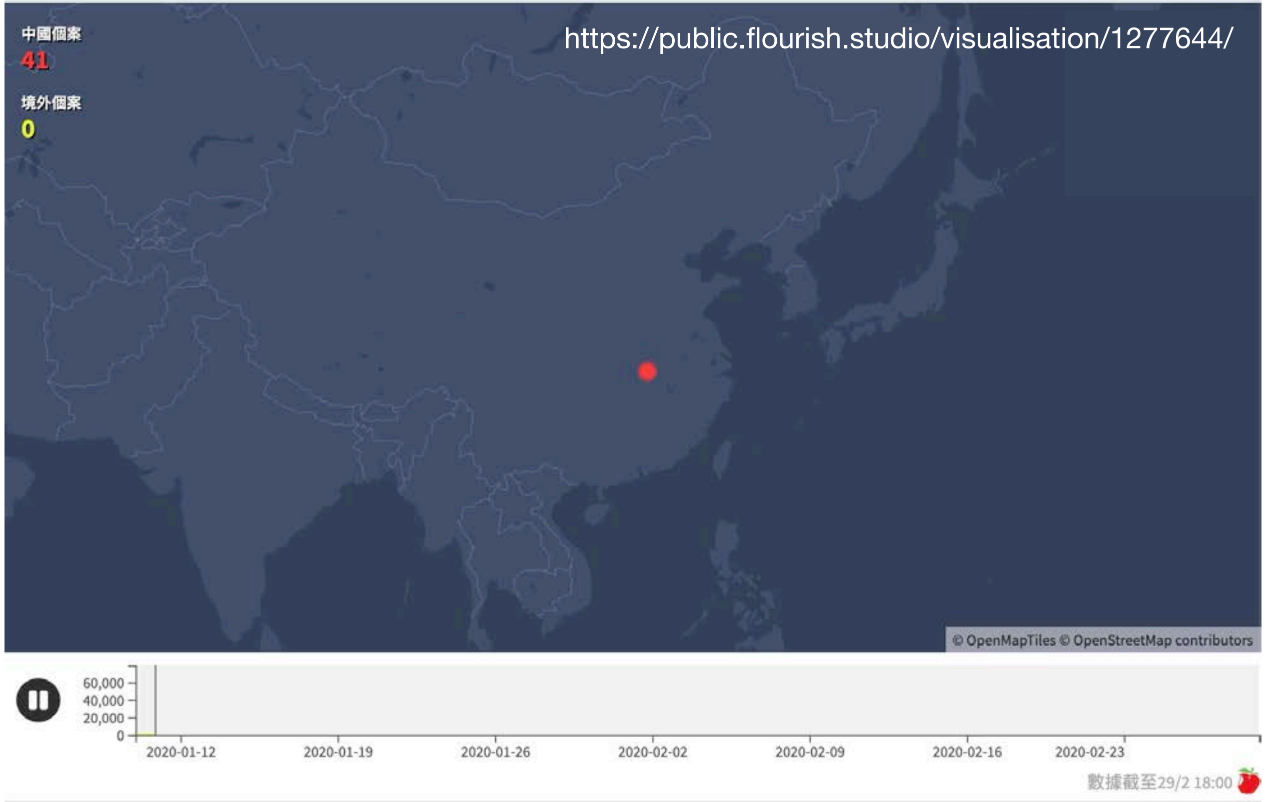
IMAGE edited by Michael Tse based on https://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2020/february/20200218_china_covid19

Modelling the Spread of COVID-19 with Travel Data and Incomplete Infection Data

C K Michael Tse
City University of Hong Kong

Collaborating with Choujun Zhan, South China Normal University

COVID-19 Early Facts



- **2019 New Coronavirus Disease (COVID-19)** began to spread from early December 2019 in **Wuhan**
- First confirmed case: December 8, 2019
- Human-to-human transmission found to occur in mid December 2019
- Active control started in late January 2020, with official instruction from central sent down on January 7, 2020.*
- Wuhan was locked down on January 24, 2020. Other cities began to impose stringent travel control.

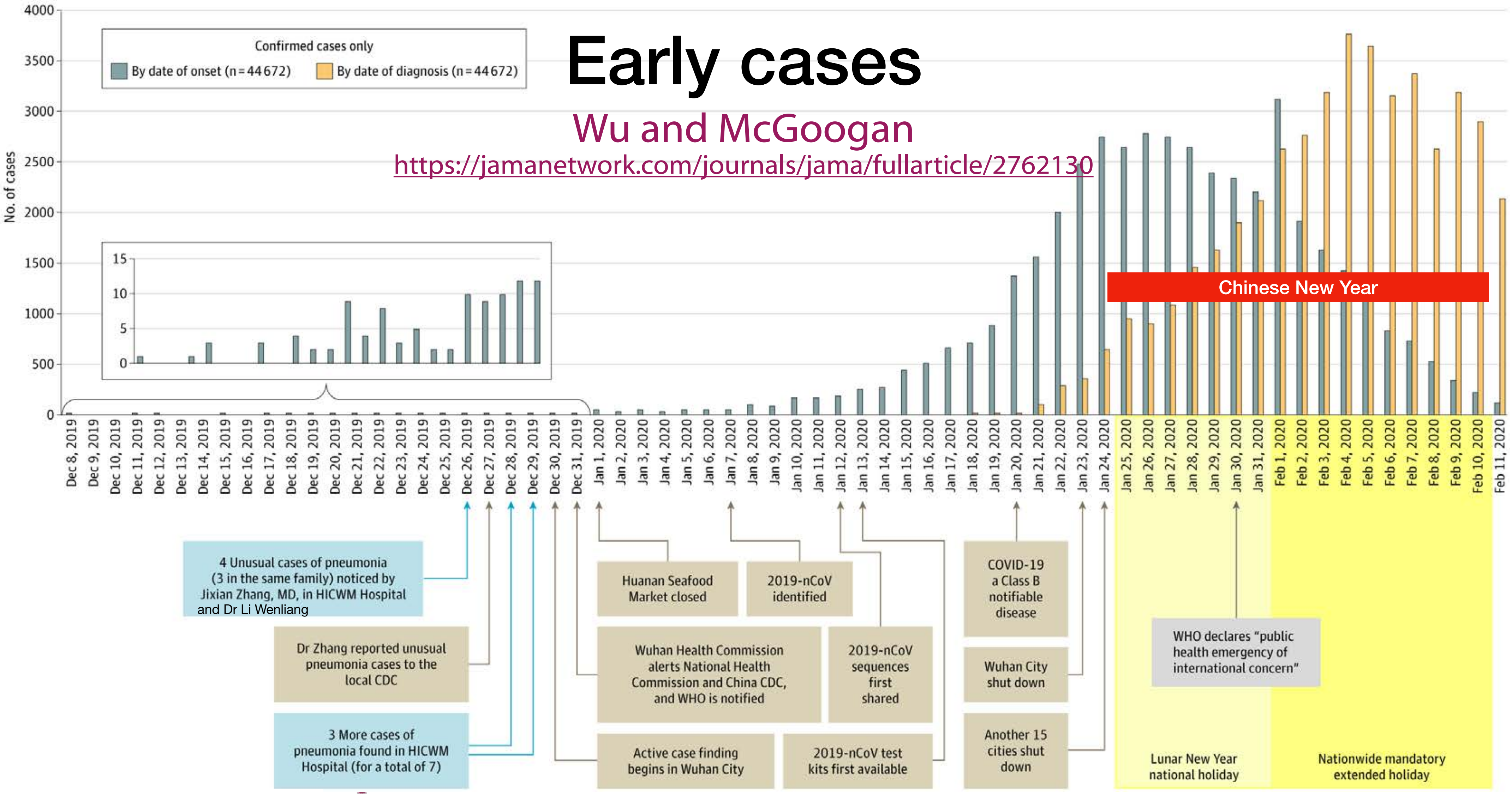
- Italy reported first case on January 31, 2020
- France also traced the first case back in December 2019

*State media Qiu Shi Magazine reported President Xi's speech at the Standing Committee of the Central Political Bureau of CCP on Feb 3, 2020, sending instruction down on Jan 7, 2020 regarding control of the new coronavirus outbreak in Wuhan.

Early cases

Wu and McGoogan

https://jamanetwork.com/journals/jama/fullarticle/2762130



Hong Kong (second wave)

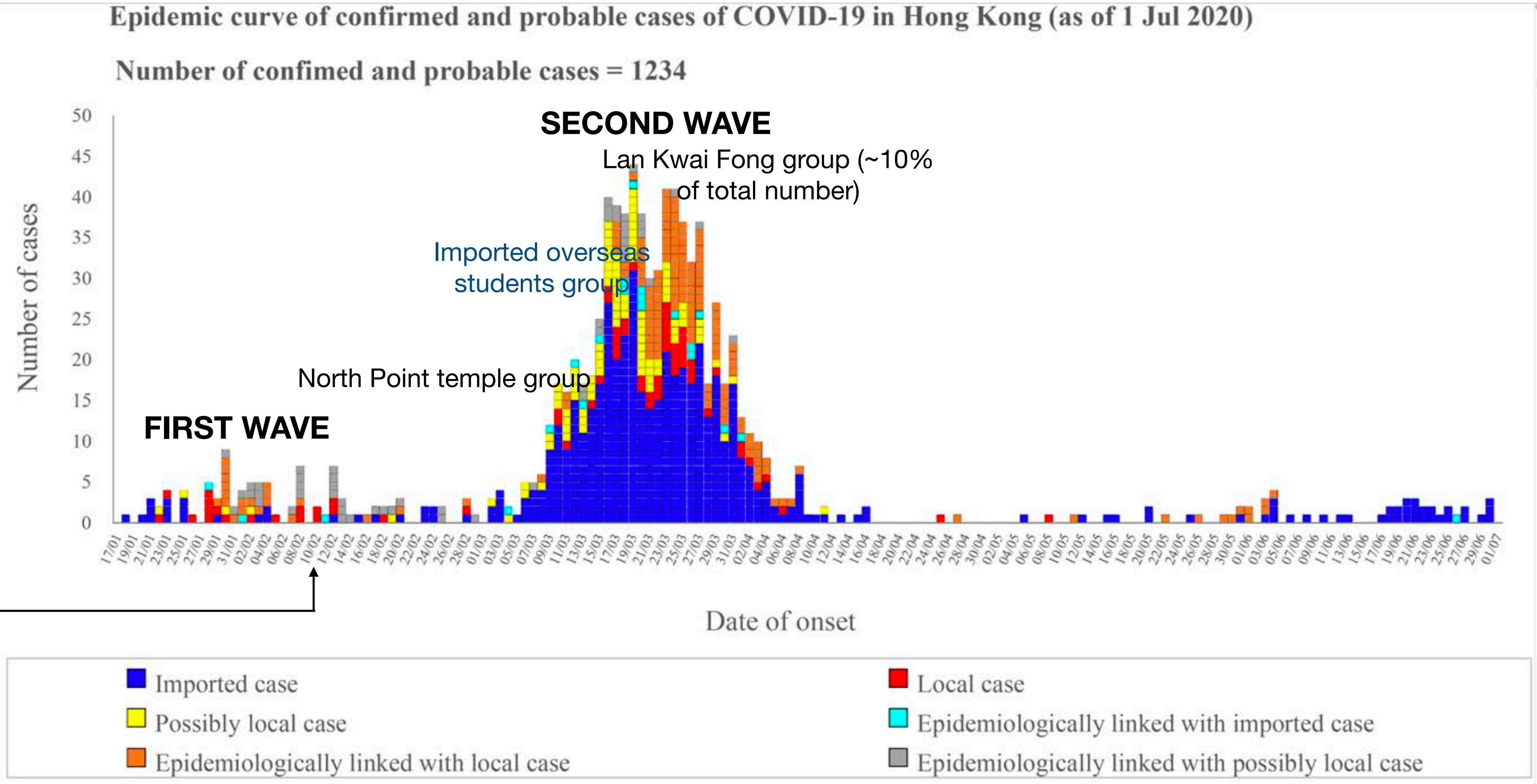
South China Morning Post

Pet dog of Covid-19 patient infected, Hong Kong health officials confirm

- Pomeranian has low-level infection and experts unanimously conclude human-to-animal transmission
- But experts also cite previous experience with S suggesting cats and dogs will not fall ill or trans virus to people

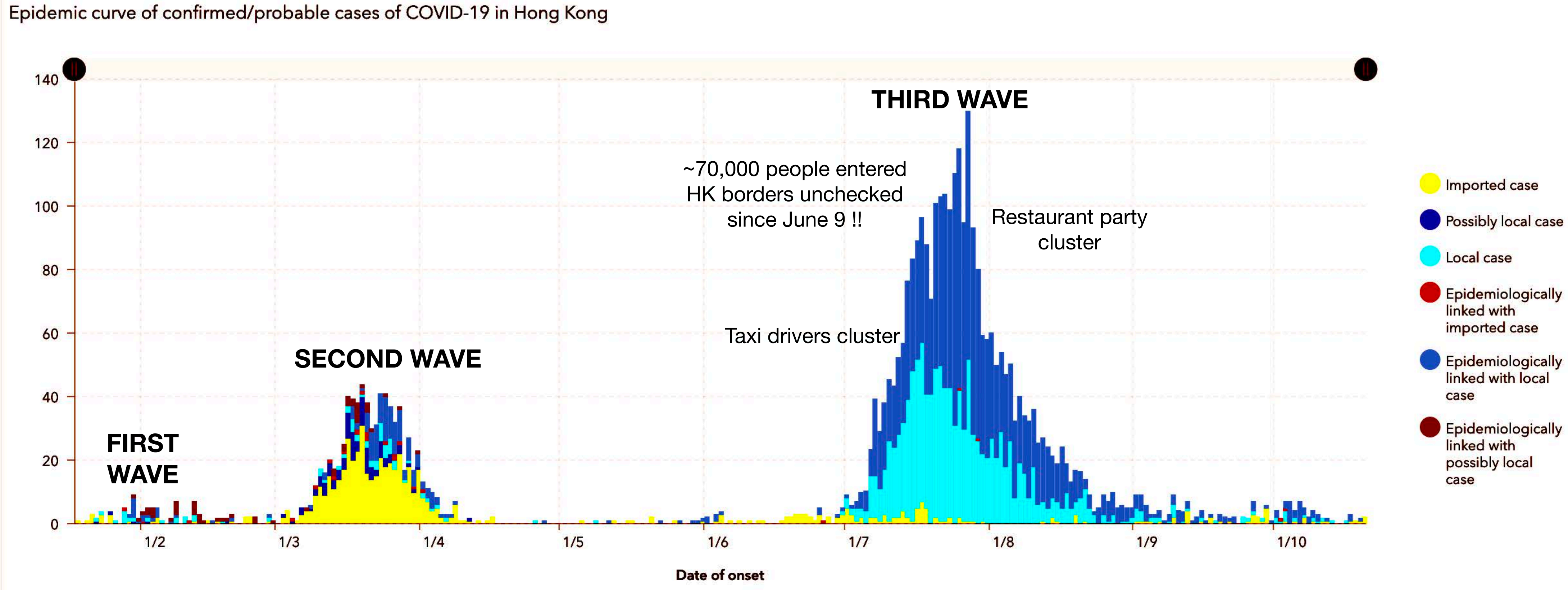
Cancellation of events

25 Eason Chan shows (Dec 2019)
6 Zhou Jie Lun shows (Jan 2020)
23 Andy Lau shows (Jan-Feb 2020)
numerous concerts, theatres, …
  ⇒ ~1 million audience
1 CNY Firework show
1 CNY Parade
  ⇒ Several thousands
… …
had prevented mass gatherings and massive intercity travel during the peak travel time.



Data: Department of Health, Hong Kong

Hong Kong (third wave)



FIRST WAVE

SECOND WAVE

THIRD WAVE

~70,000 people entered HK borders unchecked since June 9 !!

Restaurant party cluster

Taxi drivers cluster

Data: Department of Health, Hong Kong

Spreading escalated by the *spring travel rush*


https://www.storm.mg/article/2212024

- 30 days surrounding the Lunar New Year
- Migrant workers and students travel from major cities to country towns for family reunions, and return to the cities at the end of the holiday period. Holiday goers also travel to and from tourist cities.
- Ministry of Transport estimates ~3 billion trips taken during the spring travel rush.
- Wuhan, being a major transport hub and having a large number of higher education institutions as well as manufacturing plants, is among the cities with the largest outflow and inflow traffic during the spring travel rush.

The epicenter

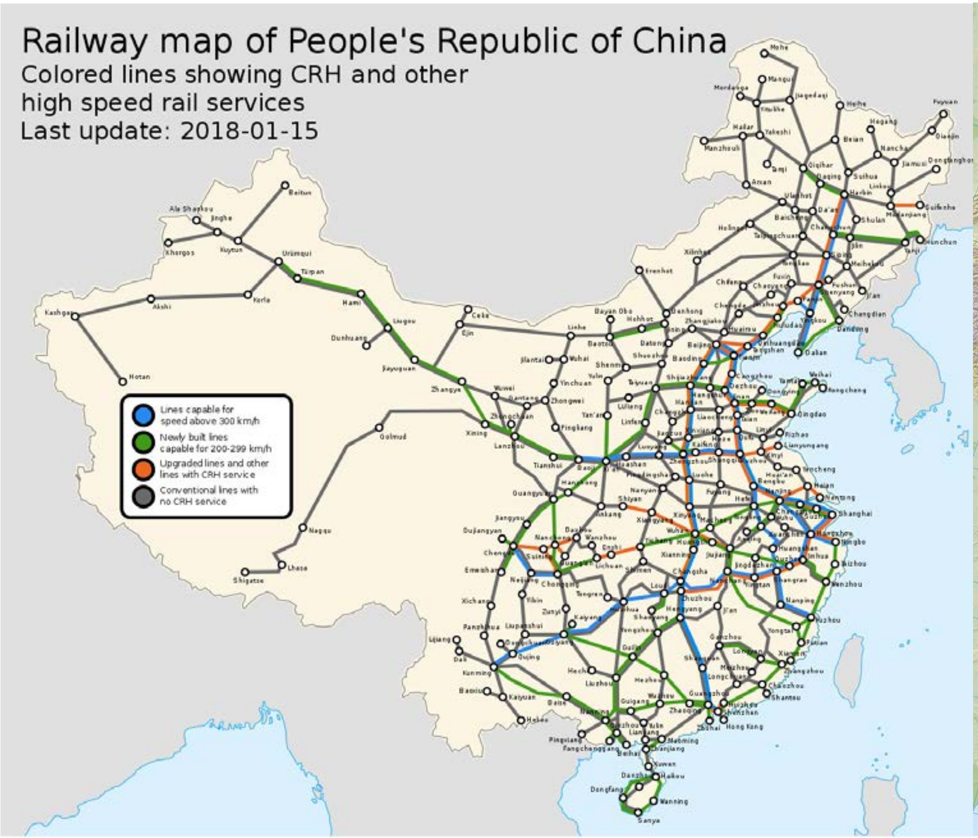
WUHAN

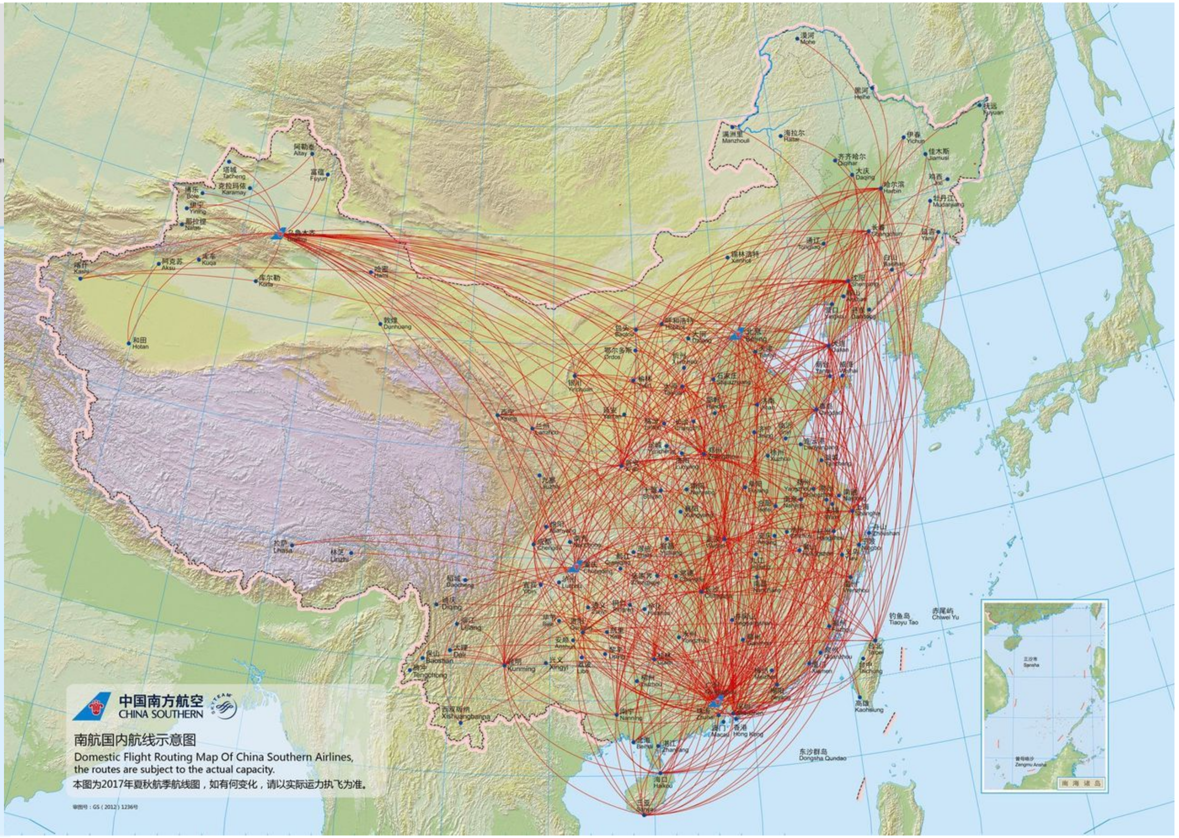
A very important city in China

- 11 million people
- ***major traffic hub***
- center of higher education, manufacturing plants, industry and business


Photo by Michael Tse 2018

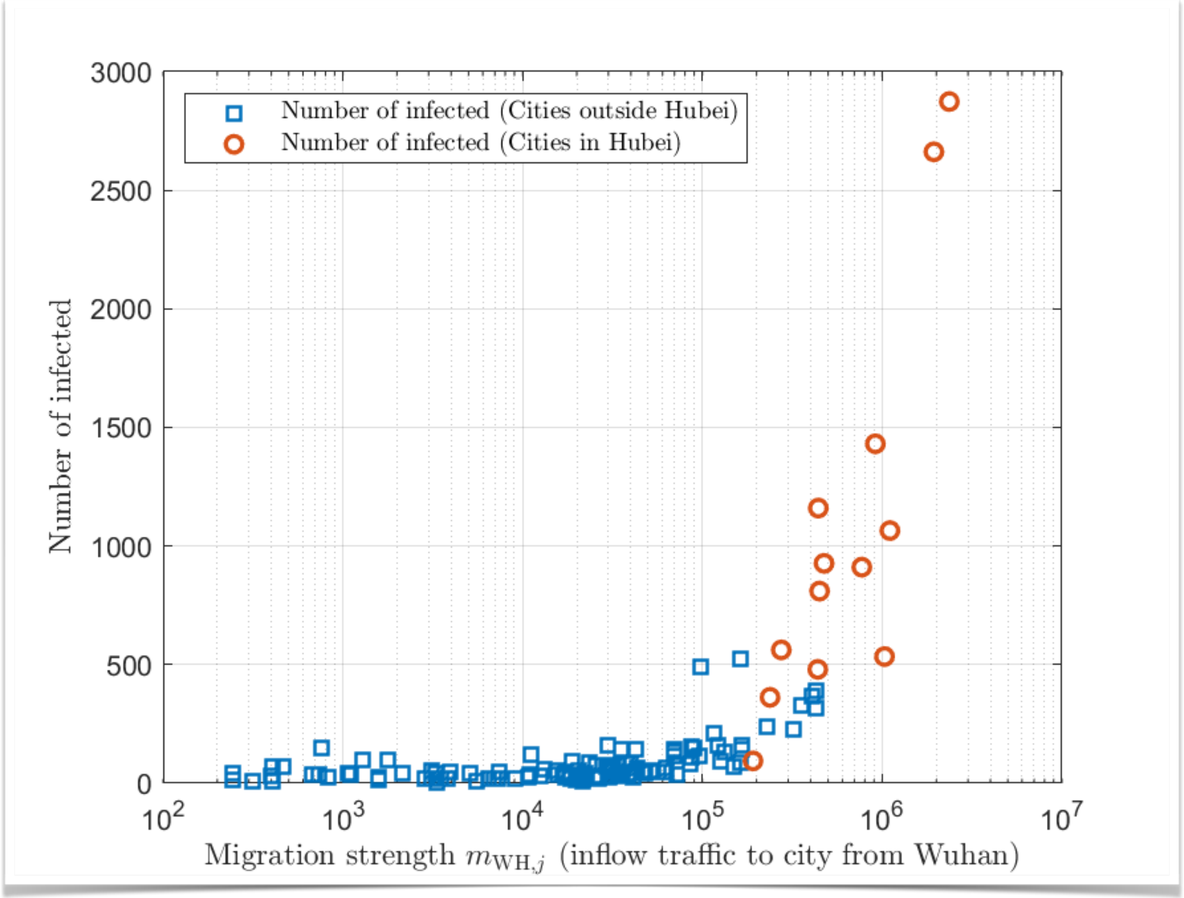
Gathering of people and **intercity travel** of infected and exposed individuals within China have been the main drives that escalated the spreading of the virus. *How to get indicative travel volume data?*





Impact of Intercity Travel

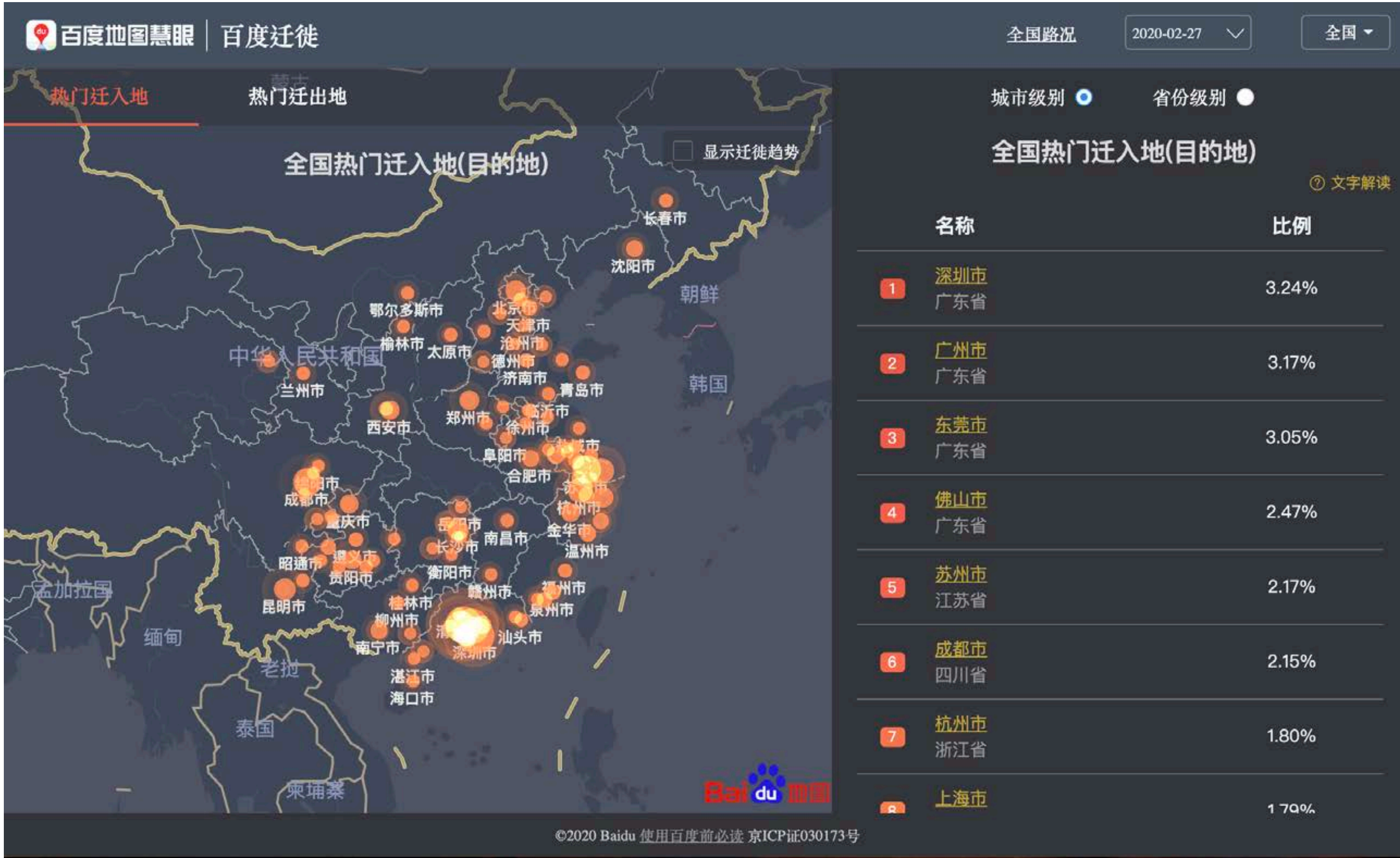
- Number of infections in a city farther from Wuhan increases almost ***linearly*** with the *Baidu Migration Strength* (inflow strength) of the city from Wuhan



Data

Baidu Migration Data

Mobile-app based human migration tracking data system provides <u>indicative</u> inflow and outflow volume of travellers between cities



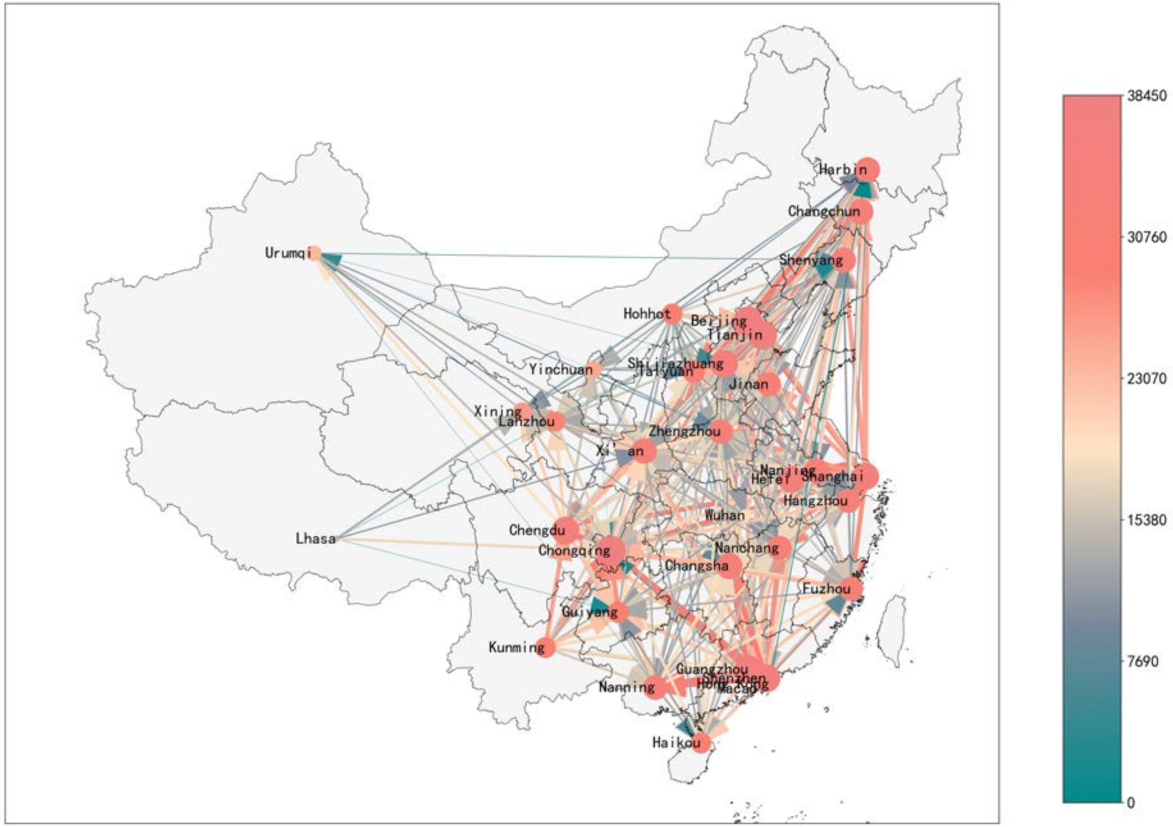
全国热门迁入地(目的地)

	名称	比例
1	深圳市 广东省	3.24%
2	广州市 广东省	3.17%
3	东莞市 广东省	3.05%
4	佛山市 广东省	2.47%
5	苏州市 江苏省	2.17%
6	成都市 四川省	2.15%
7	杭州市 浙江省	1.80%
8	上海市	1.79%

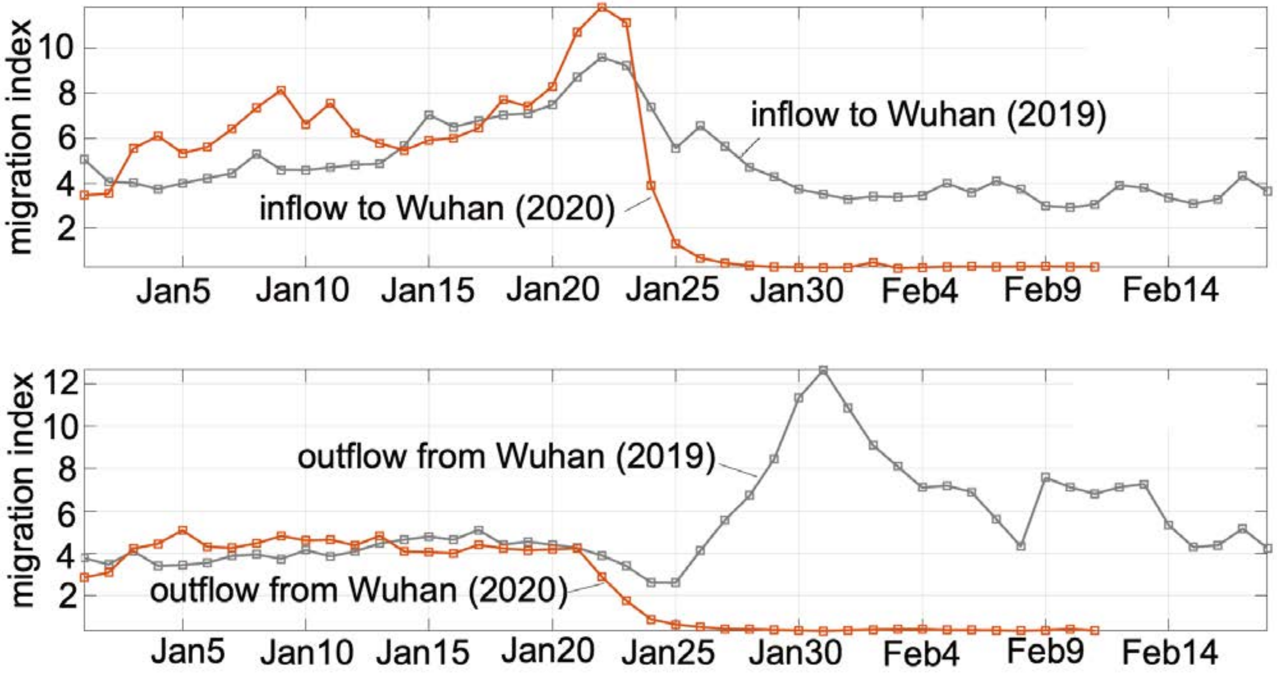
Data collection

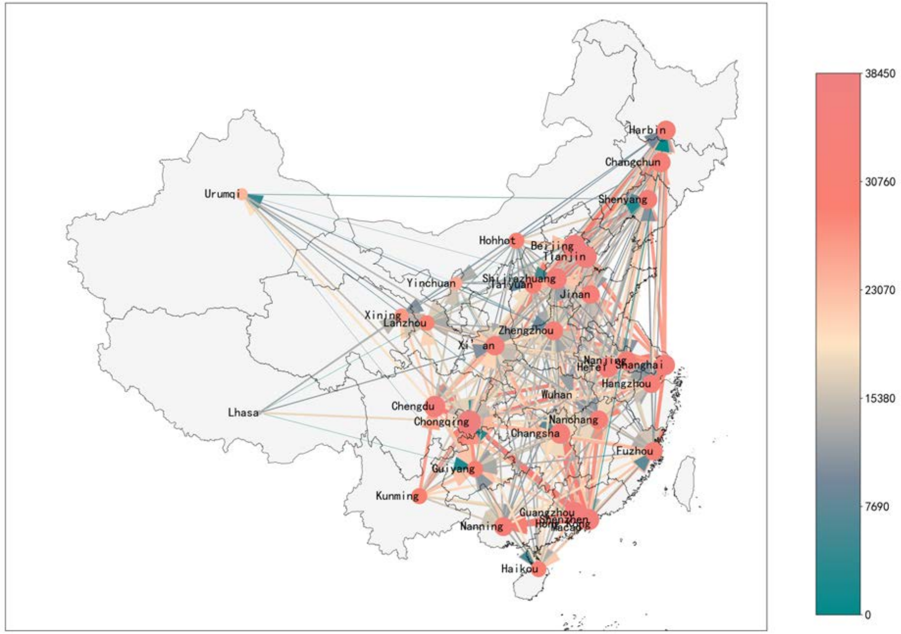
- Baidu Migration data for 367 cities (or administrative regions) in China.

- The data provide the migration strengths of cities which are indicative measures of the human traffic volume moving in and out of 367 cities and administrative regions.



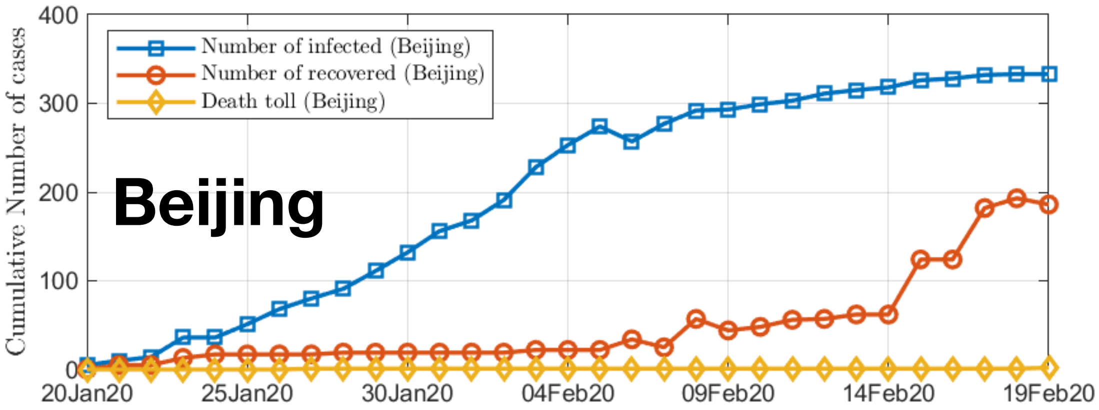
Wuhan inbound and outbound traffic

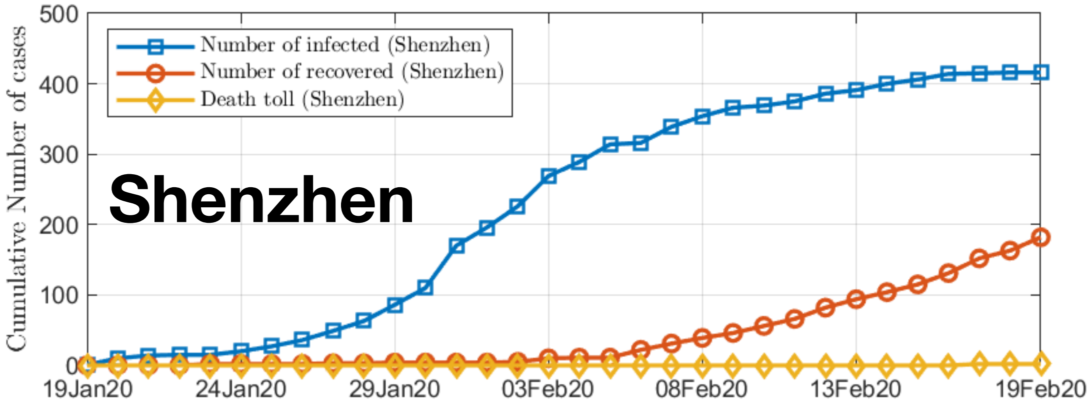


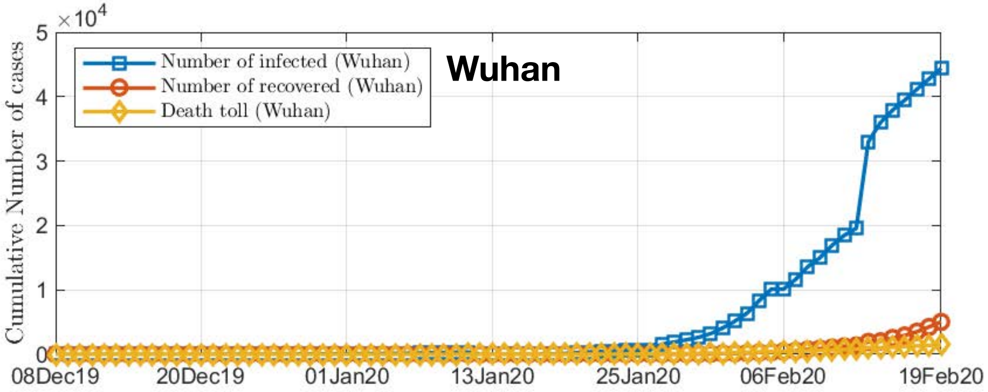


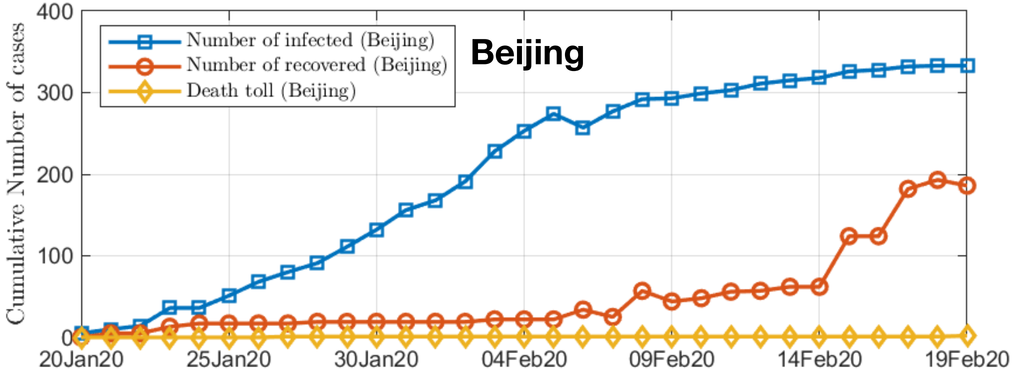
Data of Infection Cases

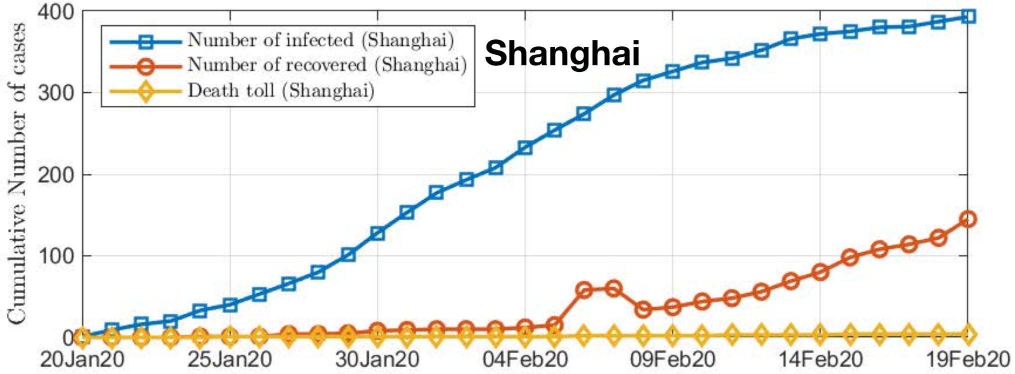
- **National Health Commission of China**
- Daily data, including the daily total number of confirmed cases in each city, daily total cumulative number of confirmed cases in each city, daily cumulative number of recovered cases in each city, and daily cumulative death toll in each city.
- Data collected for **367** cities

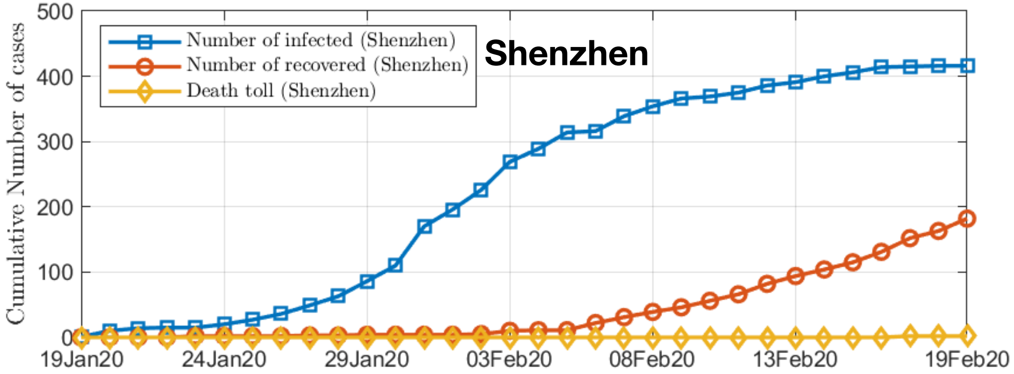


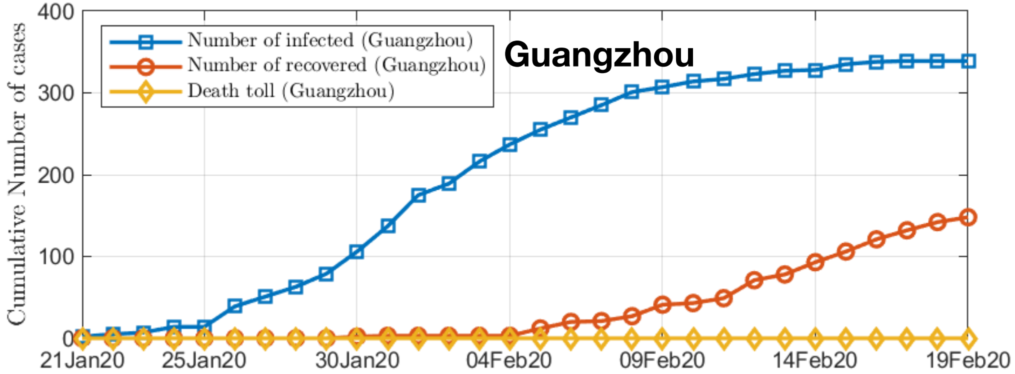


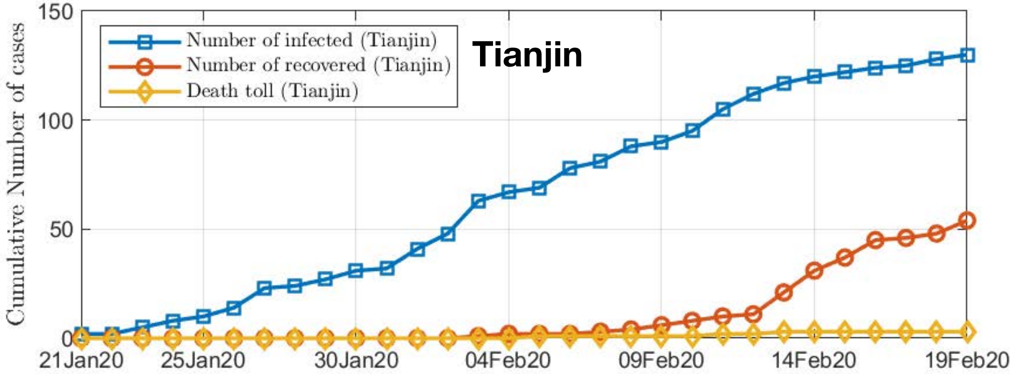

Wuhan
×10^4
Cumulative Number of cases
Number of infected (Wuhan)
Number of recovered (Wuhan)
Death toll (Wuhan)
0
1
2
3
4
5
08Dec19
20Dec19
01Jan20
13Jan20
25Jan20
06Feb20
19Feb20


Beijing
Cumulative Number of cases
Number of infected (Beijing)
Number of recovered (Beijing)
Death toll (Beijing)
0
100
200
300
400
20Jan20
25Jan20
30Jan20
04Feb20
09Feb20
14Feb20
19Feb20


Shanghai
Cumulative Number of cases
Number of infected (Shanghai)
Number of recovered (Shanghai)
Death toll (Shanghai)
0
100
200
300
400
20Jan20
25Jan20
30Jan20
04Feb20
09Feb20
14Feb20
19Feb20


Shenzhen
Cumulative Number of cases
Number of infected (Shenzhen)
Number of recovered (Shenzhen)
Death toll (Shenzhen)
0
100
200
300
400
500
19Jan20
24Jan20
29Jan20
03Feb20
08Feb20
13Feb20
19Feb20

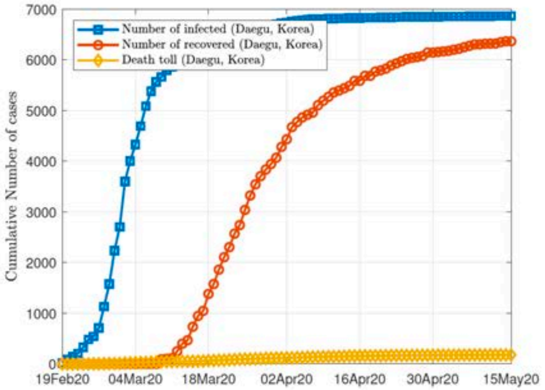

Guangzhou
Cumulative Number of cases
Number of infected (Guangzhou)
Number of recovered (Guangzhou)
Death toll (Guangzhou)
0
100
200
300
400
21Jan20
25Jan20
30Jan20
04Feb20
09Feb20
14Feb20
19Feb20


Tianjin
Cumulative Number of cases
Number of infected (Tianjin)
Number of recovered (Tianjin)
Death toll (Tianjin)
0
50
100
150
21Jan20
25Jan20
30Jan20
04Feb20
09Feb20
14Feb20
19Feb20

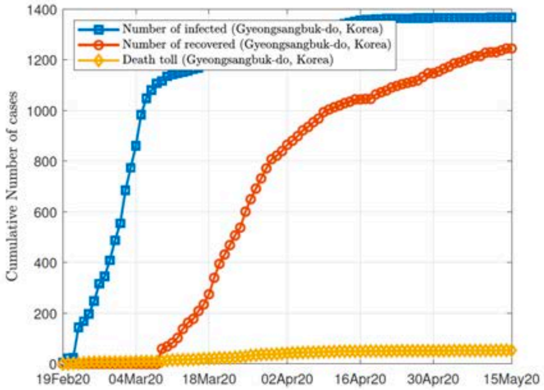
Other countries

South Korea
Italy
Iran
Japan
USA
UK

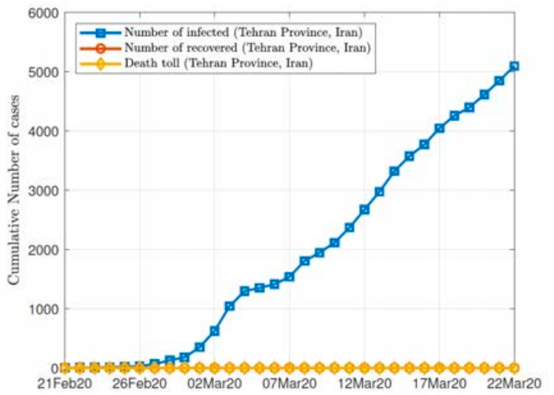
Over 500 cities



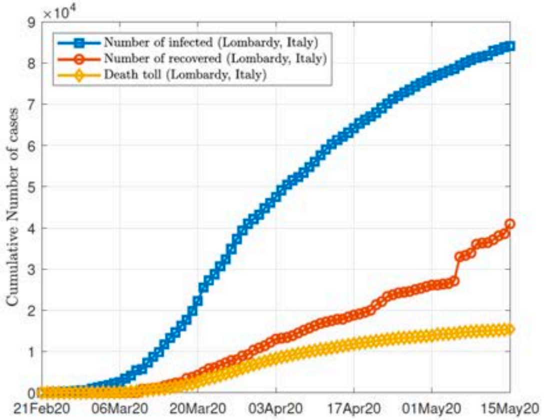
(a) Daegu



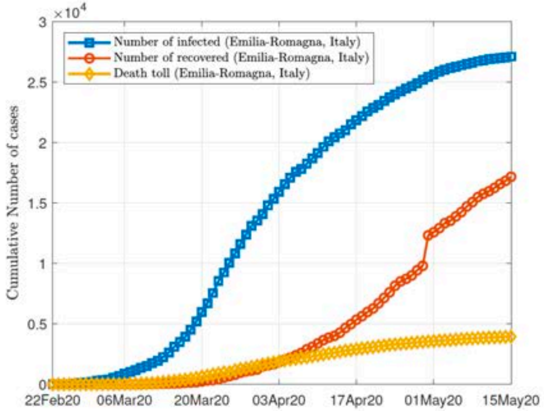
(b) Gyeongsangbuk-do



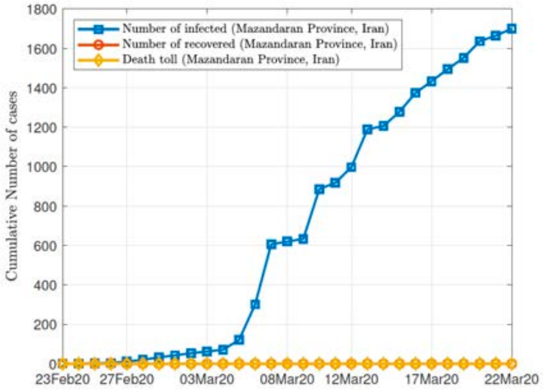
(e) Tehran



(c) Lombardy



(d) Emilia-Romagna



(f) Mazandaran

Problem of Information Latency and Loss

- One reason for Rapid Spread of Virus:
  - *Delayed information — unawareness of the real situation* in a population
- Most notable information latency lies in the number of confirmed cases reported, which depends on the ability of the particular country or city to perform tests as well as the possible bureaucracy in the local system of reporting.

$$I > C \text{ or } I \gg C$$

  - The number of confirmed cases is ***certainly not the true number*** of infected individuals at any given time

Model

Issue 1: Travel Data Integration

Issue 2: Information Latency

Review of SEIR Model

On day *t*, there are

- **S**usceptible $S_i(t)$
- **E**xposed $E_i(t)$
- **I**nfected $I_i(t)$
- **R**ecovered/ removed $R_i(t)$

City *i* with population = P_i

Eventual percentage of infection = δ_i

Total susceptible population $N_i^s = \delta_i P_i$

$$N_i^s(t) = S_i(t) + E_i(t) + I_i(t) + R_i(t)$$

IMAGE: https://zabulionis.files.wordpress.com/2014/06/people-network.jpg

Basic SEIR Dynamics

For a single population (network), we can write the state equations as

$$\begin{aligned}\dot{S}(t) &= -\beta S(t)I(t)\\ \dot{E}(t) &= \beta S(t)I(t) - kE(t)\\ \dot{I}(t) &= \kappa E(t) - \gamma I(t)\\ \dot{R}(t) &= \gamma I(t)\end{aligned}$$

$$\Rightarrow \left\{\begin{aligned}\Delta S(t) &= -\beta S(t-1)I(t-1)\\ \Delta E(t) &= \beta S(t-1)I(t-1) - \kappa E(t-1)\\ \Delta I(t) &= \kappa E(t-1) - \gamma I(t-1)\\ \Delta R(t) &= \gamma I(t-1)\end{aligned}\right.$$

Parameters

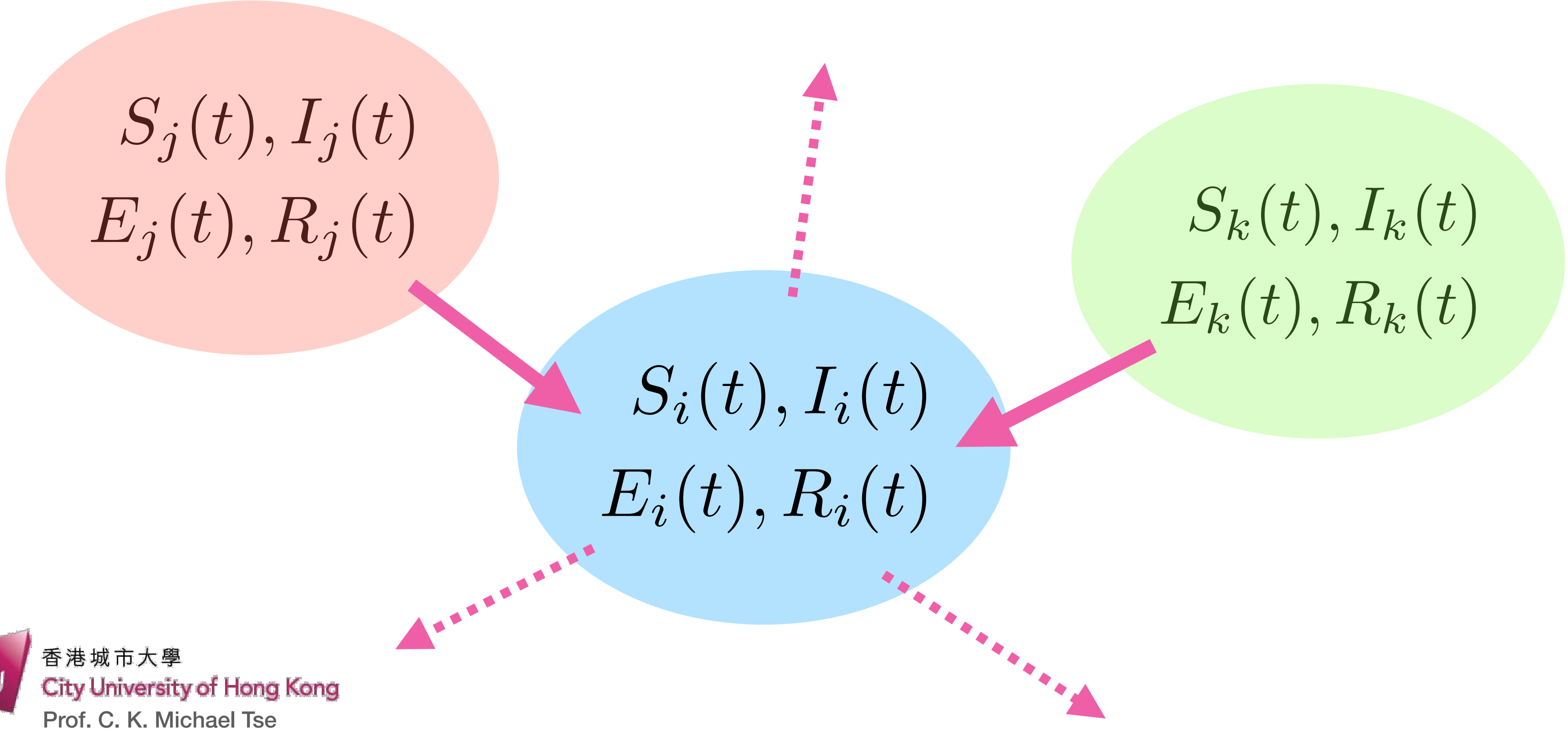
- Infection rates: $S \rightarrow E$ and $E \rightarrow I$, i.e., β_i and κ_i for each city
- Recovery rate: $I \rightarrow R$, i.e., γ_i for each city
- Eventual infection percentage: δ_i for each city

Data

- Population: P_i for each city
- Infected: I_i , R_i for each city

With intercity travel

e.g., 267 Chinese cities are interacting via inflow and outflow traffic



$S_j(t), I_j(t)$

$E_j(t), R_j(t)$

$S_k(t), I_k(t)$

$E_k(t), R_k(t)$

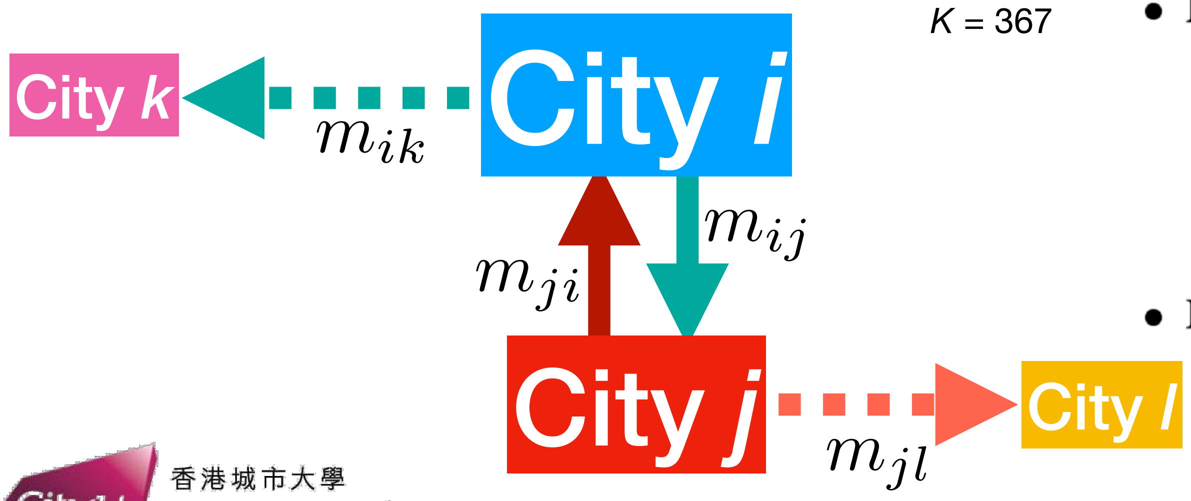
$S_i(t), I_i(t)$

$E_i(t), R_i(t)$

Intercity migration matrix

$$M(t) = \begin{bmatrix} m_{11}(t) & m_{12}(t) & \cdots & m_{1K}(t) \\ m_{21}(t) & m_{22}(t) & \cdots & m_{2K}(t) \\ \vdots & \vdots & \ddots & \vdots \\ m_{N1}(t) & m_{N2}(t) & \cdots & m_{KK}(t) \end{bmatrix}$$

$K = 367$



- M records migration from one city to another. Movement within a city is not counted, i.e., $m_{ii}(t) = 0$ for all i.
- M is non-symmetric as traffic from one city to another is not necessarily reciprocal at any given time, i.e., $m_{ij}(t) \neq m_{ji}(t)$.
- Number of outflow migrants of city i at time t is

$$m_i^{(\text{out})}(t) = \sum_{i=j}^{K} m_{ij}(t).$$

- Number of inflow migrants of city i at time t is

$$m_i^{(\text{in})}(t) = \sum_{j=1}^{K} m_{ji}(t).$$

From city *i* to city *j*

$$\Delta I_{ij}^{\text{in}}(t) = \frac{I_i(t) m_{ij}(t)}{P_i}$$

Leaving city *j*

$$\Delta I_j^{\text{out}}(t) = \frac{I_j(t) \sum_{i=1}^{N} m_{ji}(t)}{P_j}$$

Likewise for the exposed populations

Assumptions / Conditions

- Susceptible and Exposed individuals can move from one city to another.
- Infected individuals, once confirmed, do NOT move to another city.
- Recovered individuals also do NOT move to another city.
- All parameters are theoretically different in different cities.
  - For computational convenience, we only assume recovery rates and infection percentages to be different in different cities.

Travel-Data Integrated SEIR Model

MODEL
Set of Equations

PARAMETERS
(unknown)

daily changes

$\Delta S = ...$

$\Delta E = ...$

$\Delta I = ...$

$\Delta R = ...$

$\Delta P = ...$

α_j

β_j

γ_j

κ_j

δ_j

travel data

$$M(t) = \begin{bmatrix} m_{11}(t) & m_{12}(t) & \cdots & m_{1K}(t) \\ m_{21}(t) & m_{22}(t) & \cdots & m_{2K}(t) \\ \vdots & \vdots & \ddots & \vdots \\ m_{N1}(t) & m_{N2}(t) & \cdots & m_{KK}(t) \end{bmatrix}$$

S

$$\Delta S_j(t) = -\frac{\beta_j(t)}{N_j^s(t)} I_j(t) S_j(t) - \frac{\alpha_j(t)}{N_j^s(t)} E_j(t) S_j(t) + \sum_{i=1}^{N} \left(\frac{S_i(t) m_{ij}(t)}{P_i(t)} \right) - \frac{S_j(t) \sum_{i=1}^{N} m_{ji}(t)}{P_j(t)}$$

E

$$\Delta E_j(t) = \frac{\beta_j(t)}{N_j^s(t)} I_j(t) S_j(t) + \frac{\alpha_j(t)}{N_j^s(t)} E_j(t) S_j(t) - \kappa_j(t) E_i(t) + \sum_{i=1}^{N} \left(\frac{E_i(t) m_{ij}(t)}{P_i(t)} \right) - \frac{E_j(t) * \sum_{i=1}^{N} m_{ji}(t)}{P_j(t)},$$

I

$$\Delta I_j(t) = \kappa_j(t) E_i(t) - \gamma_j(t) I_j(t) + k_I \left(\sum_{i=1}^{N} \left(\frac{I_i(t) m_{ij}(t)}{P_i(t)} \right) - \frac{I_j(t) * \sum_{i=1}^{N} m_{ji}(t)}{P_j(t)} \right)$$

R

$$\Delta R_j(t) = \gamma_j(t) I_j(t)$$

Population change:

$$\Delta P_j(t) = \sum_{i=1}^{N} m_{ij}(t) - \sum_{i=1}^{N} m_{ji}(t)$$

$$\Delta N_j^s(t) = k_I \left(\sum_{i=1}^{N} \left(\frac{I_i(t) m_{ij}(t)}{P_i(t)} \right) - \frac{I_j(t) * \sum_{i=1}^{N} m_{ji}(t)}{P_j(t)} \right) + \sum_{i=1}^{N} \left(\frac{E_i(t) m_{ij}(t)}{P_i(t)} \right) - \frac{E_j(t) * \sum_{i=1}^{N} m_{ji}(t)}{P_j(t)} + \sum_{i=1}^{N} \left(\frac{S_i(t) m_{ij}(t)}{P_i(t)} \right) - \frac{S_j(t) \sum_{i=1}^{N} m_{ji}(t)}{P_j(t)}$$

Here we do not discuss the contact topology.

Parameter Extraction

- Set of parameters: $\theta_j = \{\alpha_j, \beta_j, \gamma_j, \kappa_j, \delta_j, I_{j,0}\}$ for city *j*
- Initial infected and exposed numbers: $\lambda_I I_j(t_0)$ and $\lambda_E I_j(t_0)$

Modeling to be completed by finding all parameters that match the historical data.

The number of exposed individuals, $E_j(t)$, is also unknown.

Method

- *Constrained nonlinear optimisation* problem

$$P_0:\min_{\Theta}\sum_{j=0}^{N}\left\|w_j(I(t_j)-\hat{I}(t_j))\right\|_l$$

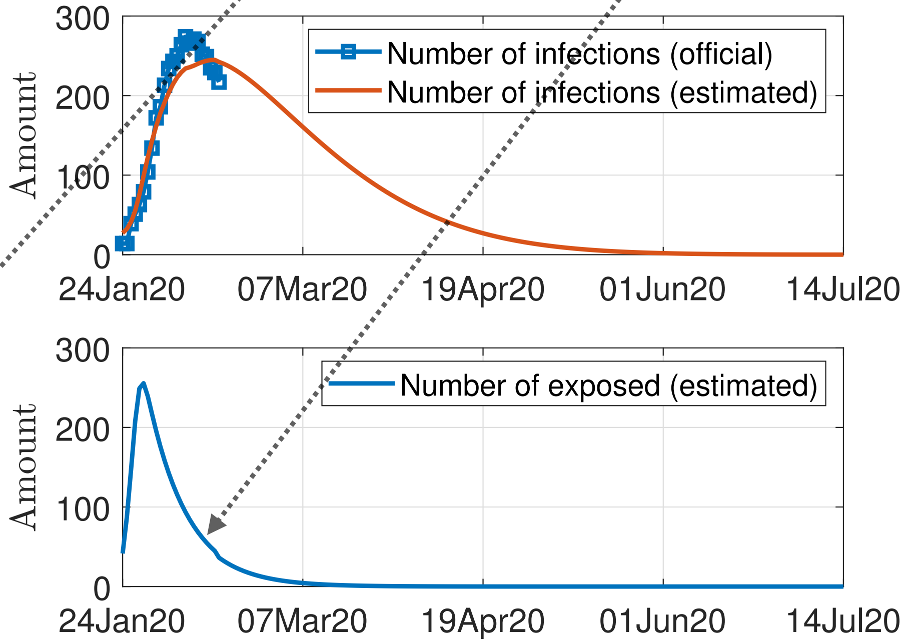
wh

$$s.t.\begin{cases}(i) & \hat{x}(t+1)=\hat{x}(t)+F(\hat{x}(t)),\\(ii) & \Theta_U\geq\Theta\geq\Theta_L,\end{cases}$$

where $\hat{x}$ is the set of estimated parameters, and $F(\,.\,)$ is the model.

- $\hat{x}(t)=[\hat{I}(t),\hat{R}(t),E(t),S(t),P(t),N^s(t)]$

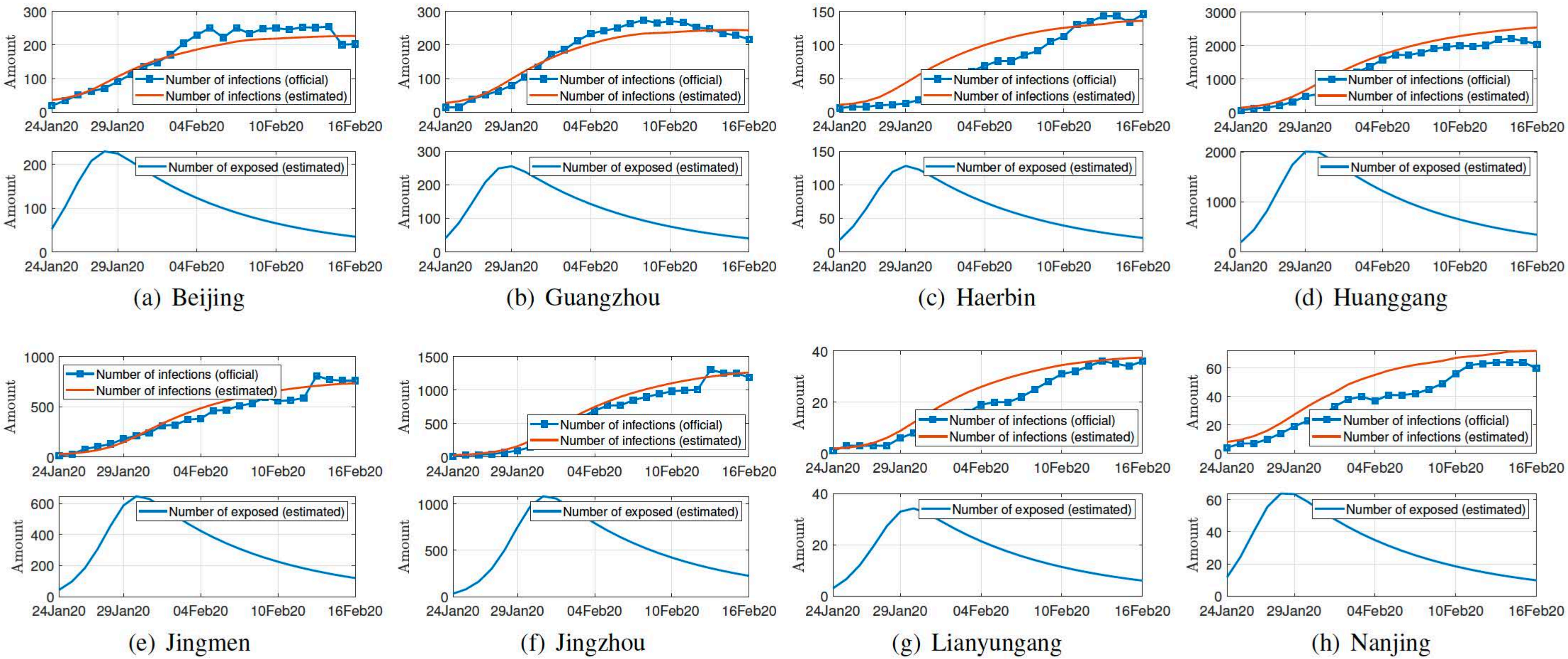
DATA FITTING with MIN ERROR
to extract all parameters and unknown variables



In other words, the model generates the solution that is ***as close to the data as possible***.

Fitting Results

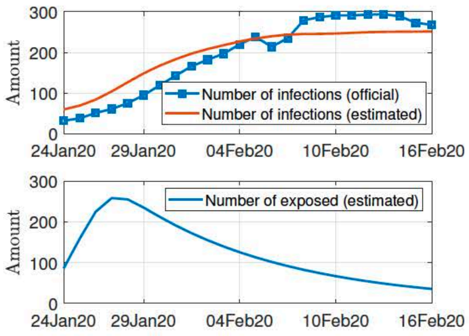
No of infected cases
||
Not the total cumulated
(i.e., less the recovered)



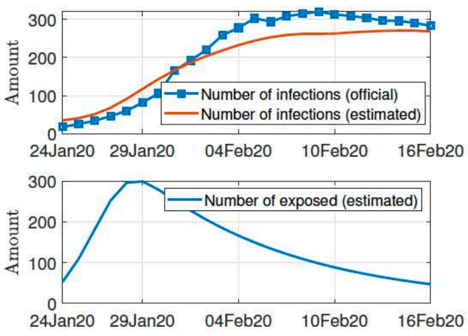
(a) Beijing (b) Guangzhou (c) Haerbin (d) Huanggang

(e) Jingmen (f) Jingzhou (g) Lianyungang (h) Nanjing

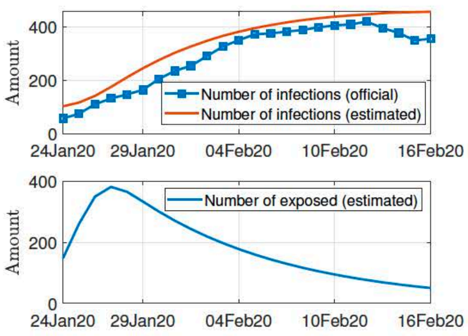
PERFORMED FEBRUARY 17, 2020



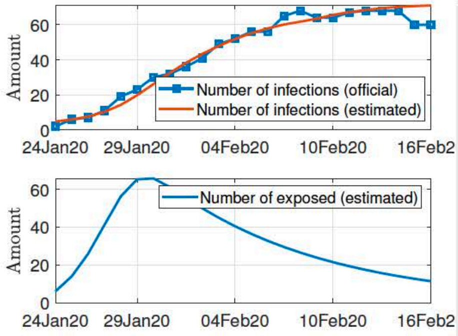
(i) Shanghai



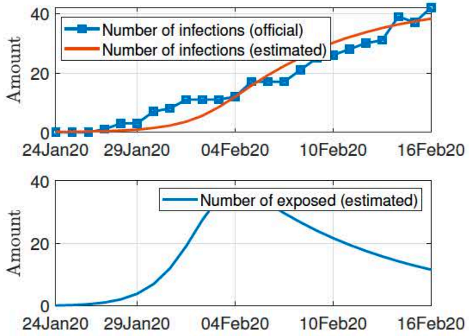
(j) Shenzhen



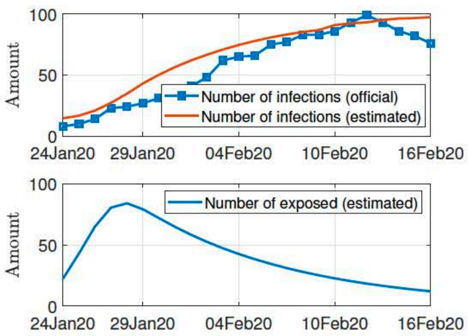
(k) Chongqing



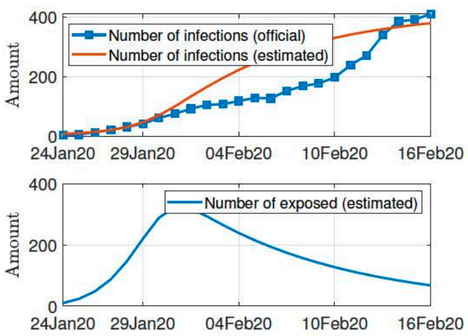
(l) Suzhou



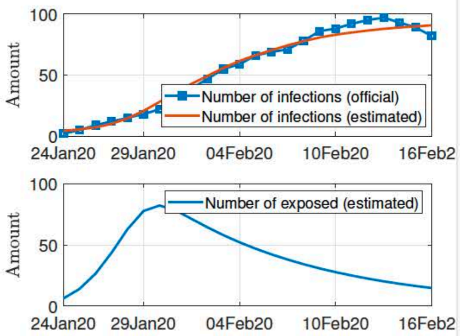
(m) Tangshan



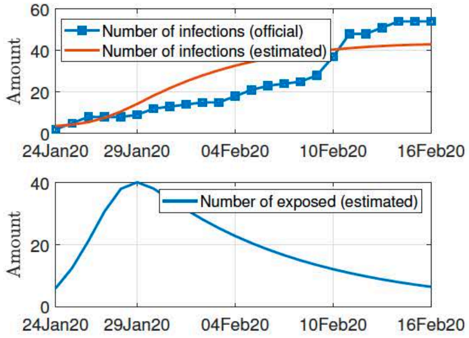
(n) Tianjin



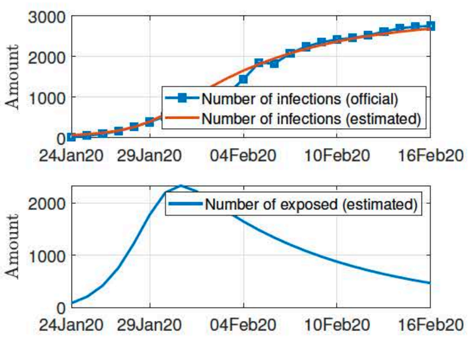
(o) Tianmen



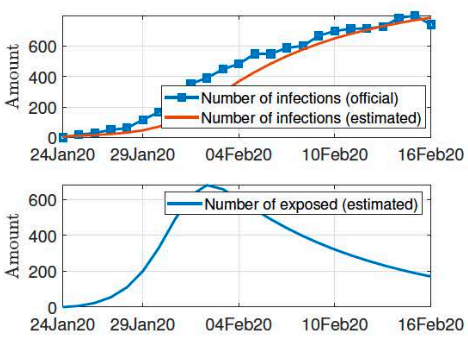
(p) Xian



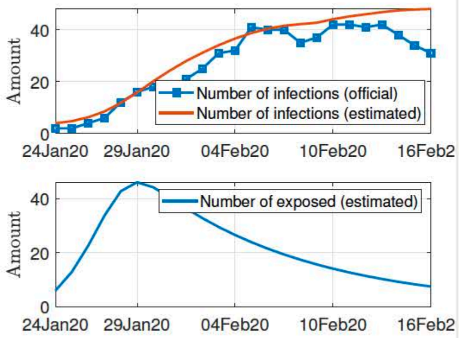
(q) Hong Kong



(r) Xiaogan



(s) Yichang

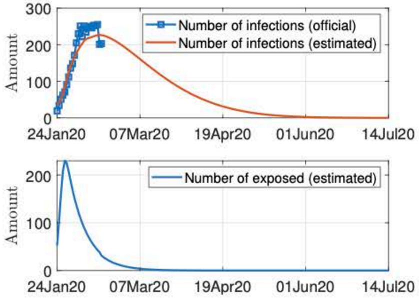


(t) Zhongshan

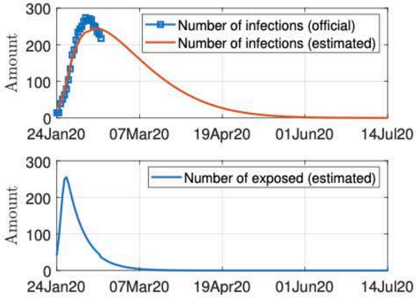
Predictions

- Once we have the parameters, the model can be used for prediction assuming the validity of the same set of parameters.

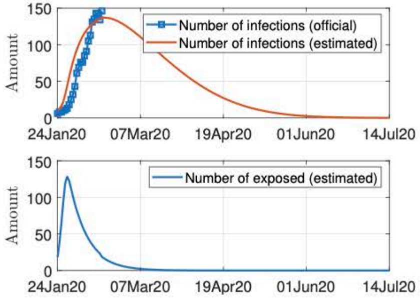
No of infected cases
‖
Not the total cumulated
(i.e., less the recovered)



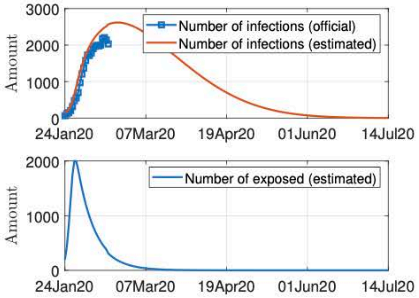
(a) Beijing



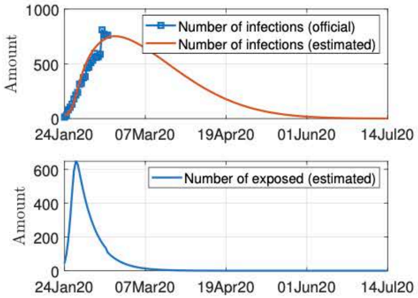
(b) Guangzhou



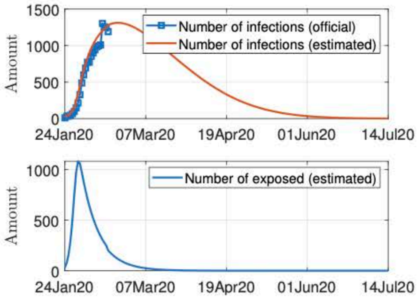
(c) Haerbin



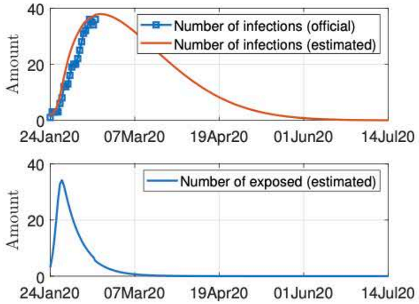
(d) Huangang



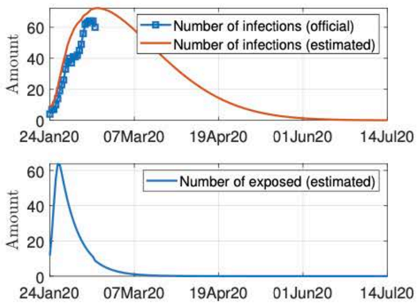
(e) Jingmen



(f) Jingzhou

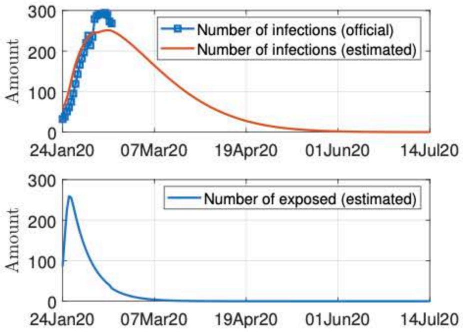


(g) Lianyungang

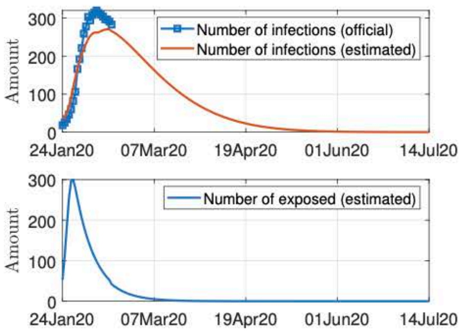


(h) Nanjing

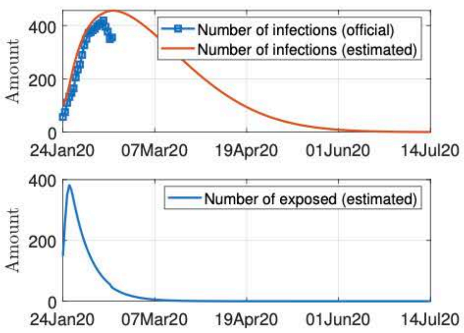
PERFORMED FEBRUARY 17, 2020



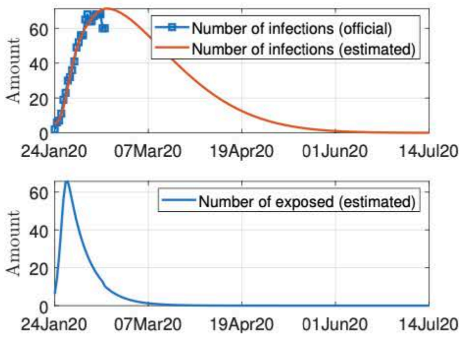
(i) Shanghai



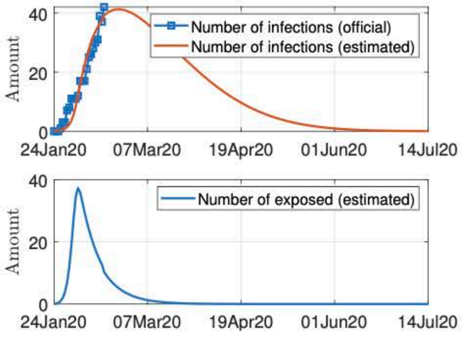
(j) Shenzhen



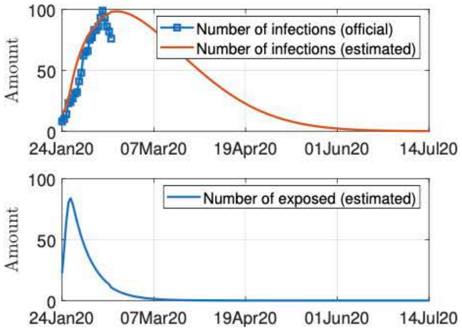
(k) Chongqing



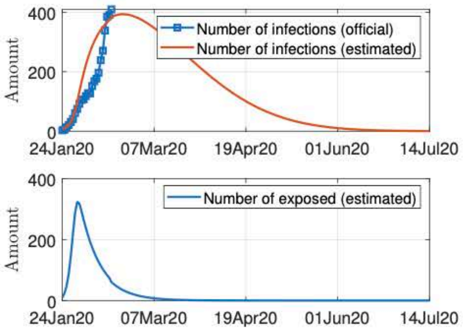
(l) Suzhou



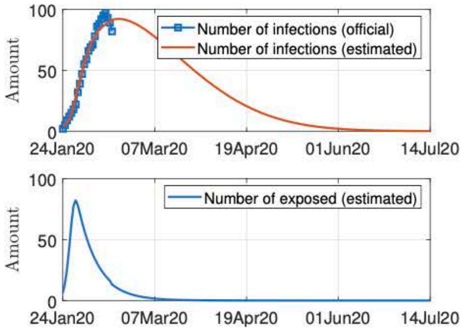
(m) Tangshan



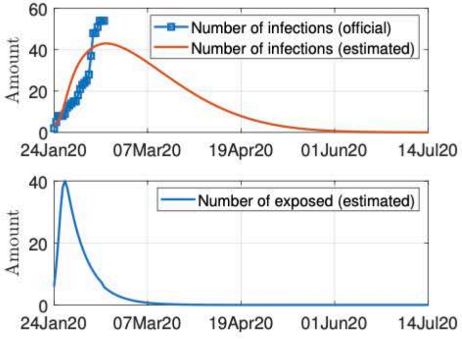
(n) Tianjin



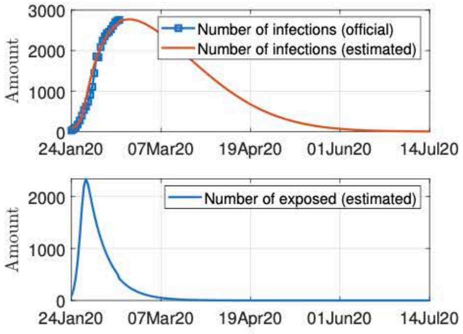
(o) Tianmen



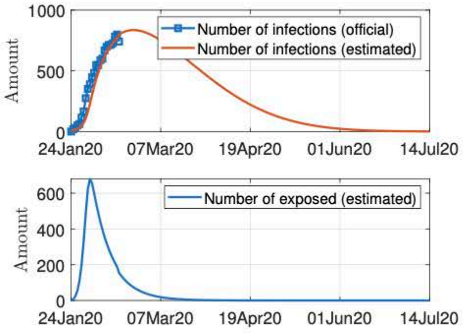
(p) Xian



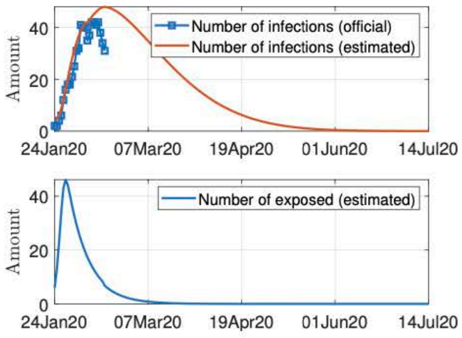
(q) Hong Kong



(r) Xiaogan

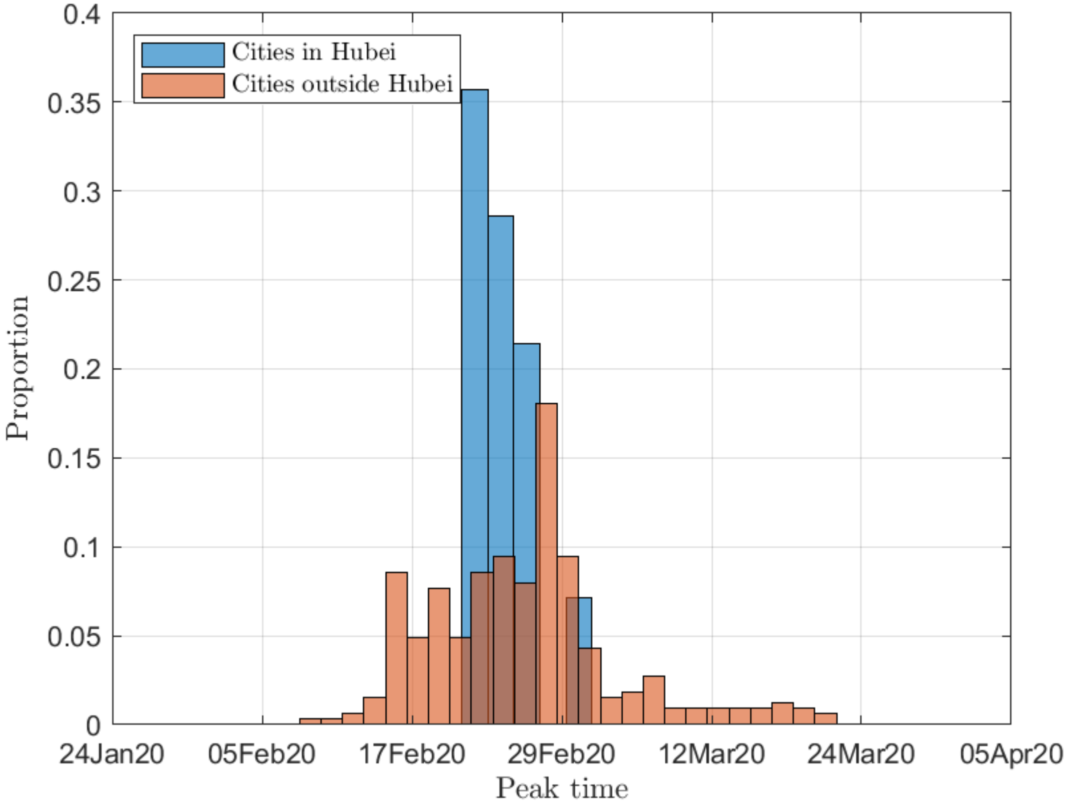


(s) Yichang



(t) Zhongshan

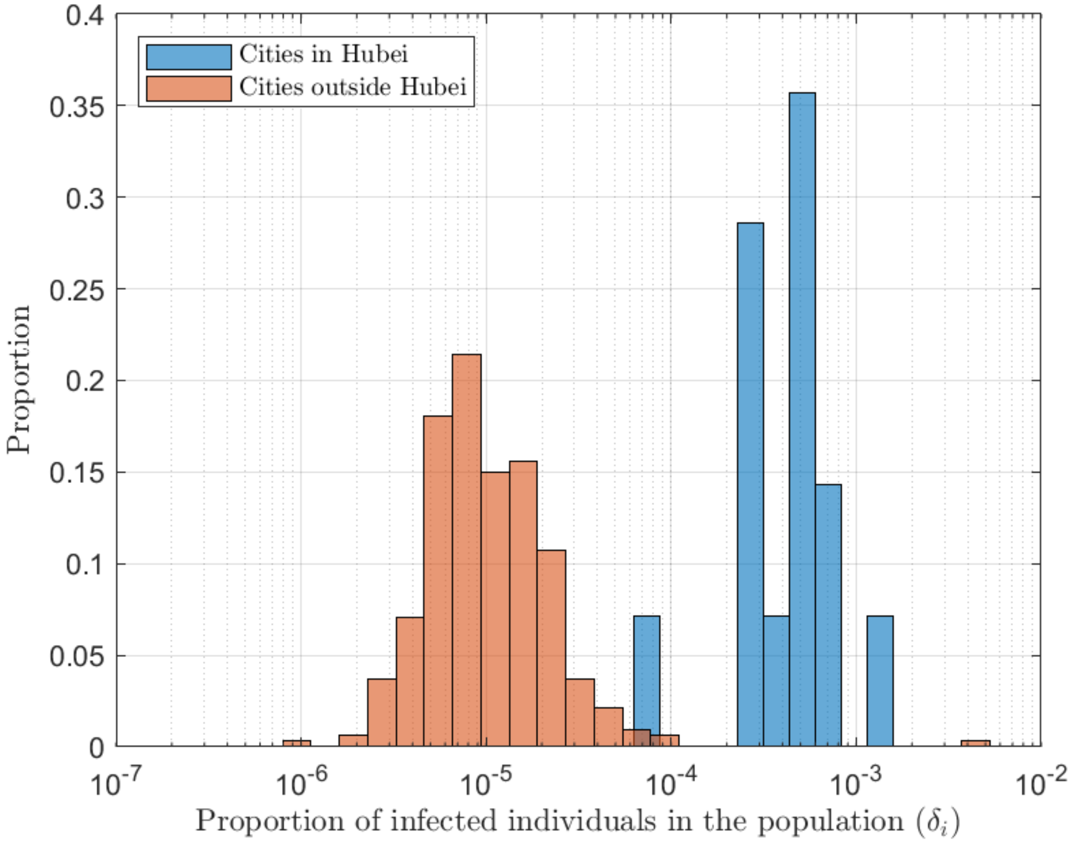
Findings



PEAK TIME

For most cities, the infection numbers would peak between mid February to early March 2020

PERFORMED FEBRUARY 17, 2020



January 24, 2020, to February 16, 2020

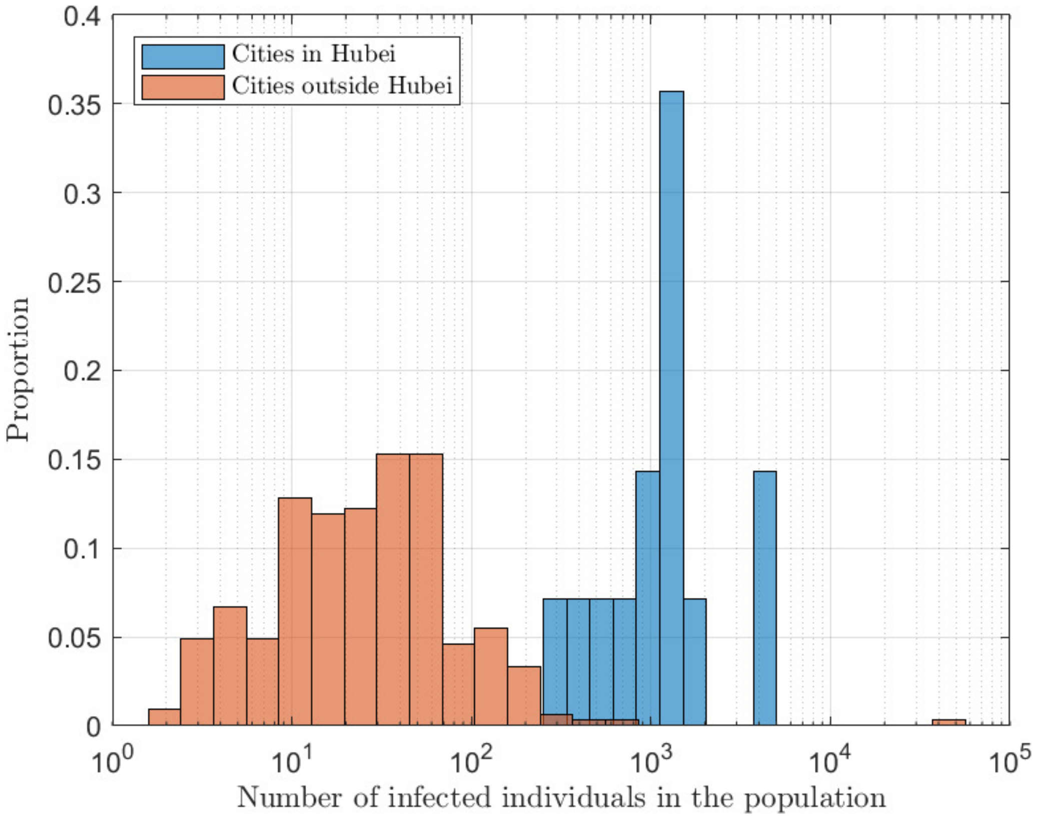
PROPORTION OF INFECTED IN EACH CITY δ_i

At the end, about 0.8%, less than 0.1% and less than 0.01% of the population would get infected in Wuhan, Hubei Province and the rest of China, respectively.

PERFORMED FEBRUARY 17, 2020

NUMBER OF INFECTED IN EACH CITY

Translating to actual figures, for most cities outside and within Hubei Province (except Wuhan), the total number of infected individuals was expected to be fewer than 300 and 4000, respectively



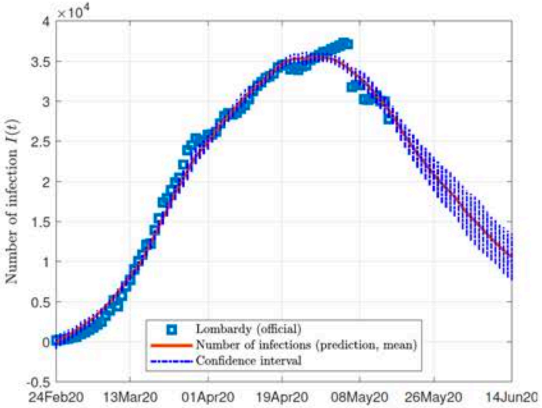
January 24, 2020, to February 16, 2020

PERFORMED FEBRUARY 17, 2020

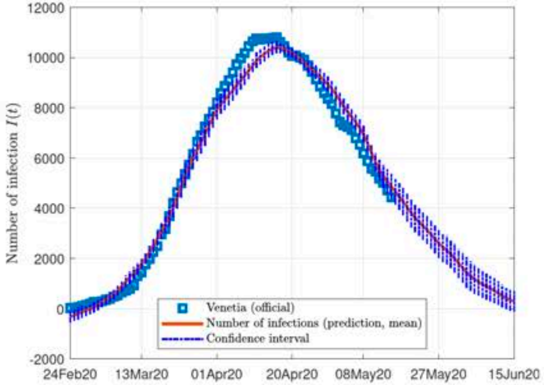
Other countries by data-driven prediction

South Korea
Italy
Iran
Japan
USA
UK

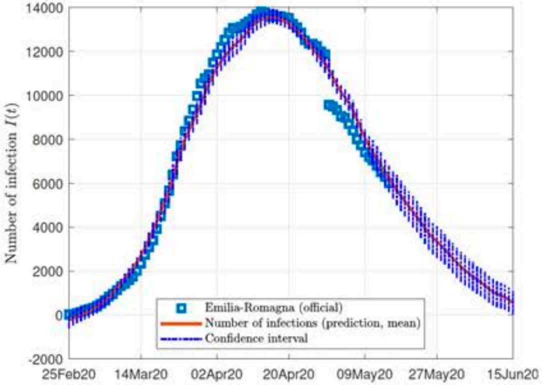
Over 500 cities



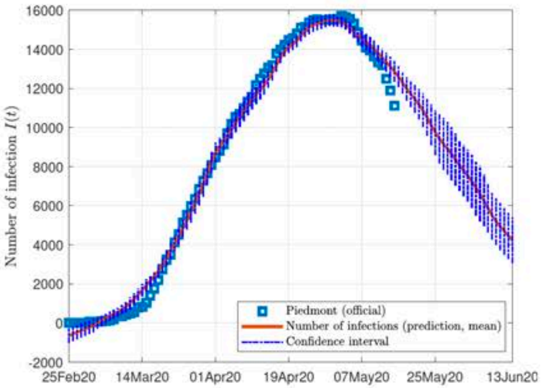
(a) Lombardy



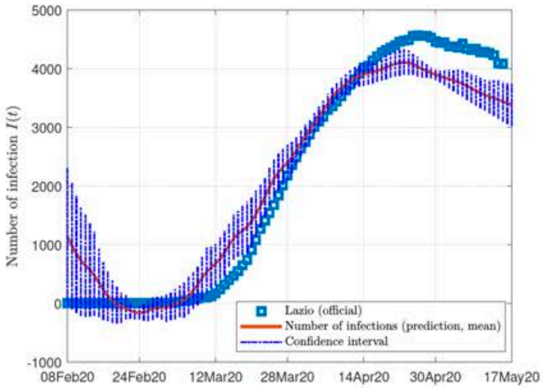
(b) Venetia



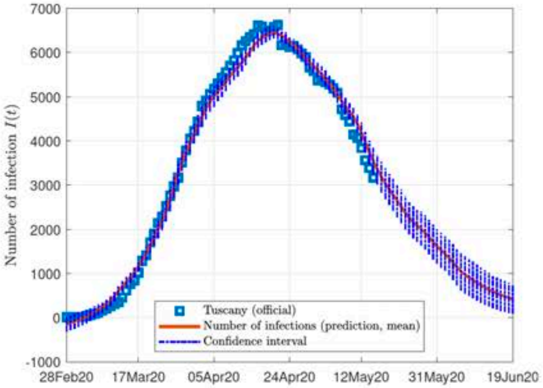
(c) Emilia-Romagna



(d) Piedmont



(e) Lazio



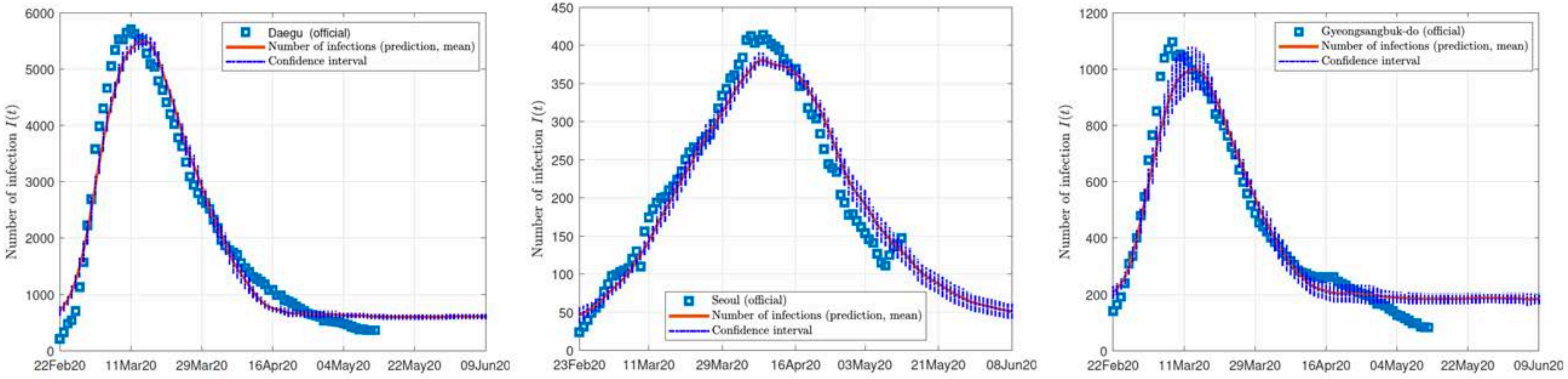
(f) Tuscany

PERFORMED MARCH 8, 2020; UPDATED MAY 17, 2020

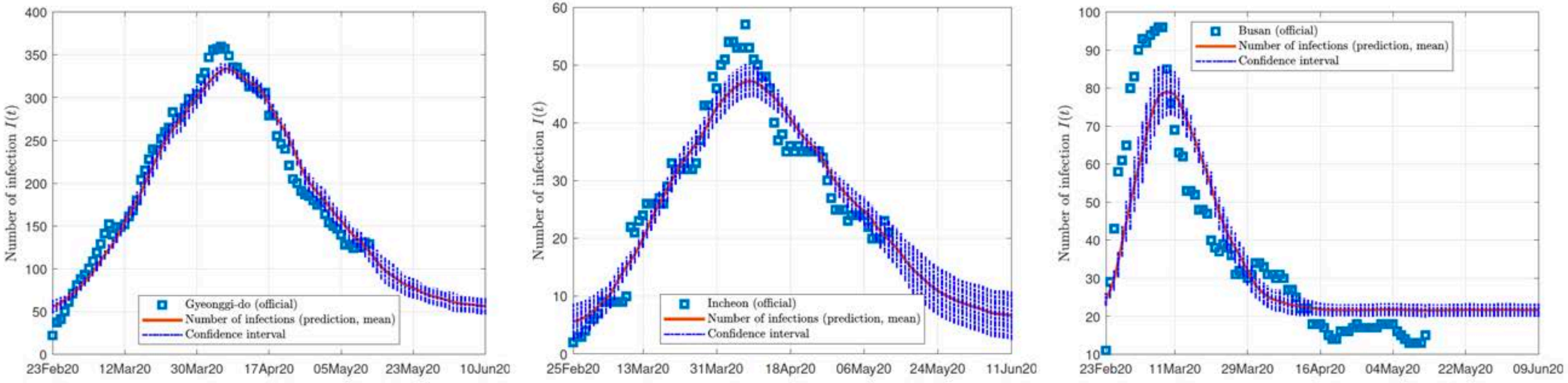
Other countries by data-driven prediction

South Korea
Italy
Iran
Japan
USA
UK

Over 500 cities



(a) Daegu (b) Seoul (c) Gyeongsangbuk-do



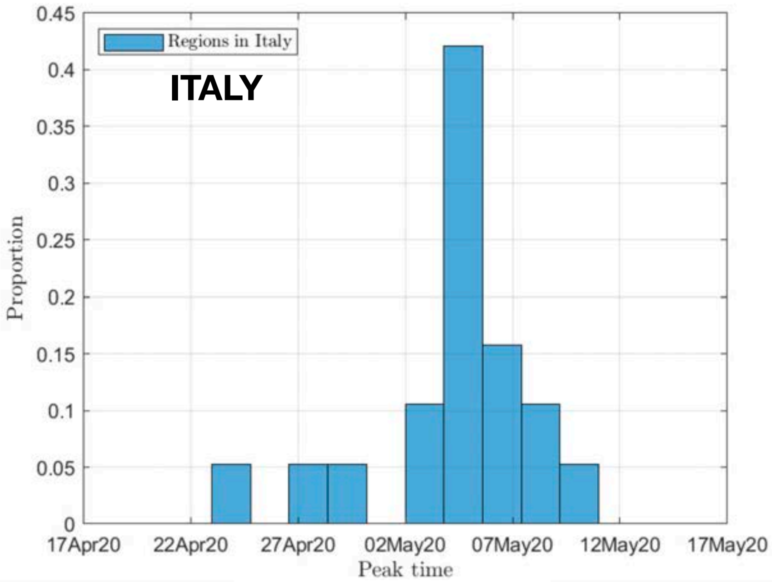
(d) Gyeonggi-do (e) Incheon (f) Busan

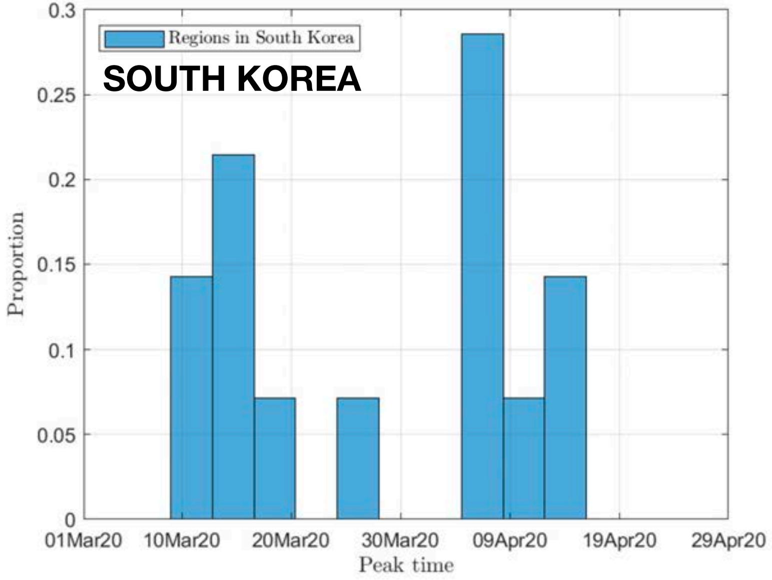
PERFORMED MARCH 8, 2020; UPDATED MAY 17, 2020

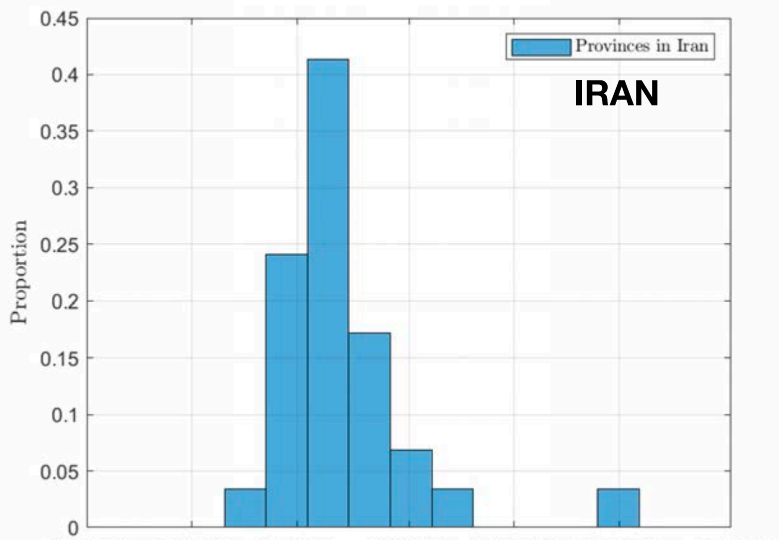
PEAK TIME

South Korea — end by April

Italy, Iran — end by May

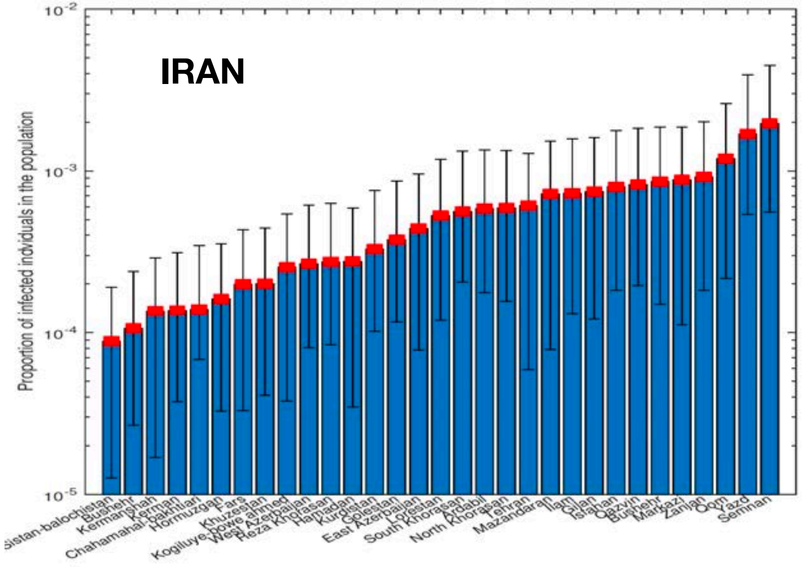


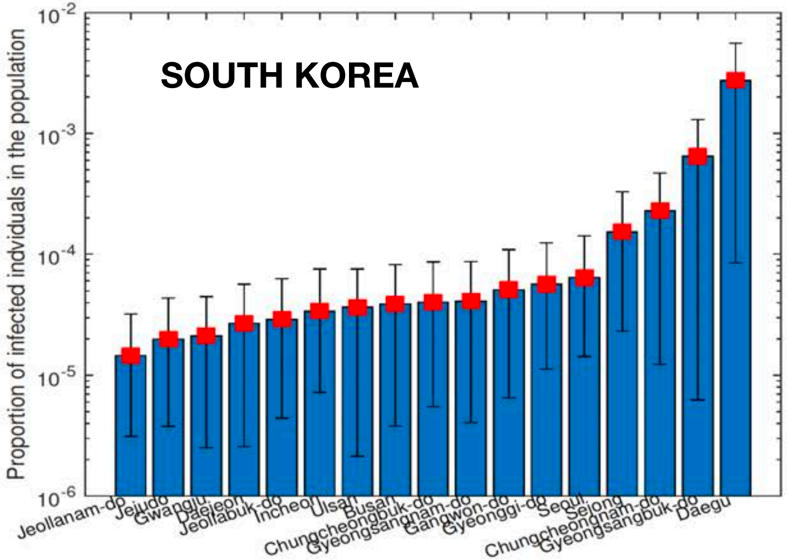


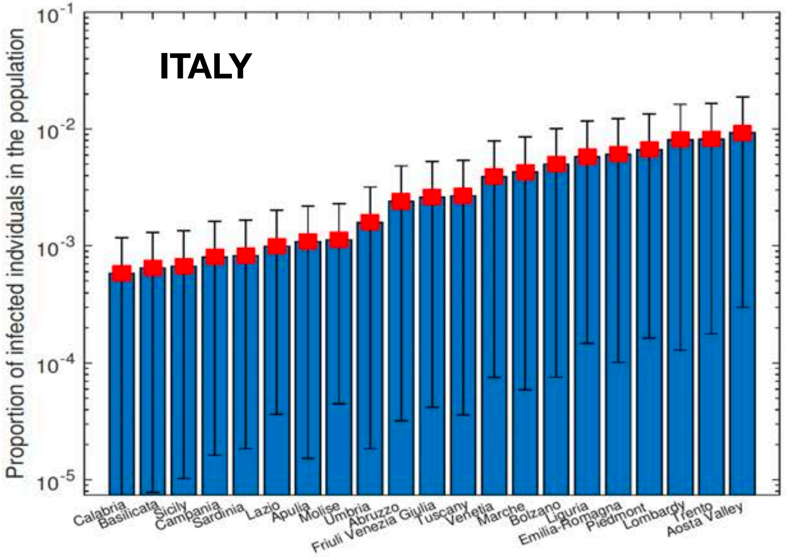


PERFORMED MARCH 8, 2020; UPDATED MAY 17, 2020

PROPORTION OF INFECTED INDIVIDUALS
IN THE POPULATION







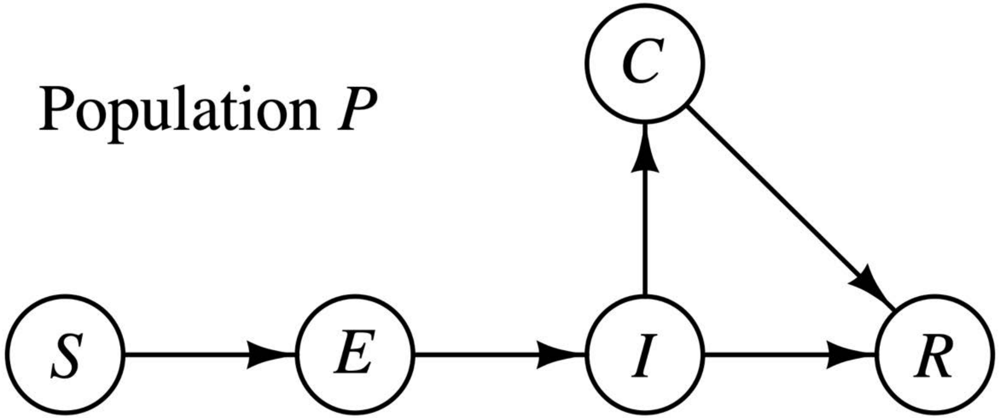
PERFORMED MARCH 8, 2020; UPDATED MAY 17, 2020

Next Challenge: Latency / Loss of Information

Most notable information latency — the number of confirmed cases reported, which depends on

- the ability of the particular country or city to perform tests
- the possible bureaucracy in the local system of reporting

$$I \gg C$$



OUR LATEST WORK (COMPLETED JULY 27) SHOWED THAT C / I ≈ 0.23, i.e.,

Unconfirmed : confirmed ≈ 3.309 : 1

based on data from Australia, Canada, Italy, Japan, Spain, UK and USA

SEICR MODEL

Assume: the number of individuals eventually infected is set initially at $N_j^s(t_0) = \delta_j P_j$ (δ_j being constant), implying effective measures in place to limit the upper bound of the infected number.

But the model may consider inactive or less active intervention: each infected case *adds* to the eventual infected number. The additional S_j and N_j are:

$$\boxed{k_j^{(c)}\left(\Delta I_j(t)+\Delta E_j(t)\right)\left(1-\frac{N_j^s(t)}{k_h P_j(t)}\right)}$$

$k_j^{(c)}$ is an inverse indicator of the level of active intervention implemented, and corresponds quantitatively to an increase in the number of eventual infected individuals for each additional infected or exposed individual in region *j*

$$\Delta I_j(t) = (1-\lambda_j-\gamma_j^{(I)})I_j(t)+\kappa E_i(t)+k_I\left(\sum_{i=1}^{N}\left(\frac{I_i(t)m_{ij}(t)}{P_i(t)}\right)-\frac{I_j(t)\sum_{i=1}^{N}m_{ji}(t)}{P_j(t)}\right)$$

$$\Delta E_j(t) = \frac{\beta_j I_j(t)+\rho_j C_j(t)}{N_j^s(t)}S_j(t)+\frac{\alpha_j E_j(t)}{N_j^s(t)}S_j(t)-\kappa E_i(t)+\sum_{i=1}^{N}\left(\frac{E_i(t)m_{ij}(t)}{P_i(t)}\right)$$
$$-\frac{E_j(t)*\sum_{i=1}^{N}m_{ji}(t)}{P_j(t)}$$

$$\Delta S_j(t) = -\frac{\beta_j I_j(t)+\rho_j C_j(t)}{N_j^s(t)}S_j(t)-\frac{\alpha_j E_j(t)}{N_j^s(t)}S_j(t)+\sum_{i=1}^{N}\left(\frac{S_i(t)m_{ij}(t)}{P_i(t)}\right)$$
$$-\frac{S_j(t)\sum_{i=1}^{N}m_{ji}(t)}{P_j(t)} \quad + \boxed{k_j^{(c)}\left(\Delta I_j(t)+\Delta E_j(t)\right)\left(1-\frac{N_j^s(t)}{k_h P_j(t)}\right)}$$

$$\Delta C_j(t) = \lambda_j I_j(t)-\gamma_j^{(C)}C_j(t)$$

$$\Delta R_j(t) = \gamma_j^{(I)}I_j(t)+\gamma_j^{(C)}C_j(t)$$

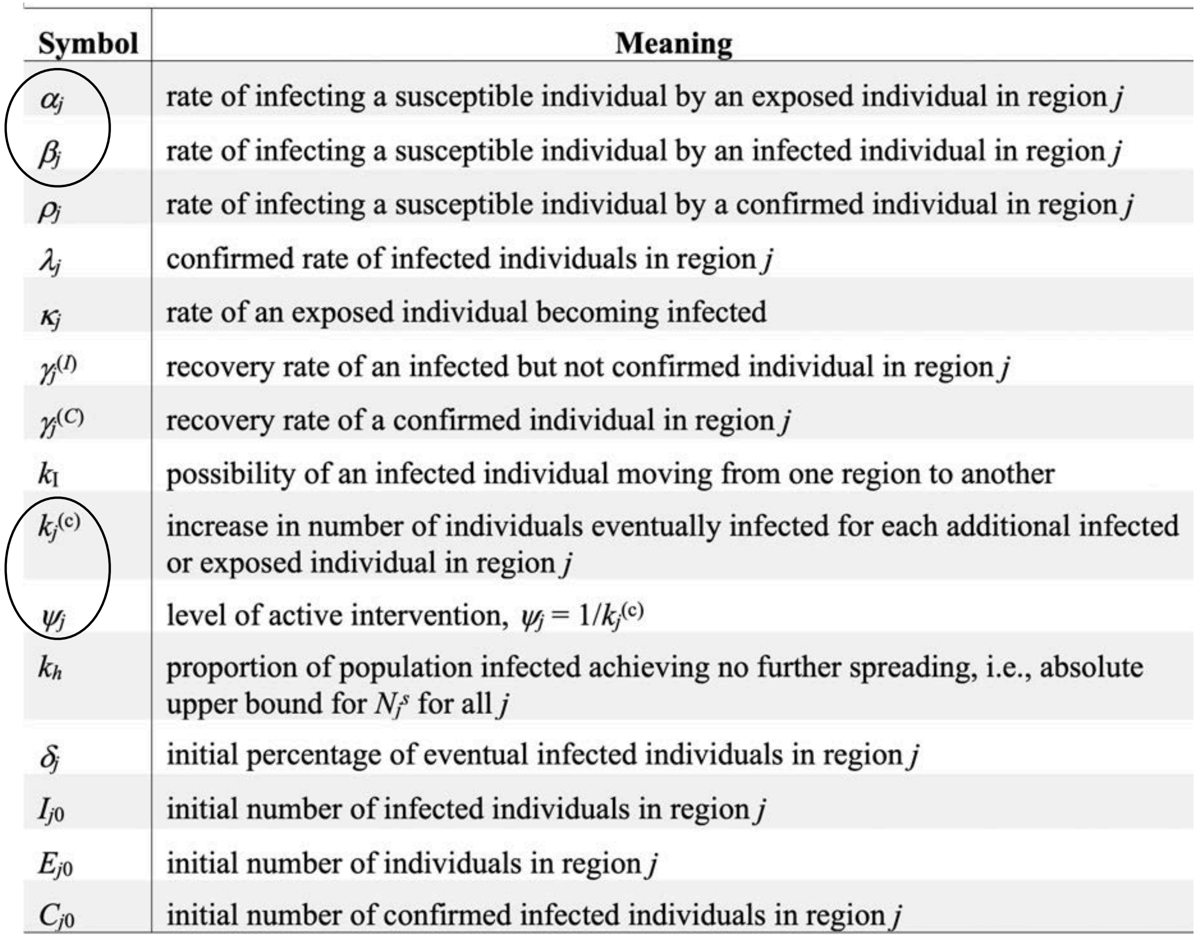
$$\Delta P_j(t) = \sum_{i=1}^{N}m_{ij}(t)-\sum_{i=1}^{N}m_{ji}(t)$$

$$\Delta N_j^s(t) = k_I\left(\sum_{i=1}^{N}\left(\frac{I_i(t)m_{ij}(t)}{P_i(t)}\right)-\frac{I_j(t)\sum_{i=1}^{N}m_{ji}(t)}{P_j(t)}\right)+\sum_{i=1}^{N}\left(\frac{E_i(t)m_{ij}(t)}{P_i(t)}\right)$$
$$-\frac{E_j(t)\sum_{i=1}^{N}m_{ji}(t)}{P_j(t)}+\sum_{i=1}^{N}\left(\frac{S_i(t)m_{ij}(t)}{P_i(t)}\right)-\frac{S_j(t)\sum_{i=1}^{N}m_{ji}(t)}{P_j(t)}$$
$$+\boxed{k_j^{(c)}\left(\Delta I_j(t)+\Delta E_j(t)\right)\left(1-\frac{N_j^s(t)}{k_h P_j(t)}\right)}$$

Full Parameter Set

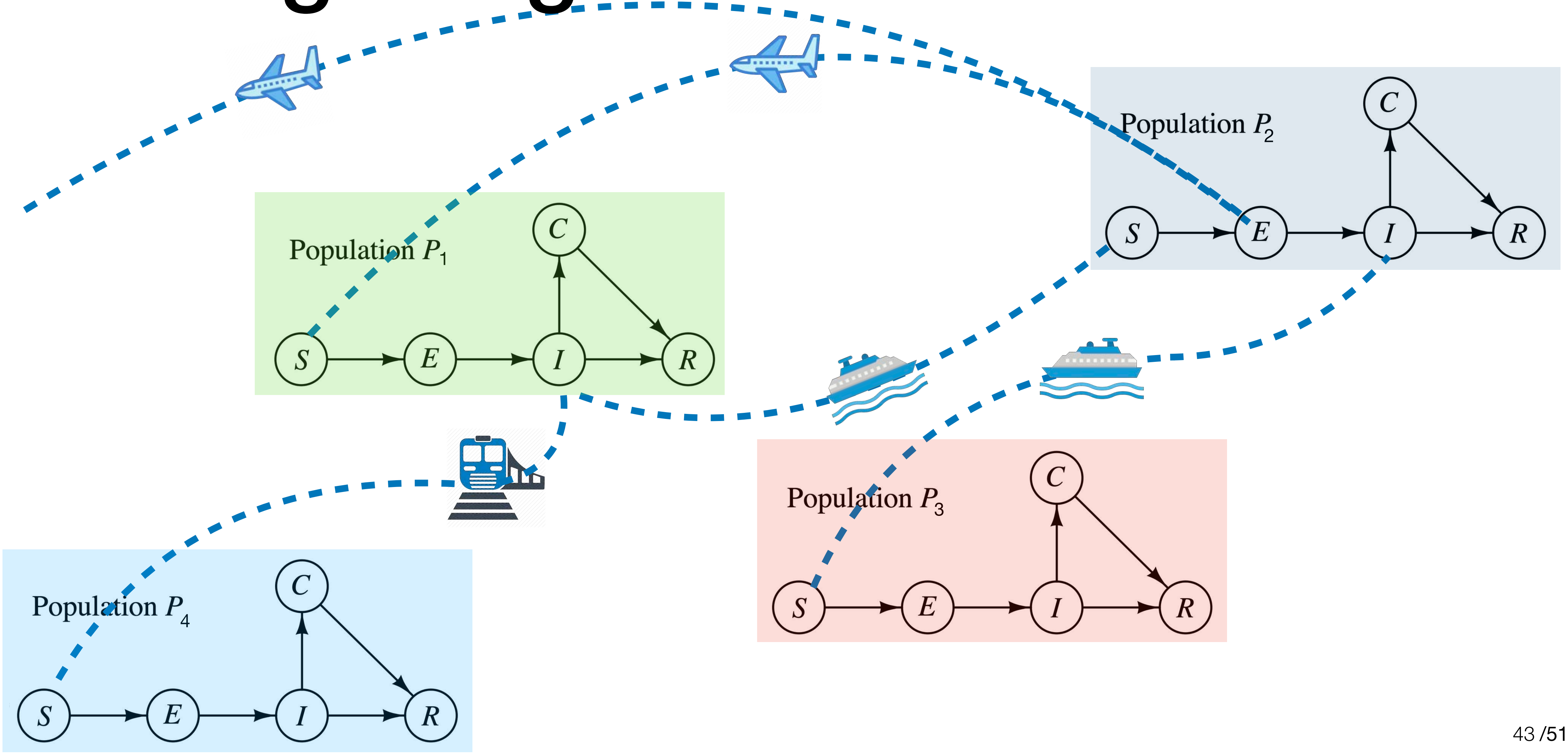
inherent rates:
social behavior
personal hygiene
mask wearing

level of intervention:
quarantine
border closure
cordon sanitaire



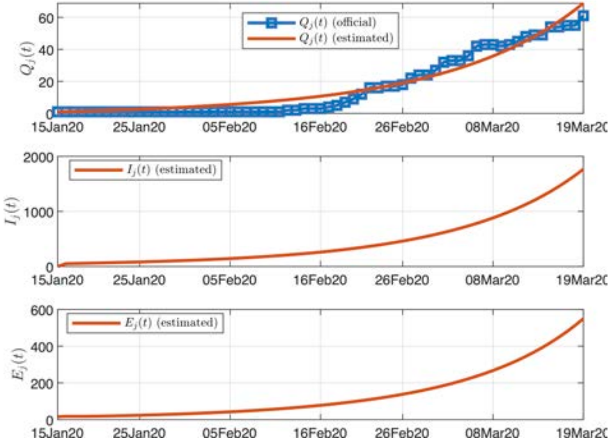
Symbol	Meaning
α_j	rate of infecting a susceptible individual by an exposed individual in region *j*
β_j	rate of infecting a susceptible individual by an infected individual in region *j*
ρ_j	rate of infecting a susceptible individual by a confirmed individual in region *j*
λ_j	confirmed rate of infected individuals in region *j*
κ_j	rate of an exposed individual becoming infected
$\gamma_j^{(I)}$	recovery rate of an infected but not confirmed individual in region *j*
$\gamma_j^{(C)}$	recovery rate of a confirmed individual in region *j*
k_{I}	possibility of an infected individual moving from one region to another
$k_j^{(c)}$	increase in number of individuals eventually infected for each additional infected or exposed individual in region *j*
ψ_j	level of active intervention, $\psi_j = 1/k_j^{(c)}$
k_h	proportion of population infected achieving no further spreading, i.e., absolute upper bound for N_j^s for all *j*
δ_j	initial percentage of eventual infected individuals in region *j*
I_{j0}	initial number of infected individuals in region *j*
E_{j0}	initial number of individuals in region *j*
C_{j0}	initial number of confirmed infected individuals in region *j*

Integrating Travel Data w SEICR



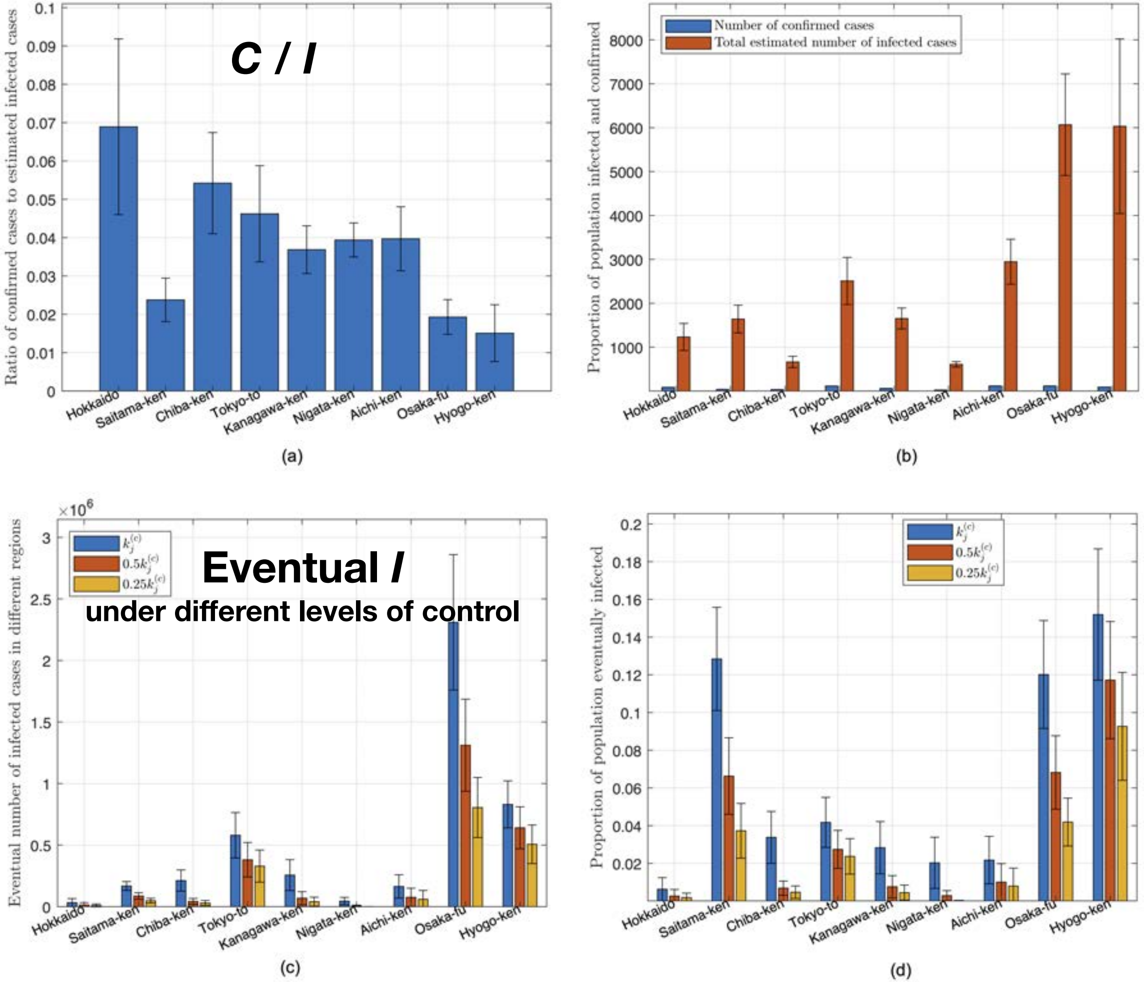
Prediction for 46 Regions (Prefectures) in Japan

Data-fitting example for Kanagawa-ken (one simulation out of 100 runs)



Actual infected cases can be 20-fold as many as the confirmed number.

$$I > 20C$$



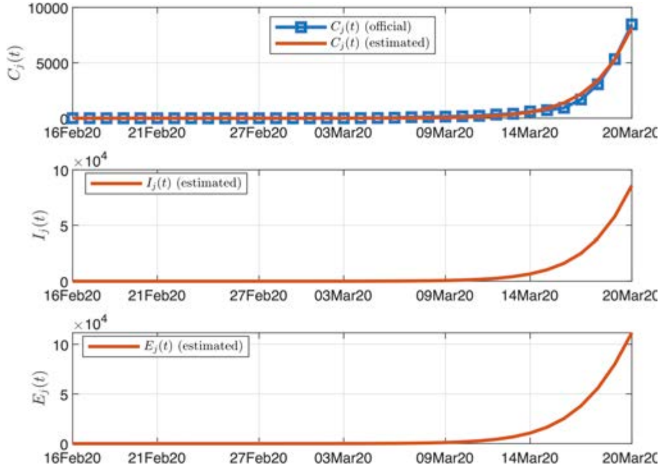
C / I

Eventual *I*

under different levels of control

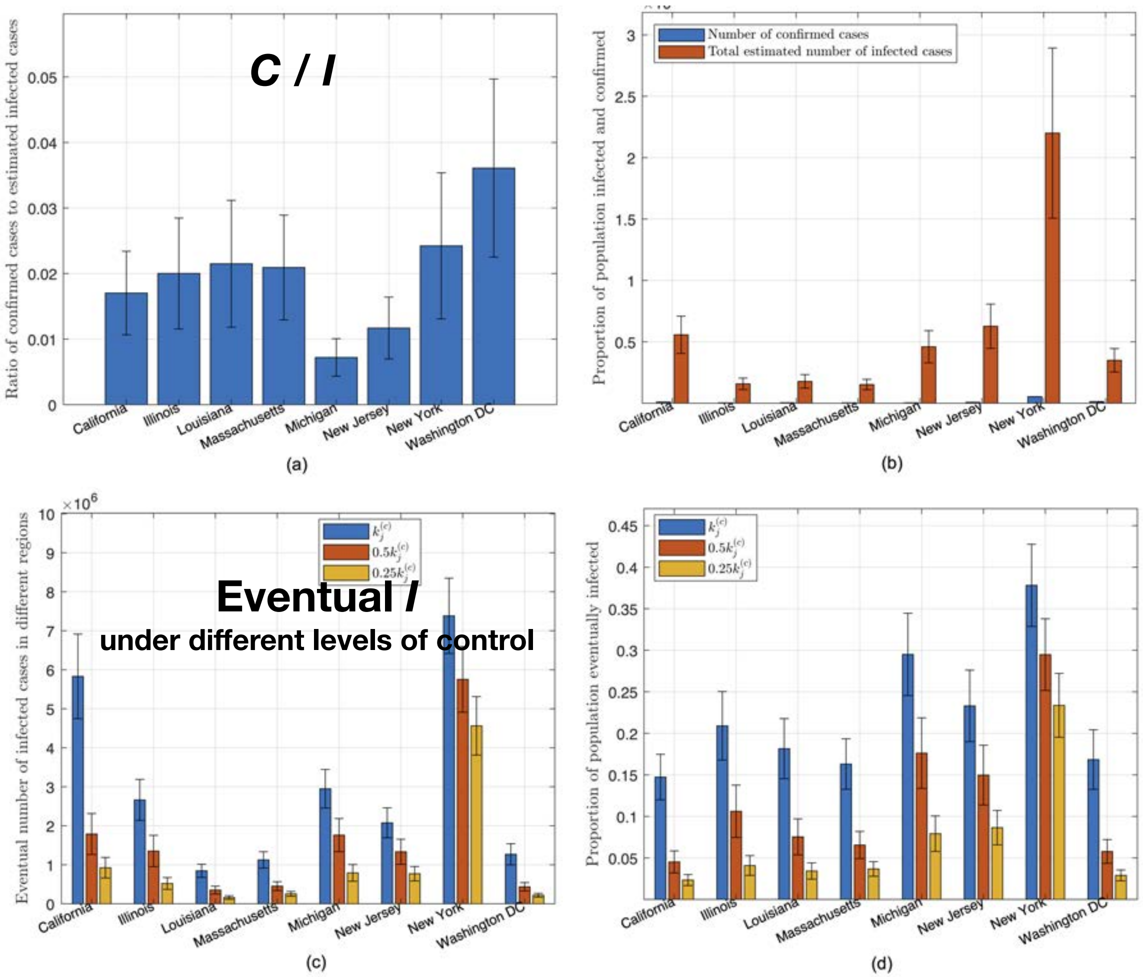
Prediction for 50 states (incl. DC) in USA

Data-fitting example for New York State
(one simulation out of 100 runs)



Actual infected cases can be 5-fold as many as the confirmed number.

$$I > 5C$$



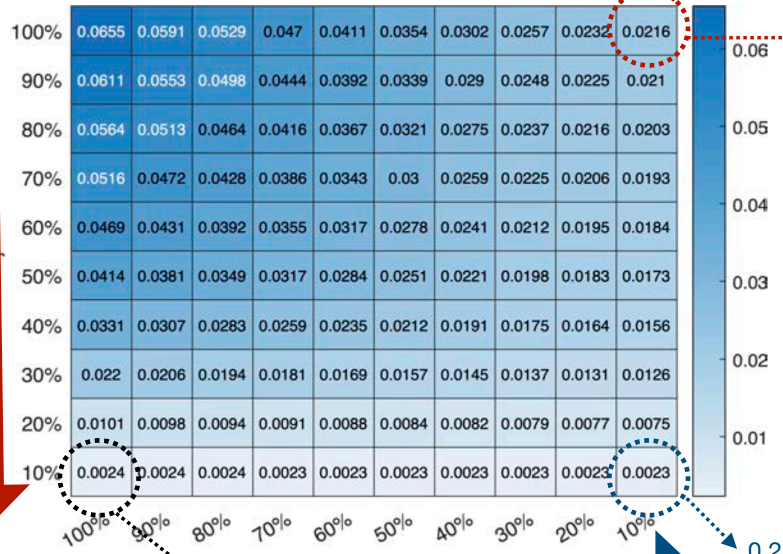
C / I

Eventual *I*
under different levels of control

香港城市大學
City University of Hong Kong

Inherent rates vs intervention

JAPAN



	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%
100%	0.0655	0.0591	0.0529	0.047	0.0411	0.0354	0.0302	0.0257	0.0232	0.0216
90%	0.0611	0.0553	0.0498	0.0444	0.0392	0.0339	0.029	0.0248	0.0225	0.021
80%	0.0564	0.0513	0.0464	0.0416	0.0367	0.0321	0.0275	0.0237	0.0216	0.0203
70%	0.0516	0.0472	0.0428	0.0386	0.0343	0.03	0.0259	0.0225	0.0206	0.0193
60%	0.0469	0.0431	0.0392	0.0355	0.0317	0.0278	0.0241	0.0212	0.0195	0.0184
50%	0.0414	0.0381	0.0349	0.0317	0.0284	0.0251	0.0221	0.0198	0.0183	0.0173
40%	0.0331	0.0307	0.0283	0.0259	0.0235	0.0212	0.0191	0.0175	0.0164	0.0156
30%	0.022	0.0206	0.0194	0.0181	0.0169	0.0157	0.0145	0.0137	0.0131	0.0126
20%	0.0101	0.0098	0.0094	0.0091	0.0088	0.0084	0.0082	0.0079	0.0077	0.0075
10%	0.0024	0.0024	0.0024	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023

Stepping up intervention

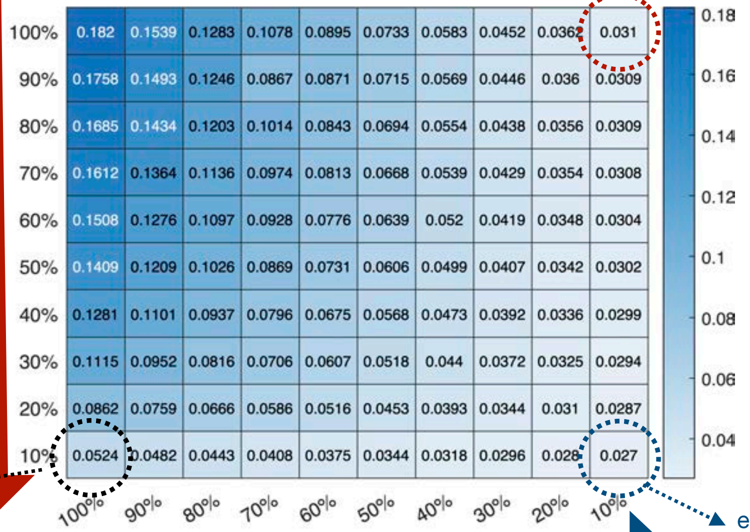
Inherent infection rates

Maintaining personal hygiene, mask wearing, level of vigilance

0.23% eventual infected population

Inherent rates more important for USA

USA



	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%
100%	0.182	0.1539	0.1283	0.1078	0.0895	0.0733	0.0583	0.0452	0.0362	0.031
90%	0.1758	0.1493	0.1246	0.0867	0.0871	0.0715	0.0569	0.0446	0.036	0.0309
80%	0.1685	0.1434	0.1203	0.1014	0.0843	0.0694	0.0554	0.0438	0.0356	0.0309
70%	0.1612	0.1364	0.1136	0.0974	0.0813	0.0668	0.0539	0.0429	0.0354	0.0308
60%	0.1508	0.1276	0.1097	0.0928	0.0776	0.0639	0.052	0.0419	0.0348	0.0304
50%	0.1409	0.1209	0.1026	0.0869	0.0731	0.0606	0.0499	0.0407	0.0342	0.0302
40%	0.1281	0.1101	0.0937	0.0796	0.0675	0.0568	0.0473	0.0392	0.0336	0.0299
30%	0.1115	0.0952	0.0816	0.0706	0.0607	0.0518	0.044	0.0372	0.0325	0.0294
20%	0.0862	0.0759	0.0666	0.0586	0.0516	0.0453	0.0393	0.0344	0.031	0.0287
10%	0.0524	0.0482	0.0443	0.0408	0.0375	0.0344	0.0318	0.0296	0.028	0.027

Stepping up intervention

Inherent infection rates

Maintaining personal hygiene, mask wearing, level of vigilance

2.7% eventual infected population with drastic measures

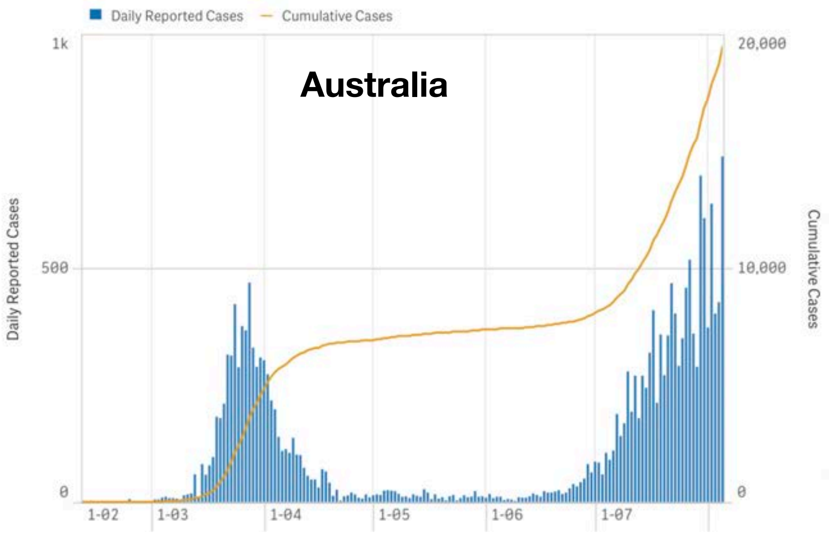
Active intervention more effective for Japan

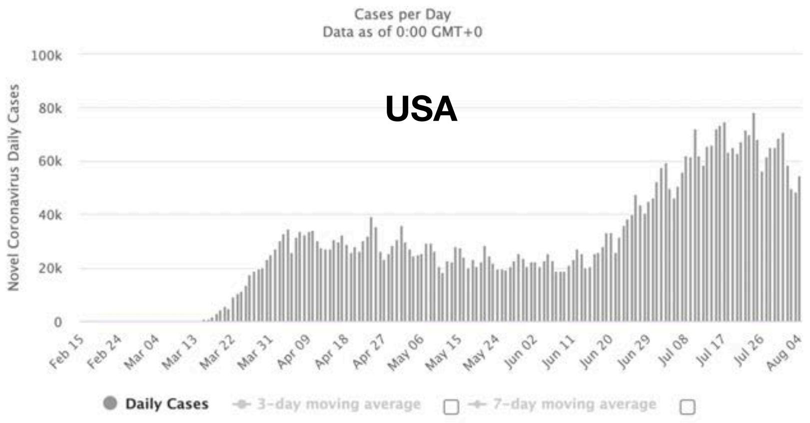
Trends

Good News:

The **effectiveness of medical treatment** is expected to improve and the recovery rate will increase in time, making the decline rate more rapid.

Predictions of first waves are highly consistent so far,





Bad News

Borders remain practically open, and intercity travel is still active, though small. Imported cases will persist. Matter of time!!!!

Asymptomatic and presymptomatic transmissions mean that "unidentified" patients are the main infectious hosts.

Conclusions

- Single population:
  - Gatherings of people - superspreader events
- Multiple populations:
  - Intercity migration
- Value of model prediction:
  - Provides essential information on the expected severity and duration of the epidemic for illuminating social and non-pharmaceutical preventive interventions.

Drastically increase TESTING CAPACITY so as to identify the “invisible” infectious hosts.

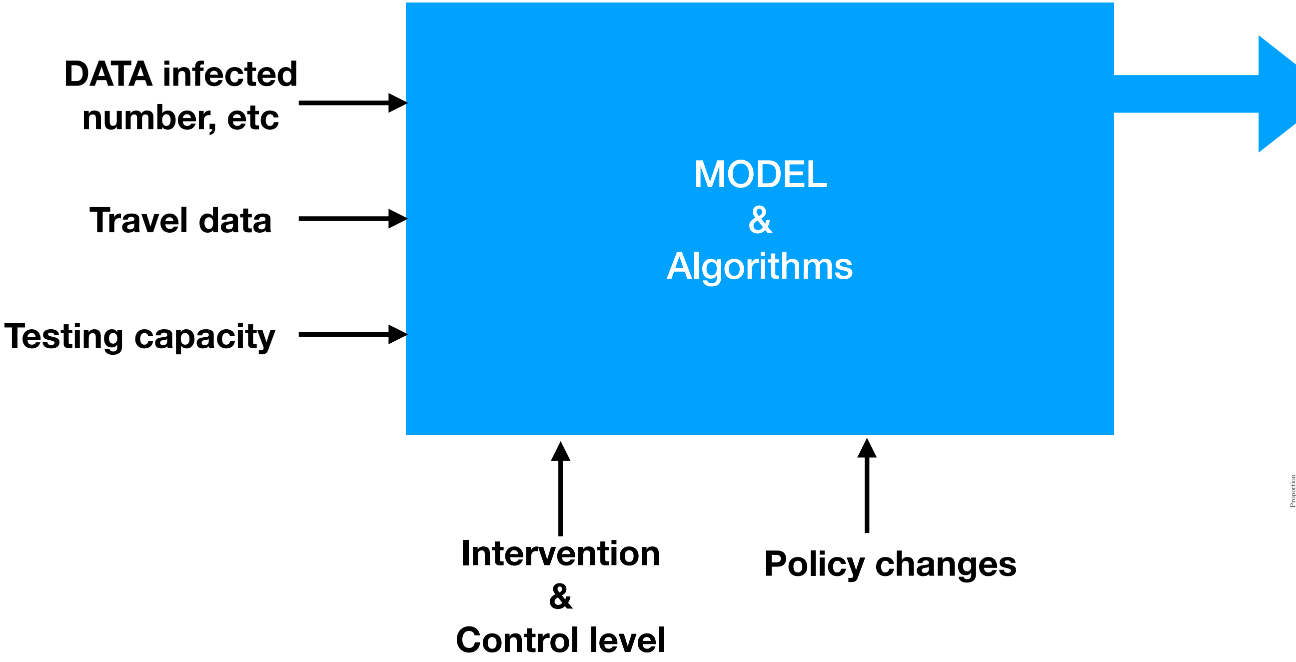
Our latest study shows that if testing capacity is raised 10-fold, the number of infected cases will reduce by 50%.

Right now, the confirmed cases are only 23% of the total infected cases.

Mass testing may be the way to go.

Further Development

- Computerised System for ANALYSIS and PREDICTON



DATA infected number, etc

Travel data

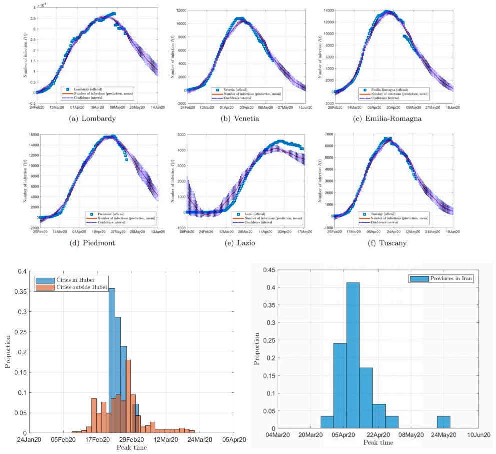
Testing capacity

MODEL
&
Algorithms

Intervention
&
Control level

Policy changes

PREDICTION



References

ON COVID-19

- C. Zhan, C. K. Tse, Y. Fu, Z. Lai, and H. Zhang, “Modeling and Prediction of the 2019 Coronavirus Disease Spreading in China Incorporating Human Migration Data,” PLOS ONE, 2020. doi: https://doi.org/10.1101/2020.02.18.20024570 [completed February 18, 2020]

- C. Zhan, C. K. Tse, Z. Lai, T. Hao, and J. Su, “Prediction of COVID-19 Spreading Profiles in South Korea, Italy and Iran by Data-Driven Coding,” PLOS ONE, July 2020. doi: https://doi.org/10.1101/2020.03.08.20032847 [completed March 8, 2020; revised on May 17, 2020]

- C. Zhan, C. K. Tse, Z. Lai, X. Chen, and M. Mo, “General Model for COVID-19 Spreading with Consideration of Intercity Migration, Insufficient Testing and Active Intervention: Modeling Study of Pandemic Progression in Japan and USA,” JMIR Public Health and Surveillance, July 2020. doi: 10.2196/preprints.18880 URL: https://preprints.jmir.org/preprint/18880 [completed March 25, 2020]

OUR WEBSITE

https://www.ee.cityu.edu.hk/~chitse/COVID-19/

COVER IMAGE by Michael Tse based on https://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2020/february/20200218_china_covid19

MODELLING AND PREDICTION OF SARS-CoV-2 SPREADING
With Human Migration Data and Data-Driven Coding

Send all inquiries to

Prof Michael Tse
Dept. of Electrical Engineering
City University of Hong Kong
Email: chitse@cityu.edu.hk
cktse@ieee.org
Phone: 3442 2619

Our Team

Prof Choujun Zhan
South China Normal University

Prof C K Michael Tse
City University of Hong Kong

Student Team:
Xiaoyun Chen
Yuxia Fu
Tianyong Hao
Zhikang Lai
Mingshan Mo
Haijun Zhang

Research funded by

National Science Foundation of China Project 61703355

Guangdong Youth University Innovative Talents Project 2016KQNCX223

City University of Hong Kong under Special Fund 9380114

Background

The Coronavirus Disease 2019 (SARS-CoV-2 or COVID-19) began to spread since mid December 2019 from Wuhan, which has been widely regarded as the epicentre of the epidemic, to almost all provinces throughout China and 80+ other countries. The SARS-CoV-2 outbreak began to occur and escalate in a special holiday period in China (about 20 days surrounding the Lunar New Year), during which a huge volume of intercity travel took place, resulting in outbreaks in multiple regions connected by an active transportation network. Thus, in order to understand the SARS-CoV-2 spreading process in China, it is essential to examine the human migration dynamics, especially between the epicentre Wuhan and other Chinese cities.

Baidu Migration Data of Inflow and Outflow Migration Strengths of Wuhan

Model and Prediction for China (February 18, 2020)

We utilise the human migration data collected from Baidu Migration, which is a mobile-app based human migration tracking data system providing historical indicative daily volume of travellers to/from and between 367 cities in China. We combine, in this study, intercity travel data collected from Baidu Migration with the SEIR model to build a new dynamic model for the spreading of COVID-19 in China. Using official historical data of infected, recovered and death cases in 367 cities, we perform fitting of the data to estimate the best set of model parameters, which are then used to estimate the number of individuals exposed to the virus in each city and to predict the extent of spreading in the coming months. Our study shows that provided such migration control and other stringent measures continue to be in place, the number of infected cases in various Chinese cities will peak between mid February to early March 2020, with about 0.8%, less than 0.1% and less than 0.01% of the population eventually infected in Wuhan, Hubei Province, and the rest of China, respectively, and no new cases to be expected from mid March. Moreover, for most cities in and outside Hubei Province (except Wuhan), the total number of infected individuals will be less than 4000 and 300, respectively. Since January 24, 2020, very strict migration control has been imposed in

Thank you!