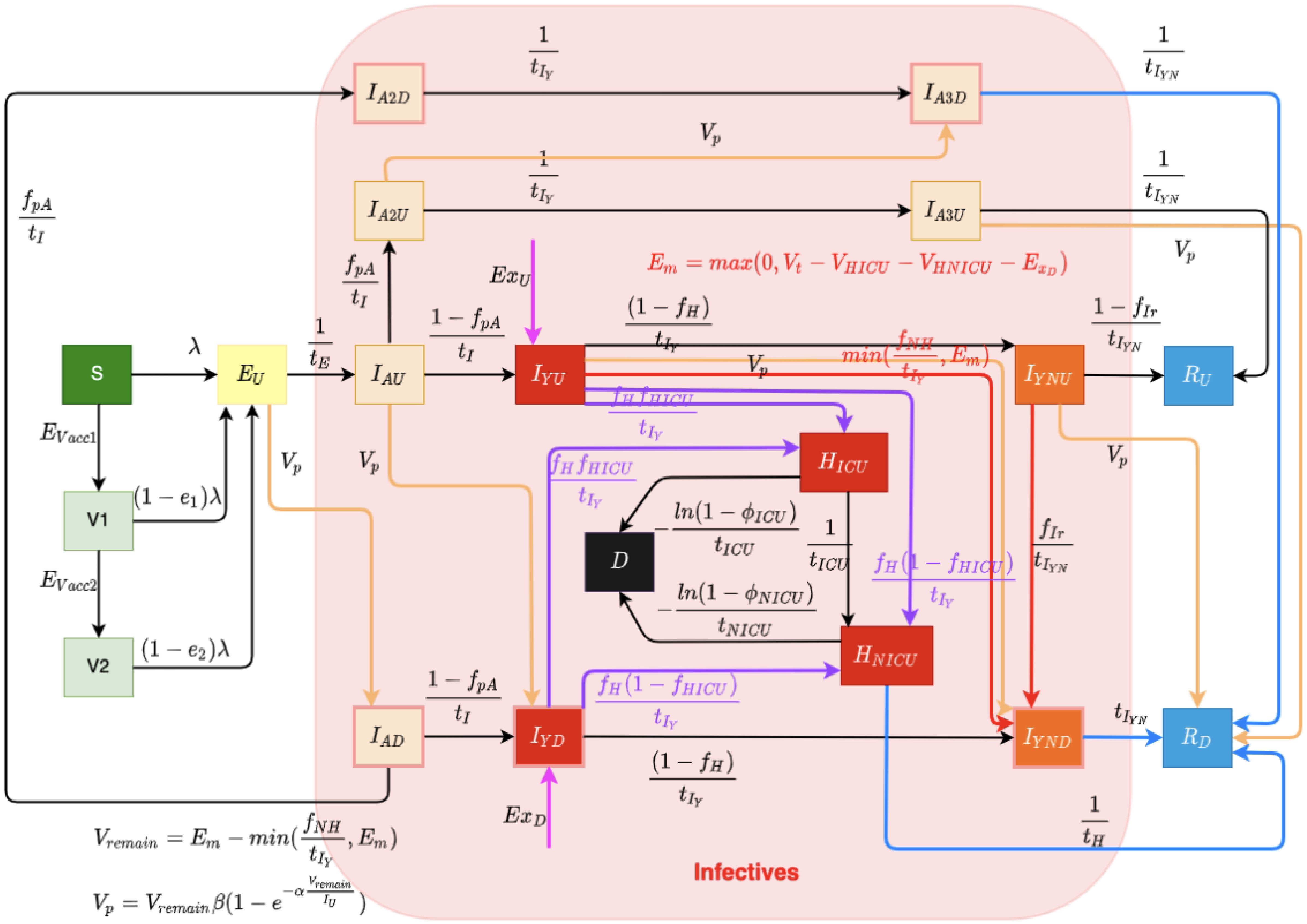
Highest-level Points

- These are **not curve-fitting models**. Instead, they ask "what's going on?" in terms of the underlying situation (in light of theory & observations to the present).
- The model uses Bayesian probability & dynamic models to identify a coherent understanding – consistent with clinical/epidemiologic understanding of COVID-19 – about the current situation that best explains what is observed across many types of data (e.g., cases, test volumes, hospital admissions, hospital census, etc.).
    - Any one type of data reflects a different facet of this underlying situation
- This process involves AI inference from observed data in ways that square with understanding of COVID-19 natural history & epidemiology (as captured in model).
- Because the models infer the underlying situation consistent with theory, they can project forward with & without additional interventions/measures.
- These projections are not projecting forward curves – they are projecting the "momentum" of the situation.

Adaptive Planning: Observing Unfolding Evidence



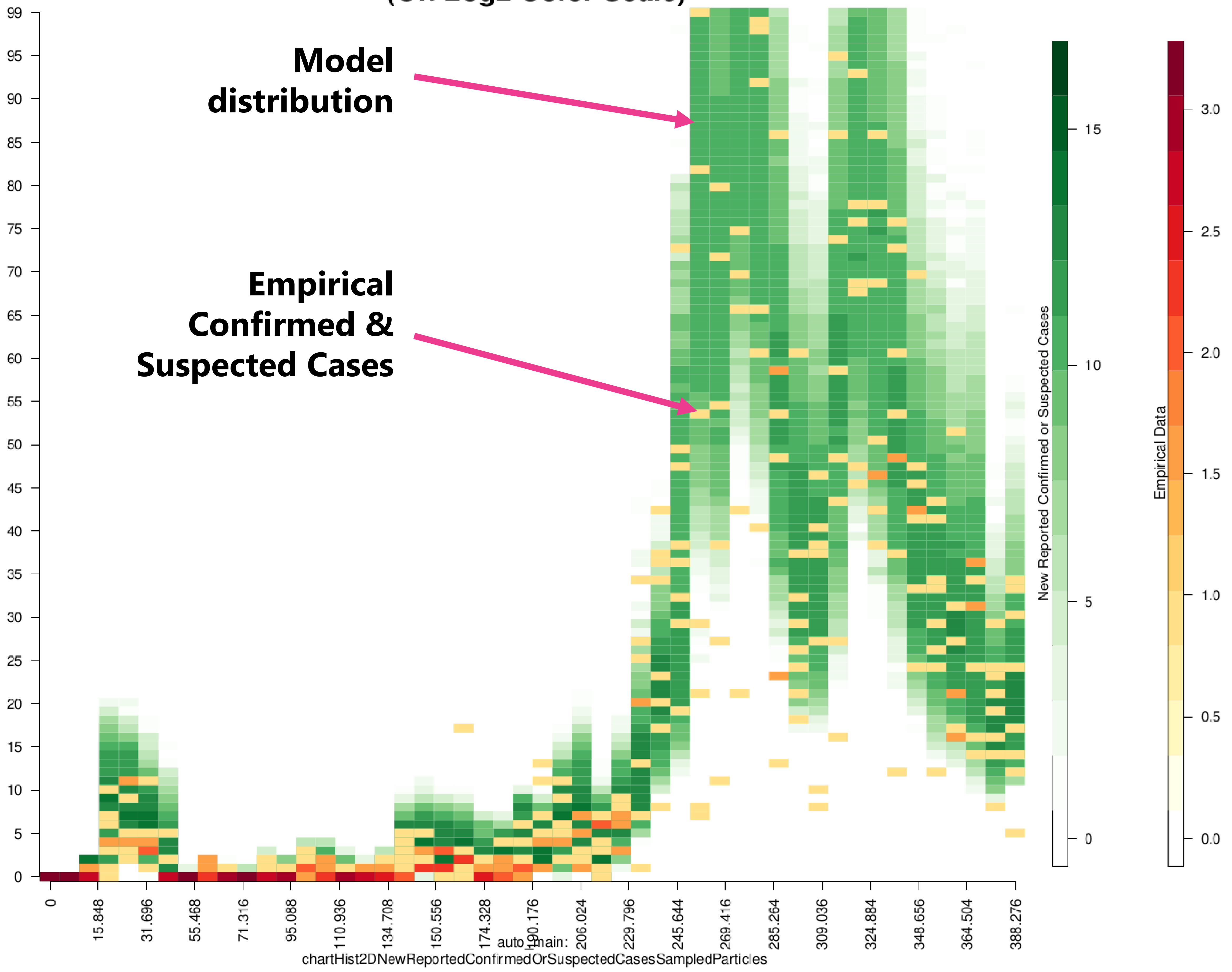
Transmission Models



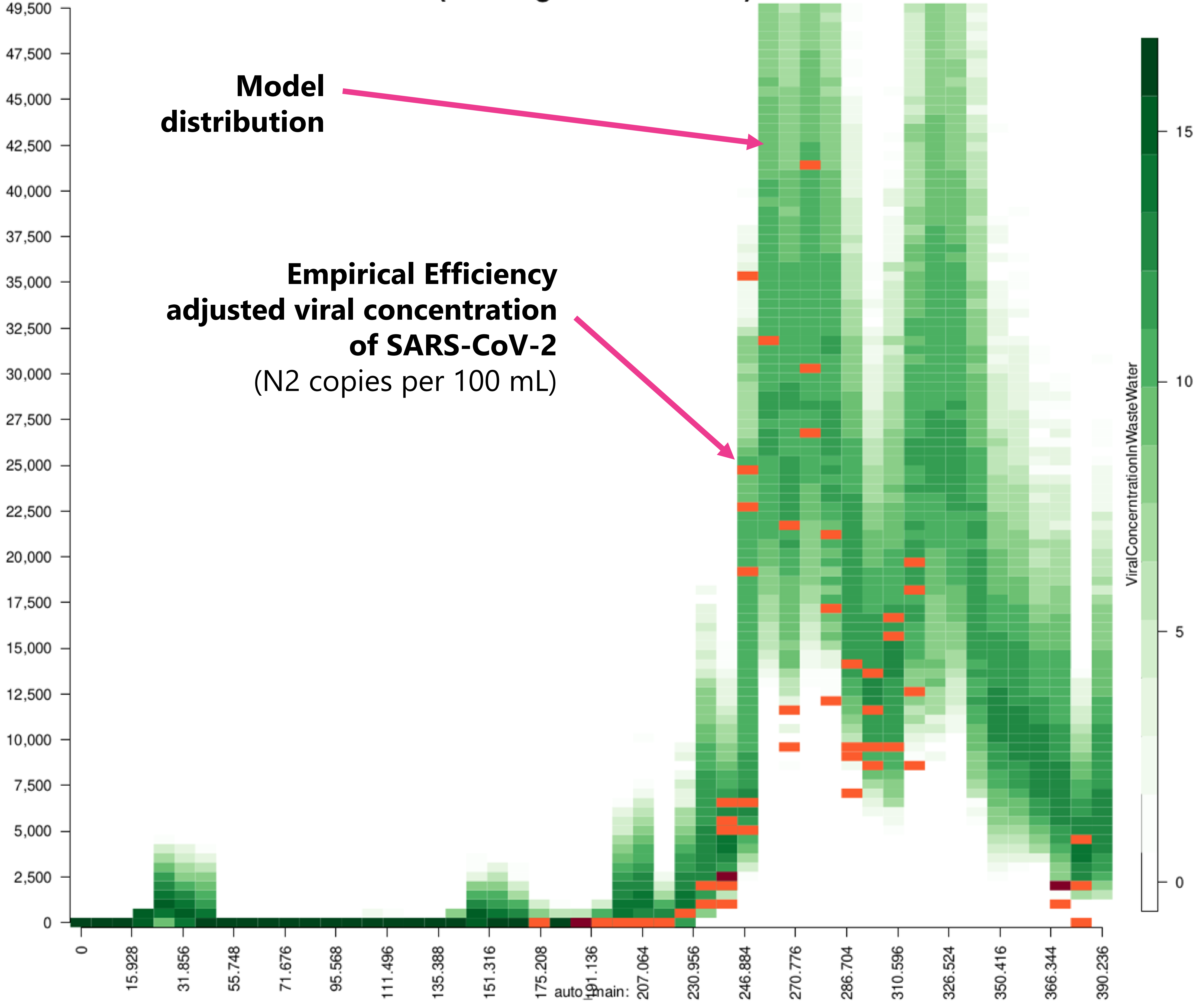
Uses of the Particle Filter Model & PMCMC Models

- **Population tomography**: Providing a consensus probabilistic interpretation (via a joint distribution) the situation now and in the past
- **Projection/Forecasting**: Projection forward from now with model dynamics and "status quo" or diffusive assumptions concerning active testing, contact patterns, etc.
- **Backcasting**: Historical reconstruction based on earlier & later data
- **Policy evaluation**: Evaluation of intervention portfolios, exogenous scenarios or other "what if" possibilities using a consistent picture from the latest evidence

New Reported Confirmed or Suspected Cases
(On Log2 Color Scale)

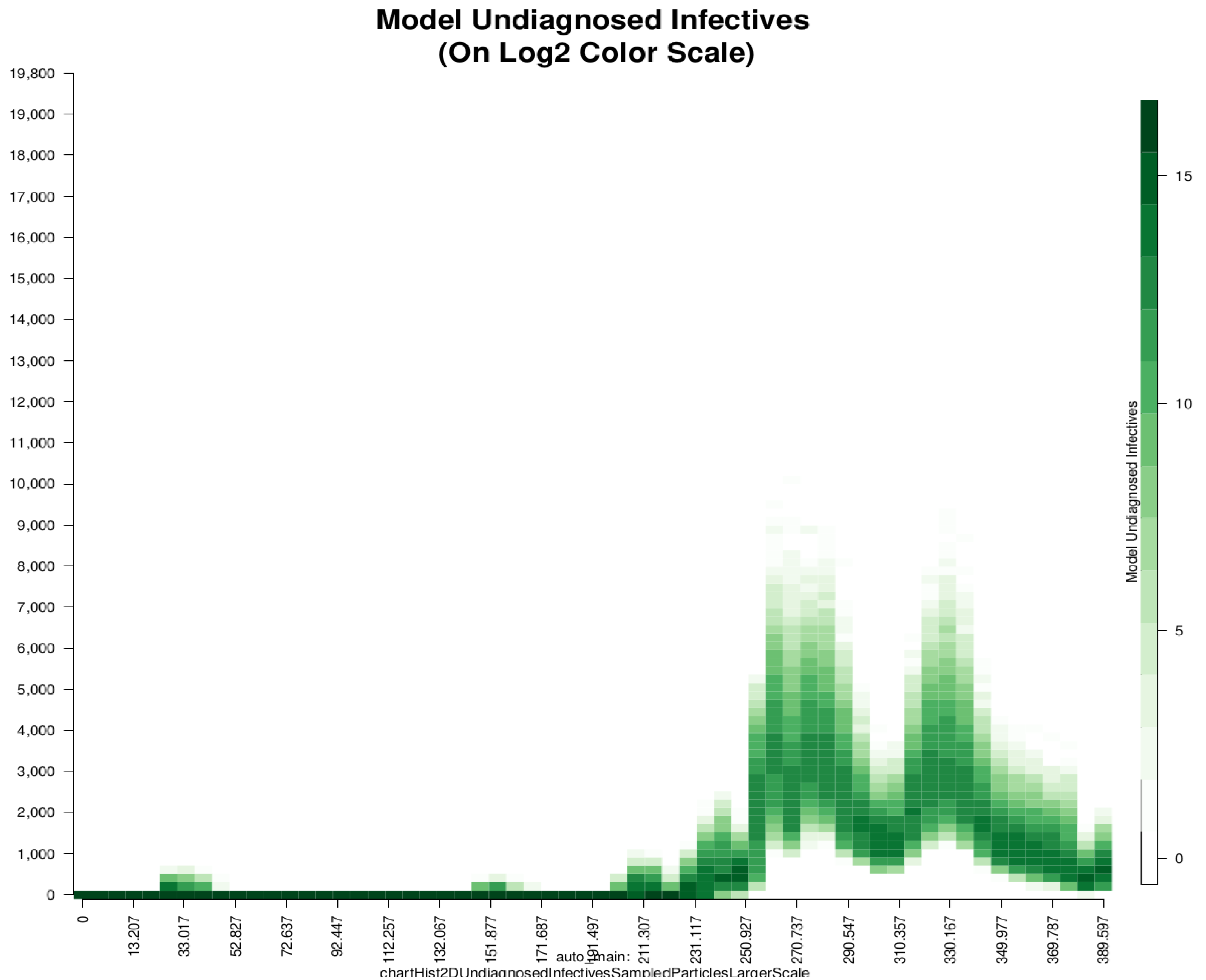


Wastewater Concentration Data Time Series



Generative View of Unmeasured Quantities

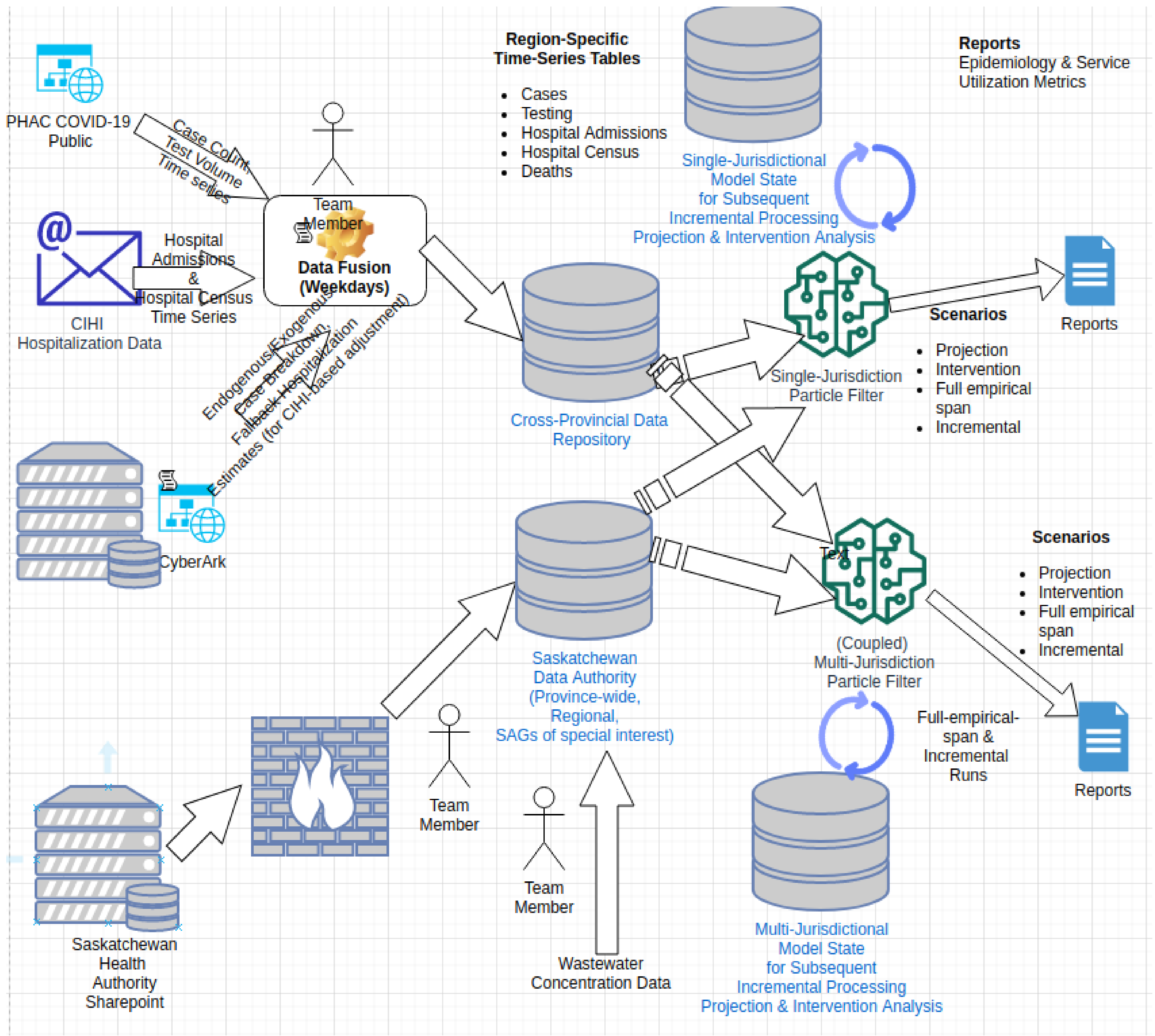
Latent state:
Undiagnosed infectives



Advantages

- Obtain unified, system-wide picture consistent with natural history of infection from diverse data sources
- Capacity to estimate both underlying epidemiological state & parameters
- Ready ability to layer in support for new data sources
- Tolerance for missing, and various quality levels of data
- Applicability to wide variety of epidemiological models
- Whole greater than the sum of the parts: Capacity to sharpen the estimates for any one infection by considering data from **multiple pathogens with common drivers**
- Viability of scalable real-time model-assisted epidemiological & behavioural surveillance frameworks consuming diverse data, including for multiple pathogens

Industrial Strength Framework Deployed during Pandemic



Conclusions

- Models gain much additional value if used for service provision with ongoing updates
- Combining Bayesian AI/machine learning algorithms theory based models allows for “always updated” models to understand current situation & project forward
- Such methods are synergistic with large-scale data collection using high-velocity versions of traditional (e.g., testing) and novel (e.g., wastewater) information
- In the presence of aggregate dynamic models, particle filtering for COVID-19 can perform well both at the national, regional and local levels
- Such methods supports integration of diverse time series, including WW, SM & Sear
- With contemporary parallel & distributed computing, daily updating is readily possible
- Appropriate reporting pipelines can allow for scalable, efficient data ingestion & reporting and interactive exploration to inform decision making
- Whole is greater than sum of the parts: Early work suggests that use of data from multiple pathogens with common risk factors sharpens analysis of any one risk factor

Discussion Period

Any questions?

Please use the **Q&A tab** to submit your questions for our speakers. You can “**like**” other people’s questions to push them up in priority.

Closing Remarks

Thank You!

Join us on Tuesday, January 30, 2024 (1:00-2:00pm ET) for the next seminar!

Please complete our **survey** that will be shared shortly after the seminar. Scan the QR code.

Seminar recording and presentation slides will be posted on https://nccid.ca/ within two weeks.

Visit https://nccid.ca/surveillance-advances-seminar-series/ for more information about the Surveillance Advances seminar series.