SCHOOL OF INFORMATICS
AND COMPUTING
INDIANA UNIVERSITY
Bloomington

Pervasive Technology Institute

Predicting the behavior of techno-social systems

Alessandro Vespignani

CNetS
Center for Complex Networks
and Systems Research

Collaborators....

CNetS -Center for Complex Networks and Systems Research, Indiana University

- *D. Balcan*
- *B. Goncalves*
- *H. Hu*
- *Nicola Perra*
- *K.Borner*
- *J.Sherman*

INSTITUTE FOR SCIENTIFIC INTERCHANGE FOUNDATION

- *V. Colizza*
- *P. Bajardi*
- *C. Poletto*
- *J. Ramasco*
- *M. Tizzoni*
- *V. Van den Broeck*

UPC

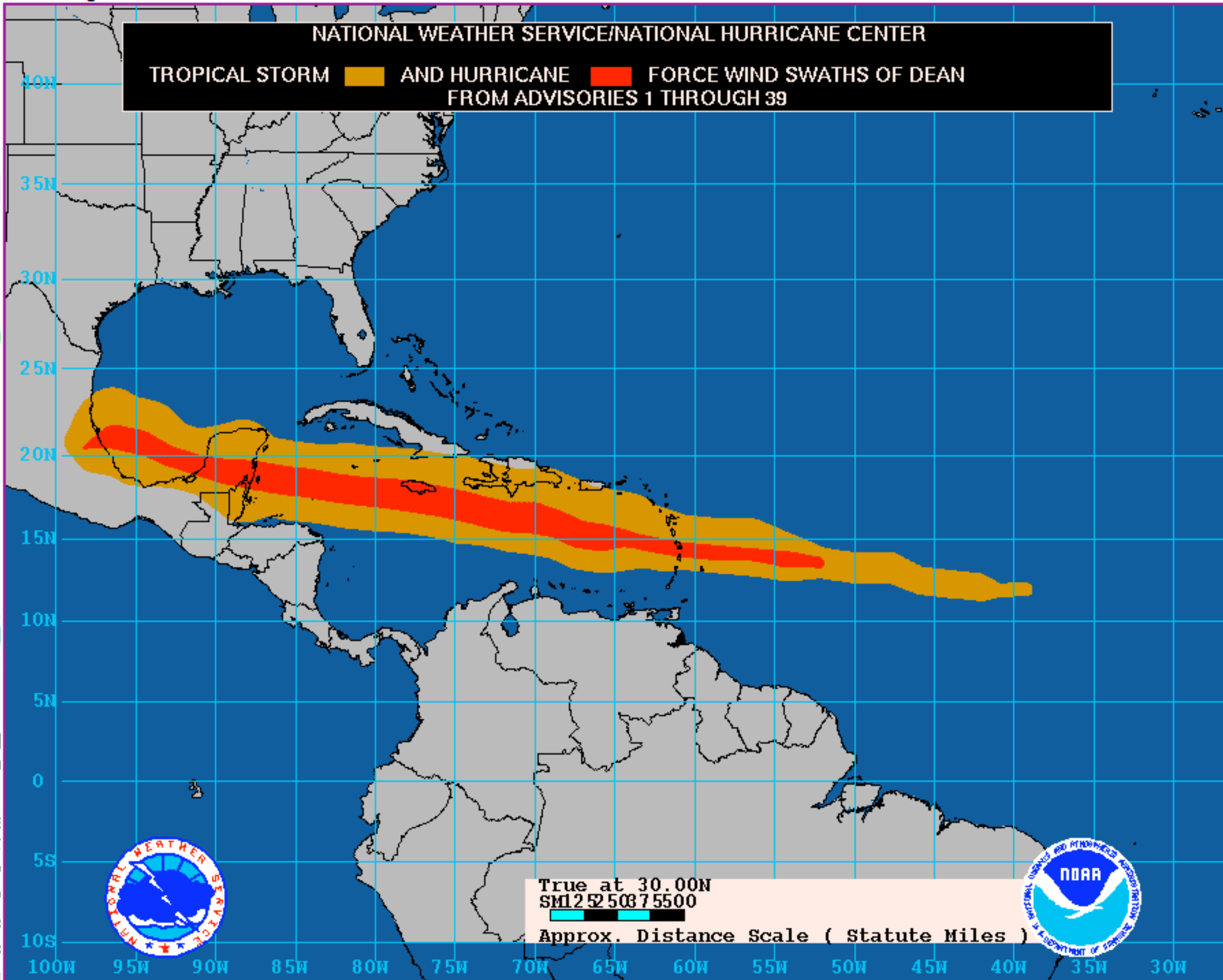
CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

cea

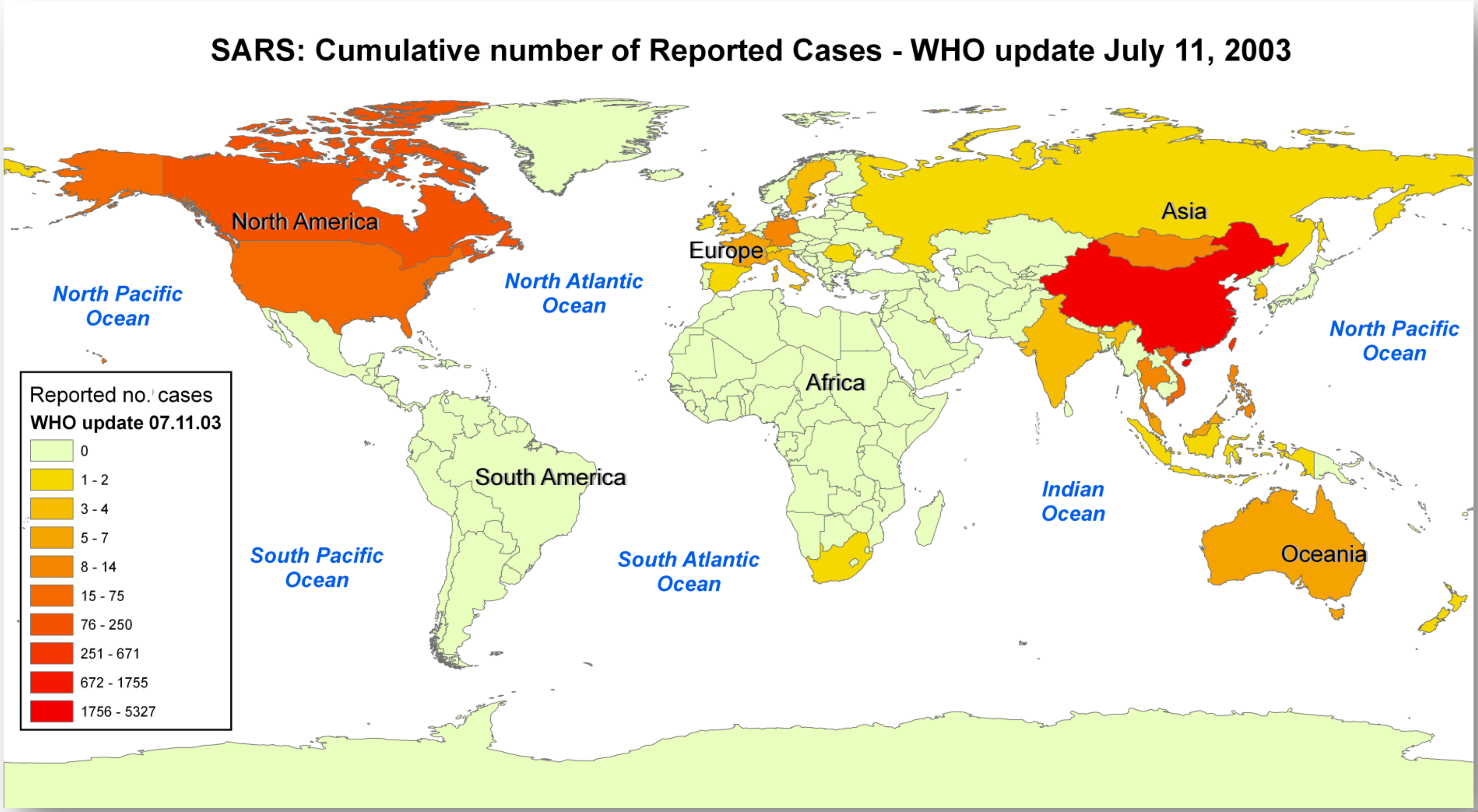
FBK FONDAZIONE BRUNO KESSLER RICERCA SCIENTIFICA E TECNOLOGICA

- *M. Ajelli*
- *A. Barrat*
- *M. Barthelemy*
- *S. Merler*
- *R. Pastor Satorras*
- *A.J.Valleron*

NIH, DTRA, NAKFI, Lilly foundation,Abbott FET-program


NATIONAL WEATHER SERVICE/NATIONAL HURRICANE CENTER
TROPICAL STORM AND HURRICANE FORCE WIND SWATHS OF DEAN
FROM ADVISORIES 1 THROUGH 39
40N
35N
30N
25N
20N
15N
10N
5N
0
5S
10S
100W
95W
90W
85W
80W
75W
70W
65W
60W
55W
50W
45W
40W
35W
30W
True at 30.00N
Approx. Distance Scale (Statute Miles)
NOAA

Why we do not forecast pandemic evolution



Collective Bio-Techno-social problems ….

In other words

The complete temperature analysis of the sea surface, and satellite images of atmospheric turbulence are easier to get than the large scale knowledge of commuting patterns or the quantitative measure of the propensity of a certain social behavior, or the spreading rate of a given pathogen.

Complexity+complications

- *large numbers of heterogeneous individuals*
- *over multiple time and size scales*
- *Non-linearity, threshold effects, discreteness, cooperation*
- *huge richness of cognitive/social science problems*
- *Technological infrastructures*
- *Bio-medical understanding*

Limits in the forecast of techno-social systems...

•Lack of large scale-data on human behavior and social adaptation (>million individuals, worldwide dimension).

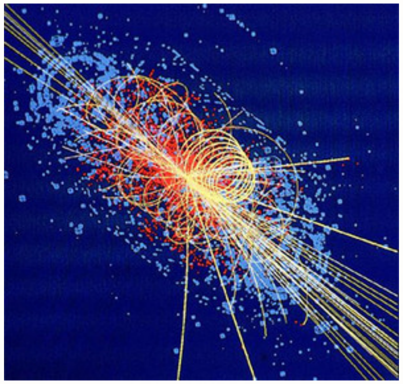
•Undersatnding of the feedback loop
system behavior->prediction -> information+adaptation -> new system behavior...

•Formal models and understanding of social adaptation/reaction during social disruptive events.

•Large scale characterization of interdependencies between technology, infrastructures and social behavior.

WHY A MAJOR FUNDING EFFORT: IT AND THE SOCIETY "TOMOGRAPHY"

- Data integration and assimilation
- Mobile telephone and devices..
- Web 2.0, Proxy networks
- Pervasive and embedded technology/sensors



"Social collider" where the trace of human behavior can be harvested, characterized, rationalized….

Computational predictive tools to inform policy making, drawing scenarios, communicate to the public opinion, anticipate impact, produce knowledge

UNPRECEDENTED AMOUNT OF DATA…..

- Transportation infrastructures
- Behavioral Networks
- Census data
- Commuting/traveling patterns

  - Different scales:
    - International
    - Intra-nation (county/city/municipality)
    - Intra-city (workplace/daily commuters/individuals behavior)

MAJOR OBSTACLES..

- Many data layers (method/language divide)
- New computational tools (skill divide)
- Expertise/cultural divide
- Data sharing-practice divide
- Data Collection
- Ethics/security

  - Different scales:
    - International
    - Disciplines
    - methodology

ETHICAL AND PRIVACY ISSUES

- Geo-localization ☑
- Telephone logs ☑☑
- E-mail/social networks ☑☑☑
- Health data ☑☑☑☑
- Research focus:

From “social atom” or “social molecules” (i.e. small social groups) to the quantitative analysis of “social aggregate states”.

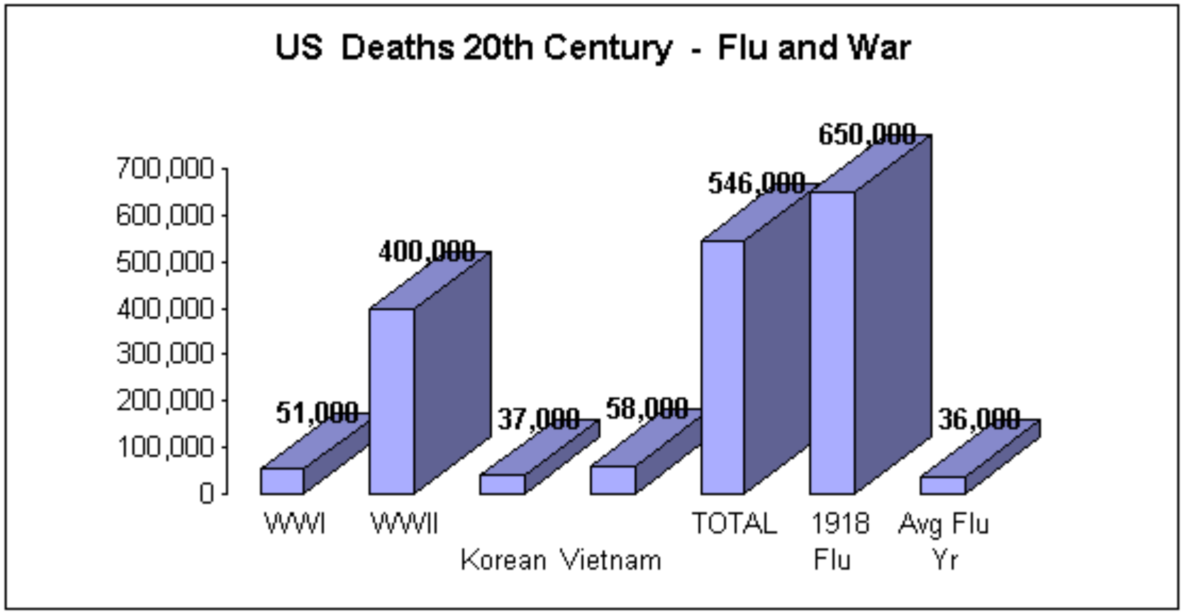
“social aggregate states”= large-scale social systems consisting of millions of individuals that can be characterized statistically in space (geographic and social) and time.

1918 Pandemic

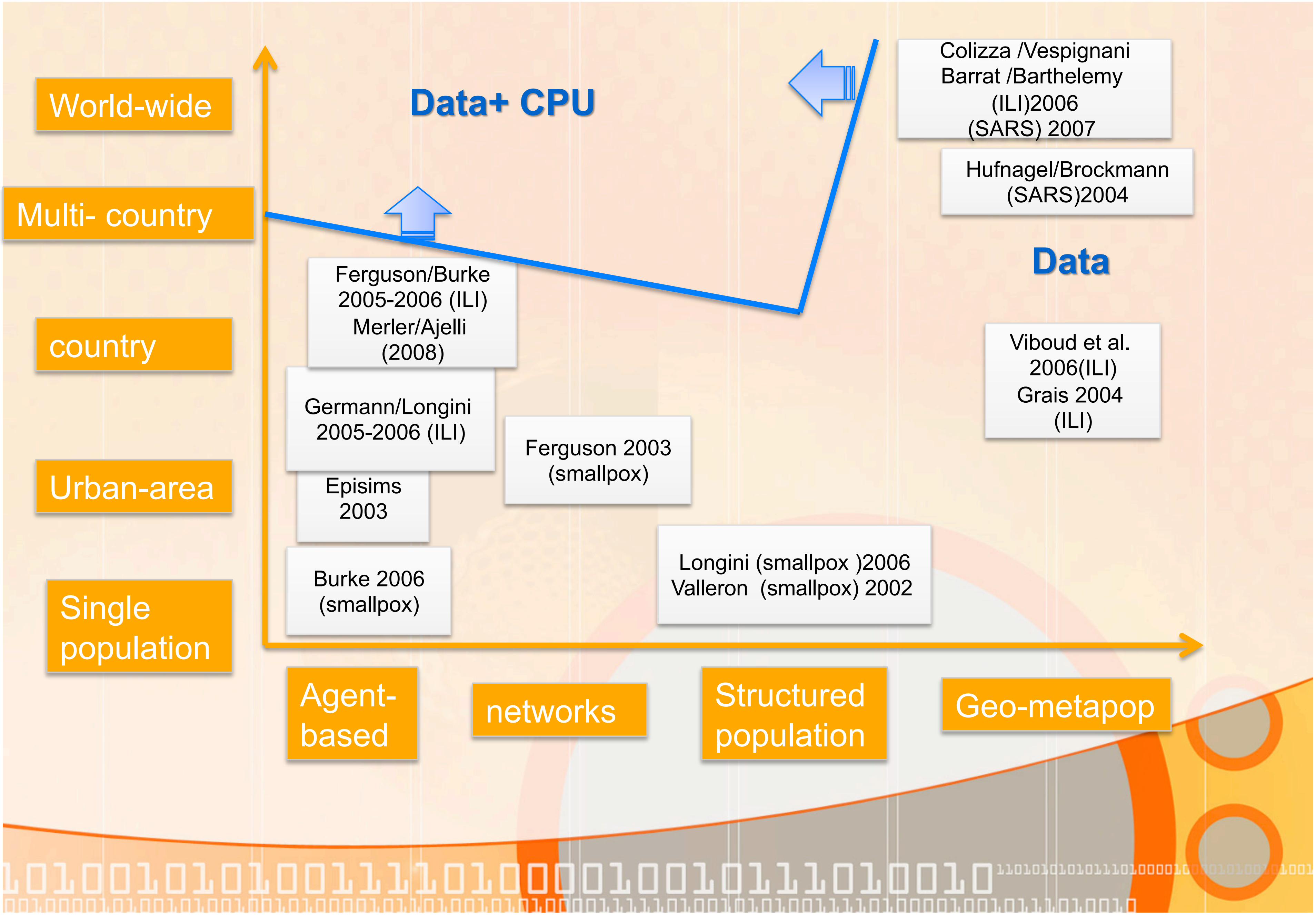
•An estimated 50 million people, about 3% of the world's population (approximately 1.6 billion at the time), died of the disease.

•An estimated 500 million, or 1/3 were infected.[5]

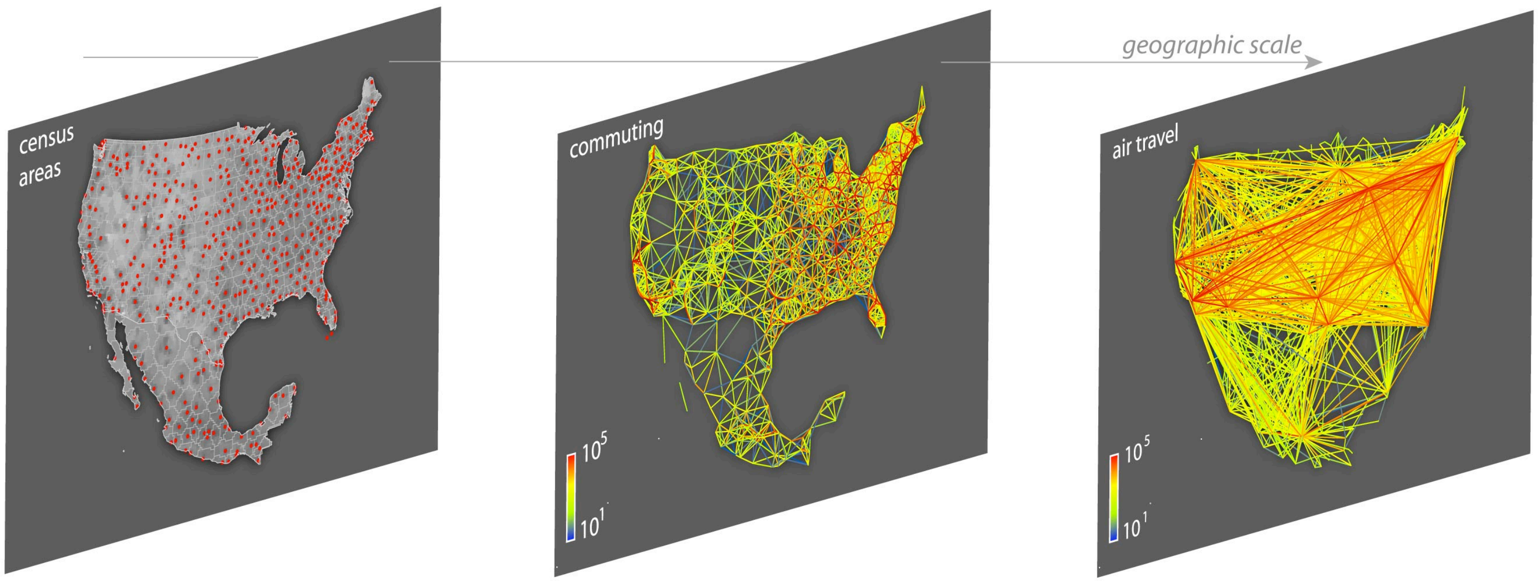
•In the U.S., about 28% of the population suffered, and 500,000 to 675,000 died

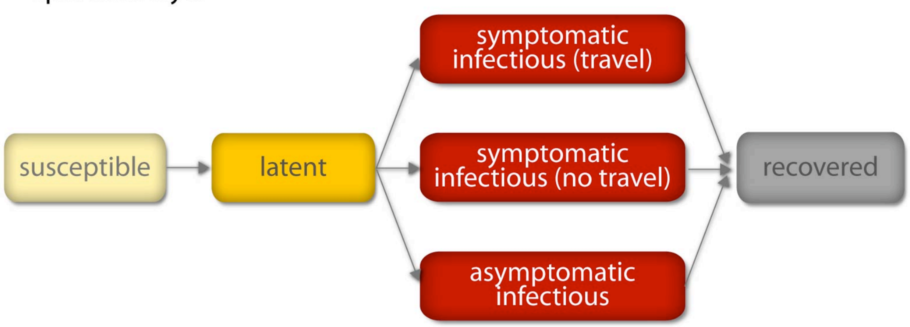


Pandemics & Travel


World-wide
Multi- country
country
Urban-area
Single population
Data+ CPU
Colizza /Vespignani
Barrat /Barthelemy
(ILI)2006
(SARS) 2007
Hufnagel/Brockmann
(SARS)2004
Data
Ferguson/Burke
2005-2006 (ILI)
Merler/Ajelli
(2008)
Germann/Longini
2005-2006 (ILI)
Ferguson 2003
(smallpox)
Viboud et al.
2006(ILI)
Grais 2004
(ILI)
Episims
2003
Burke 2006
(smallpox)
Longini (smallpox)2006
Valleron (smallpox) 2002
Agent-based
networks
Structured population
Geo-metapop

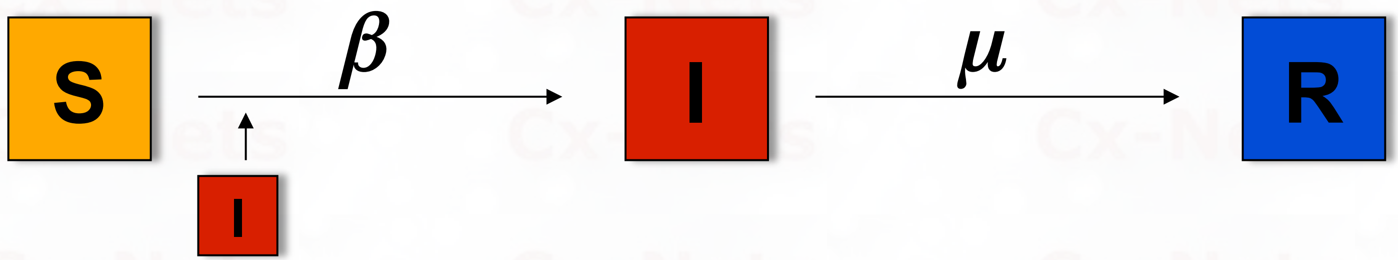
Large scale computational model....GLEaM

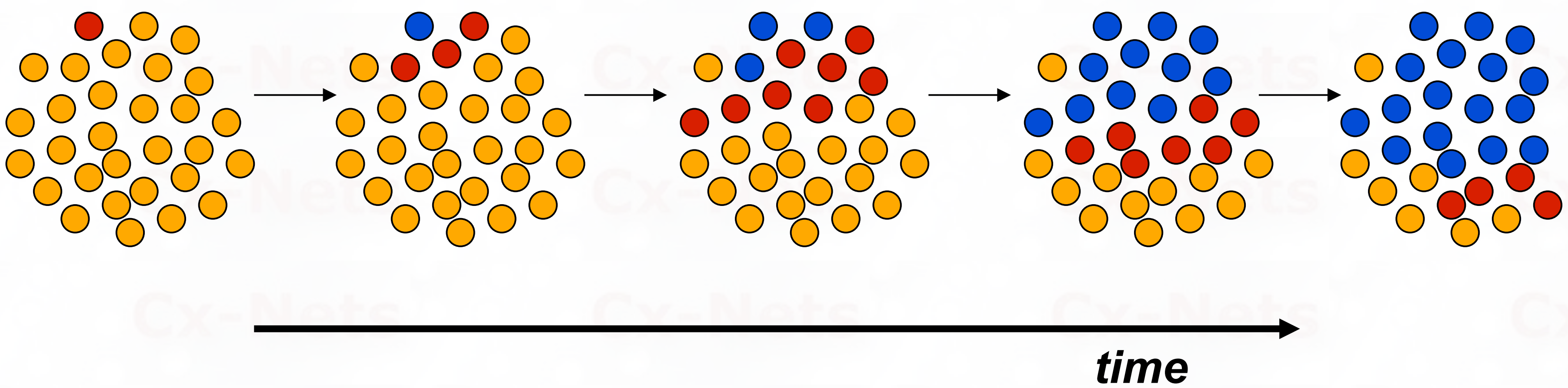




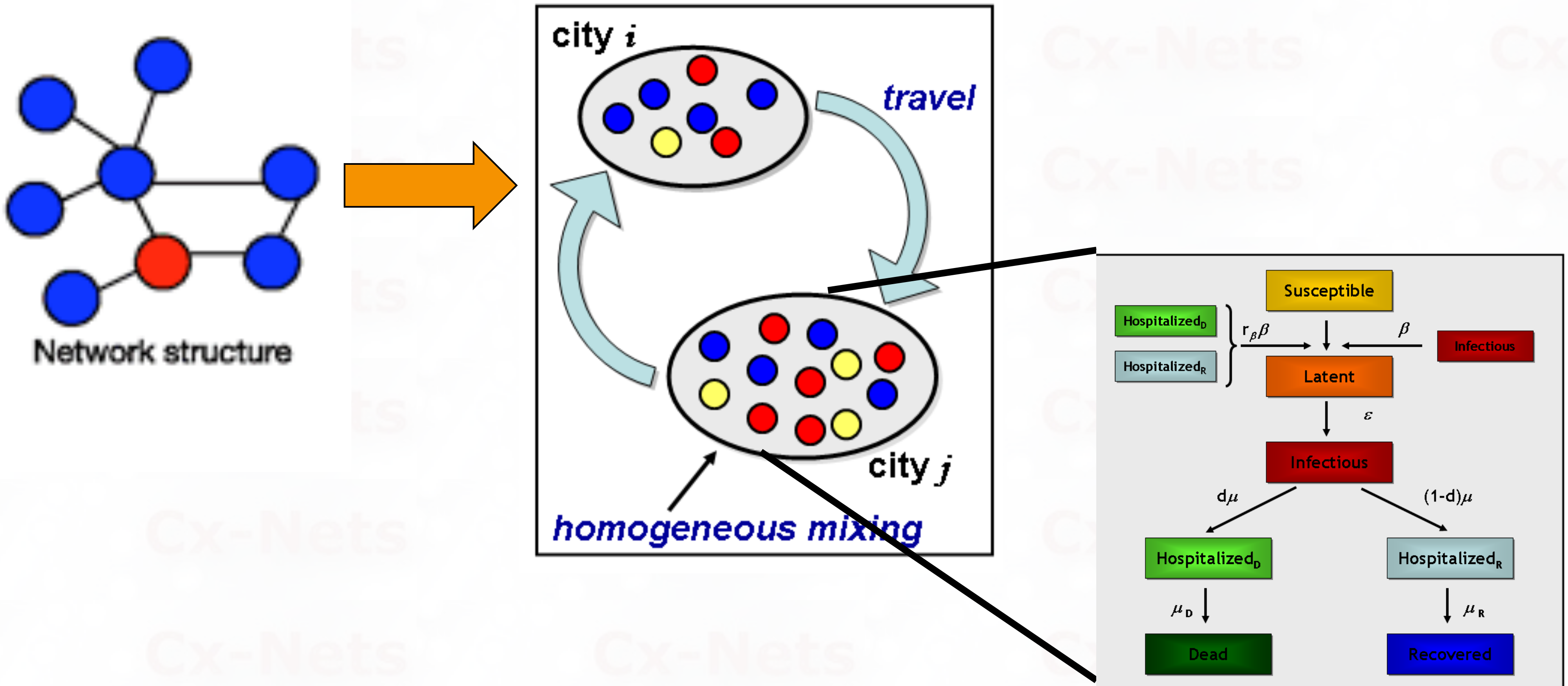
Parameter	Value	Description
β	from R_0	transmission probability
ε^{-1}	1.9 [1.1-2.5] d	average latency period
μ^{-1}	3 [3-5] d	average infectious period
p_t	50%	probability of traveling for infectious individuals
p_a	33%	probability of being asymptomatic
r_β	50%	relative infectiousness of asymptomatic infectious individuals

INTRA-POPULATION: THE **SIR** MODEL



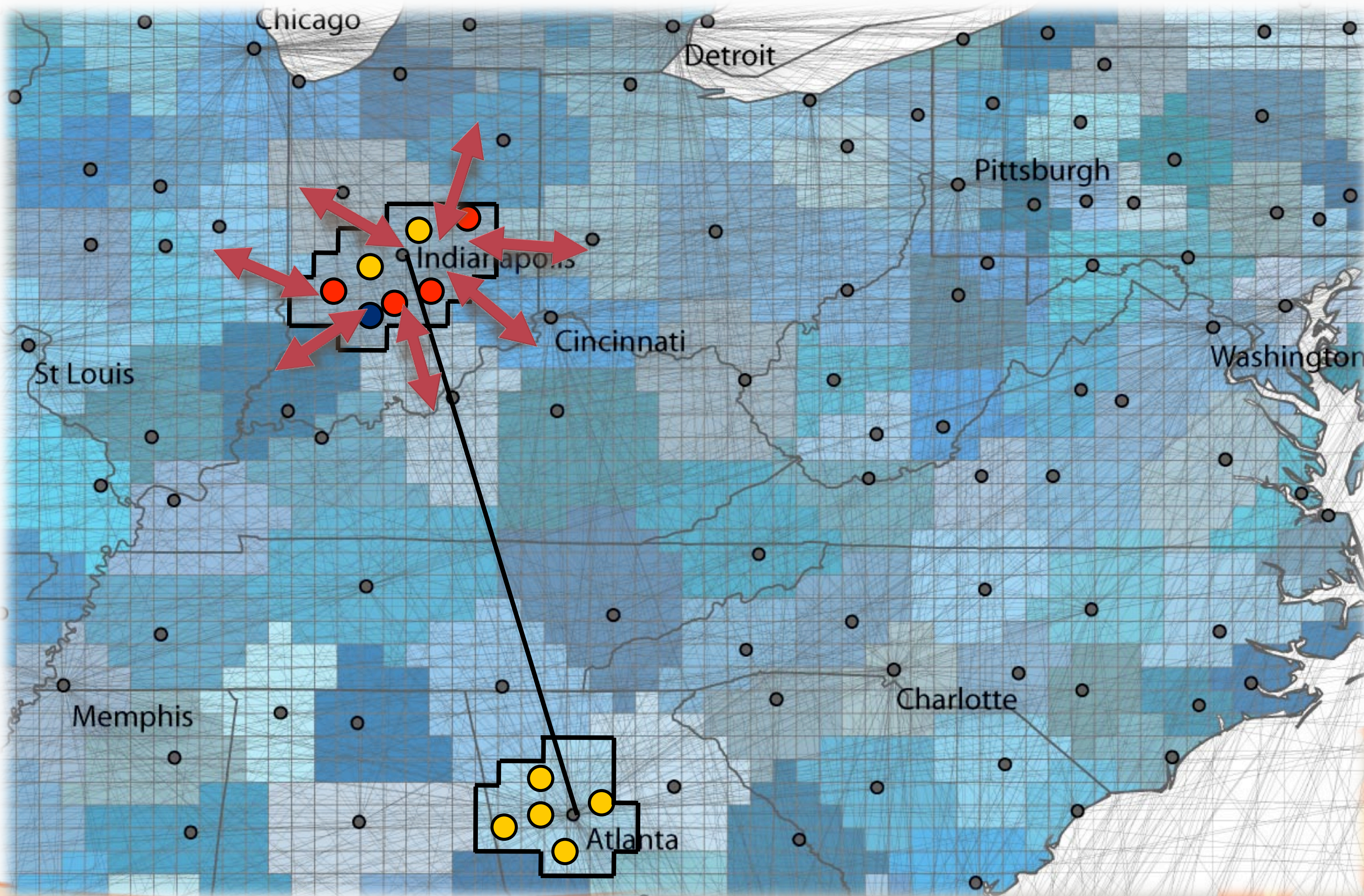


Mechanistic metapopulation models...

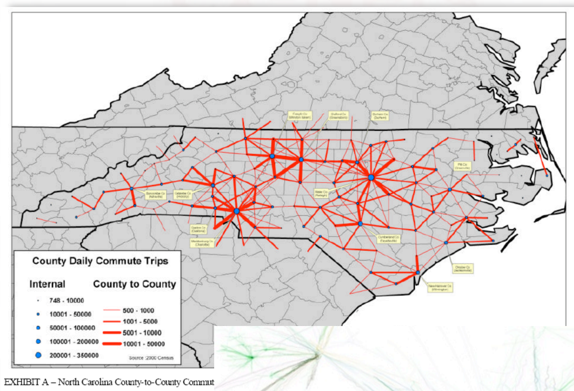


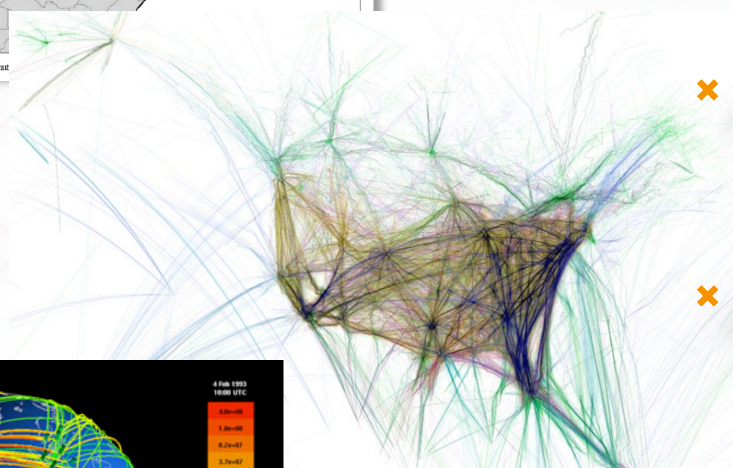
Intra-population infection dynamics

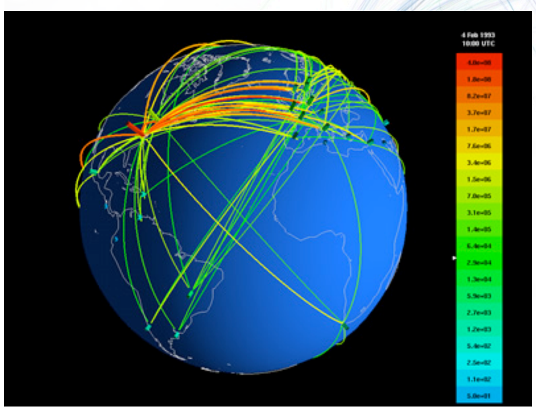
Long and Short range mobility (I)



COMPLEX MOBILITY AND TRAFFIC NETS….





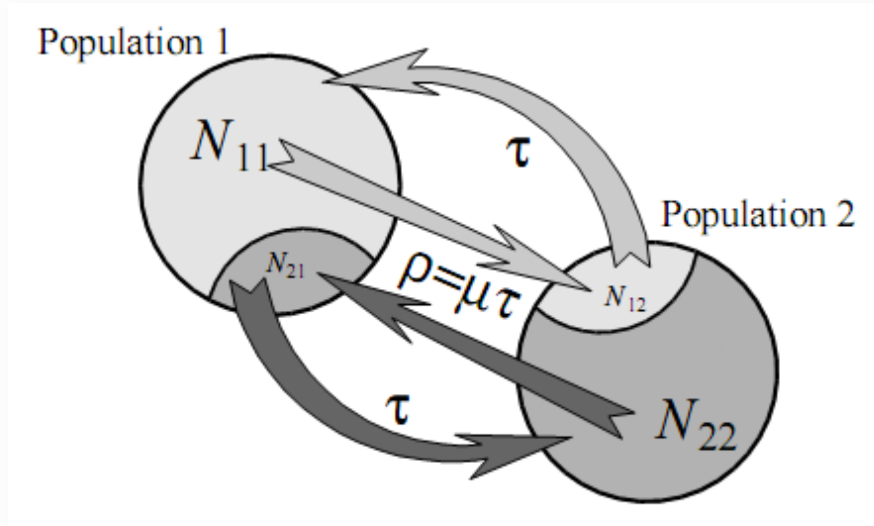


- spatially structured models depend on the notion of identifiable populations in a meta-population.
- For instance, cities, town and villages (on a smaller scale schools, workplaces) and homes, cell tower, WiFi base station area range, etc.
- intuitive segmentations is often fuzzy, ill defined, arbitrary and lacks a systematic backbone.
- How much geography, cultural diversity, spatial and technological variability is encoded in the topology and mobility/diffusion of multi-scale networks
- Which are the topology and traffic/ flows statistical properties of real world networks.

Time scale separation technique

(Sattenspiel & Dietz 1995)
(Keeling & Rohani, 2002)

- *Theoretical description of reaction diffusion processes with memory and "return rates".*
- *define $\mu = \rho / \tau$ (leaving rate / returning rate) ,*
- *N_{ij}: population in i who is visiting region j*
- *N_{ii}: population in i who is actually present in the home region*

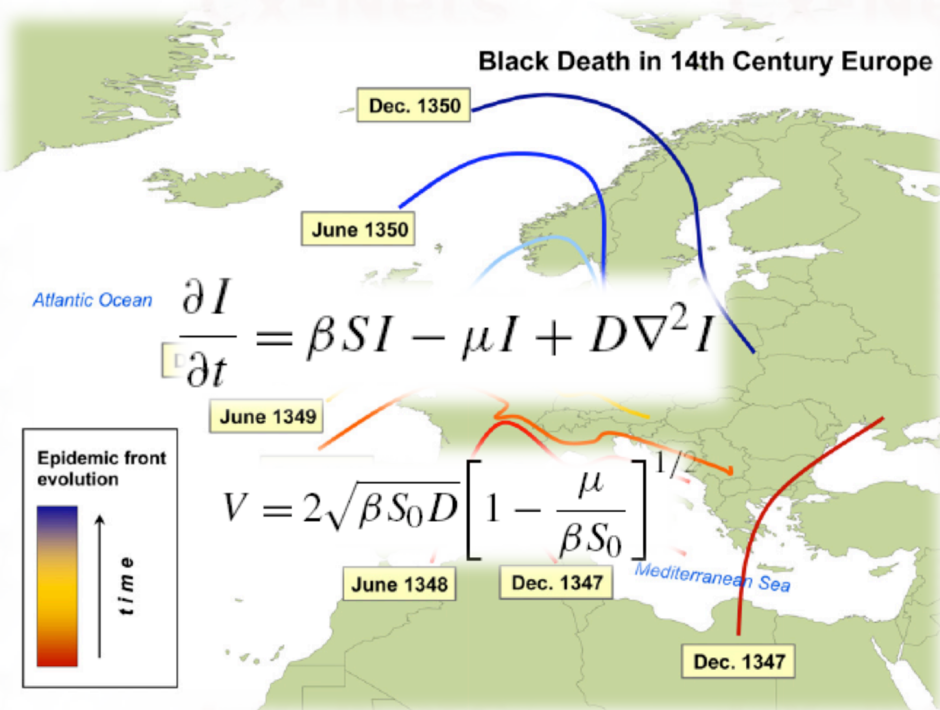


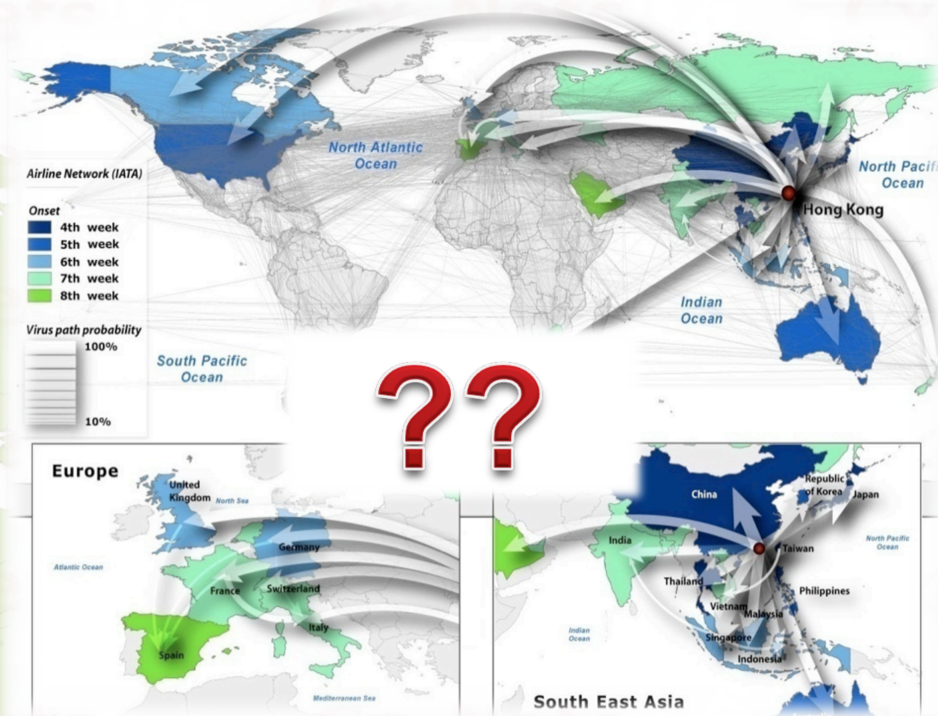
Time scale separation μ <<1 ; heterogeneous population sizes

$$
\begin{aligned}
p_\ell^{\mathcal{D}}(S \to L) \quad = \quad & \frac{\beta_\ell}{(1+\widetilde{\mu}_\ell)N_\ell^*}\left[\frac{r_\beta I_\ell^a + I_\ell^t}{1+\widetilde{\mu}_\ell} + I_\ell^{nt} + \sum_k \frac{\widetilde{\mu}_k \widetilde{\nu}_{k\ell}}{1+\widetilde{\mu}_k}(r_\beta I_k^a + I_k^t)\right]\Delta t \\
+ \quad & \sum_k \frac{\beta_k \widetilde{\mu}_\ell \widetilde{\nu}_{\ell k}}{(1+\widetilde{\mu}_\ell)N_k^*}\left[\frac{r_\beta I_k^a + I_k^t}{1+\widetilde{\mu}_k} + I_k^{nt}\right]\Delta t
\end{aligned}
$$

BLACK DEATH IN1347: A CONTINUOUS DIFFUSION PROCESS

SARS EPIDEMICS: A DISCRETE NETWORK DRIVEN PROCESS



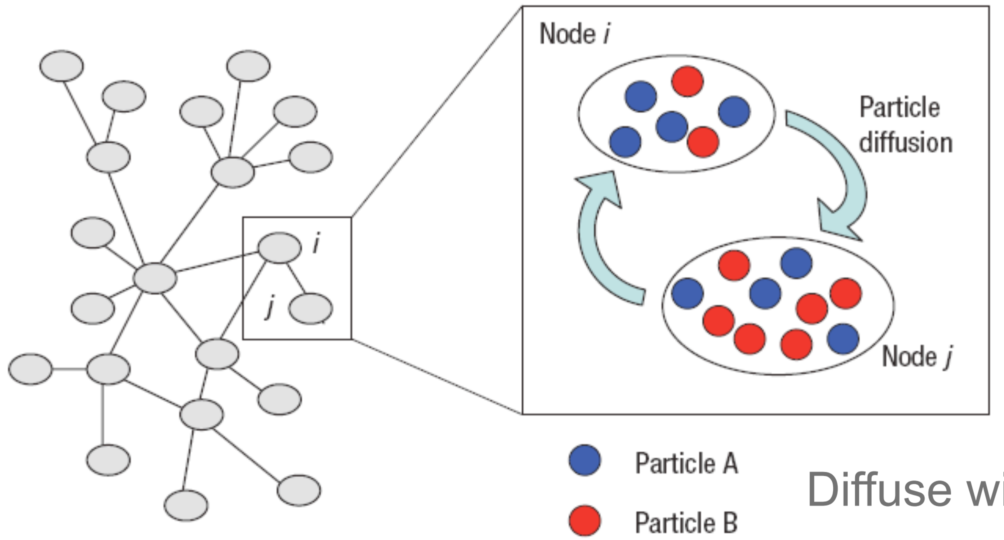


NETWORKS AND NETWORK THINKING

Complex networks approach

•Understanding of non-linear/tipping points due to the complexity of the network system

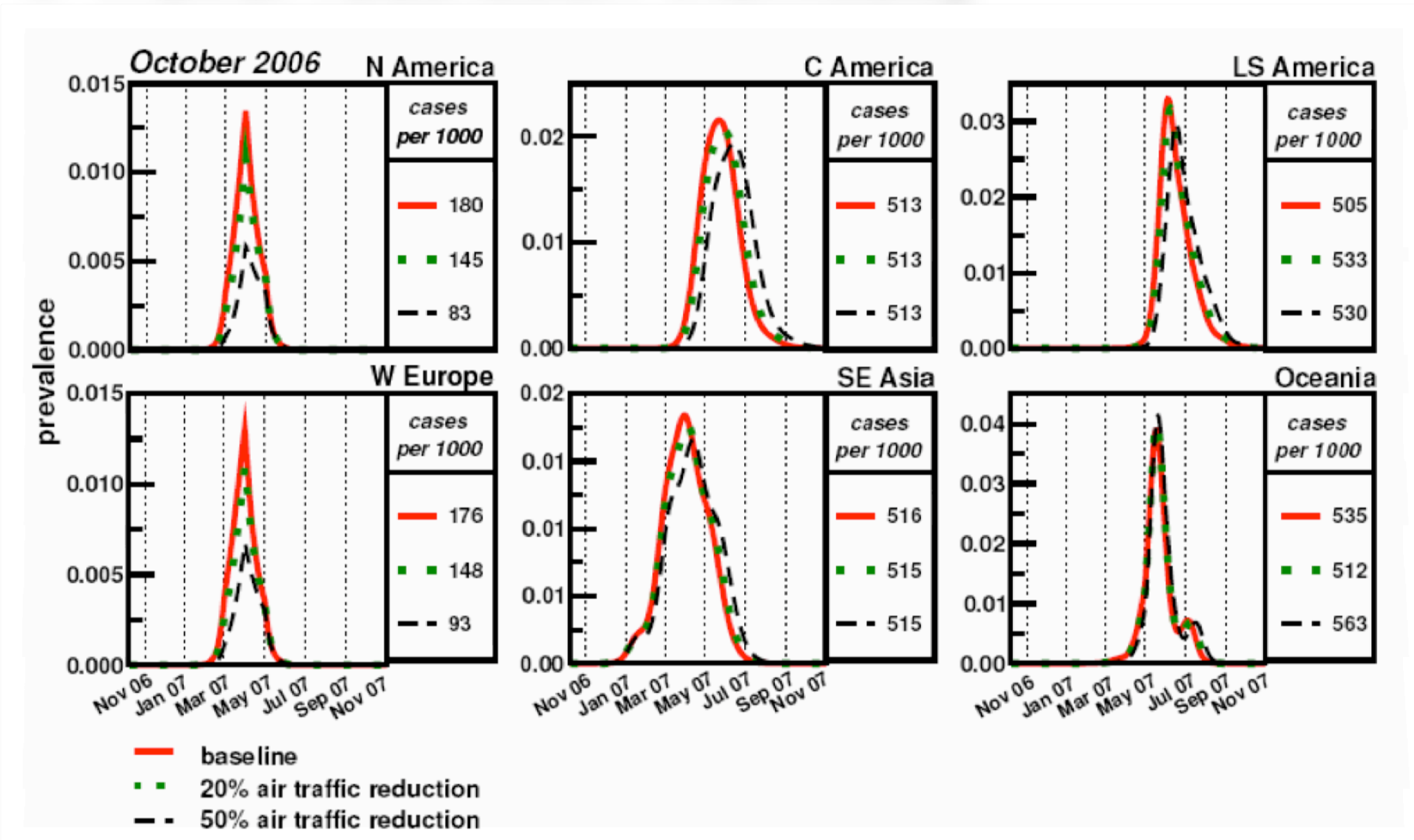
- •Global Threshold and Mobility reduction,
- •Epidemic pathways,
- •mitigation-vs-containment



Diffuse with rate p

PROTOTYPICAL RESULT

- Travel restriction inefficacy



THEORETICAL FINDING….VS REAL WORLD DATA

$$p_c \bar{N} \geq \frac{\langle k \rangle^2}{\langle k^2 \rangle} \frac{\mu R_0^2}{2(R_0 - 1)^2}$$

	Airport network
Average population *N*	**≈ 10^5**
$<k>^2/<k^2>$	**≈ 10^{-1} -10^{-2}**
μ	**1/3**
R_0	**1.1**
p_c	**≈10^{-5}-10^{-6}**
p_{real}	**≈10^{-3} -10^{-4}**

Tenfold traffic reduction is needed to achieve containment effects

Colizza & Vespignani,
Journal of Theoretical Biology2008
Physical Review Letters 2007

What's in GLEaM...

- Refined census data (up to 2.5 arc/min resolution)

- IATA/OAG airline database

- Commuting data for about ~35 countries in 5 continents (definition of a world-wide commuting networks)

- Disease structure and population heterogeneity.

- Multiscale force of infection

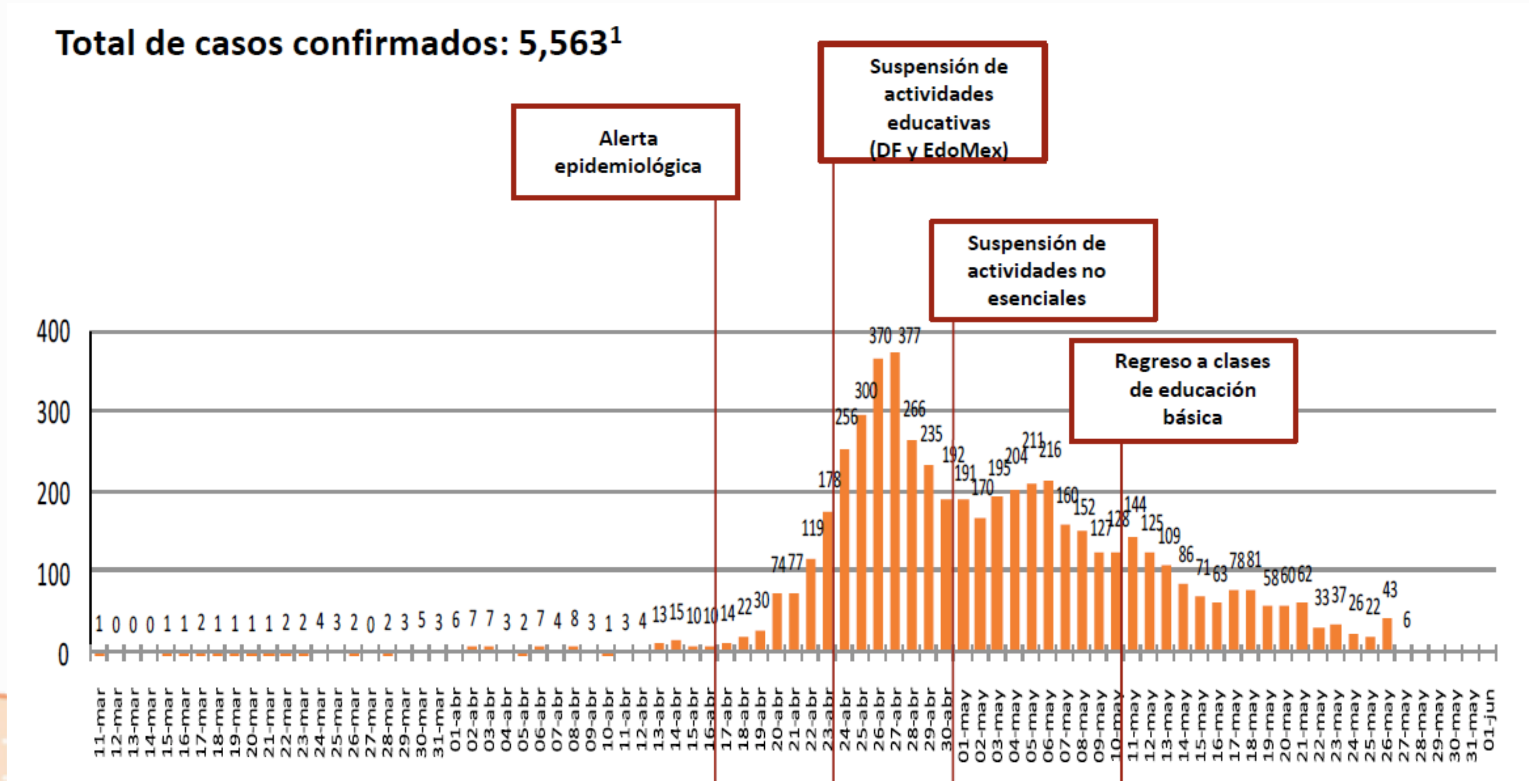
- Intervention scenarios (vaccination, contact tracing, secondary infections, quarantine..)

H1N1 epidemic (Feb. 2009, Mexico):

Can we do projections in real time ?

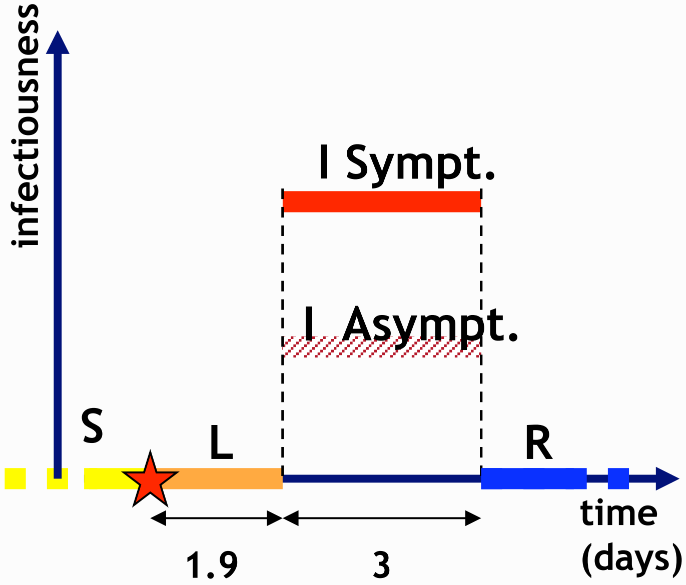
What we need to do projections?

Are projections reliable?

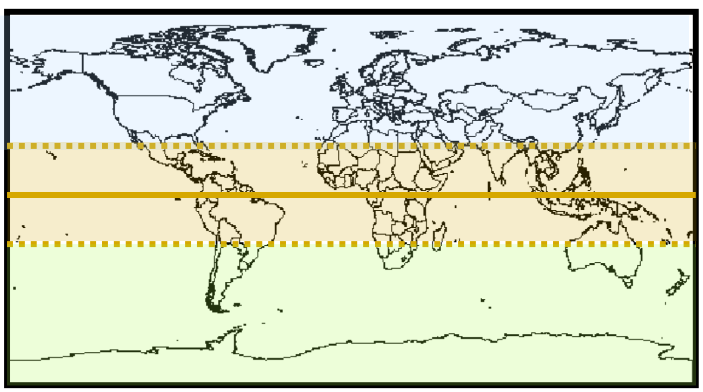


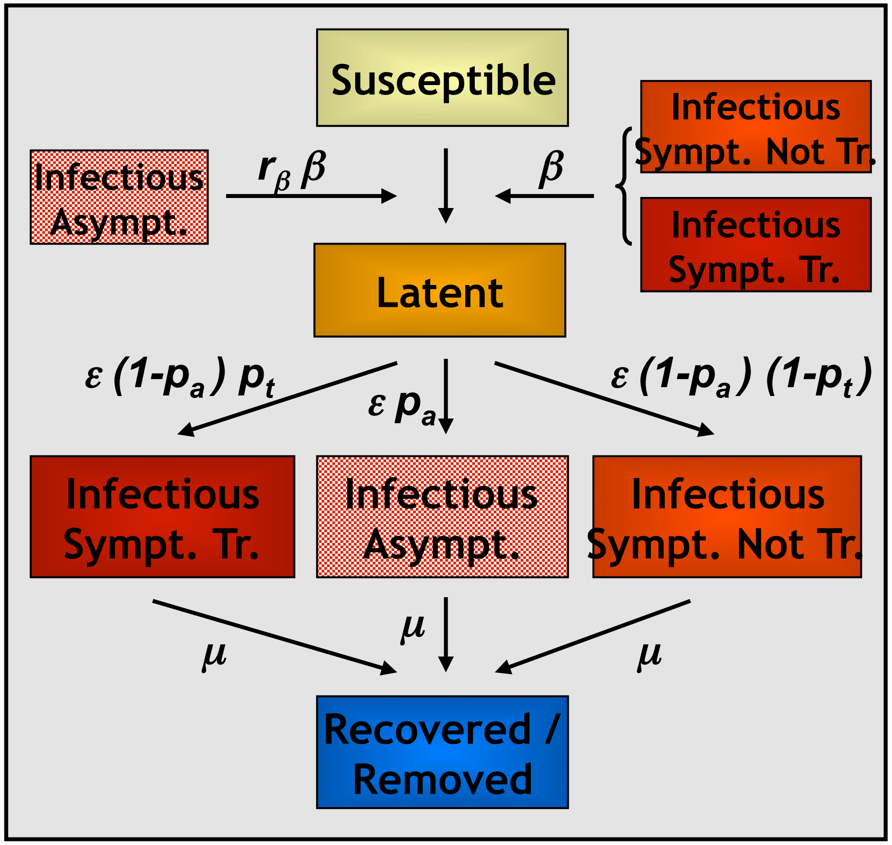
Influenza Transmission

Scenarios and policy making



Seasonality



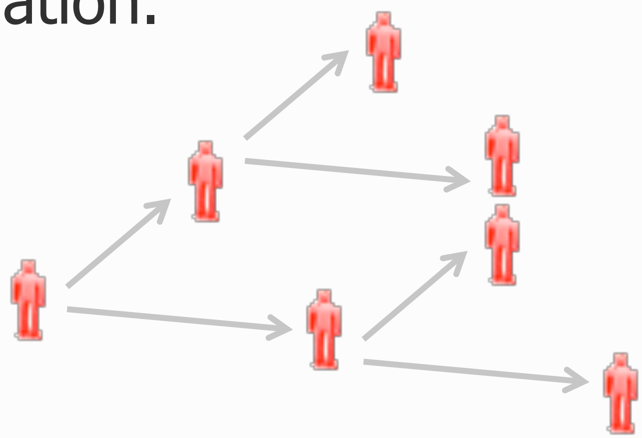


Longini *et al. Am. J. Epid.* (2004)

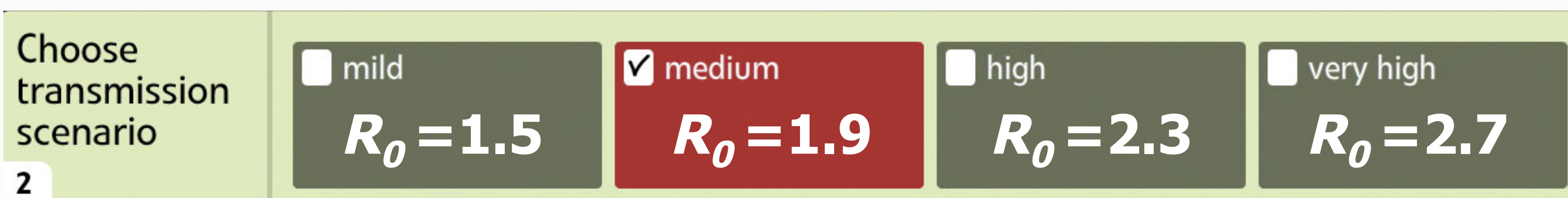
Epidemic evolution

reproductive number *R_0* : average # of infectious individuals generated by one infected in a fully susceptible population.

e.g. R_0=2



pandemic influenza: *R_0 unknown*



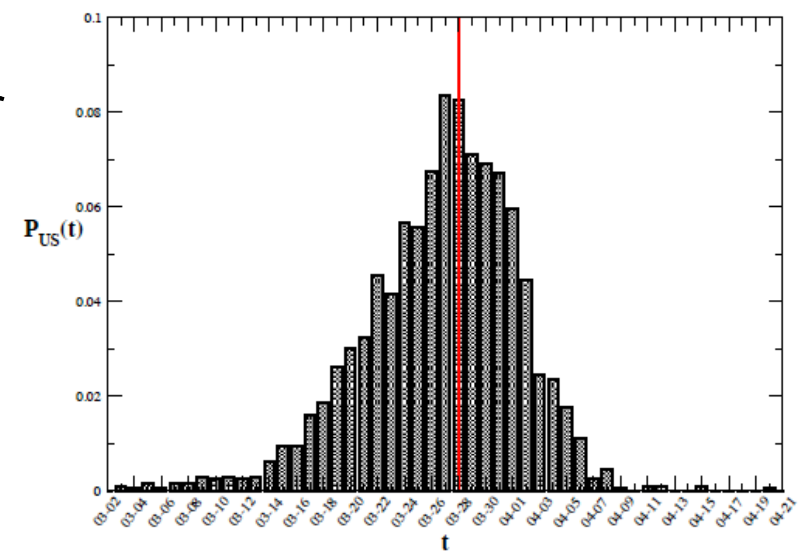
Monte Carlo Likelihood analysis (BMC Medicine 7,45 (2009))

•Probability of arrival in each country depend depend on the disease parameter R_0

•For each set of parameter we can use a Monte Carlo estimate of the arrival time distribution.

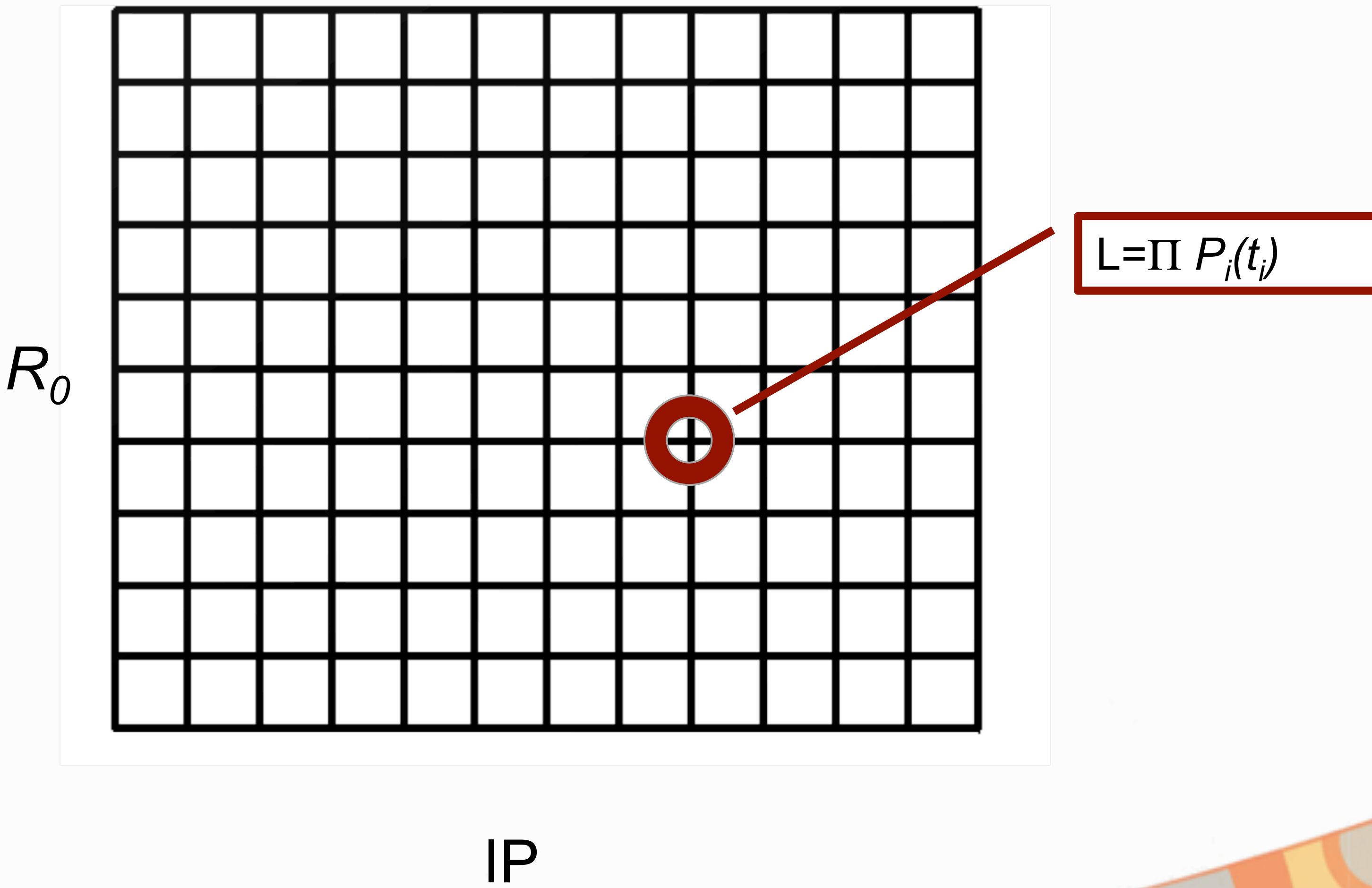
•We can therefore construct the likelihood function and then maximize it to find the best estimate of R_0

•Just one problem:
We need to simulate 10^6 epidemics for a total of 50 Terabyte of data and 2Million minutes of state-of-the-art CPU time.

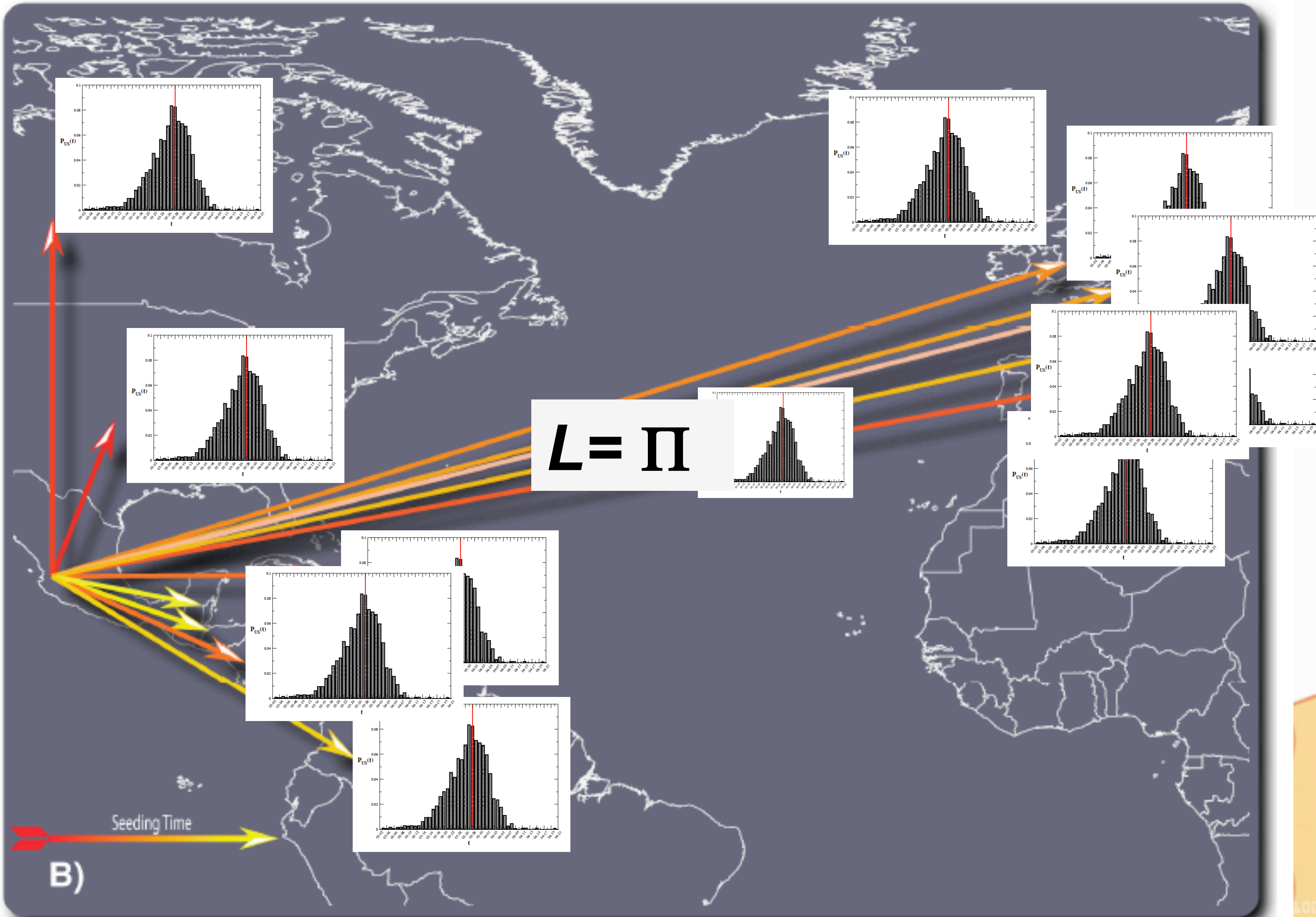


Country	Onset of symptoms	Flight arrival	Confirmed on
United States	March 28 [20]	–	April 21 [20]
Canada	April 11 [21]	April 8 [22]	April 23 [23]
El Salvador	–	April 19 [24]	May 3 [25]
United Kingdom	April 24 [26]	April 21 [27]	April 27 [23]
Spain	April 25 [28]	April 22 [29]	April 27 [23]
Cuba	–	April 25 [30]	May 13 [23]
Costa Rica	April 25 [31]	April 25 [31]	May 2 [23]
Netherlands	–	April 27 [32]	April 30 [32]
Germany	April 28 [33]	–	April 29 [23]
France	–	–	May 1 [34]
Guatemala	May 1 [35]	–	May 5 [36]
Colombia	–	–	May 3 [37]

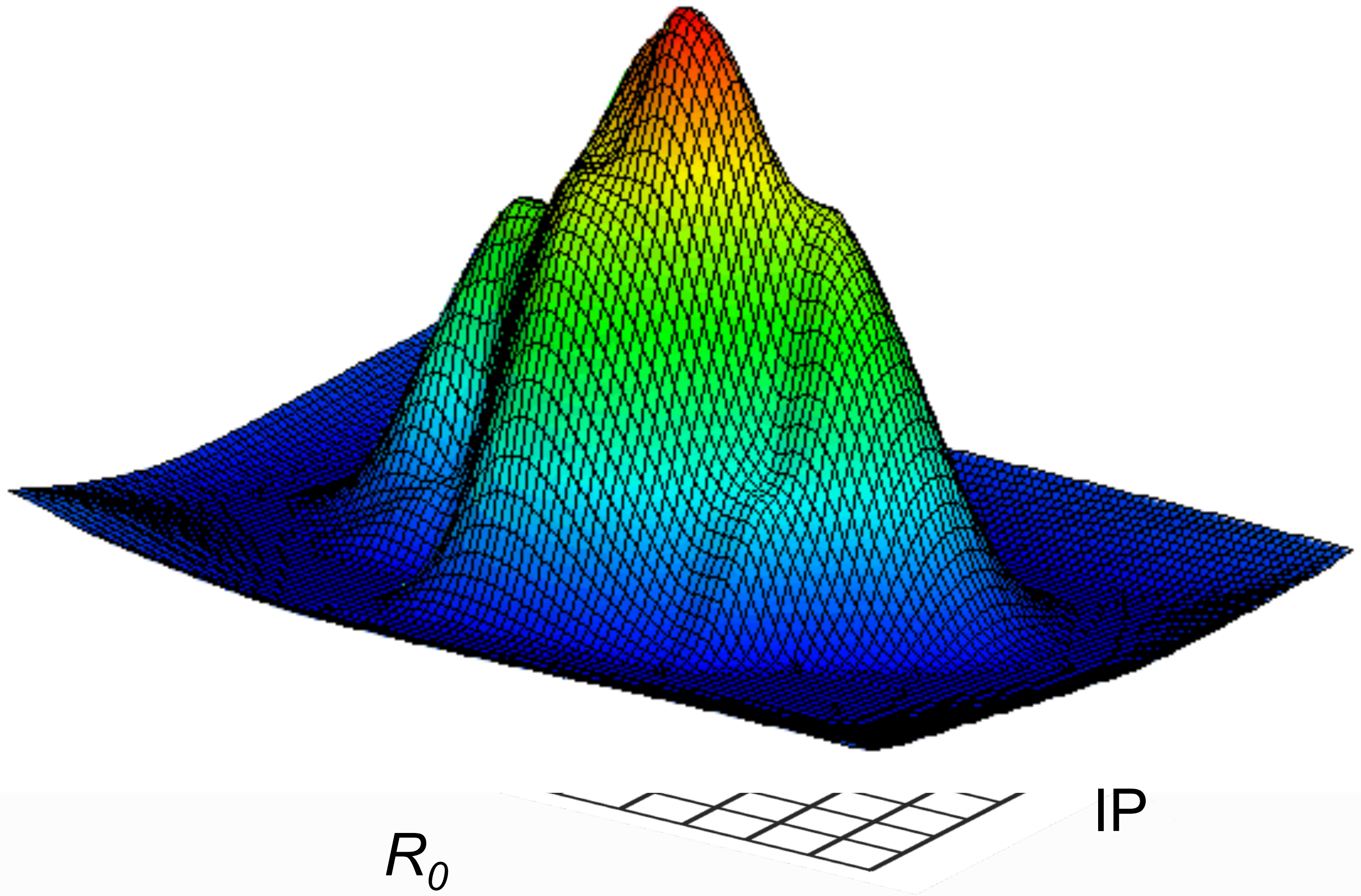
Monte Carlo likelihood sampling



Parameter set likelihood



Monte Carlo likelihood sampling



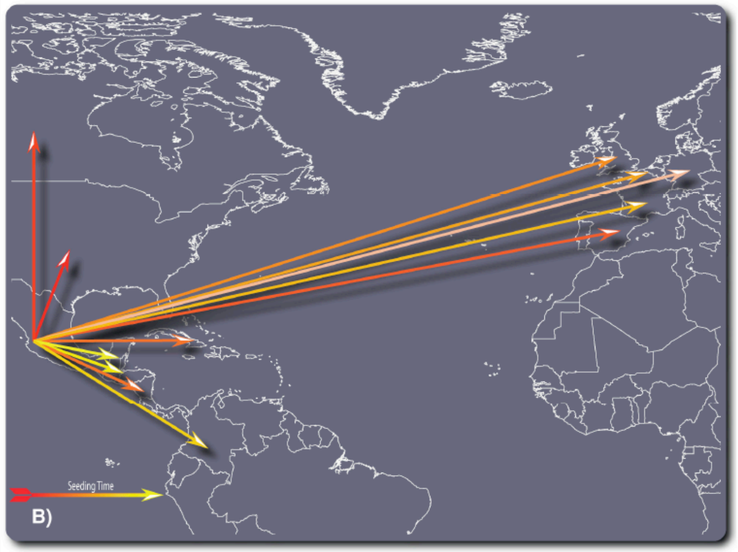
Parameter estimates

Parameter	Best Estimate	Confidence Interval	Description
R_0			number
G_t			ne (days)
μ^{-1}			riod (days)
α_{min}			ısonality

As at			on
ε^{-1}			od (days)
α_{max}			seasonality rescaling

Month	$R(t)$ in Northern hemisphere
May	[1.19-1.49]
June	[1.07-1.33]
July	[1.05-1.24]
August	[1.07 - 1.33]
September	[1.19 - 1.49]

Anticipate the spreading pattern



•Repeated stochastic realizations ~ 10^5

•For each subpopulation pair we define p_{ij} as the probability of infection transmission from *I* to *j*.

•Define a distance

•The minimum spanning tree is then calculated using Chu-Liu-Edmunds Algorithm.

A test and ....what happened in Mexico

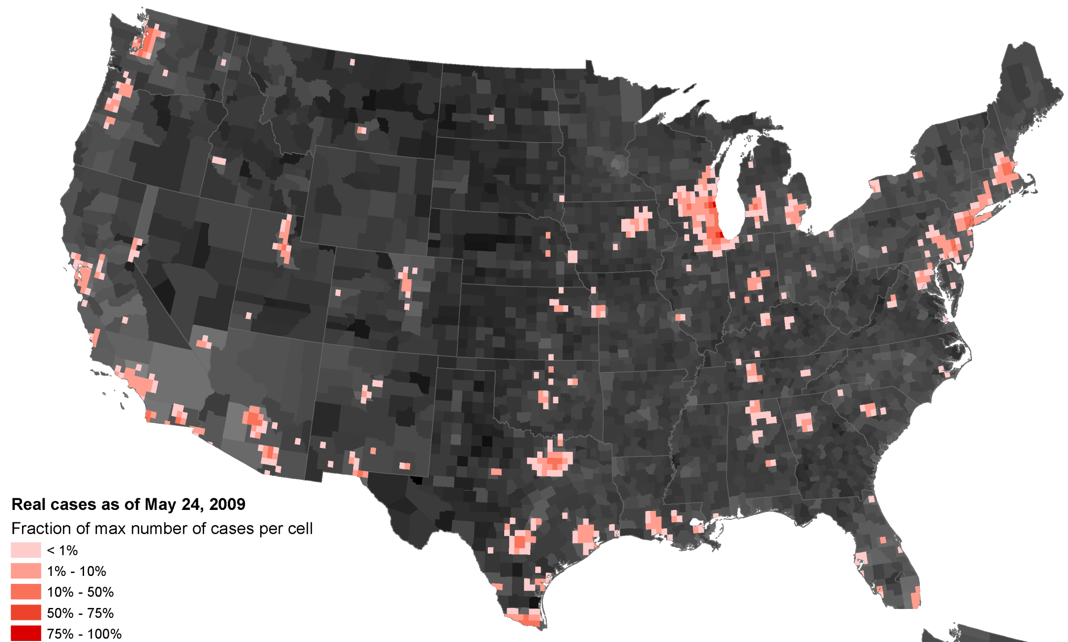
Number imported cases (May 8th)	USA	UK	France	Germany	Brazil
Simulation Results	0 - 534	0 - 44	0 - 62	0 - 55	0 - 45
Surveillance data	85 (170)	17	11	9	3

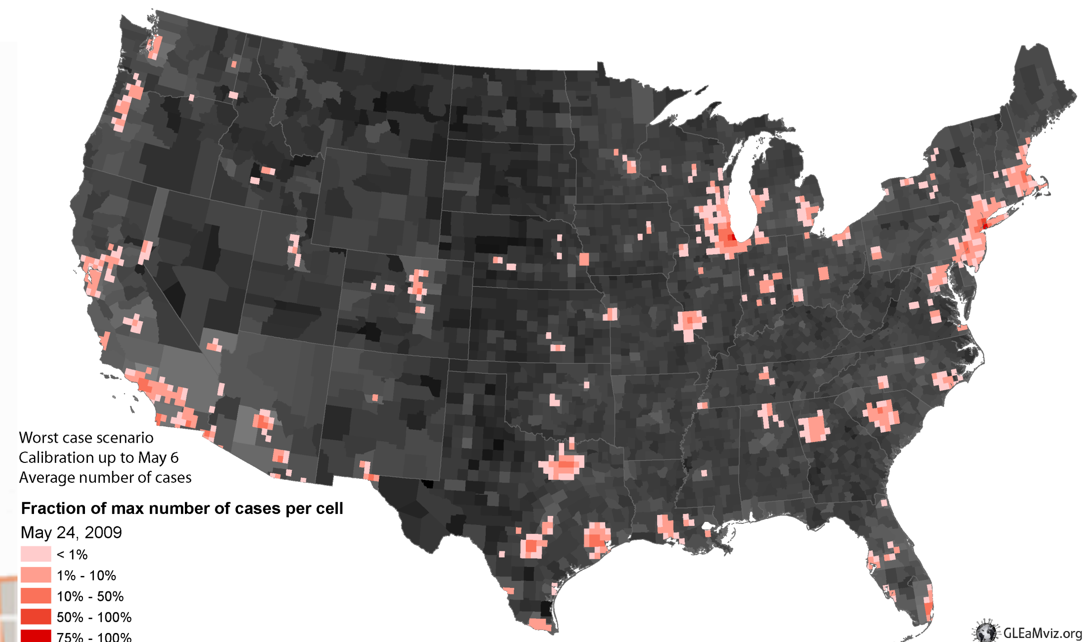
Table 1. - Cumulative number of imported cases from Mexico shown as the 99% reference range over 2,000 realizations on May 8 for a few countries. The simulations are obtained with the best estimate

	Number of symptomatic cases in Mexico (Apr. the 30th)
Simulation Results	[121,000 - 1,394,000]
Lower bound range of Ref. [6]	113,000-375,000
Estimate of Ref. [5]*	2,000 – 280,000
Mexican official report [7] (confirmed cases)	3,350

Table 2. Predictions of GLEaM for the size of the epidemic in Mexico on April 30 in thousands of cases and comparison with other approaches and with empirical data. The simulations are obtained with

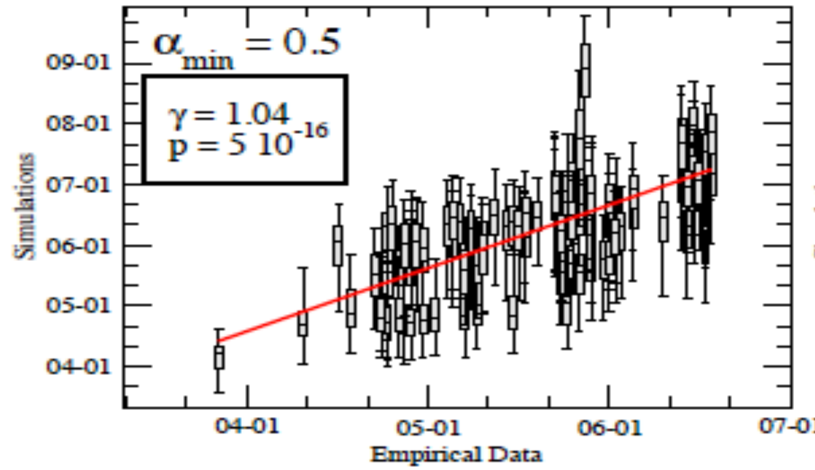
Anticipate where and when





Where and when continued…..

Estimation of the seasonality rescaling against the arrival time in 96 countries.



Region	Estimated activity peak time
North America	[Sept 25 - Nov 09]
Western Europe	[Oct 14 - Nov 21]
Lower South America	[Jul 30 - Sep 06]
South Pacific	[Jul 28 - Sep 17]

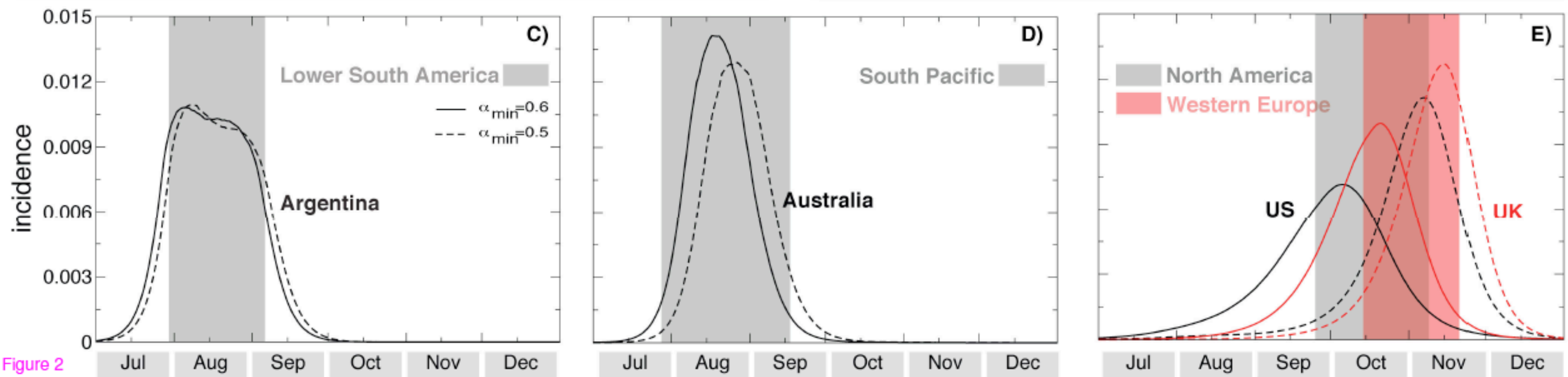
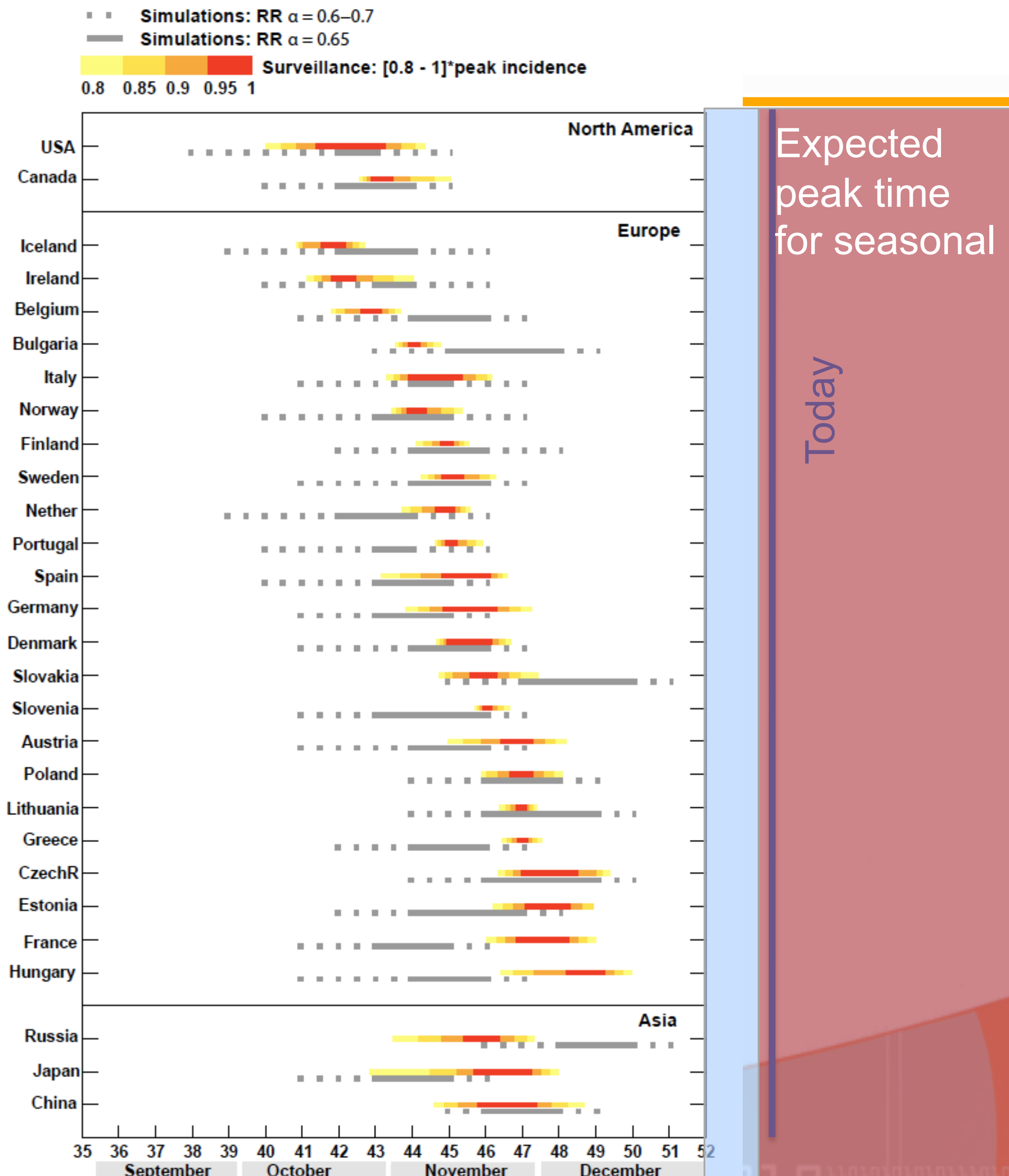
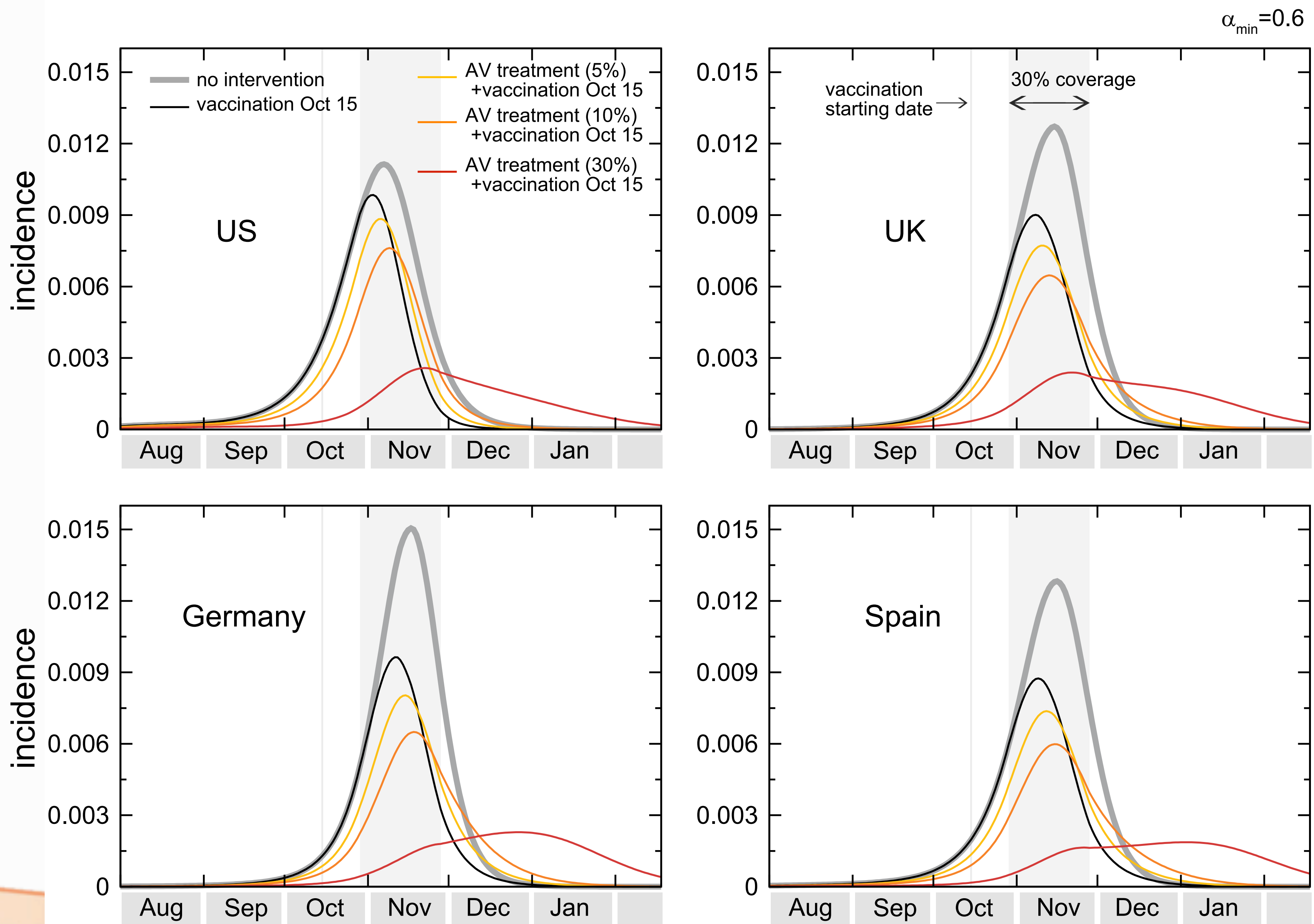


Figure 2

When…

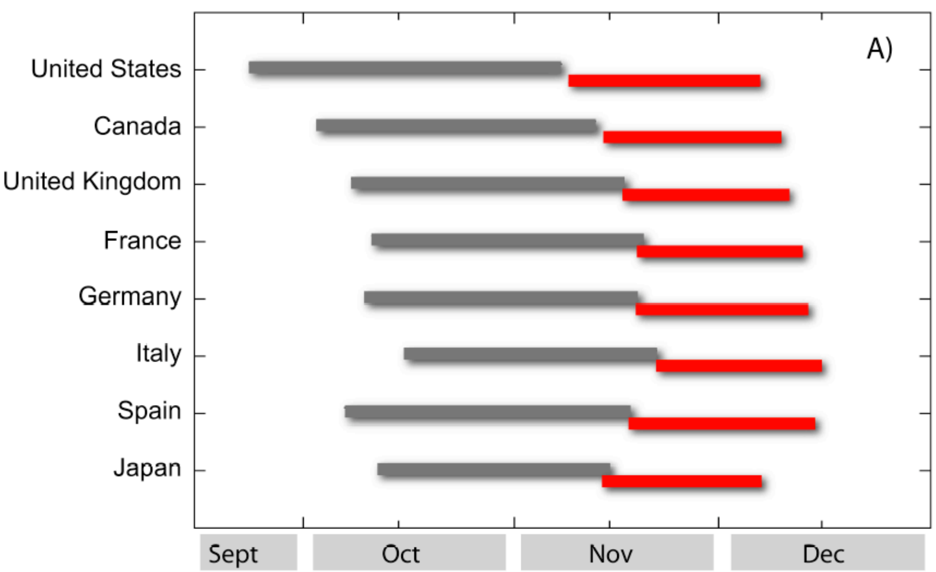

Simulations: RR α = 0.6–0.7
Simulations: RR α = 0.65
Surveillance: [0.8 - 1]*peak incidence
0.8 0.85 0.9 0.95 1
North America
USA
Canada
Europe
Iceland
Ireland
Belgium
Bulgaria
Italy
Norway
Finland
Sweden
Nether
Portugal
Spain
Germany
Denmark
Slovakia
Slovenia
Austria
Poland
Lithuania
Greece
CzechR
Estonia
France
Hungary
Asia
Russia
Japan
China
35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
September
October
November
December
Expected peak time for seasonal
Today

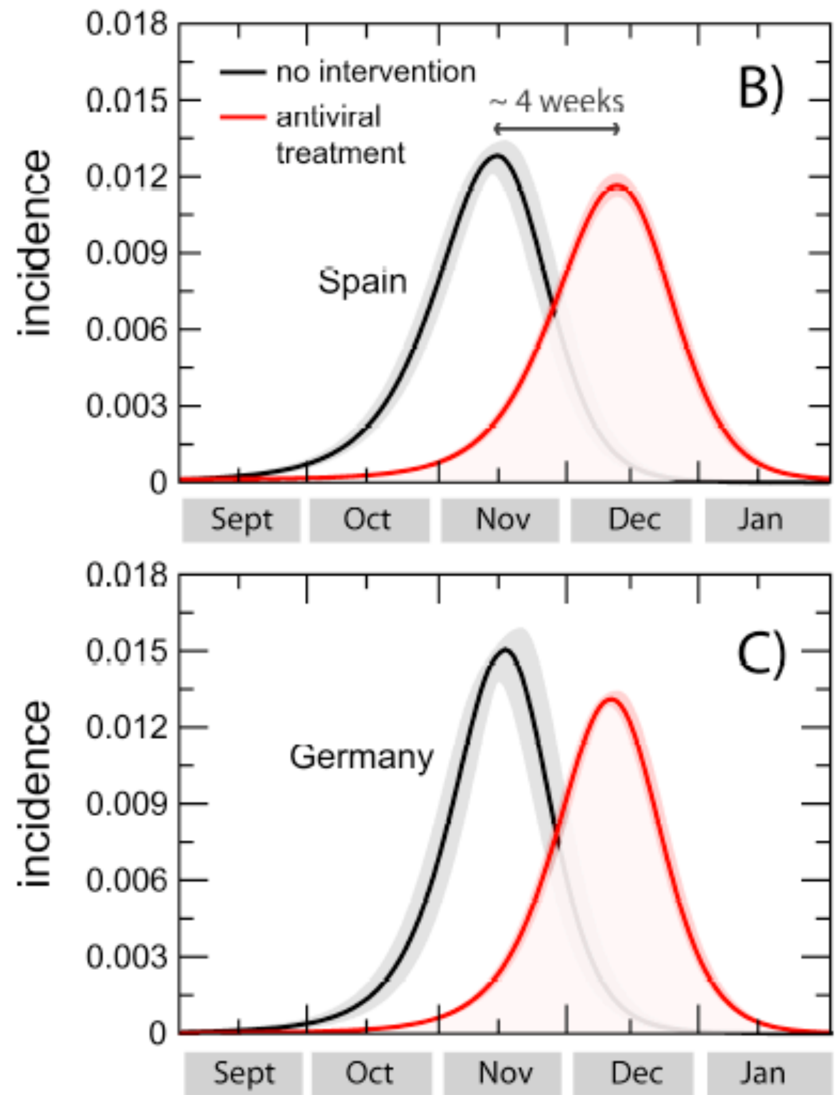
Vaccination campaigns…..



Where and when continued…..

Delay with the use of systematic AV treatments (30%)

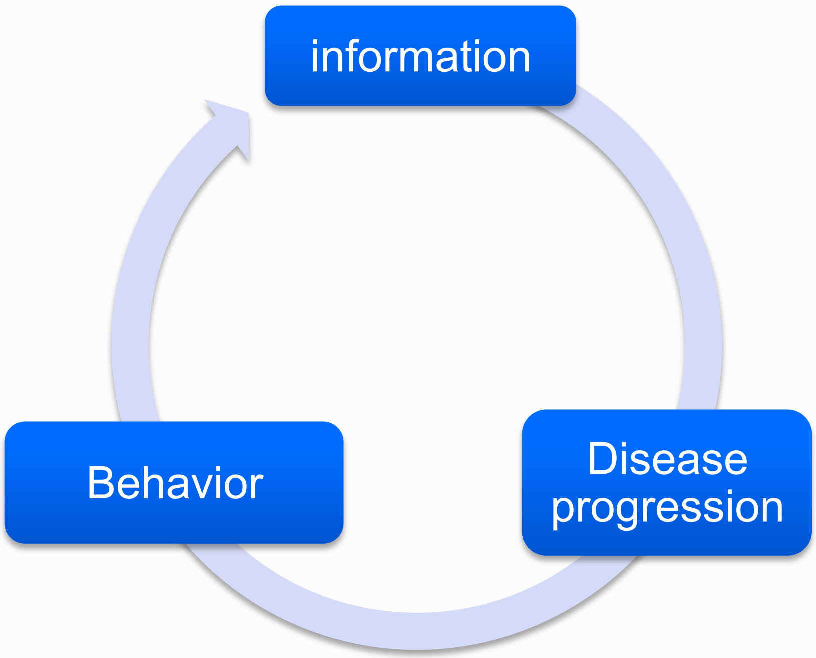




What's next for...

•Undersatnding of the feedback loop
system behavior->prediction -> information+adaptation -> new system behavior...

•Formal models and understanding of social adaptation/reaction during social disruptive events.



Recent papers

- PNAS, 106, 21484 (2009)
- BMC-*Medicine* 7, e45 (2009)
- Emerging Health Threats, **2** e11 (2009)
- Science **325**, 425 (2009)
- Journal of Theoretical Biology 251, 450-467 (2008)
- Plos *Medicine*, **4**, e13 (2007)
- Nature Physics, **3,** 276-282 (2007)
- Physical Review Letters **99**, 148701 (2007).
- PNAS, **103**, 2015 (2006)

More Information/paper/data

Cx-Nets

http://cxnets.googlepages.com