



传染病研究领域的时空数据与AI

石宏濤

计算机学院，北京航空航天大学，北京，中国



目录



01 ▶ 总体介绍

02 ▶ 时空数据在传染病建模中扮演的角色

03 ▶ 应用：武汉疫情预测仿真

石泓浩，北航计算机学院，计算机科学与技术专业

主要科创成果：

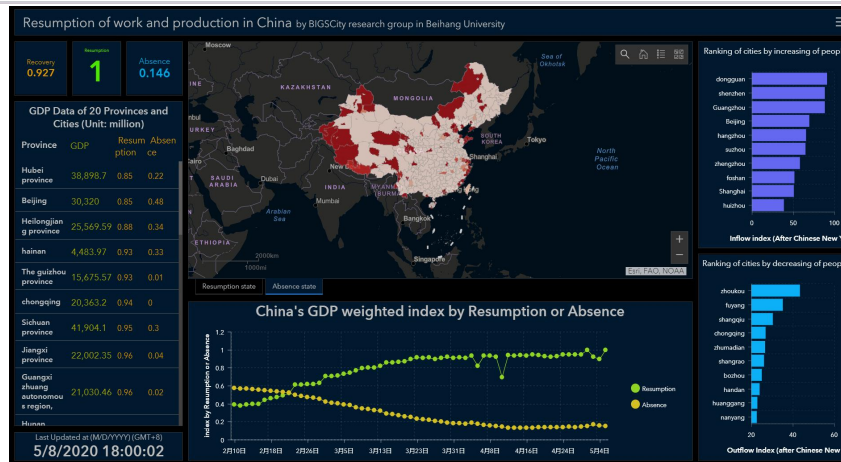
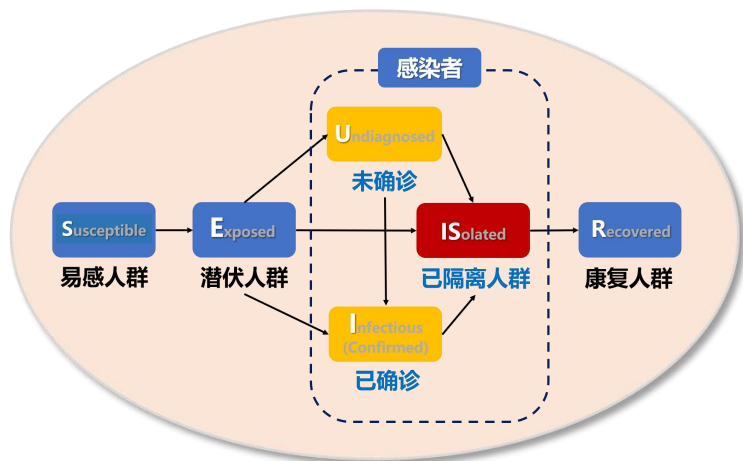
围绕大数据的表征与应用，传染病系统模型设计展开。涵盖预测、研判、反事实推断、最优控制，物理知识与AI混合的生成模型等；

发表学术论文5篇。其中2篇第一作者，1篇学生第一作者；

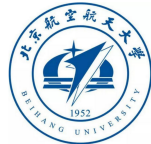
获第17届挑战杯全国大学生课外学术科技作品竞赛全国特等奖（第一作者），

及挑战杯，互联网+省部级奖项若干；

申请4项专利（均为第一作者），完成软著两项。

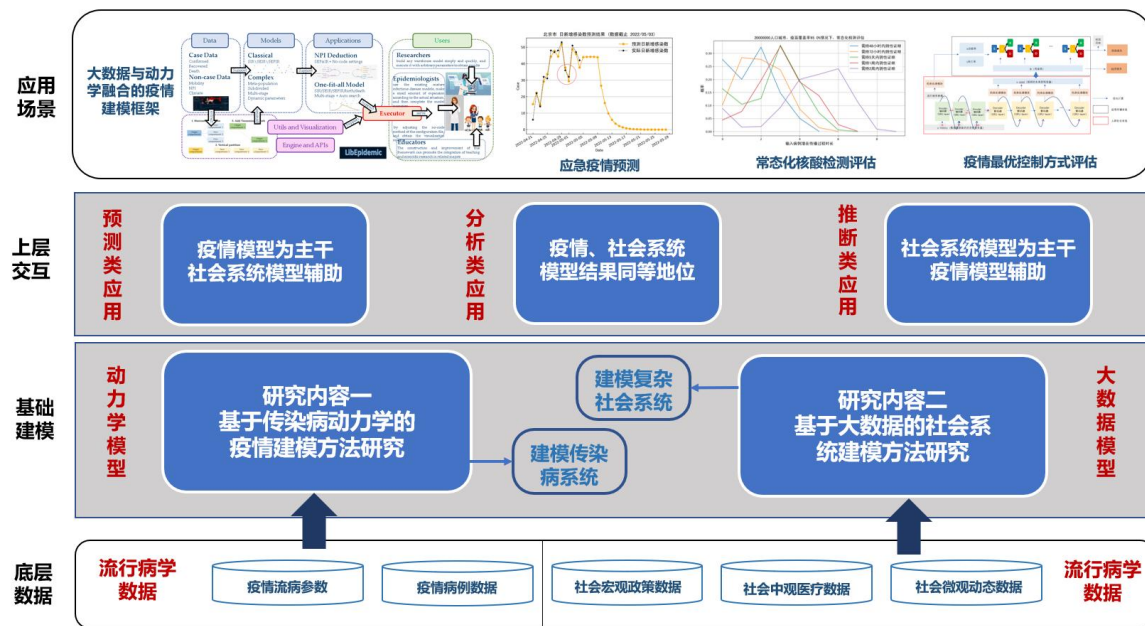
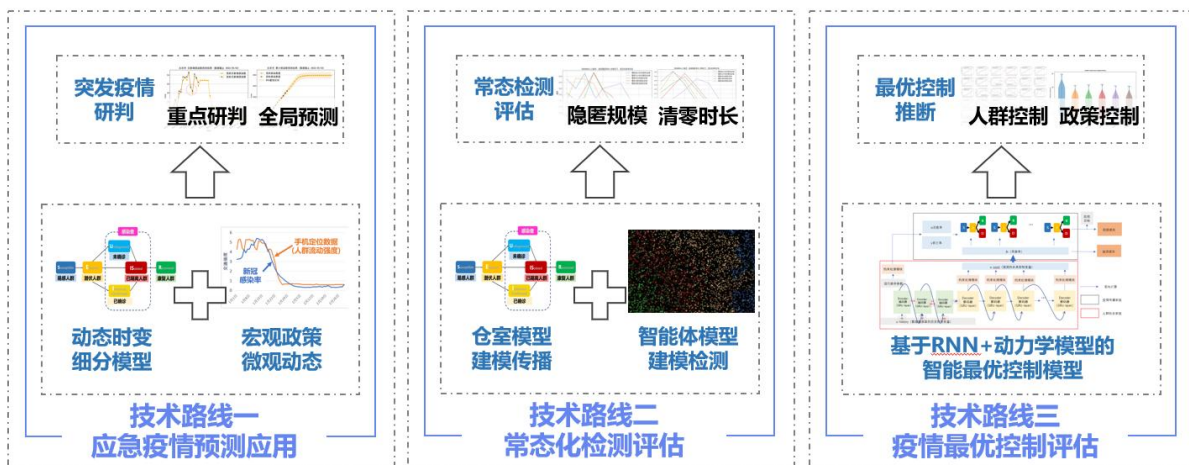


传染病研究领域的时空数据



在数据量不充足的时代，研究者使用一个加权平均的统计值来作为单一的参数可以达到优秀的模型性能。但**数据获取和应用手段的提升**，令仓室模型解决这些局限性成为了可能。**仓室细分和动态时变模型**即试图在仓室模型框架内部解决这一局限性，**智能体模型和混合AI模型**试图直接构建不含有这类局限性，却仍可以完成仓室模型建模目的的新模型。

这些改进通常对应着更复杂、高维度的参数。这些参数无法像SEIR等中的参数一样简单算出，而是需要对多源大数据表征，最终学习出适配的参数。



目录

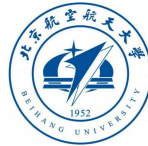


01 ▶ 领域总体介绍

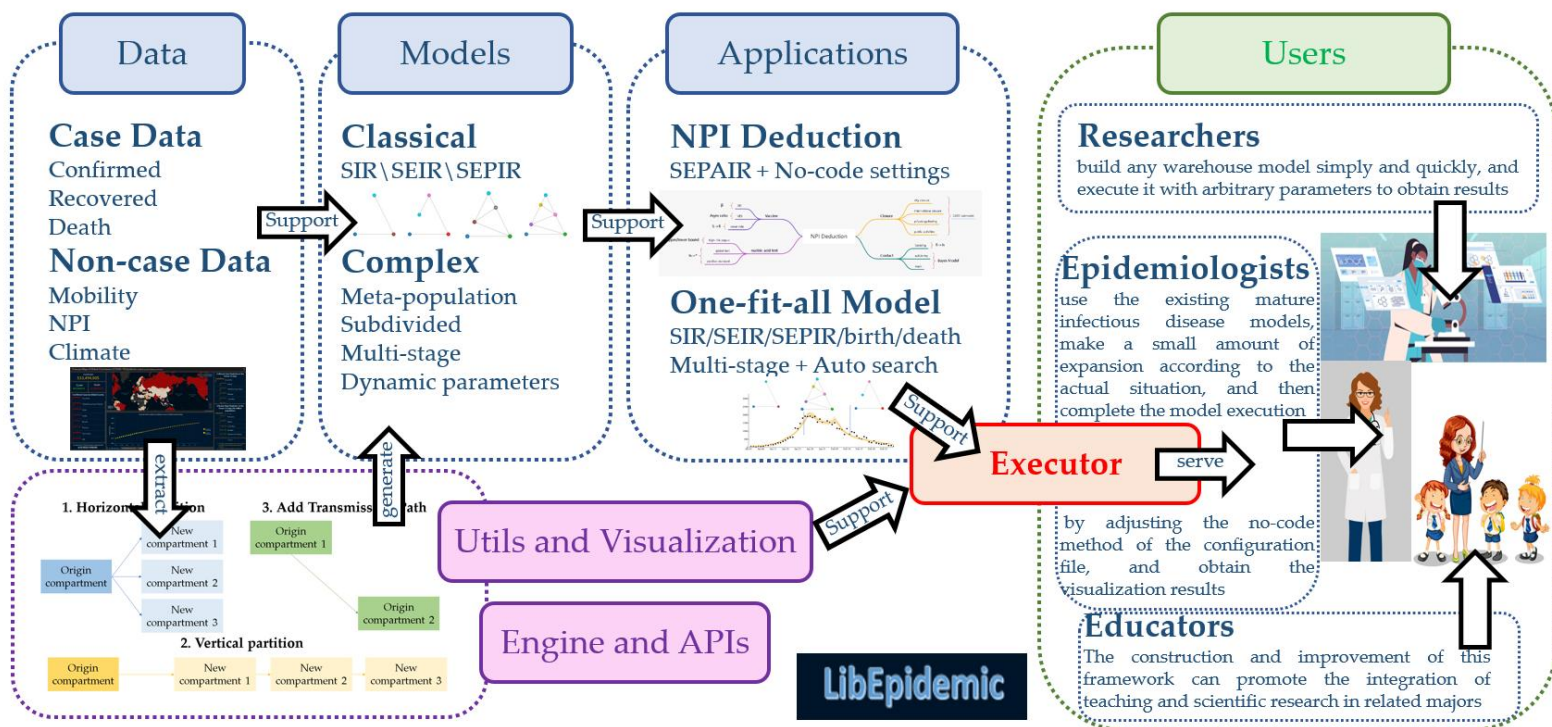
02 ▶ 时空数据在传染病建模
中扮演的角色

03 ▶ 应用：武汉疫情预测仿真

有害微生物传染病建模 – 传统传染病建模



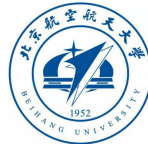
在数据量不充足的时代，研究者使用一个加权平均的统计值来作为单一的参数可以达到优秀的模型性能。但**数据获取和应用手段的提升**，令仓室模型解决这些局限性成为了可能。这些改进通常对应着更复杂、高维度的参数。这些参数无法像SEIR等中的参数一样简单算出，而是需要对多源大数据表征，最终学习出适配的参数。这一过程将被从建模传染病系统本身中被剥离出来，形成针对社会系统的建模。



H. Shi, J. Wang, J. Cheng, X. Qi, H. Ji, C. J. Struchiner, D. A. M. Villela, E. V. Karamov, and A. S. Turgiev, "Big data technology in infectious diseases modeling, simulation and prediction after the COVID-19 outbreak: a survey," *Intelligent Medicine*, 2023.

H. Shi, Q. Tian, J. Wang, and J. Cheng, "LibEpidemic: an open-source framework for modeling infectious disease with bigdata," in *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM'22)*, 2022, pp. 4980-4984.

时空数据在传染病建模中扮演的角色



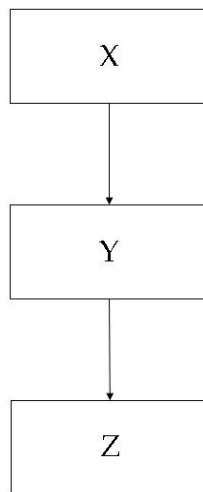
仓室细分通过利用更精细化的数据，来将现有仓室模型的仓室细分为多个。
按照细分形式可以分为横向仓室细分模型、纵向仓室细分模型和多种群模型。

$$\frac{dX}{dt} = -v_{xy}$$

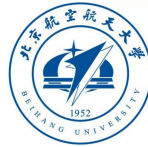
$$\frac{dY}{dt} = v_{xy} - v_{yz}$$

$$\frac{dZ}{dt} = v_{yz}$$

$$X + Y + Z \equiv N$$



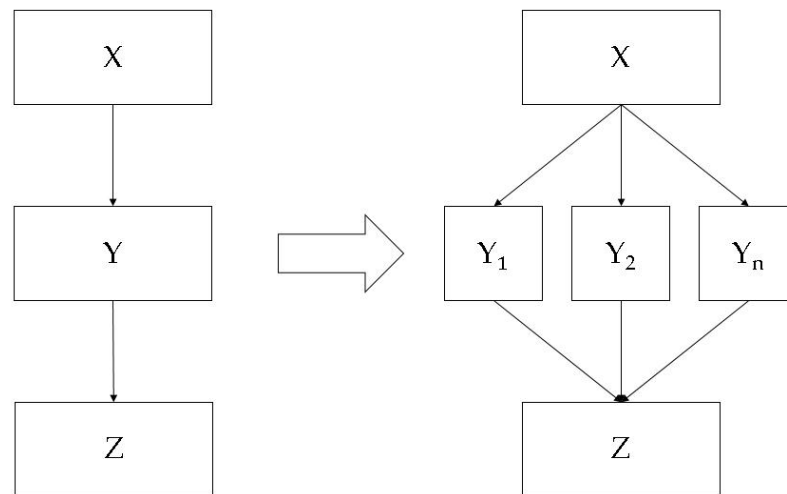
时空数据在传染病建模中扮演的角色



横向仓室细分将一个原有的仓室拆分为 n 个新仓室，新仓室之间没有转化关系。

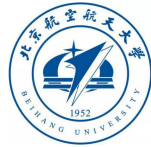
原有仓室的前序仓室成为每一个新仓室的前序仓室，原有仓室的后续仓室成为每一个新仓室的后续仓室。

$$\begin{aligned}\frac{dX}{dt} &= -v_{xy} \\ \frac{dY}{dt} &= v_{xy} - v_{yz} \\ \frac{dZ}{dt} &= v_{yz} \\ X + Y + Z &\equiv N\end{aligned}$$



$$\begin{aligned}\frac{dX}{dt} &= -\sum_{i=1}^n v_{xy_i} \\ \frac{dY_i}{dt} &= v_{xy_i} - v_{y_i z} \\ \frac{dZ}{dt} &= \sum_{i=1}^n v_{y_i z} \\ X + \sum_{i=1}^n Y_i + Z &\equiv N\end{aligned}$$

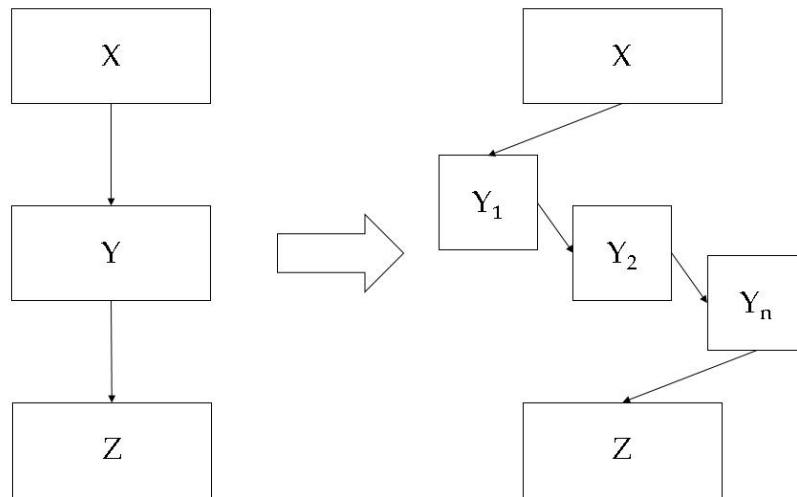
时空数据在传染病建模中扮演的角色



纵向仓室细分将一个原有的仓室拆分成 n 个新仓室，新仓室之间依次首位相连形成一个链。

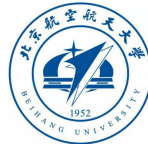
原有仓室的前序仓室成为第一个新仓室（链头）的前序仓室，原有仓室的后续仓室成为最后一个新仓室（链尾）的后续仓室。

$$\begin{aligned}\frac{dX}{dt} &= -v_{xy} \\ \frac{dY}{dt} &= v_{xy} - v_{yz} \\ \frac{dZ}{dt} &= v_{yz} \\ X + Y + Z &\equiv N\end{aligned}$$



$$\begin{aligned}\frac{dX}{dt} &= -v_{xy_1} \\ \frac{dY_1}{dt} &= v_{xy_1} - v_{y_1y_2} \\ \frac{dY_i}{dt} &= v_{y_{i-1}y_i} - v_{y_iy_{i+1}}, \text{ where } 1 < i < n \\ \frac{dY_n}{dt} &= v_{y_{n-1}y_n} - v_{y_nz} \\ \frac{dZ}{dt} &= v_{y_nz} \\ X + \sum_{i=1}^n Y_i + Z &\equiv N\end{aligned}$$

时空数据在传染病建模中扮演的角色

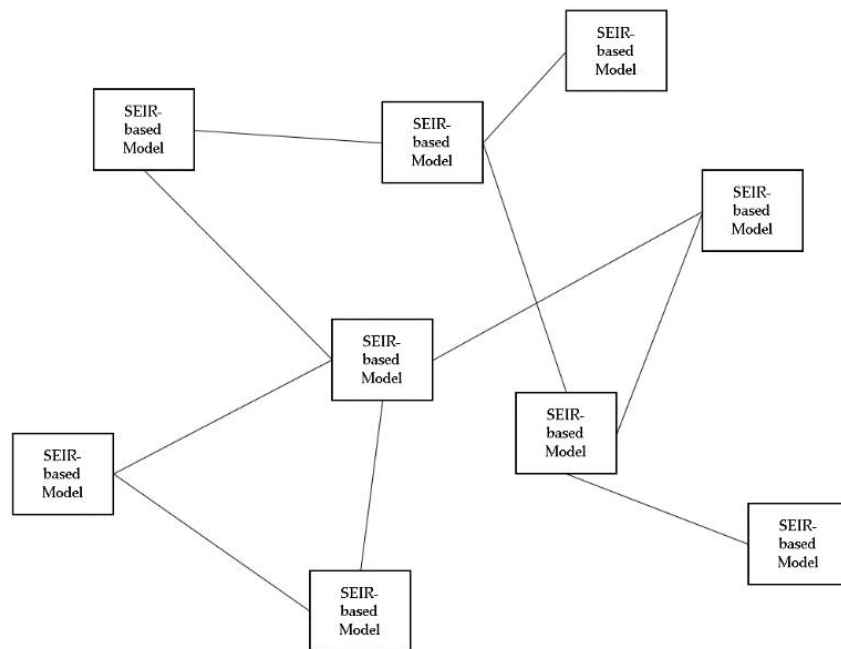


多种群模型通常基于空间特征和社交网络展开，而社交网络所代表的社交距离和接触矩阵是控制有害微生物事件的最关键因素。

多种群模型将社交网络上的每个节点视为一个种群，并认为每个种群间的人群流动是均匀的。

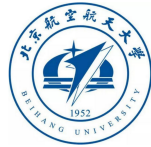
多种群模型利用了人口流动数据，在空间特征和社交网络建模的角度上超越了仓室细分模型。

$$\begin{aligned}\frac{dX}{dt} &= -v_{xy} \\ \frac{dY}{dt} &= v_{xy} - v_{yz} \\ \frac{dZ}{dt} &= v_{yz} \\ X + Y + Z &\equiv N\end{aligned}$$



$$\begin{aligned}\frac{dS}{dt} &= -\frac{\text{multiply}(S, \beta \cdot I)}{N} \\ S &= (S_1, S_2, \dots, S_n)^T \\ \beta &= (\beta_{ij})_{n \times n} \\ I &= (I_1, I_2, \dots, I_n)^T \\ \text{where } \leftarrow & \\ x &= (x_1, x_2, \dots, x_n)^T \\ y &= (y_1, y_2, \dots, y_n)^T \\ \text{multiply}(x, y) &= (x_1 \cdot y_1, x_2 \cdot y_2, \dots, x_n \cdot y_n)^T\end{aligned}$$

时空数据在传染病建模中扮演的角色



动态时变模型通过承接获取的与时间有关的数据，通过随时间更改模型设置或更改参数值来时间动态性建模。

动态时变模型按照时变的尺度可以分为**多阶段模型**和**时变参数模型**。

多阶段模型设计更复杂，需要更多的数据甚至逻辑推演去支撑，但建模能力也更强。

时变参数模型在牺牲了一定的时变灵活性的条件下，仅需更少的数据即可建模动态性。

参数时变前:↵

$$\mathbf{X}(t) = \text{Epi}(\mathbf{X}(0), \mathbf{u}, t)$$

例如在SEIR中:↵

$$\mathbf{X}(t) = (S(t), E(t), I(t), R(t))^T$$

$\text{Epi}()$ 是SEIR模型对应的微分方程组↵

$$\mathbf{u} = (\beta, \alpha, \gamma)^T$$

参数时变后:↵

$$\mathbf{X}(t) = \text{Epi}(\mathbf{X}(0), \mathbf{u}(t), t)$$

多阶段模型后: ↵

$$\mathbf{X}(t_n) = \text{Epi}(\mathbf{X}(t_{n-1}), \mathbf{u}(t_n), t)$$

$$\mathbf{u}(t_n) \text{ may related to } \mathbf{X}(t_{n-1})$$

$$t_n > t_{n-1} > \dots > t_0$$

目录



01 ▶ 领域总体介绍

02 ▶ 时空数据在传染病建模中扮演的角色

03 ▶ 应用：武汉疫情预测仿真

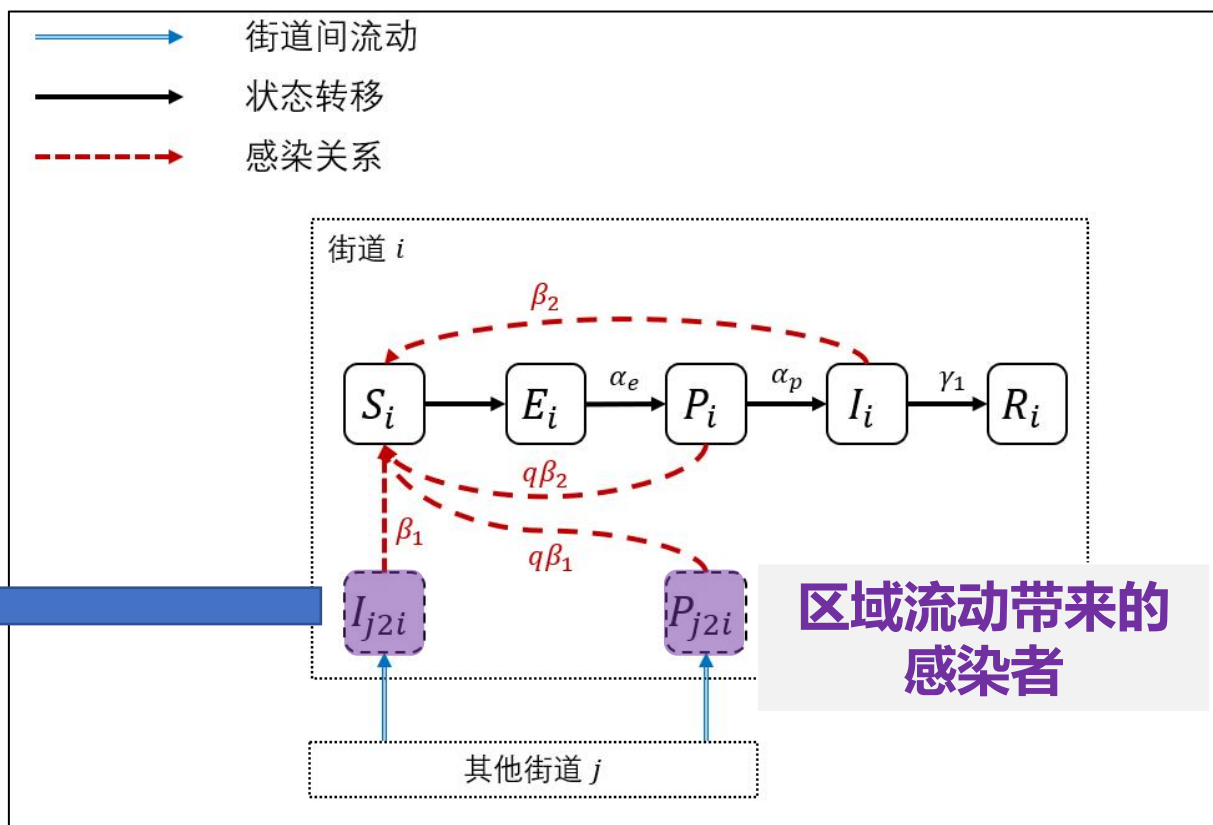
重点：建模城市内部区域间的流动关系

J. Wang, H. Shi, J. Ji, X. Lin, and H. Tian, "High-resolution data on human behavior for effective COVID-19 policy-making—Wuhan city, Hubei province, China, January 1–February 29, 2020," *China CDC Weekly*, 2023.

$$\begin{aligned}\frac{dS_i}{dt} &= -\beta_2 \frac{S_i}{N_i} I_i - \beta_1 \frac{S_i}{N_i} \sum_{j \neq i} I_{j2i} - q\beta_2 \frac{S_i}{N_i} P_i - q\beta_1 \frac{S_i}{N_i} \sum_{j \neq i} P_{j2i}, \\ \frac{dE_i}{dt} &= \beta_2 \frac{S_i}{N_i} I_i + \beta_1 \frac{S_i}{N_i} \sum_{j \neq i} I_{j2i} - q\beta_2 \frac{S_i}{N_i} P_i - q\beta_1 \frac{S_i}{N_i} \sum_{j \neq i} P_{j2i} - \gamma_e E_i, \\ \frac{dP_i}{dt} &= \gamma_e E_i - \gamma_p P_i, \\ \frac{dI_i}{dt} &= \gamma_p P_i - \sigma_t I_i, \\ \frac{dR_i}{dt} &= \sigma_t I_i.\end{aligned}$$

区域间感染

区域内部感染
(与单种群模型相同)

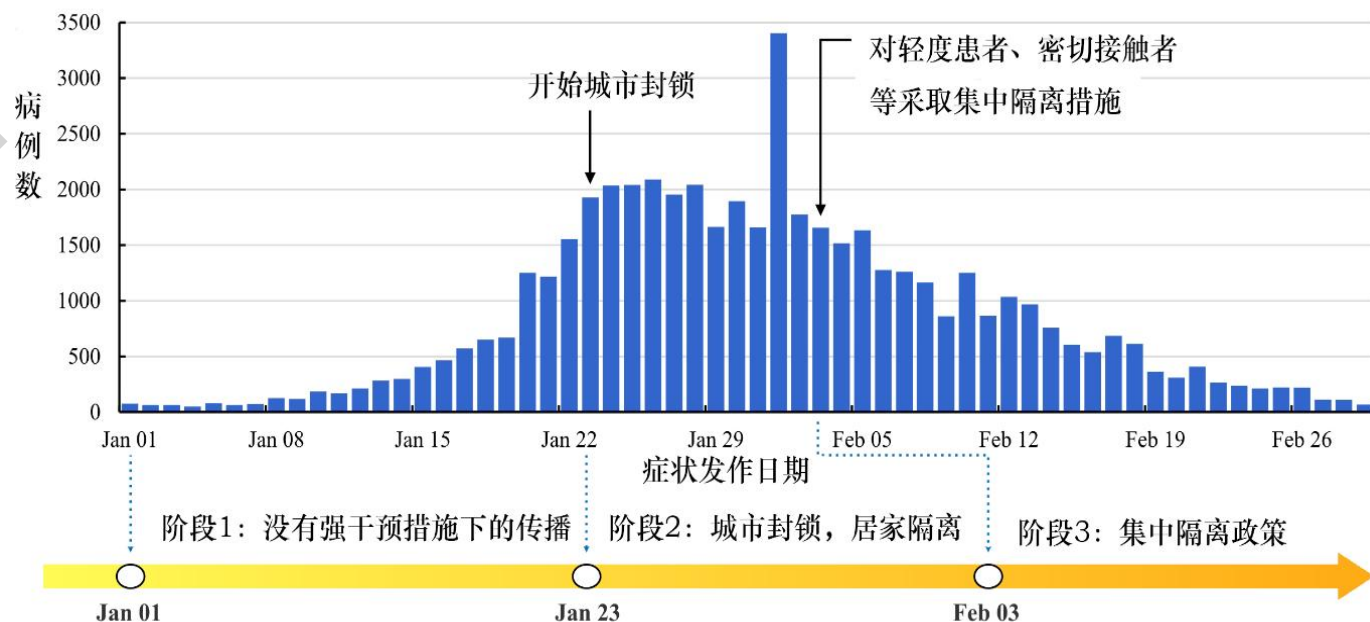


任务说明

- 使用复合种群SEPIR模型拟合历史疫情数据
- 设计多组场景，评估疫情防控措施的效果
 - 限制人口流动
 - 缩短传染期

实验数据

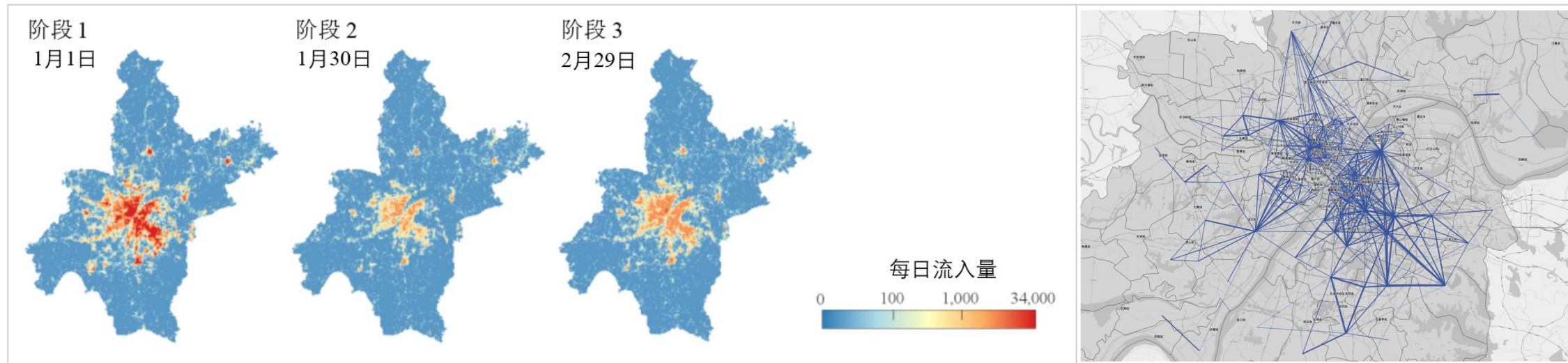
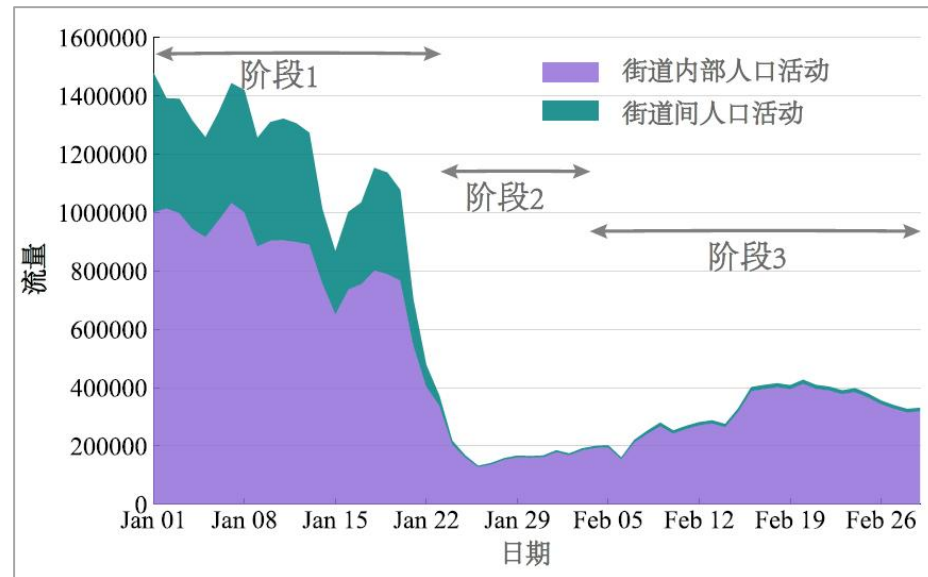
- 病例数据
 - 2020年1月1日至2月29日期间武汉市病例数据，包含发病时间、诊断时间等字段
 - 根据疫情防控措施实施情况将模型拟合分为三个阶段，以1月23日、2月3日为分界点



实验数据

• 人口活动数据

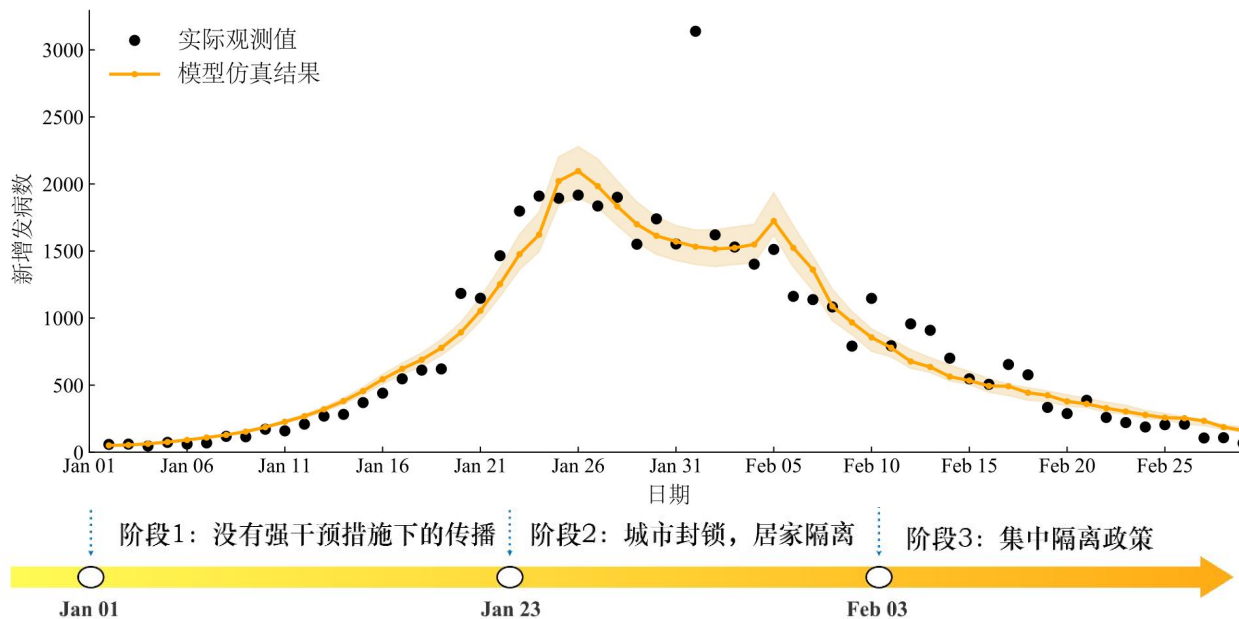
- 使用手机信令数据作为度量
- 时间：2020年1月1日至2月29日
- 范围：武汉全域，覆盖**50%**以上的人口（约**582万人**）
- 精度：**500米** × **500米**网格级别，24小时
- 以街道办为尺度重新整理



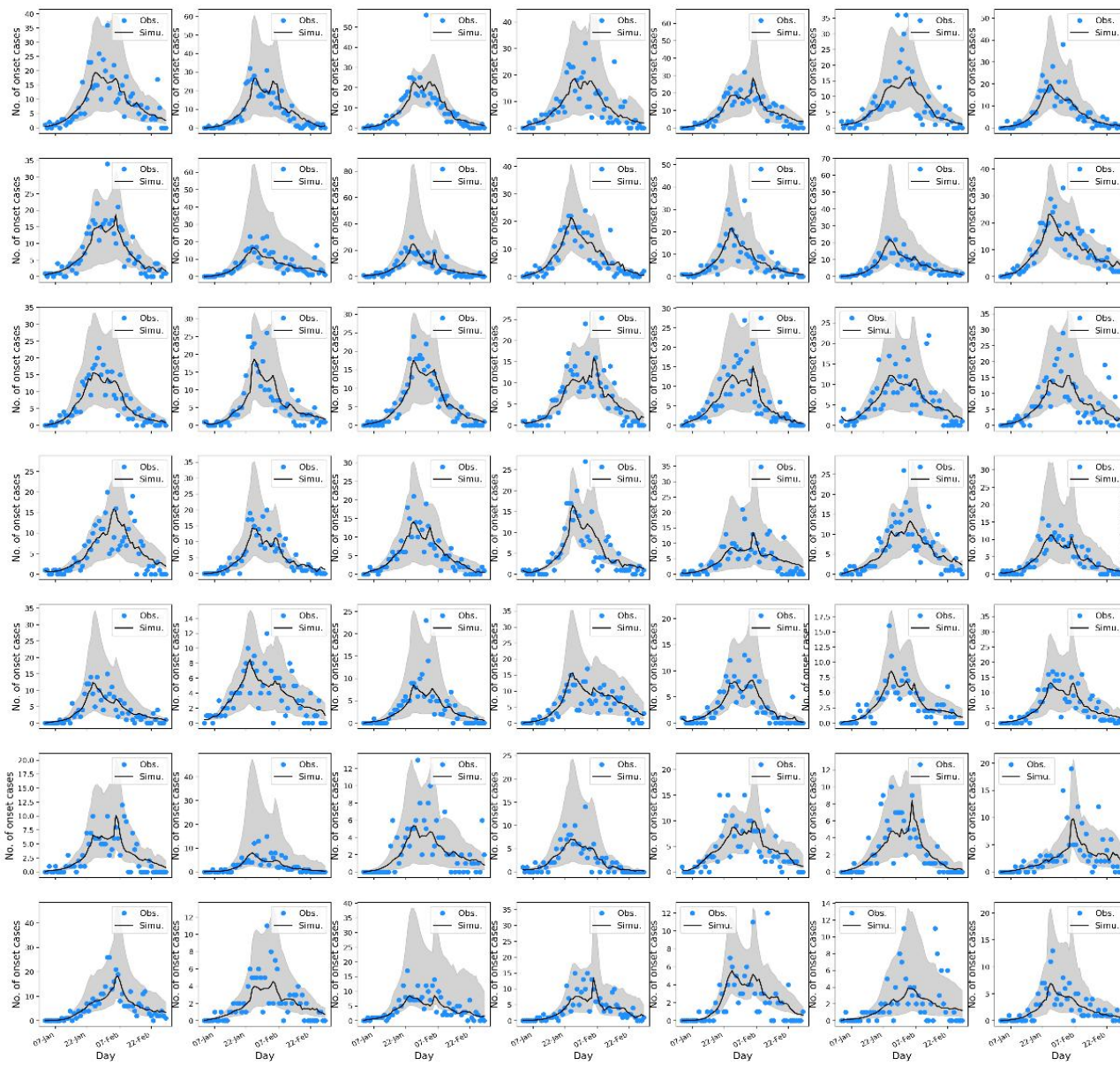
拟合结果

- 时间范围：2020年1月1日 – 2月29日
- 99个街道办、三个阶段分别进行拟合
- 使用MCMC各采样10000次
- 平均RMSE=5.07

武汉全市拟合结果



武汉部分街道办拟合结果

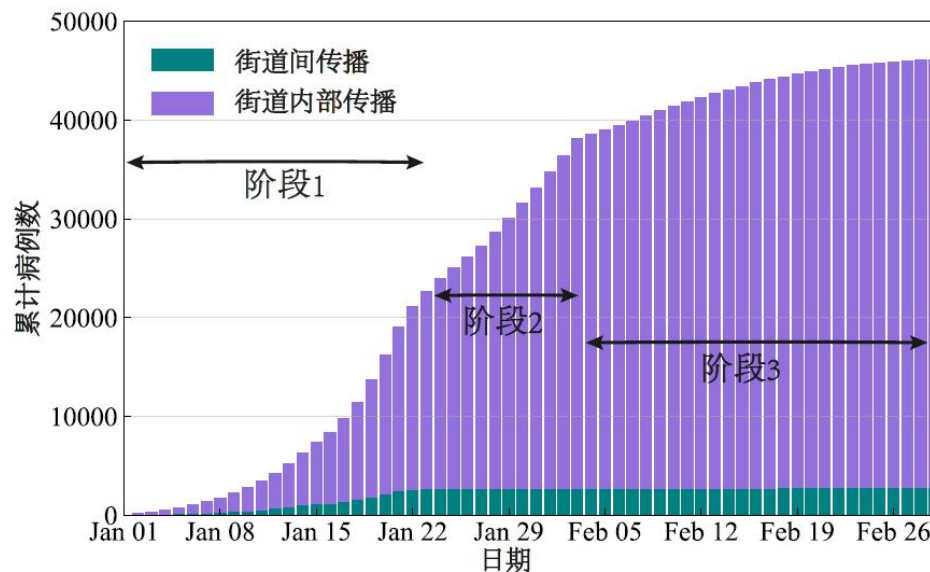


城市内部活动对疫情影响分析

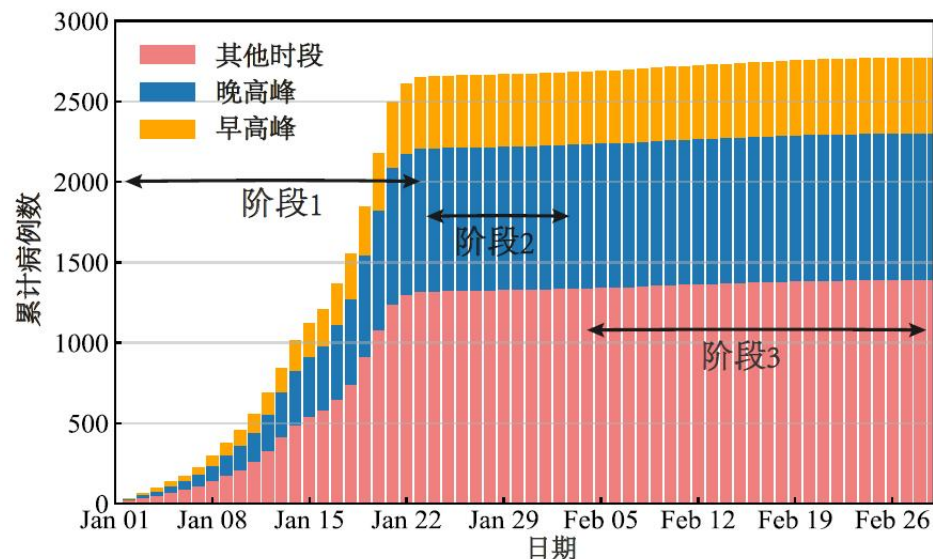
- 街道内部传播造成的病例数约为街道间传播造成病例数的**7.6倍**
- 街道间传播：**封城后停止**
- 街道内部传播：**实施集中收治后停止**
- 街道间传播：病例数占比/时段流量占比
 - 晚高峰时段(7-9时)：**1.35**
 - 早高峰时段(16-18时)：**0.90**
 - 其他时段：**0.88**

晚高峰时段传播带来的影响更为显著

街道内部传播和街道间传播的影响

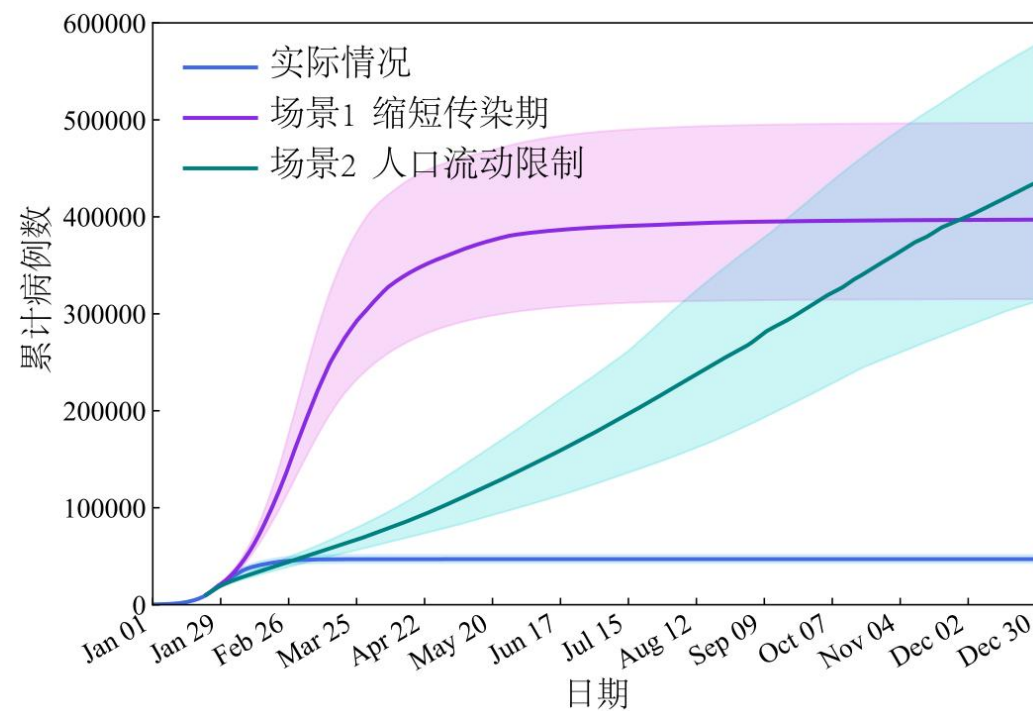
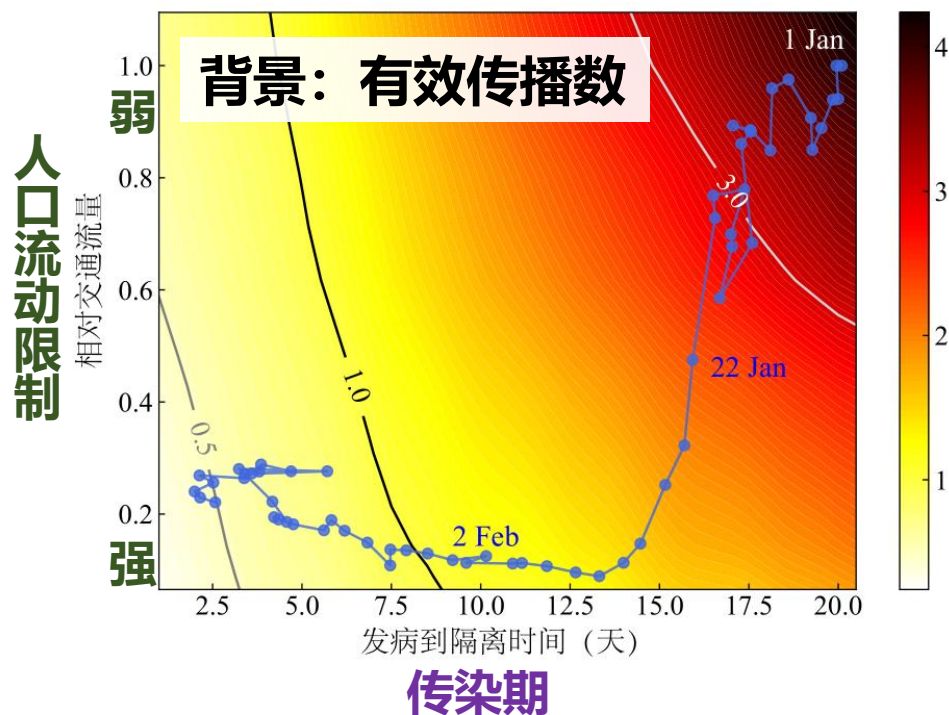


街道间传播中早高峰、晚高峰和随机流动时段的影响



武汉市疫情防控措施评估

- 研究人口流动限制（相对交通流量）、缩短传染期（发病到隔离时间）对疫情发展的影响



- 首先通过**限制人口流动**，在短期内遏制疫情蔓延，为进一步行动**争取了时间**；
- 后期推进集中收治等措施**缩短传染期**，达到疫情的**完全控制**。



谢谢

石泓浩

计算机学院，北京航空航天大学，北京，中国

shihonghao@buaa.edu.cn

<https://github.com/shh2000>