COMPUTING FOR HEALTH

In silico approaches for health sciences

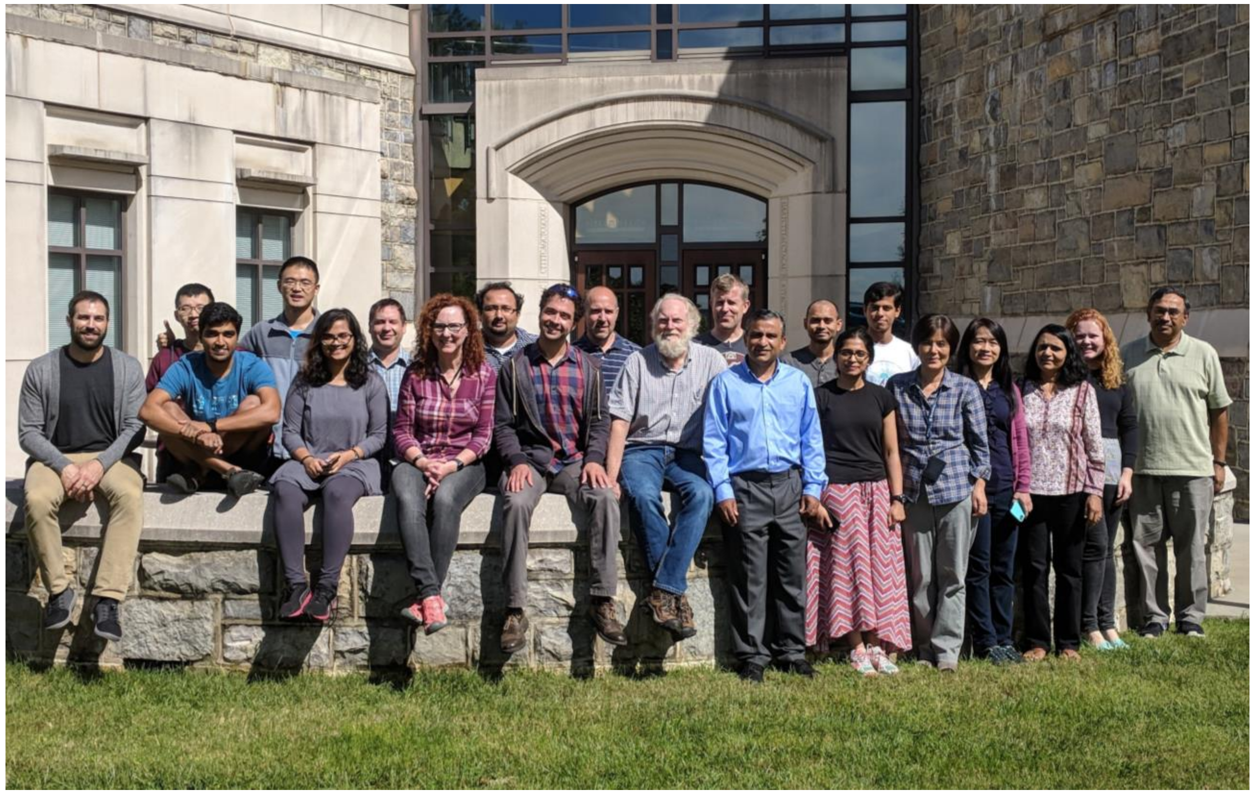
SRINI VENKATRAMANAN

02 JANUARY 2019

University of Virginia | Biocomplexity Institute & Initiative

ACKNOWLEDGMENTS

- Staff and Students of Network Systems Science & Advanced Computing Division (BII, UVA)
- Biocomplexity Institute of Virginia Tech
- Funding agencies, external collaborators and industry partners



University of Virginia VIRGINIA TECH

National Institutes of Health

USAID

CENTERS FOR DISEASE CONTROL AND PREVENTION

AccuWeather

OUTLINE

- INTRODUCTION TO DIGITAL HEALTH
- CASE STUDIES IN COMPUTATIONAL MODELING
- LESSONS (BEING) LEARNT

INTRODUCTI
ON

A BRIEF TIMELINE OF HEALTH/MEDICINE

Before Computing

Invention of the Microscope (1590)

William Harvey - the circulatory system (1628)

Edward Jenner - small pox vaccine (1796)

John Snow - Investigation of London cholera epidemic (1855)

Louis Pasteur - germ theory of disease (1870)

Wilhelm Röntgen - Xray for medical imaging (1895)

Willem Einthoven - electrocardiography (1903)

Alexander Fleming - Penicillin (1928)

Rune Elmqvist - Implantable pacemaker (1958)

Raymond Damadian - Magnetic Resonance Imaging (1971)

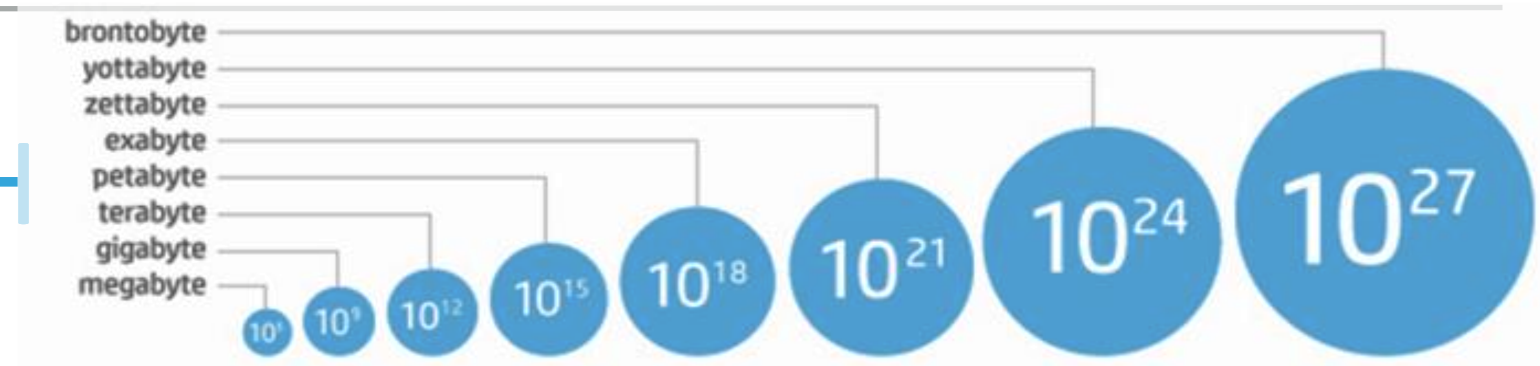
Hood and Smith - Automated DNA sequencing (1985)

Jacques Marescaux - First telesurgery (2001)

FDA approves AI-based diagnostic of diabetic retinopathy (2018)

After Digitization

BIG DATA IN HEALTH



- Genomics: Doubling every 7 months. Expected to produce 2 to 40 exabytes per year by 2025
  - About 1 billion people expected to have genome sequenced by 2025. [1]
- EMR/EHR data: Upto 2 zettabytes per year by 2020. 95% penetration in U.S. hospitals [4]
- Wearable sensors: Expected 1 billion connected wearables by 2020 [5]
- Social data: 5.2 billion daily Google searches in 2017. 1.32 billion daily active users in Facebook (June 2017) [2]
- Internet of Things: Upto 20 billion devices by 2020. (12 billion consumer IoT) [3]
- Remote sensing: NASA EOSDIS manages about 10 petabytes of data, adding 6.4TB every day. [6]

A CONFLUENCE OF JARGONS

- Big Data: Artificial Intelligence, Data Science, Predictive Analytics
- Health: Precision Medicine, Telemedicine, OneHealth

CHALLENGES

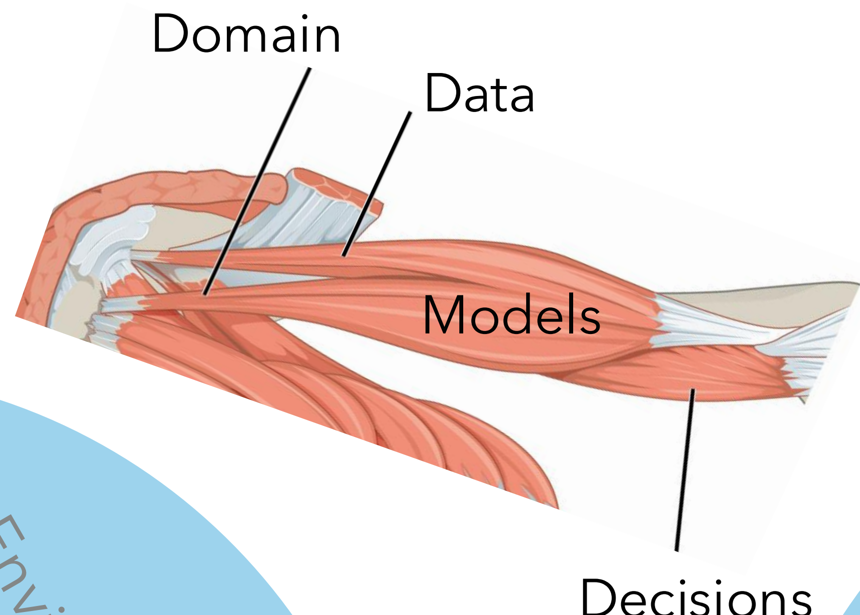
5V's: Volume, Velocity, Veracity, Variety, Value

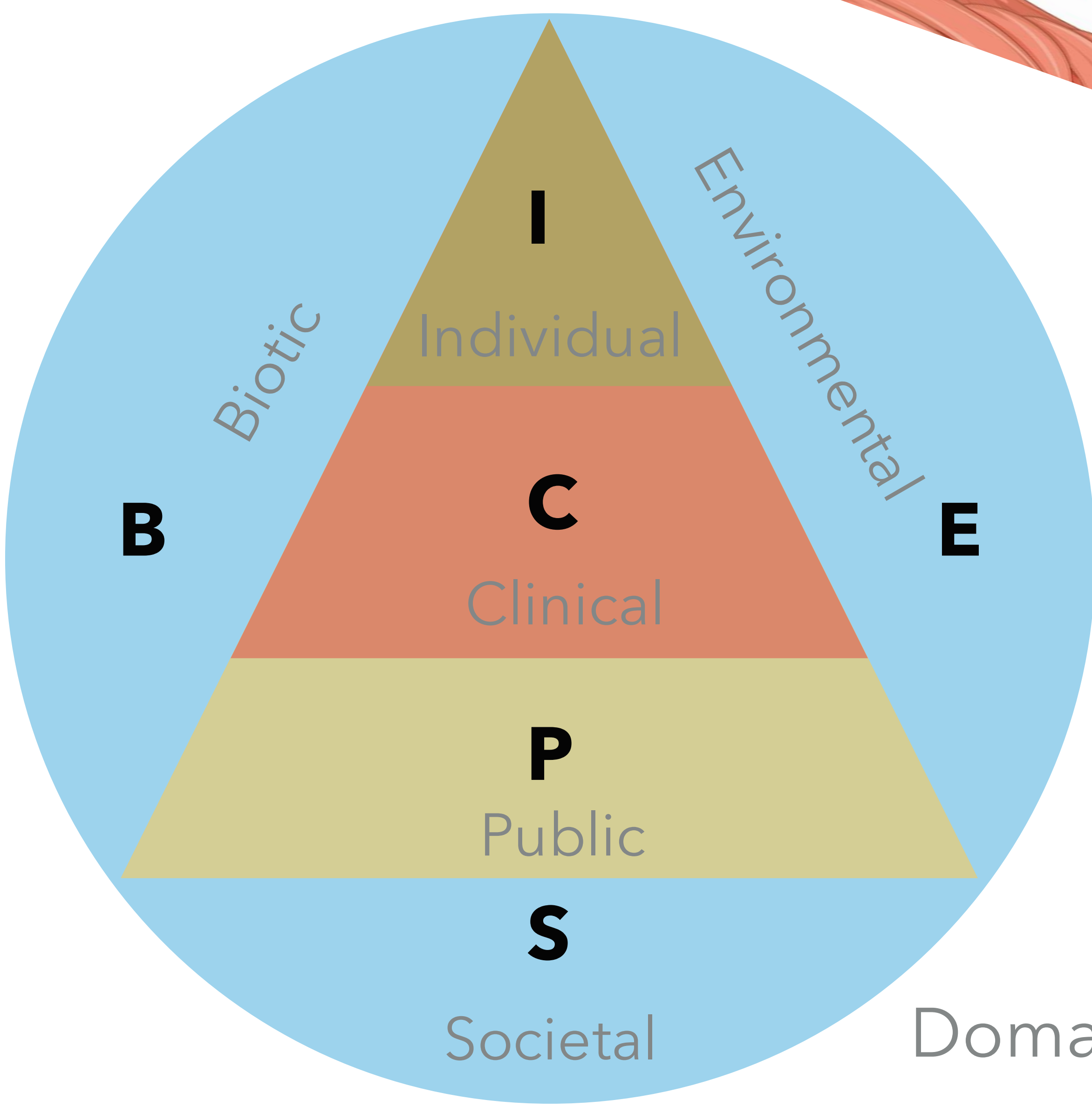
- How to handle the amount of data being generated, both in terms of magnitude and rate of change?
- How to verify and track the provenance of different datasets?
- How to integrate the varied datasets into a coherent view of the system?
- How to use the data meaningfully to guide decision-making?

especially for digital health

CEPS MODEL FOR DIGITAL HEALTH

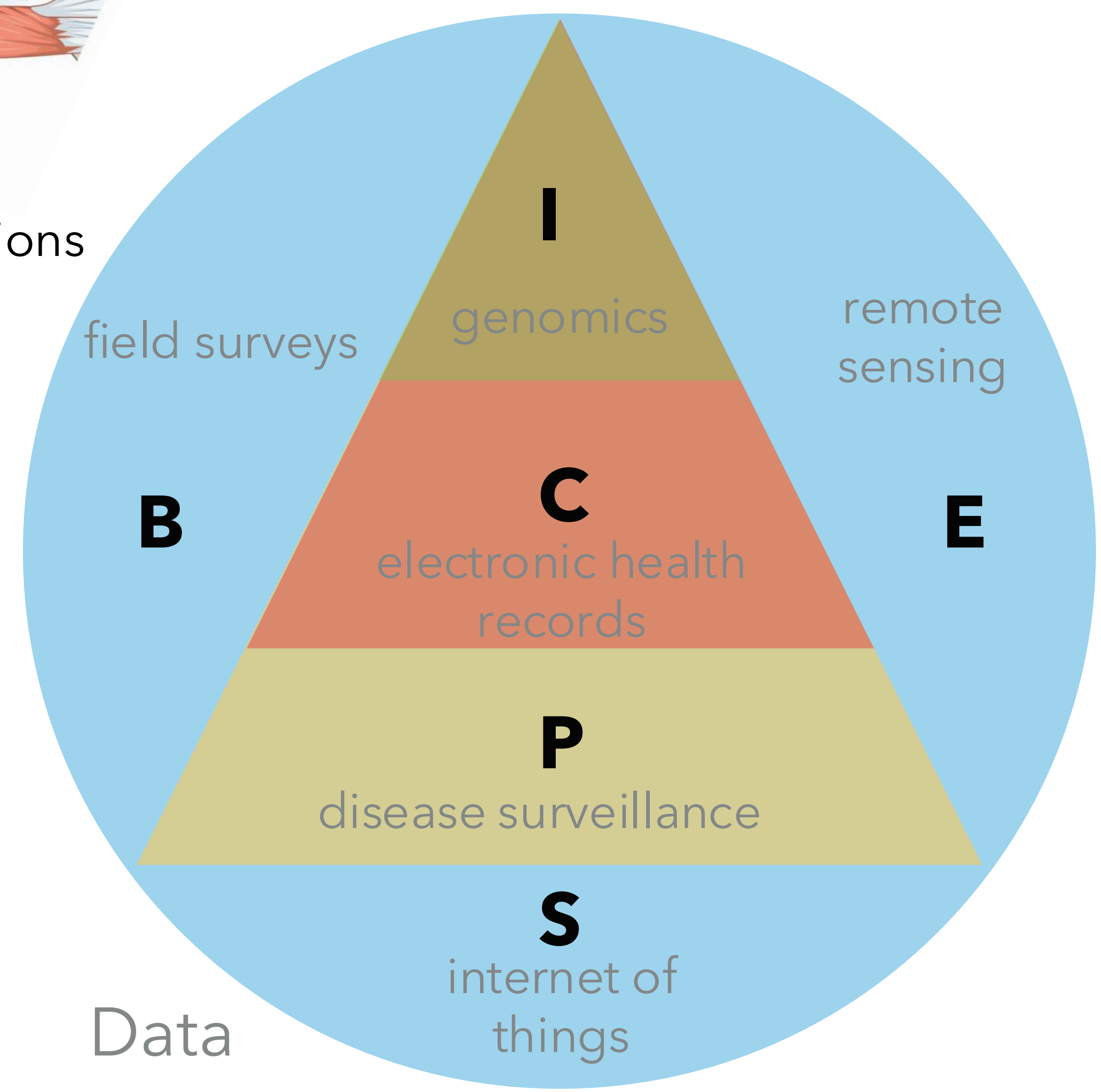
biceps - literally 'two-headed', from *bi-* 'two' + *ceps* (from *caput* 'head')





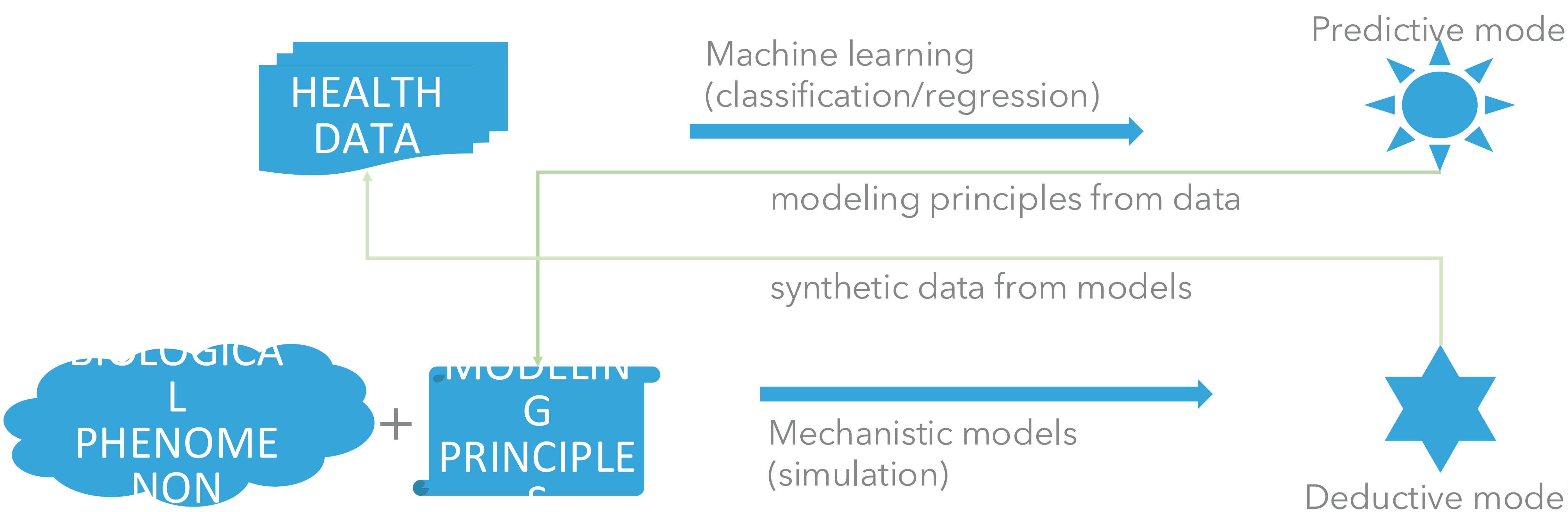
Domain

Data



COMPUTATIONAL MODELS

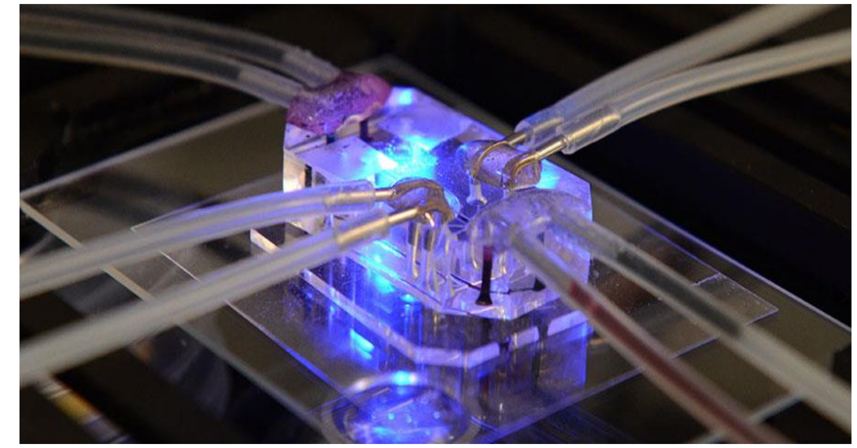
Machine learning models	Mechanistic models
Establishing a statistical (correlation) relationship between inputs and outputs	Establishing a mechanistic (causal) relationship between inputs and outputs
Requires large datasets, often homogeneous linked records	Can work with small datasets coming from diverse sources
Can only make predictions that relate to patterns in the training data	Once validated, can be used for prediction where experiments are difficult to perform



Baker, Ruth E., et al. "Mechanistic models versus machine learning, a fight worth fighting for the biological community?." *Biology letters* 14.5 (2018): 20170660.

RECENT KEY TRENDS

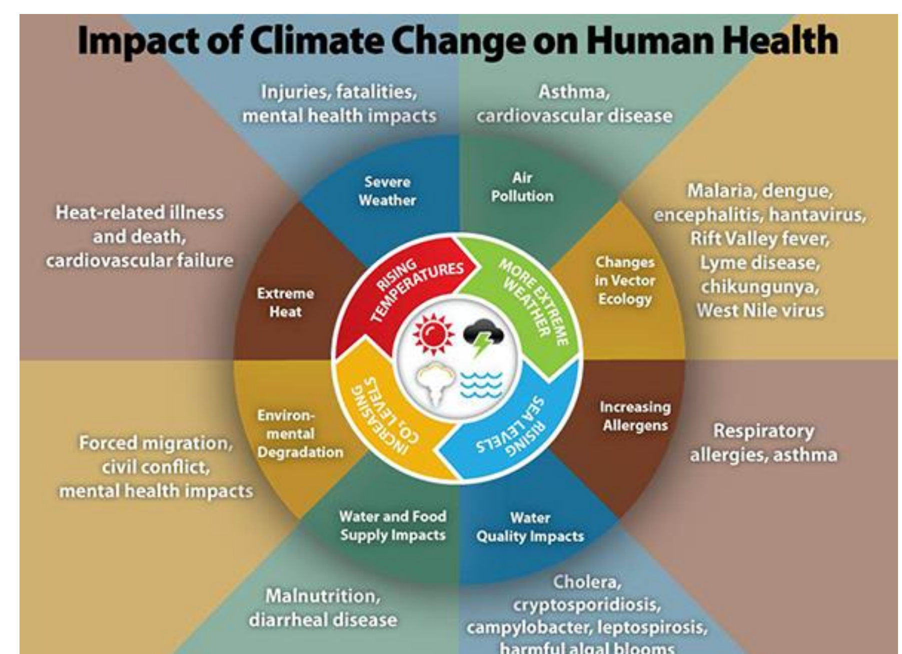
- I - Drug discovery with Tissue Chip
- C - Disease progression modeling, realtime bedside analytics
- P - Randomized trials via simulation, healthcare resource optimization
- B - Zoonotic diseases, antibiotic resistance
- E – Impact of climate change
- S - Health impacts of infrastructure failures



https://ncats.nih.gov/tissuechip

The One Health Triad

http://www.onehealthinitiative.com/



https://www.cdc.gov/climateandhealth/

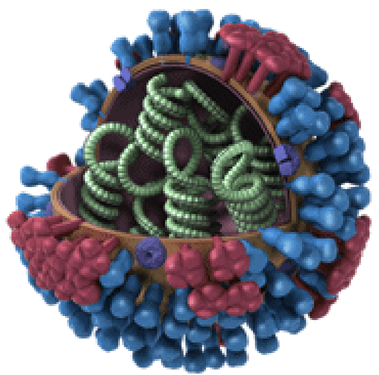
CASE STUDIES

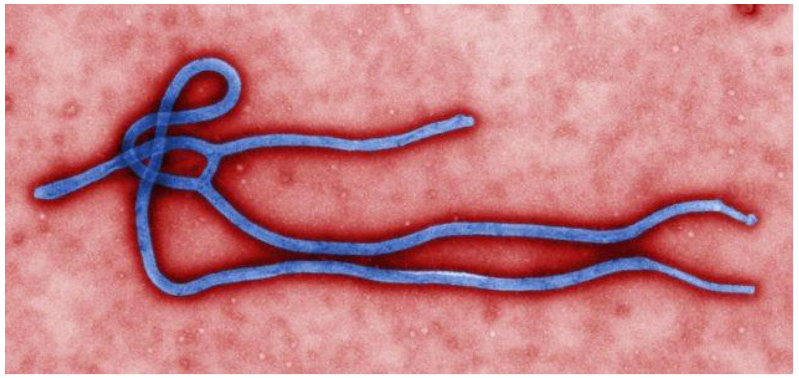
DISEASES WITHOUT BORDERS

(Population Scale)

INFLUENZA & EBOLA: A STUDY IN CONTRAST

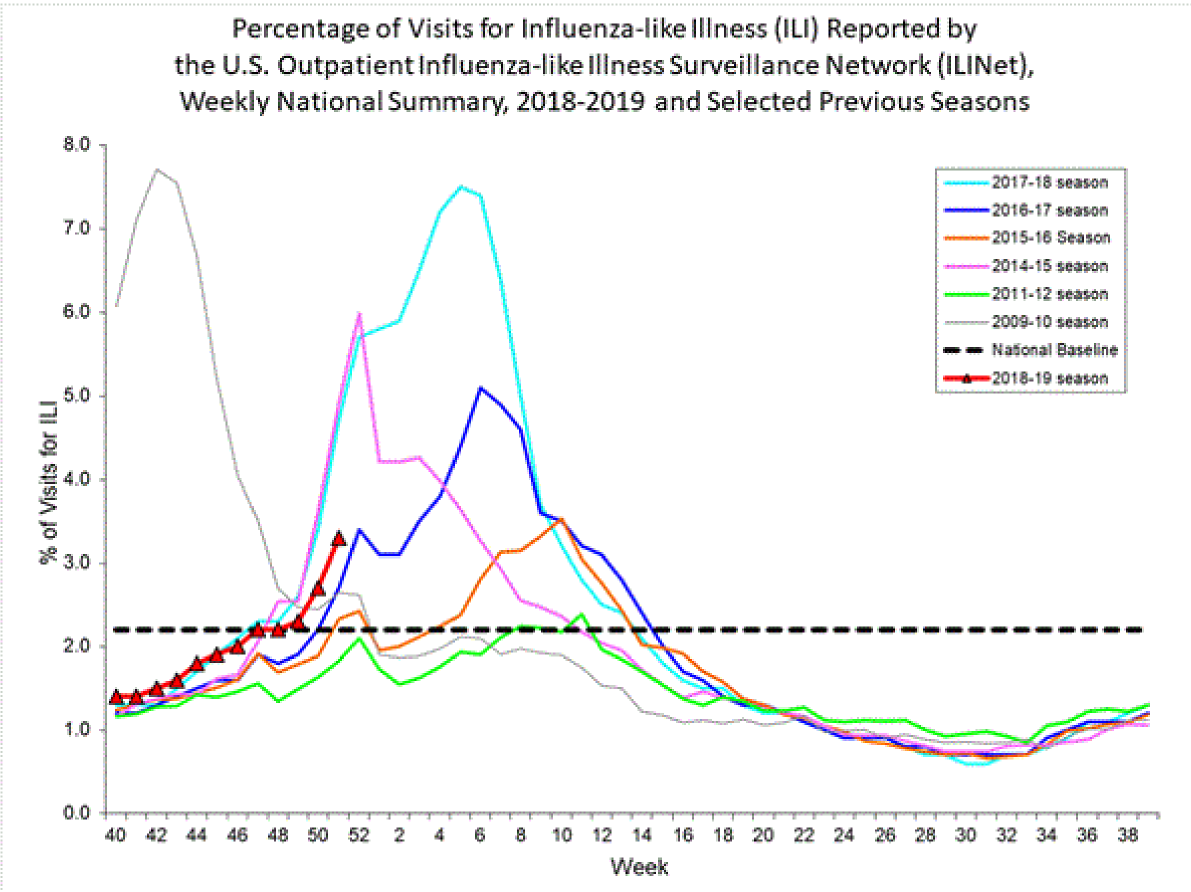
INFLUENZA	EBOLA
Seasonal in nature. Has pandemic potential.	Sporadic outbreaks. Potential international risk.
Hospitalization rate dependent on age, virus strain.	High case fatality rate (60%-90% for ZEBOV)
Annual economic burden in US $27-87 billion	2014-15 outbreak caused $2.2 billion GDP loss, international response costed $3.6 billion
Existing mechanisms for control (vaccines, antivirals, strategic national stockpile)	Vaccines being tested for the first time in DRC. Safe burials and contact tracing key to control.
Global presence. High data availability (resolution, dimensions)	Mostly in sub-Saharan Africa. Limited surveillance.



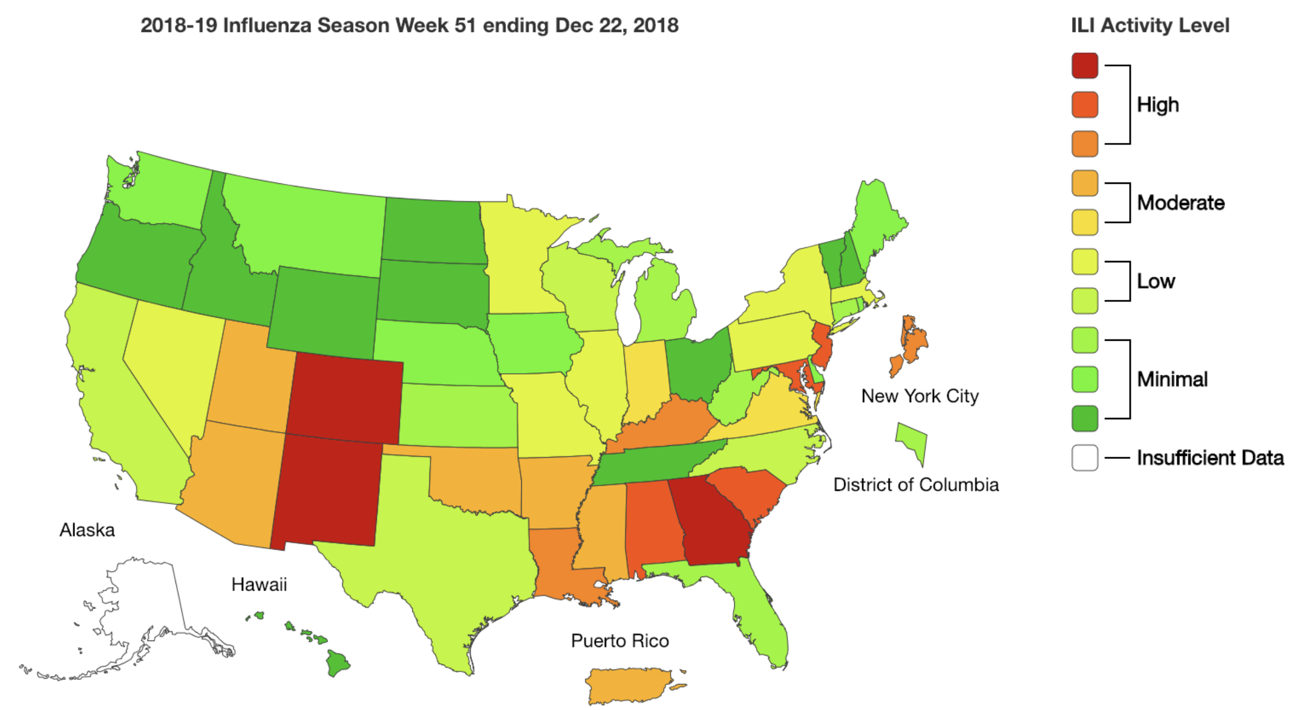


INFLUENZA SURVEILLANCE IN THE UNITED STATES

Weighted ILI%



Influenza Activity Indicator



https://www.cdc.gov/flu/weekly/index.htm

sickweather
flu near you

Participatory

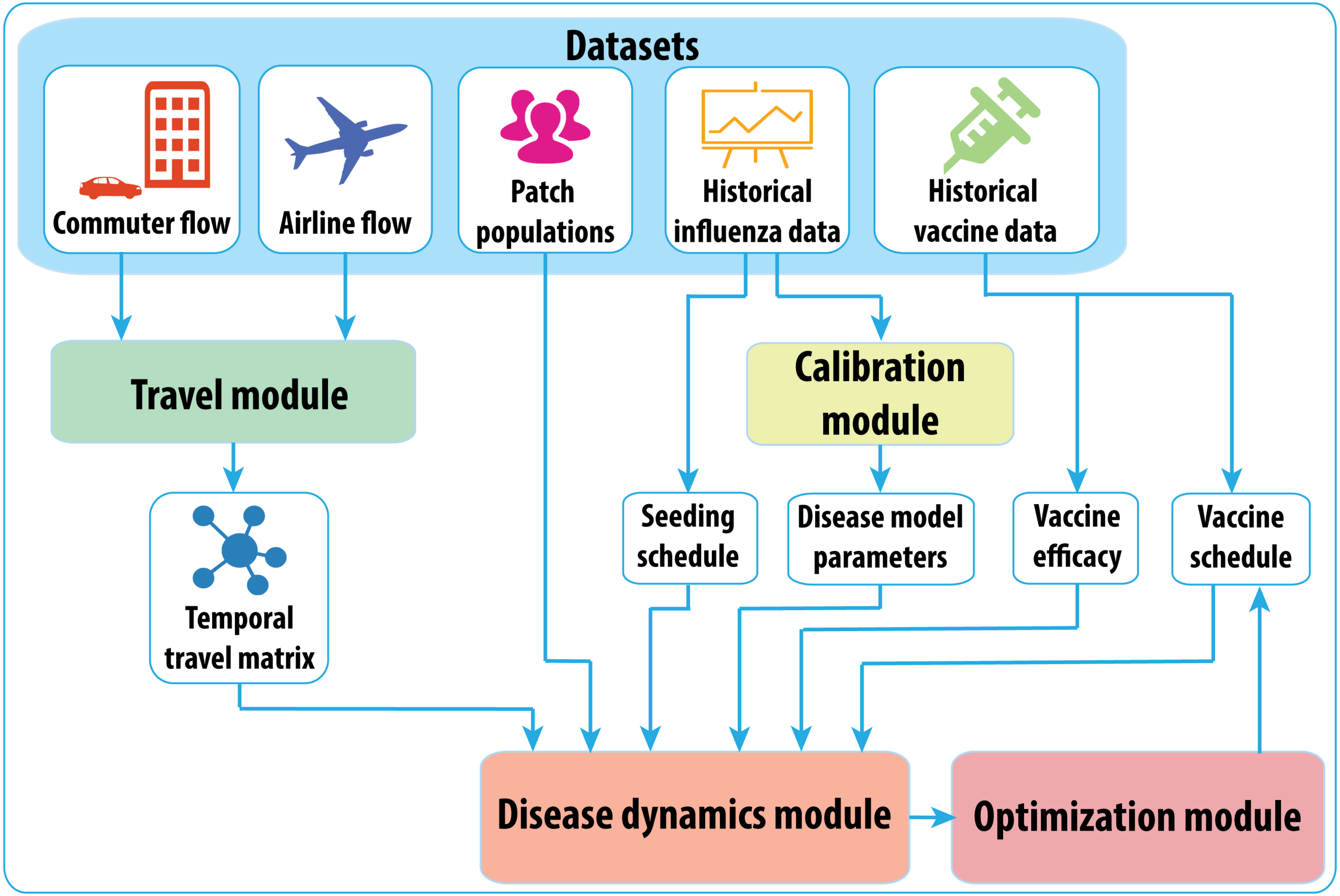
Google
WebMD

Search-based

#FLU
HealthMap
Global Health, Local Information

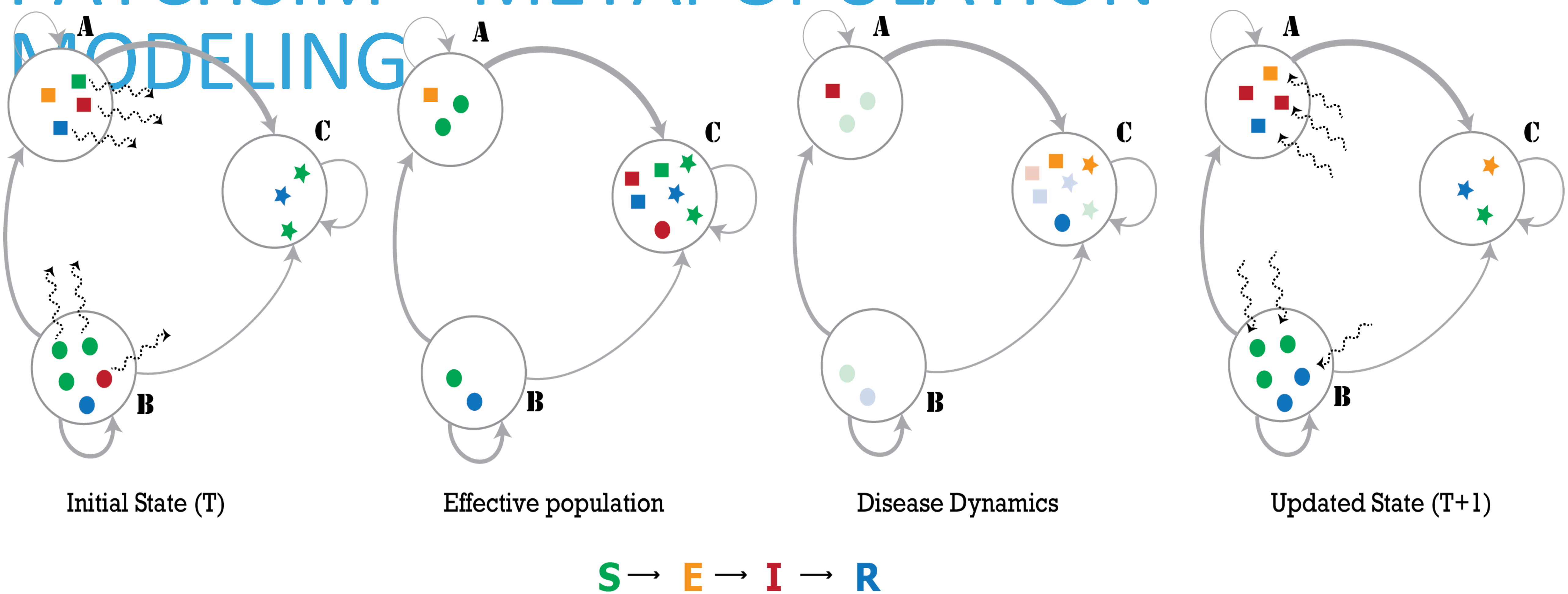
Social/News media

NATIONAL INFLUENZA MODEL



Venkatramanan, Srinivasan, et al. "Spatio-temporal optimization of seasonal vaccination using a metapopulation model of influenza." *Healthcare Informatics (ICHI), 2017 IEEE International Conference on*. IEEE, 2017.

PATCHSIM – METAPOPULATION MODELING



The large circles represent the patches (A,B,C) in the simulation, and the grey edges represent the travel network, colors represent the disease state the individuals are in (susceptible, exposed, infected or recovered), The wavy dashed arrows show movement of individuals.

Panel 1: Individuals move moved from home to another patch.
Panel 2: This creates another effective population for the disease dynamics
Panel 3: Nodes update their respective disease states based on dynamics. Panel 4: Individuals return to their home patch

https://github.com/srinivvenkat/PatchSim

ACCUWEATHER PARTNERSHIP

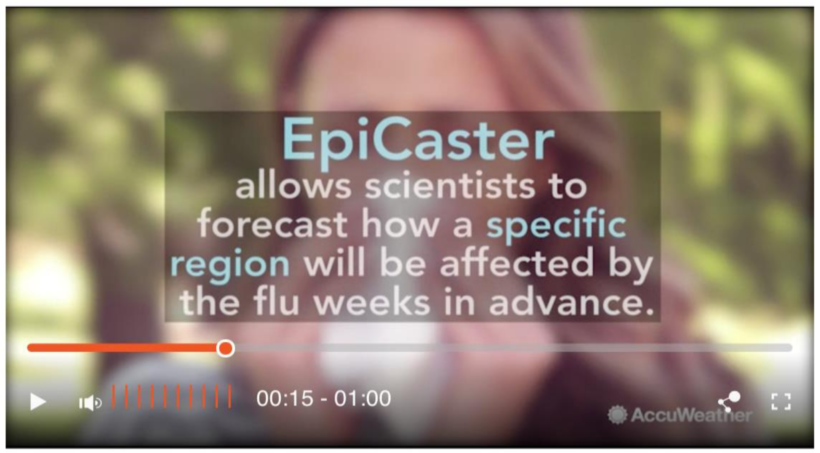
Innovative partnership will extend cutting-edge flu forecasting research to millions

December 1, 2017

A new partnership between the Biocomplexity Institute of Virginia Tech and AccuWeather could dramatically change the level of flu knowledge and awareness for people across the U.S.

AccuWeather to offer users innovative flu forecasts from Virginia Tech's Biocomplexity Institute

By Ashley Williams, AccuWeather staff writer
December 07, 2017, 9:36:12 AM EST



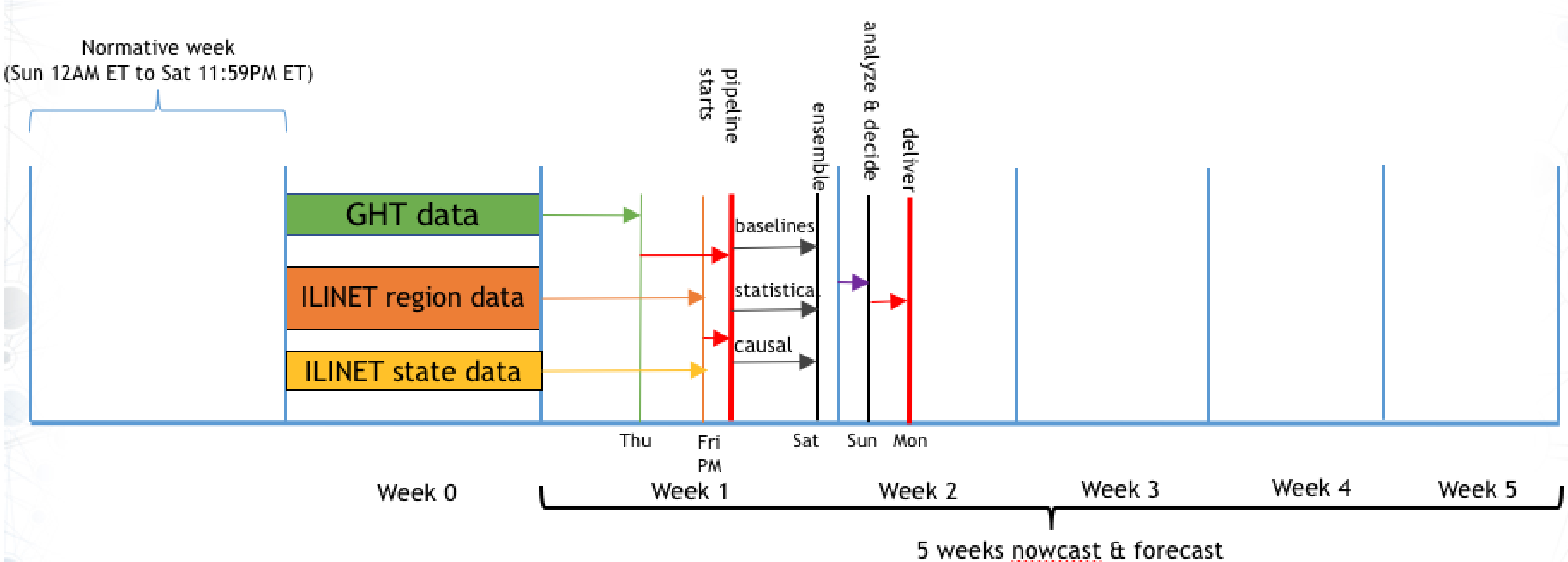
AccuWeather | UNIVERSITY of VIRGINIA | Biocomplexity Institute & Initiative

HOLIDAYS & THE FLU: EARLY FORECASTS TO PREPARE YOUR BUSINESS

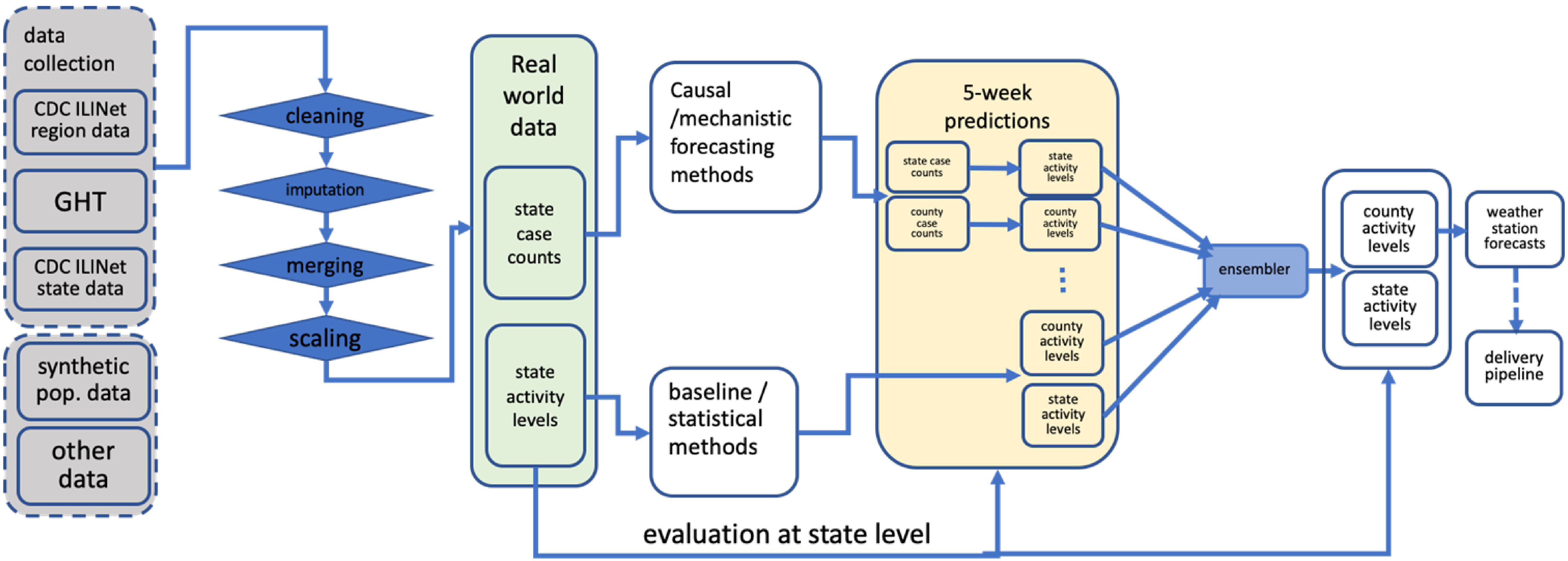
AccuWeather Flu Forecasts | UVA - AccuWeather

Webinar at 2:00 PM EST | December 4, 2018

Jiangzhuo Chen, Srini Venkatramanan, Bryan Lewis and Madhav Marathe on behalf of Network Systems Science & Advanced Computing Division of Biocomplexity Institute & Initiative, University of Virginia

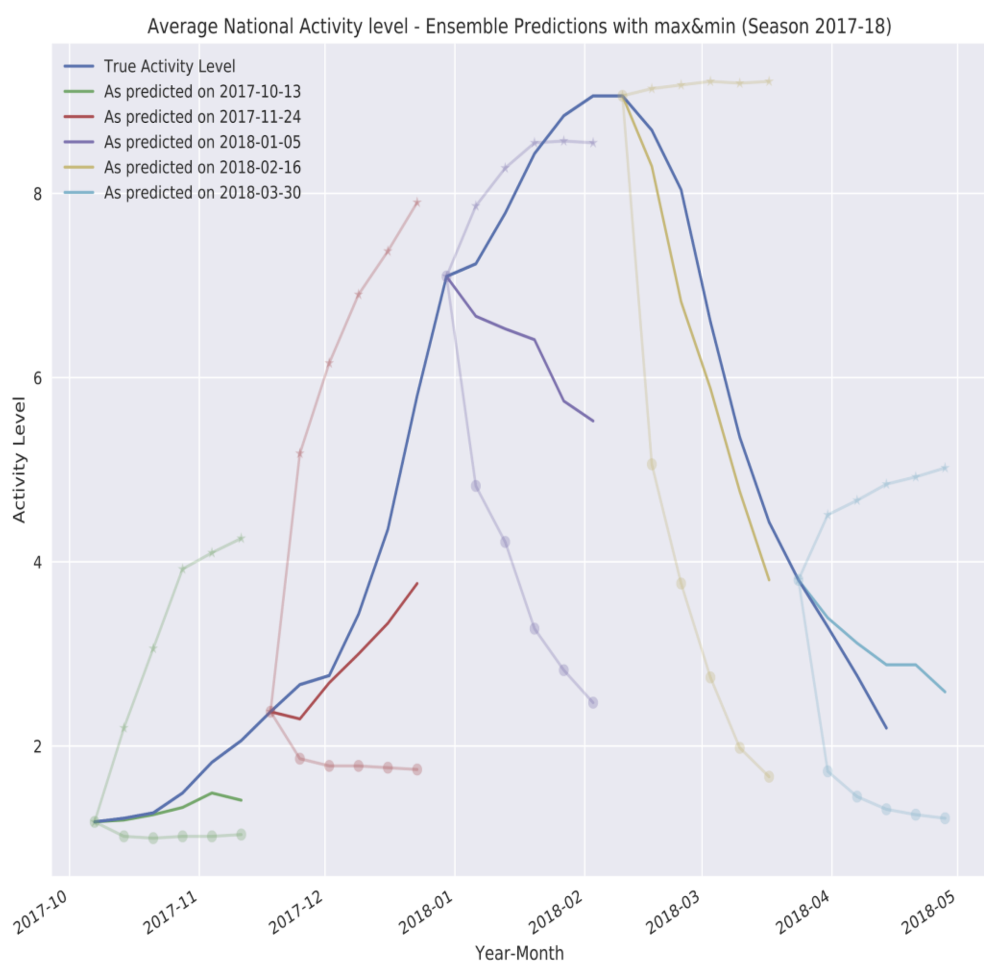


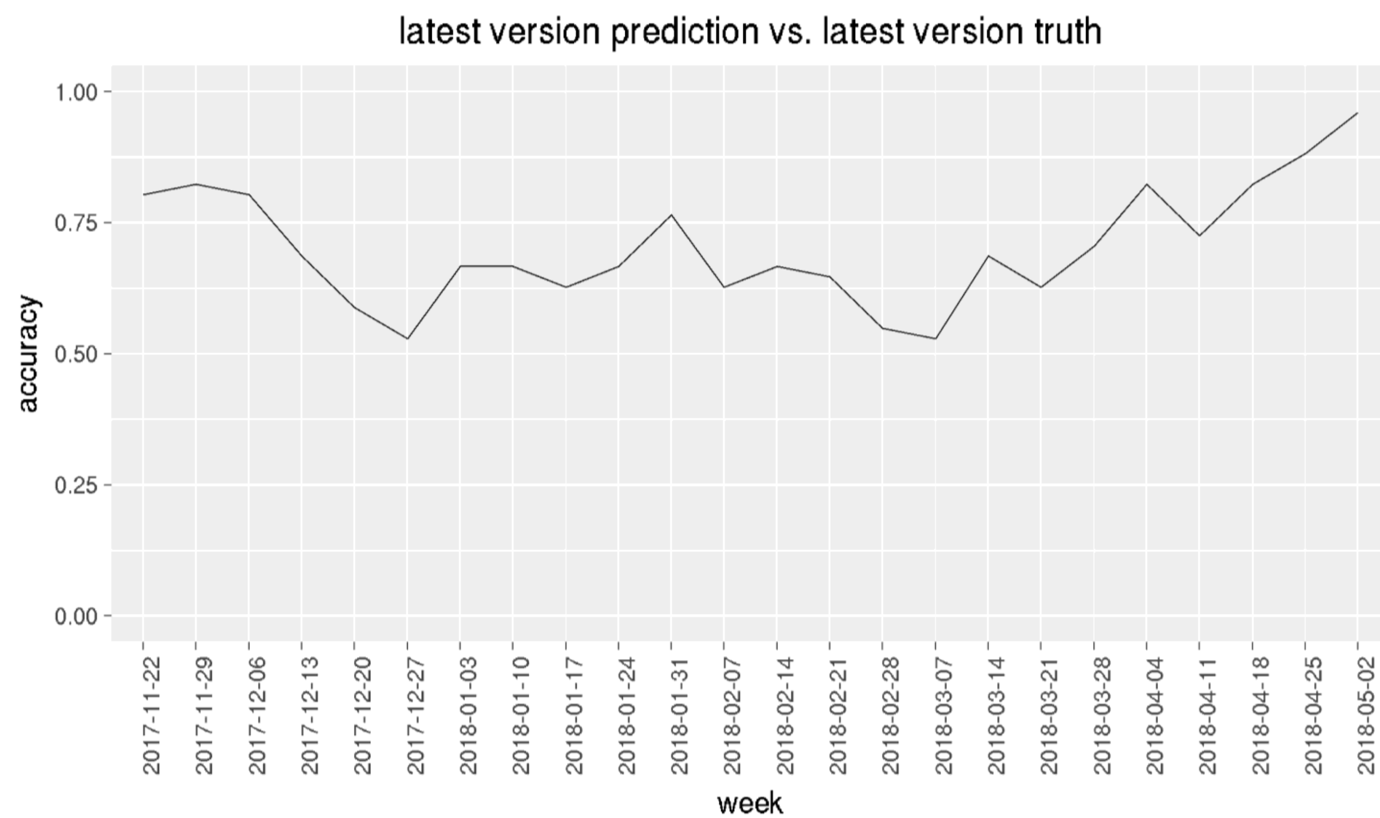
ACCUWEATHER PIPELINE



- **Statistical Methods**
  - **Markov method**: Learning transition probability matrix from a given activity level to the next $P_{h,t}(Y_s(t+1)|Y_s(t))$
  - **Ordinal regression:** Middle-ground between multi-class classification and regression, trained on historical activity levels
  - **Ensembler:** Majority voting across models
- **Mechanistic Modeling**
  - Metapopulation model at the level of counties using commuter/airline data and realistic vaccine uptake information
  - Calibrated using Bayesian particle filtering

FORECAST ACCURACY – 2017-18 SEASON



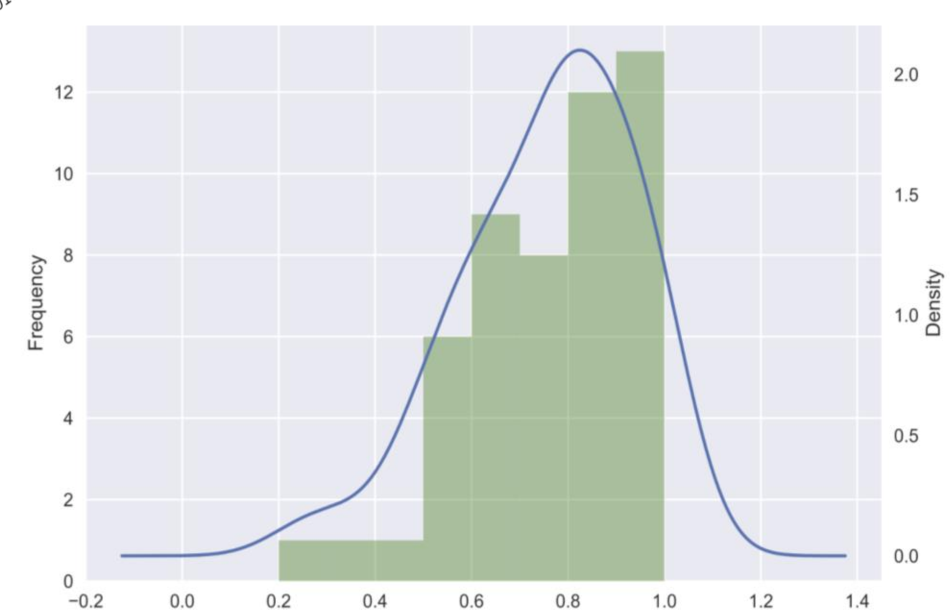


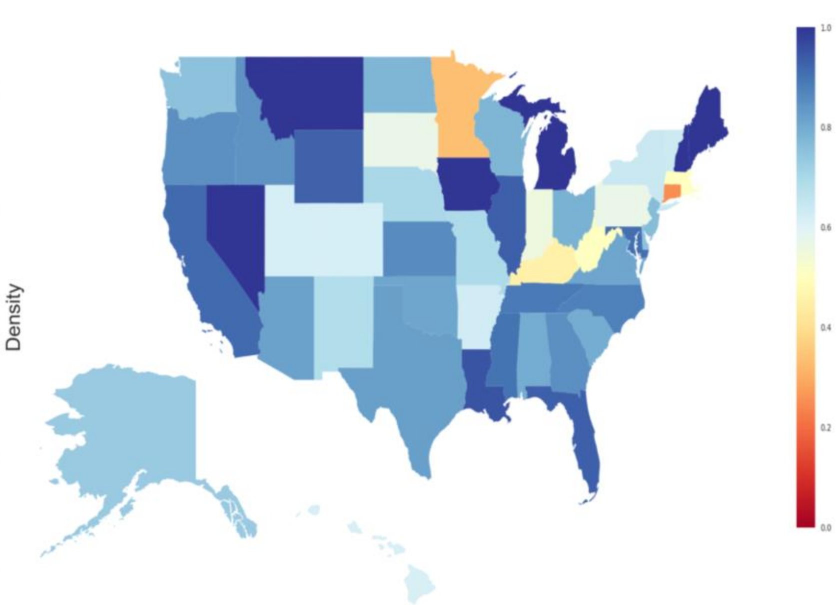
State-level forecast performance (season average)

- Average accuracy of 75% across states and weeks.

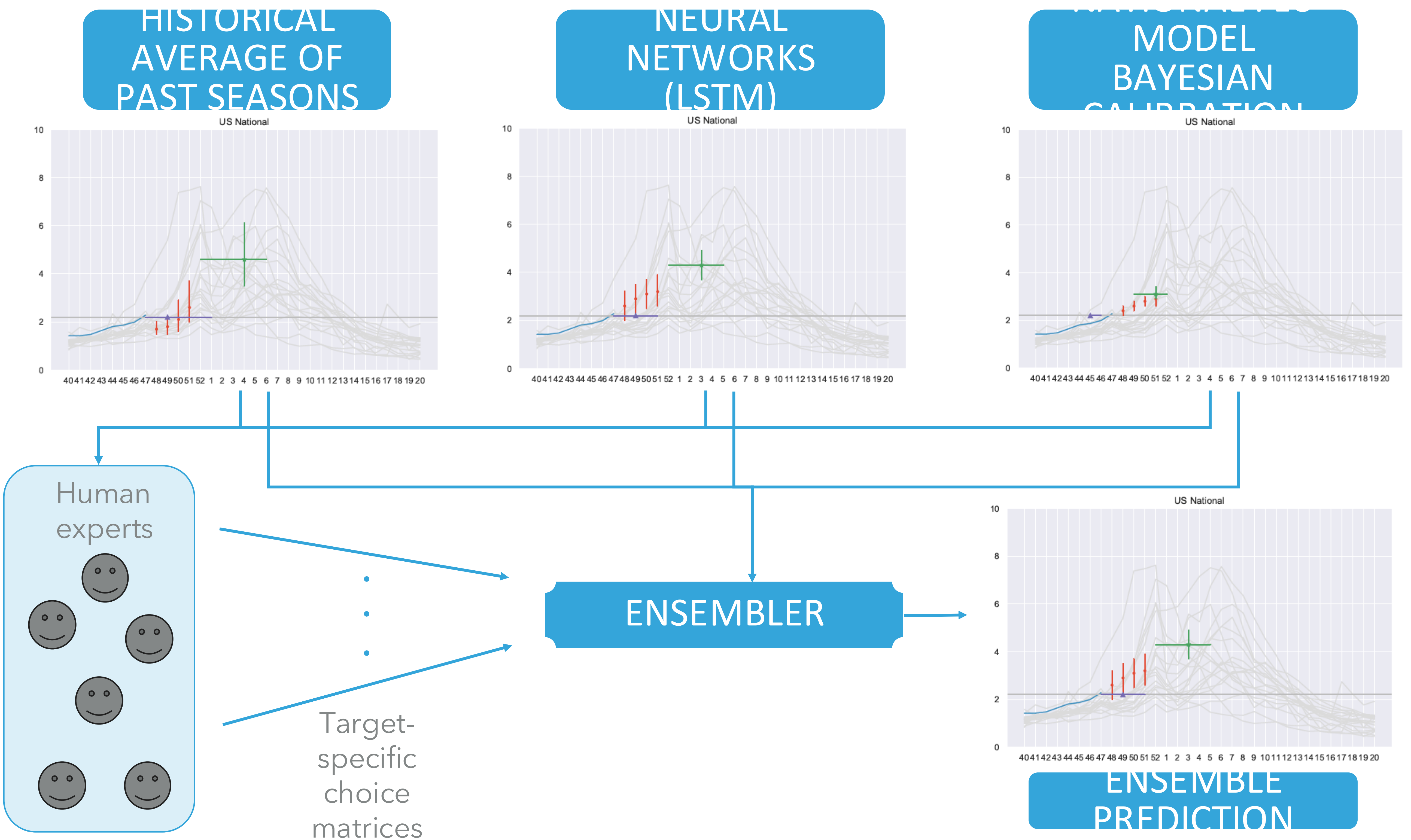
- Performed better in some states than others (better surveillance, more predictable seasons)

What are the upper bounds?



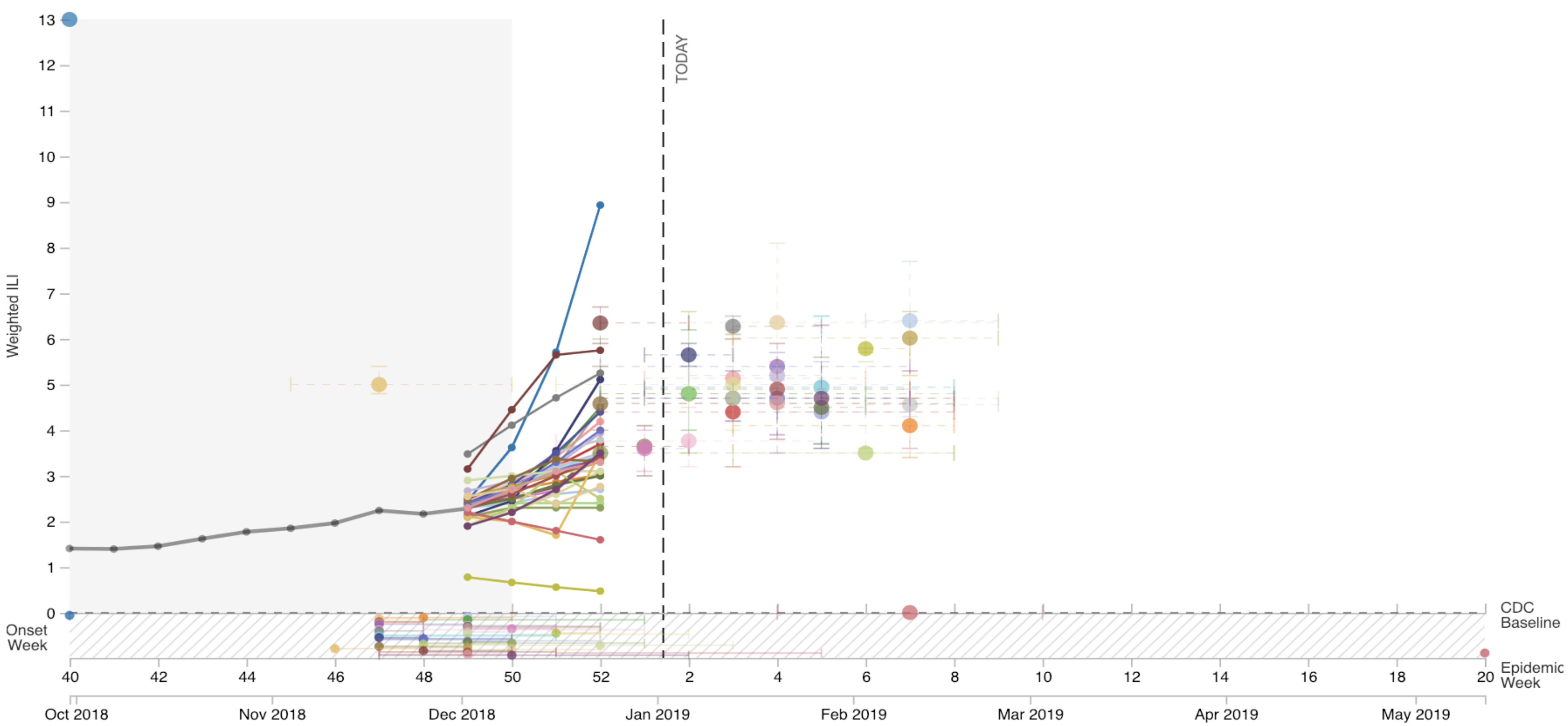


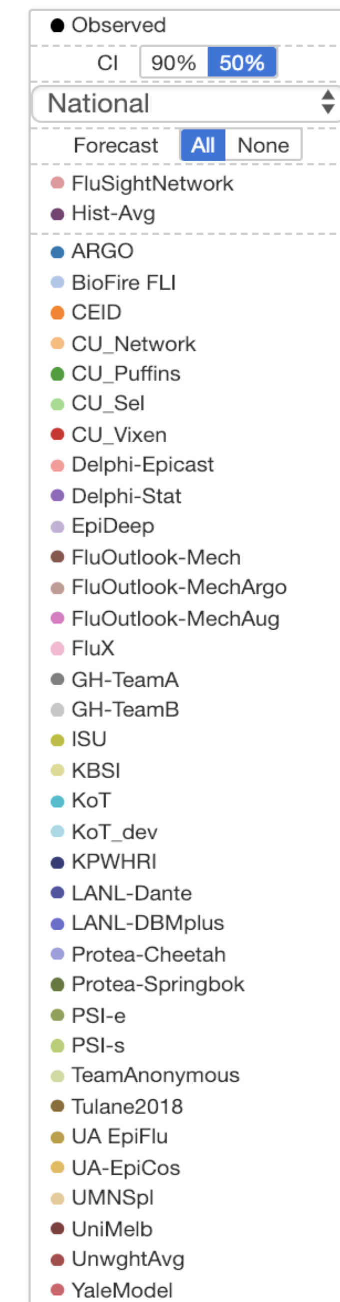
CDC FORECASTING APPROACH



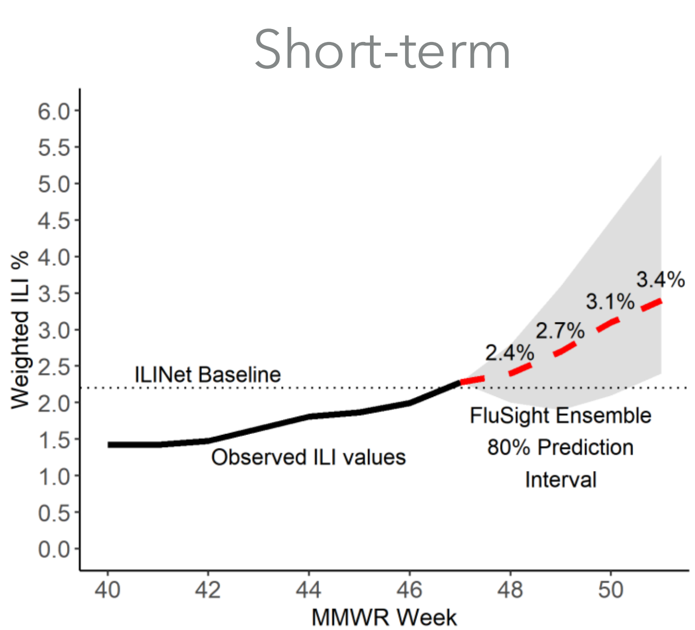
CDC FORECASTING – FLUSIGHT NETWORK

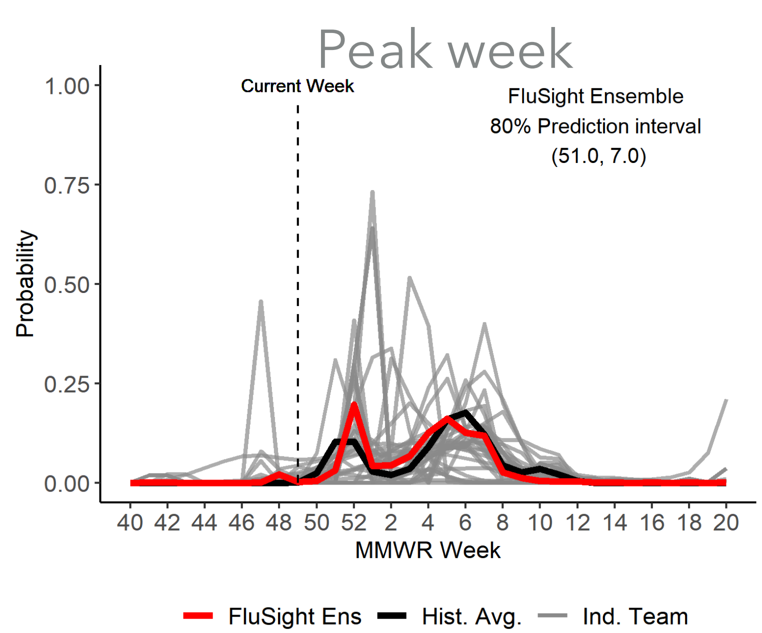
https://predict.cdc.gov/

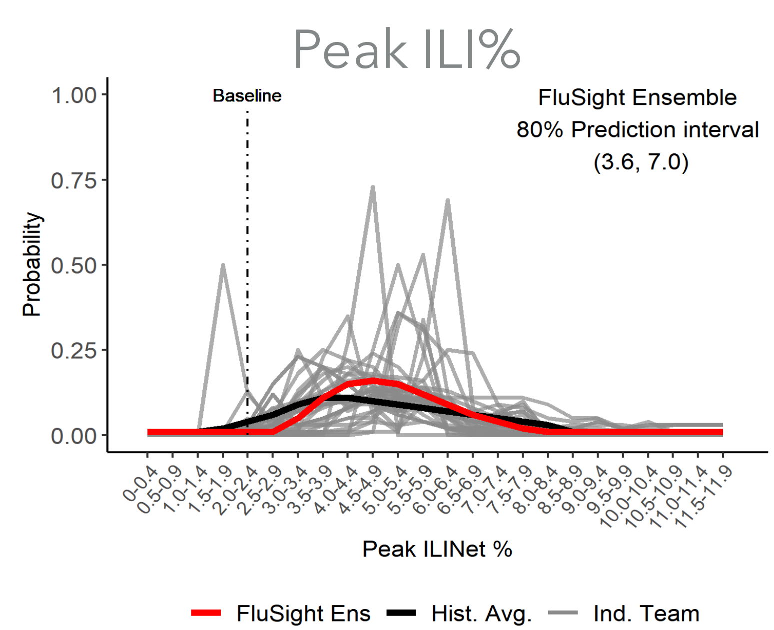




Ensemble predictions

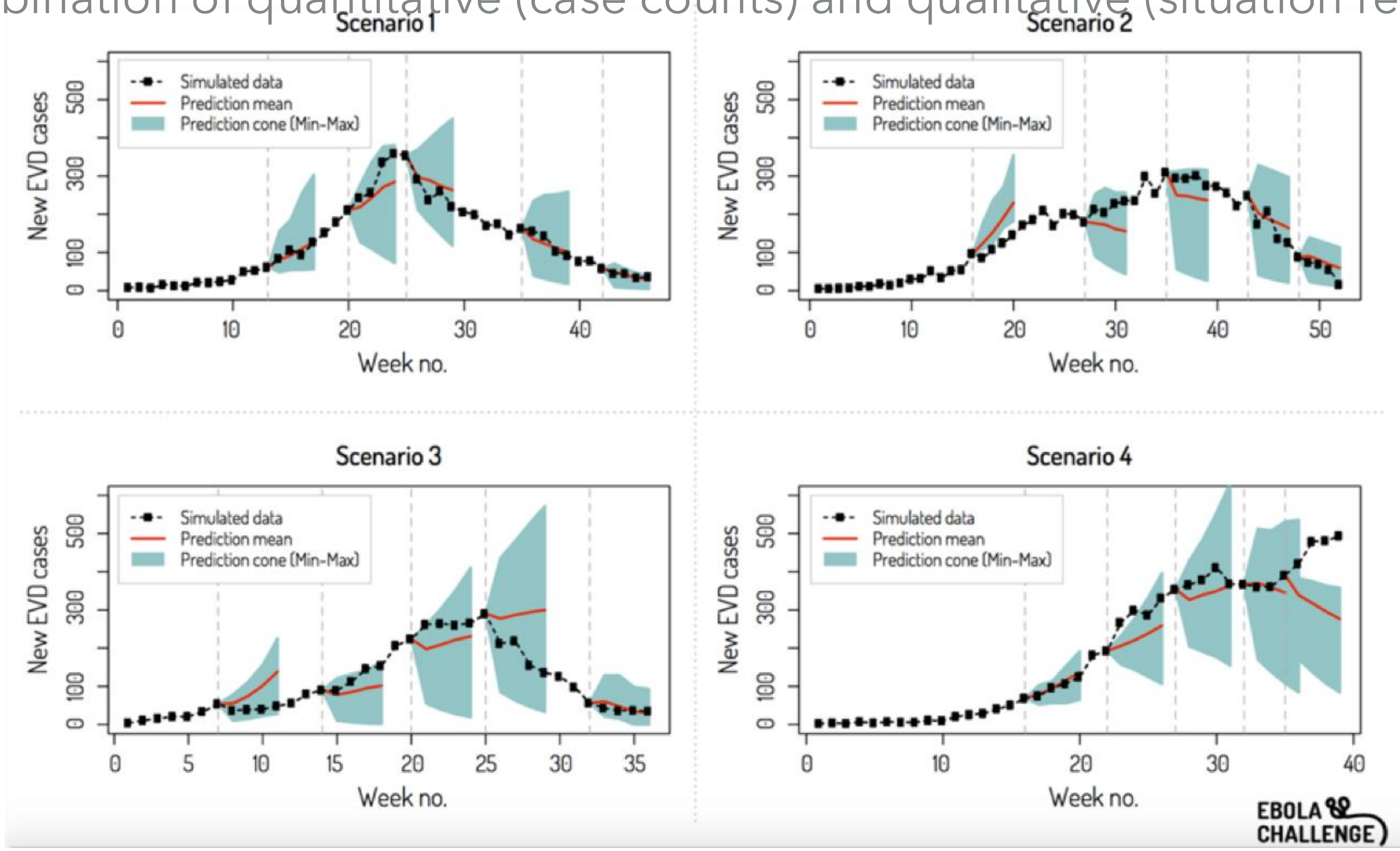






EBOLA FORECASTING CHALLENGE

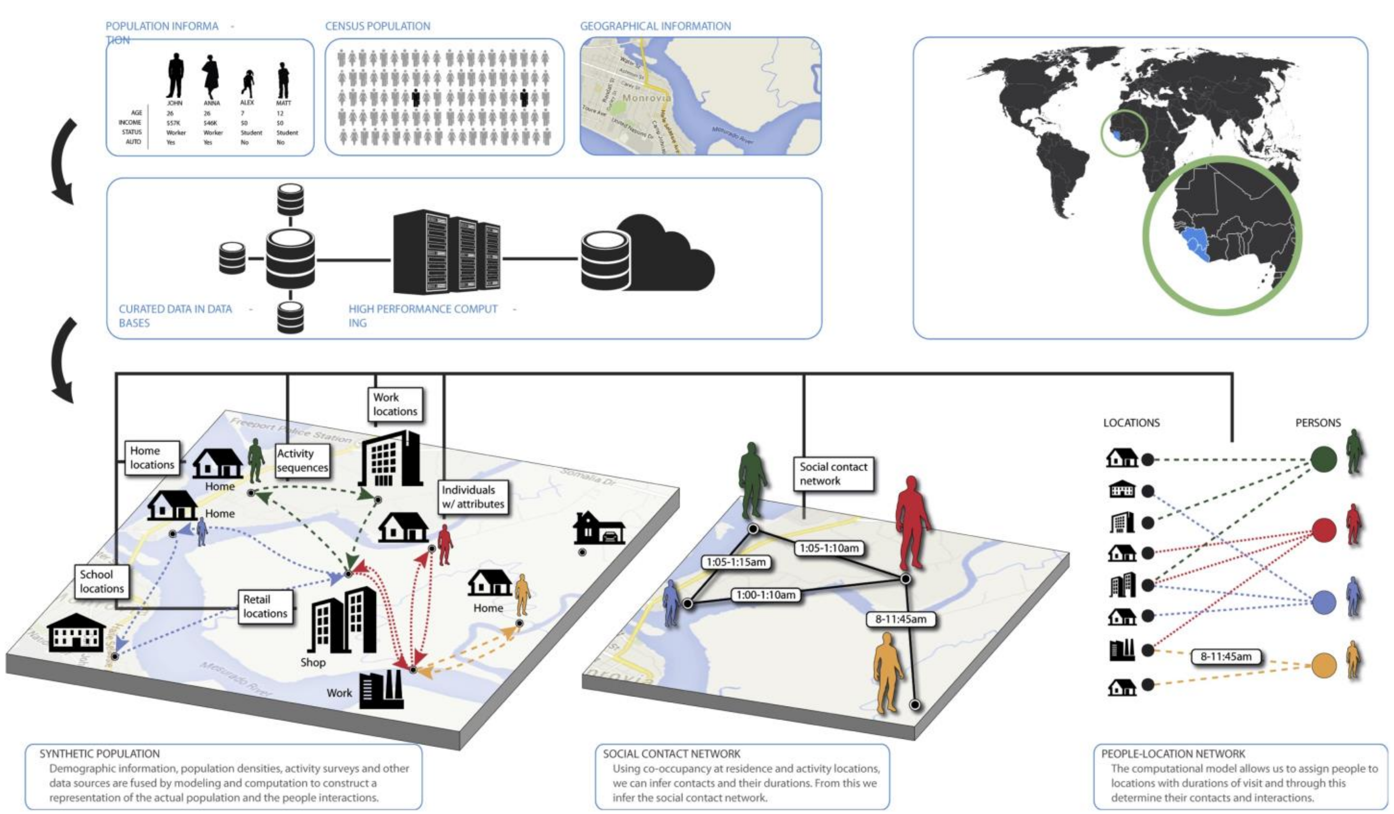
- Four synthetic scenarios of Ebola in Liberia (varying in intensity, interventions, data availability)
- Forecasting short-term and long-term targets at 5 different timepoints
- Combination of quantitative (case counts) and qualitative (situation reports) data



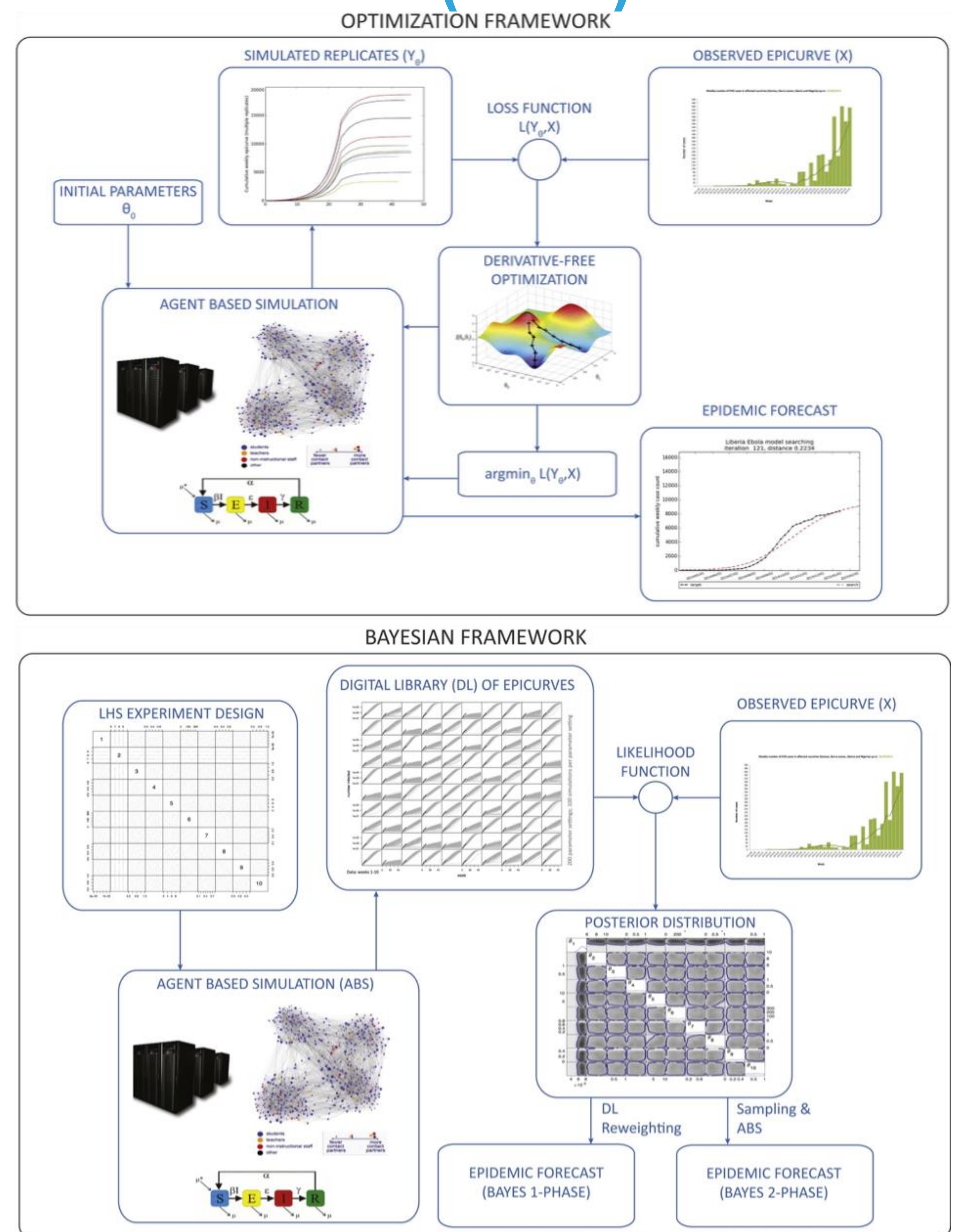
Viboud, Cécile, et al. "The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt." *Epidemics* 22 (2018): 13-21.

DETAILED AGENT BASED MODELING (EFC)

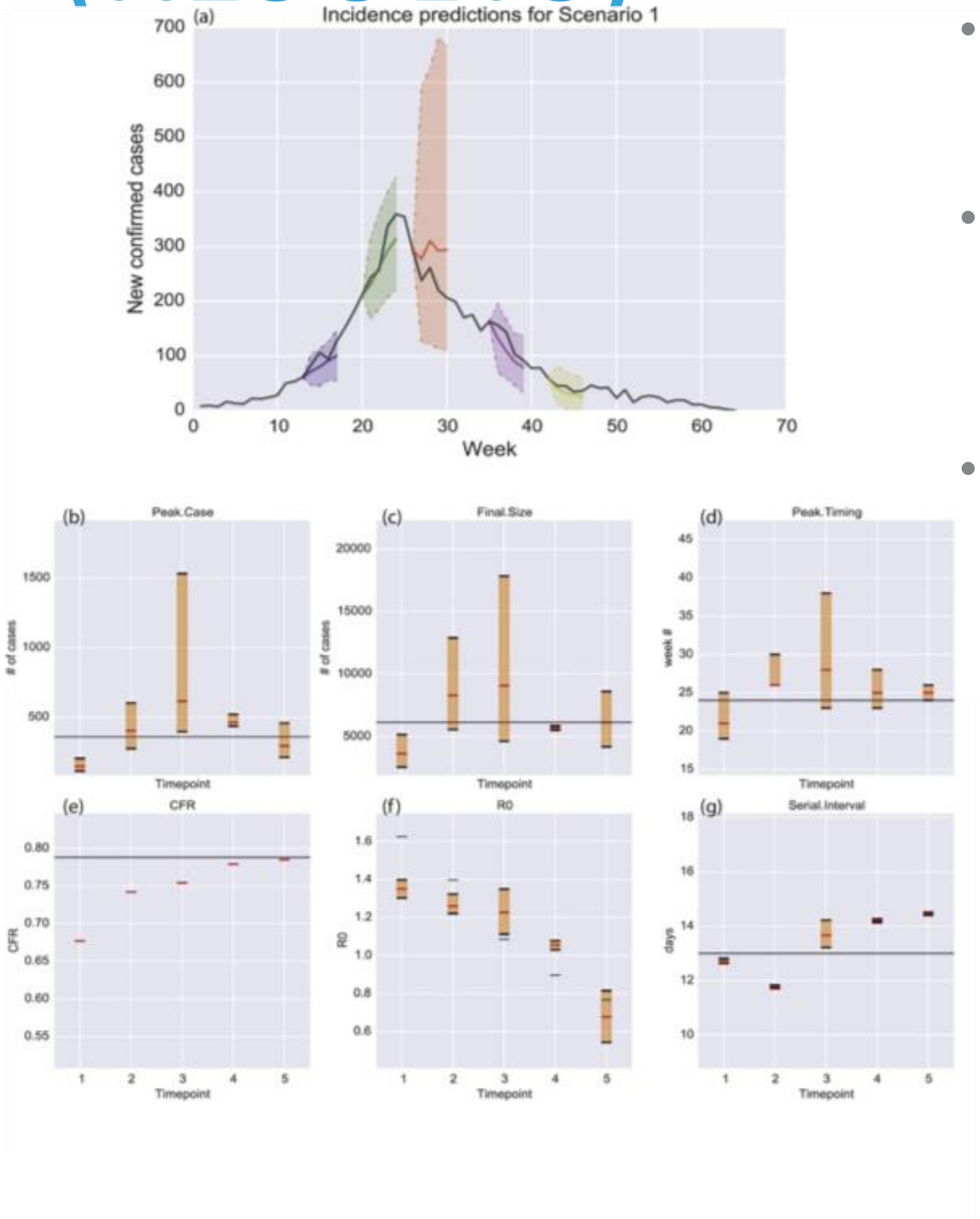
- Our approach: Calibrated agent-based models using synthetic populations
- Calibration using two different approaches: derivative-free optimization and bayesian framework.



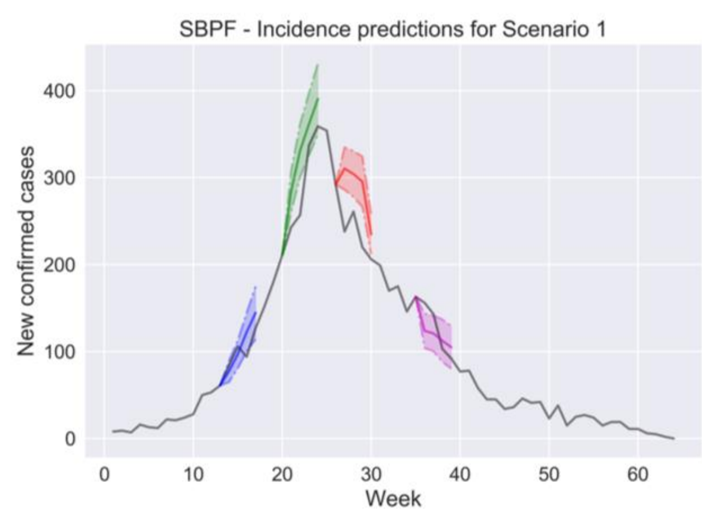
Venkatramanan, Srinivasan, et al. "Using data-driven agent-based models for forecasting emerging infectious diseases." *Epidemics* 22 (2018): 43-49.

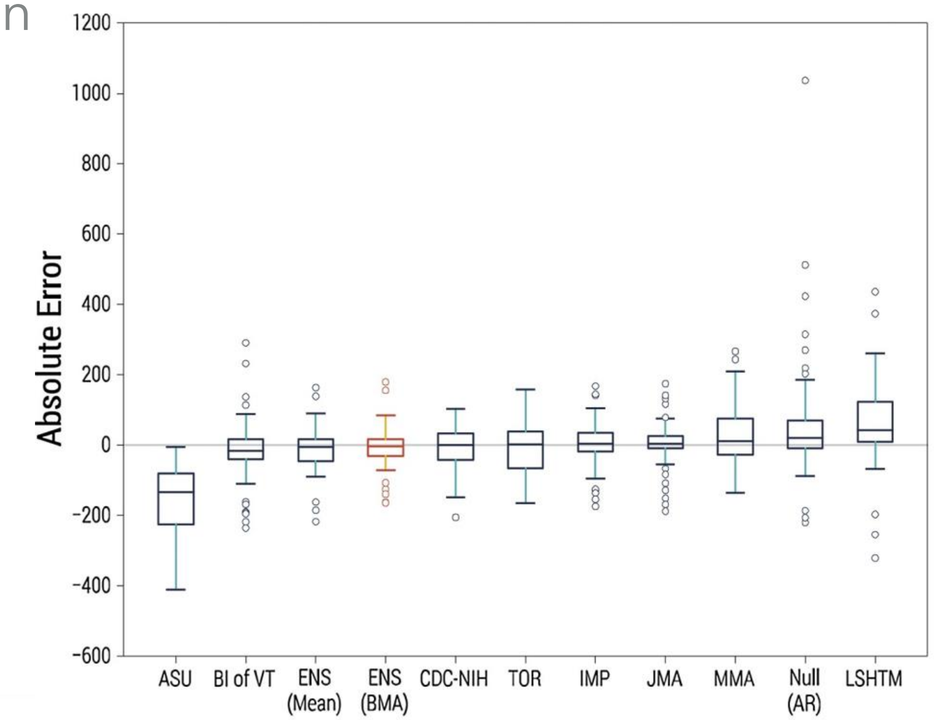


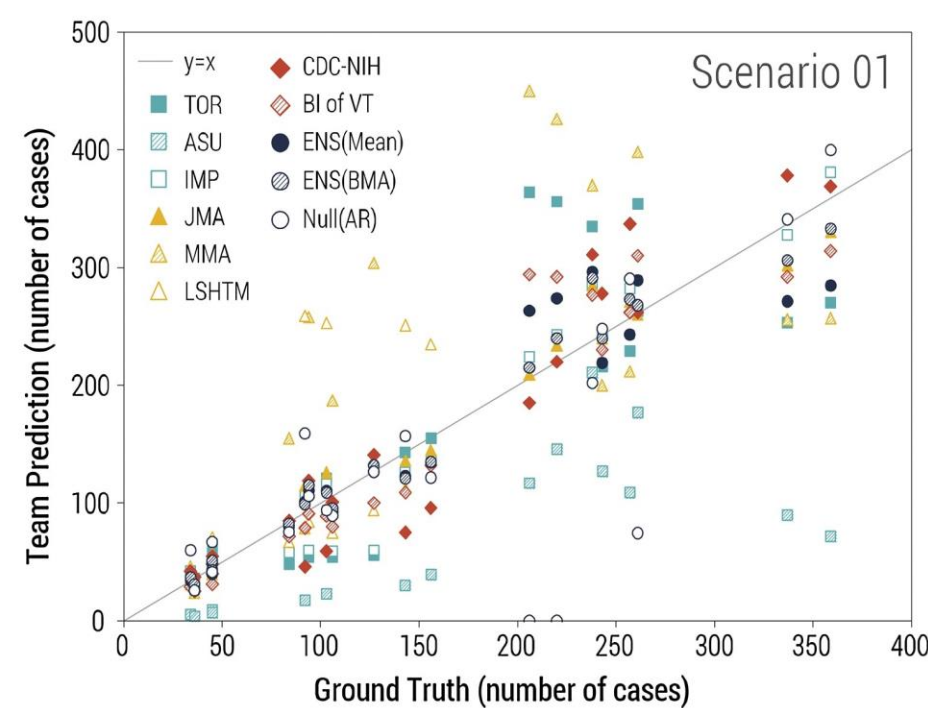
EBOLA FORECASTING CHALLENGE (RESULTS)



- Among the top three teams in the challenge.
- Consistently provided high resolution (county) level forecasts.
- Further improved performance by using smart-beam particle filter

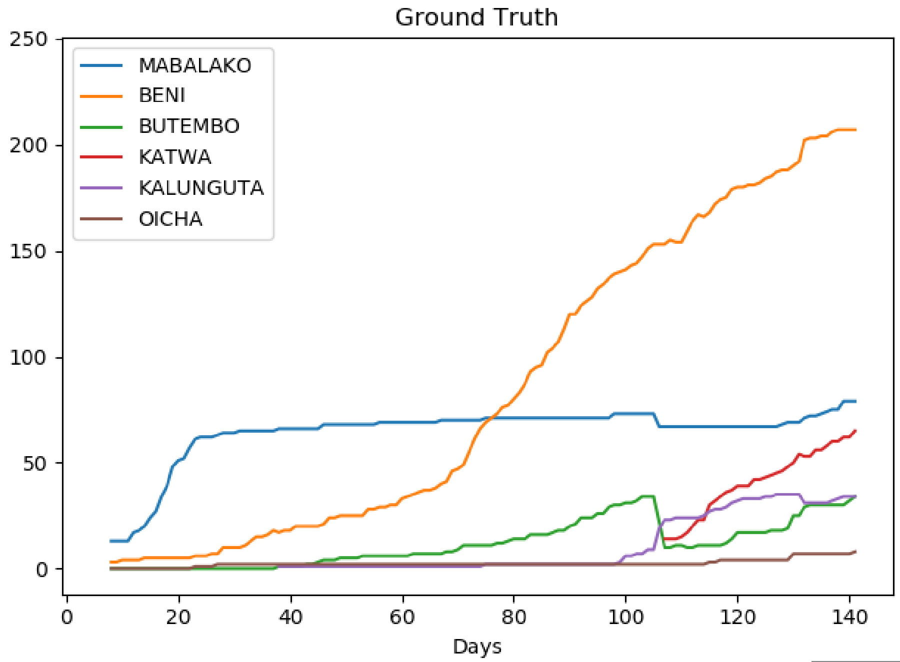


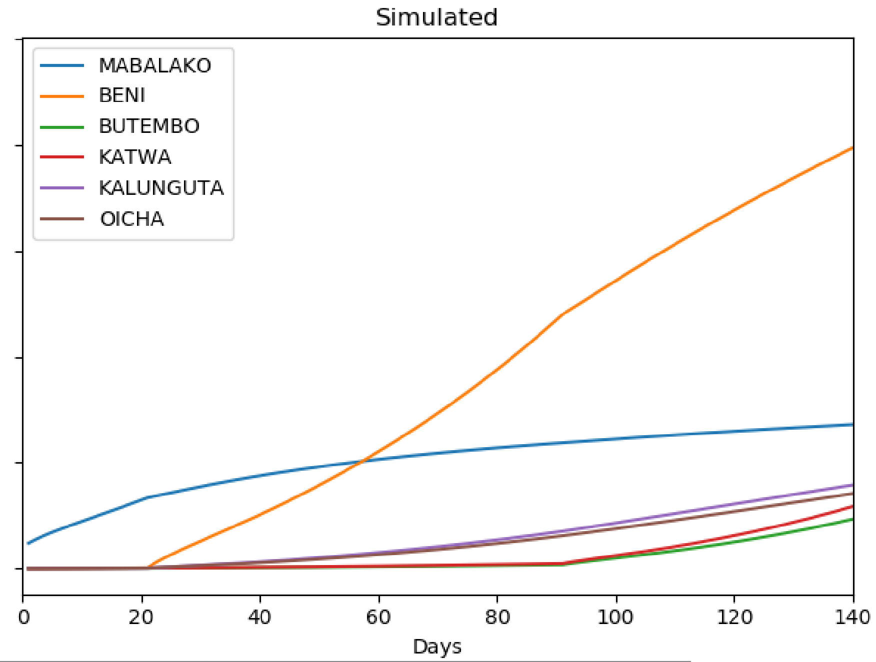


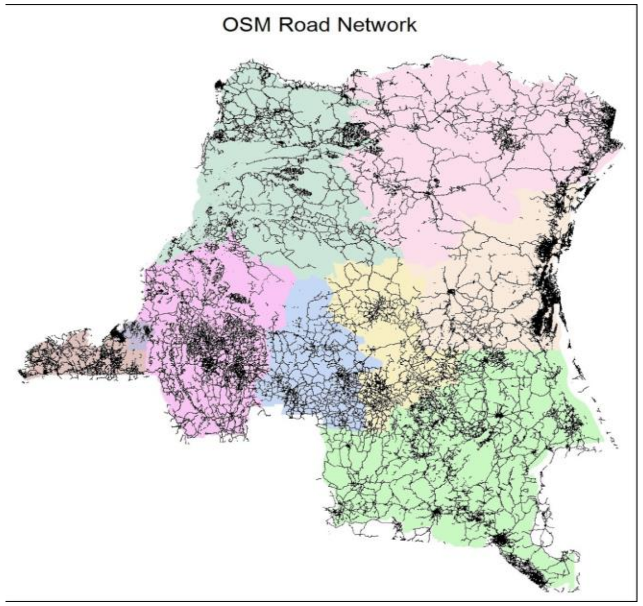


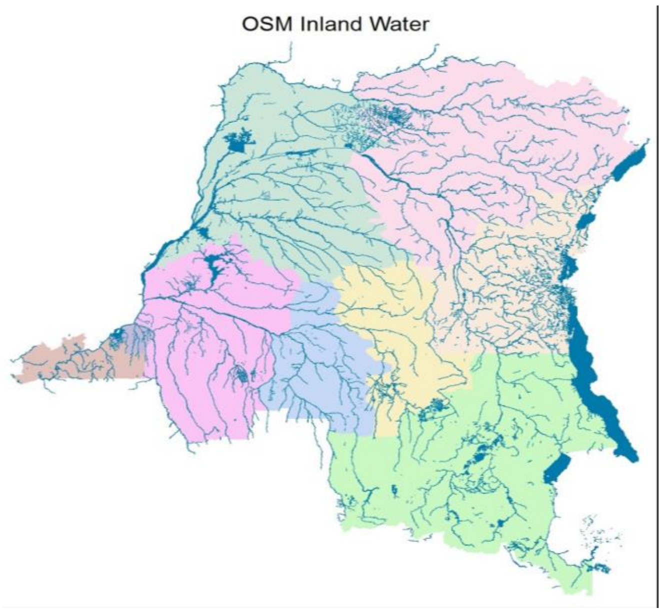
Tabataba, Farzaneh S., et al. "Epidemic Forecasting Framework Combining Agent-Based Models and Smart Beam Particle Filtering." *Data Mining (ICDM), 2017 IEEE International Conference on*. IEEE, 2017.

DRC EBOLA (GIS + PATCHSIM)

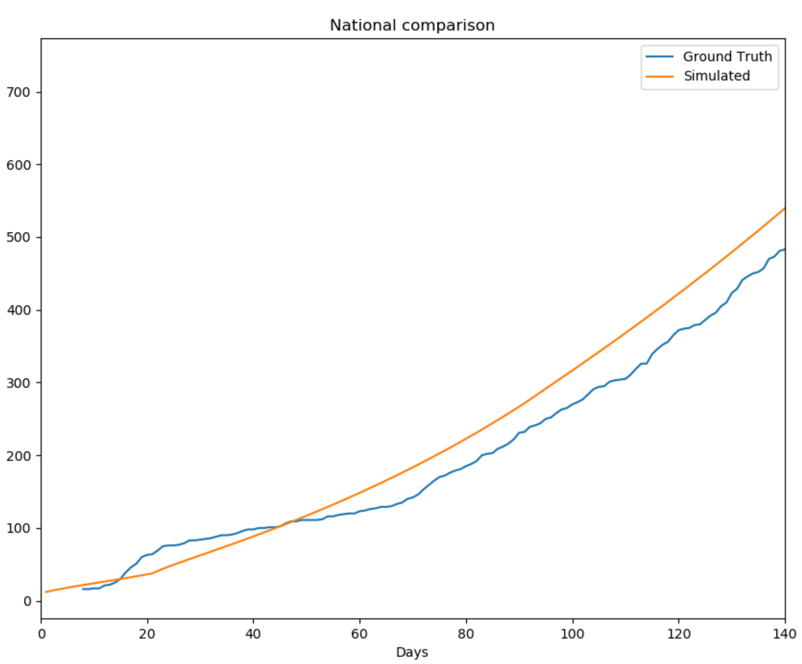


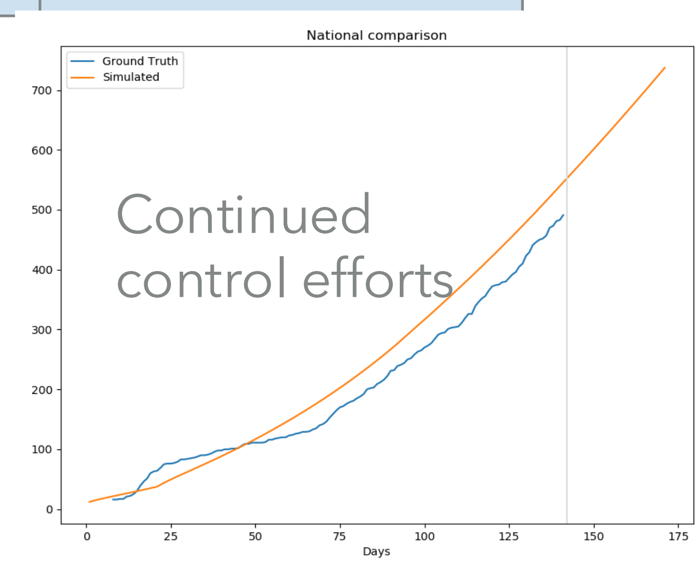


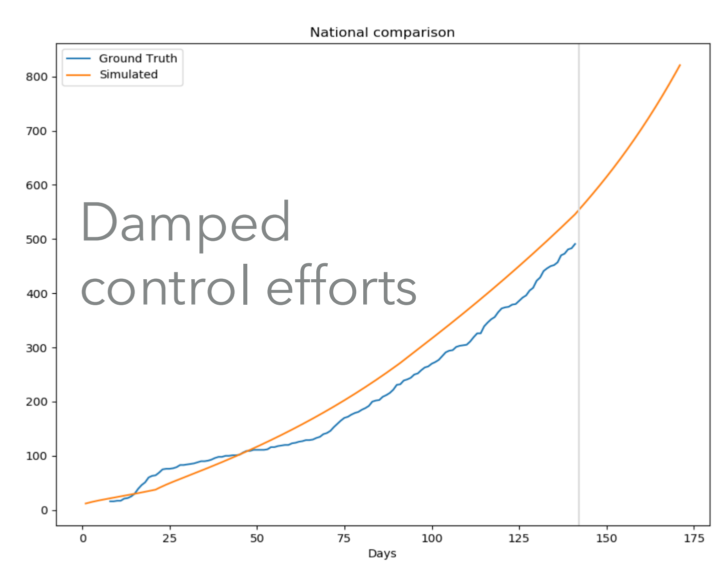




Health Zone	Risk Index
BENI	20.5%
KATWA	13.2%
BUTEMBO	10.3%
KALUNGUTA	6.7%
OICHA	6.7%





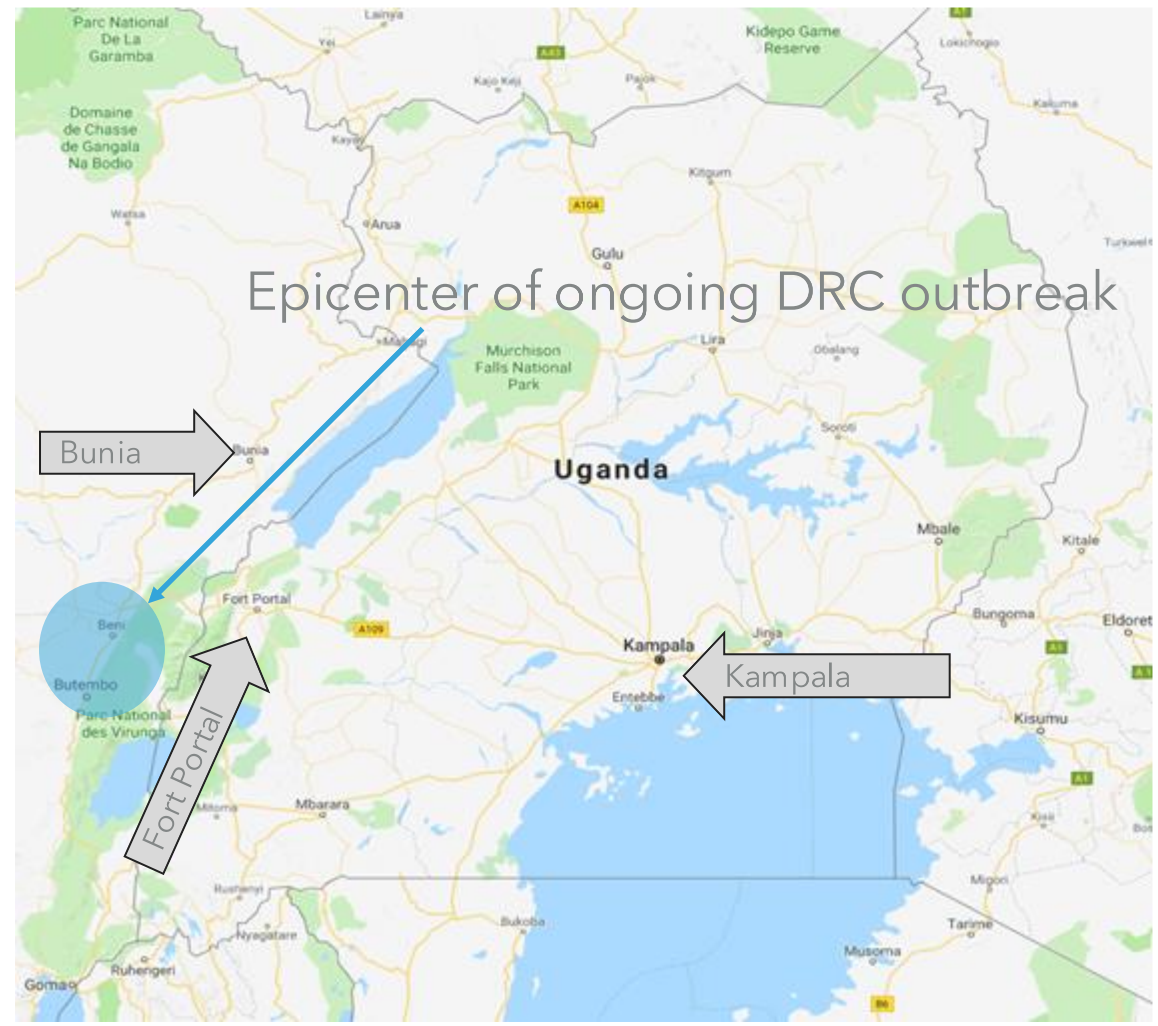


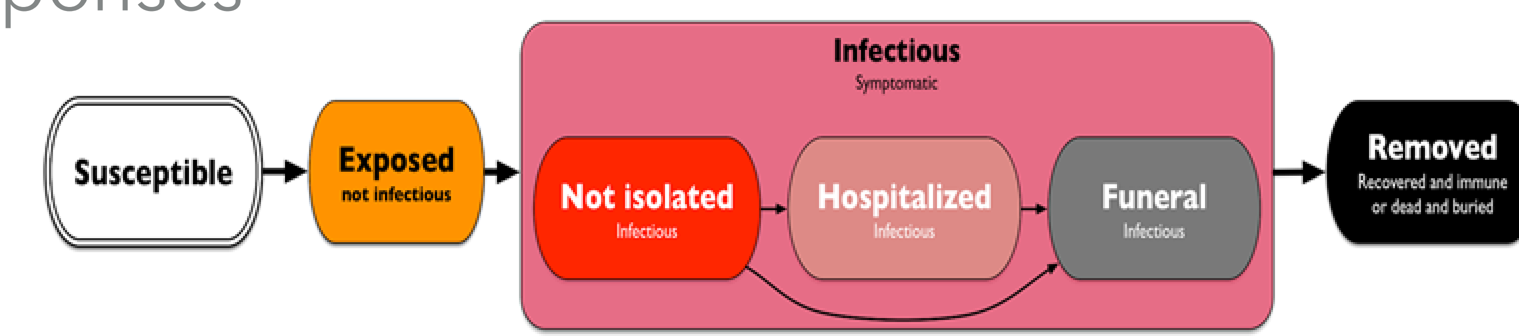
Y-axis: confirmed EVD cases X-axis: Days since Jul 31, 2018

Risk index: proportion of projected cases in next 4 weeks

EBOLA RESPONSE MODELING – UGANDA (HYPOTHETICAL)

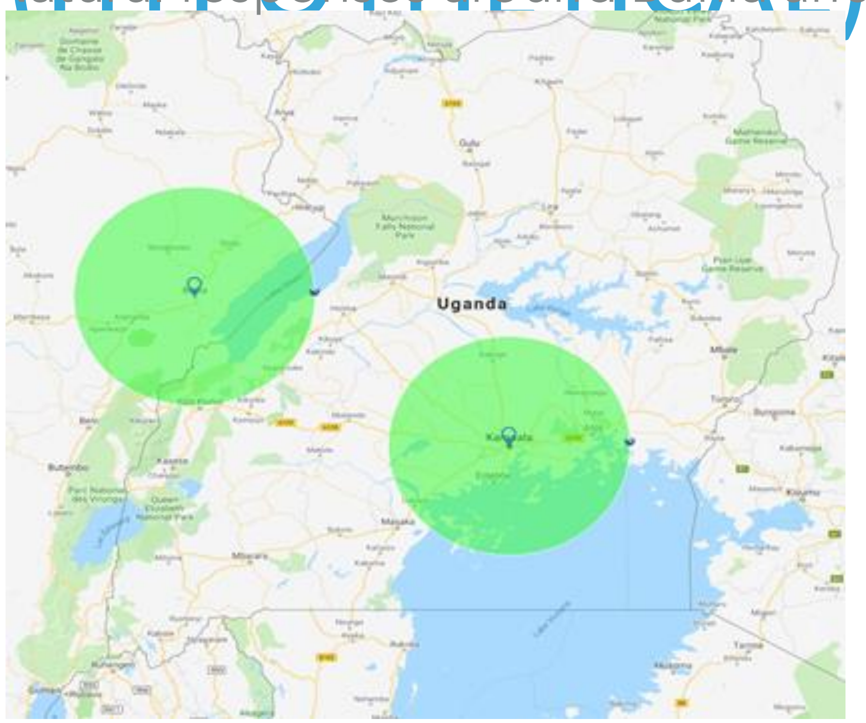
- Study done in March 2018!
- Outbreak in Bunia province of DRC, first case in Uganda 4 months after.
- Population/mobility modeling
  - LandScan aggregated to 10km X 10km grid cells
  - Mobility between patches by gravity model
- Impact of aided/delayed responses

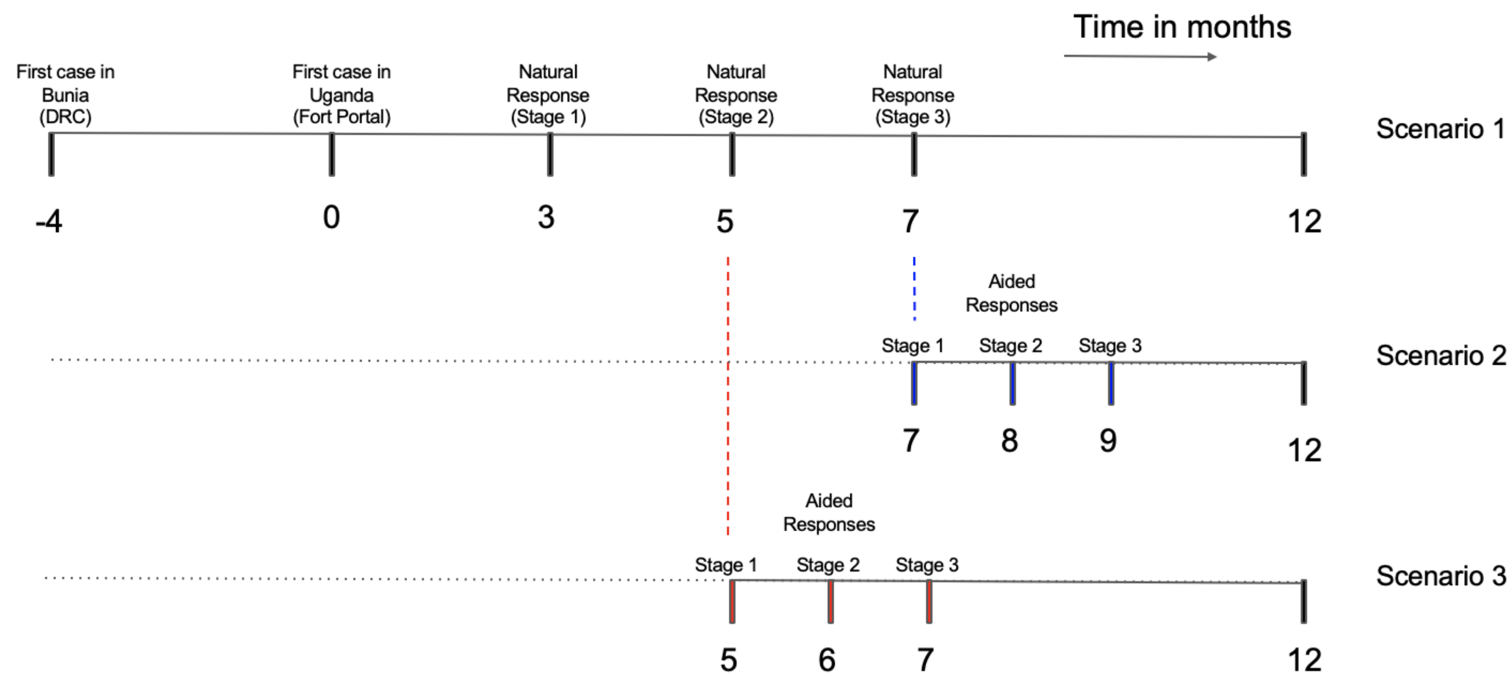




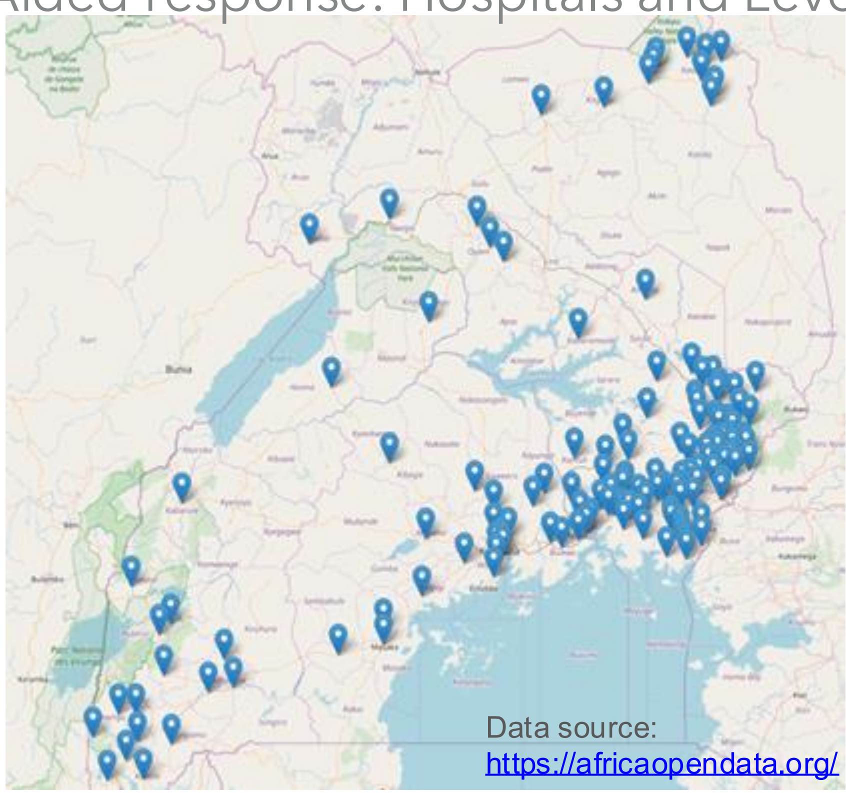
EBOLA RESPONSE MODELING – UGANDA (HYPOTHETICAL)

Natural responses around Bunia and Kampala



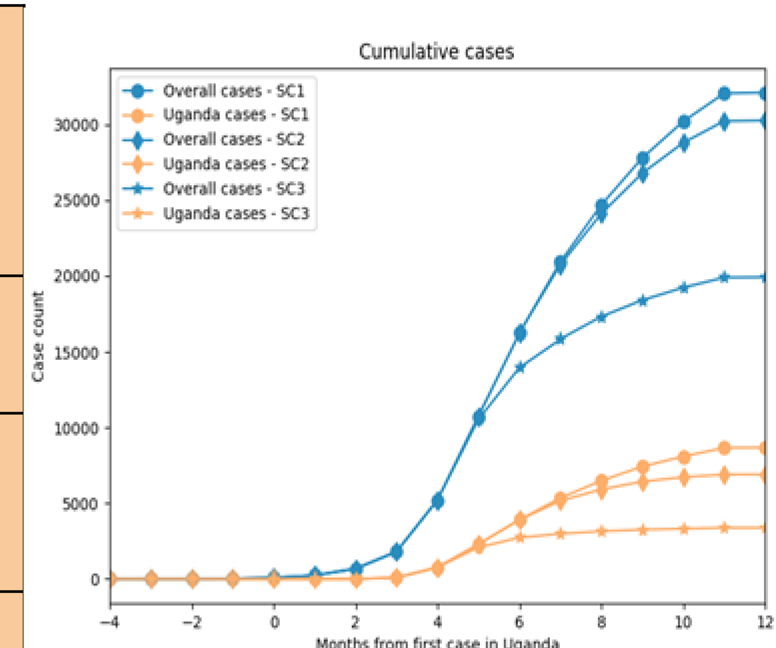


Aided response: Hospitals and Level IV HCs



Data source: https://africaopendata.org/

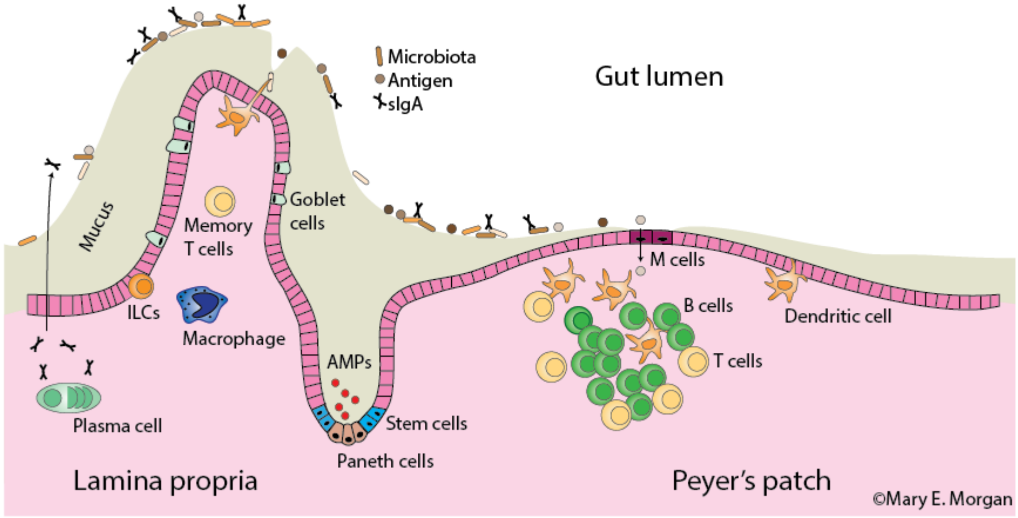
Scenario	Overall cases	% change (from scenario 1)	Uganda cases	% change (from scenario 1)
Scenario 1 (Natural)	32113		8682	
Scenario 2 (Aided, late)	30268	- 5.74%	6906	- 20.45 %
Scenario 3 (Aided, earlyl)	19906	-38.01%	3381	- 61.05 %



LISTENING TO YOUR GUT

(Organ-system Scale)

INTESTINAL IMMUNE SYSTEM MODELING



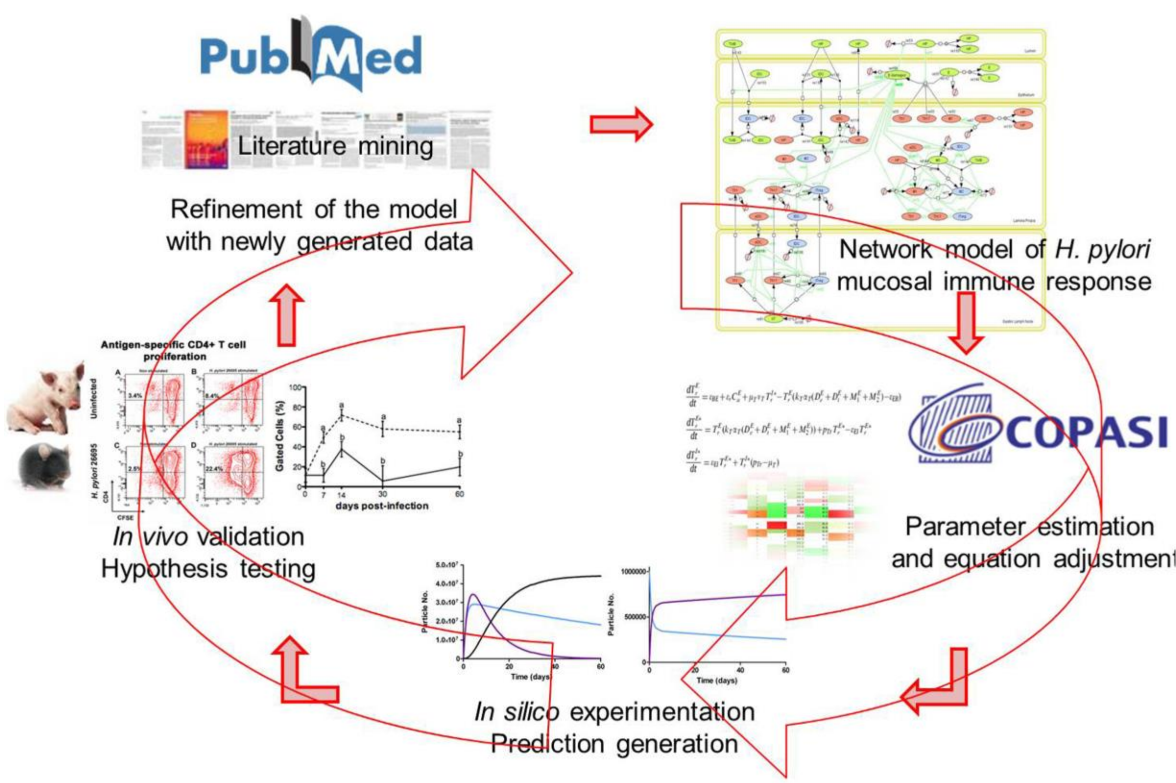
Source: theibdimmunologist.com

- Inflammation induced pathology in GI tract
- Complex interacting immune pathways
- ENISI: Enteric Immunity Simulator
  - Simulate mucosal immune responses in the gut
  - Studying immune regulation to pathogens such as *E.coli, B. hyodysenteriae, H. pylori*.
  - Representing bacteria and immune cells as agents, moving through tissue sites and engaging in context-dependent interactions.

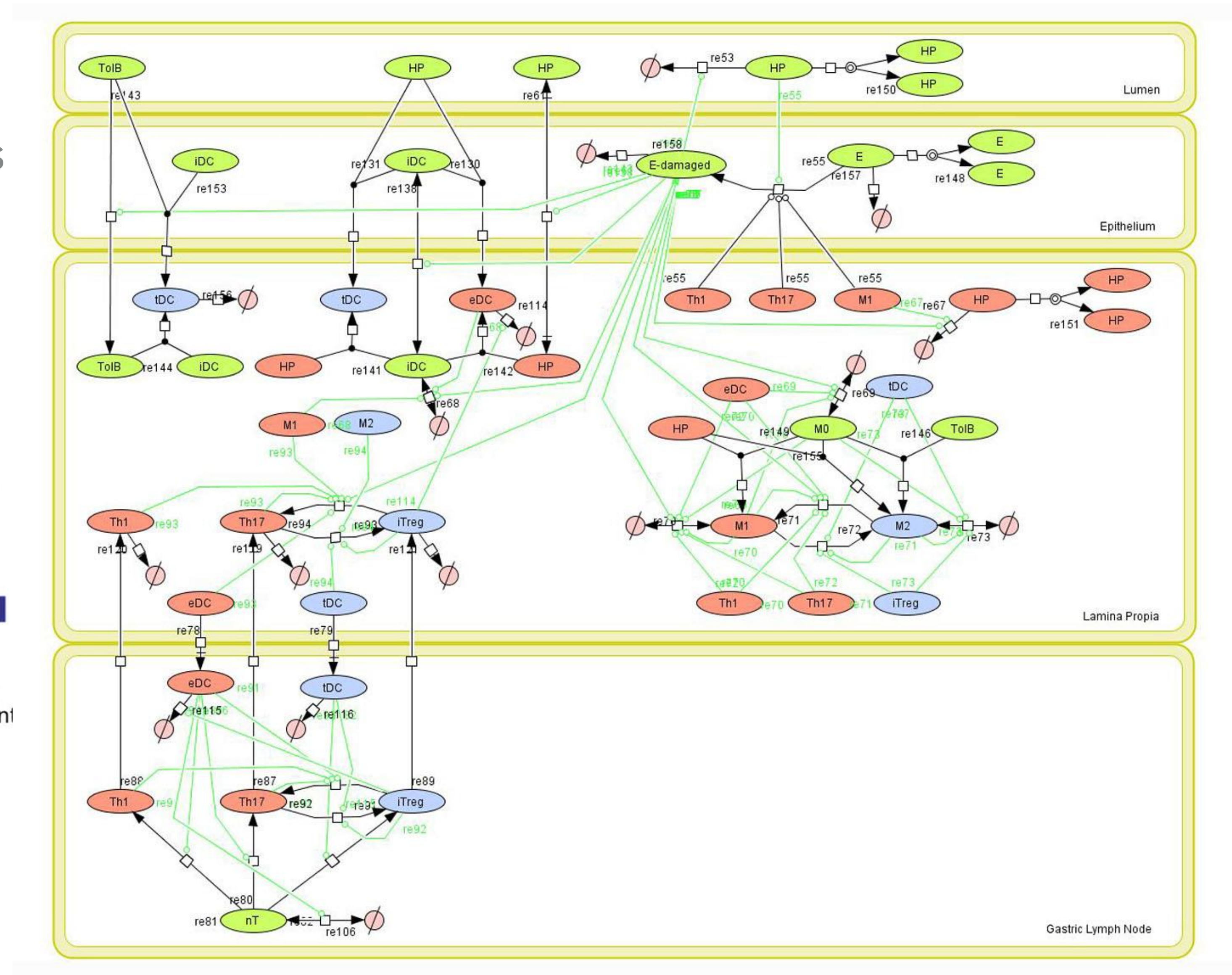
Wendelsdorf, Katherine V., et al. "ENteric Immunity SImulator: a tool for in silico study of gastroenteric infections." *IEEE transactions on nanobioscience* 11.3 (2012): 273-288.

MODELING IMMUNE RESPONSE TO H. PYLORI

- Understanding the role of T helper cells (Th1, Th17) in immune response and pathology
- Simulation models reproduce dynamics of effector and regulator pathways
- Using both ODE and ABM based approaches
- Co-evolving Graph Dynamical Systems



Integrating computational and experimental approaches



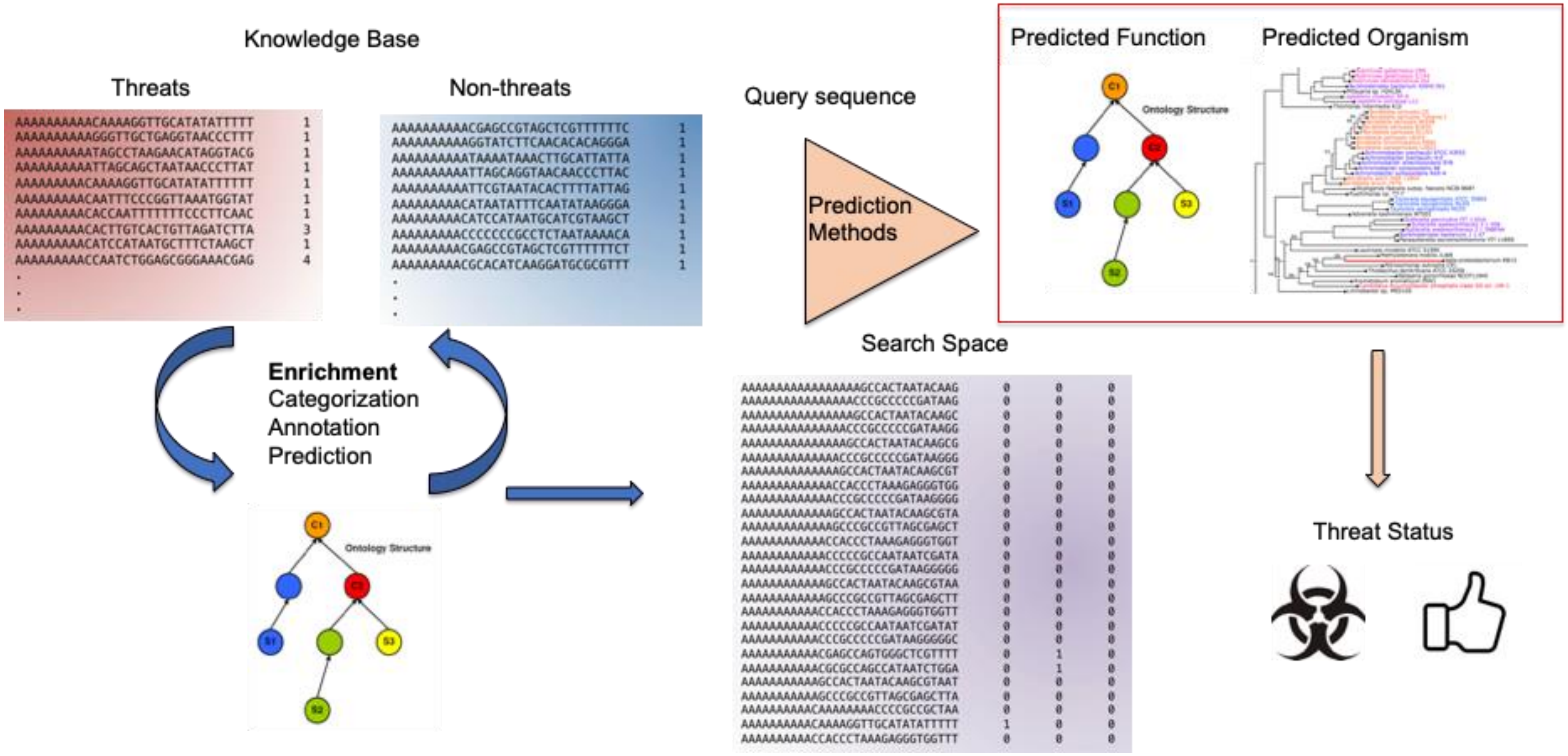
https://www.modelingimmunity.org/CellPubModels/HP_May2012/

GO FOR GENOME!

(Molecular Scale)

FUNGCAT

- Functional Genomic and Computational Assessment of Threats (IARPA program)
- Approaches and tools for screening nucleic acid sequences and functional gene annotation
- Preventing accidental creation of biological threats



Challenges: Variety of databases (UniProt, TrEMBL, NCBI)
Limited understanding of Sequence -> Structure -> Function
Limited Training data (biased)

APPROACHES

- Individual teams:
  - Data Acquisition and Integration
  - Computational Framework
  - Predictive Modeling
  - Ensemble modeling
- Deep neural networks for predicting Gene Ontology and Organism (sequence embedding with word2vec, auto-encoders)
- Clustering based on sequence similarity (at protein level)
- Building heterogeneous ensembles (with diverse base predictors)

Wang, Linhua, et al. "Large-scale protein function prediction using heterogeneous ensembles." *F1000Research* 7 (2018).

BEING

LESSONS

LEARNT

LESSONS (BEING) LEARNT - TECHNICAL

- Ensembling approaches prevail over others
- Theory guided Data science (TGDS) bridges the domain experts and data scientists
- Human in the loop is unavoidable. Wisdom of (expert) crowds.
- Data summarization (visualizations) is key to make decisions (or even make sense of the data)

LESSONS (BEING) LEARNT - OPERATIONAL

- Benchmarks/Contests help align targets and metrics. Beware of group-think!
- Team Science is essential to address real-world problems. Rethinking credit sharing practices.
- Open source environment (papers, code, data) accelerates research and ensures reproducibility. Incentive mechanisms!
- Automated pipelines must be built with scalability, robustness, development and testing in mind.
- Translational research: Start from the problem. Not from the data. From papers to practice.

LESSONS (BEING) LEARNT - ETHICAL

- Secure/Safe:
  - Ensuring protection of PII. Algorithms robust to adversarial attacks.
  - Healthcare technology must first do no harm.
- Accessible:
  - Reaching unserved/under-served population.
  - Mobile, connected, cost-effective, rapidly deployable.
- Fair:
  - Accounting for health disparities in minorities. Equality vs Equity.
- Explainable:
  - Causally linking the decision made to input data.

QUESTIONS?

srini@virginia.edu
@sriniv_venkat

https://biocomplexity.virginia.edu/