CENTER FOR INFERENCE & DYNAMICS OF INFECTIOUS DISEASES

Northeastern

Pandemics in the age of Big Data

Alessandro Vespignani

@alexvespi

MOBS LAB

LABORATORY FOR THE MODELING OF BIOLOGICAL AND SOCIO TECHNICAL SYSTEMS

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

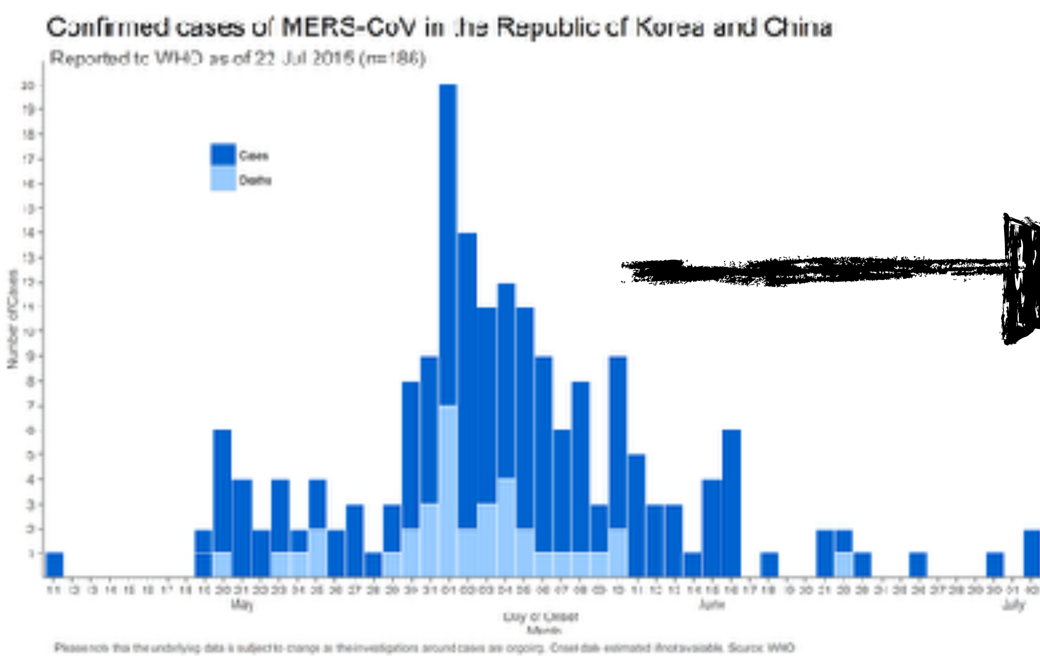
G.Chowell
(Georgia State University)

C.Viboud
(Fogarty, NIH, USA)

L.Simonsen
(George Washington University, USA)

Epidemic modeling

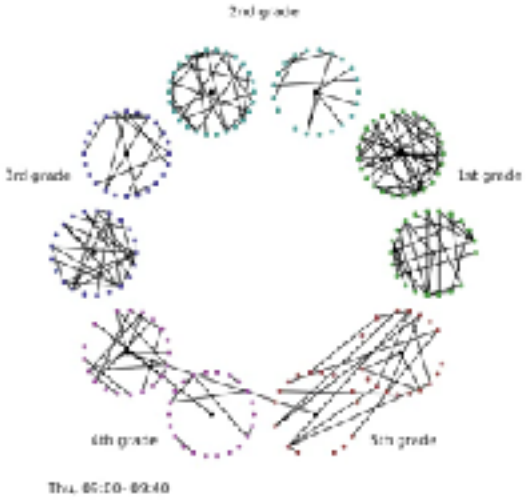
“I simply wish that, in a matter which so closely concerns the wellbeing of the human race, no decision shall be made without all the knowledge which a little analysis and calculation can provide”
Daniel Bernoulli ~1760



MOBS LAB

DATA AVAILABILITY HAS EXPLODED

Human interactions/ contact networks

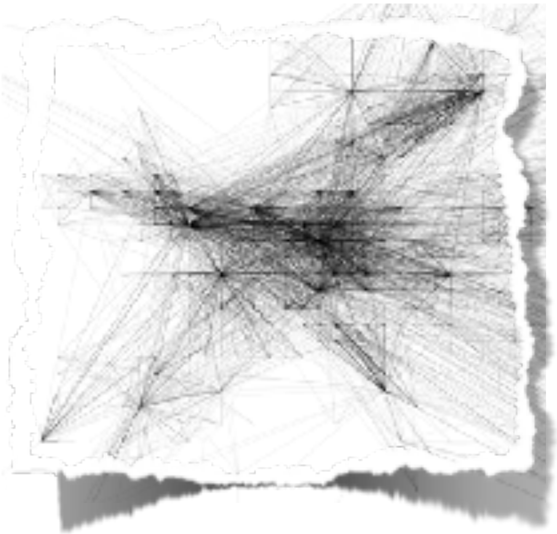


Within school contact patterns (@Sociopatterns)

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, ...

Big Data

Big data narrative, fourth paradigm etc.

Large number of pitfalls:

- no dynamical understanding
- black box (algorithmic issues)
- See Hosni, Vulpiani, Philosophy &Technology (2017) MUST READ!

~~BIG DATA~~ → NEW DATA

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
- Effective equations
- Initial conditions
- Model’s calibration

A MULTIDISCIPLINARY APPROACH TO EPIDEMIC ANALYSIS

01

MODELING

Elaborate stochastic infectious-disease models to support a wide range of epidemiological studies, covering different types of infections and intervention scenarios.

02

REAL-WORLD DATA

Real-world data on population and mobility networks are integrated in structured spatial epidemic models to generate data-driven simulations of the worldwide spread of infectious diseases.

03

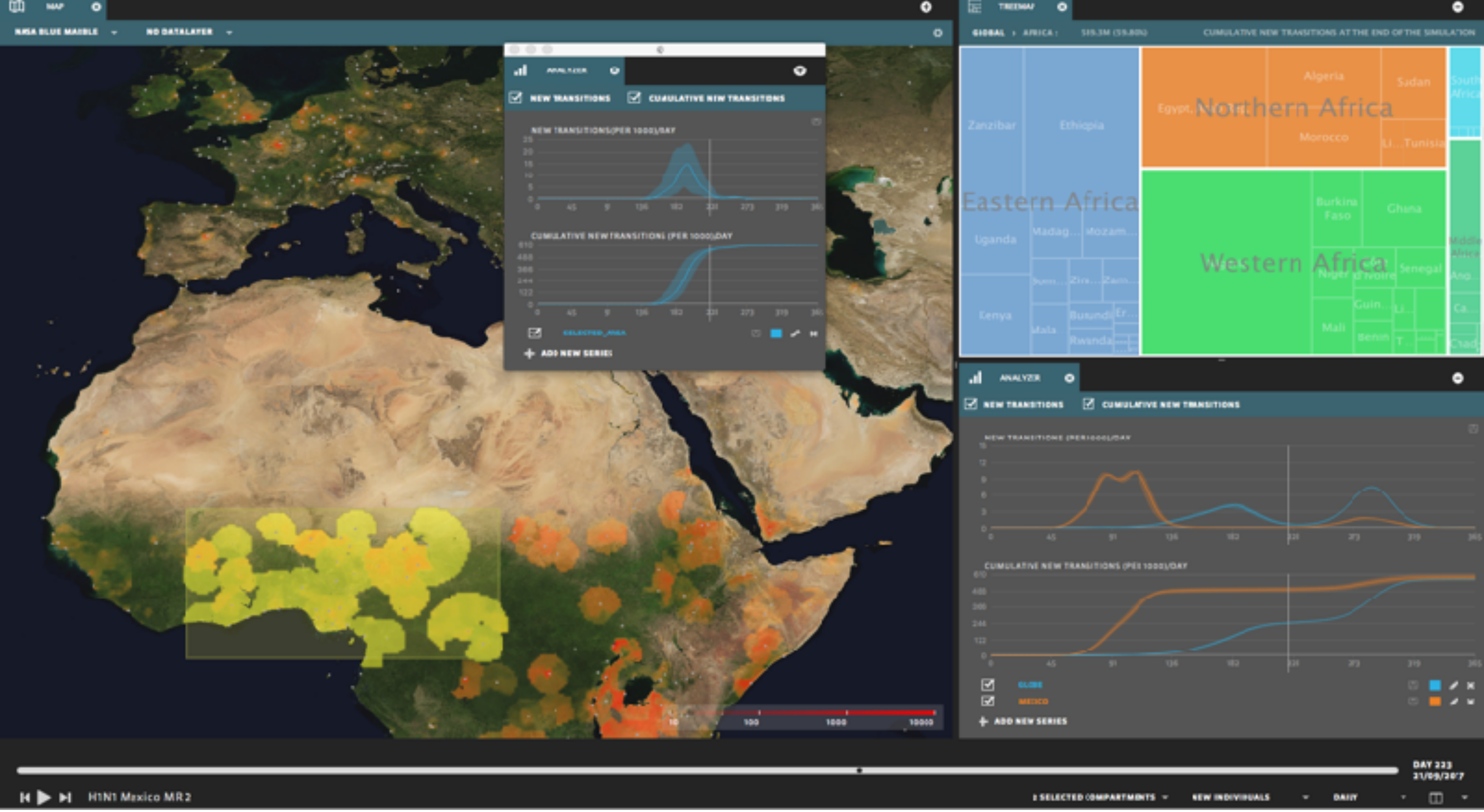
COMPUTATIONAL THINKING

The computer is the laboratory. Models run on high performance computers to create in-silico experiments that would be hardly feasible in real systems, to guide our understanding of typical non-linear behavior and tipping points of epidemic phenomena.

04

TOOLS DEVELOPMENT

Computational tools help in modeling the spread of a disease, understanding observed epidemic patterns, and studying the effectiveness of different intervention strategies. These tools are available to researchers, healthcare professionals, and policy makers.



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

ARPA BE THE FUTURE

CENTER FOR INFERENCE & DYNAMICS OF INFECTIOUS DISEASES

MIDAS

NATIONAL INSTITUTES OF HEALTH

MOBS LAB

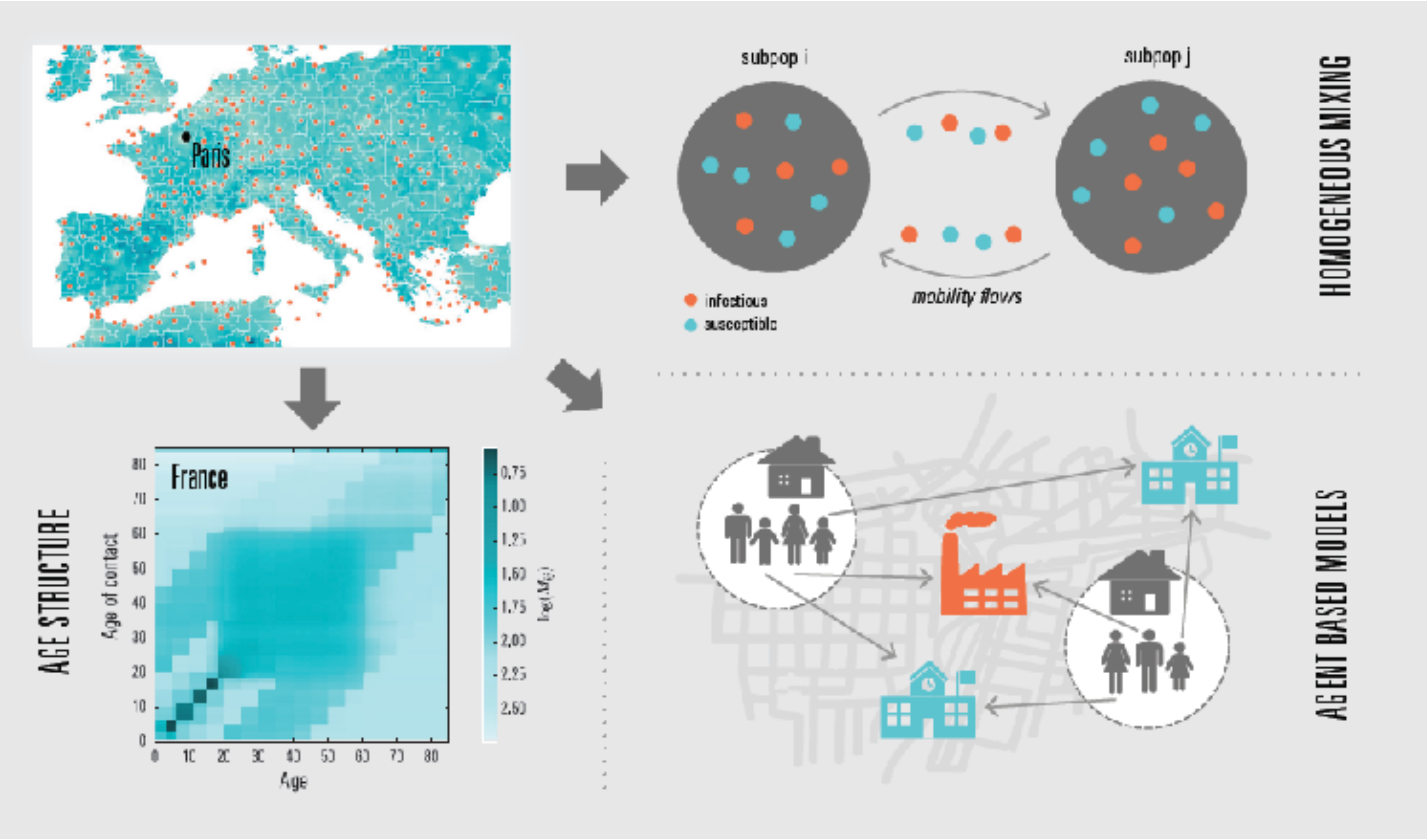
Stochastic Inter population dynamics

Figure 2.9 | Directed flight connections from New York

In orange we show the locations reachable from New York via one single flight. *Inset in orange we show the cells within a circle of 100 miles from New York connected by short range mobility.*

MOBS LAB

Multiple schemes for the stochastic intra-population contagion dynamic



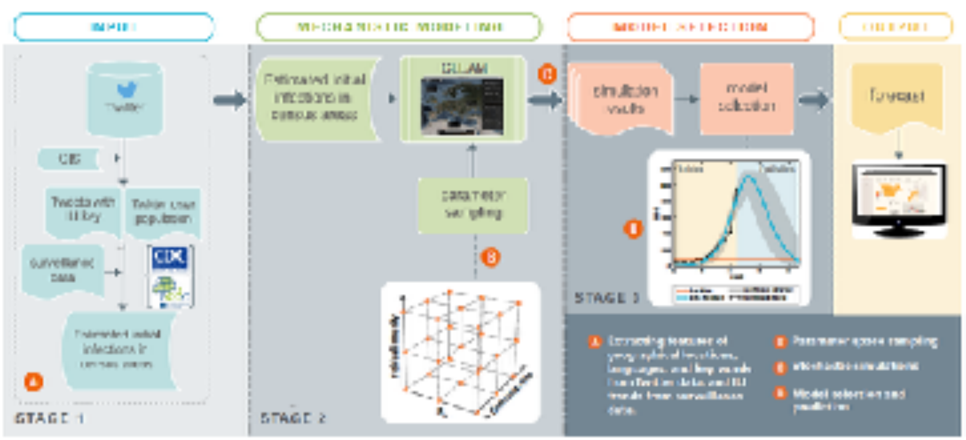
Applications

- SARS Epidemic (retrospective)
- 2009 H1N1pdm
- 2014 West Africa Ebola Epidemic
- Zika epidemic in Latina America
- CDC yearly challenge since 2012

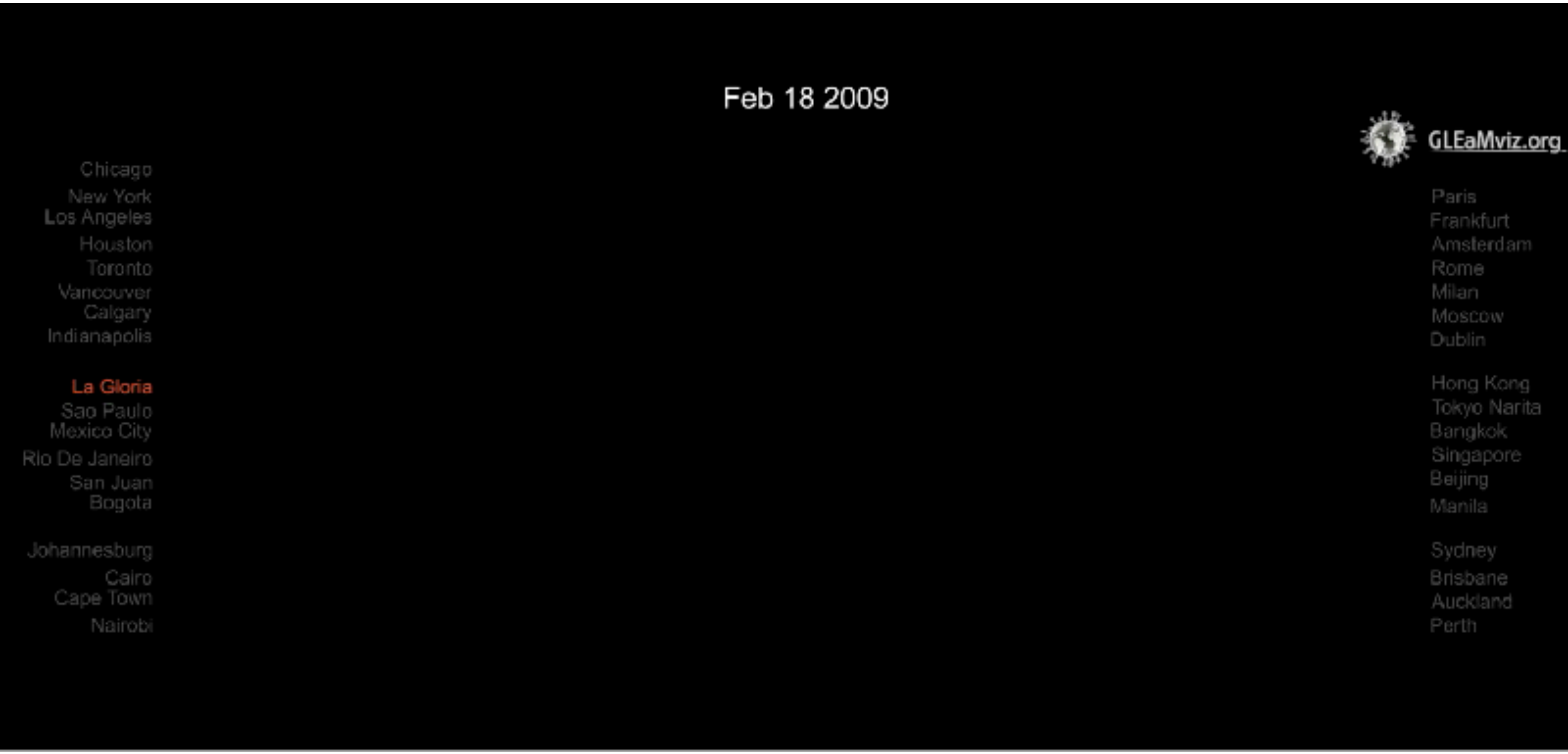
This work has been in collaboration with the CIDID, MIDAS network and a number of national and International Institution.

Research & Development

- Predictability limits/frameworks (Ebola synthetic challenge)
- Synthetic contact matrices
- Forecasting Mechanistic approach with the help of novel data streams.



H1N1 ppm invasion tree



Spatial stochastic individual based model

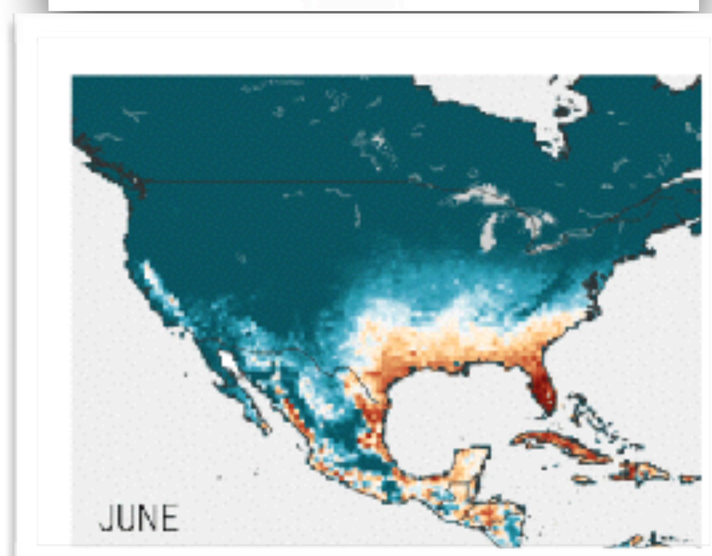
Zhang et al. PNAS 2017 ; doi:10.1073/pnas.1620161114

- 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.



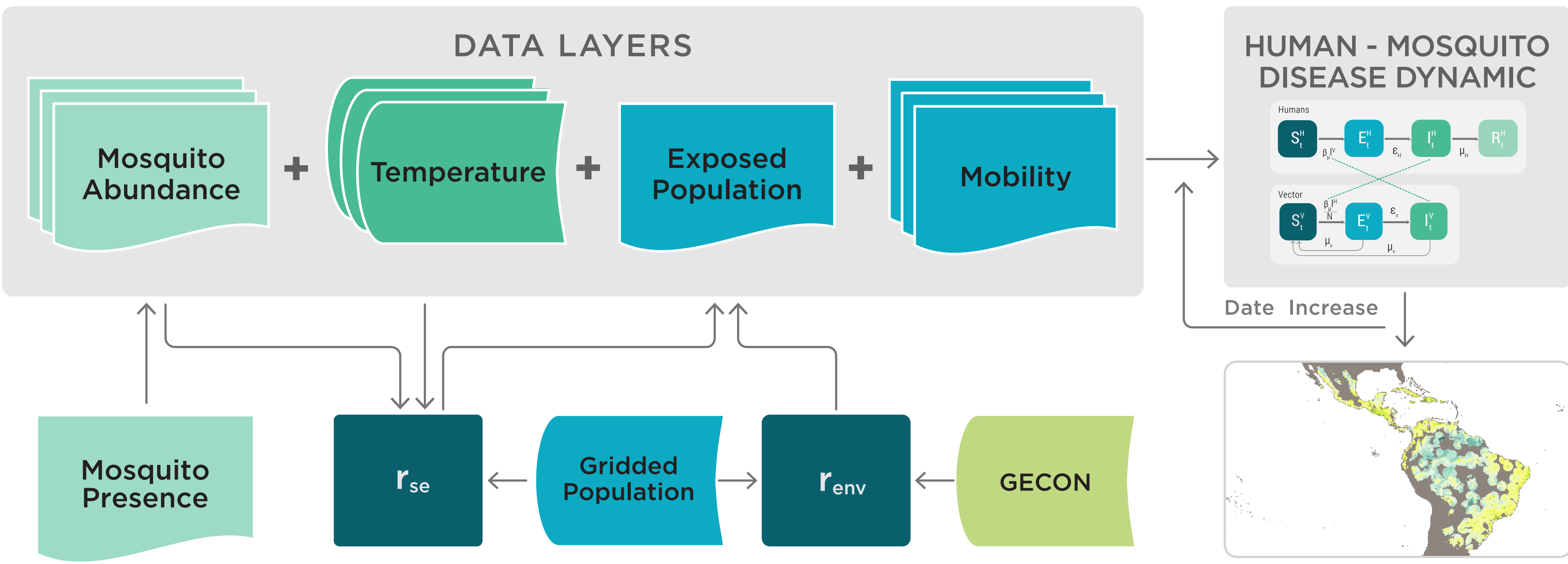
MOBS LAB

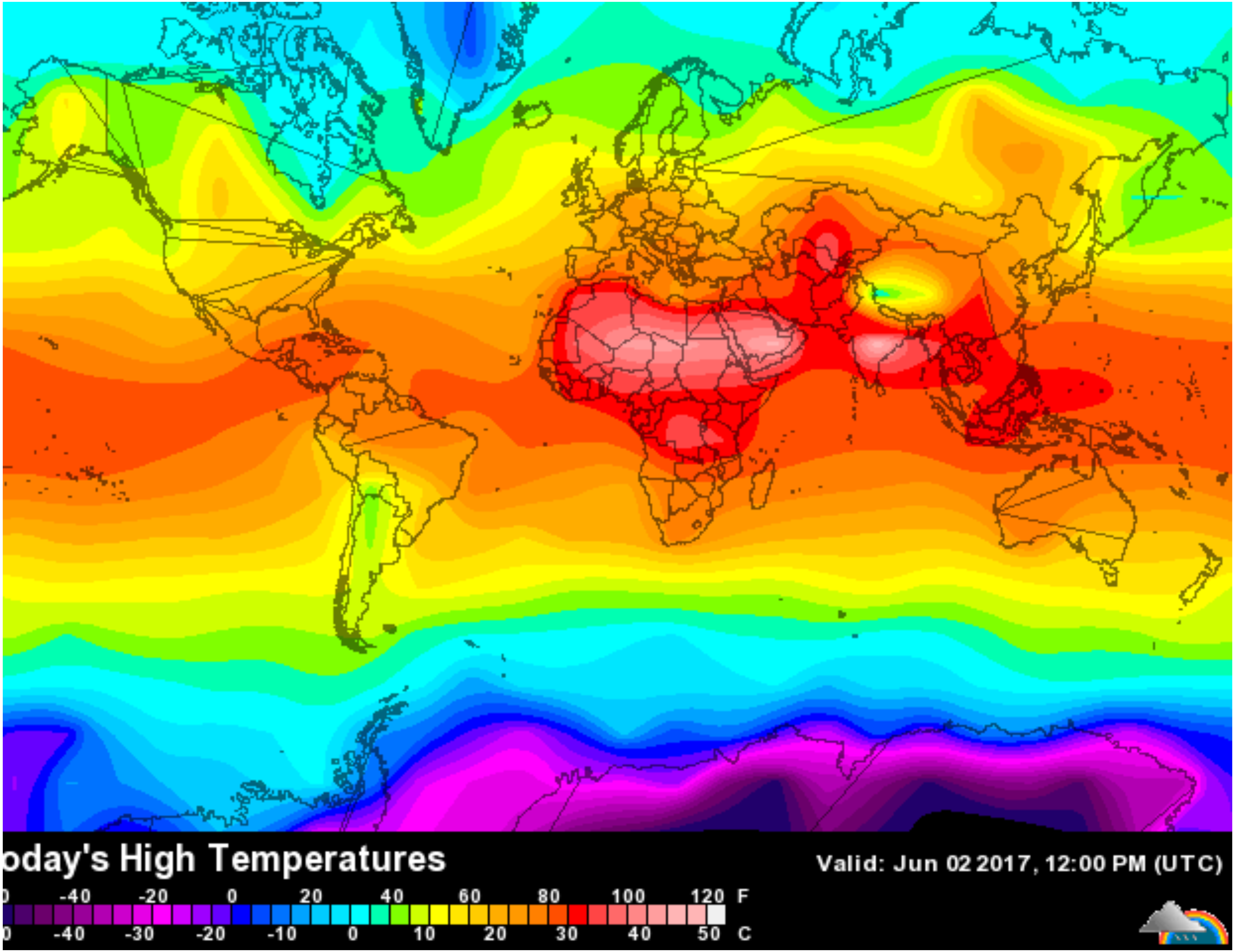
Model structure

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

Mosquitoes abundance + local climate drivers + socio-economic indicators

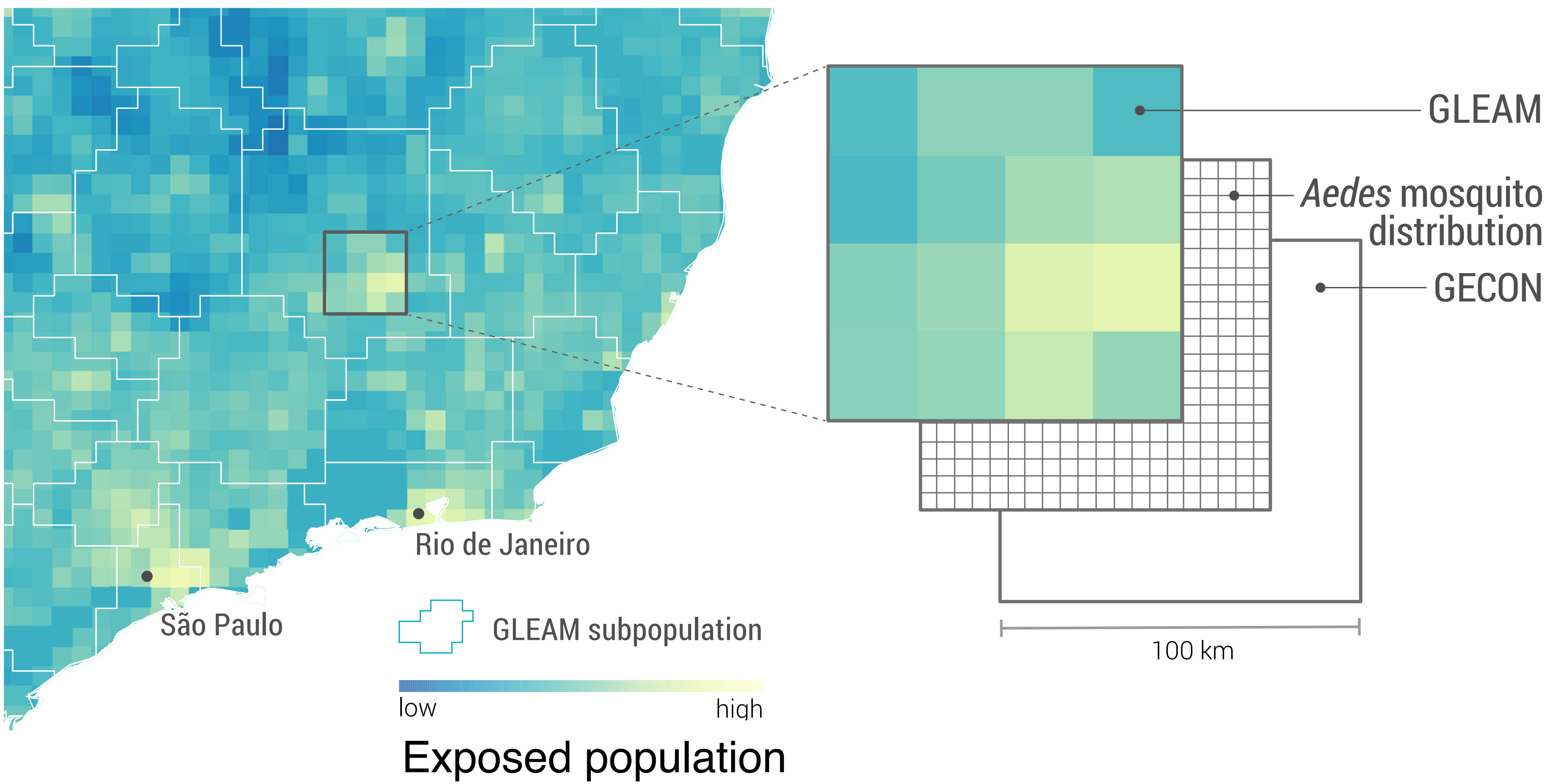
Dynamic stochastic model providing time evolution of the epidemic




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

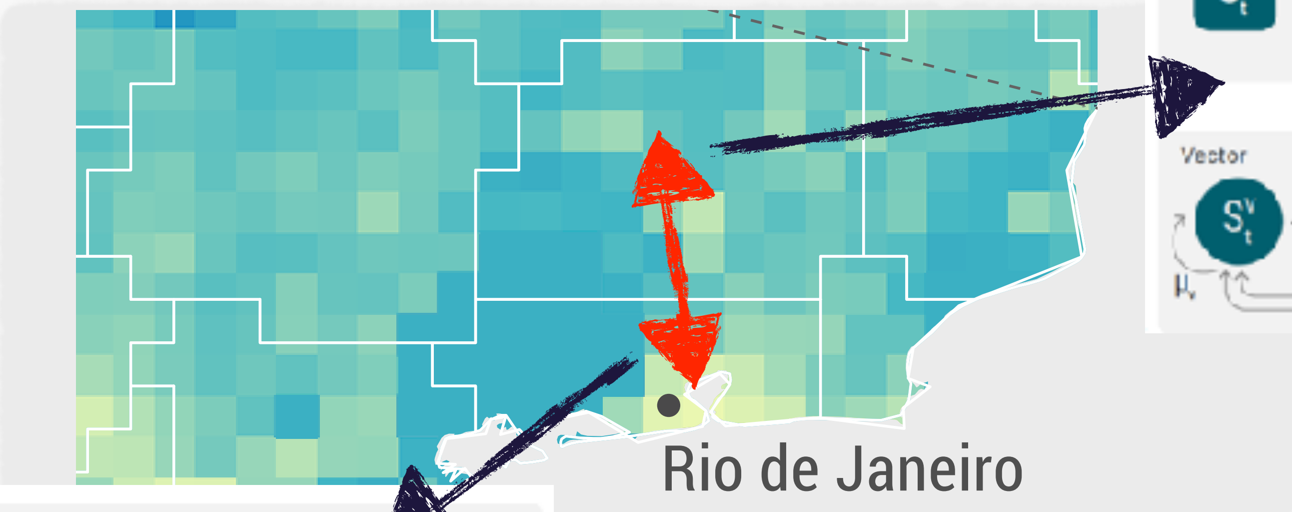
Model resolution

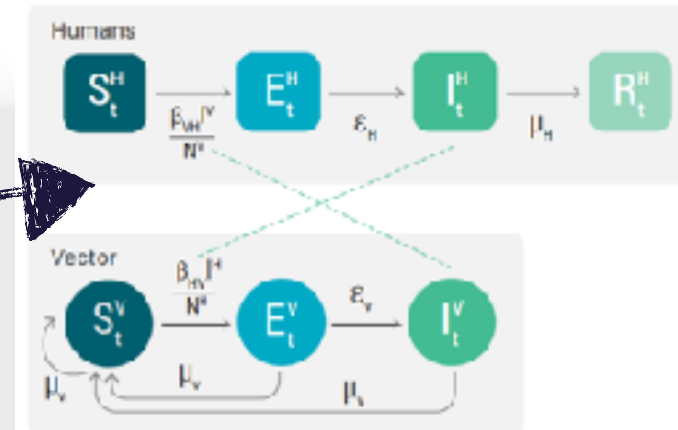
25 km x 25 km within census areas in all the Americas.
A few quantities can be projected up to 1km x 1 km

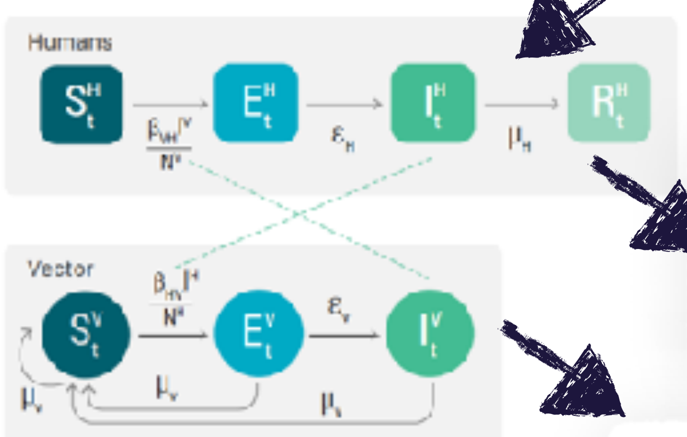


MOBS LAB

EPIDEMIC DYNAMICS





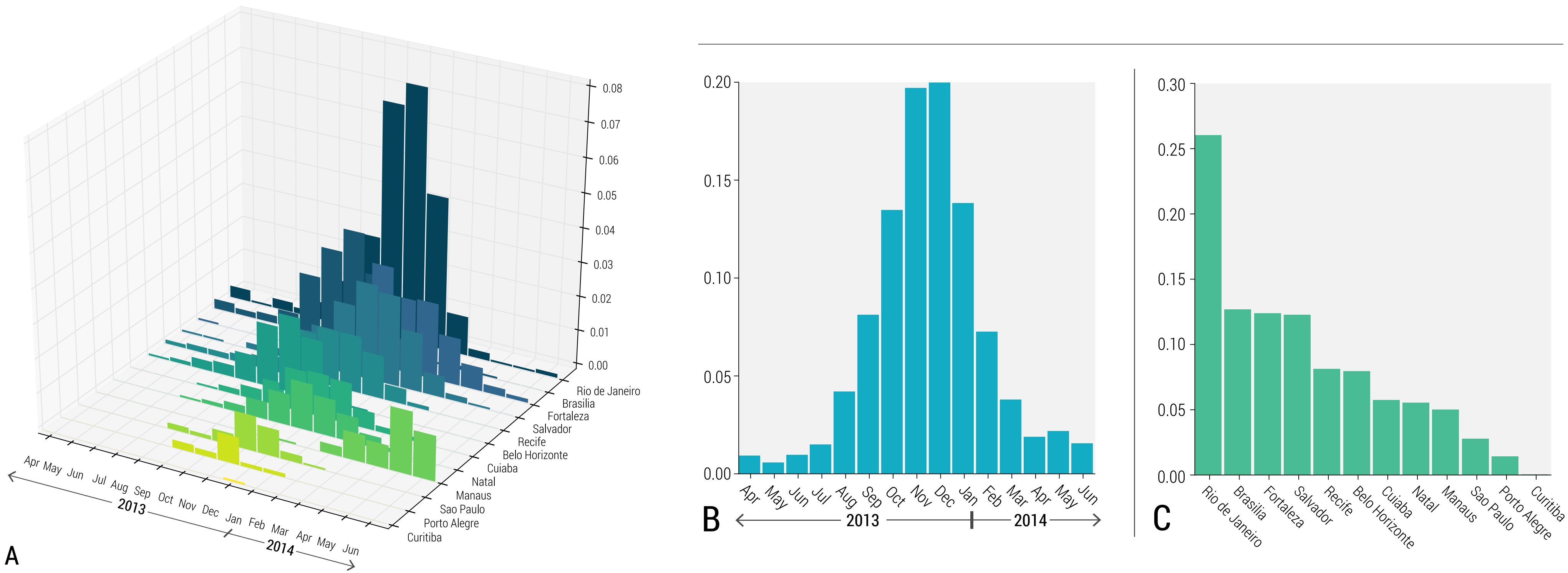


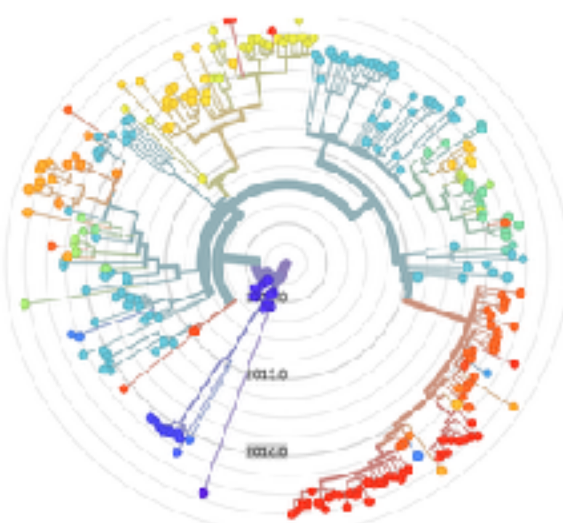
$$
\begin{aligned}
S^H_{t+1} &= S^H_t - Bin(S^H_t, \lambda^H_t) \\
E^H_{t+1} &= E^H_t + Bin(S^H_t, \lambda^H_t) - Bin(E^H_t, \epsilon_H) \\
I^H_{t+1} &= I^H_t + Bin(E^H_t, \epsilon_H) - Bin(I^H_t, \mu_H) \\
R^H_{t+1} &= R^H_t + Bin(I^H_t, \mu_H),
\end{aligned}
$$

$$
\begin{aligned}
S^V_{t+1} &= S^V_t - Bin(S^V_t, \lambda^V_t) + Bin(I^V_t, \mu_V) + Bin(E^V_t, \mu_V) \\
E^V_{t+1} &= E^V_t - Bin(E^V_t, \mu_V) - Bin(E^V_t, \epsilon_V) + Bin(S^V_t, \lambda^V_t) \\
I^V_{t+1} &= I^V_t + Bin(E^V_t, \epsilon_V) - Bin(I^V_t, \mu_V),
\end{aligned}
$$

MOBS LAB

Epidemiological explanations: Monte-Carlo analysis of ZIKAV introduction in the Americas

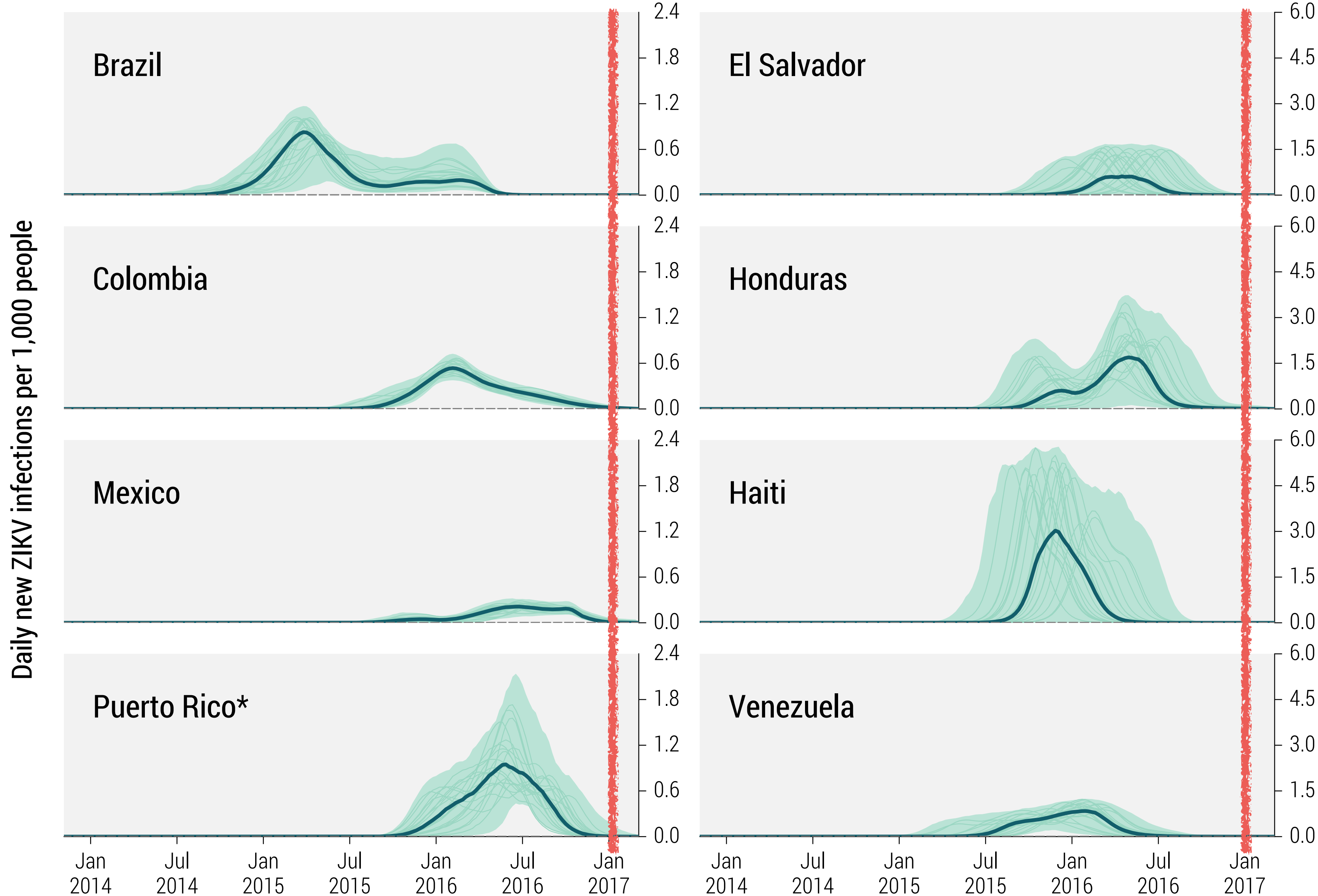




Model provides posterior distributions for the place and date of introduction in Brazil in good agreement with Phylogenetic analysis

MOBS LAB

Epidemic declining in 2017



MOBS LAB

Number of infections

Attack Rate and Microcephaly Cases, Model Projection (Median with 95%CI)

Date of Projection	Attack Rate %
By Feb. 01, 2016	4.33 [2.73 - 6.83]
By Feb. 28, 2017	11.9 [10.5 - 13.5]

Date of Projection	Microcephaly Cases* (0.95 % first trimester risk)	Microcephaly Cases* (2.19 % first trimester risk)	Microcephaly Cases* (4.52 % first trimester risk)
By Feb. 01, 2016	0 [0 - 4]	0 [0 - 10]	1 [0 - 20]
By Dec. 10, 2017	219 [194 - 248]	504 [447 - 572]	1041 [922 - 1180]

*Estimates for microcephaly cases are based on risk calculation provided in Michael A. Johansson et al. New England Journal of Medicine (2016); Cauchemez et al. The Lancet 387, 2125-2132 (2016). Estimates may change as new and/or more precise risk multipliers will be available.

**The incidence profile is the median incidence calculated each week from the stochastic ensemble output of the model and it may not be representative of specific epidemic realizations.

Colombia

Attack Rate and Microcephaly Cases, Model Projection (Median with 95%CI)

Date of Projection	Attack Rate %
By Feb. 01, 2016	1.93 [0.02 - 6.83]
By Feb. 28, 2017	19.6 [13.3 - 27.6]

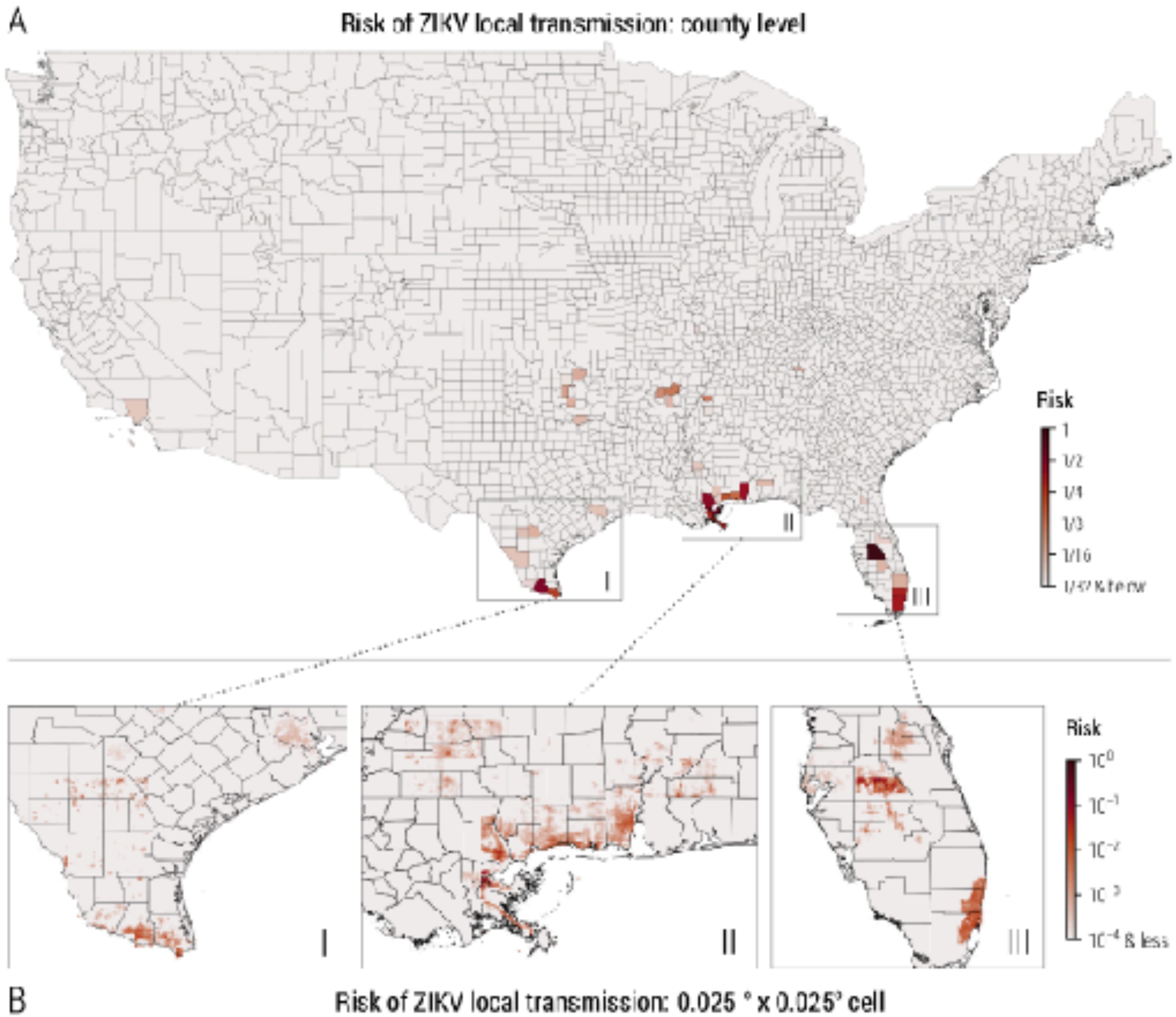
Date of Projection	Microcephaly Cases* (0.95 % first trimester risk)	Microcephaly Cases* (2.19 % first trimester risk)	Microcephaly Cases* (4.52 % first trimester risk)
By Feb. 01, 2016	0 [0 - 0]	0 [0 - 0]	0 [0 - 0]
By Dec. 10, 2017	19 [13 - 26]	43 [29 - 60]	88 [60 - 124]

*Estimates for microcephaly cases are based on risk calculation provided in Michael A. Johansson et al. New England Journal of Medicine (2016); Cauchemez et al. The Lancet 387, 2125-2132 (2016). Estimates may change as new and/or more precise risk multipliers will be available.

**The incidence profile is the median incidence calculated each week from the stochastic ensemble output of the model and it may not be representative of specific epidemic realizations.

Puerto Rico

Risk of local transmission in the US



risk	population at risk	% of total population
> 1/2	2.27 million	0.71 %
>1/4	7.68 million	2.40 %
> 1/8	8.20 million	2.56 %
> 1/16	[illegible] million	8.50 %

MOBS LAB

COMPUTATIONAL IMPLEMENTATION

HPC

Large-scale numerical simulations

- About 15 million runs (overall)
- More than 1M instances deployed
- About 125 years of computational time

HPDA

From raw data to actionable summary tables

- Hundred of TB of data analyzed
- Each scenario up to more than 10 TB of raw data

STILL A MAJOR HURDLE FOR REAL-TIME GLOBAL MODELING

WHERE IS THE COMPUTATIONAL PROBLEM

Statistical ensemble

Each large-scale numerical simulations is a single instance of a stochastic process.

- Thousands (or more) of runs to extract statistical patterns
- The finer the scale, the larger the required number of realizations
- Real time requirement

Model calibration

Calibration on real data is executed via MCL or MCMC scanning high dimensional parameters space

- Millions (or more) of single realizations exploring the level of match with real data.
- Real time requirement

Real-time analysis of large-scale/highly detailed models requires increasing amount of computational resources.

Supercomputing culture must be more and more a key component of the pandemic modeling and analysis research.

Interdisciplinary endeavor not possible without the contribution of the HPC