CENTER FOR INFERENCE & DYNAMICS OF INFECTIOUS DISEASES

Northeastern

Data Science and Epidemiology: more than forecast

Alessandro Vespignani

@alexvespi

MOBS LAB

LABORATORY FOR THE MODELING OF BIOLOGICAL AND SOCIO-TECHNICAL SYSTEMS

Mathematical epidemiology

"I simply wish that, in a matter which so closely concerns the wellbeing of the human race, no decision shall be made without all the knowledge which a little analysis and calculation can provide"
Daniel Bernoulli ~1760

MATHEMATICAL -> COMPUTATIONAL

Numerical Weather models

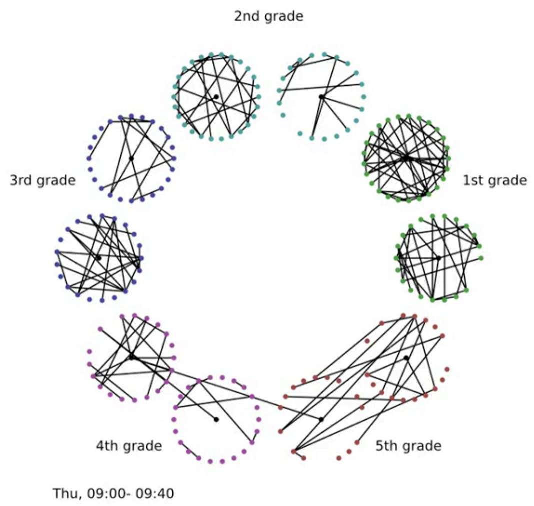
- 1920 Richardson integrate manually equations of the atmosphere
- 1950 First numerical weather forecast (24h computation for a 24h forecast)
- 1955 Numerical weather prediction models became operational by the USWB
- 2000 Government and Commercial entities routinely forecast up to three weeks

Numerical Epidemic models

- 1930 Reed-Frost define a simple chain binomial model that they integrate with a “sandbox’ computer
- 1952 First Reed-Frost numerical implementation
- 1980-2000 progress toward the definition of large-scale individual models
- 2005 Large scale agent-based models early approaches
- 2015 Operational tests

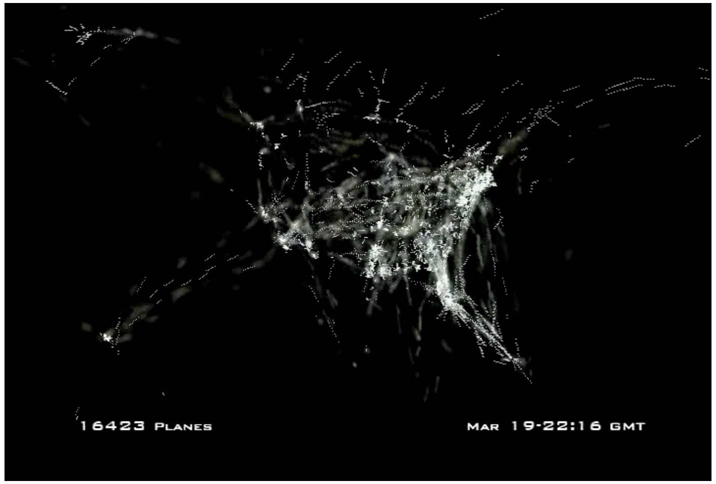
SHIFTING GEAR: DATA AVAILABILITY

Human interactions/ contact networks



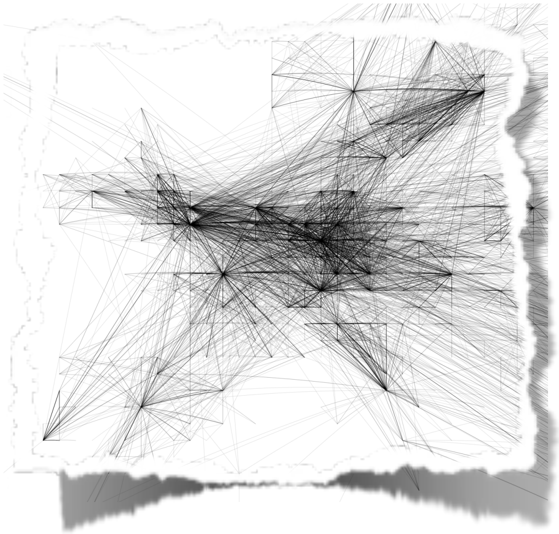
Within school contact patterns (@Sociopatterns)

Mobility and epidemic spreding



Multiscale integration of mobility networks in the analysis of potentially pandemic pathogens spread.

Networks heterogeneity and complexity



Hubs, community, clustering, heavy tails, ...

Novel digital data streams

Active data collection

INFLUWEB

flusurvey

grippenet.fr

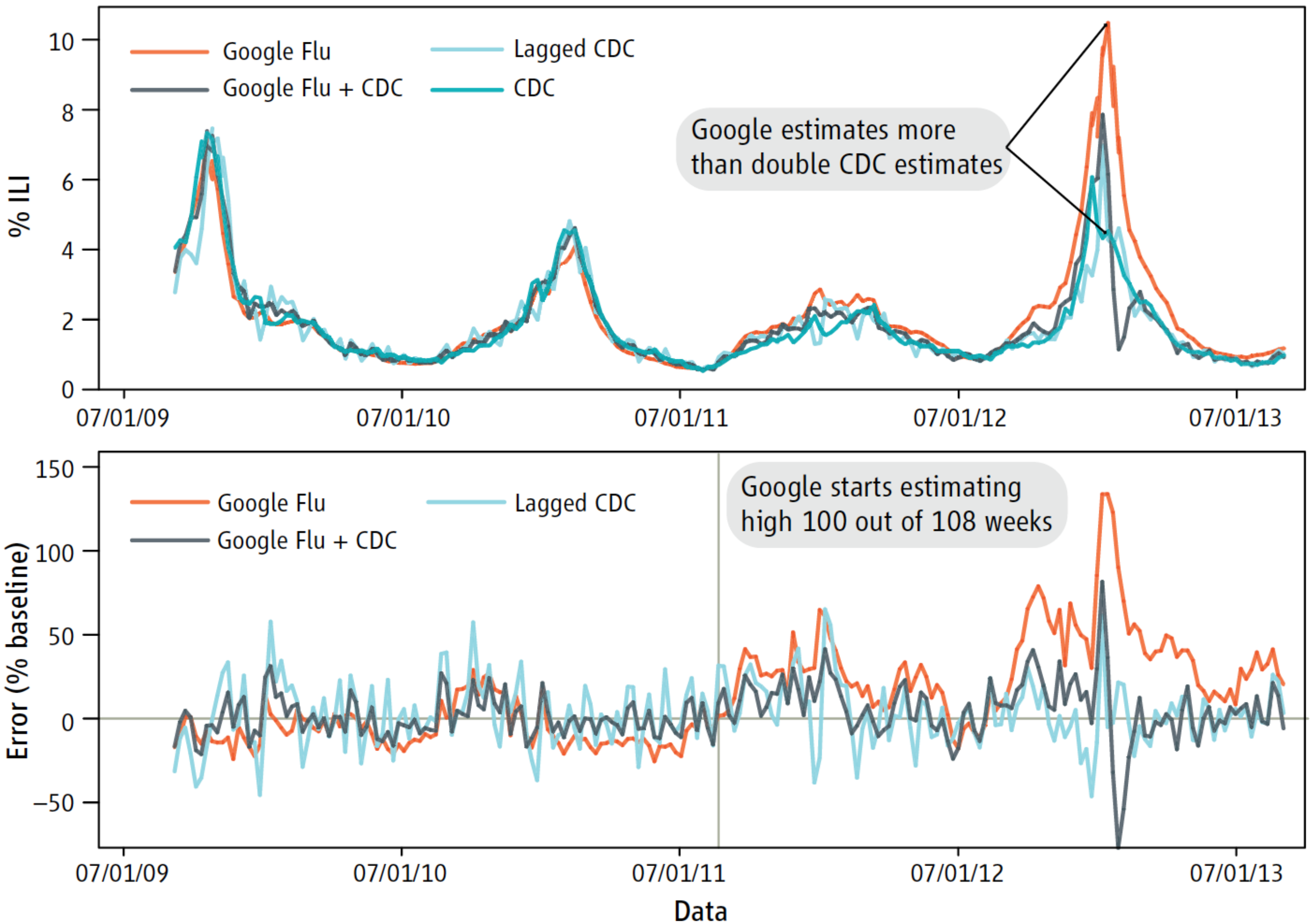
flu near you
do you have it in you?

FluTracking.net

Influenzanet

Passive data collection

Google Flu Trend (paradigm)



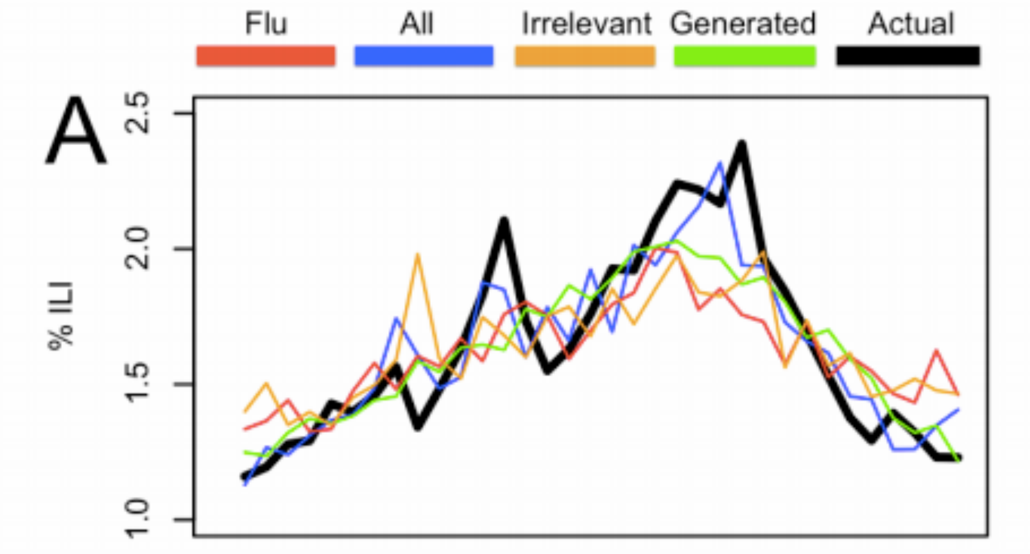
Use surrogate signal in algorithm trained on historical data (generally CDC time series) to achieve lead time (real time data collection, time-series extrapolation)

Case study on GFT and other non-generative models simple Lagged regression can be 90% “good” (Lazer et al. Science 2014).

Red-team - Blue team issues

Media hype

Twitter, OpenTable, Wikipedia, ……..



Statistical biases, "zombies" etc

- Well discussed in the literature
- Similarities & difference with GFT.

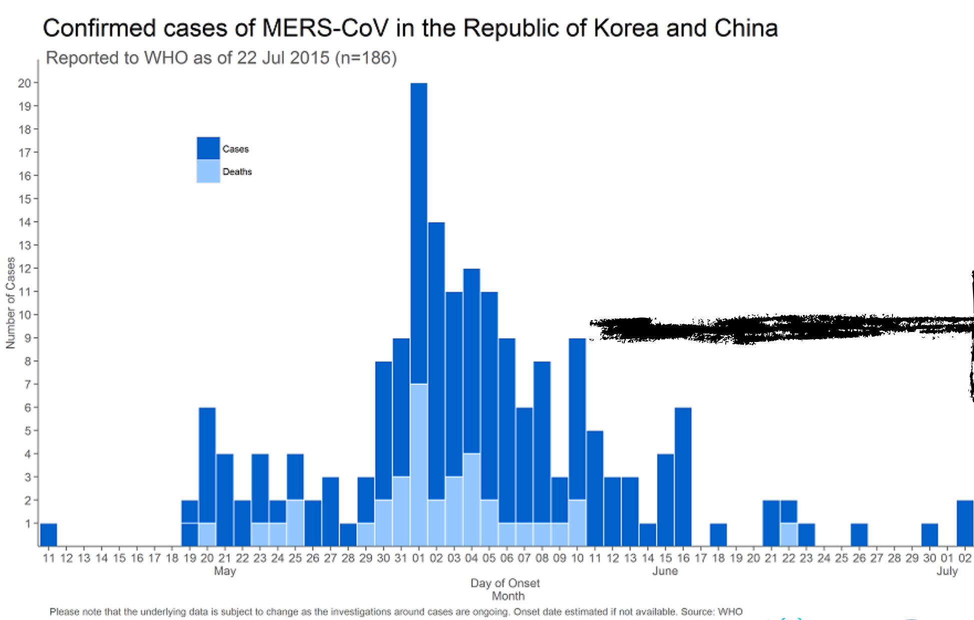
- Salathe'; Culotta; Dredze etc. Etc. (since the first paper by Signorini et al.);
- Word selection
- Linear regression, Multiple linear regression; SVM Regresssion; EFS
- High-level geographical resolution
- Full natural language processing

Big data narrative, “fourth paradigm”, “end of theory” etc.

“AI is changing how we do science”, “as far as it works” etc.

Potential number of pitfalls

- Lack of microlevel understanding (Black box effect, Causal inference, microscopic processes, observables…)
- Intrinsic Biases, Data incompleteness, noise
- More data not necessarily better modeling
- Inductive approaches to dynamical systems are dangerous
  See Hosni, Vulpiani, Philosophy &Technology (2017) MUST READ!



MOBS LAB

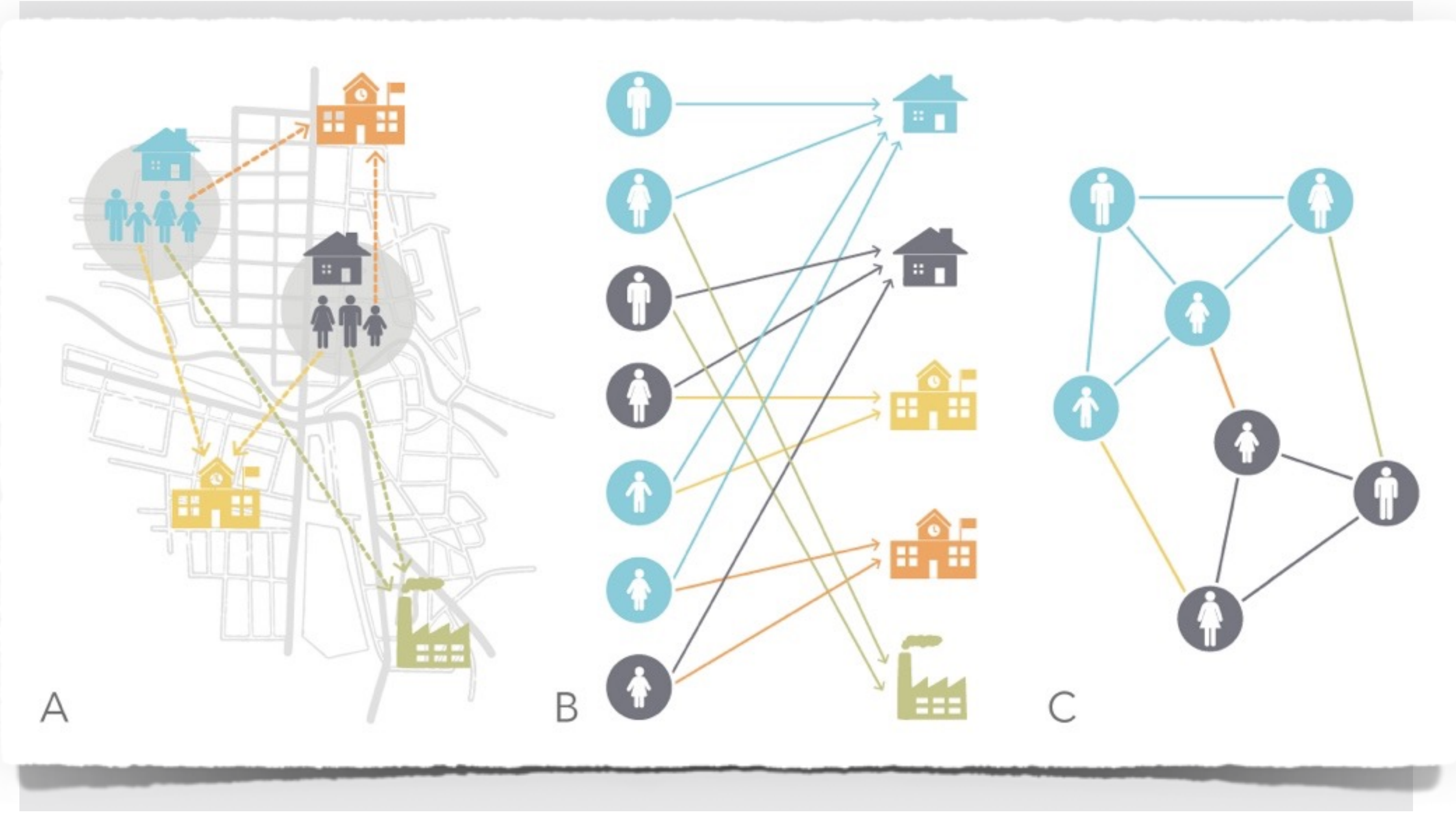
Actionable modeling with new data (big, or small)

- The focus is on understanding these data sets in a scientific sense and more deeply the real world processes which produced the data (Theory)
- Mechanistic approach (apparent reductionism)
- Effective equations
- Initial conditions

[WWW.GLEAMVIZ.ORG](http://www.gleamviz.org)

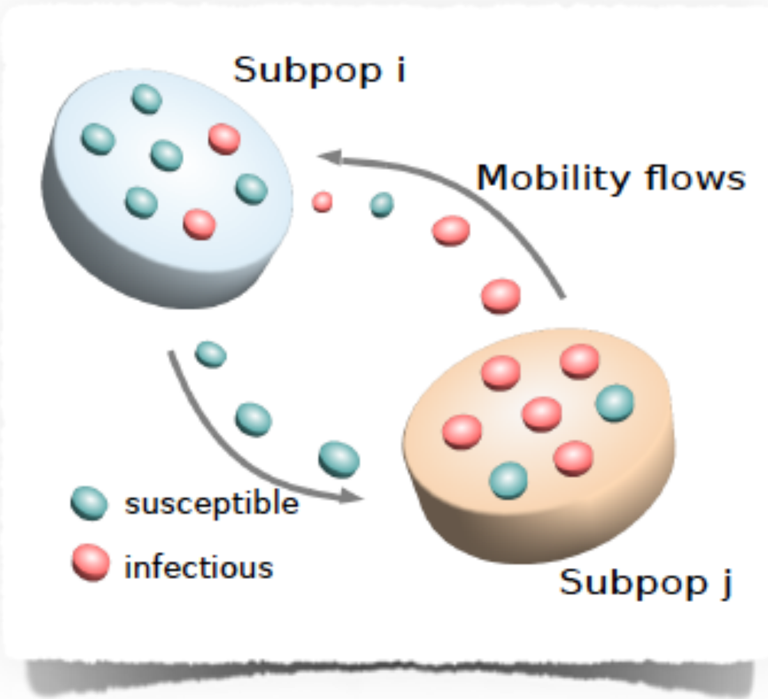
Stochastic Inter population dynamics

Multiple schemes for the stochastic intra-population contagion dynamic



MOBS LAB

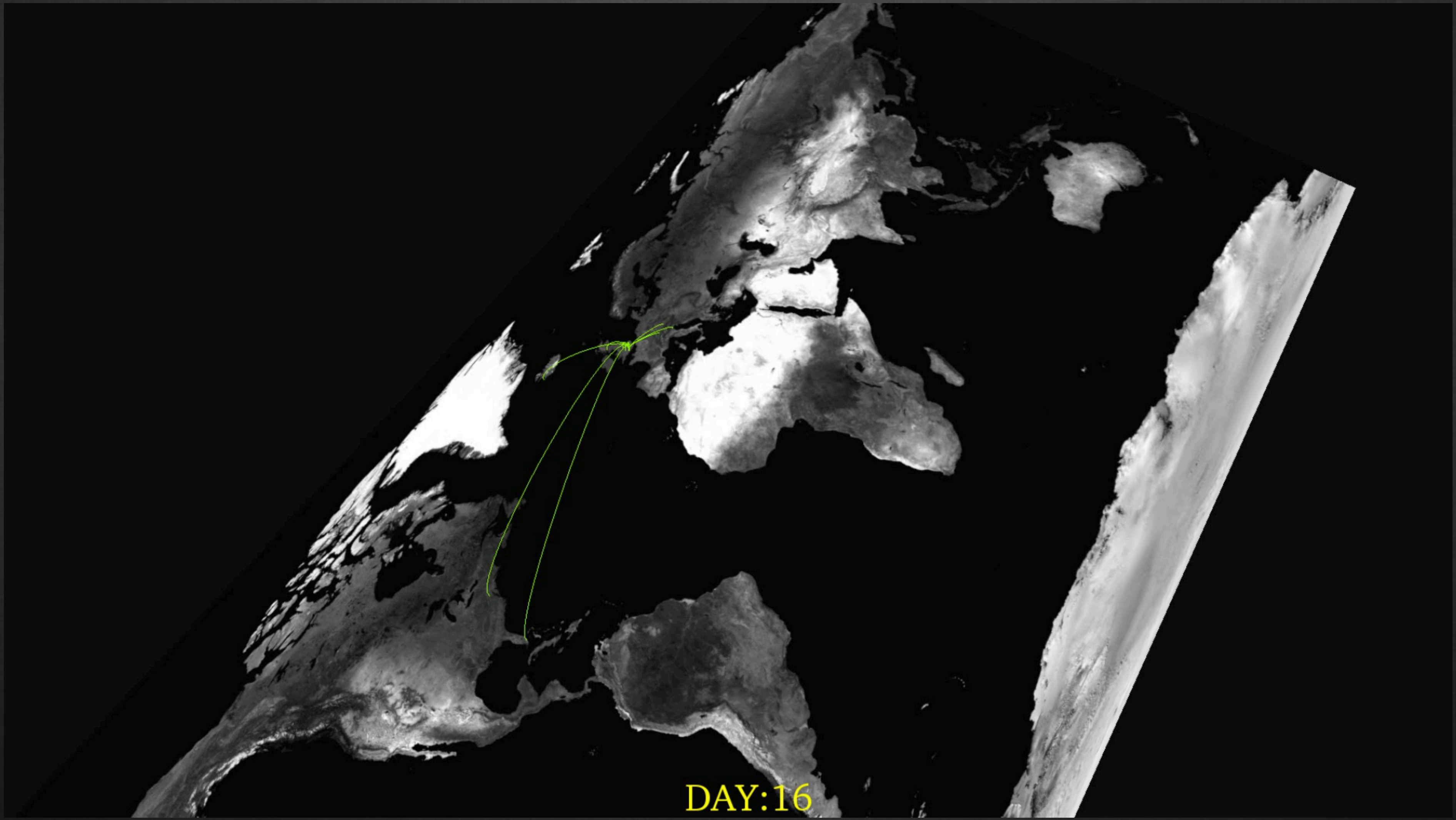
Reaction-diffusion on a network



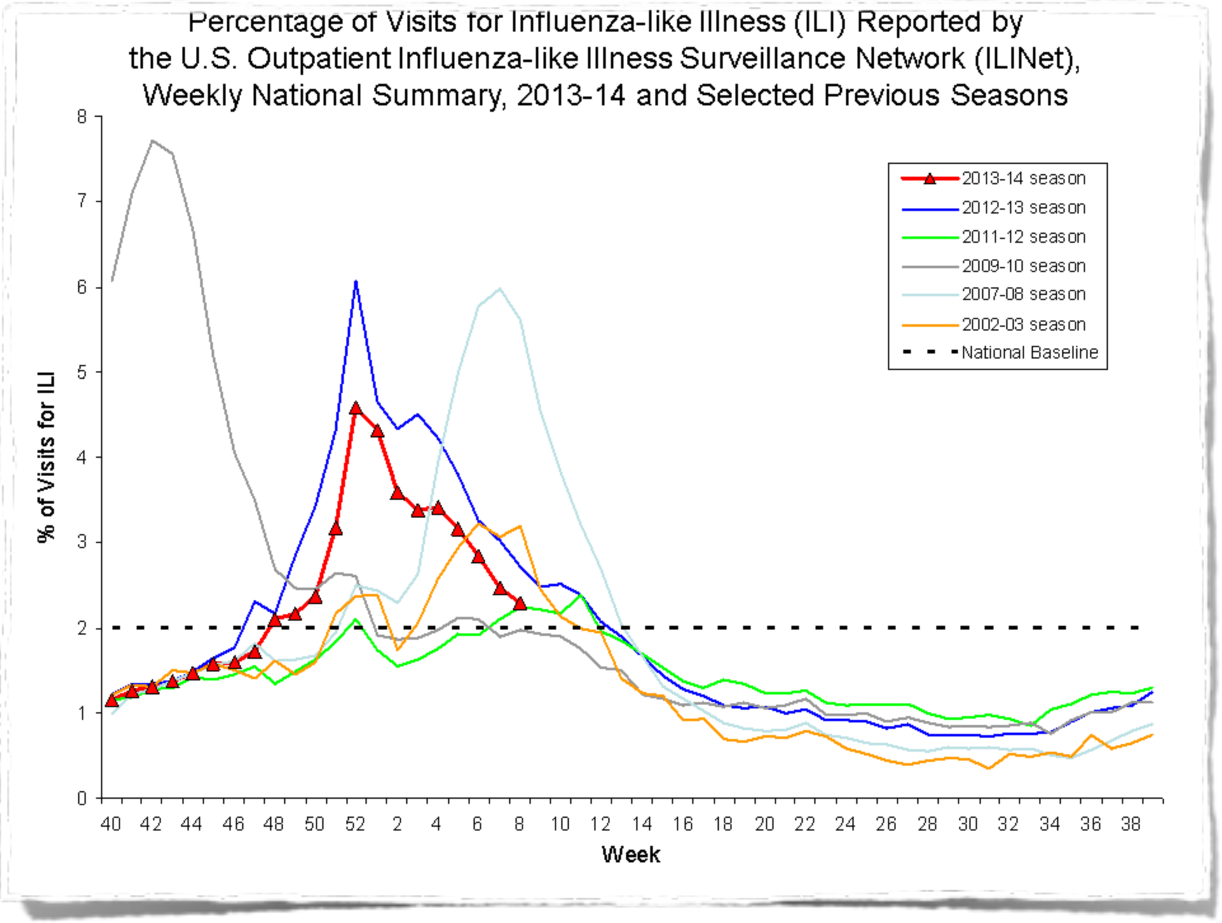
- REPEAT
  - CALL RANBin$(S, \beta I/N)$ and RANBin(I, μ)
  - S=S-RANBin$(S, \beta I/N)$
  - I=I+RANBin$(S, \beta I/N)$-RANBin(I, μ)
  - R=R+RANBin(I, μ)
  - $t = t + \Delta t$
  - PRINT S, I, R, t
- UNTIL $I = 0$

Not always more details better modeling/forecast. Context got the questions/scale needed.

Effective equations are not simple approximations (ex.: time-scale separation of fast-slow degrees of freedom through a B-O scheme).

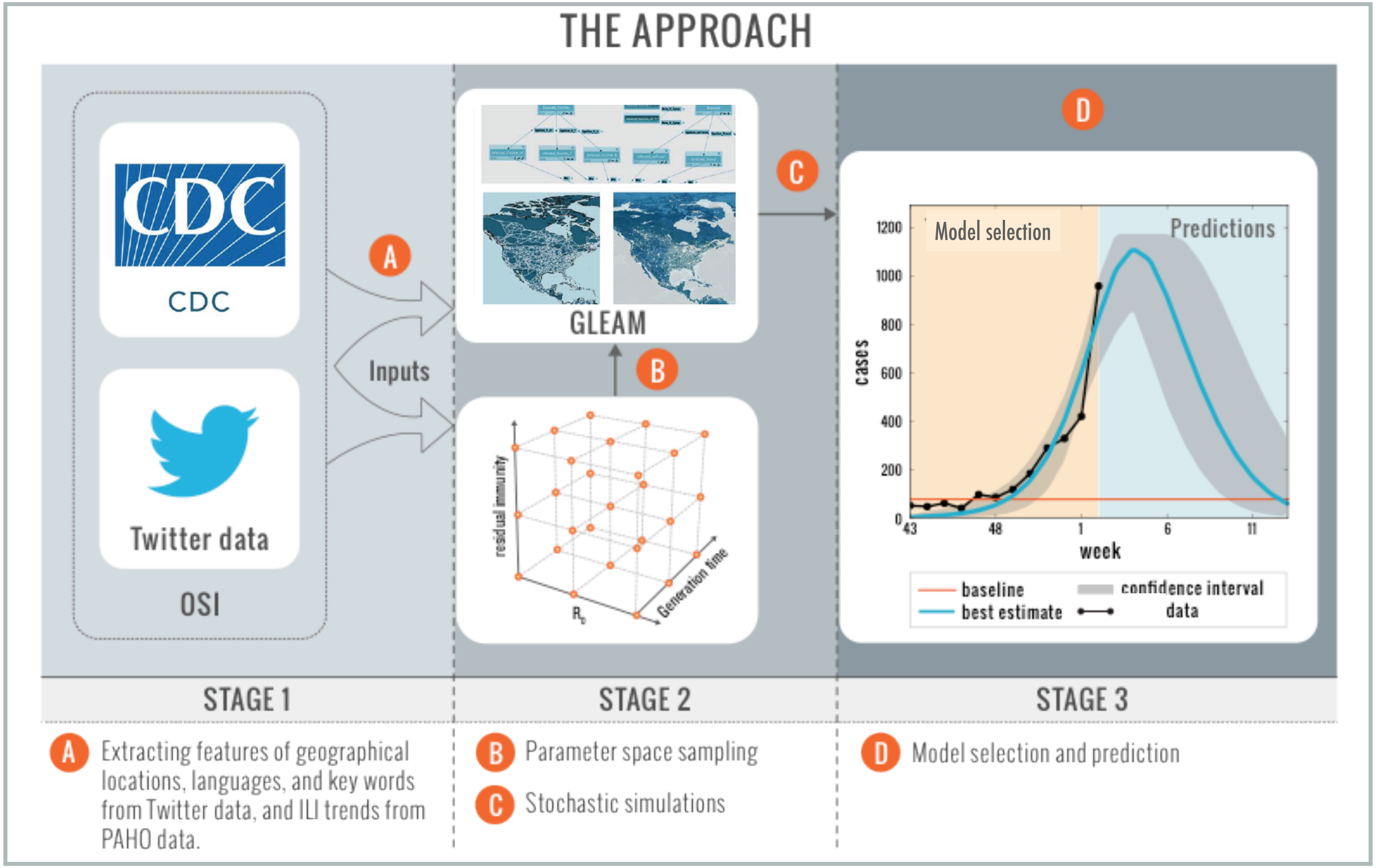

DAY:16

Seasonal Influenza



MOBS LAB

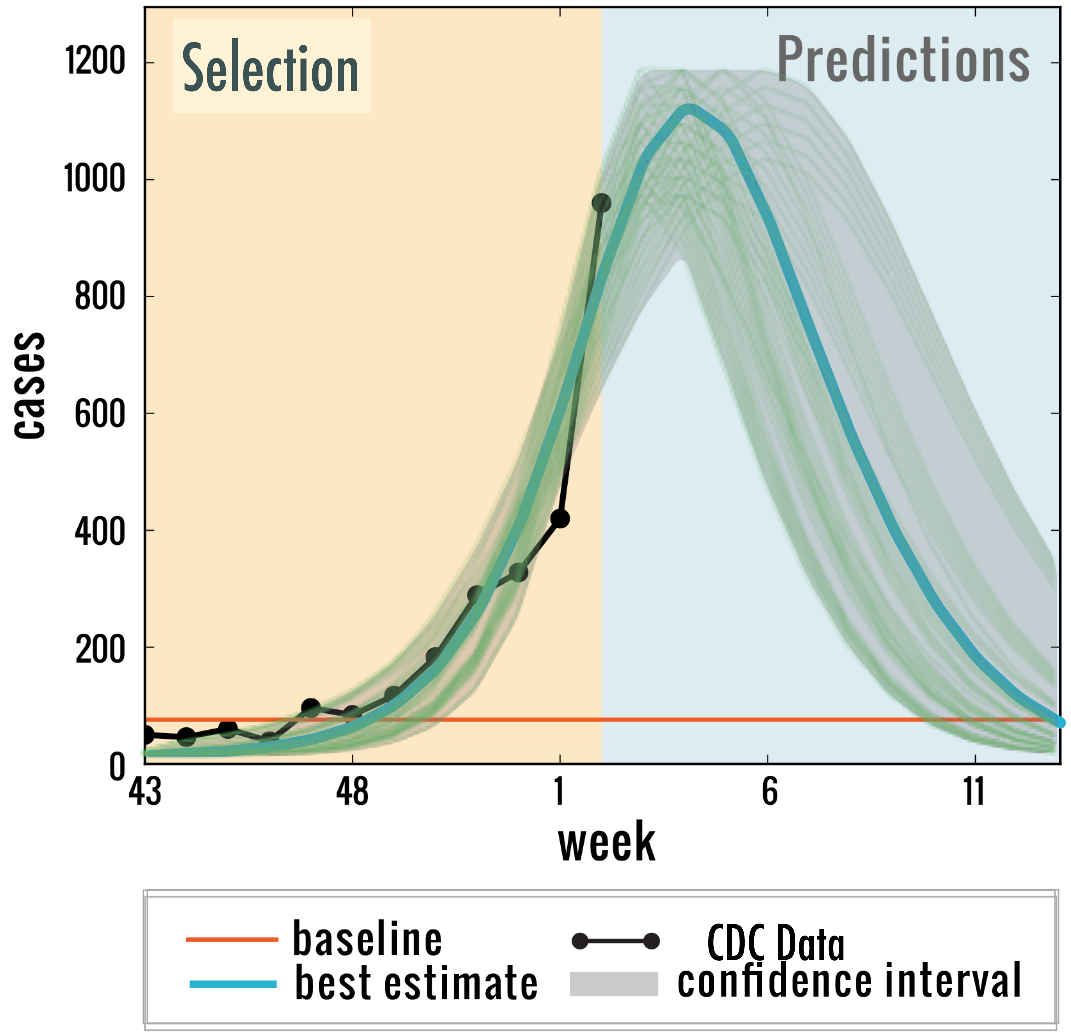
Generative modeling approach



MOBS LAB

Zhang,et al. WWW2017

Model selection

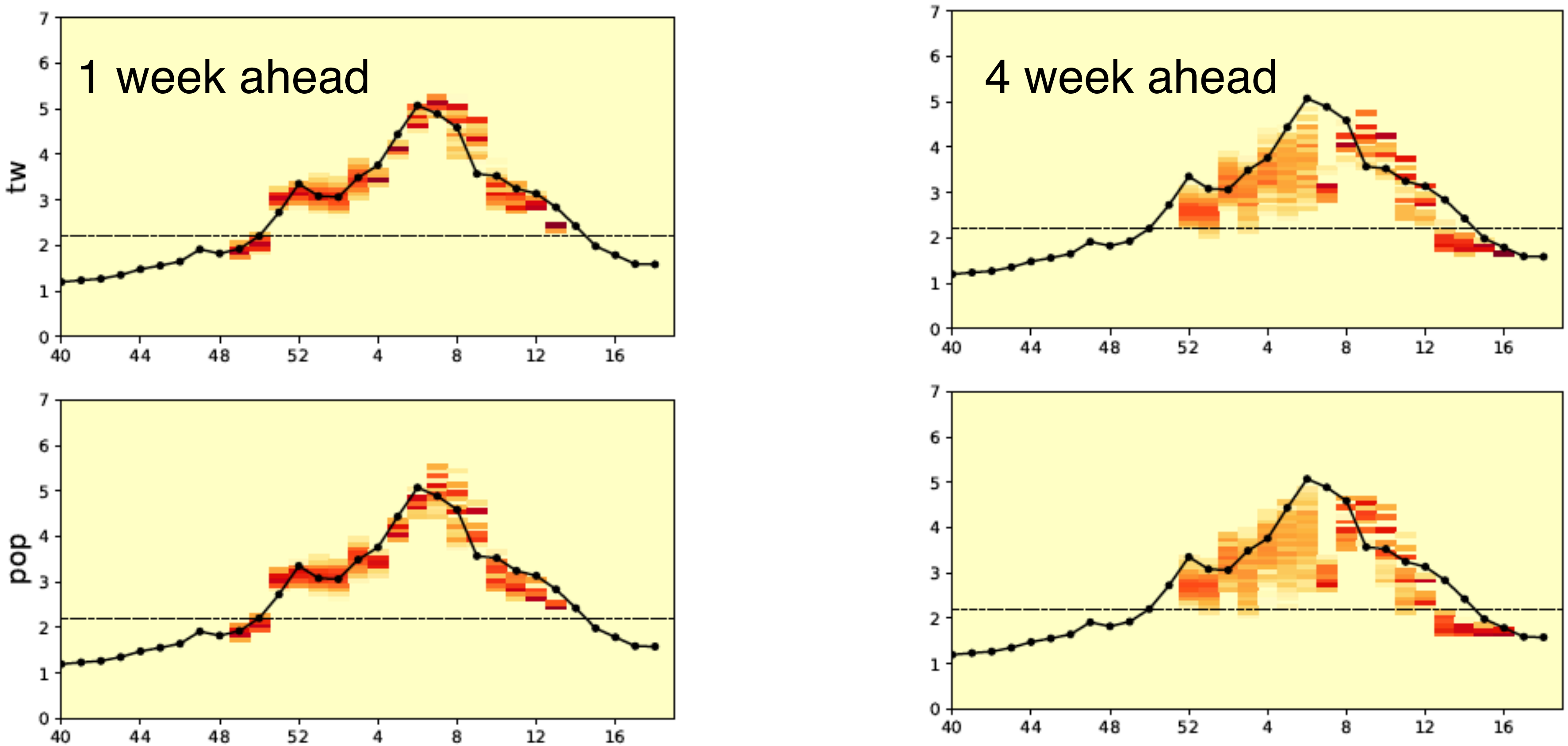


Information criterion (AIC) for model selection

MOBS LAB

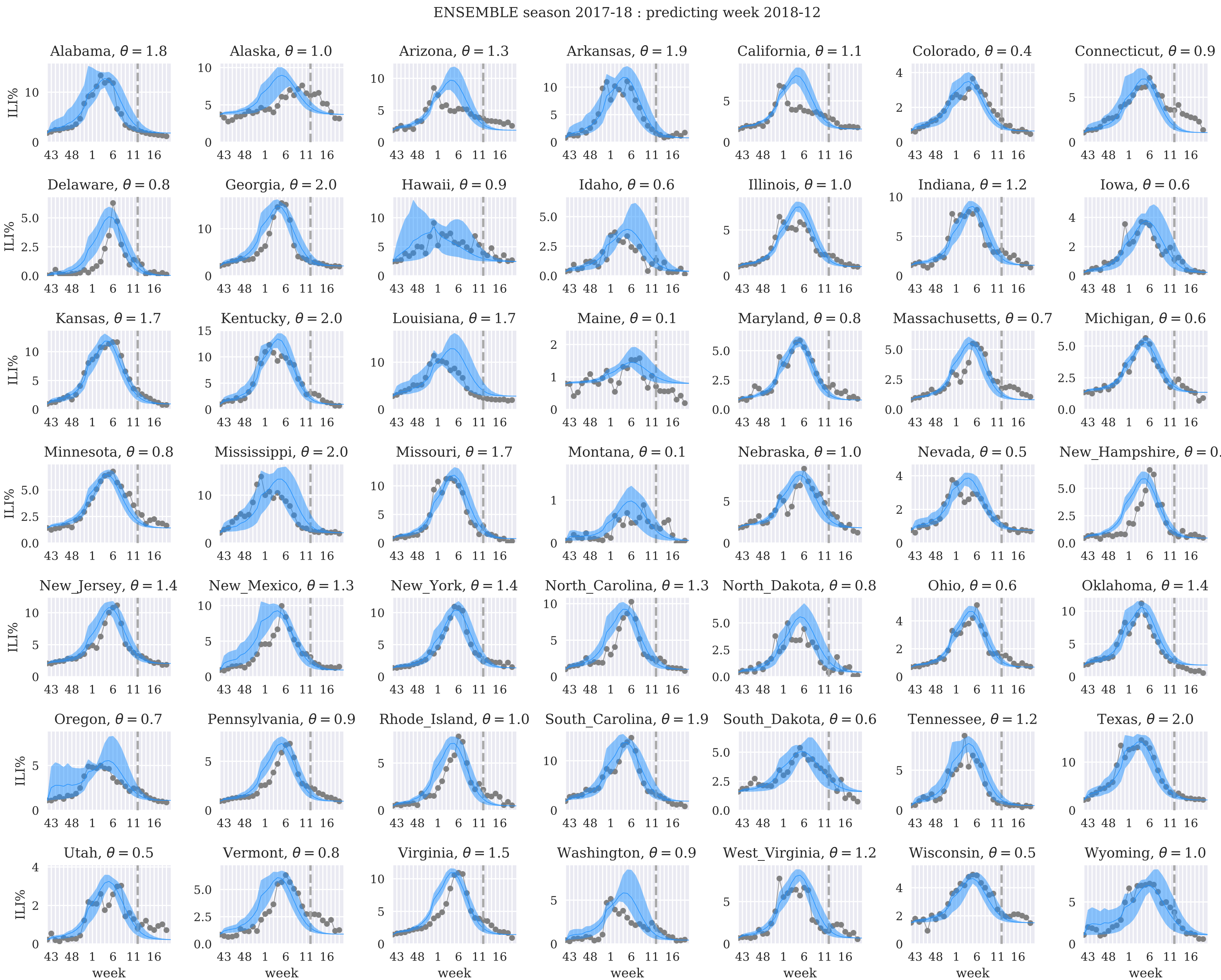
Time horizon and quality of predictions

			PearsonCorr				MAPE			
season	country	model	1-wlp	2-wlp	3-wlp	4-wlp	1-wlp	2-wlp	3-wlp	4-wlp
13/14	US	emm	0.90	0.78	0.73	0.78	0.13	0.18	0.23	0.23
13/14	US	emmAug	**0.96**	**0.95**	**0.90**	**0.86**	**0.07**	**0.09**	**0.13**	**0.17**
17/18	US	emm	0.97	0.91	0.82	0.75	0.09	0.14	0.18	**0.20**
17/18	US	emmAug	**0.99**	**0.95**	**0.89**	**0.83**	**0.07**	**0.11**	**0.16**	**0.20**



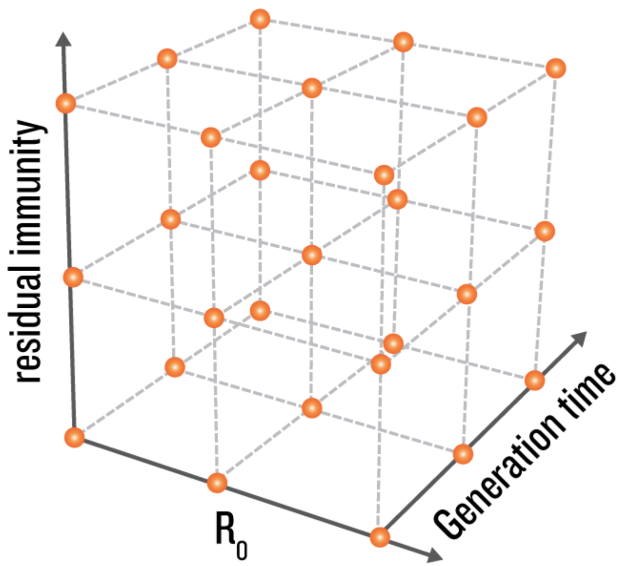
MOBS LAB

Bonus results (I)



Bonus results (II)

Q. Zhang, et al.
WWW '17, 311-319 (2017) (ACM DL).



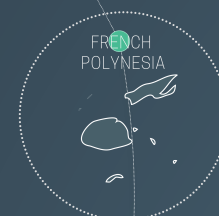
R_{eff}	Gt	imm (%)
1.50-1.74	4.0-6.1	15-40

	season	USA	Italy	Spain
R^{eff}	14/15	1.80 [1.50, 2.20]	1.50 [1.40, 1.50]	2.00 [1.80, 2.20]
	15/16	1.30 [1.20, 1.40]	1.20 [1.10, 1.30]	1.30 [1.20, 1.30]
residual immunity	14/15	0.15 [0.05, 0.35]	0.20 [0.05, 0.40]	0.15 [0.00, 0.30]
	15/16	0.30 [0.10, 0.40]	0.25 [0.00, 0.40]	0.10 [0.05, 0.35]
average infectious time	14/15	4.00 [2.50, 5.00]	3.60 [2.80, 5.00]	3.30 [2.50, 4.00]
	15/16	5.00 [3.60, 5.00]	3.30 [2.00, 5.00]	3.30 [2.50, 4.00]

First isolated in Zika Forest, Uganda
1947

Circulated in Central Africa and South Asia
1960s/1970s

Yap Island Outbreak, Micronesia
2007



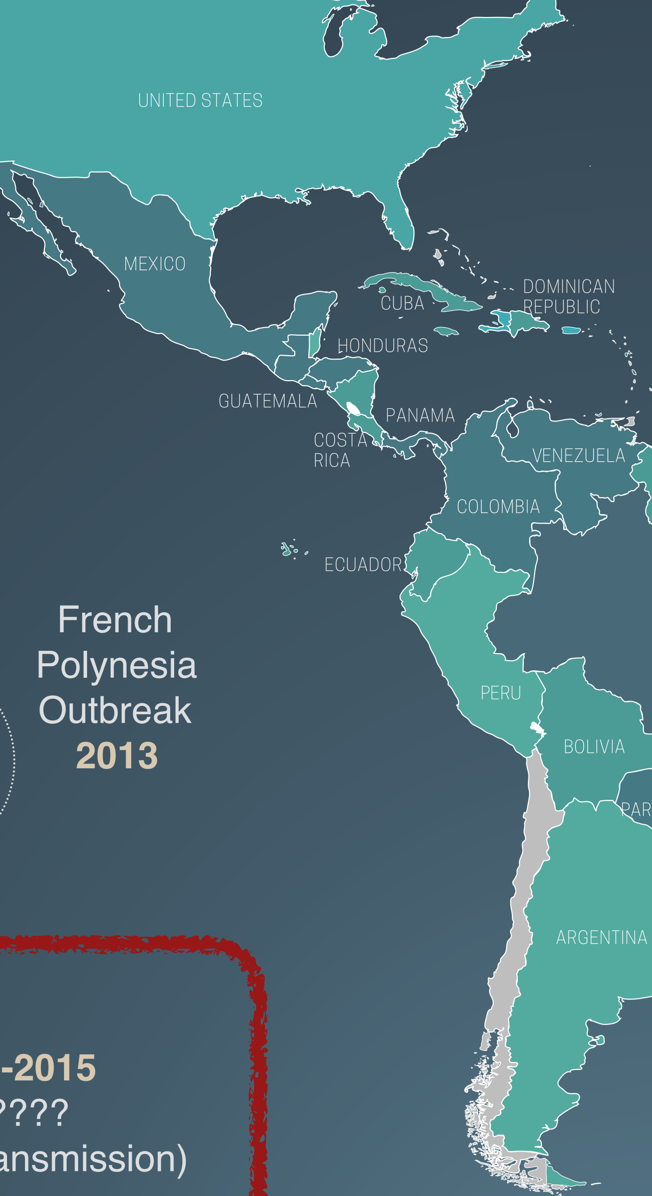
French Polynesia Outbreak
2013

2014-2015
??????
(Cryptic transmission)

Oct 2015
Unusual increase of microcephaly in Brazil, Zika link suspected

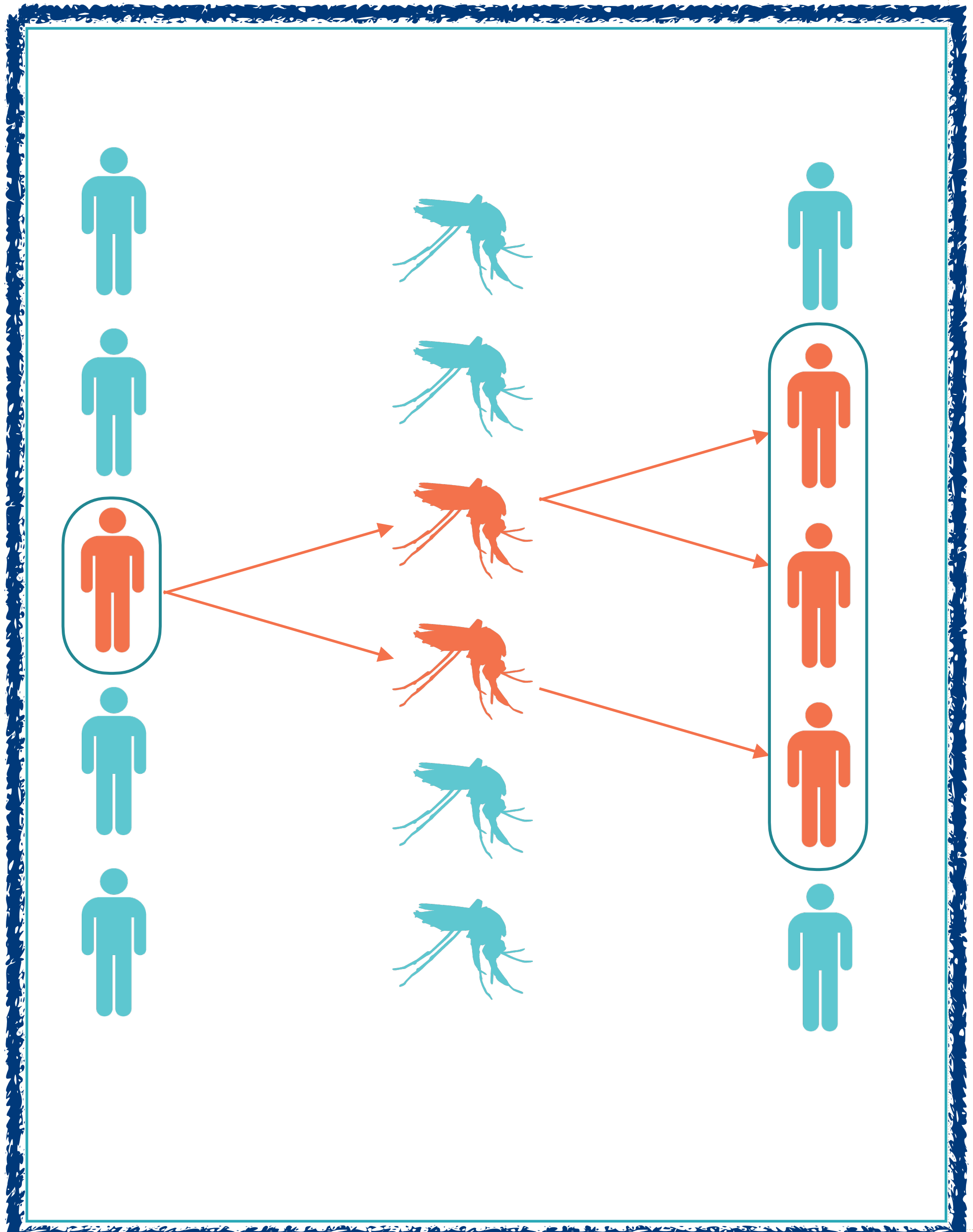
Feb 2016
WHO declared Zika as PHEIC

2013-2017
Zika marched through most of American countries



MOBS LAB

Zika Virus (ZIKV)



- RNA virus from the Flaviviridae family, genus Flavivirus.
- Generally mild disease characterized by low grade fever, rash, and/or conjunctivitis.
- Only approximately ~ 10% - 20% of those infected are symptomatic.
- ZIKV is spread primarily through infected Aedes mosquitoes.
- Plus, sexual and perinatal/vertical transmission are possible and the potential for transmission by transfusion is present

Spatial stochastic individual based model

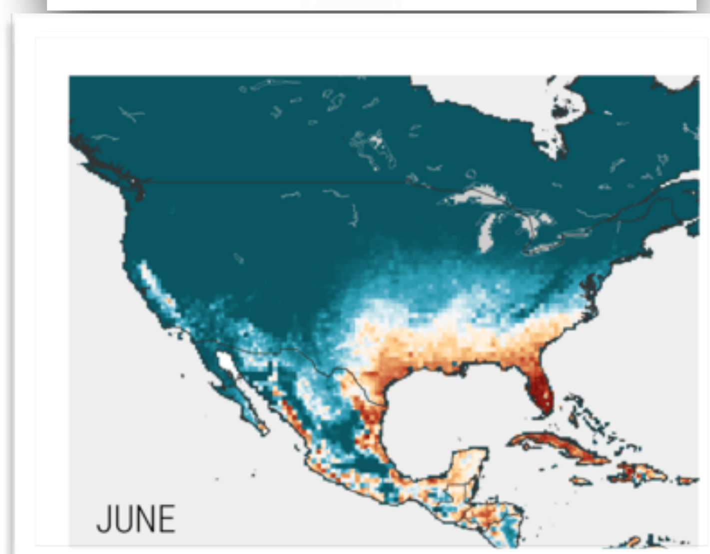
Zhang et al. PNAS 2017 ; doi:10.1073/pnas.1620161114

- Introduce explicitly the coupling of traveling patterns (case importation and colonization) on the disease progression

- Introduce seasonal drivers of Mosquito transmission in the epidemic dynamic.

- Introduce effect of socio-economic drivers

- Interplay of traveling pattern, outbreak initial conditions, disease dynamic and seasonal driving in defining the epidemic progression at the regional level.



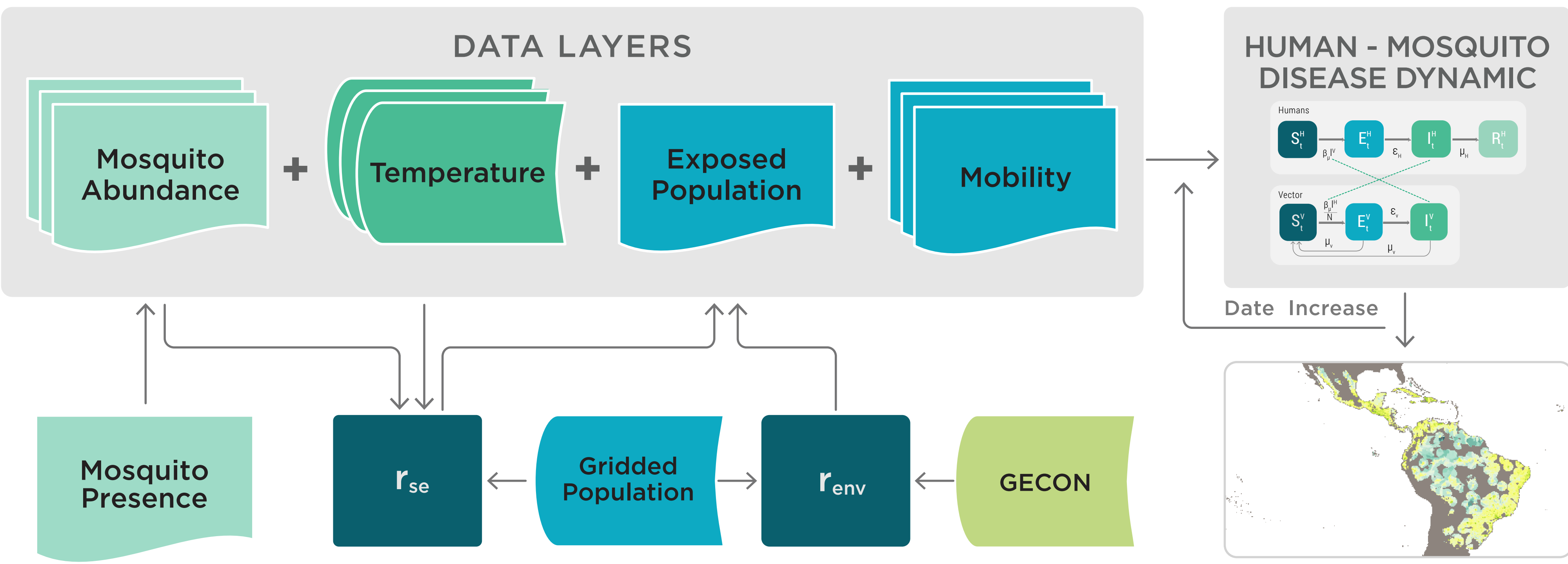
MOBS LAB

Model structure

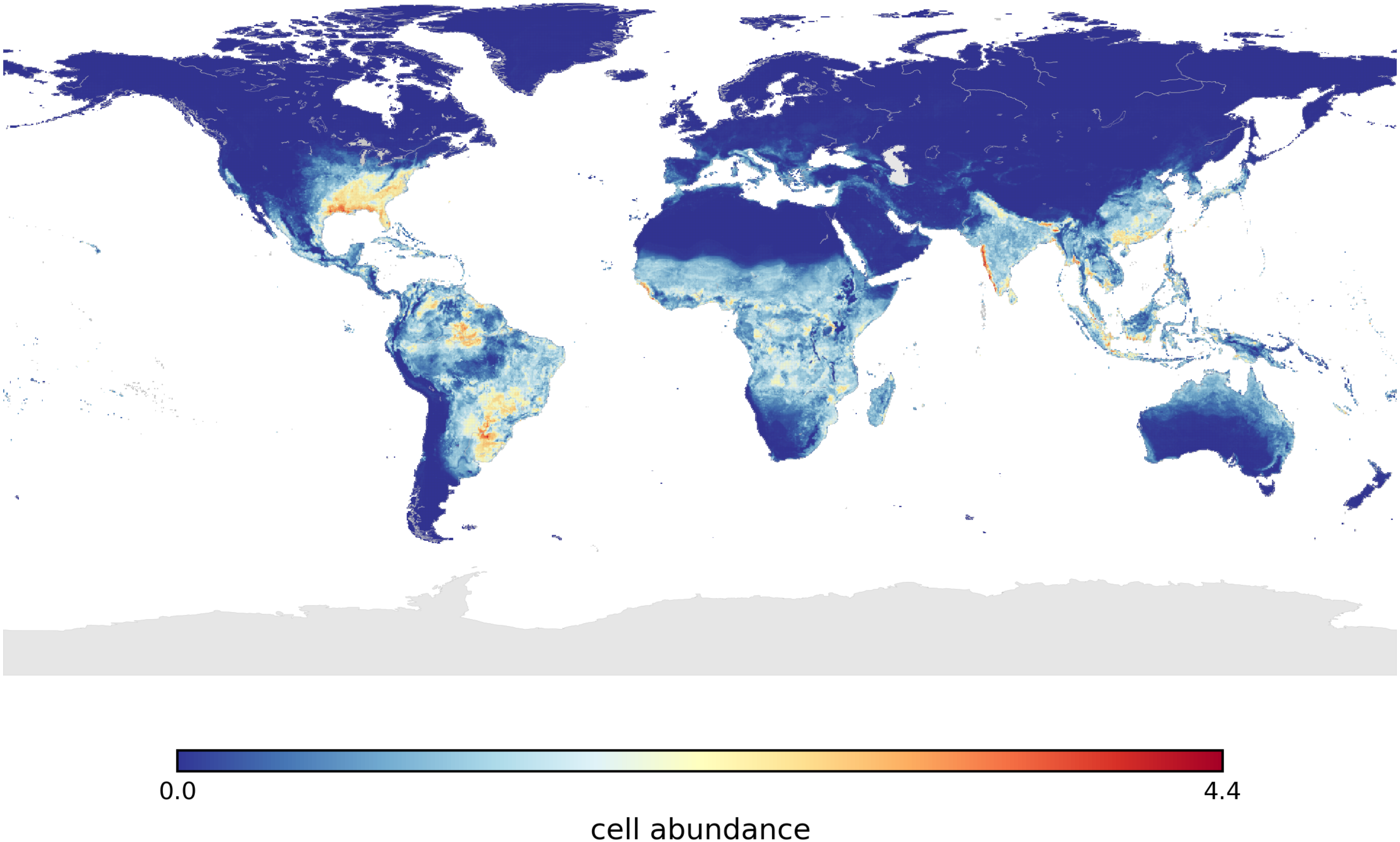
Explicit modeling of airline traffic national/ international + commuting patterns and local mobility

Mosquitoes abundance + local climate drivers + socio-economic indicators

Dynamic stochastic model providing time evolution of the epidemic



MOBS LAB



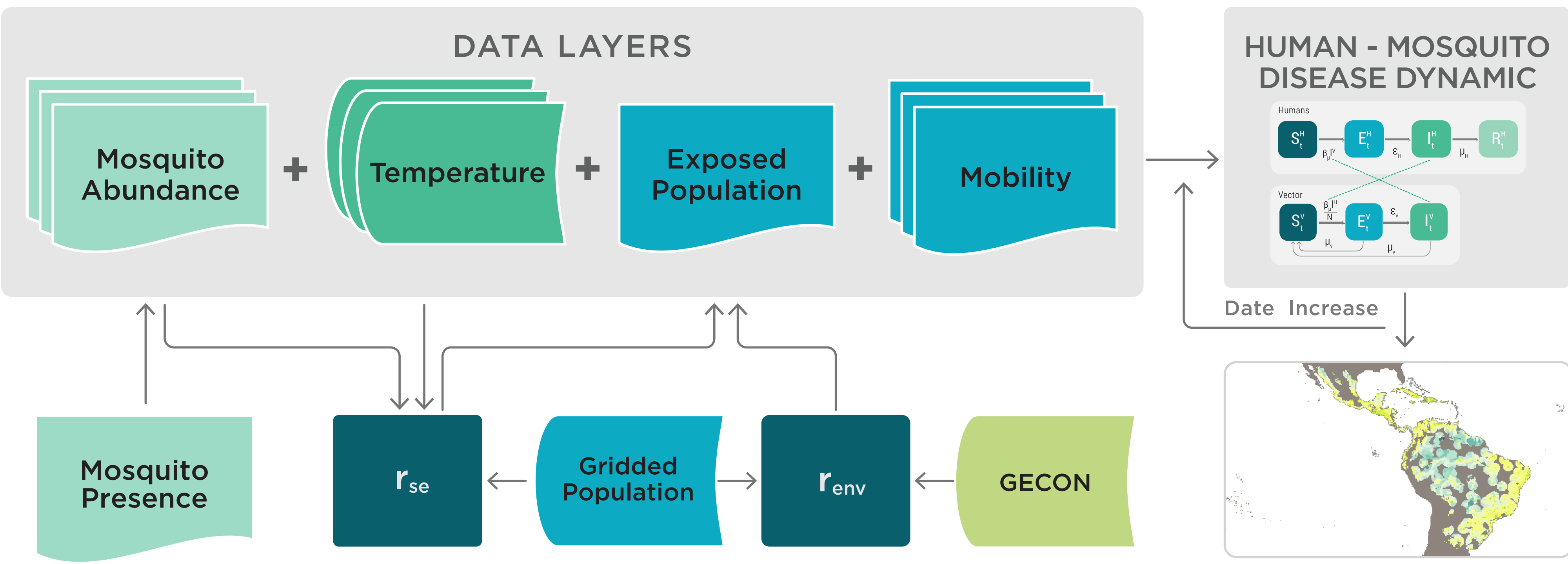
MOBS LAB

Model structure

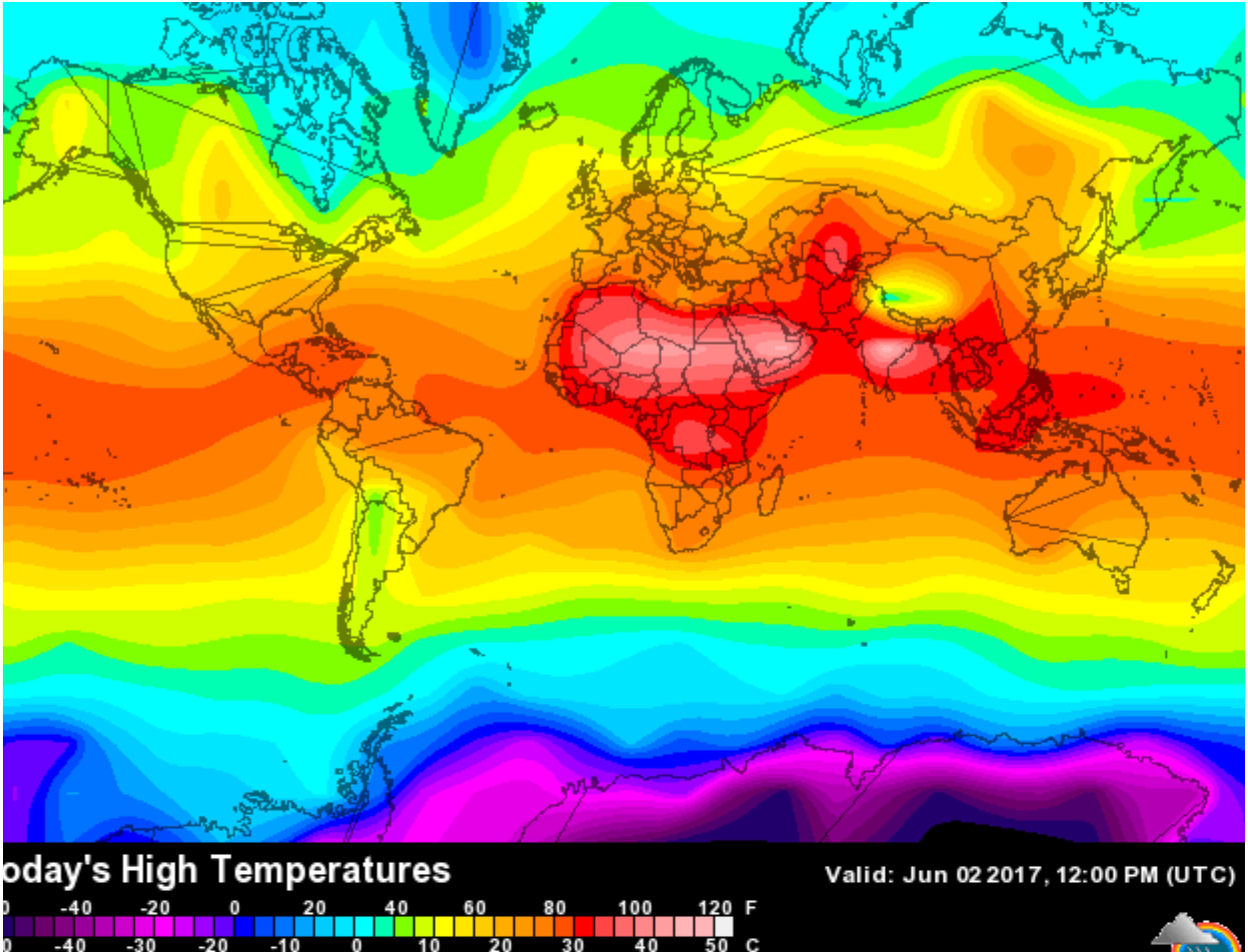
Explicit modeling of airline traffic national/ international + commuting patterns and local mobility

Mosquitoes abundance + local climate drivers + socio-economic indicators

Dynamic stochastic model providing time evolution of the epidemic



MOBS LAB

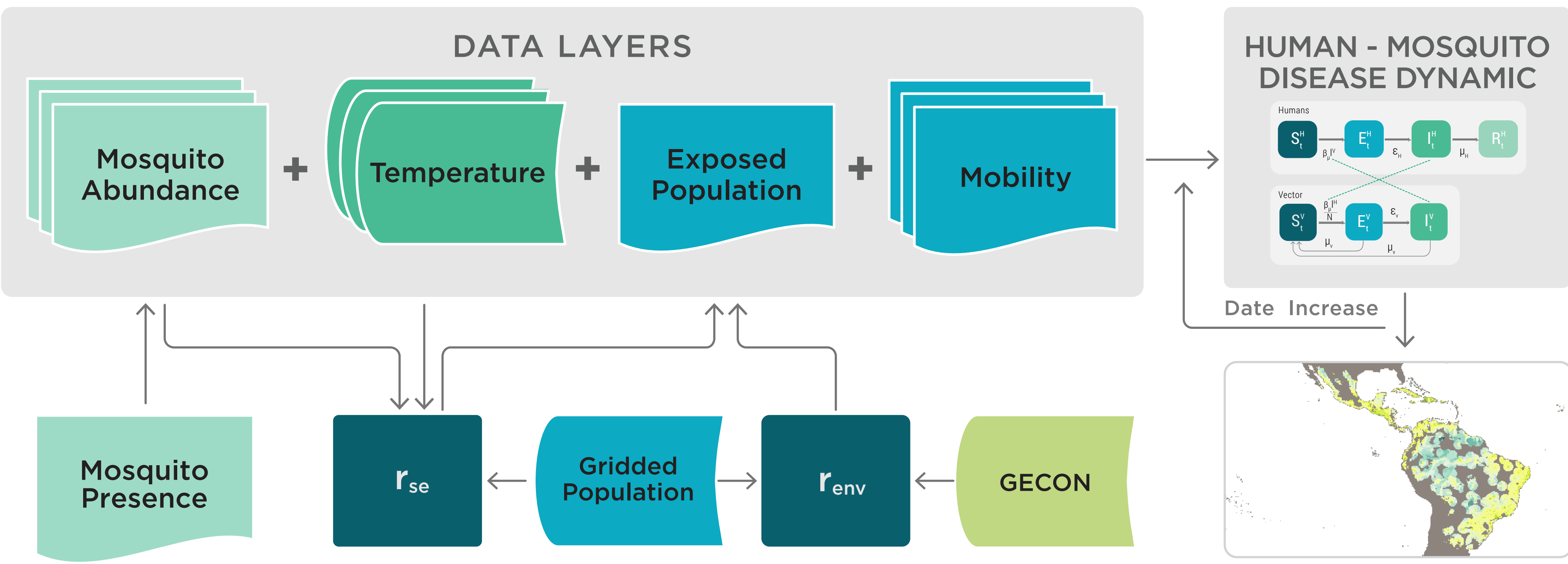

oday's High Temperatures
Valid: Jun 02 2017, 12:00 PM (UTC)
0 -40 -20 0 20 40 60 80 100 120 F
0 -40 -30 -20 -10 0 10 20 30 40 50 C

Model structure

Explicit modeling of airline traffic national/ international + commuting patterns and local mobility

Mosquitoes abundance + local climate drivers + socio-economic indicators

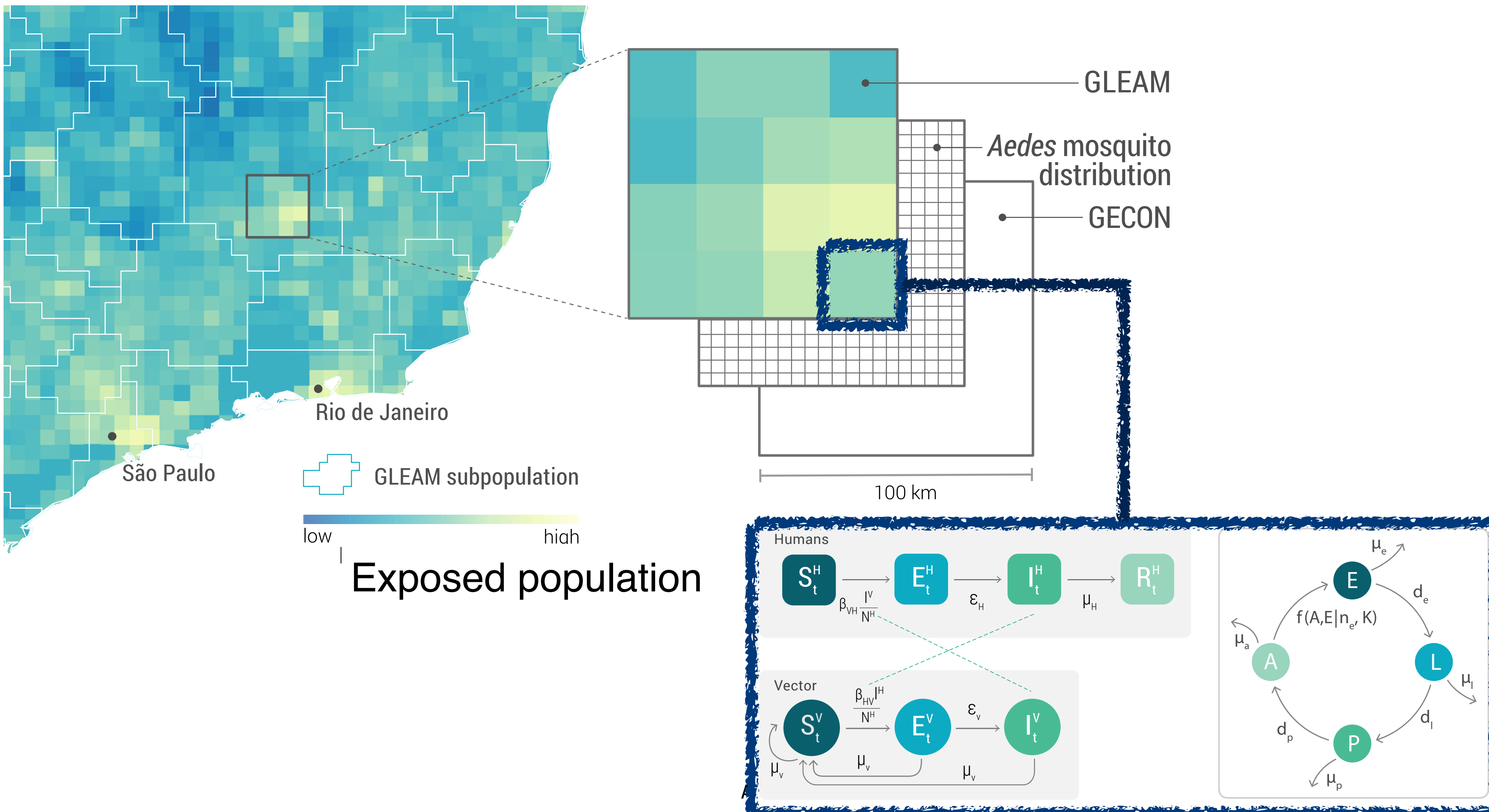
Dynamic stochastic model providing time evolution of the epidemic



MOBS LAB

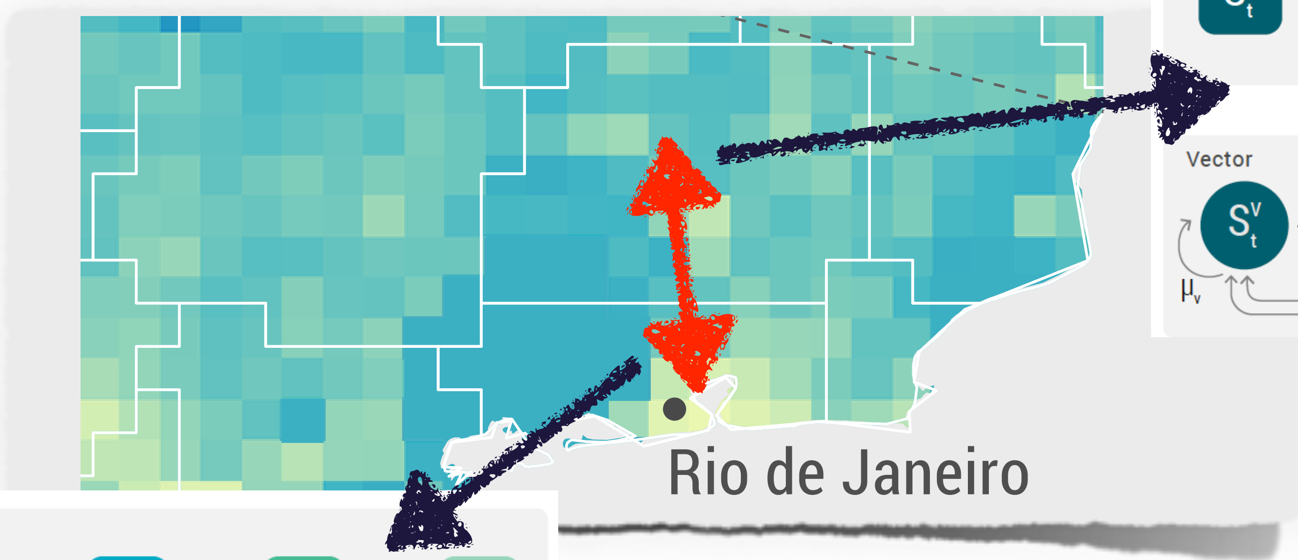
Model resolution

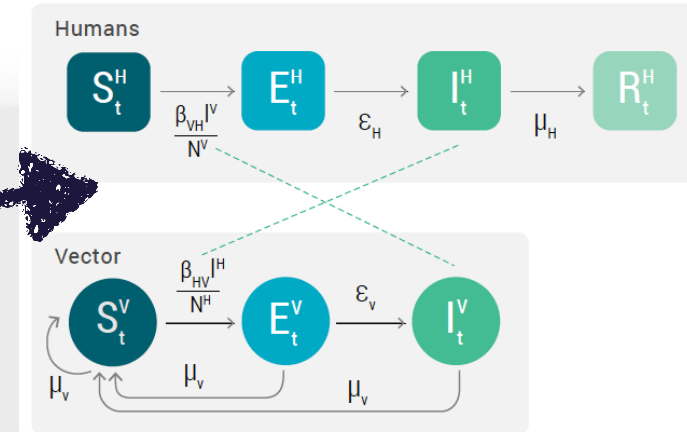
25 km x 25 km within census areas in all the Americas.
A few quantities can be projected up to 1km x 1 km

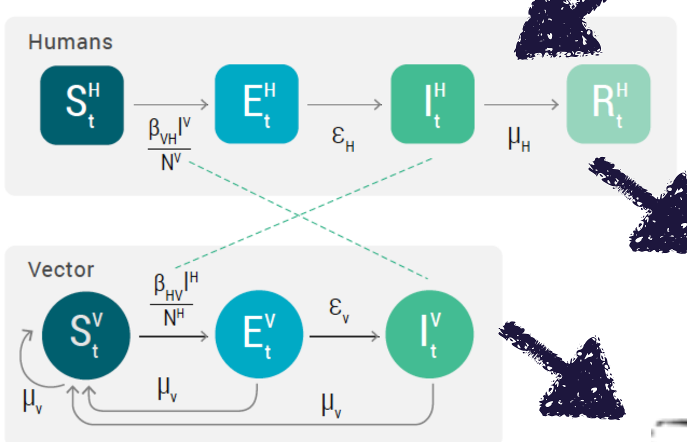


MOBS LAB

EPIDEMIC DYNAMICS





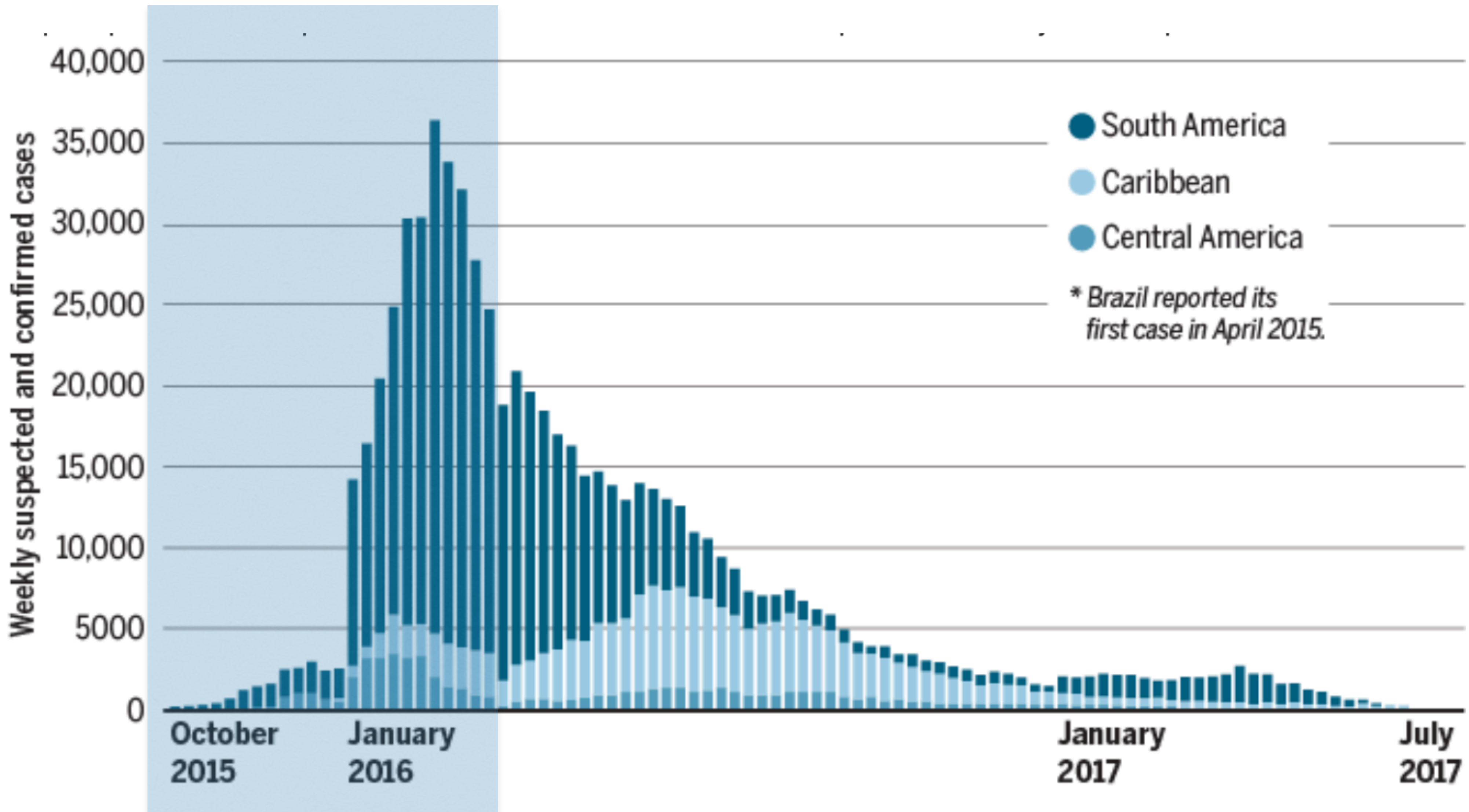


$$
\begin{aligned}
S_{t+1}^H &= S_t^H - Bin(S_t^H, \lambda_t^H) \\
E_{t+1}^H &= E_t^H + Bin(S_t^H, \lambda_t^H) - Bin(E_t^H, \epsilon_H) \\
I_{t+1}^H &= I_t^H + Bin(E_t^H, \epsilon_H) - Bin(I_t^H, \mu_H) \\
R_{t+1}^H &= R_t^H + Bin(I_t^H, \mu_H),
\end{aligned}
$$

$$
\begin{aligned}
S_{t+1}^V &= S_t^V - Bin(S_t^V, \lambda_t^V) + Bin(I_t^V, \mu_V) + Bin(E_t^V, \mu_V) \\
E_{t+1}^V &= E_t^V - Bin(E_t^V, \mu_V) - Bin(E_t^V, \epsilon_V) + Bin(S_t^V, \lambda_t^V) \\
I_{t+1}^V &= I_t^V + Bin(E_t^V, \epsilon_V) - Bin(I_t^V, \mu_V),
\end{aligned}
$$

MOBS LAB

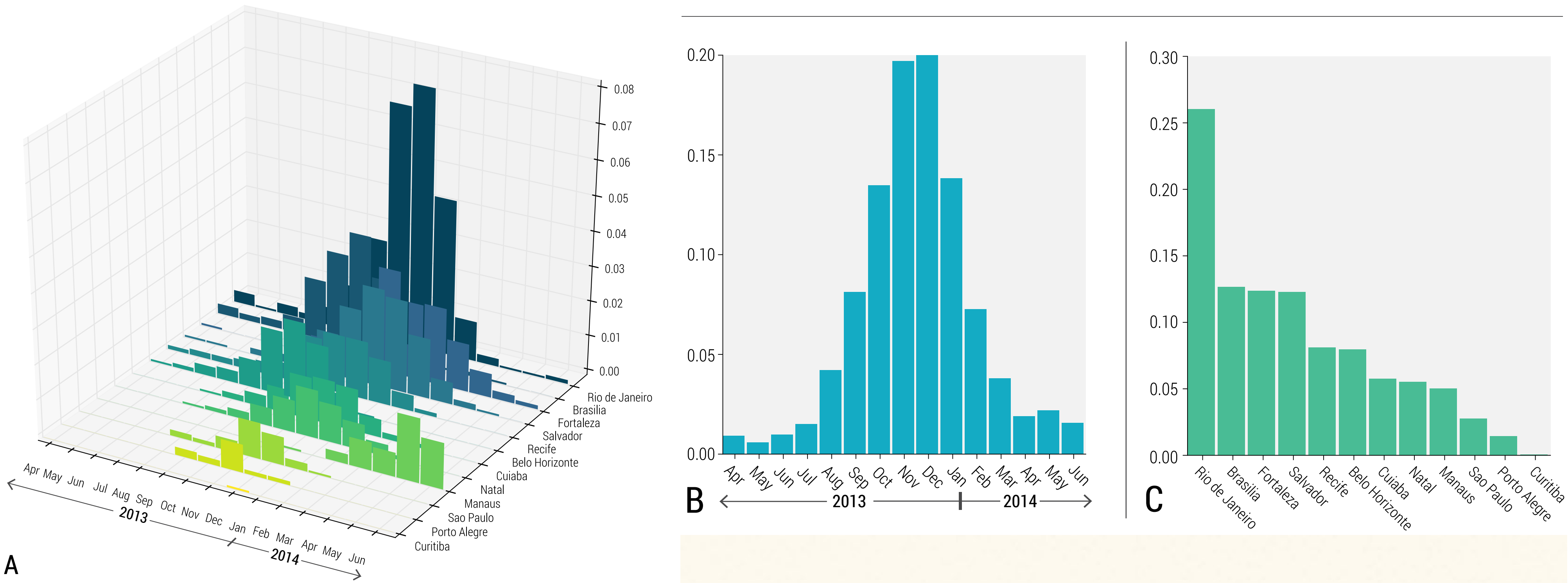
In order to understand the future of Zika epidemiology one needs to understand its past



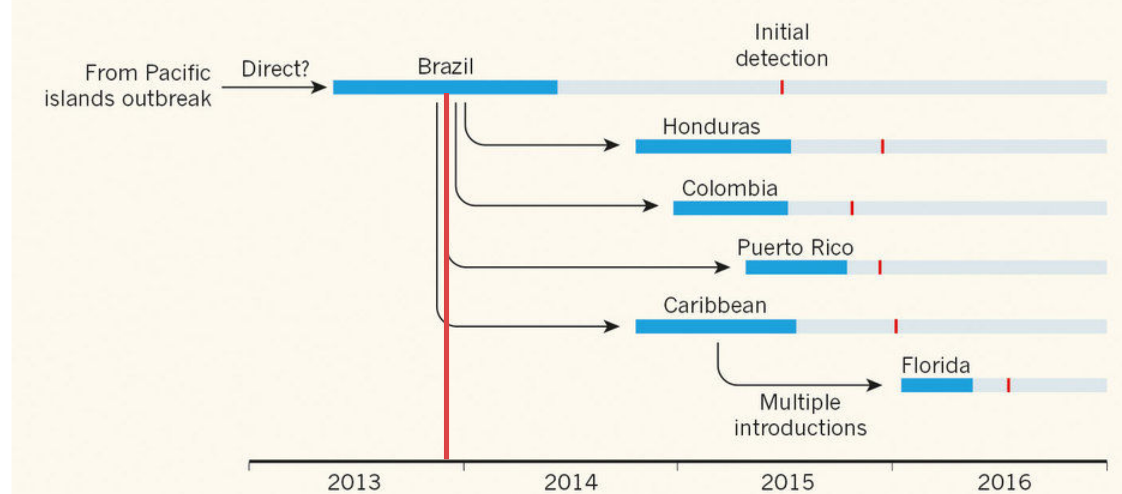
Source: Science, 2017

MOBS LAB

The Time Machine: Monte-Carlo estimates of ZIKAV introduction in the Americas

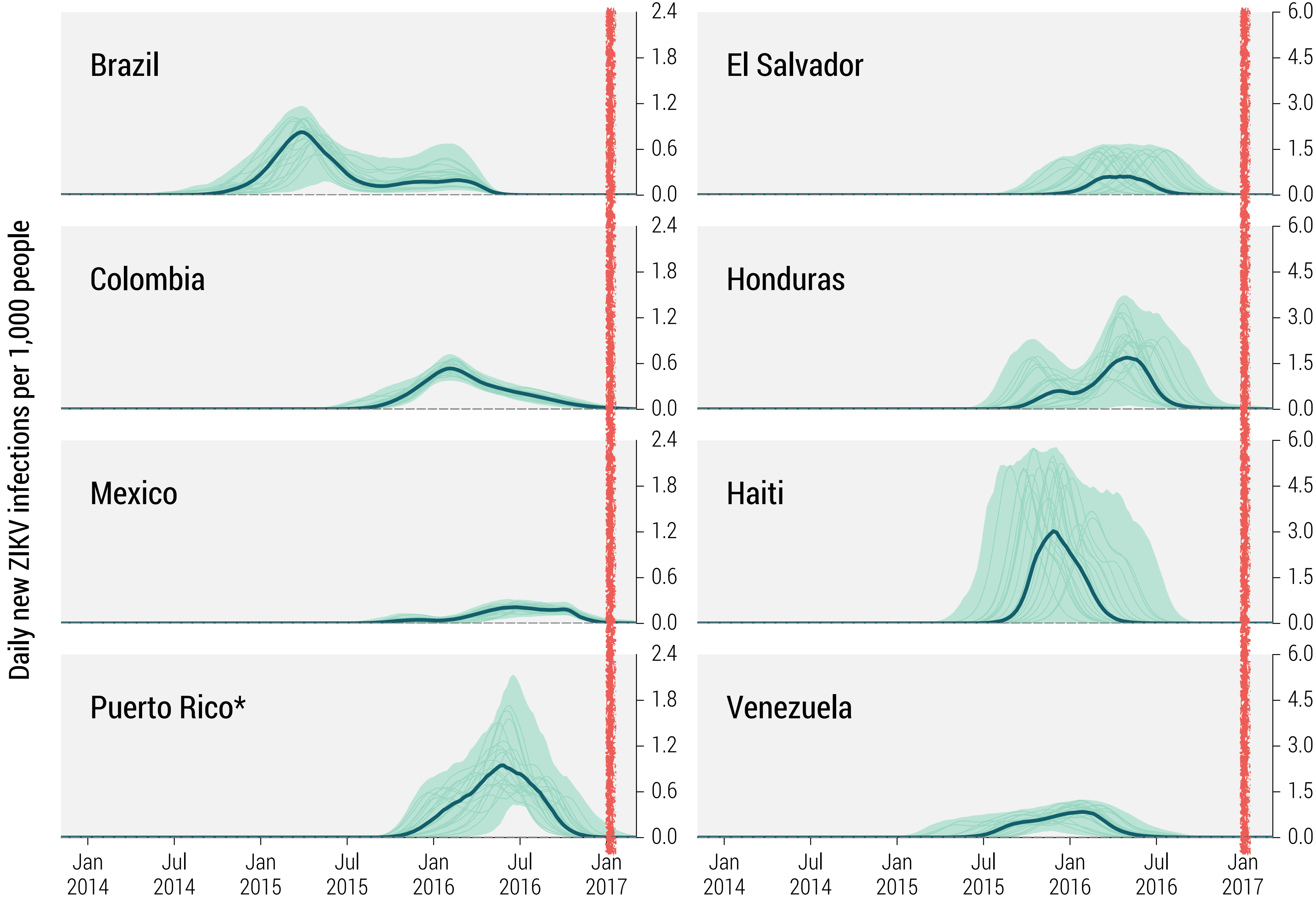


Modeling provides posterior distributions for the place and date of introduction in Brazil in good agreement with Phylogenetic analysis



Worobey, Michael. "Epidemiology: Molecular mapping of Zika spread." *Nature* 546.7658 (2017): 355-357.

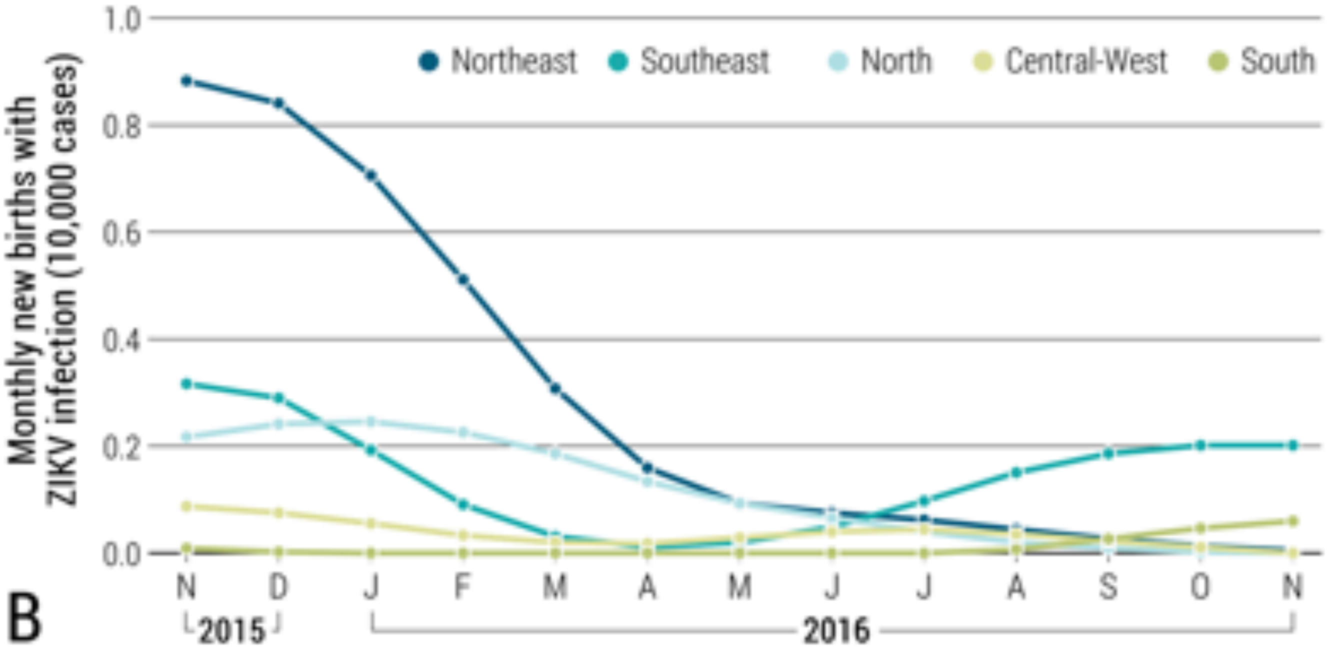
Epidemic declining in 2017



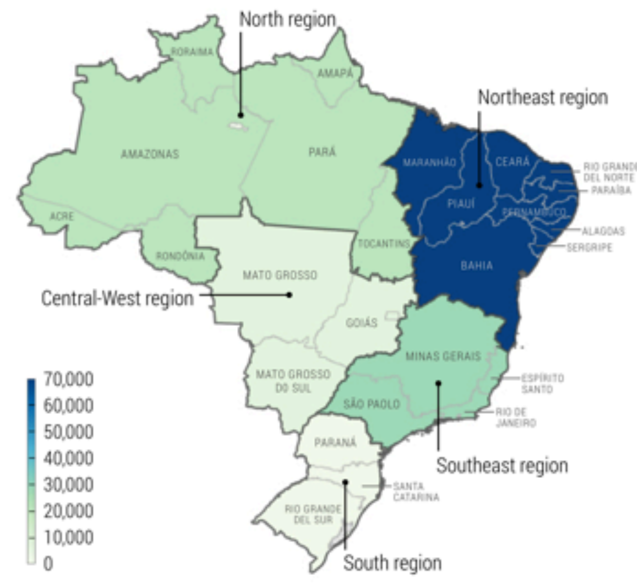
MOBS LAB

What happened in Northeast Brazil?

Researchers still don't understand why Northeast—a political region comprising nine Brazilian states—had so many cases of microcephaly and related birth defects between November 2015 and May. Across the country, roughly one-third of reported cases were confirmed as related to the Zika v



Brazil regional cumulative births with first trimester ZIKV infection by 2016-11-19



R=0.850 (p < 0.0001)					
explanatory variable	coefficient	p	[0.025	0.975]	% variance explained
log(population)	1.25	<0.001	0.90	1.60	62.8%
fraction of days with average temperature > 20℃	2.97	<0.001	2.01	3.94	46.5%

Determine areas at risk of observing Zika virus infections during 2017.

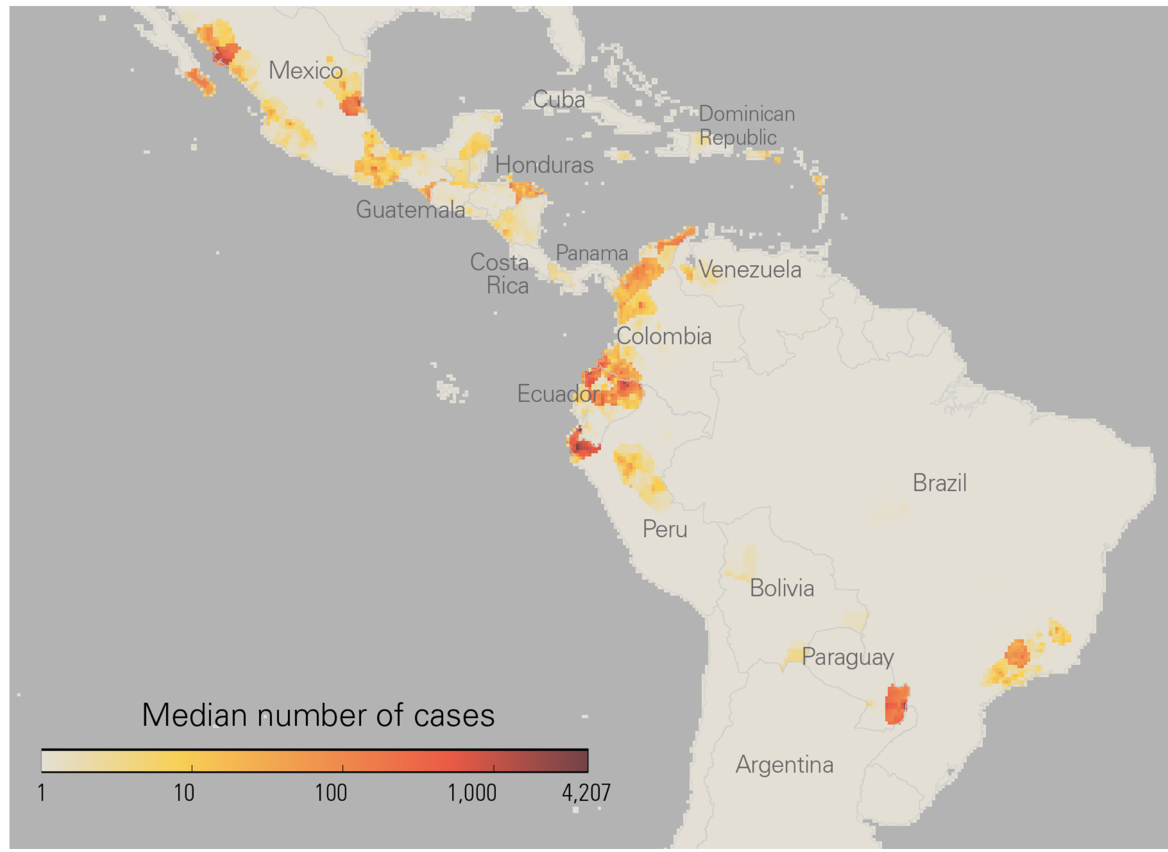
Analysis and Predictions for Vaccination Trials

NIH & CDC +3 modeling team:

Coordinators/advisors: Marc Fischer, Kiersten Kugeler, Michael Johansson, Grace Chen, Dean Follman, Rebecca Prevots, Jennifer Kwan, Shelby Daniel-Wayman, Jason Asher, Andrew Monaghan.

Modeling team 1 (MT1): Alessandro Vespignani, Qian Zhang, Kaiyuan Sun, Ana Pastore y Piontti, Matteo Chinazzi, Ira Longini and M. Elizabeth Halloran.
Modeling team 2 (MT2): Alex Perkins, Amir Siraj, Christopher Barker and Robert Reiner.
Modeling team 3 (MT3): Justin Lessler, Isabel Rodriguez-Barraquer, Neil Ferguson and Derek Cummings.

bioRχiv beta



MOBS LAB

Data & Modeling is more than forecast

- situational awareness
- intervention planning
- projections
- epidemiological explanations
- Structured reasoning

[MIDAS collaboration paper: Lofgren et al. Mathematical models: A key tool for outbreak response; PNAS 111 (51): 18095 (2014)]

MOBS LAB

CHARTING THE NEXT PANDEMIC

Modeling Infectious Disease Spreading in the Data Science Age

This book provides an introduction to the computational and complex systems modeling of the global spreading of infectious diseases. The latest developments in the area of contagion processes modeling are discussed, and readers are exposed to real world examples of data-model integration impacting the decision-making process. Recent advances in computational science and the increasing availability of real-world data are making it possible to develop realistic scenarios and real-time forecasts of the global spreading of emerging health threats.

The first part of the book guides the reader through sophisticated complex systems modeling techniques with a non-technical and visual approach, explaining and illustrating the construction of the modern framework used to project the spread of pandemics and epidemics. Models can be used to transform data to knowledge that is intuitively communicated by powerful infographics and for this reason, the second part of the book focuses on a set of charts that illustrate possible scenarios of future pandemics. The visual atlas contained allows the reader to identify commonalities and patterns in emerging health threats, as well as explore the wide range of models and data that can be used by policy makers to anticipate trends, evaluate risks and eventually manage future events.

Charting the Next Pandemic puts the reader in the position to explore different pandemic scenarios and to understand the potential impact of available containment and prevention strategies. This book emphasizes the importance of a global perspective in the assessment of emerging health threats and captures the possible evolution of the next pandemic, while at the same time providing the intelligence needed to fight it. The text will appeal to a wide range of audiences with diverse technical backgrounds.

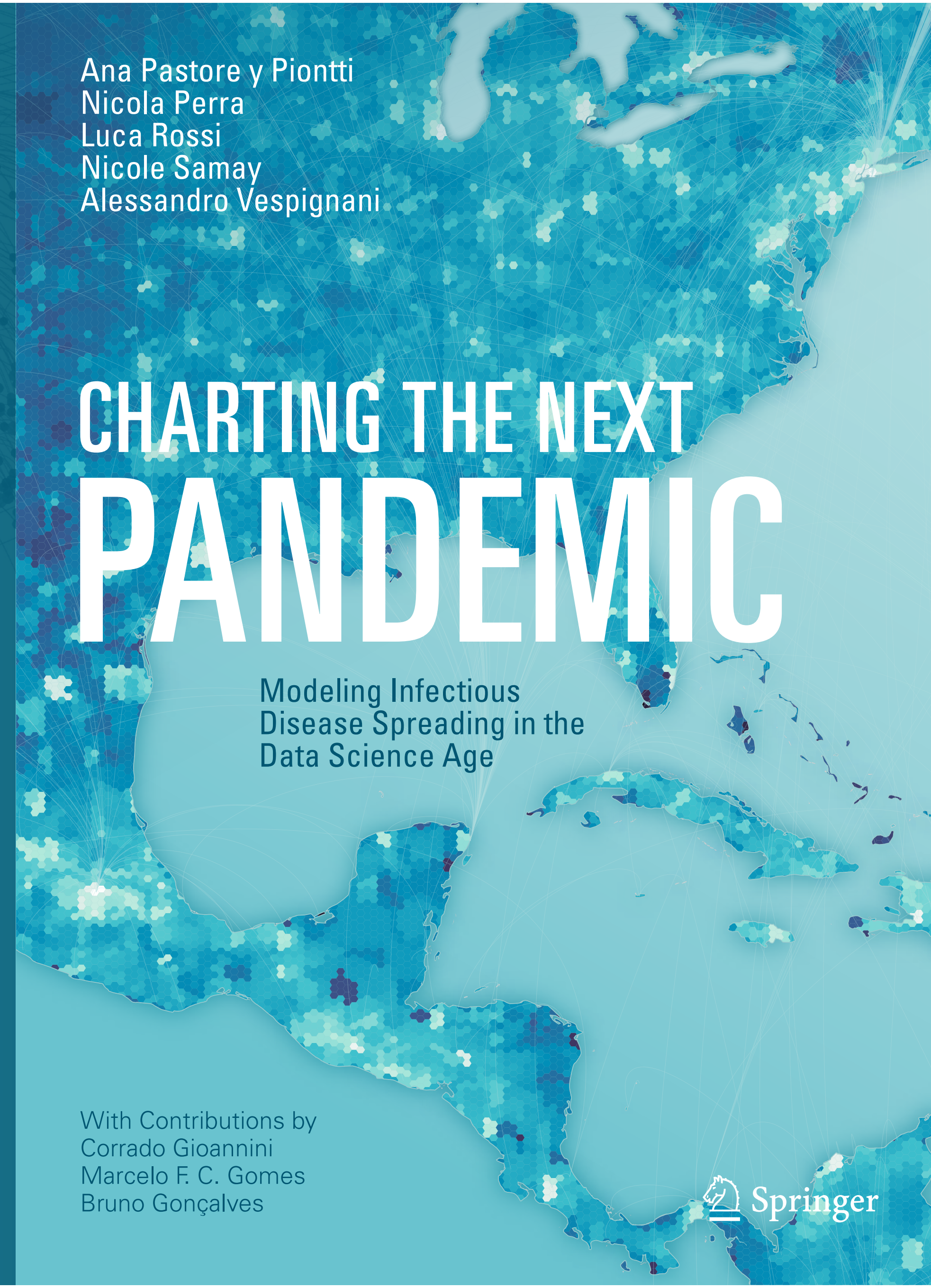
ISBN 978-3-319-93289-7

9 783319 932897

▸ springer.com

Pastore y Piontti · Perra
Rossi · Samay · Vespignani

CHARTING THE NEXT PANDEMIC



Ana Pastore y Piontti
Nicola Perra
Luca Rossi
Nicole Samay
Alessandro Vespignani

CHARTING THE NEXT PANDEMIC

Modeling Infectious Disease Spreading in the Data Science Age

With Contributions by
Corrado Gioannini
Marcelo F. C. Gomes
Bruno Gonçalves

Springer

MOBS LAB

M.Ajelli, M. Chinazzi, M.Litvinova,
D.Mistry, A. Pastore y Piontti,
K.Sun, S.Haque, N. Samay,
Q.Zhang,
(Northeastern University, USA)

M.E. Halloran
(Fred Hutchinson Cancer Research Center, USA)

N. Dean, D. Rojas,
I.M. Longini
(University of Florida, USA)

N.Perra
(Greenwich University, UK)

S. Merler, L. Fumanelli, P. Poletti
FBK, Trento, Italy

C.Gioannini, L.Rossi, M.Quaggiotto,
M.Tizzoni, D Perrotta, D.Paolotti,
P. Milano, M.Selim, E.Ubaldi,
(Scientific Interchange Foundation, Italy)

C.Poletto, V.Colizza
INSERM, Paris

G.Chowell
(Georgia State University)

C.Viboud
(Fogarty, NIH, USA)

L.Simonsen
(George Washington University, USA)