

DOI: 10.13718/j.cnki.xdzk.2022.02.024

基于 SEIRD-GEOCA 的 COVID-19 疫情 时空分布模型构建与情景模拟

荆磊^{1,2}, 刘明皓^{1,2}, 陈春^{3,4}, 曹逸凡^{1,2}, 刘天林^{1,2}

1. 重庆邮电大学 计算机科学与技术学院, 重庆 400065; 2. 重庆邮电大学 空间信息研究中心, 重庆 400065;
3. 重庆交通大学 建筑与城市规划学院, 重庆 400074; 4. 重庆交通大学 生态人居与绿色交通研究中心, 重庆 400074

摘要: 当前如何通过模型的构建来揭示新型冠状病毒肺炎(Corona Virus Disease 2019, COVID-19)的传播机制, 提高对疫情传播的时空预测预警能力仍面临很大挑战. 为充分发挥 SEIRD 模型和 CA 模型各自在时间和空间仿真方面的优势, 依据空间现象的地理相似性原理, 构建了一种基于 SEIRD-GEOCA 的传染病时空耦合模型, 并在模型中融入了蔓延扩散和随机迁移扩散策略, 采用重庆市 2020 年 1 月 21 日到 2020 年 2 月 29 日的感染者轨迹数据和相关驱动因子数据实现了对重庆市 COVID-19 感染者时空分布的情景模拟. 情景模拟结果显示, 在空间上将呈现出爆发式扩散和局域扩散的明显差别; 且实施不同强度的干预, COVID-19 呈现出了不同的时空扩散效应. 研究表明, SEIRD-GEOCA 模型能揭示传染病的传播机制, 并能较好地模拟 COVID-19 疫情的时空扩散规律; 面对 COVID-19 疫情, 政府是否采取干预措施以及实施干预的强度都将对疫情的时空扩散产生较大影响.

关键词: 新型冠状病毒肺炎; SEIRD-GEOCA 模型; 时空扩散;
动态模拟

中图分类号: R181.8+1; TP391 文献标志码: A 开放科学(资源服务)标识码(OSID):
文章编号: 1673-9868(2022)02-0207-12



Spatio-temporal Distribution Model Construction and Scenario Simulation of COVID-19 Epidemic Using SEIRD-GEOCA

JING Lei^{1,2}, LIU Minghao^{1,2}, CHEN Chun^{3,4},
CAO Yifan^{1,2}, LIU Tianlin^{1,2}

1. School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
2. Spatial Information Research Center, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
3. School of Architecture and Urban planning, Chongqing Jiaotong University, Chongqing 400074, China;
4. Ecological Habitat and Green Transportation Research Center, Chongqing Jiaotong University, Chongqing 400074, China

Abstract: At present, how to construct the models to reveal the transmission mechanism of Corona Virus

收稿日期: 2021-04-08
基金项目: 国家自然科学基金资助项目(42071218); 重庆市自然科学基金资助项目(cstc2019jcyj-msxmX0139).
作者简介: 荆磊, 硕士研究生, 主要从事新冠疫情期间空大数据研究.
通信作者: 刘明皓, 教授, 主要从事城市动态模拟与地理智能计算方面的研究.

Disease 2019 (COVID-19), and improve the spatio-temporal prediction and early warning capabilities of the epidemic spread is still facing great challenges. In order to fully use the respective advantages of the SEIRD model and the CA model in temporal and spatial simulation, based on the principle of geographic similarity of spatial phenomena, a SEIRD-GEOCA based infectious disease spatio-temporal coupled model was constructed, and strategies of contact diffusion and random migration were incorporated into the model. Using the data of the infected persons in Chongqing from January 21, 2020 to February 28, 2020 and related driving factor data, the scenario simulation of the temporal and spatial distribution of COVID-19 infected persons in Chongqing was realized. The scenario simulation results show that obvious differences in the spatial distribution between outbreak and local spread. In addition, the implementation of different intervention intensities showed different spatio-temporal diffusion effects. Studies have shown that the SEIRD-GEOCA model can reveal the transmission mechanism of infectious diseases and better simulate the spatio-temporal spread of COVID-19. In the face of the epidemic, whether the government takes intervention measures and the intensity of the intervention will affect the spatio-temporal spread of the epidemic.

Key words: Corona Virus Disease 2019; SEIRD-GEOCA model; spatio-temporal diffusion; dynamic simulation

新型冠状病毒肺炎(Corona Virus Disease 2019, COVID-19)的快速传播,对全球政治、经济和社会造成了严重危害^[1]. 传染病的传播是个复杂的时空扩散过程,对传染病扩散建模有助于理解传染病的传播机理与内在规律^[2]. 学术界针对 COVID-19 传播进行了大量的模拟研究,但如何通过模型的构建揭示传染病的传播机制,提高对疫情传播的时空预测预警能力仍面临很大挑战. 目前数量预测模型主要有传染病动力学模型,这类模型根据人群的 S, E, I, R 状态分为多个仓室[S, E, I, R 分别表示易感者(susceptible)、接触者(exposed)、感染者(infected)、康复人群(removed)],采用微分方程刻画人群数量的变化,主要用来预测疫情的拐点和规模. 依据病原体、传染性、感染后免疫、传染源、传播途径、易感人群等流行病学特征,传染病动力学模型又可分为 SI, SIS, SIR, SEIR 等基本模型,分别适用于感染者不可恢复和治愈^[3]、感染者治愈后可能反复感染^[4]、感染者具有免疫力但潜伏期短且症状强烈^[5]以及具有强免疫力但潜伏期较长^[6]这 4 种情况. 根据 COVID-19 的流行病学特征(人传人、症状前传播、呼吸道吸出物、飞沫、接触物和粪便传播、易感者与年龄或性别无关、感染后免疫等),目前揭示 COVID-19 传播机理的动力学模型研究仍然以 SEIR 模型为主^[7]. 数量预测方面,传统数学模型主要考虑参数如何准确地获取,而机器学习主要考虑驱动因子的相关性,另外,发挥不同时间序列模型优势的数量预测整合模型也相继被开发出来^[8]. 除了数量预测,揭示传染病在空间上的扩散规律与机理也是模型研究的重要方向,这类模型主要包括多智能体^[9]、CA 模型等. SEIRD-agent 模型用来模拟人类活动(如人为干预措施)对 COVID-19 疫情防控的效果^[10-11]. 元胞自动机(Cellular Automata, CA)是一种时间、空间、状态都离散,空间相互作用和时间因果关系为局部的网格动力学模型^[12-16]. 近年来,CA 与传染病预测模型相结合被广泛应用于传染病传播的模拟研究中^[17-18]. Almuzakki 等^[19]利用 SEIR-CA 模型模拟疫情在人群中的传播方式和疫苗接种策略对传播方式的影响;Cissé 等^[20]利用 SEIR-CA 模型证明邻近元胞的空间分布对疾病传播的实际影响;Medrek 等^[21]利用 SEIR-CA 模型模拟栅格与人数的对应关系,但并没有反映前一日感染者对后一日感染者的迭代过程;Gwizdaa^[22]利用 SEIR-CA 从个体行为对群体的影响出发,利用网络中邻居之间接触人数的变化对于干预措施进行分析.

归纳起来,目前对于传染病的数量预测、聚集特征与扩散机制的模型构建主要集中在以下两方面. 一是从传染病的传统模型出发,对一定统计区域的传染病人进行推估,较少考虑细粒度的空间交互,如 SEIR 模型,这类模型尽管有较好的时间分辨率,但空间分辨率很差;二是从 CA 出发,CA 模型尽可能地将研究区域划分为较小的格网来模拟传染病在不同情景下的空间扩散情况. 在一定程度上揭示了

流行机理与内在规律,但大多数研究没有将传染病的扩散与生物生存环境的地理相似性有效结合起来,不具有真正的生物地理学意义,本质上是一种数值仿真模拟.鉴于其在模拟复杂系统时强大的时空演变过程能力,通过与机器学习和人工智能等算法的结合,构建高时空分辨率的耦合模型,是传染病传播模型需进一步探讨的方向.另外,传染病的空间扩散,既存在空间上随距离衰减的连续蔓延的扩散方式,也存在从高级节点向低级节点城市渗透的空间跳跃式等级扩散方式^[23-24].传染病的扩散与蔓延是在一定时间、空间尺度上的传播过程,其时空关系相互依存、不可分割^[25-26].相关研究也发现传染病感染者不仅具有很强的时间关联性,而且感染者的空间分布与其生活的其他空间属性也具有强相关性^[27-30],因此构建时空耦合模型,并将上述空间扩散机制融入传染病扩散模型中用来揭示传染病的空间传播机制具有十分重要的理论和实践意义.

本研究①为充分发挥 SEIR 模型和 CA 模型在时间分辨率预测和空间分辨率仿真方面的优势,依据空间现象地理相似性原理,构建了一种基于 SEIRD-GEOCA 的传染病时空耦合模型;②在模型中融入接触扩散和随机迁移扩散策略,采用重庆市感染者轨迹数据和统计数据实现对重庆市 COVID-19 疫情时空分布的情景模拟和预测,揭示 COVID-19 疫情的时空扩散规律和传播机理;③实现不同感染源与干预情景下的时空扩散模拟与效应评估.

1 模型构建原理与方法

本研究构建了 SEIRD-GEOCA 时空耦合模型用于模拟 COVID-19 疫情的时空扩散规律,该模型包括数量预测模块、空间映射模块和时空耦合模块.数量预测模块在经典的 SEIR 模型基础上引入了有无封城的情景,用来表示随着时间变化新增的感染者数量;空间映射模块利用 CA 模型将感染者分配到空间栅格上,将感染源分为本地潜在感染源和外部输入感染源,加入了有无排查政策;时空耦合模块将前 2 部分进行整合,实现每日疫情地图的动态迭代.空间适宜性概率充分考虑地理相似性原理,即某种传染病病毒的传播和生存与“生物岛”的大小、岛与岛的关联性、自然环境与文化环境的相似性等密切相关.

1.1 SEIRD 模块

经典 SEIR 模型将人群分为易感人群、接触人群、感染人群和康复人群,但不能很好地表现政府干预措施的影响.该模型假设人群中所有个体都有被感染的概率,当被感染个体痊愈后,会产生抗体.考虑到防治传染病的隔离措施,在传统 SEIR 模型上,将 R 细分为 R 和 D ,并调整封城前后每人每天接触的人数,形成新的传染病动力学模型 SEIRD 模型^[7].

如图 1 中 SEIRD 模型所示, S, E, I, R, D 分别表示易感人群、接触人群、感染人群、康复人群(治愈人群)、死亡人群; α 代表每人每天接触的人数, β 代表被感染的概率, $\alpha \times \beta$ 代表感染力; σ 代表 E 转为 I 的概率; γ 代表 I 转为 R 的概率(治愈率); κ 代表 I 转为 D 的概率(致死率).

SEIRD 模型如下所示:

$$\left\{ \begin{array}{l} S_{t+1} = S_t - \frac{\beta_1 \times \alpha_t \times I_t \times S_t}{N} - \frac{\beta_2 \times \alpha_t \times E_t \times S_t}{N} \\ E_{t+1} = E_t + \frac{\beta_1 \times \alpha_t \times I_t \times S_t}{N} + \frac{\beta_2 \times \alpha_t \times E_t \times S_t}{N} - \sigma E_t \\ I_{t+1} = I_t + \sigma E_t - \gamma I_t - \kappa I_t \\ R_{t+1} = R_t + \gamma I_t \\ D_{t+1} = D_t + \kappa I_t \end{array} \right. \quad (1)$$

式中: S_t 代表 t 时刻易感人数,对应统计数据中的健康人数, $S = N - E - I - R - D$; E_t 代表 t 时刻接触人数,对应统计数据中的新增疑似人数; I_t 代表 t 时刻感染人数,对应统计数据中的新增确诊人数; R_t 代表 t 时刻治愈人数,对应统计数据中的累计治愈人数; D_t 代表 t 时刻死亡人数,对应统计数据中的累计死亡人数; N 代表总人口数,对应研究区域内的总人口数; t 代表时间; α 代表每人每天接触的人数,与防控政策

有关. 封城隔离前, $\alpha=15$, 封城隔离后, $\alpha=3^{[6]}$; β_1 代表感染人群的感染率; β_2 代表潜伏人群的感染率; σ 代表潜伏人群(E)转为感染人群(I)的概率; γ 代表感染人群(I)转为治愈人群(R)的概率(即治愈率); κ 代表感染人群(I)转为死亡人群(D)的概率(即致死率).

1.2 地理相似性原理与 GEOCA 模块构建

GEOCA 模块将研究区域划分为若干个均质空间. 为了简化计算, GEOCA 状态设置为感染与非感染 2 种状态, 其中感染格网还记录该格网在病毒生命周期范围内(7 d)感染者的累计数量. 格网转换规则最终受全局影响因素(如适宜性、约束性和随机性)和局部影响因素(如邻域)的影响. 通过建立影响疾病传播的空间驱动因子与疾病分布的相关关系, 借助机器学习等相关算法, 学习观测时段内某传染性疾病的不同空间位置发生的概率, 形成传染性疾病的空间适宜性概率地图. 同时通过整合传染性疾病的空间适宜性、约束性概率、邻域概率和随机概率等形成综合概率. GEOCA 模块中如何完成新增患者数量的空间配置十分关键. 本研究采用 2 种策略(蔓延扩散策略和随机迁移扩散策略)完成新增传染性疾病的空间配置. 蔓延扩散策略依据传染邻域概率的高低来完成新增传染者个体的空间配置. 随机迁移扩散策略依据某格网空间综合概率越高, 新增传染性疾病的个体映射到该格网空间的可能性越大的原则完成新增传染者个体的空间配置. 根据地理学第三定律, 地理环境越相似, 在该环境中地理个体的目标特征相近的概率越大. 借助地理相似性原理可判断某区域与感染区域是否具有相似的环境特征, 一旦种子点落入该环境, 则该区域感染该病毒的可能性也比较大. 可通过机器学习方法获得观测期间新冠病毒的空间适宜性(见图 1).

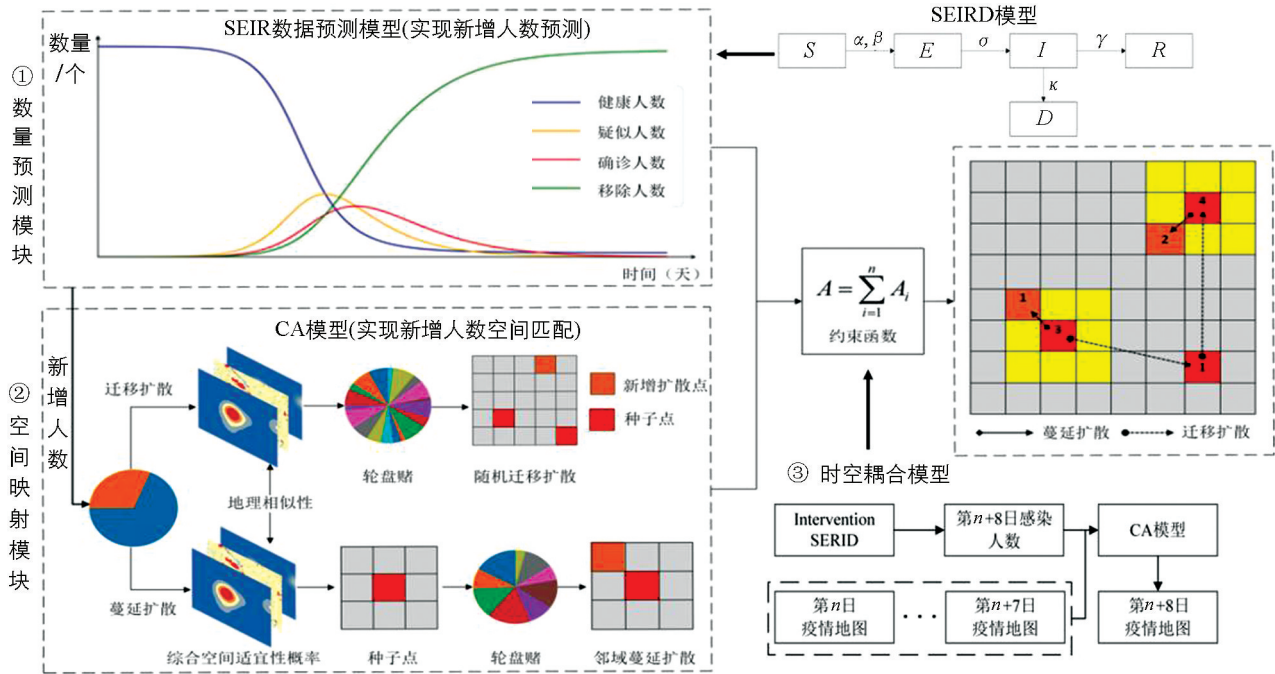


图 1 GEOCA 模型图

CA 模型中每个元胞的总体转换概率取决于疫情发展的适宜性 P_s 、邻域效应 Ω 与限制因素 P_c . 疫情发展的适宜性通过随机森林算法计算得到. 随机森林利用多棵决策树对样本进行训练预测, 可用于回归和分类预测算法中, 属于集成学习的 bagging 算法^[24-25], 如式(2)所示.

$$H(x) = arg_Y \max \sum_{j=1}^n I(h_j(x) = Y) \tag{2}$$

其中, $h(*)$ 表示决策树模型, Y 是单颗决策树的分类结果, $I(*)$ 是分类结果指示函数, n 代表随机森林子树的数量. $arg_Y \max$ 表示 $\sum_{j=1}^n I(h_j(x) = Y)$ 取最大值时 Y 的取值. 元胞 i 被分为第 k 种类别的概率如式(3)所示.

$$P_{i,k}(x) = \frac{\sum_{j=1}^n I(h_j(x_i) = Y_k)}{n} \quad (3)$$

在疫情扩张模拟研究中,非感染元胞 i 在时刻 t 转换为感染元胞的概率可表示为

$$P_{s_{i,k}^t} = \frac{\sum_{j=1}^n I(h_j(x_i) = Inf)}{n} \quad (4)$$

$P_{s_{i,k}^t}$ 表示在时刻 t 非感染元胞 i 的疫情发展适应性, Inf 表示感染的元胞。

邻域效应是 CA 模型的重要组成部分,代表 $n \times n$ 邻域范围内的中心元胞与周边元胞的相互作用。具体表现为中心元胞 $n \times n$ 邻域范围内感染元胞的密度。元胞 i 在时刻 t 的邻域效应可表示为

$$\Omega_i^t = \frac{\sum_{n^2} con(S_i^t = Inf)}{n^2 - 1} \quad (5)$$

其中, S_i^t 表示元胞 i 在时刻 t 的状态, $con()$ 是条件函数,当元胞状态为感染时其值为该元胞的感染者人数,否则值为 0。

限制因素约束了特殊区域向感染者所在栅格的转换。在实际疫情扩张过程中,考虑到地理和政策因素,诸如水体和无人区等不会转变为感染栅格。限制因素可表示为

$$P_{c_i^t} = con(S_i^t = LU_k) \quad (6)$$

其中, LU_k 表示第 k 种土地利用类型;当元胞 i 在时刻 t 满足扩张约束条件时, $con()$ 取值为 1,否则取值为 0。本研究对河流、湖泊、无人区进行限制,规定其不会转换为城市。

因此,元胞 i 在时刻 t 的总体转换概率可表示为

$$P_{i,k}^t = P_{s_{i,k}^t} \times \Omega_i^t \times P_{c_i^t} \quad (7)$$

CA 模型的最终停止条件为

$$A = \sum_{i=1}^n A_i \quad (8)$$

其中, A 代表将新增加的 n 个感染者分配到空间栅格的疫情地图图层。 A_i 代表第 i 个感染者所在元胞的位置。

1.3 时空耦合模块

本研究利用 CA 模型的动态演化特征,将 CA 模型用于新型冠状病毒感染者的空间分布模拟中,充分发掘其空间动态演化规则。SEIRD 模型产生每日感染者的数量,CA 模型分配感染者。时空耦合模块通过构建约束函数将上述 2 种模型耦合,形成 CA 迭代与转换规则,完成模型的运算(见图 1)。

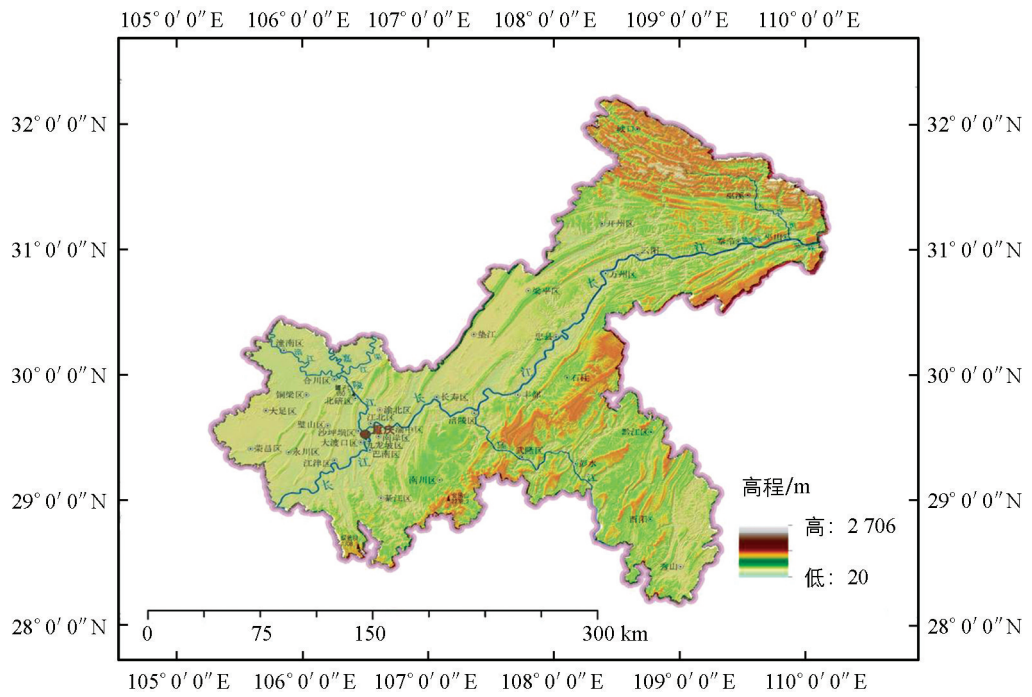
时空耦合模型的最终停止条件可表示为

$$A = \sum_{i=1}^n \sum_{j=1}^m A_{ij} \quad (9)$$

其中, A 代表预测期间累计 n 天时间范围内待分配感染者最终被分配的疫情地图图层; A_{ij} 代表第 i 天第 j 个感染者所在元胞的位置。

2 数据来源和处理

重庆市位于中国西南部,地跨东经 $105^{\circ}11' \sim 110^{\circ}11'$ 、北纬 $28^{\circ}10' \sim 32^{\circ}13'$,是青藏高原与长江中下游平原的过渡地带。东邻湖北、湖南,南靠贵州,西接四川,北连陕西,总面积约 8.24 万平方千米,辖 26 个市,8 个县,4 个自治县。重庆市邻近湖北,与湖北人员来往频繁,受 COVID-19 疫情影响较大。疫情期间,重庆市先后设立新型冠状病毒肺炎治疗定点医院 48 个,故选择重庆作为研究区域。



图片来源于网络基础地理信息中心标准地图服务网站“中国分省地图—重庆市”，审图号为 GS(2019)3333。

图 2 研究区域示意图

2.1 数据来源

COVID-19 疫情数据主要来自国家卫生健康委员会官网(<http://www.nhc.gov.cn/>)、世界卫生组织(<https://www.who.int/westernpacific>)、MIDAS network 等官方公布的统计数据。从 MIDAS network 的统计数据可看出，重庆市 COVID-19 疫情新增确诊者主要分布在 2020 年 1 月 21 日至 2 月 29 日，故选取该时间段数据对重庆市 COVID-19 疫情进行分析模拟预测。本研究使用的 COVID-19 数据包括统计数据 and 感染者轨迹数据，统计数据涉及新增疑似、新增确诊、累计死亡、累计治愈等，数据来自美国匹兹堡大学公共卫生研究生院公共卫生动力学实验室(Public Health Dynamics Lab, PHDL)。感染者轨迹数据通过华龙网获取。

考虑到 COVID-19 的传播与人类活动密切相关，本文选取自然环境和社会经济 2 类数据作为影响其传播的主要驱动因子(见表 1)。

2.2 数据处理

将表中的异构数据统一研究区域范围、空间坐标系和分辨率。9 类 POI 数据采用网络爬虫软件爬取后，进行核密度分析获得核密度分布图层。由于 COVID-19 感染者的分布与用地类型有一定的相关性，水域和无人区完全不可能出现 COVID-19 感染者，故本研究将用地类型重新分为 3 类，其中未利用土地和水域等无人区分为 1 类，耕地、林地、草地分为 1 类，城乡居民用地分为 1 类。无人区作为模型构建过程中的约束概率层。高程数据来自 ASTERGDEM，分辨率为 30m，经过掩膜提取得到重庆市的高程图层；借助 ARC-GIS 经过坡度计算得到坡度图层。距河流、铁路、高速公路、主干道、二级道路的距离数据来源于开放街道地图(Open Street Map, OSM)，在 ARCGIS 中经过欧氏距离分析得到各交通线路的欧氏距离图层。重庆市 COVID-19 感染者活动轨迹数据来源于华龙网，首先将 COVID-19 感染者文本信息中涉及的小区、街道、村、乡镇、县等位置数据通过网络爬虫软件在高德地图中爬取其对应的带地理坐标的点数据，然后进行坐标转换(GCJ-02→WGS-84)，从而得到不同时间 COVID-19 感染者的位置分布图层。

表 1 空间模型驱动因子的选择和数据说明

数据类别	驱动因子	数据名称	数据年份	数据来源	数据类型	数据处理说明
自然环境	高程	Dem	2011	ASTER GDEM	栅格	无
	坡度	Slope	2011			坡度
	用地类型	Landuse	2020	中国科学院资源 环境与数据中心		无
	河流	River	2020	OSM (Open Street Map)	矢量	欧氏距离
社会经济	医院	Hospital	2020	高德地图	点	核密度分析
	学校	School			矢量	欧氏距离
	宾馆	Hotel			栅格	重分类
	购物	Shop			点	核密度分析
	餐饮	Restaurant			矢量	欧氏距离
	风景名胜	Spot			栅格	重分类
	娱乐场所	Entertainment			点	核密度分析
	银行	Bank			矢量	欧氏距离
	车站站点	Station			栅格	重分类
	铁路	Railway		OSM	点	核密度分析
	高速公路	Motorway			矢量	欧氏距离
	主干道	Primary way			栅格	重分类
	二级道路	Second way			点	核密度分析
	人口密度	Pop density		World pop	矢量	欧氏距离
感染者轨迹数据	感染者	Track	2020.1.21— 2020.2.29	华龙网	点	提取轨迹点并 进行核密度分析

3 模型实施

3.1 整合模型构建

3.1.1 SEIRD 模型构建

采集重庆市 2020.1.21—2020.2.29 的 COVID-19 感染者统计数据进行探索性分析(图 3a),将 COVID-19 感染者统计数据和采用 SEIRD 模型模拟的结果进行对比(图 3b),从数据走势可看出,实际数据曲线趋势与 SEIRD 曲线趋势相同,表明 SEIRD 模型能够有效反映 COVID-19 感染者数量的发展趋势. COVID-19 存活时间为 2~14 d,将平均潜伏期定为符合报道的 7 d,死亡率设置为 3%^[6]. 在采取严格的防控措施下,感染者数量峰值显著降低,疫情高峰更早地到来.

图 3b 中,蓝色曲线反映实际感染者数量随天数的变化情况,红色曲线代表模拟的感染者数量在有干预的情况下随天数的变化情况. 采用 SEIRD 模型,对重新调优后的模拟感染者曲线进行格网搜索,寻找最优参数. 实验得出,在 β_1 为 0.02, β_2 为 0.03, σ 为 0.2, γ 为 0.29 时,这条曲线拟合效果最好.

3.1.2 融入不同扩散策略的 GEOCA 模型构建

1) 基于地理相似性原理的概率图层生成

COVID-19 疫情的传播与人类活动密切相关. 本文选用随机森林(利用多棵树对样本进行训练并预测的一种分类器)算法,以重庆市为研究区域,将高程、坡度、用地类型、河流作为自然环境驱动因子,将医院、学校、宾馆、购物、餐饮、风景名胜、娱乐场所、银行、车站站点、铁路、高速公路、主干道、二级道路、人口密度作为社会经济驱动因子,将 2020 年 1 月 21 日到 2020 年 2 月 29 日整个时间范围内疫情感染者的

空间位置作为标签,进行回归,得到适宜性概率层(P_s),如图 1 综合空间适宜性概率所示.

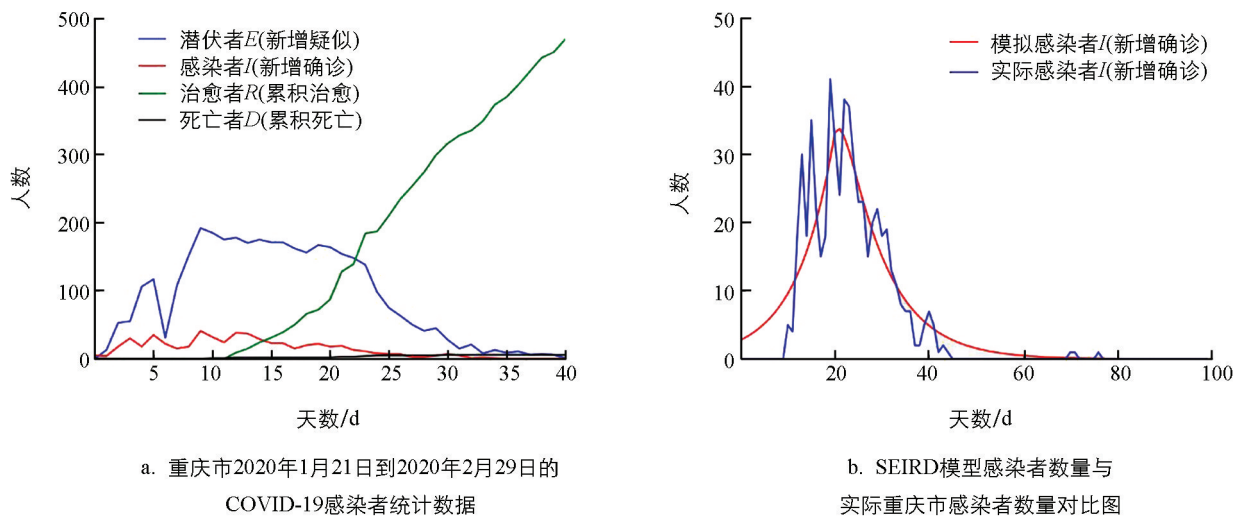


图 3 COVID-19 统计分析与 SEIRD 感染者数量模拟

随机森林可利用袋外数据计算驱动因子的重要性. 各个驱动因子中银行核密度对感染者分布影响最大, 其次各种兴趣点以及人口是影响疫情感染者分布的重要因素, 说明社会经济因素对疫情感染者分布具有主要影响力. 另外, 高程坡度、用地类型等对感染者分布影响较弱, 原因在于重庆以丘陵和山地为主, 坡度面积较大, 城市基本在复杂的地形地貌上进行扩张, 因此感染者的空间分布对地理环境相对不敏感.

2) 空间扩散策略的实现

如何将预测的新增感染者映射到某一格网空间, 实现从数量预测到空间位置预测的转变, 是该时空耦合模型构建的关键. 空间扩散建模需模拟传染病的 2 种空间扩散机制: 一种是输入型病例, 另一种是在已有病例基础上的接触蔓延(“生根发芽”). 本研究采用 2 种策略, 一种是蔓延扩散策略, 另一种是随机迁移扩散策略.

① 蔓延扩散策略

在某一格网位置一旦发现有 COVID-19 感染者, 病毒便依托已有感染者沿着交际链、交通链按一定概率在“生物岛”内近距离接触扩散. 它的传播与人际交往、防护措施和病毒防范意识等密切相关, 以聚会或亲朋交往等近距离接触传播为主. 其观测区域是已有 COVID-19 感染者分布格网的近邻区域. 鉴于蔓延扩散以“在地型”病例或已“落地生根”的输入型病例近距离蔓延为特征, 因此预测的新增感染者位置落在何处同时受传染周期内已有的 COVID-19 感染者的位置分布和邻域概率影响. 本研究拟将已有 COVID-19 感染者栅格作为候选点, 使得在该点邻域范围内, 综合概率图层最大者, 新增感染者落在该点的可能性最大, 即已有的 COVID-19 感染者栅格邻域范围内概率大者优先配置位置, 直到新增感染者全部分配完毕, 完成蔓延扩散策略(见图 1). 随机迁移扩散和蔓延扩散在时间上将同时发生. 如 COVID-19 患者发生位置移动, 则扩散方式由蔓延扩散转变为随机迁移扩散.

② 随机迁移扩散策略

因 COVID-19 接触者从感染到爆发具有潜伏期, 且在潜伏期已经具有传染性, 使得病毒从源头通过中间宿主输入到“生物岛”, 在病毒爆发直至采取隔离措施之前, 接触者并不知道该病毒是什么, 来自何方, 去往何处. 因此这里的随机扩散是指在对病毒位置信息双盲情形下, 病毒传播宛如“随风飘散的蒲公英”, 蒲公英种子点到底花落谁家, 一切随机(病毒的初始传播具有很强的随机性和危害性). 随机迁移扩散主要表现为通过交通链在“生物岛”间的远距离输入性传播, 它类似蛙跳在不同的“生物岛”间跳跃性迁移扩散. 随机扩散在传染病早期表现比较明显, 即使后期采取核酸检测、禁足限行等措施, 大大减少了传播的强度, 但也常有“漏网之鱼”, 如时有发生输入性病例就是随机迁移扩散的表现.

由于远距离流通形式(物流、人流)的存在, 增强了种子点在“生物岛”之间或岛内不确定性位置上传播的可能性, 使得采取随机迁移扩散策略成为必要. 随机迁移扩散策略的观测区域是整个研究区域的各“生

物岛”。鉴于预测的新增感染者个体可能通过输入的方式以不同的概率进入“生物岛”的任何格网空间,因此本研究借助轮盘赌的思想,在 COVID-19 感染空间适宜性图层上叠加随机因子,实现新增感染者在空间上的随机迁移扩散策略(见图 1)。

当然,随着时间的推移,预测的感染者数量与位置将不断得到纠正,通过元胞的不断更新迭代,得到若干日的疫情分布图层。时间越长学习的知识越充分,预测能力越强。为了实现一日感染者空间分布图层,将适宜性概率层与用地类型重分类后得到的约束性概率层(P_{con})相乘得到全局概率层(P_{glo}),根据前一日疫情地图图层,按栅格的概率选取一个感染者所在的栅格作为种子点(i, j),即可得到该种子点在全局概率层中的综合概率,由 $ROI(P_{ROI})$ 相乘邻域概率层(P_{nei})得到,按照栅格概率随机选取感染栅格,重复上述过程,待当日感染者全部分配完成后,即完成一次 CA 迭代。COVID-19 疫情的传播表现为外界输入和继发性蔓延机制,前者表现为随机突变点,发生在疫情早期,后者表现为从本地种子点蔓延到周边栅格,发生在疫情中后期。

3.1.3 时空耦合模型构建

将 SEIRD 模型与 GEOCA 模型相结合,在封城有排查,平均潜伏期为 7 d 的情景下,利用前 7 日的感染者图层结合第 8 日感染者的数目,产生第 8 日的疫情地图图层,依次迭代,即可得到整个时间段的疫情地图图层,从而完成时空耦合模型的构建。

3.2 模型校正与检验

3.2.1 对比方案设计

- 本研究基于以下假设:
- 1) 使用不同的图层(感染者核密度图层、实际人数栅格图层以及有无感染者的 01 栅格图层)作为邻域分析图层,将对结果产生影响,且差异有统计学意义。
 - 2) 模拟过程中,是否采取干预措施以及采取不同的干预措施将对结果产生影响且差异有统计学意义。
 - 3) 将新闻报道中的 2~14 d 的潜伏期设为 7 d。

选取 2020 年 1 月 21 日到 2020 年 2 月 29 日的 COVID-19 感染者统计数据 and 点数据,针对上述假设设计了 5 种对比方案,并分别与实际疫情地图图层进行对比分析,其具体方案为:

方案一(封城有排查 kernel): 采取封城有排查的干预措施,用感染者核密度图层作为邻域分析图层,简称 kernel,采用 SEIRD-GEOCA 模型模拟感染者时空分布。

方案二(封城有排查 pop): 采取封城有排查的干预措施,用实际人数栅格分布图层作为邻域分析图层,简称 pop,采用 SEIRD-GEOCA 模型模拟感染者时空分布。

方案三(封城有排查 01): 采取封城有排查的干预措施,用感染者 01 栅格分布图层作为邻域分析图层,简称 01,采用 SEIRD-GEOCA 模型模拟感染者时空分布。

方案四(封城无排查): 采取封城无排查的干预措施,用实际人数栅格图层作为邻域分析图层,采用 SEIRD-GEOCA 模型模拟感染者时空分布。

方案五(无干预): 不采取干预措施,用实际人数栅格图层作为邻域分析图层,采用 SEIRD-GEOCA 模型模拟感染者时空分布。

这里的有排查是指针对疫情发展状况,积极采取必要的隔离(限行、禁足、集中隔离)、病例调查、红外温度计测温 and 核酸检测等措施。在模型中表现为通过调整种子点的生命周期,使得种子点在病毒潜伏期内能感染其他人,但超过 7 d,由于对种子点采取了积极的隔离和治疗等措施,从而使得种子点自动消亡。无排查是指针对疫情发展状况,未及时采取必要的排查措施,在模型中表现为种子点在病毒潜伏期内不会失活,种子点将会一直传染他人。封城措施是指针对疫情发展状况,政府采取了果断的城市封锁政策,防止疫情在城外快速扩张。在模型中表现为让种子点每天接触的人数从无任何防护措施的情况下接触 15 人/日,调整到 3 人/日。通过调控种子点是否消亡区分有无排查,通过调整 SEIRD 模型接触人数来区分是否封城。

3.2.2 对比方案设计

分别从栅格尺度和区县尺度,利用均方误差(Mean Square Error, MSE)、平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Square Error, RMSE)对模型的精度进行检验。

均方误差是各数据偏离真实值差值平方和的平均数，也就是误差平方和的平均数；平均绝对误差是指观测值与真实值的误差绝对值的平均值；均方根误差是指观测值与真实值偏差的平方和与观测次数 n 比值的平方根。

4 结果分析

选取重庆市 2020 年 1 月 21 日到 2020 年 2 月 29 日的感染者为研究对象，通过 SEIRD-GEOCA 模型得到不同情形下感染者的空间分布模拟结果，同时将不同情形下的模拟结果和真实环境下的 COVID-19 感染者时空分布进行对比(表 2)。同时，采用核密度制图法直观展示模拟结果，并与观测期间真实环境下的 COVID-19 感染者时空分布进行对比(见图 4)。

表 2 不同情景模拟结果与真实情况的对比

		方案一	方案二	方案三	方案四	方案五
		封城有排查 kernel	封城有排查 pop	封城有排查 01	封城无排查	无干预
栅格尺度	MSE	0.001 188	0.001 207	0.001 21	0.001 253	0.358 328
	MAE	0.002 788	0.002 771	0.002 781	0.002 853	0.047 956
	RMSE	0.034 47	0.034 742	0.034 788	0.035 391	0.598 605
区县尺度	MSE	20.256 74	20.622 45	21.401 705	18.674 855	170 372.627
	MAE	3.474	3.376	3.383	3.446	243.475
	RMSE	4.500 749	4.541 195	4.626 198	4.321 441	412.762 192

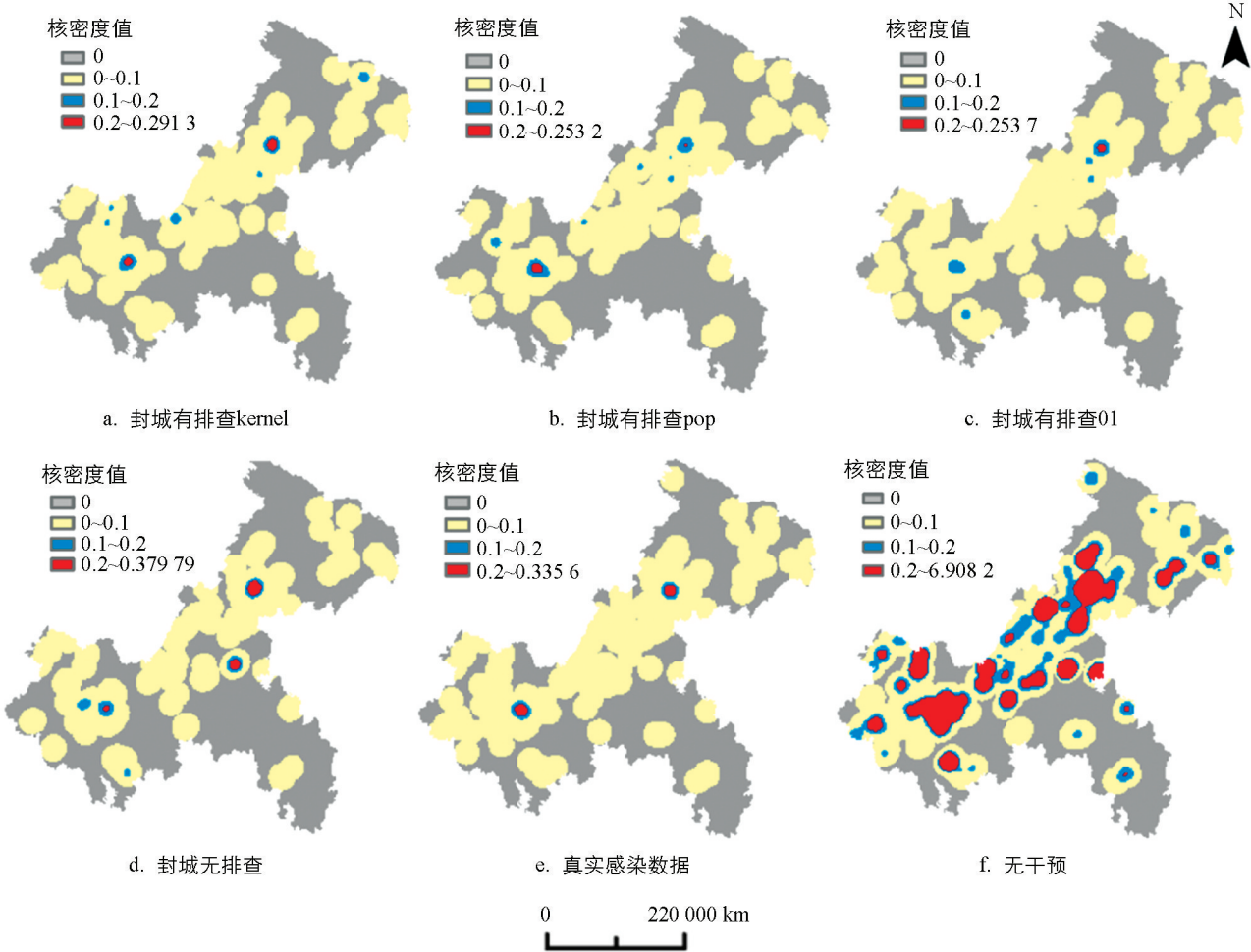


图 4 不同方案下 COVID-19 感染者空间分布模拟结果对比图

方案一、二、三表明, MSE, MAE, RMSE 三者差距并不显著, 且上述三类误差的最大偏离度为 5.52%。但相较于实际人数栅格图层和有无感染者的 01 栅格图层, 选取感染者核密度图层作为邻域分析图层能获得更好的可视化效果(方案一)(图 4a)。根据表 2 各项指标可看出, 方案一中各精度项指标比其他方案稍优。方案二的扩散中心围绕着以主城区为核心的渝西南地区, 与实际情况稍有偏离(图 4b), 方案三的扩散中心则围绕着以万州区为核心的渝东北地区, 与实际情况偏离程度最大(图 4c)。因此, 方案一更显合理。方案一、四对比可看出, 在封城的前提下, 相较于有排查情形, 无排查情形呈现出团簇现象, 表明封城无排查会产生超级传播者, 在人口较为集中的区域传播。方案一、五对比可看出, 有无干预对 COVID-19 疫情空间扩散影响显著, 防控措施可有效地遏制 COVID-19 疫情大面积传播。重庆重点防控区域为以万州区为核心的渝东北地区 and 以主城区为核心的渝西南地区。从数量来看, 无干预比有干预在观测期间里数量上要多很多倍, 无干预使得局面失控, 整个重庆市形成了以各区县为中心的众多热点区域。

5 结 论

本研究赋予 CA 模型生物地理学意义, 构建了一种基于 SEIRD-GEOCA 的 COVID-19 疫情时空分布模型, 在模型中融入蔓延扩散和随机迁移扩散策略, 设计了 5 种方案用于验证干预措施, 同时验证邻域分析图层的不同是否对传染病时空模拟结果产生影响。

结果显示: ① 采用 SEIRD-GEOCA 时空耦合模型, 能很好地模拟 COVID-19 疫情的时空扩散; ② 是否采取封城措施在空间分布上呈现出明显的爆发式扩散和局部区域扩散的差别; ③ 实施不同的干预措施呈现出了不同的时空扩散效应。封城措施下, 有排查能有效避免产生超级传播者。采取干预措施, 感染者数量大幅度降低, 空间上的扩散被约束在很小的区域。

传统的 SEIR 模型只提供了数量预测, 并不能给出空间扩散上的指导。SEIRD-GEOCA 模型在历史观测数据基础上, 借助 SEIRD 模型预测新增确诊病例, 采用随机种子点与已有的概率地图图层结合来预测未来空间的扩散情况, 研究范围内精度较高, 可以较好地满足 COVID-19 疫情的时空预测。

参考文献:

- [1] 周成虎, 裴韬, 杜云艳, 等. 新冠肺炎疫情大数据分析与区域防控政策建议 [J]. 中国科学院院刊, 2020, 35(2): 200-203.
- [2] GRASSLY N C, FRASER C. Mathematical Models of Infectious Disease Transmission [J]. Nature Reviews Microbiology, 2008, 6(6): 477-487.
- [3] RANA P S, SHARMA N. Mathematical Modeling and Stability Analysis of a SI Type Model for HIV/AIDS [J]. Journal of Interdisciplinary Mathematics, 2020, 23(1): 257-273.
- [4] KERMACK W O, MCKENDRICK A G. Contributions to the Mathematical Theory of Epidemics-III. Further Studies of the Problem of Endemicity [J]. Society for Mathematical Biology, 1991, 53(1/2): 89-118.
- [5] BAI Z G. Threshold Dynamics of a Periodic SIR Model with Delay in an Infected Compartment [J]. Mathematical Biosciences and Engineering, 2015, 12(3): 555-564.
- [6] YANG Z, ZENG Z, WANG K, et al. Modified SEIR and AI Prediction of the Epidemics Trend of COVID-19 in China under Public Health Interventions [J]. Journal of Thoracic Disease, 2020, 12(3): 165-174.
- [7] 曹盛力, 冯沛华, 时朋朋. 修正 SEIR 传染病动力学模型应用于湖北省 2019 冠状病毒病(COVID-19)疫情预测和评估 [J]. 浙江大学学报(医学版), 2020, 49(2): 178-184.
- [8] 洪彬, 陈锦秀, 王连生, 等. 基于 SEIR-LSTM 混合模型的新型冠状病毒肺炎传播趋势分析与预测 [J]. 厦门大学学报(自然科学版), 2020, 59(6): 1034-1040.
- [9] 刘涛, 黎夏, 刘萌伟. 传染病时空传播的多智能体仿真研究 [J]. 系统仿真学报, 2009, 21(18): 5874-5877.
- [10] GHARAKHANLOU N M, HOOSHANGI N. Spatio-Temporal Simulation of the Novel Coronavirus (COVID-19) Out-

- break Using the Agent-Based Modeling Approach (Case Study: Urmia, Iran) [J]. *Informatics in Medicine Unlocked*, 2020, 20: 100403.
- [11] KOO J R, COOK A R, PARK M, et al. Interventions to Mitigate Early Spread of SARS-CoV-2 in Singapore: a Modeling Study [J]. *The Lancet Infectious Diseases*, 2020, 20(6): 678-688.
- [12] 周成虎, 孙战利, 谢一春. 地理元胞自动机研究 [M]. 北京: 科学出版社, 1999.
- [13] BATTY M, XIE Y. From Cells to Cities [J]. *Environment and Planning B: Planning and Design*, 1994, 21(7): S31-S48.
- [14] 刘明皓, 尚尉, 张奇瑞, 等. GA-logistic 回归参数优化与城市用地动态变化模拟——以成渝经济带为例 [J]. *地理与地理信息科学*, 2018, 34(1): 12-17.
- [15] LIU X P, LIANG X, LI X, et al. A Future Land Use Simulation Model (FLUS) for Simulating Multiple Land Use Scenarios by Coupling Human and Natural Effects [J]. *Landscape and Urban Planning*, 2017, 168: 94-116.
- [16] WANG C L, JIANG Q O, SHAO Y Q, et al. Ecological Environment Assessment Based on Land Use Simulation: a Case Study in the Heihe River Basin [J]. *Science of the Total Environment*, 2019, 697: 133928.
- [17] 郑三强, 韩晓卓. 多因素制约下的 SIR 传染病模型的元胞自动机仿真模拟研究 [J]. *广东工业大学学报*, 2018, 35(5): 51-59.
- [18] 陈长坤, 童蕴贺. 基于元胞自动机的传染病跨区域传播模型研究 [J]. *武汉理工大学学报(信息与管理工程版)*, 2018, 40(4): 359-363, 382.
- [19] ALMUZAKKI M Z, NURAINI N. A Vaccination Strategy to SEIR-CA Model [C] // *AIP Conference Proceedings* 1723. Bandung, Indonesia: AIP Publishing LLC, 2016: 030002.
- [20] Cissé B, YACOUBI S E, TRIDANE A. Impact of Neighborhood Structure on Epidemic Spreading by Means of Cellular Automata Approach [J]. *Procedia Computer Science*, 2013, 18: 2603-2606.
- [21] MEDREK M, PASTUSZAK Z. Numerical Simulation of the Novel Coronavirus Spreading [J]. *Expert Systems With Applications*, 2021, 166: 114109.
- [22] GWIZDAA T. Viral Disease Spreading in Grouped Population [J]. *Computer Methods and Programs in Biomedicine*, 2020, 197: 105715.
- [23] 宗跃光, 王莉, 曲秀丽. 基于蒙特卡罗模拟法的北京地区非典时空变化特征 [J]. *地理研究*, 2004, 23(6): 815-824.
- [24] 曹志冬, 王劲峰, 高一鸽, 等. 广州 SARS 流行过程的空间模式与分异特征 [J]. *地理研究*, 2008, 27(5): 1139-1149, 1226.
- [25] AZEVEDO L, PEREIRA M J, RIBEIRO M C, et al. Geostatistical COVID-19 Infection Risk Maps for Portugal [J]. *International Journal of HealthGeographics*, 2020, 19(1): 25.
- [26] 曹培明, 李晓旭, 严晓峰, 等. 重庆市 223 例新型冠状病毒肺炎病例的回顾性流行病学分析 [J]. *西南大学学报(自然科学版)*, 2020, 42(3): 10-16.
- [27] 童昀, 马勇, 刘海猛. COVID-19 疫情对中国城市人口迁徙的短期影响及城市恢复力评价 [J]. *地理学报*, 2020, 75(11): 2505-2520.
- [28] 王卷乐, 张敏, 韩雪华, 等. COVID-19 疫情防控中的中国公众舆情时空演变特征 [J]. *地理学报*, 2020, 75(11): 2490-2504.
- [29] 叶玉瑶, 王长建, 张虹鸥, 等. 基于人口流动的广东省 COVID-19 疫情风险时空分析 [J]. *地理学报*, 2020, 75(11): 2521-2534.
- [30] 李钢, 王皎贝, 徐婷婷, 等. 中国 COVID-19 疫情时空演化与综合防控 [J]. *地理学报*, 2020, 75(11): 2475-2489.