Srinivasan Venkatramanan

Modeling in the Time of Ebola

Using HPC Simulations to Understand Infectious Disease Dynamics

Virginia Tech Invent the Future
NDSSL Network Dynamics & Simulation Science Laboratory
Biocomplexity Institute

Part I

Epidemiology - Then & Now

What is Epidemiology?

- from Greek from ***epi*** "among, upon" + ***demos*** "people" + ***logos*** "study of"
- **Epidemiology:** study of the *distribution* and *determinants* of health-related states or events in specified populations, and the application of this study to the *control* of health problems
  - Now applies to non-communicable diseases as well as social and behavioral outcomes
  - Computational/mathematical epidemiology - deals with the development of computational/mathematical methods, tools and techniques to support epidemiology

Evolution of Epidemic Modeling

BEGINNINGS OF FORMAL EPIDEMIC MODELING

1796 — SMALLPOX // Virus

Edward Jenner's research led to the development of vaccines.

Daniel Bernoulli mathematical models demonstrated the benefits of inoculation from a mathematical perspective.

Disease status today: eradicated.

1854 — CHOLERA // Bacteria

John Snow was the first to link the London cholera epidemic to a particular water source.

Disease status today: endemic; occuring in poverty-stricken countries.

1897 — MALARIA // Parasite

Ronald Ross and George Macdonald developed a mathematical model of mosquito-borne pathogen transmission.

Anderson McKendrick studied with Ross on anti-malarial operations, pioneering many discoveries in stochastic processes.

Disease status today: controlled in US; still prevalent in Africa, India.

1946 — TUBERCULOSIS // Bacteria

Albert Schatz discovered the antibiotic streptomycin under the direction of Selman Waksman.

Streptomycin was the first antibiotic that could be used to cure TB.

Disease status today: drug resistant TB strains persist since the 1980s.

1981 — HIV // Virus

Luc Montagnier discovered HIV and Robert Gallo determined HIV is the infectious agent responsible for AIDS.

The use of social network models have been initiated with the goal of controlling the virus.

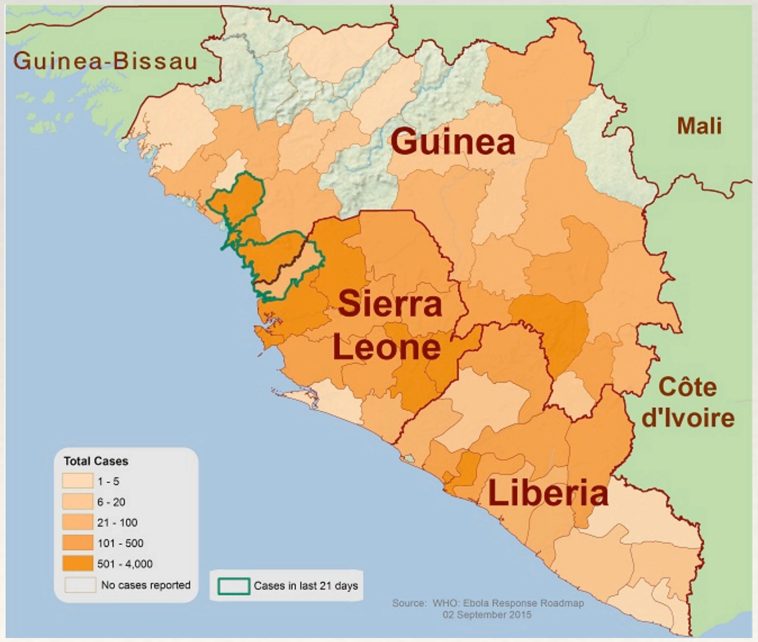
Disease status today: no cure.

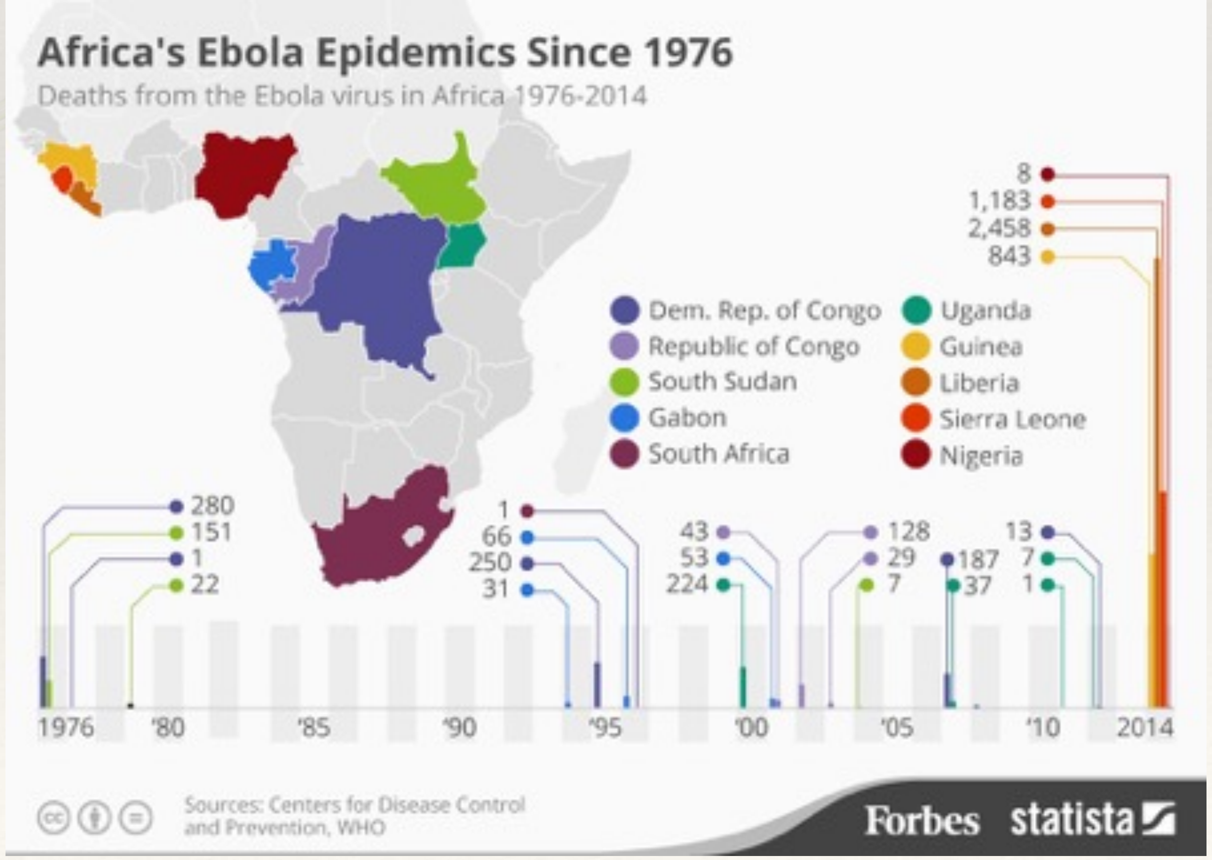
Modern Trends in Epidemiology

- ***Improved disease surveillance***
  - crowdsourced surveillance via participatory apps
  - social media and search trends
- ***Detailed computational models*** for both local and global spread
- ***Active interdisciplinary effort*** involving experts from public health, social science, mathematics, computer science, etc.
- ***Modern challenges*** - highly connected mobile population; emerging diseases

Ebola - 2014 outbreak - West Africa

- Largest and most complex Ebola outbreak till date
- 5 countries; 28000 cases; 10500 deaths (WHO, Sept 2015)
- Zoonotic in nature (fruit bats and monkeys)
- Passed among humans through bodily fluids
- Average incubation period ~ 12 days
- Symptom onset to recovery/death ~ 9 days
- Overall case-fatality ratio ~ 54%





VirginiaTech Invent the Future

NDSSL Network Dynamics & Simulation Science Laboratory

On the ground

- Severe load on an already weak healthcare infrastructure
- Doctors Without Borders/Médecins Sans Frontières (MSF)
  - *"Pushed to the limit and Beyond"*, MSF Report, 2015
  - Nearly 500 healthcare workers died in Guinea, Liberia, Sierra Leone
- ***What worked?***
  - Placement of Ebola treatment units (ETU)
  - Personal protective equipment (PPE) for Healthcare workers
  - Improved surveillance and contact tracing
  - Community engagement - Safe burial practices
  - No effective vaccines developed during the outbreak

- Long term socio-economic and health impacts still being ascertained
- Spanish Flu (1918) …, SARS(2003), H1N1 (2009), Ebola (2014), Zika (2016)? ….
- *"Lessons from Ebola: We're Not Ready for the Next Epidemic"*, Bill Gates, 2015

Part II

Models - Simple, Complicated & Complex

Why Model?

"All models are wrong. Some are useful."

- Added complexity when modeling systems involving humans
  - Impossible to accurately capture human behavior
  - *"Self-negating prophecies"*: Forecast information feeds back into behavior
- Models are critical during both ***war time*** and ***peace time*** efforts
- ***"War time"*** (during the epidemic)
  - Immediate **risk assessment**: short-term & long-term forecasts, by region and by community
  - Determine the **load on critical healthcare infrastructure**
  - **Decide among interventions** through what-if scenarios
  - Suggest **resource deployment** strategies (e.g., vaccinations, emergency clinics, healthcare personnel)

Why Model?

- *"Peace time"* (post/pre-epidemic)
  - Advance **tools, techniques** and **understanding** for quick response from the modeling community
  - Recruit and train **emergency personnel** for rapid deployment
  - **Retrospective studies** to identify what went right/wrong
  - Spot **system vulnerabilities** and improve resilience
  - Identify **data gaps** and improve monitoring
  - **Educate the masses** - preparedness vs panic

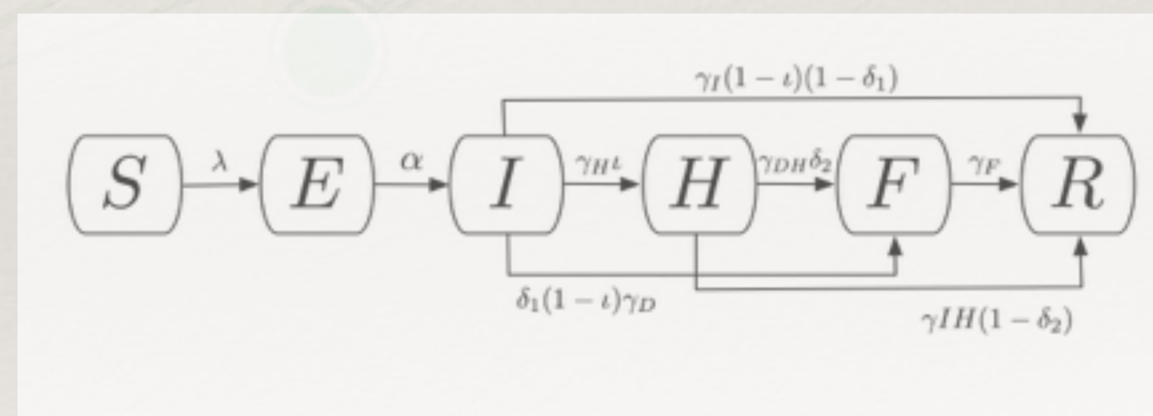
Rivers, Caitlin. "Ebola: models do more than forecast." Nature 515.7528 (2014): 492-492.

Brown, David. "Computer modelers vs. ebola." Spectrum, IEEE 52.6 (2015): 62-67.

Compartmental Models

- SI, SIR (***simple***)
  - Elegant analysis and interesting insights
  - Quick to setup; critical for initial assessment
  - Simplifying assumptions
  - Very few, abstract control knobs for policy planners

- SE_mI_nR, Legrand model (***complicated***)
  - More compartments, more parameters
  - More realistic
  - No analytical results, tough to gain insights
  - Still operates at population level

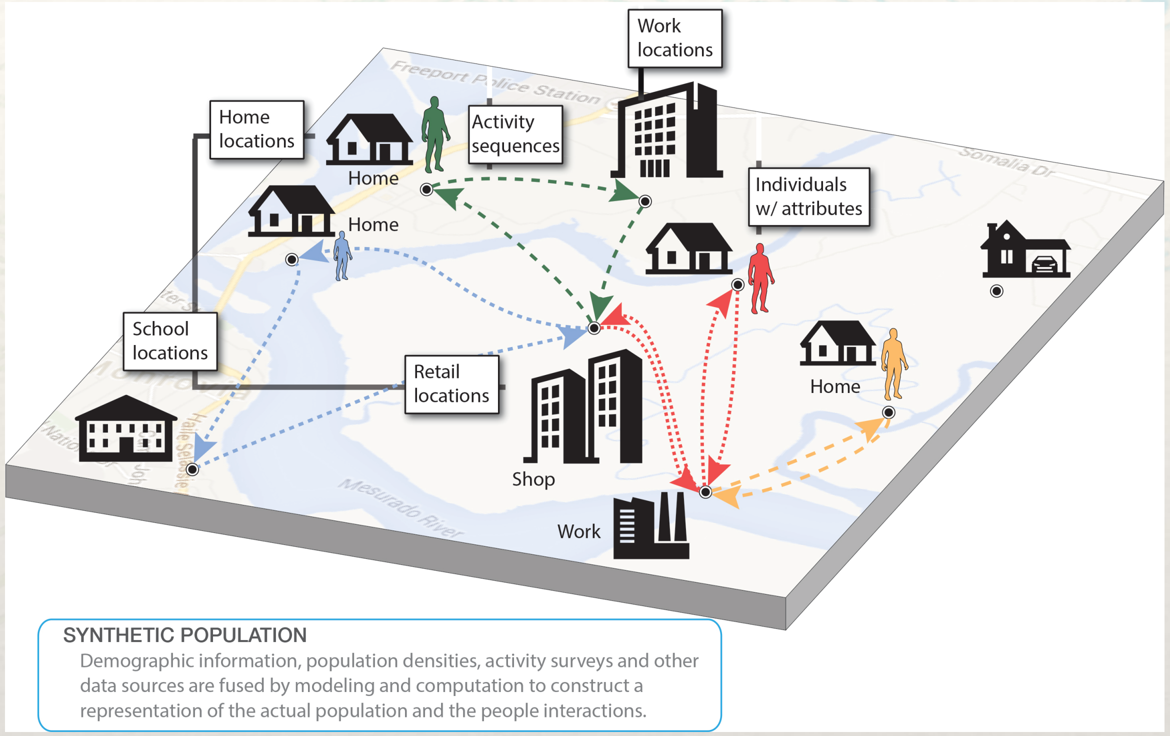


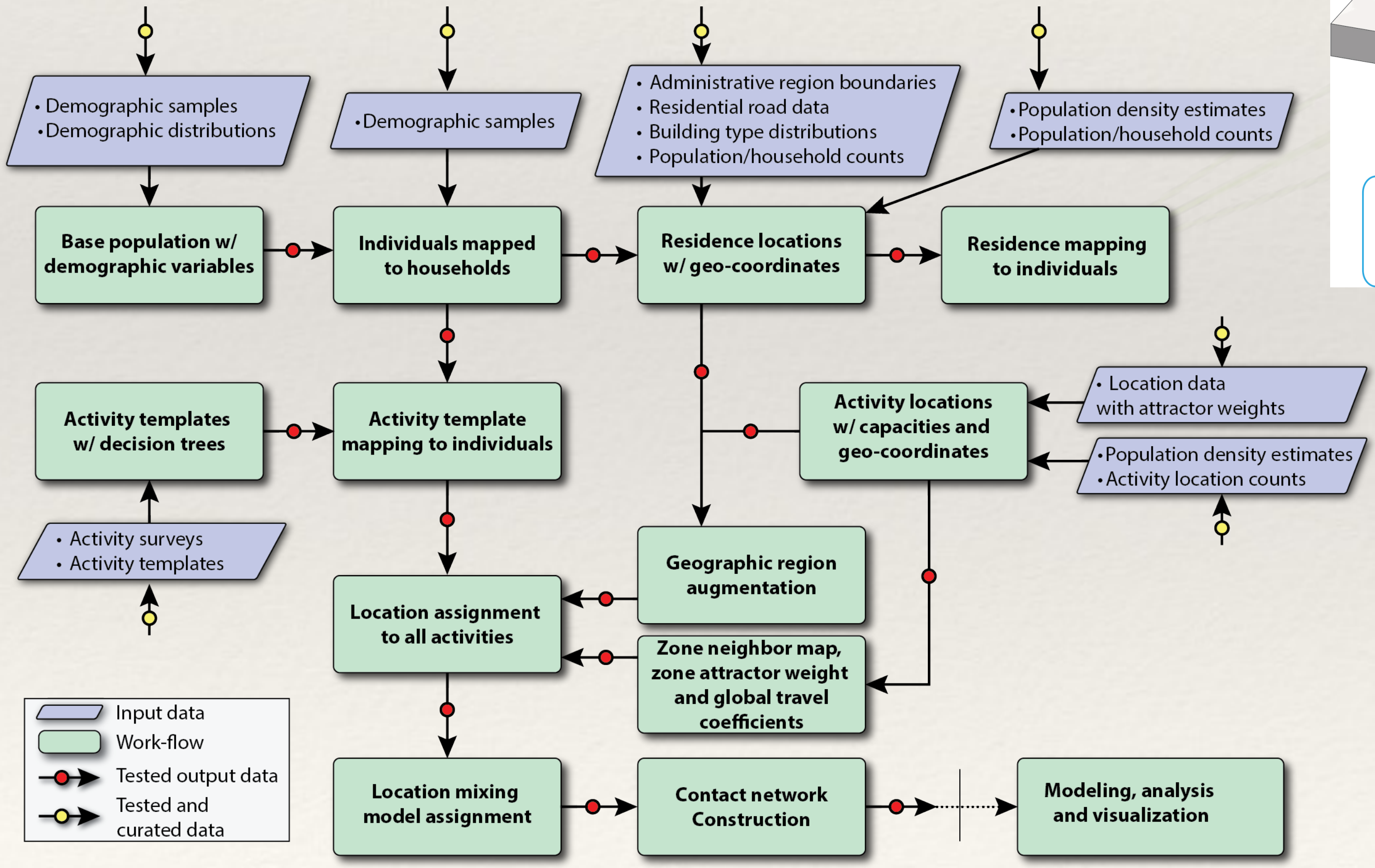
Network Models

- Contagions on ***complex networks*** (scale-free, small-world, etc.)
- Operates at the individual level and allows detailed analyses
  - transmission trees, local and global interventions
- Direct network construction often unrealistic
  - Social science approach - Surveys (HIV) - not scalable
  - Computer science approach -Passive collection (OSNs) - not always possible
  - *"Given a network G=(V,E) …"?*
- ***Our approach: Behavior & Data-driven network models***
  - Synthetic population of agents with coded activities/behavior
  - Networks emerge from agents' behavior (***truly complex***)

Synthetic Population

- Who, where, when and what?
  - individuals with household structure
  - home and other activity locations
  - activity sequences
  - statistically identical to US Census

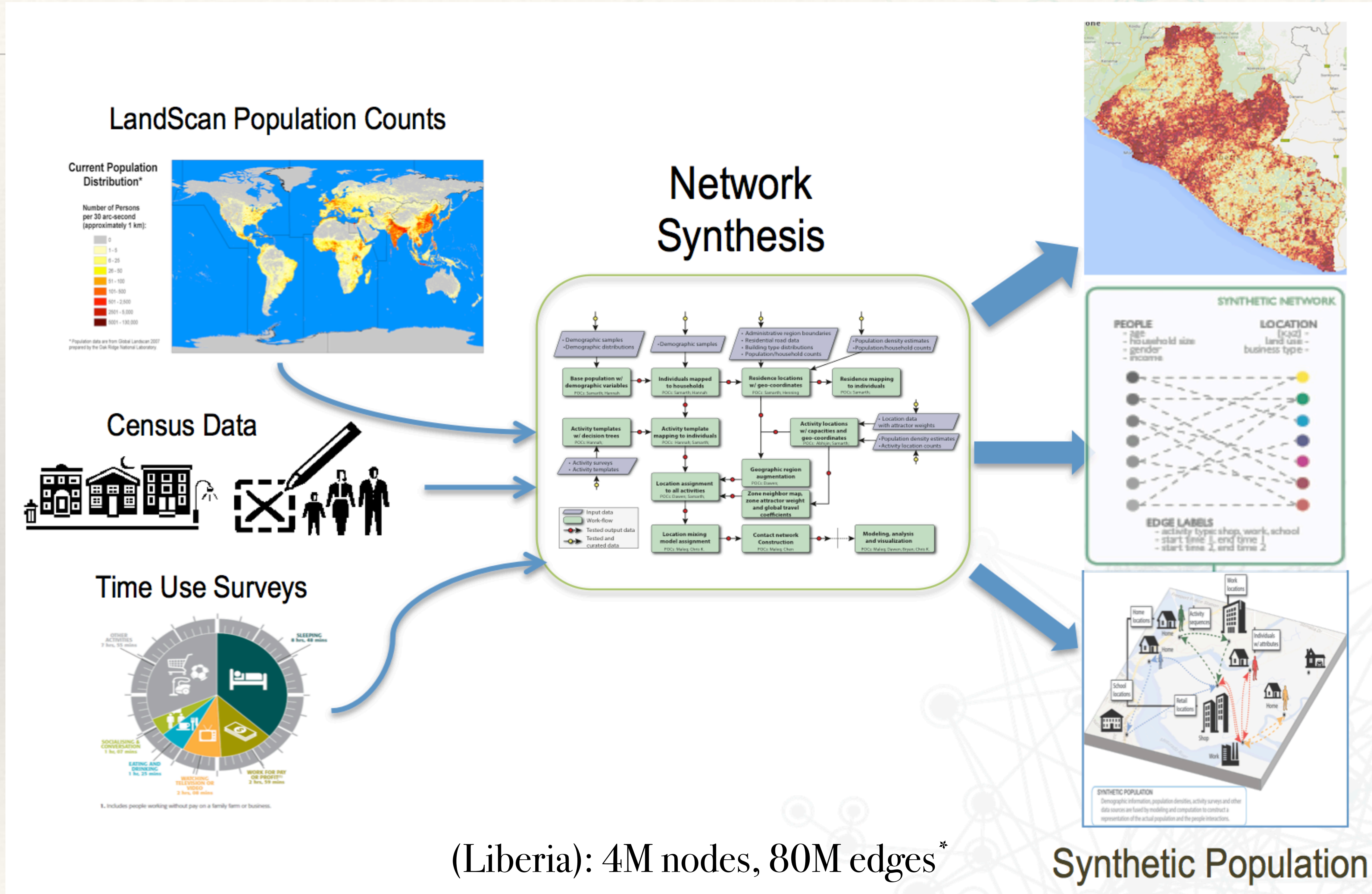




VirginiaTech
Invent the Future

NDSSL
NETWORK DYNAMICS & SIMULATION SCIENCE LABORATORY

Sample Scenario



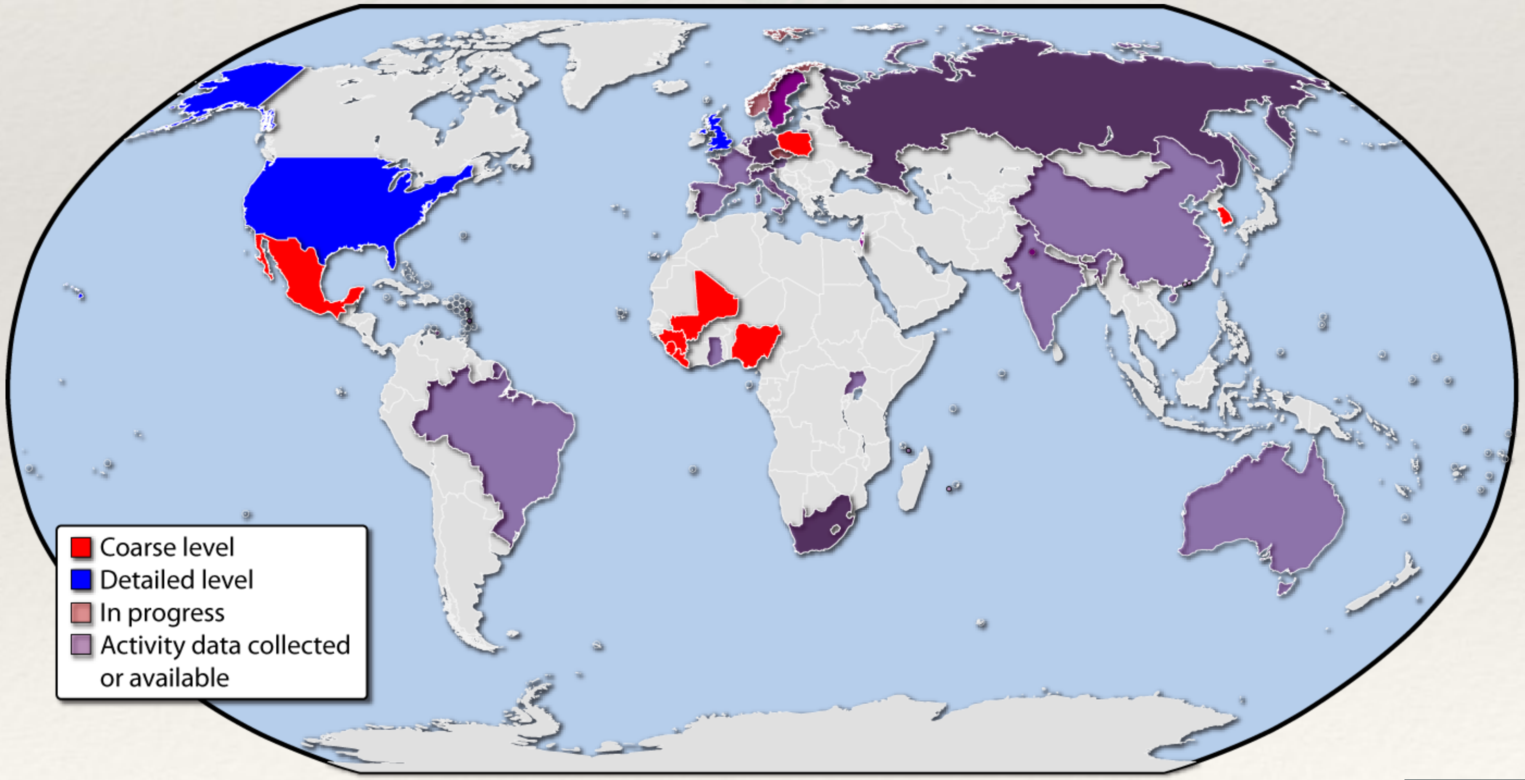
* Data available at: https://www.bi.vt.edu/ndssl/featured-projects/ebola/

VirginiaTech Invent the Future

NDSSL Network Dynamics & Simulation Science Laboratory

Current state-of-the-art

- *Coarse populations:*
  - Computational time for final version of current countries: 5 days
    - assuming extensive and powerful hardware
  - Number of countries for which data has been collected: 45
  - Storage: rule of thumb 1-4 GB/million individuals
  - Current size ~ 1.4 TB



- *Challenges:*
  - Continual update of the population
  - Easy integration of new data sources
  - Customization and extensions for specific studies

Simulation Environment

- **Fast and situation-specific simulation** of processes on the synthetic population
- Critical for supporting government agencies during ongoing epidemics
- Based on the mathematical abstraction of **Graph Dynamical Systems (GDS)**
- Rigorous verification & validation for interactionist models
    - ***Verification:*** Have we built the system right?
    - ***Validation:*** Have we built the right system?
- Different flavors: EpiFast, Indemics, EpiSimdemics
- ***EpiFast:***
    - Parallel algorithm and implementation for simulating standard SEIR model
    - Scalable, computationally fast and allows for various dynamic interventions
    - Single stochastic replicate of Ebola epidemic for 400 days, in the Liberia population (4M nodes, 80M edges) takes approximately 30 seconds on two 12-core Westmere-EP X5670 2.93 GHz Dual Processor Nodes with 48GB Memory and 1 TB Internal Hard Drive.

Part III

Calibration Methodology

Model Calibration

- Parameter estimation:

  Given the ground truth (epidemic curve),

  estimate model parameter values that best explain the data.

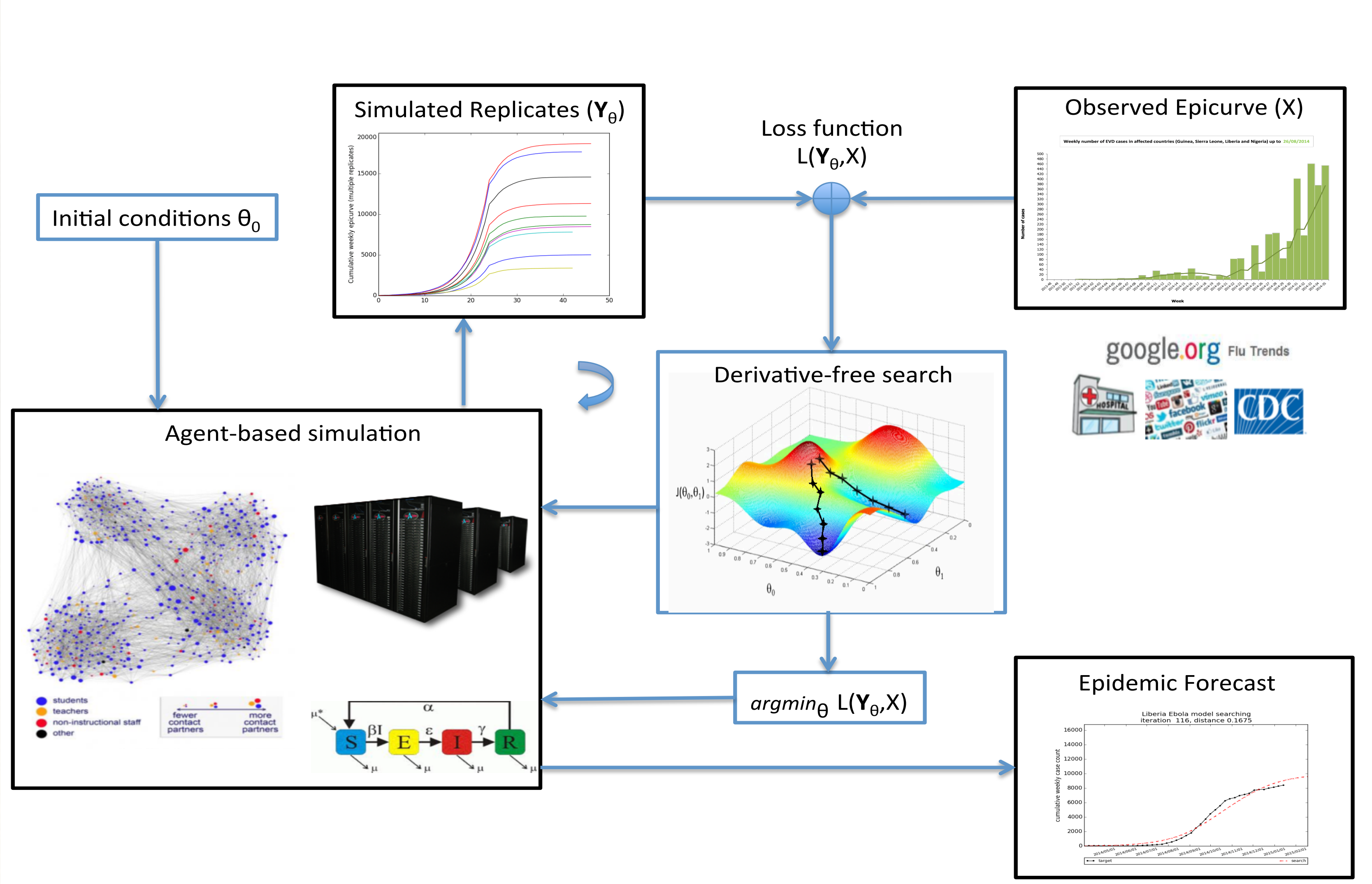
- Useful for situation assessment and critical for forecasting
- *Challenges:*
  - Ground truth: single instance from a stochastic system; often noisy
  - Model: High dimensional, highly non-linear, non-convex response surface

- **Our approaches:**
  - **Optimization framework**
    - Fewer simulation runs, quicker to estimate
    - Results are much easier to interpret for policy planners
    - May converge to local minima
    - Difficult to capture parameter uncertainty
  - **Bayesian framework**

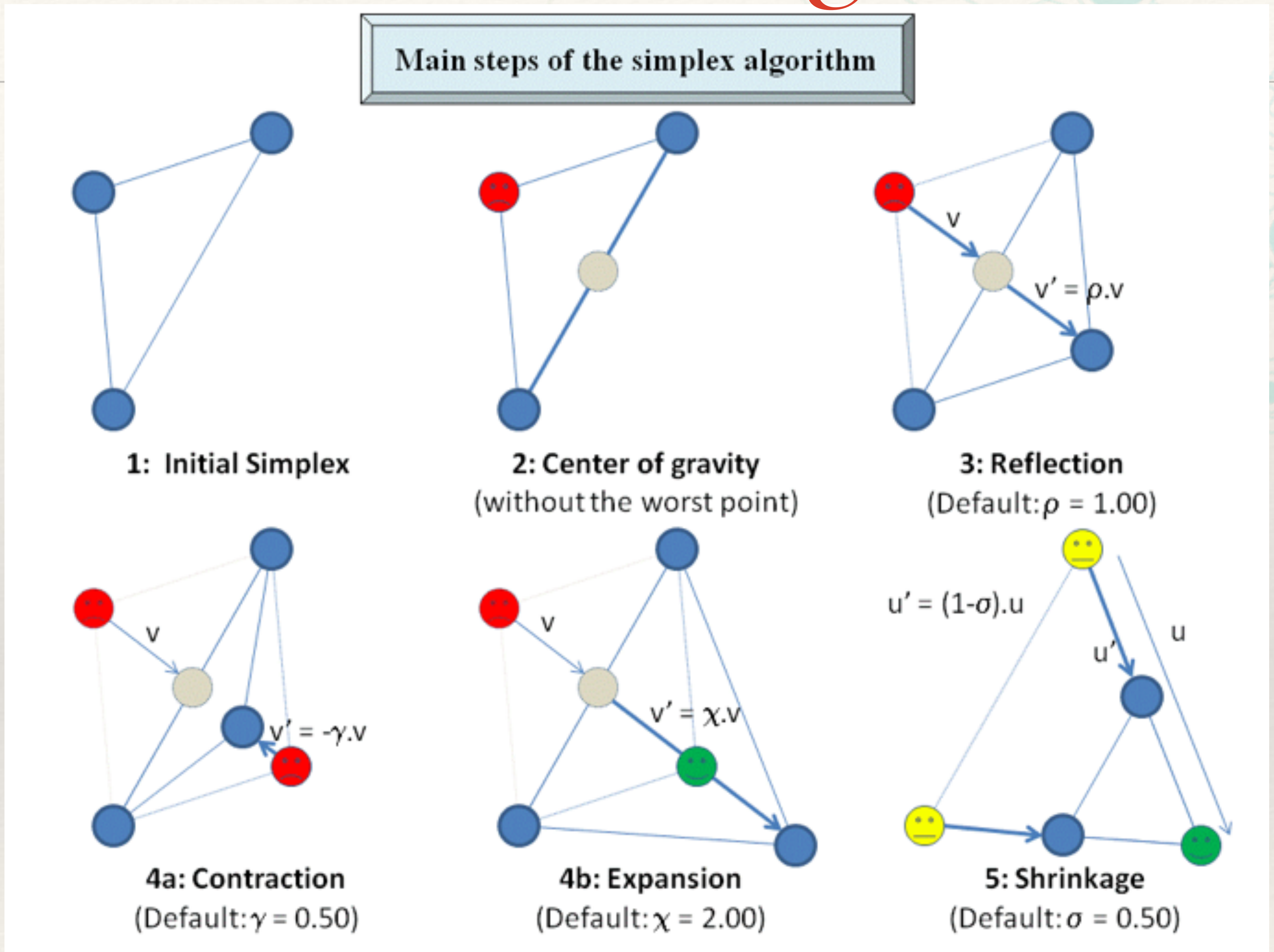
Optimization Framework

- *Search method:* Nelder-Mead simplex method (derivative-free)
- *Objective function:* average weighted-L1 distance between the simulated replicates and observed incremental epicurve

$$\min_\theta \frac{1}{R}\sum_{r=1}^{R}\left(\sum_{i=1}^{N}\alpha^{N-i}|Y_{\theta,r}(i)-X(i)|\right)$$

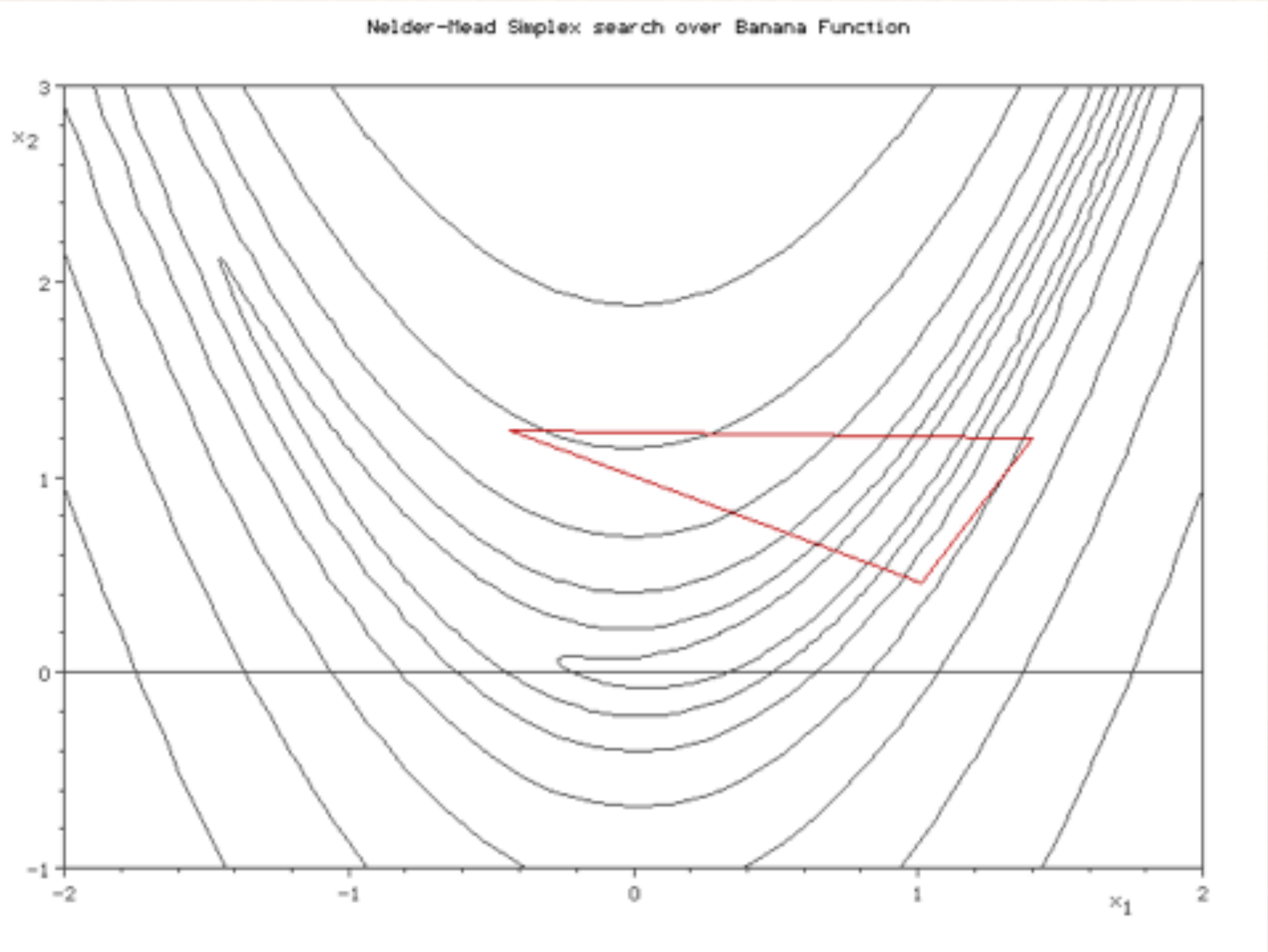


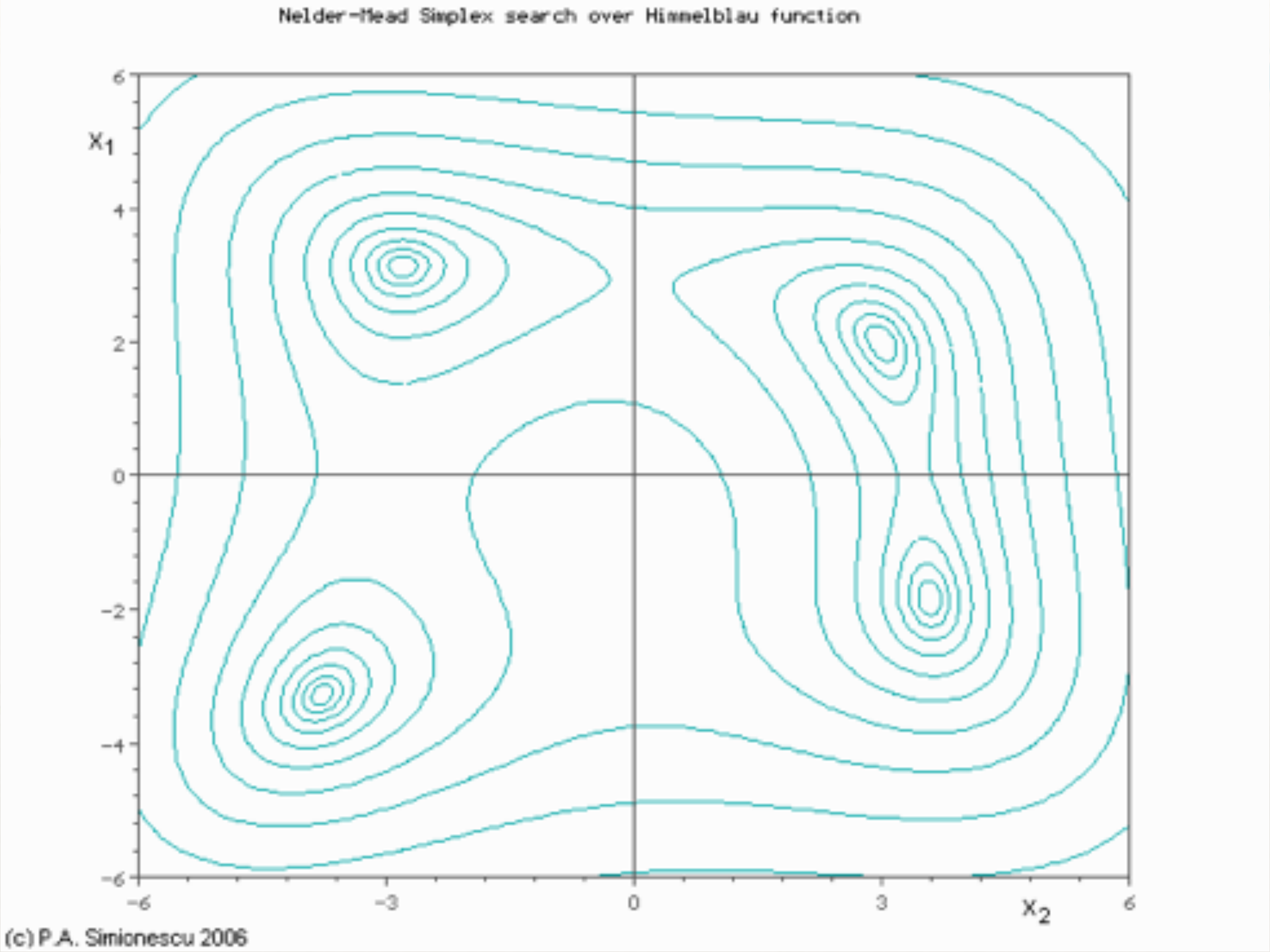
Nelder-Mead Algorithm



Nelder, John A.; R. Mead (1965). "A simplex method for function minimization". Computer Journal 7: 308–313.

Nelder-Mead Animation

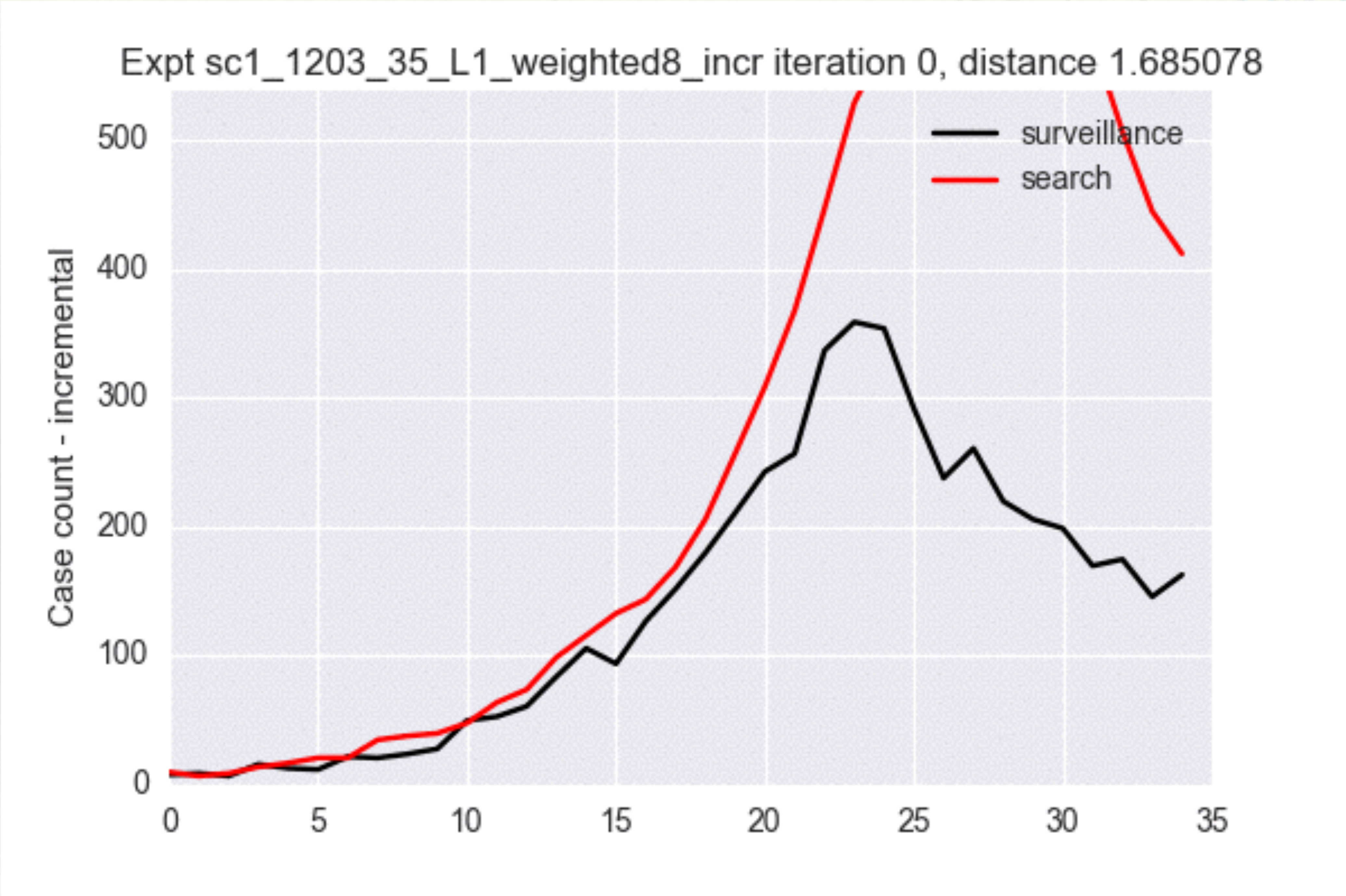




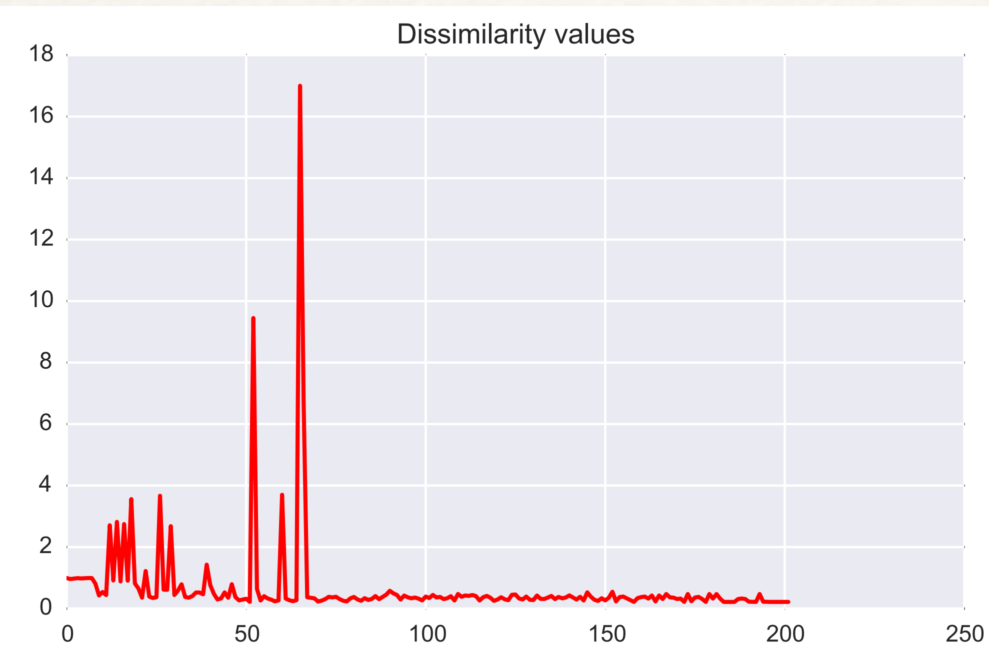
$$f(x,y) = (a-x)^2 + b(y-x^2)^2$$

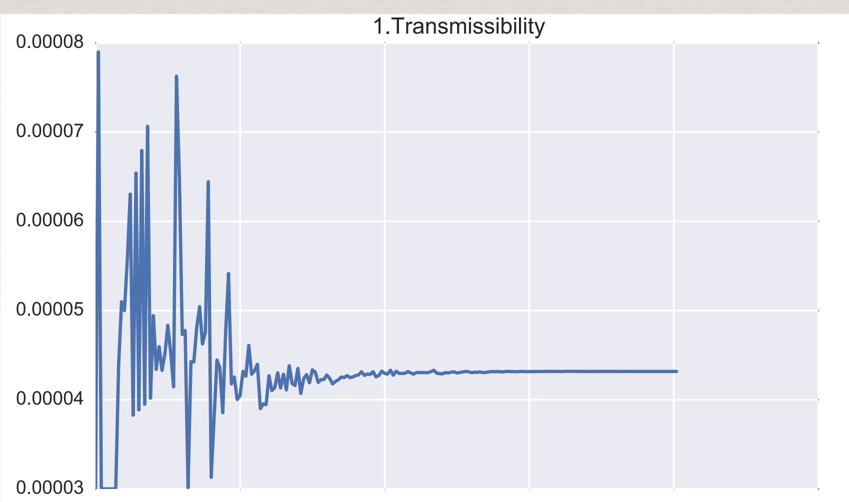
$$f(x,y) = (x^2+y-11)^2 + (x+y^2-7)^2.$$

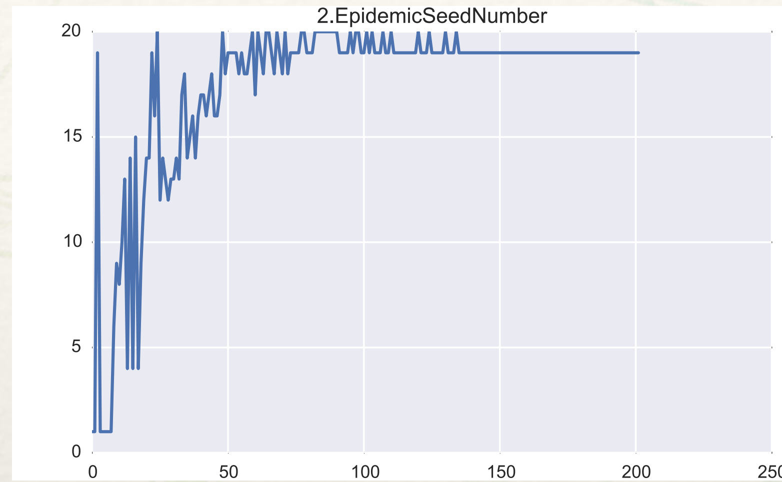
Search Evolution

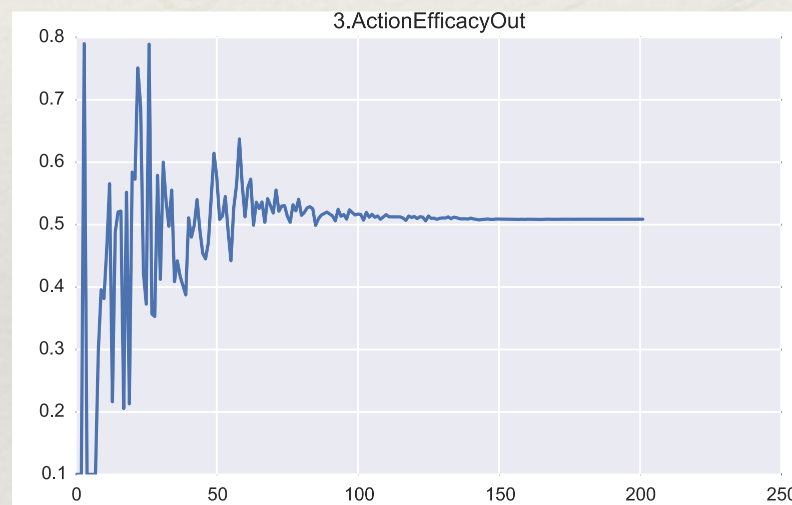


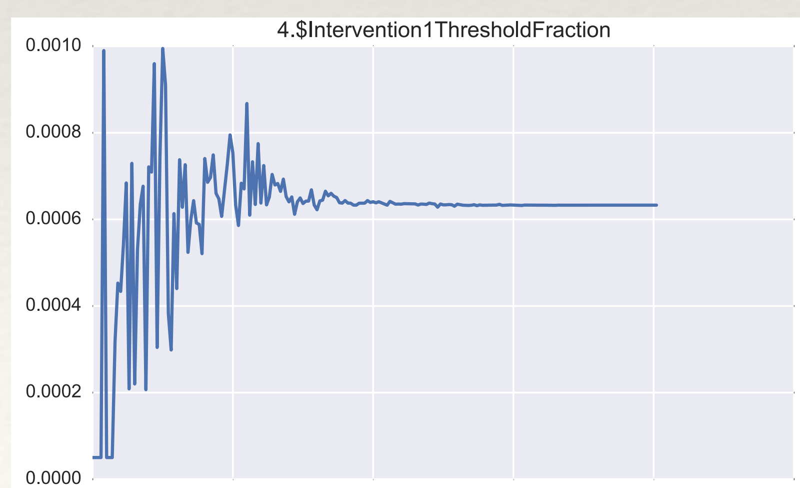
Parameter Evolution

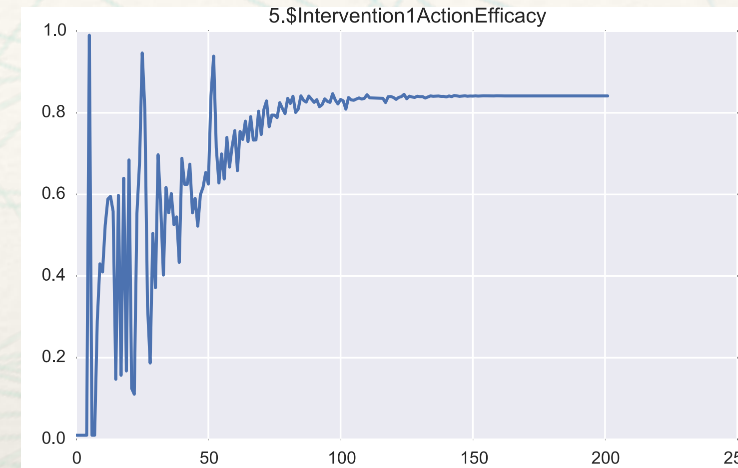


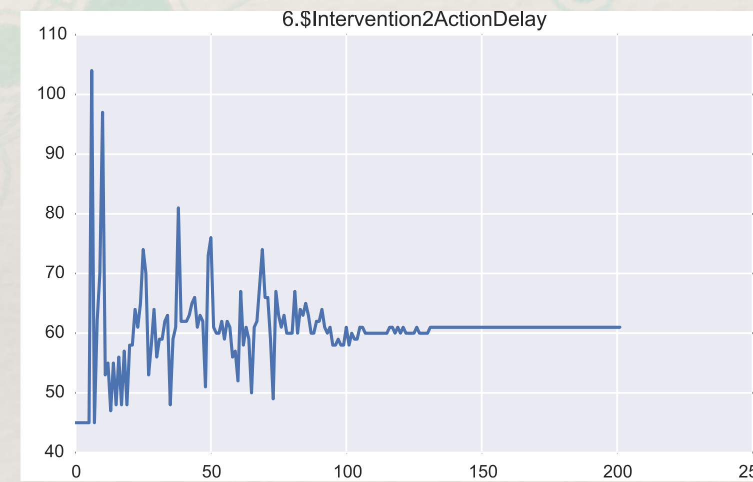


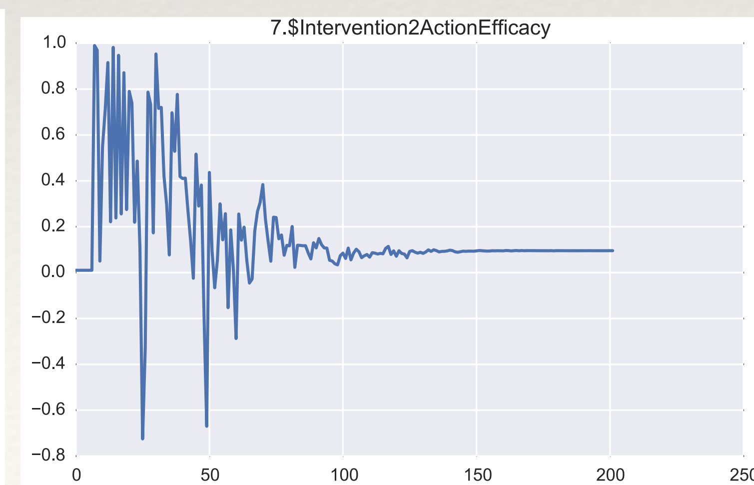






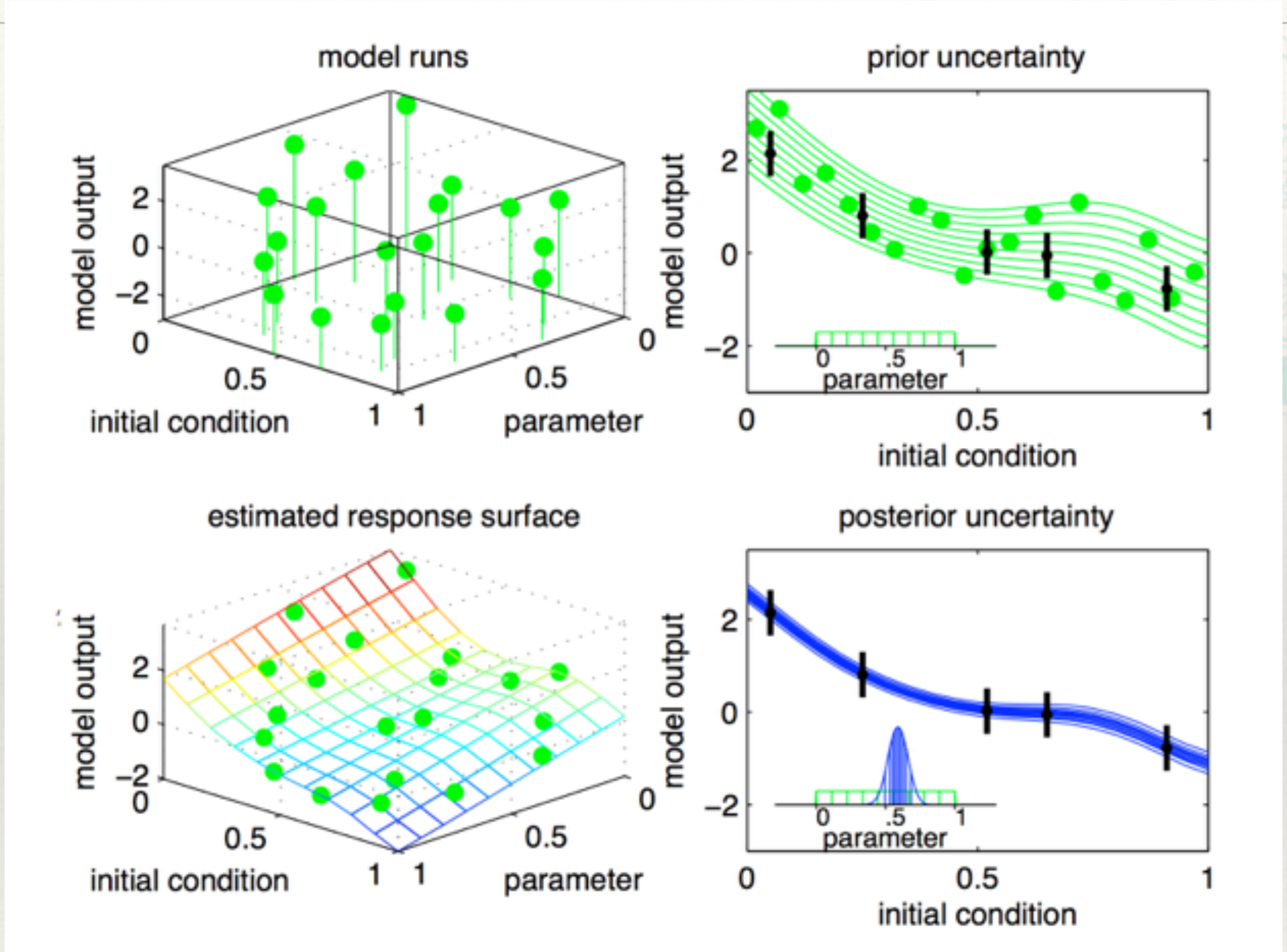






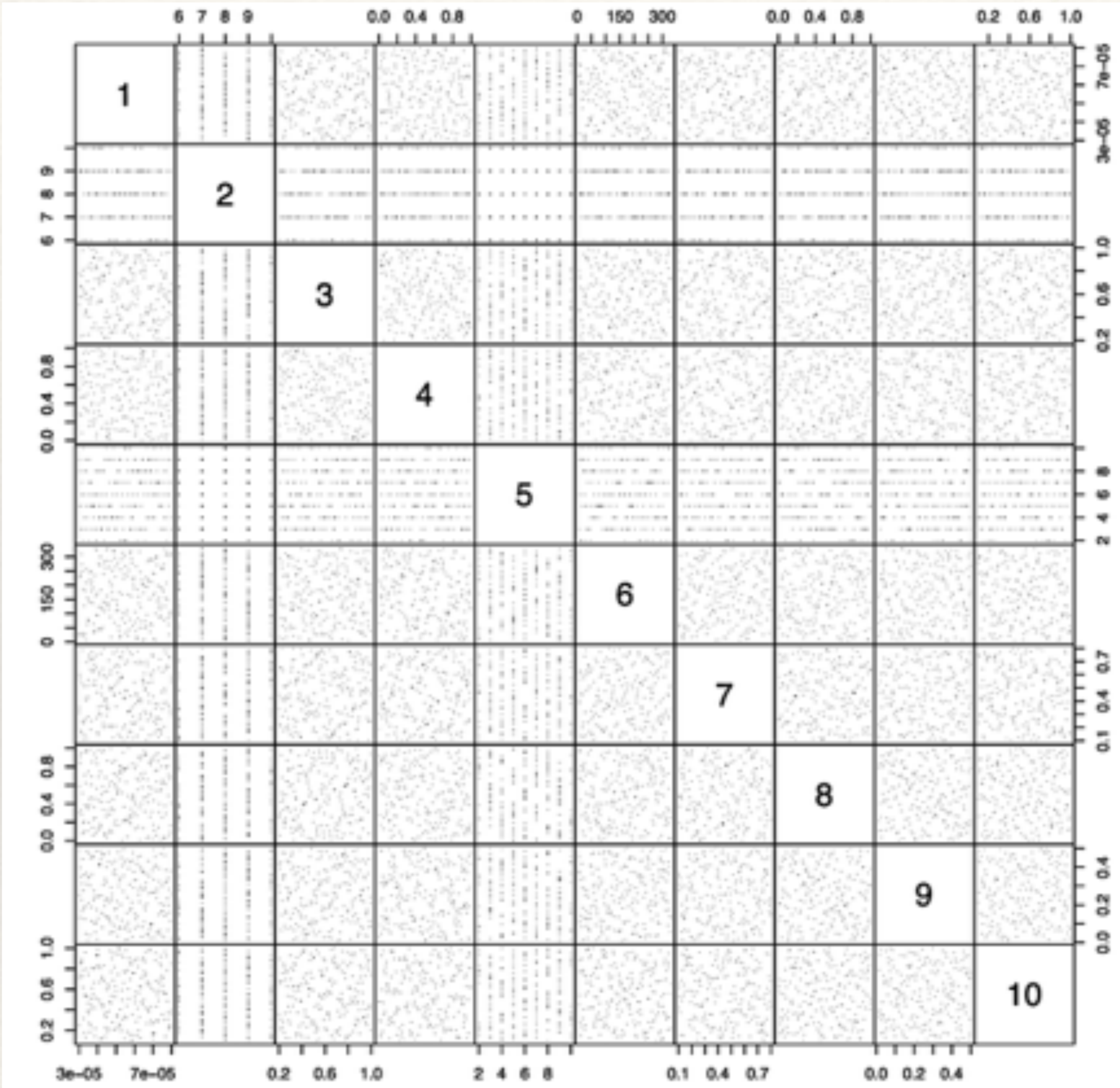
VirginiaTech
Invent the Future

Bayesian Framework

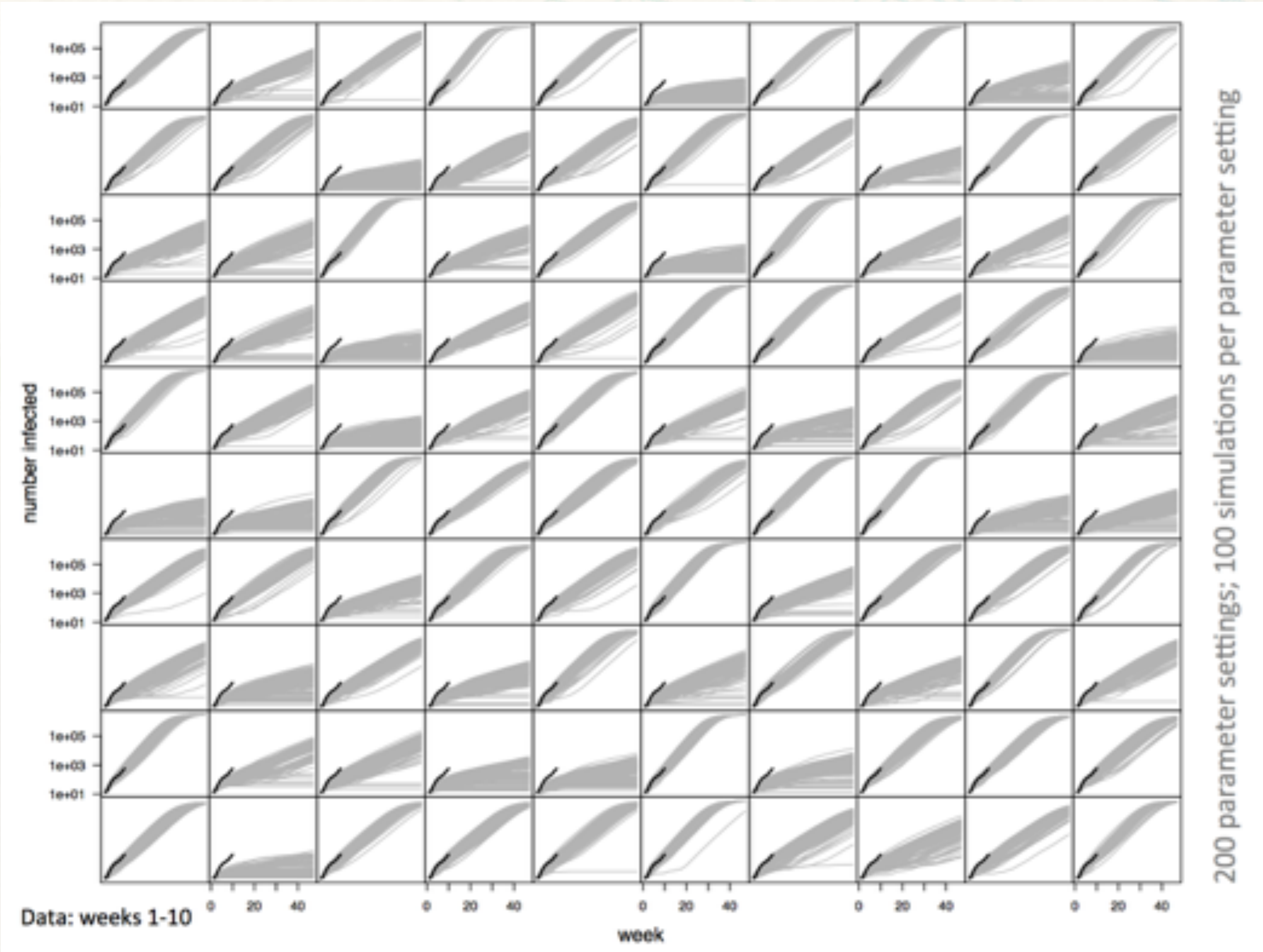


Higdon, Dave, et al. "Computer model calibration using high-dimensional output." Journal of the American Statistical Association (2012).

Bayesian Framework



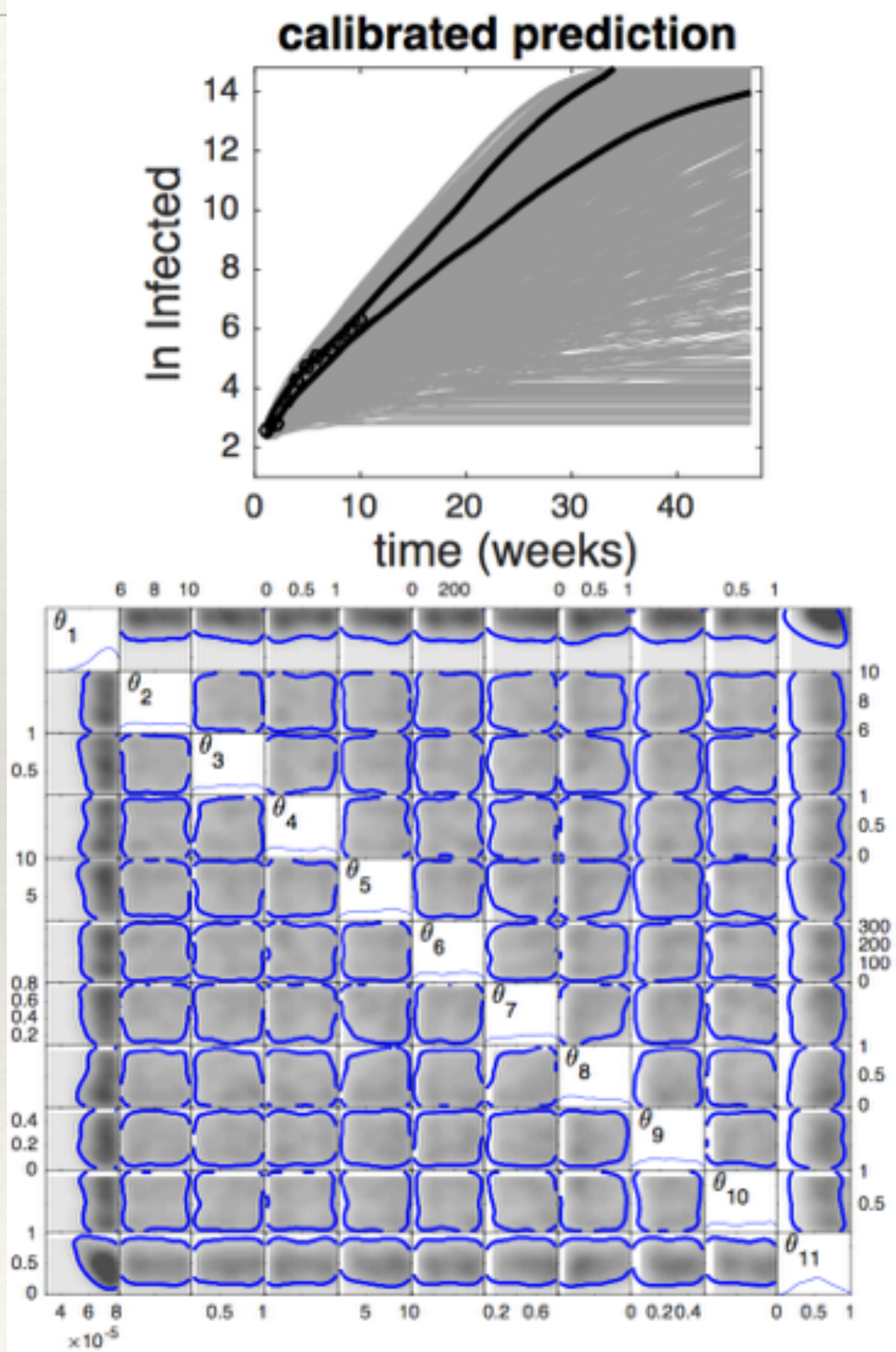
2-D space filling LHS design



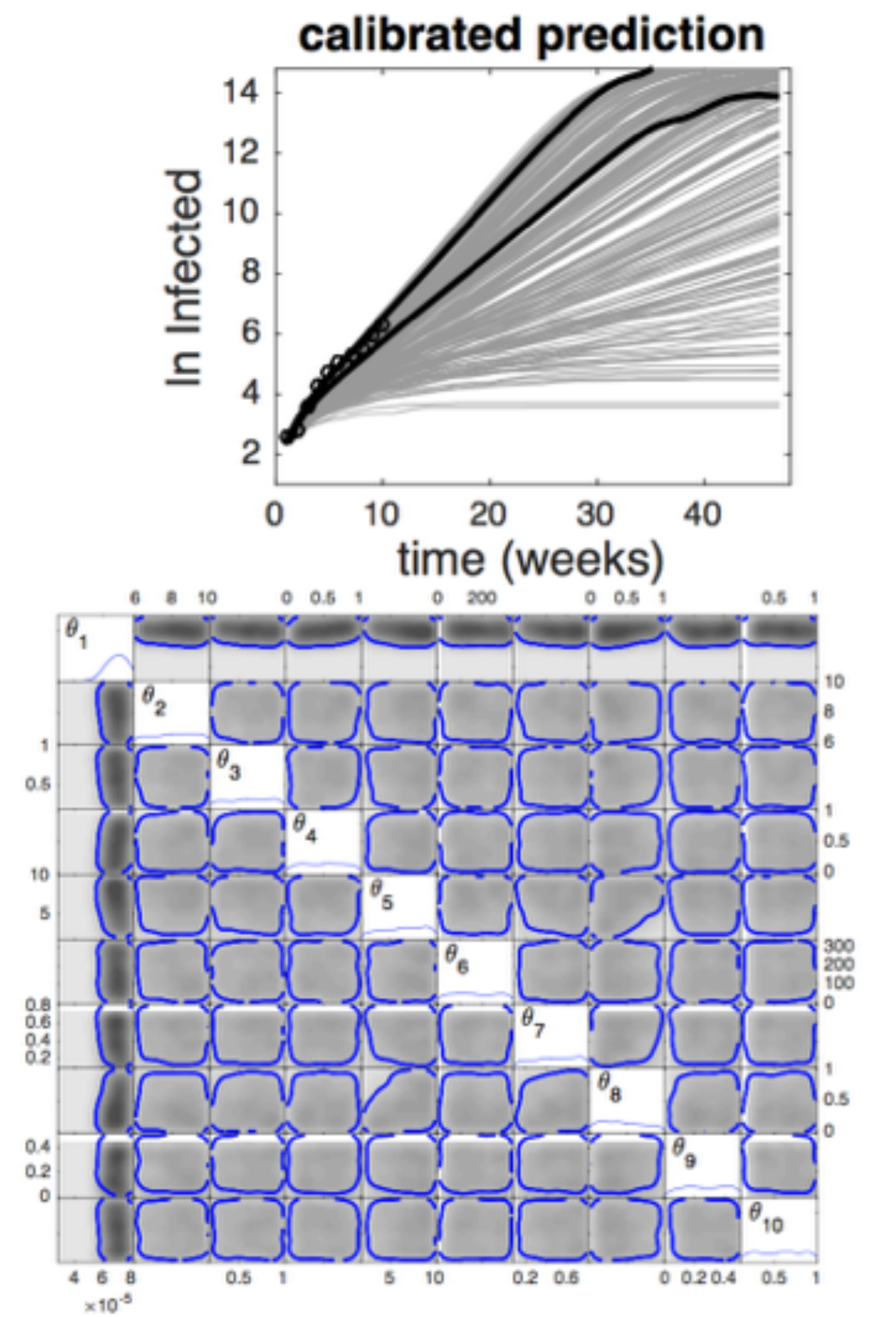
Digital Library of epicurves

Posterior Distribution

Accounting for randomness



Using median simulation



VirginiaTech
Invent the Future

NDSSL

Part IV

Ebola Forecasting Challenge

Ebola Challenge

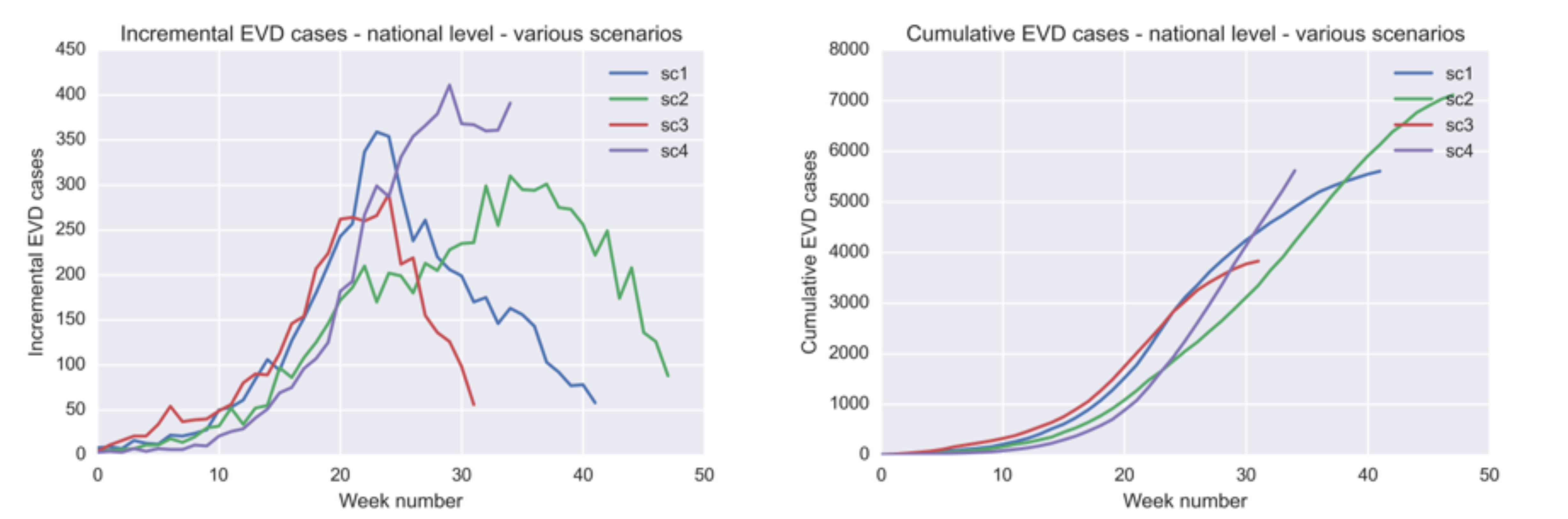
- *Organizer:* **RAPIDD** (Research and Policy for Infectious Disease Dynamics) program, NIH
- *Teams*: some of whom supported the CDC/WHO efforts during 2014 outbreak
    - Virginia Tech, Columbia Univ, McMasters Univ, Univ. of Toronto, Imperial College London, Arizona State Univ.
    - Bruno Kessler Foundation, Univ. of Tokyo, Univ. of Copenhagen, London School of Tropical Medicine & Hygiene
    - Variety of models: non-parametric, SIR-type ODE models, ABM
- *Collaborators:* Northeastern University, CDC, White House OSTP Prediction WG
- *Aim:*
    - Training the modeling community in peace time to improve predictive capabilities
    - Illustrating how data quality and availability affect accuracy and reliability
- Synthetic Ebola datasets generated using an ABM calibrated for Liberia
    - *Merler et al., Lancet Infectious Diseases, vol. 15, 204 (2015)*

Scenarios & Timepoints

- 4 scenarios - a single typical realization per scenario
  - different disease parameters, place of origin, behavioral changes and varying interventions
  - different data availability: 1 data-rich scenario, 3 fog-of-war scenarios
- ***Timeline:*** Sep'15 to Dec'15 (data released over 5 time points - 2 weeks each)

Scenario

Timepoint (weeks)	1	2	3	4
1	13	16	7	16
2	20	27	14	22
3	26	35	20	27
4	35	43	25	32
5	42	48	32	35



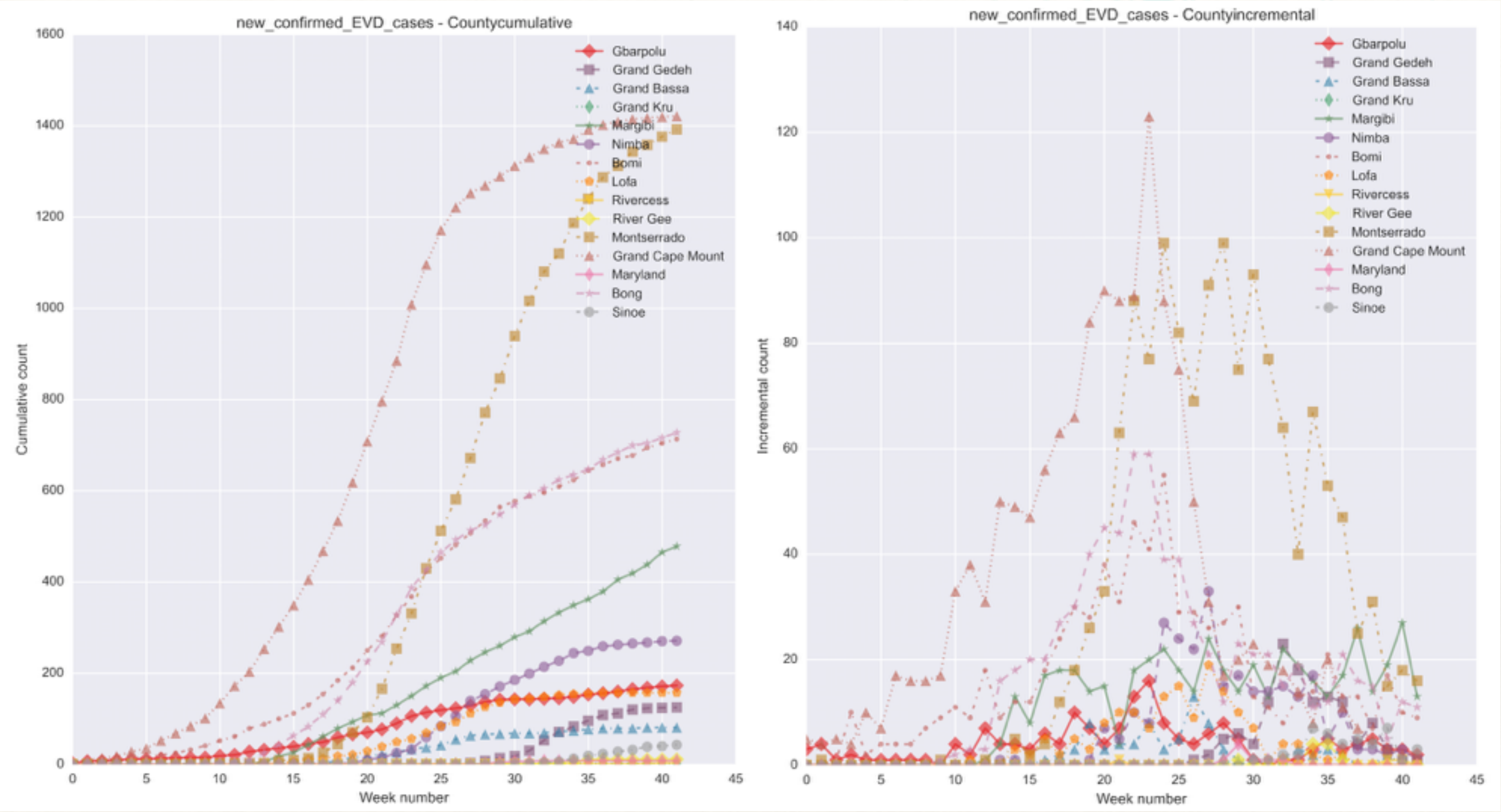
VirginiaTech
Invent the Future

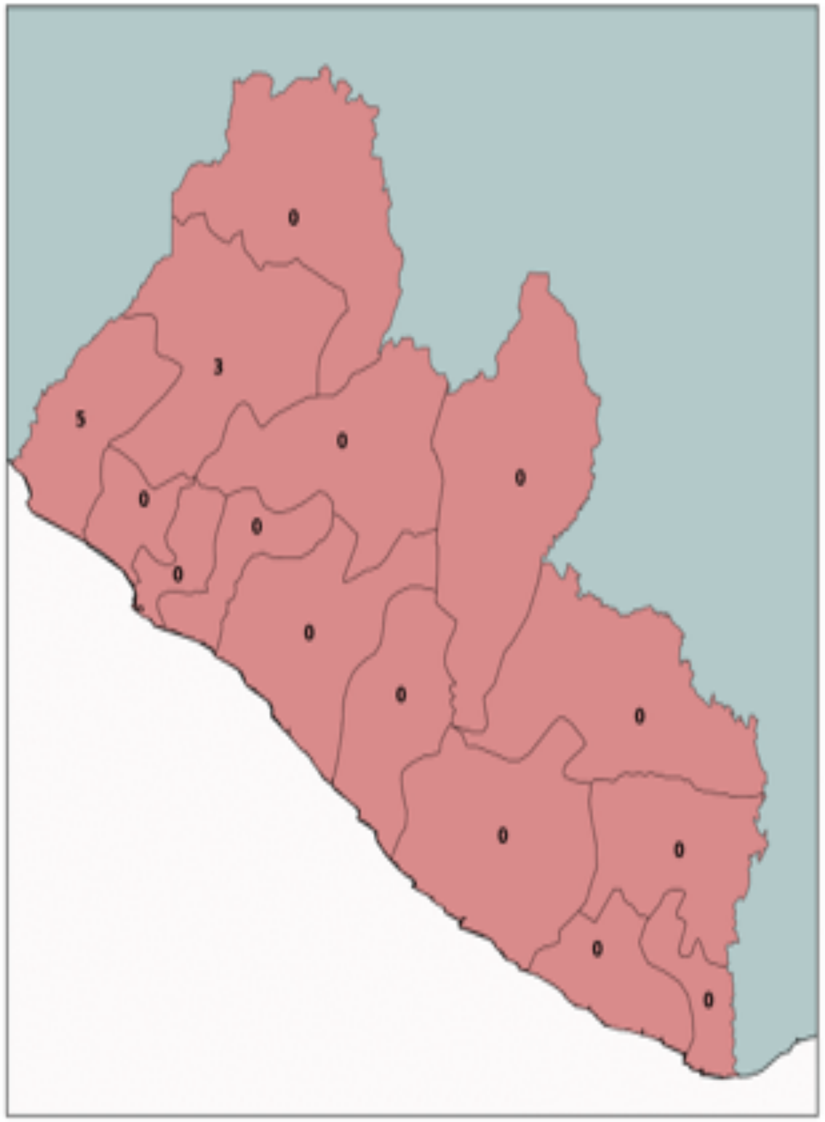
NDSSL
NETWORK DYNAMICS & SIMULATION SCIENCE LABORATORY

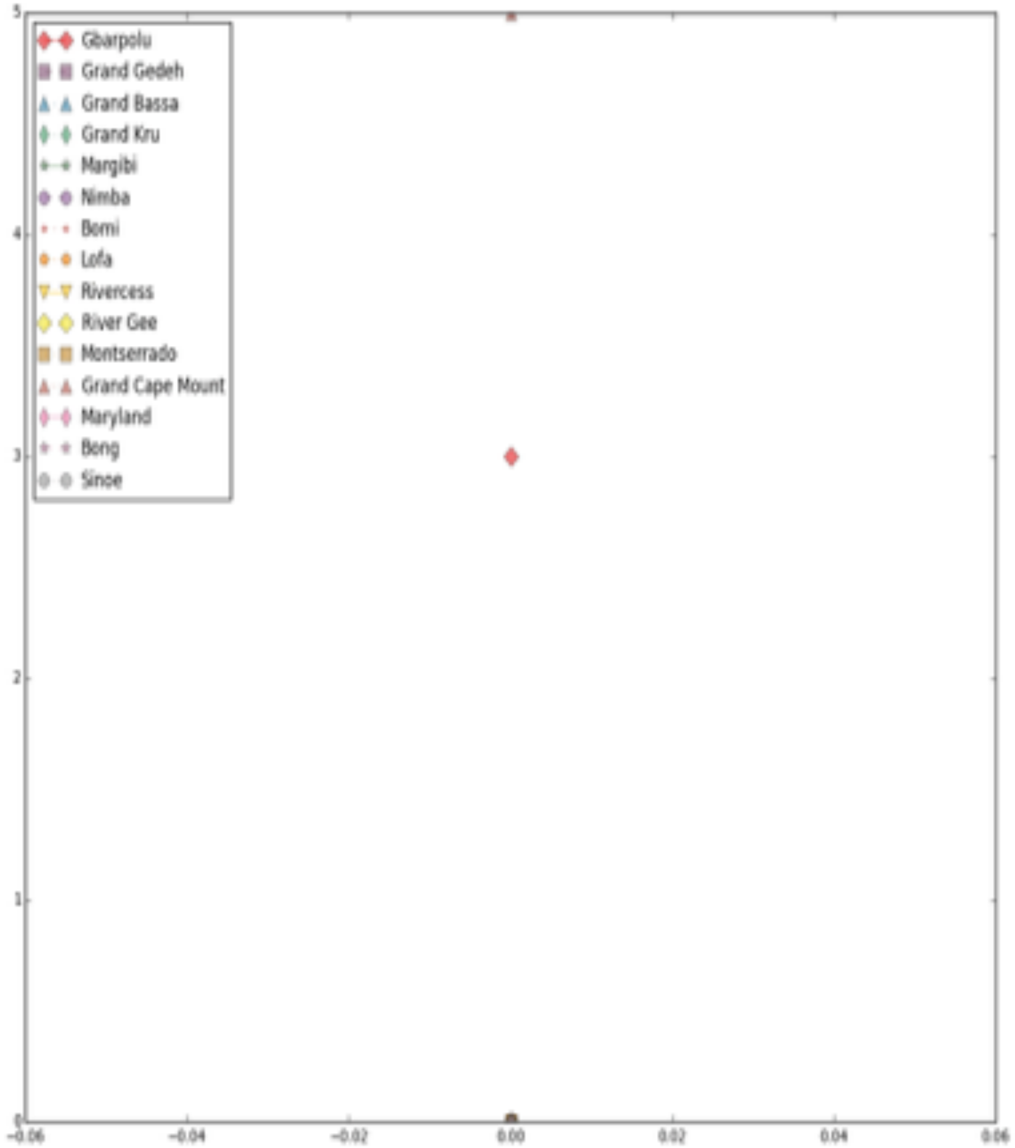
Dataset Description

- ***Epidemic curves (weekly, national)***
  - Ebola Viral Disease (EVD) cases, deaths
  - EVD cases, deaths among healthcare workers (HCW)
- Data-rich scenario
  - ***Patient database***
    - more detailed records for a subset of cases
    - Useful for constructing transmission trees
  - ***County specific epidemic curves***
- ***Situation report*** document describing interventions
  - Contact tracing intensity, Implementation of safe burials, ETU operations
- Some ***aggregate demographic statistics*** for the ABM population

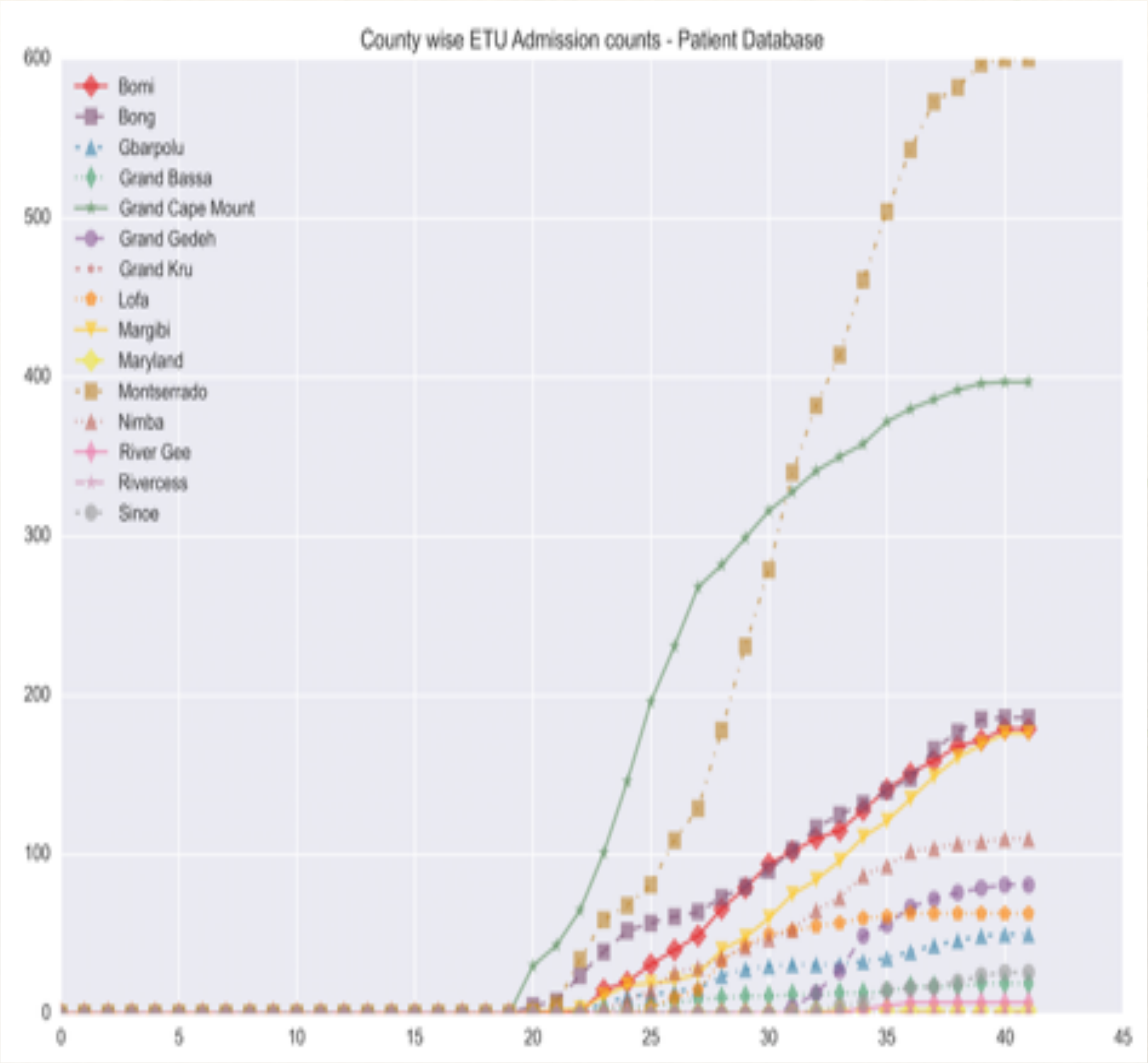
Scenario 1 - county-wise epicurves

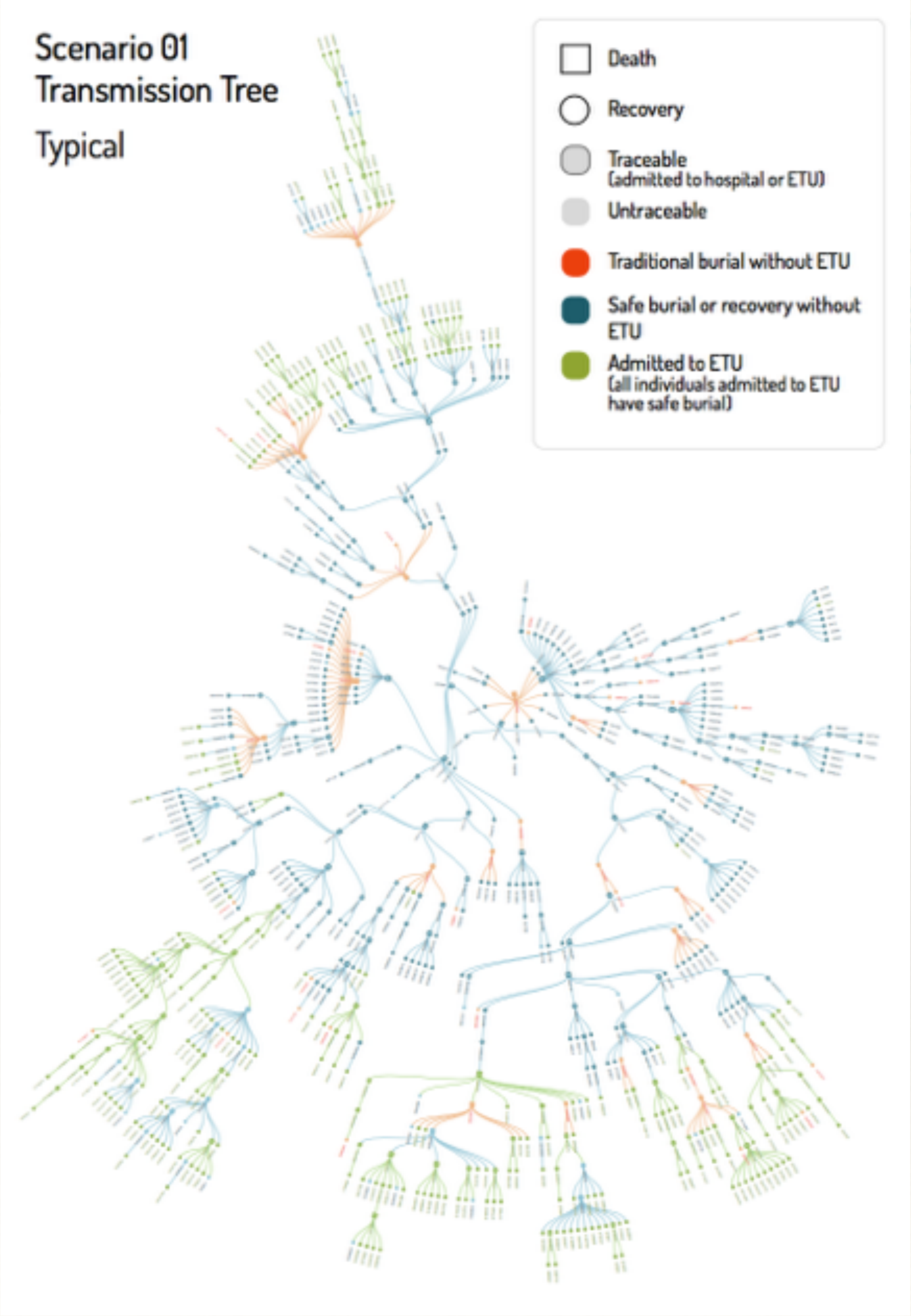



0
3
5
0
0
0
0
0
0
0
0
0
0
0
0


Gbarpolu
Grand Gedeh
Grand Bassa
Grand Kru
Margibi
Nimba
Bomi
Lofa
Rivercess
River Gee
Montserrado
Grand Cape Mount
Maryland
Bong
Sinoe
5
4
3
2
1
0
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06

Scenario 1 - Patient database





VirginiaTech
Invent the Future

NDSSL
NETWORK DYNAMICS & SIMULATION SCIENCE LABORATORY

Scenario 1
Timepoint 5
Situation report

ETU occupancy

Safe burial practices

Community engagement

Contact tracing

ETU operation

Date: week 42
Scenario 01

- A total of 707 EVD confirmed and probable cases were reported from week 35 to week 42 (included) across Liberia; 50 cases were reported in Grand Cape Mount, 25 cases in Gbarpolu, 42 cases in Grand Gedeh, 7 cases in Grand Bassa, 129 cases in Margibi, 27 cases in Nimba, 90 cases in Bomi, 3 case in Lofa, 2 case in Rivercess, 5 cases in River Gee, 205 cases in Montserrado, 92 cases in Bong, 30 cases in Sinoe.
- There is consistent evidence from the field that the epidemic is steadily decreasing in most of the counties. ETUs occupancy is estimated at about 40% in most of the facilities.
- A total number of 662 deaths were reported from week 35 till week 42 across Liberia, among which 50 were in Grand Cape Mount, 19 in Gbarpolu, 50 cases in Grand Gedeh, 6 cases in Grand Bassa, 96 cases in Margibi, 36 in Nimba, 81 in Bomi, 5 in Lofa, 2 in Rivercess, 7 in River Gee, 203 in Montserrado, 78 in Bong, 29 in Sinoe.
- A total number of 9 health care workers(HCW) confirmed and probable cases were reported, 2 in Grand Gedeh, 1 in Nimba, 4 in Montserrado. 2 in Bong. 10 of new deaths of HCWs were reported among the fatalities count from week 35.
- An effective safe burial protocol was enforced from week 35 till week 42. It is estimated that close to 90% of the burial are safe and dignified.
- Community engagement and social mobilization programs as well as the distribution of home disinfection kits are continuing in several parts of the country. It is not yet possible to estimate the effect of these interventions, although it is certainly moderate.
- A total number of 2266 individuals has been contact traced from week 35 till week 42. Among which 110 were in Grand Cape Mount, 103 in Gbarpolu, 223 in Grand Gedeh, 10 in Grand Bassa, 508 in Margibi, 60 in Nimba, 346 in Bomi, 31 in Lofa, 14 in Rivercess, 19 in River Gee, 534 in Montserrado, 256 in Bong, 52 in Sinoe.
- All previous opened ETUs are still in operation with full capacity. Here is a list of new ETUs in operation:
  - An ETU at Montserrado with 95 beds and 125 HCWs has started operations at week 35.
  - An ETU at Grand Bassa with 50 beds and 65 HCWs has started operations at week 36.
- There is no plan for the opening of new ETUs in the next few weeks.

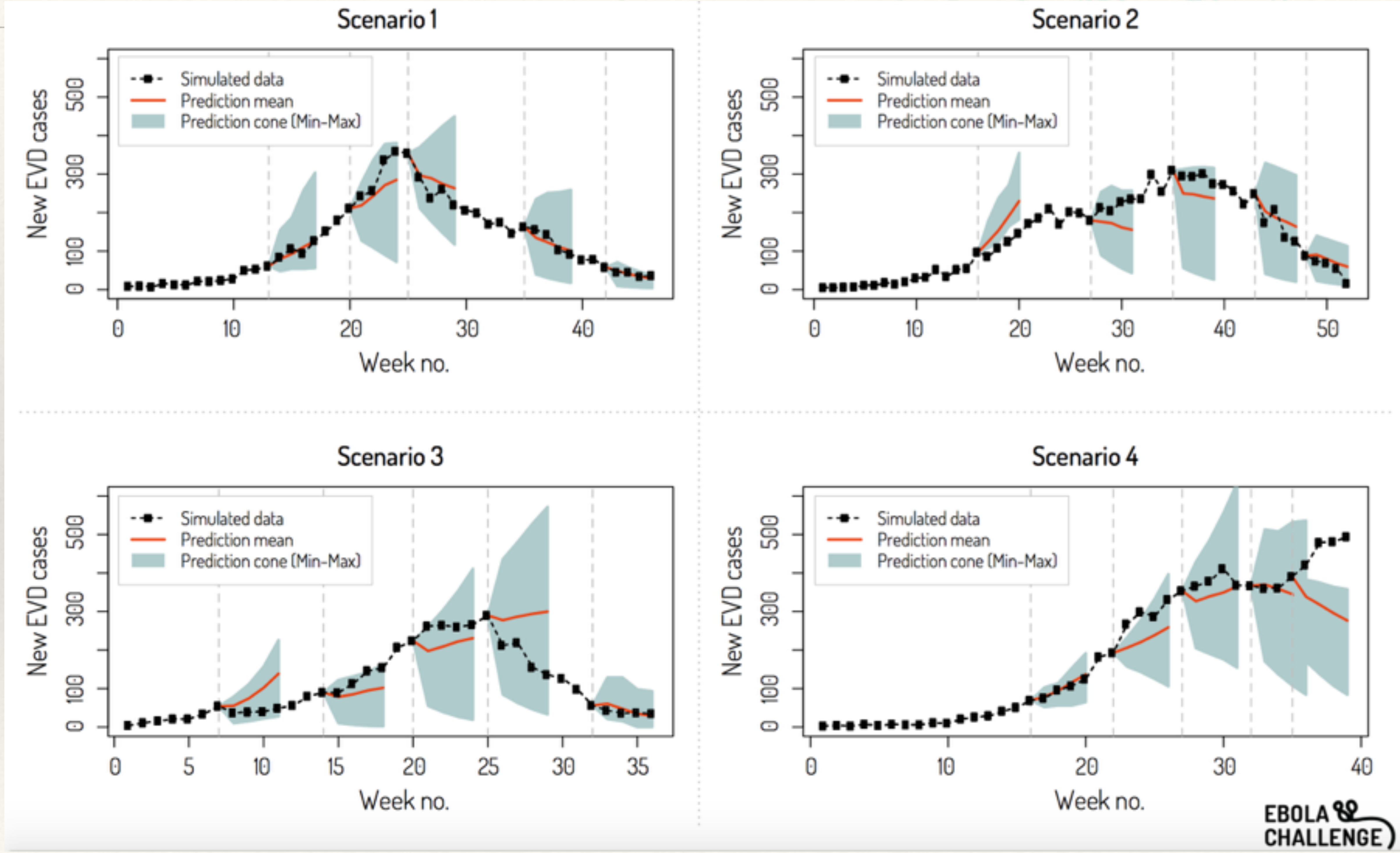
VirginiaTech
Invent the Future

NDSSL
Network Dynamics & Simulation Science Laboratory

Forecast Metrics

- ***Short-term forecasts***
  - Weekly incidences for next four weeks (national and county)
- ***Peak forecasts***
  - Peak time, Peak case count
- ***Long-term forecasts***
  - Final size of the epidemic, extinction time
- ***Disease parameters (optional)***
  - Reproduction Number, Serial interval, Case fatality rate
  - Reporting Rate
- All forecasts: Point estimates with interquartile range (25th and 75th percentile)

Ensemble predictions



Part V

Way Forward - Challenges & Opportunities

Ebola Challenge - Follow up

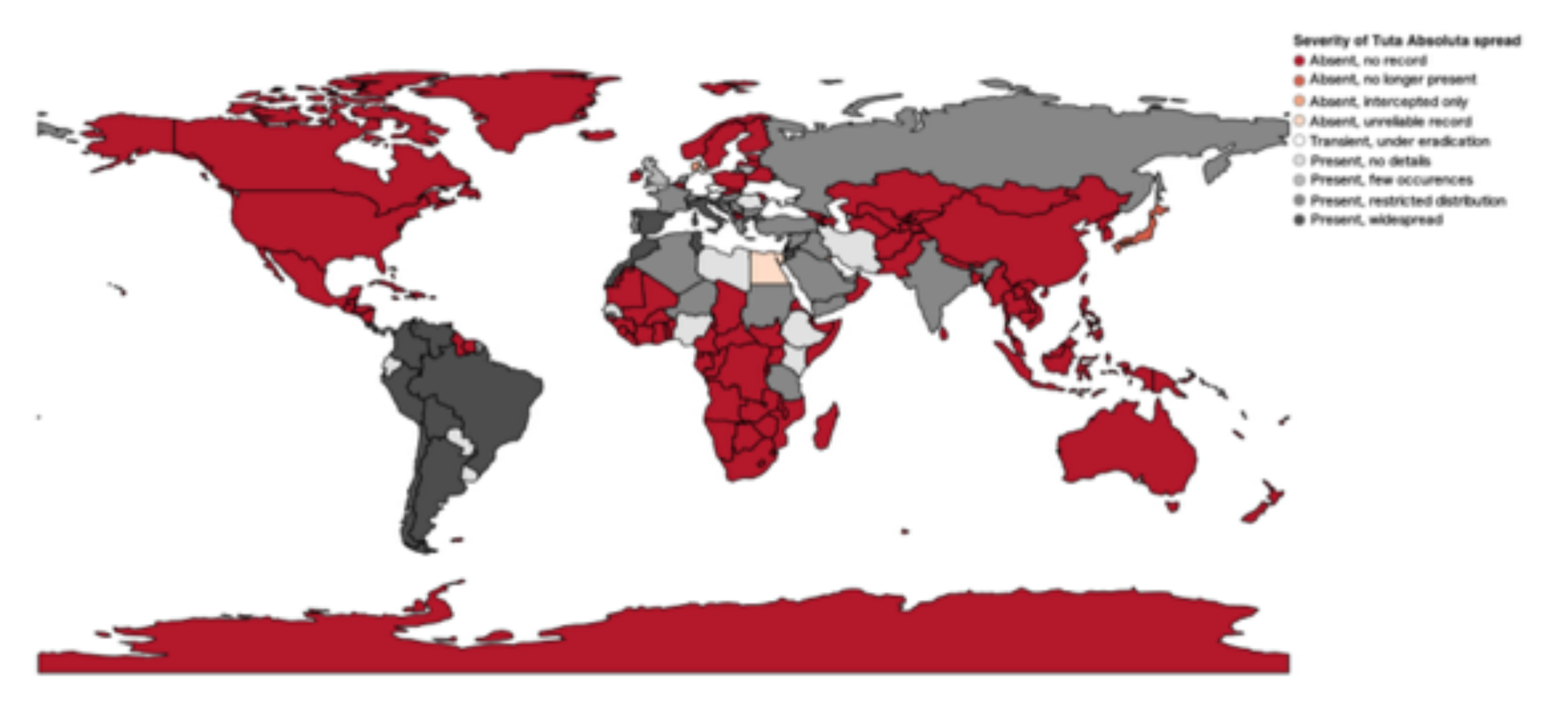
- ***Evaluation of participating methods***
  - data requirements/usage, degrees of freedom
  - Utility for policy planning
- ***Evaluation of forecasts***
  - performance across metrics, scenarios and time points
- ***Our Focus:***
  - Test other optimization/bayesian approaches
  - Improve mathematical rigor of our methods
  - Perform retrospective studies
  - Building a better ensemble of existing models

Challenges

- ***Forecasting a moving target***
  - Time varying disease and intervention parameters
  - Impact of information feedback into behavior
- ***Sensitivity analysis and Uncertainty quantification***
  - Sensitivity to noise in data, model construction
  - Model reliability versus data utility
  - Model uncertainty, parameter uncertainty, inherent stochasticity
- ***Multi-scale calibration***
  - Calibrating to multiple ground truth at various spatio-temporal scales
- ***Fundamental limits to forecasting?***
  - How far ahead can we forecast reliably?
  - Computational complexity?

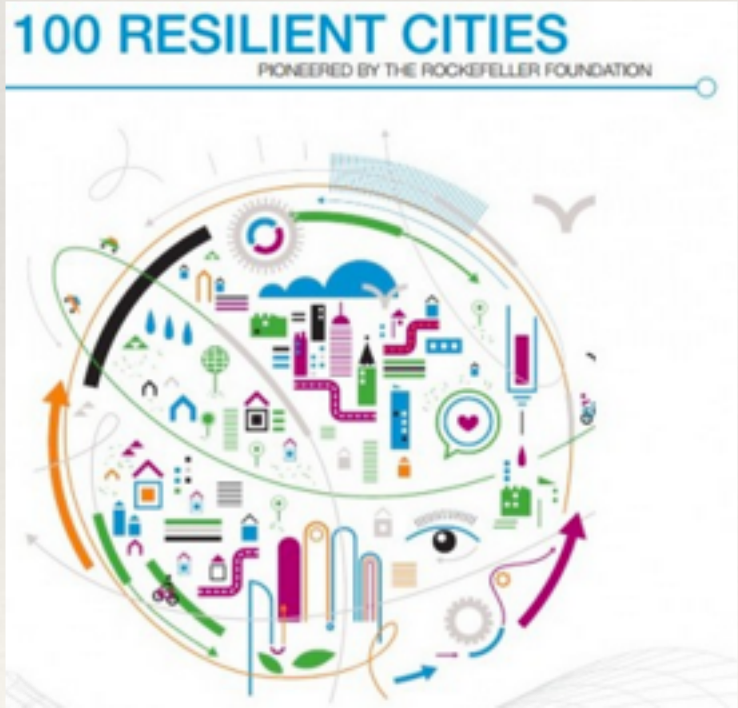
Beyond Epidemics

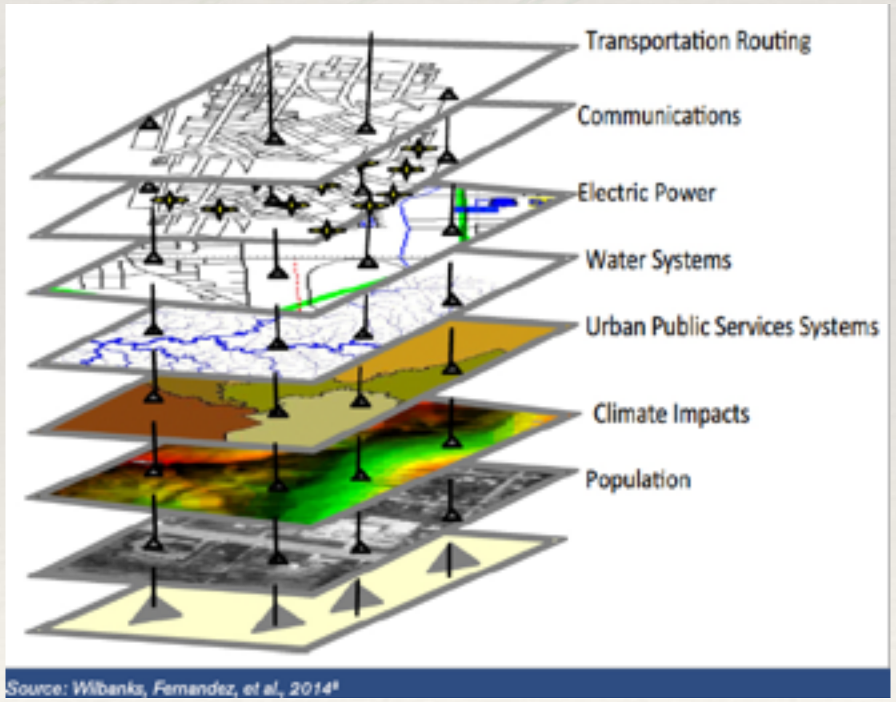
- ***Food Security***
  - Modeling global spread of invasive pests (e.g. Tomato leaf-miner)
  - Combine biotic, abiotic factors with human-induced pathways
  - Develop a comprehensive pest risk map, and integrated pest management strategies



Beyond Epidemics

- *Urban resilience*
  - Modeling city's response to natural disasters, civil unrest, terror threats
  - Incorporate infrastructure networks - transport, power grid, communication, etc.
  - Designing adaptive strategies for overall resilience





Concluding Remarks

Conclusions

- Modeling ***realistic complex socio-technical systems*** is of immediate and great importance
- Real world systems are ***really messy***
- Questions are often ***ill-defined***; Several ***knowledge and data gaps***
- Ability to ***synthesize and simulate in-silico societies quickly*** is very valuable for policy planners
- Very interesting ***computational and mathematical challenges*** emerge from handling real-world systems at this scale
- Calls for ***greater collaboration cutting across domains*** with a stress on ***addressing the problem***

Acknowledgments

Sponsors

Ebola Challenge Team

Srini Venkat
Computer Science

Bryan Lewis
Public Health

Jiangzhuo Chen
Software Systems

Dave Higdon
Statistics

Anil Vullikanti
Computer Science

Madhav Marathe
Computer Science

NDSSL Staff

VirginiaTech
Invent the Future

NDSSL
Network Dynamics & Simulation Science Laboratory

Questions?

vsriniv@vt.edu
bi.vt.edu/ndssl

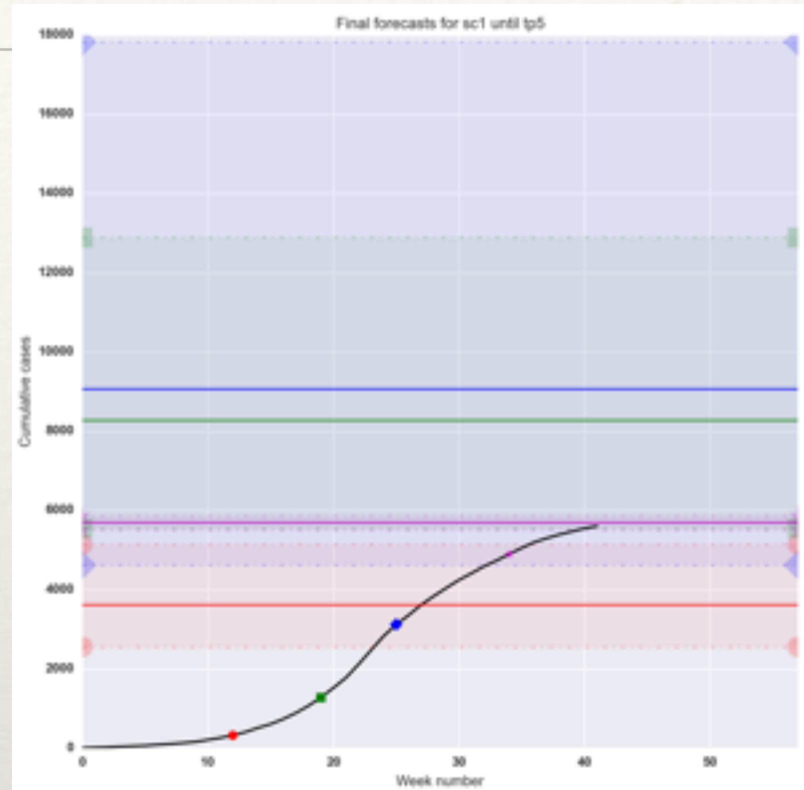
VirginiaTech
Invent the Future

NDSSL
NETWORK DYNAMICS & SIMULATION SCIENCE LABORATORY

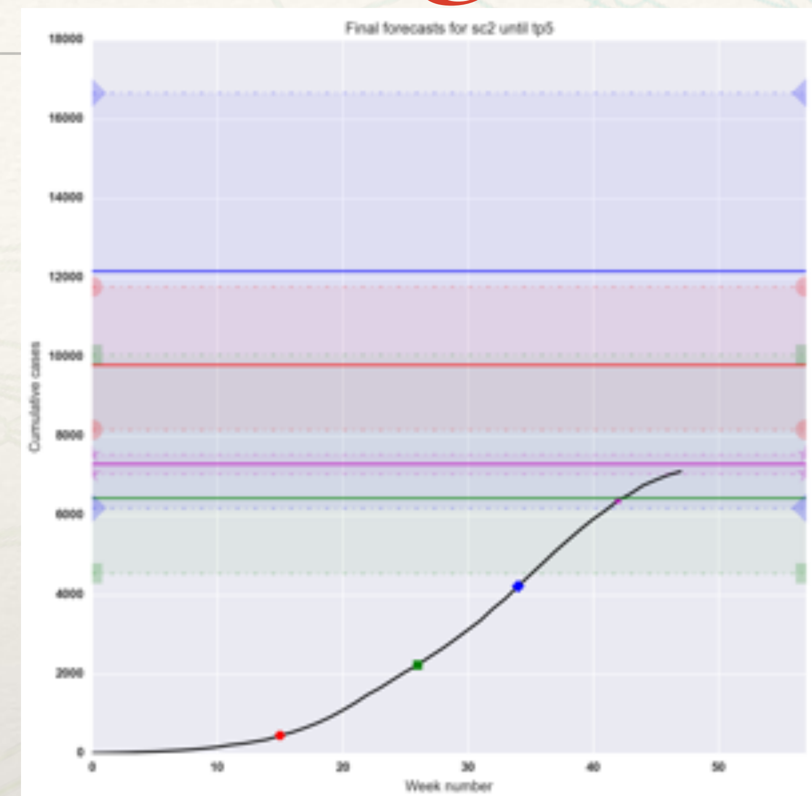
Backup slides

Our Forecasts - Long-term

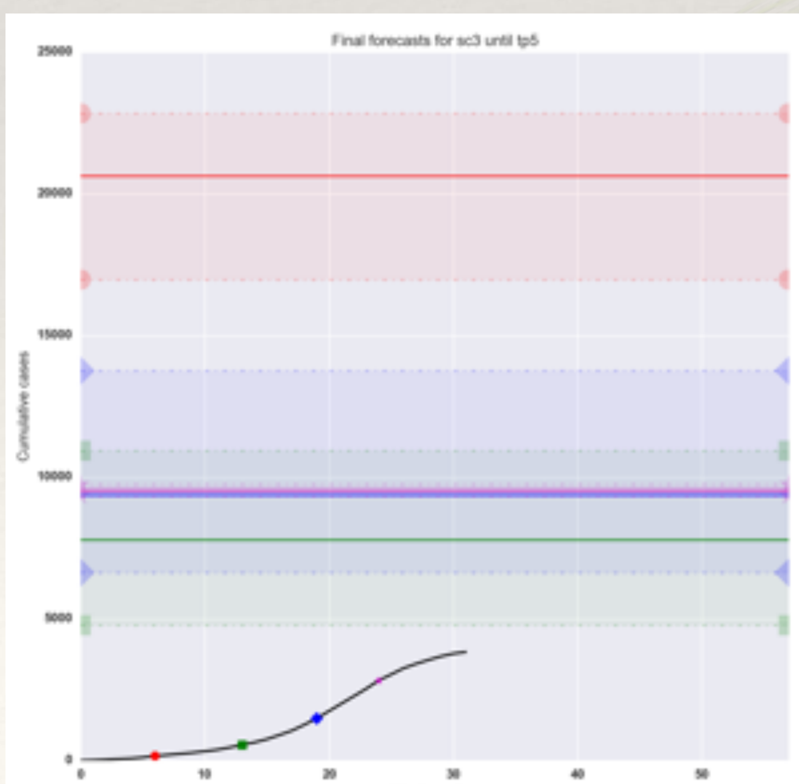
Sc. 1



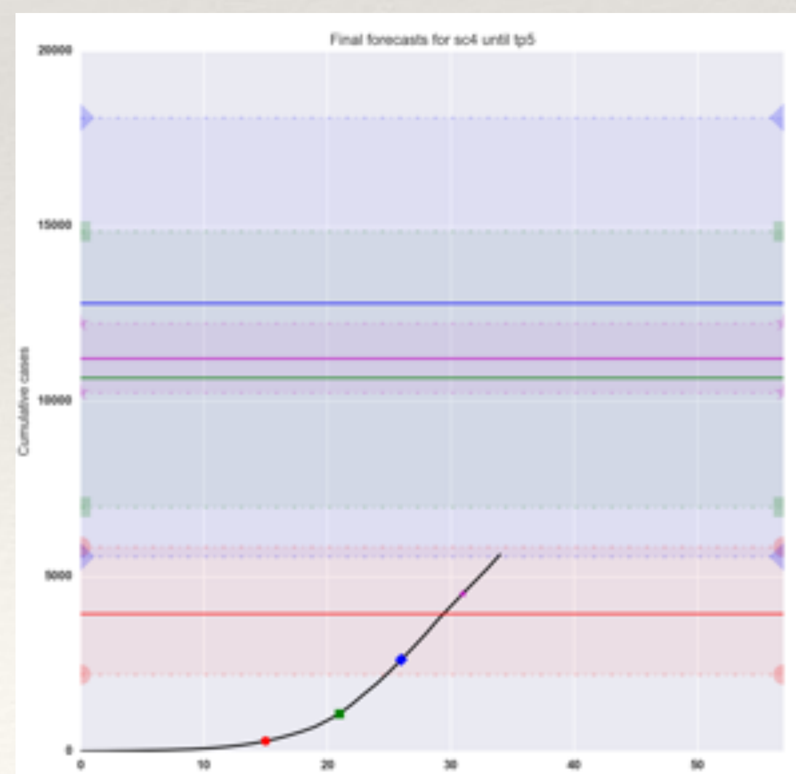
Sc. 2



Sc. 3



Sc. 4

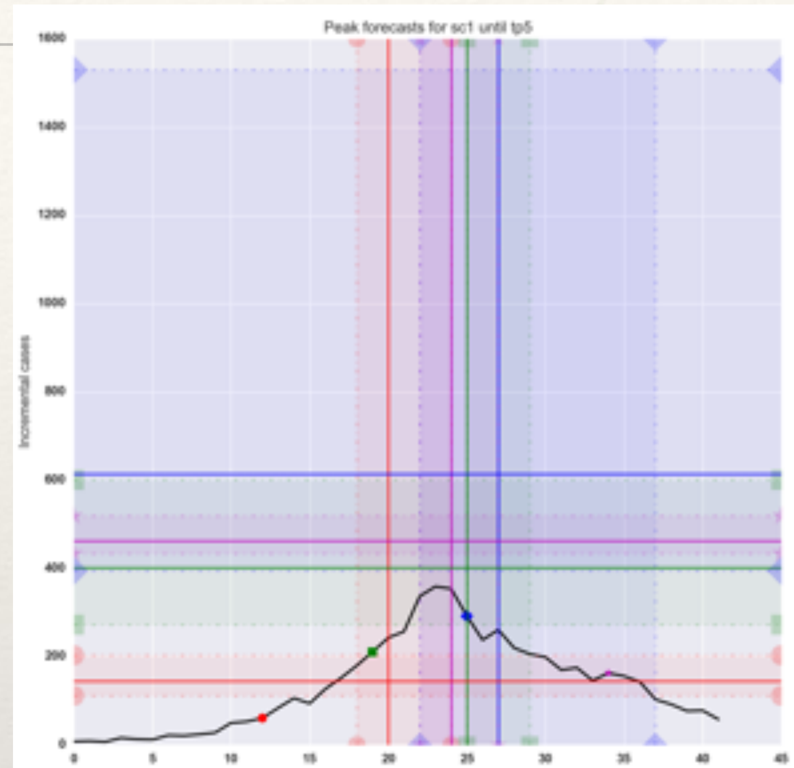


VirginiaTech
Invent the Future

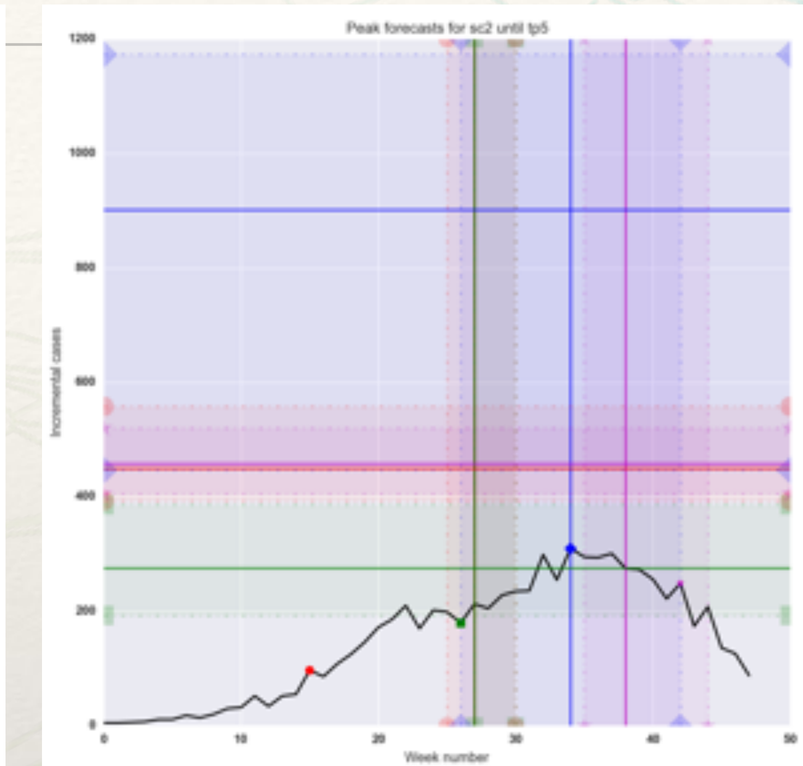
NDSSL

Our Forecasts - Peak

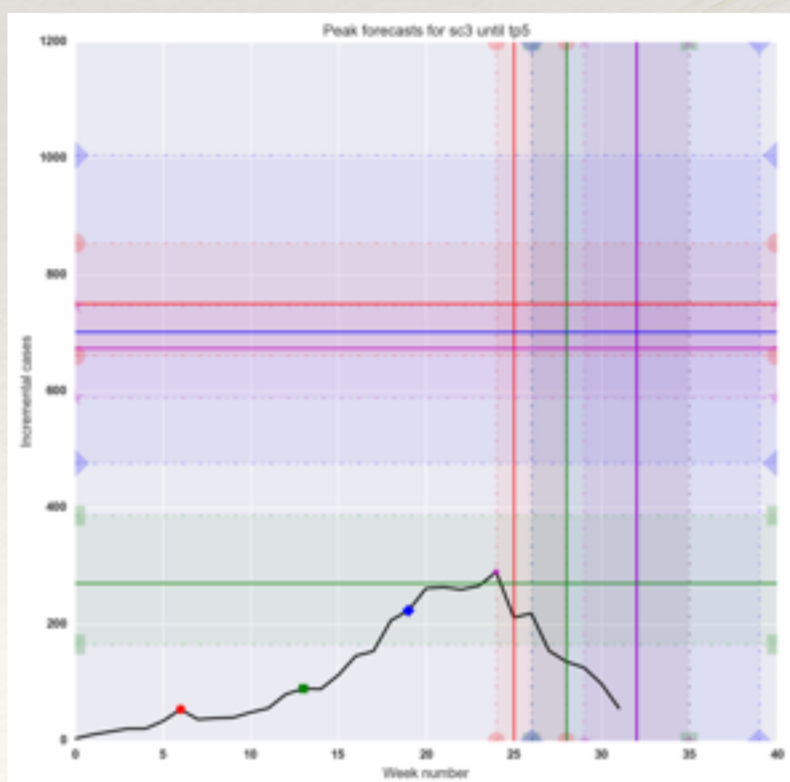
Sc. 1



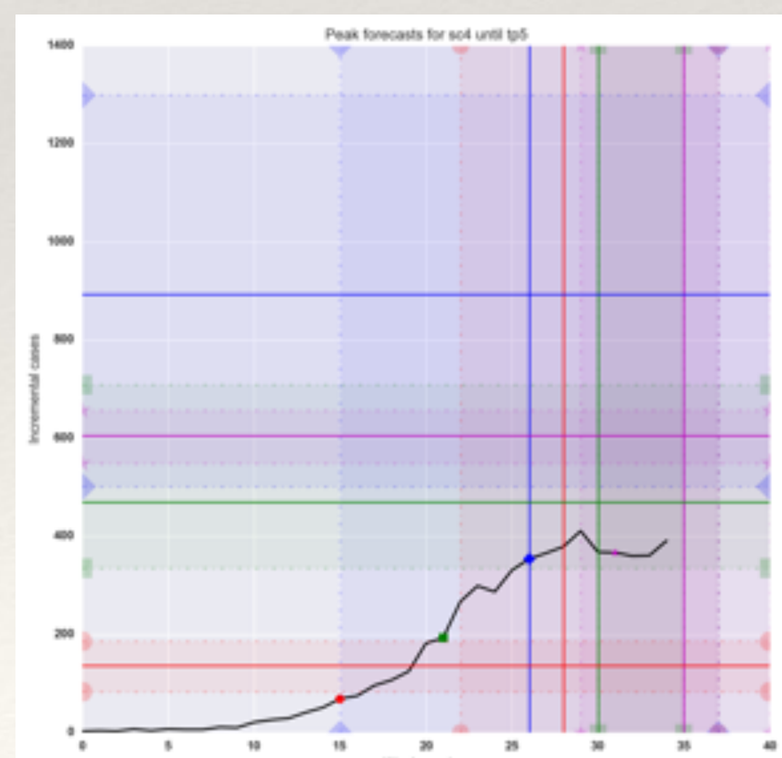
Sc. 2



Sc. 3



Sc. 4

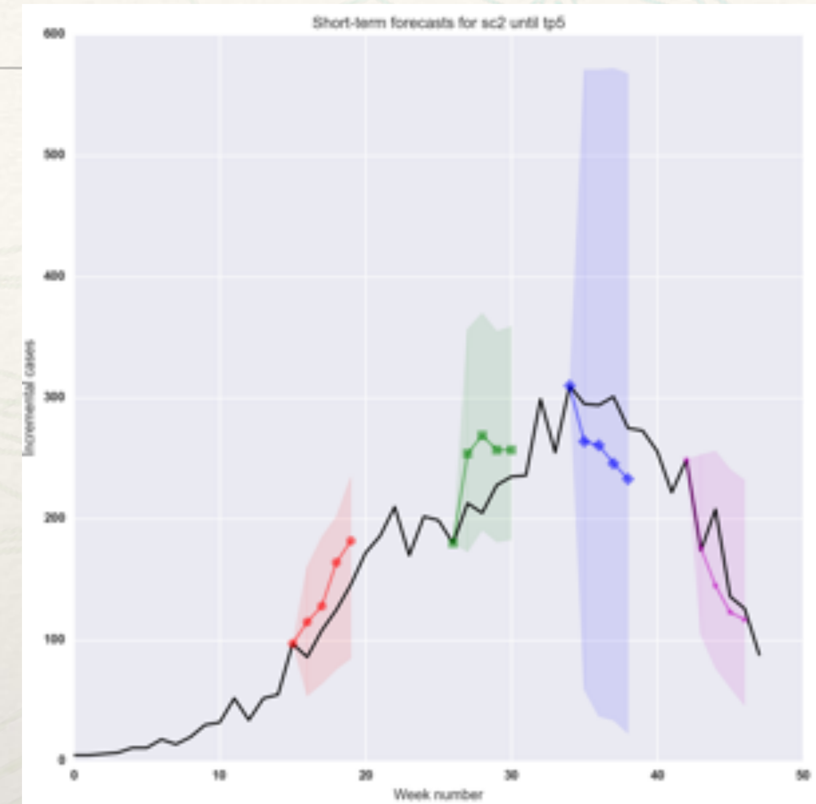


VirginiaTech
Invent the Future

NDSSL
NETWORK DYNAMICS & SIMULATION SCIENCE LABORATORY

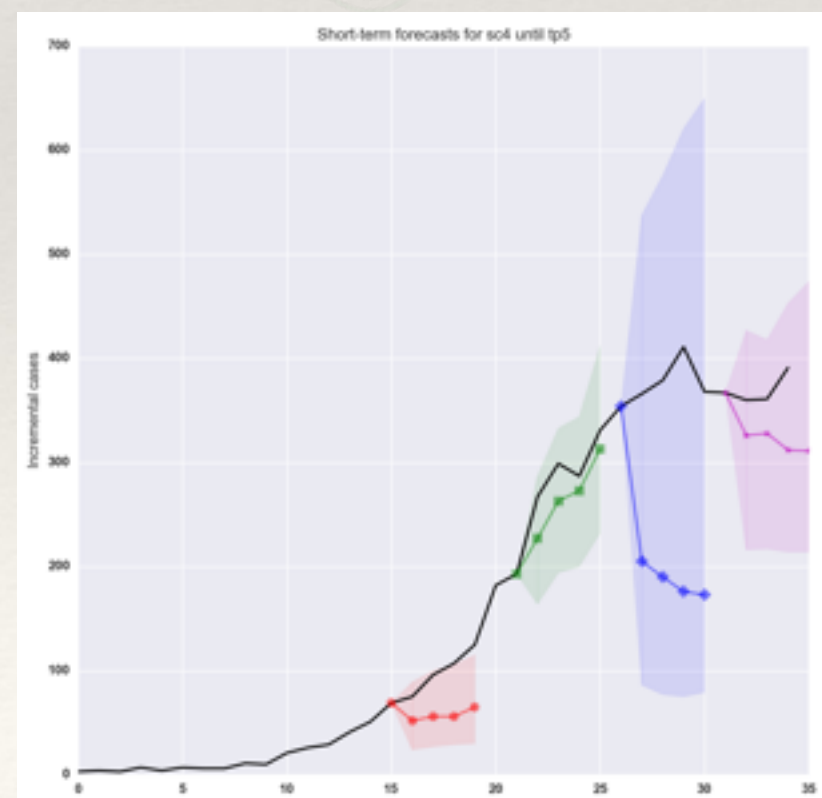
Our Forecasts - Short-term

Sc. 1



Sc. 2

Sc. 3



Sc. 4